

Sketch2Data: Recovering Data from Hand-Drawn Infographics

Supplemental Materials

1. Statistics of the collected 10 designs in our benchmark

Table 3 summarizes the statistics of each design, including the number of parameters in the parametric glyph template, visual attribute(s) associated with each parameter, the number of values of each parameter, the total number of parameter value combinations, and number of glyphs in the design. On average, each design includes 3.6 parameters, 313 possible combinations of parameter values, and 73 glyphs. This reflects significant complexity and variation in the glyphs. For example, the *Lollipop* and *Triangle* designs have 5 parameters with a total of 2,160 and 320 possible combinations, respectively. Even simpler design, such as *Sound* with only 2 parameters, offers 110 combinations. Figure 2 and 3 visualize 100 variations of glyphs generated from our *Triangle* and *Sound* parametric glyph templates, respectively. These visualizations emphasize the similarities and differences among glyphs. Despite their often small size to fit on paper, people draw these glyphs with rich details, contributing to the overall design complexity.

Figure 4 provides a visual summary of all the parametric glyph templates we used throughout our evaluation.

2. Dataset details

Table 1 summarizes the details of localization (Loc.) and parameter estimation (Param. Est) datasets. Our localization dataset contains 2,000 images for each design. The parameter estimation dataset includes 10,000 glyph images if the total number of values in its parametric template is equal or less than 20; otherwise, we generate 20,000 glyph images for the design. Both datasets are split into 70% for training, 10% for validation, and 20% for testing.

Table 1: Dataset details

| Design | Number of image | | Image size (width × height) | |
|-----------|-----------------|------------|-----------------------------|------------|
| | Loc. | Param. Est | Loc. | Param. Est |
| book | 2k | 10k | 1169 × 827 | 145 × 80 |
| boyfriend | 2k | 10k | 1000 × 660 | 65 × 100 |
| dog | 2k | 10k | 1150 × 785 | 100 × 65 |
| leaf | 2k | 10k | 1169 × 827 | 120 × 120 |
| lollipop | 2k | 20k | 800 × 525 | 100 × 100 |
| smell | 2k | 10k | 700 × 500 | 100 × 80 |
| sound | 2k | 20k | 2000 × 1680 | 80 × 90 |
| thoughts | 2k | 10k | 1000 × 700 | 130 × 130 |
| tree | 2k | 10k | 1000 × 1000 | 150 × 150 |
| triangle | 2k | 10k | 800 × 650 | 80 × 80 |

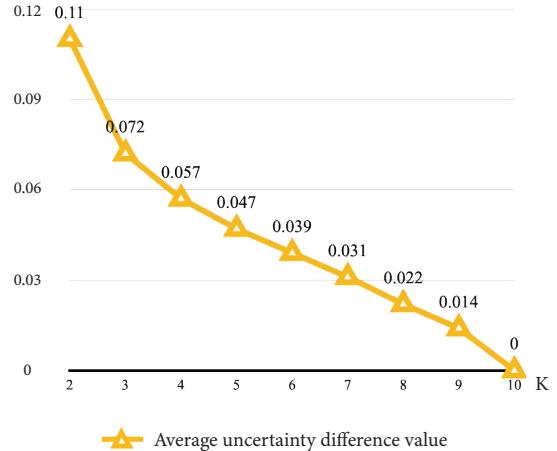


Figure 1: Average difference between the uncertainty estimated with K networks compared to the reference obtained with 10 networks. While we used 10 networks to obtain a precise estimate for our experiments, a good estimate of uncertainty can already be achieved with 5 networks.

3. Influence of K networks in uncertainty estimation

In Figure 1, we visualize the average uncertainty difference across the 10 collected designs, between an ensemble of K networks used for uncertainty estimation compared to the reference value $K = 10$, for K ranging from 2 to 10. The results show a rapid decrease in un-

certainty differences between $K = 2$ to 4. By $K = 5$, the difference is already small (0.047). To ensure robustness, we chose $K = 10$ for our experiments.

4. Detection and parameter estimation results

Table 4 summarizes the detection and parameter estimation accuracy over the 3 designs drawn by each participant. Table 5 and Table 6 details these results further. Table 7 reports the glyph detection results for three designs drawn with different tools, and Table 8 provides the results of the corresponding parameter estimation. Figure 7 to 20 show the detected, reconstructed, and ground-truth glyphs for each of the 10 designs in our benchmark, for all drawings produced by participants of our study, and for the drawings created with different tools.

5. Data augmentation

We conducted an ablation study to analyze the impact of data augmentation on parameter estimation. Table 2 presents the accuracy of each parameter across the 10 collected designs. It shows that without data augmentation, the average accuracy of parameter estimation drops from 0.94 to 0.83, highlighting the effectiveness of our data augmentation.

6. Glyph localization by open-set object detector

We have experimented with T-REX (Jiang et al., 2024), a state-of-the-art pre-trained open-set object detector, to localize glyphs in hand-drawn infographics. We used the paper’s online demo that offers an interactive visual prompt workflow. It allows users to specify the object to be detected by providing visual examples, either by drawing boxes or marking points on the image. Figure 21 and 22 provides results of this experiment on *Boyfriend* and *Lollipop* design, which reveals that while the open-set detector can locate some of the glyphs based on a few examples, providing more examples does not always improve accuracy. In particular, the detector fails to detect diverse variants of the glyph, and its bounding boxes sometimes miss small details at the border of the glyphs.

References

- Q. Jiang, F. Li, Z. Zeng, T. Ren, S. Liu, L. Zhang, T-rex2: Towards generic object detection via text-visual prompt synergy, ECCV (2024).

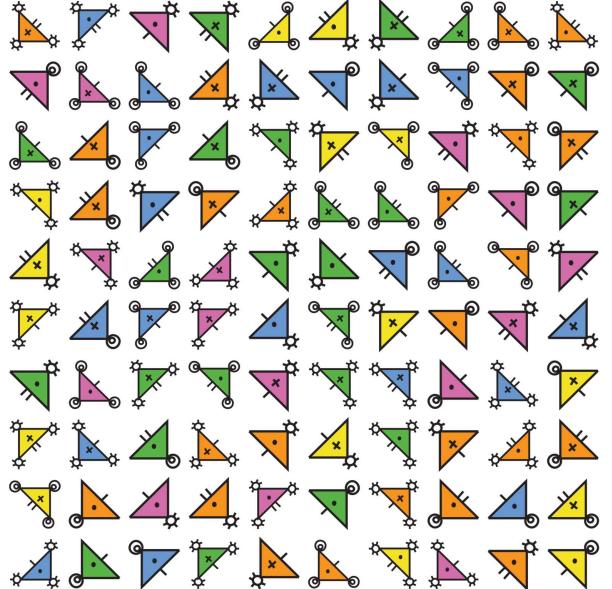


Figure 2: 100 possible variations generated from our *Triangle* parametric glyph template.

Table 2: Estimation accuracy for each parameter (P) without/with data augmentation. We show if accuracy increases \uparrow or decreases \downarrow with data augmentation.

| Design | P1 | P2 | P3 | P4 | P5 | Avg. P |
|-----------|----------------------|------------------------|----------------------|----------------------|----------------------|----------------------|
| Book | 0.39/1.00 \uparrow | 0.32/1.00 \uparrow | 0.64/1.00 \uparrow | 0.68/1.00 \uparrow | | 0.51/1.00 \uparrow |
| Boyfriend | 0.90/1.00 \uparrow | 0.88/1.00 \uparrow | 0.72/0.97 \uparrow | 0.94/1.00 \uparrow | | 0.86/0.99 \uparrow |
| Dog | 1.00/1.00 | 0.92/0.94 \uparrow | 1.00/1.00 | | | 0.97/0.98 \uparrow |
| Leaf | 1.00/1.00 | 1.00/1.00 | 0.96/1.00 \uparrow | | | 0.99/1.00 \uparrow |
| Lollipop | 0.57/0.71 \uparrow | 0.80/0.71 \downarrow | 0.63/0.74 \uparrow | 0.26/0.46 \uparrow | 0.97/1.00 \uparrow | 0.65/0.73 \uparrow |
| Smell | 1.00/1.00 | 1.00/1.00 | 1.00/1.00 | 0.98/1.00 \uparrow | | 1.00/1.00 \uparrow |
| Sound | 0.83/0.92 \uparrow | 0.65/0.88 \uparrow | | | | 0.74/0.90 \uparrow |
| Thoughts | 0.98/1.00 \uparrow | 1.00/0.98 \downarrow | 0.58/0.91 \uparrow | | | 0.85/0.96 \uparrow |
| Tree | 0.65/0.91 \uparrow | 1.00/1.00 | 0.91/1.00 \uparrow | | | 0.85/0.97 \uparrow |
| Triangle | 1.00/1.00 | 0.68/0.70 \uparrow | 0.58/0.82 \uparrow | 0.93/0.93 | 1.00/1.00 | 0.84/0.89 \uparrow |
| Avg. | — | — | — | — | — | 0.83/0.94 \uparrow |

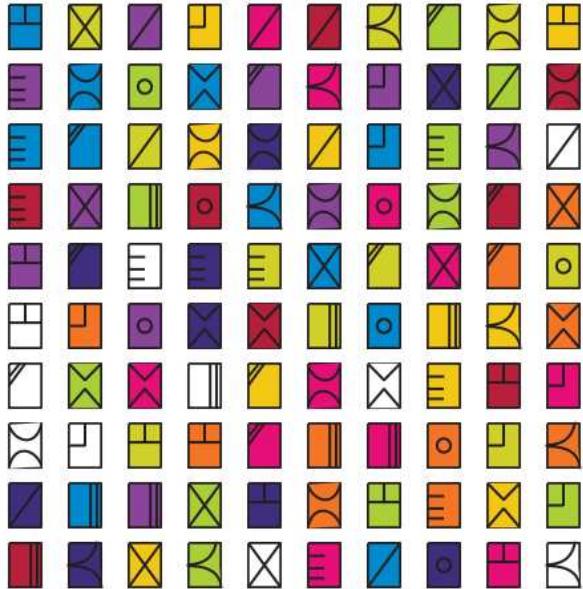


Figure 3: 100 possible variations generated from our *Sound* parametric glyph template.

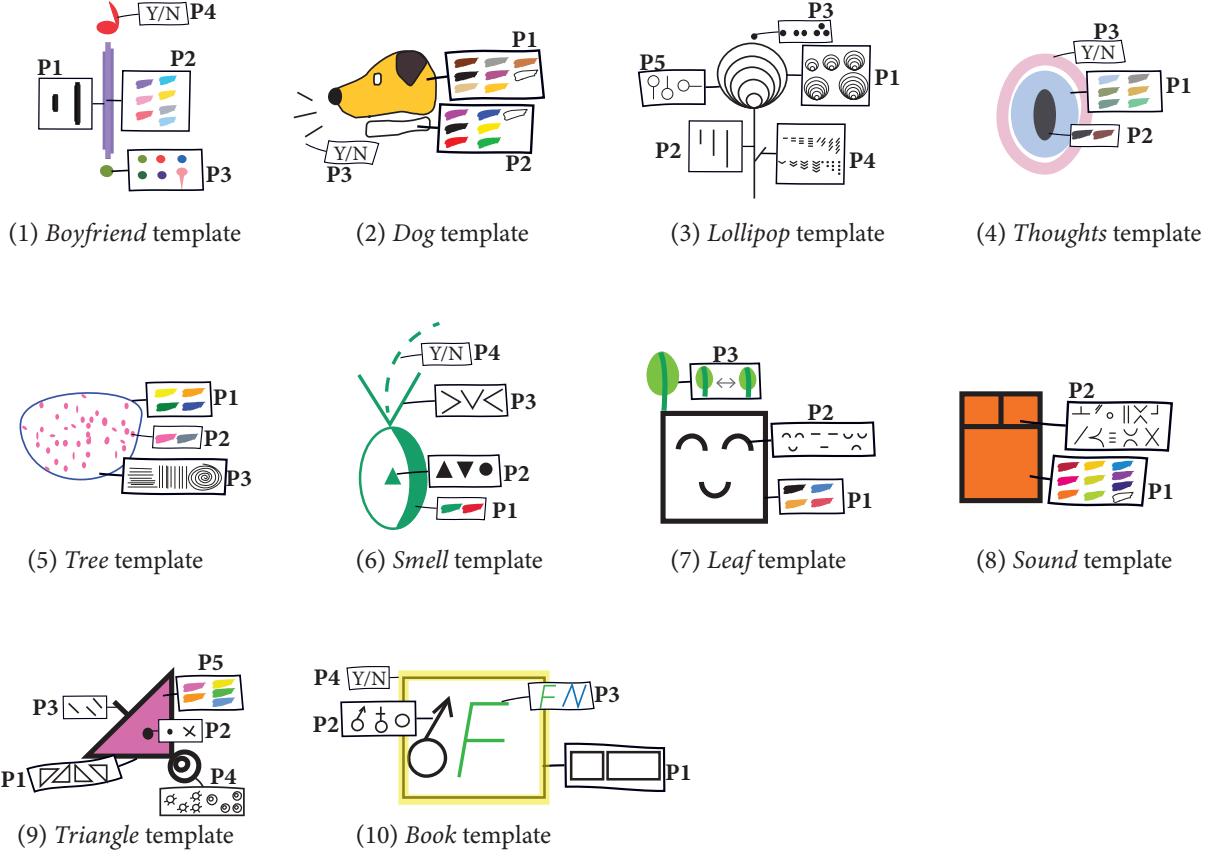


Figure 4: The parametric glyph templates for the 10 designs in our benchmark.

Table 3: Statistics of the 10 designs collected in our benchmark. For each design, we report the number of parameters, visual attribute(s) associated with each parameter, the number of values of each parameter, the total number of parameter value combinations, and number of glyphs in the design. ‘Orient.’, ‘Posit.’ and ‘Cl-Sp’ represent Orientation, Position and the combination of Color and Shape, respectively, as visual attributes.

| Design | Number of parameters | (Visual attribute, Number of values) in each parameter | | | | | Total number of combinations | Number of glyphs |
|-----------|----------------------|--------------------------------------------------------|-------------|--------------|-------------|--------------|------------------------------|------------------|
| | | P1 | P2 | P3 | P4 | P5 | | |
| Book | 4 | (Size, 3) | (Shape, 3) | (Cl-Sp, 2) | 2 | — | 36 | 29 |
| Boyfriend | 4 | (Size, 2) | (Color, 7) | (Cl-Sp, 7) | (Shape, 2) | — | 196 | 94 |
| Dog | 3 | (Color, 8) | (Color, 7) | (Shape, 2) | — | — | 112 | 63 |
| Leaf | 3 | (Color, 4) | (Shape, 3) | (Posit., 2) | — | — | 24 | 23 |
| Lollipop | 5 | (Size, 5) | (Size, 3) | (Shape, 3) | (Shape, 16) | (Orient., 3) | 2160 | 35 |
| Smell | 4 | (Color, 2) | (Shape, 4) | (Posit., 4) | (Posit., 4) | — | 128 | 45 |
| Sound | 2 | (Color, 11) | (Shape, 10) | — | — | — | 110 | 315 |
| Thoughts | 3 | (Color, 6) | (Color, 2) | (Shape, 2) | — | — | 24 | 45 |
| Tree | 3 | (Color, 4) | (Color, 3) | (Texture, 2) | — | — | 24 | 23 |
| Triangle | 5 | (Orient., 4) | (Shape, 2) | (Shape, 2) | (Shape, 4) | (Color, 5) | 320 | 57 |
| Avg. | 3.6 | — | — | — | — | — | 313 | 73 |

Table 4: Evaluation of the glyph detector (Precis. and Recall), parameter estimation (Para Est.) and uncertainty (Unc.) for each participant (PID) over the three designs used in our study (Book, Leaf and Smell).

| PID | Book | | | | Leaf | | | | Smell | | | | Avg. | | | |
|------|---------|--------|-----------|-------|---------|--------|-----------|-------|---------|--------|-----------|-------|---------|--------|-----------|-------|
| | Precis. | Recall | Para Est. | Unc. | Precis. | Recall | Para Est. | Unc. | Precis. | Recall | Para Est. | Unc. | Precis. | Recall | Para Est. | Unc. |
| 1 | 1.0 | 1.0 | 1.0 | 0.021 | 1.0 | 1.0 | 1.0 | 0.016 | 1.0 | 1.0 | 0.95 | 0.092 | 1.0 | 1.0 | 0.98 | 0.043 |
| 2 | 1.0 | 1.0 | 0.88 | 0.082 | 1.0 | 1.0 | 1.0 | 0.027 | 1.0 | 1.0 | 0.9 | 0.100 | 1.0 | 1.0 | 0.93 | 0.068 |
| 3 | 1.0 | 1.0 | 0.93 | 0.101 | 1.0 | 1.0 | 1.0 | 0.029 | 1.0 | 1.0 | 0.9 | 0.128 | 1.0 | 1.0 | 0.94 | 0.086 |
| 4 | 1.0 | 1.0 | 1.0 | 0.064 | 1.0 | 1.0 | 1.0 | 0.034 | 1.0 | 1.0 | 0.98 | 0.093 | 1.0 | 1.0 | 0.99 | 0.064 |
| 5 | 1.0 | 1.0 | 1.0 | 0.028 | 1.0 | 1.0 | 1.0 | 0.014 | 1.0 | 1.0 | 0.98 | 0.102 | 1.0 | 1.0 | 0.99 | 0.048 |
| 6 | 1.0 | 1.0 | 0.93 | 0.111 | 1.0 | 1.0 | 1.0 | 0.043 | 1.0 | 1.0 | 0.98 | 0.061 | 1.0 | 1.0 | 0.97 | 0.071 |
| 7 | 1.0 | 1.0 | 0.98 | 0.047 | 1.0 | 1.0 | 1.0 | 0.027 | 0.91 | 1.0 | 0.95 | 0.062 | 0.97 | 1.0 | 0.98 | 0.045 |
| 8 | 1.0 | 1.0 | 0.93 | 0.075 | 1.0 | 1.0 | 1.0 | 0.016 | 1.0 | 1.0 | 1.0 | 0.089 | 1.0 | 1.0 | 0.98 | 0.060 |
| 9 | 1.0 | 1.0 | 1.0 | 0.050 | 1.0 | 1.0 | 1.0 | 0.063 | 1.0 | 1.0 | 0.95 | 0.120 | 1.0 | 1.0 | 0.98 | 0.078 |
| 10 | 0.62 | 0.8 | 0.73 | 0.214 | 1.0 | 1.0 | 1.0 | 0.045 | 1.0 | 1.0 | 0.88 | 0.119 | 0.87 | 0.93 | 0.87 | 0.093 |
| 11 | 1.0 | 1.0 | 0.95 | 0.080 | 1.0 | 1.0 | 0.93 | 0.059 | 1.0 | 1.0 | 0.9 | 0.064 | 1.0 | 1.0 | 0.93 | 0.068 |
| 12 | 1.0 | 1.0 | 0.98 | 0.065 | 1.0 | 1.0 | 1.0 | 0.012 | 1.0 | 1.0 | 0.925 | 0.046 | 1.0 | 1.0 | 0.97 | 0.041 |
| Avg. | 0.97 | 0.98 | 0.94 | 0.078 | 1.0 | 1.0 | 0.99 | 0.032 | 1.0 | 1.0 | 0.90 | 0.089 | 0.99 | 0.99 | 0.96 | 0.064 |

Table 5: Evaluation of the glyph detector by participant (PID) and design: for each participant and design, we report the number of glyphs to be detected (Num), the number of true positive detection (TP), false positive detection (FP), and false negative detection (FN), from which we compute precision and recall.

| Design | PID | Num | TP | FP | FN | Precis. | Recall |
|--------|-----|-----|----|----|----|---------|--------|
| book | 1 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 2 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 3 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 4 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 5 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 6 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 7 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 8 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 9 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 10 | 10 | 8 | 5 | 2 | 0.62 | 0.8 |
| | 11 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 12 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| Avg. | | — | — | — | — | 0.97 | 0.98 |
| leaf | 1 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 2 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 3 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 4 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 5 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 6 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 7 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 8 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 9 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 10 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 11 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 12 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| Avg. | | — | — | — | — | 1.0 | 1.0 |
| smell | 1 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 2 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 3 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 4 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 5 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 6 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 7 | 10 | 10 | 1 | 0 | 0.91 | 1.0 |
| | 8 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 9 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 10 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 11 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | 12 | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| Avg. | | — | — | — | — | 0.99 | 1.0 |

Table 6: Accuracy of the estimation for each parameter (P) and each participant (PID)

| Design | PID | P1 | P2 | P3 | P4 | Avg. P |
|--------|-----|-------|-------|-------|-------|--------|
| book | 1 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| | 2 | 1.0 | 0.5 | 1.0 | 1.0 | 0.875 |
| | 3 | 1.0 | 0.9 | 0.8 | 1.0 | 0.925 |
| | 4 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| | 5 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| | 6 | 1.0 | 0.7 | 1.0 | 1.0 | 0.925 |
| | 7 | 1.0 | 0.9 | 1.0 | 1.0 | 0.975 |
| | 8 | 1.0 | 0.7 | 1.0 | 1.0 | 0.925 |
| | 9 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| | 10 | 0.75 | 0.875 | 0.75 | 0.5 | 0.719 |
| | 11 | 1.0 | 0.8 | 1.0 | 1.0 | 0.95 |
| | 12 | 1.0 | 0.9 | 1.0 | 1.0 | 0.975 |
| Avg. | | 0.979 | 0.856 | 0.962 | 0.958 | 0.939 |
| leaf | 1 | 1.0 | 1.0 | 1.0 | — | 1.0 |
| | 2 | 1.0 | 1.0 | 1.0 | — | 1.0 |
| | 3 | 1.0 | 1.0 | 1.0 | — | 1.0 |
| | 4 | 1.0 | 1.0 | 1.0 | — | 1.0 |
| | 5 | 1.0 | 1.0 | 1.0 | — | 1.0 |
| | 6 | 1.0 | 1.0 | 1.0 | — | 1.0 |
| | 7 | 1.0 | 1.0 | 1.0 | — | 1.0 |
| | 8 | 1.0 | 1.0 | 1.0 | — | 1.0 |
| | 9 | 1.0 | 1.0 | 1.0 | — | 1.0 |
| | 10 | 1.0 | 1.0 | 1.0 | — | 1.0 |
| | 11 | 1.0 | 0.8 | 1.0 | — | 0.933 |
| | 12 | 1.0 | 1.0 | 1.0 | — | 1.0 |
| Avg. | | 1.0 | 0.983 | 1.0 | — | 0.994 |
| smell | 1 | 1.0 | 1.0 | 0.9 | 0.9 | 0.95 |
| | 2 | 1.0 | 1.0 | 0.9 | 0.7 | 0.9 |
| | 3 | 1.0 | 0.8 | 0.7 | 0.7 | 0.8 |
| | 4 | 1.0 | 0.9 | 1.0 | 1.0 | 0.975 |
| | 5 | 1.0 | 1.0 | 1.0 | 0.9 | 0.975 |
| | 6 | 1.0 | 1.0 | 1.0 | 0.9 | 0.975 |
| | 7 | 1.0 | 0.9 | 1.0 | 0.9 | 0.95 |
| | 8 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| | 9 | 1.0 | 0.9 | 1.0 | 0.9 | 0.95 |
| | 10 | 1.0 | 1.0 | 1.0 | 0.5 | 0.875 |
| | 11 | 1.0 | 1.0 | 0.8 | 0.8 | 0.9 |
| | 12 | 1.0 | 1.0 | 1.0 | 0.7 | 0.925 |
| Avg. | | 1.0 | 0.961 | 0.941 | 0.825 | 0.931 |

Table 7: Evaluation of the glyph detector by drawing material and design: for each drawing material and design, we report the number of glyphs to be detected (Num), the number of true positive detection (TP), false positive detection (FP), and false negative detection (FN), from which we compute precision and recall.

| Design | Materia | Num | TP | FP | FN | Precis. | Recall |
|--------|------------|-----|----|----|----|---------|--------|
| book | pencil | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | watercolor | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | digital | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| Avg. | | — | — | — | — | 1.0 | 1.0 |
| leaf | pencil | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | watercolor | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | digital | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| Avg. | | — | — | — | — | 1.0 | 1.0 |
| smell | pencil | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| | watercolor | 10 | 10 | 2 | 0 | 0.83 | 1.0 |
| | digital | 10 | 10 | 0 | 0 | 1.0 | 1.0 |
| Avg. | | — | — | — | — | 0.94 | 1.0 |

Table 8: Accuracy of the parameter estimation for each parameter (P) over three designs drawn with different tools.

| Design | Material | P1 | P2 | P3 | P4 | Avg. P |
|--------|------------|-----|-----|-----|-----|--------|
| book | pencil | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| | watercolor | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| | digital | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| Avg. | | — | — | — | — | 1.0 |
| leaf | pencil | 1.0 | 1.0 | 1.0 | - | 1.0 |
| | watercolor | 1.0 | 1.0 | 1.0 | - | 1.0 |
| | digital | 1.0 | 1.0 | 1.0 | - | 1.0 |
| Avg. | | — | — | — | — | 1.0 |
| smell | pencil | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| | watercolor | 1.0 | 0.9 | 1.0 | 0.9 | 0.95 |
| | digital | 1.0 | 0.8 | 1.0 | 1.0 | 0.95 |
| Avg. | | — | — | — | — | 0.97 |

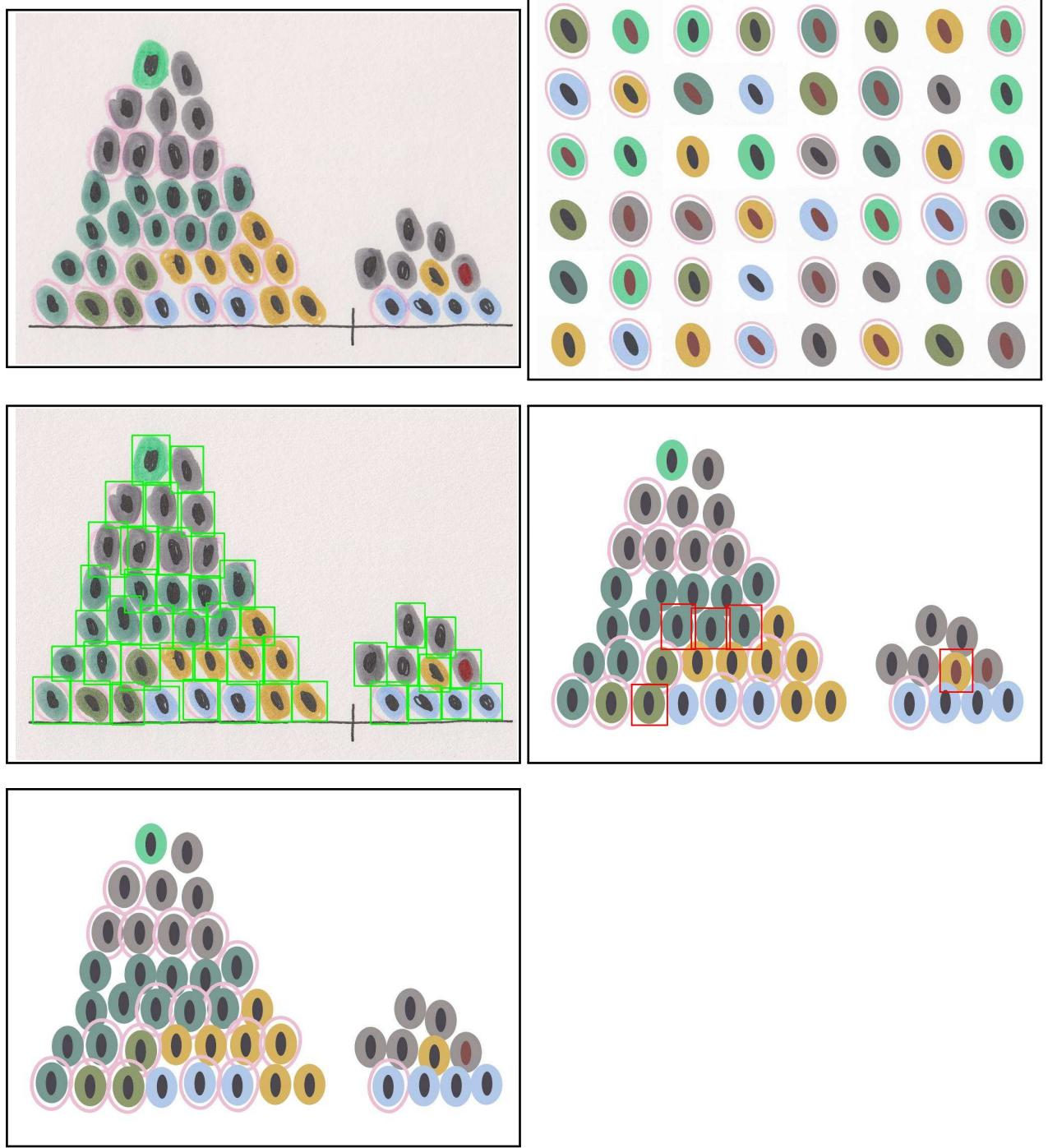


Figure 5: Results for the *Thoughts* design. Input image (top left), sampled glyphs from the training dataset (top right), detected glyphs (center left, green boxes indicate True Positive detection and red boxes indicate False Positive detection), reconstruction from the estimated parameter values (center right, red outlines indicate glyphs for which one or several parameter values are erroneous), reconstruction from the ground truth parameter values (bottom left).



Figure 6: Results for the *Sound* design. Input image (top left), sampled glyphs from the training dataset (top right), detected glyphs (center left, green boxes indicate True Positive detection and red boxes indicate False Positive detection), reconstruction from the estimated parameter values (center right, red outlines indicate glyphs for which one or several parameter values are erroneous), reconstruction from the ground truth parameter values (bottom left).



Figure 7: Results for the *Book* design. Input image (top left), sampled glyphs from the training dataset (top right), detected glyphs (center left, green boxes indicate True Positive detection and red boxes indicate False Positive detection), reconstruction from the estimated parameter values (center right, red outlines indicate glyphs for which one or several parameter values are erroneous), reconstruction from the ground truth parameter values (bottom left).

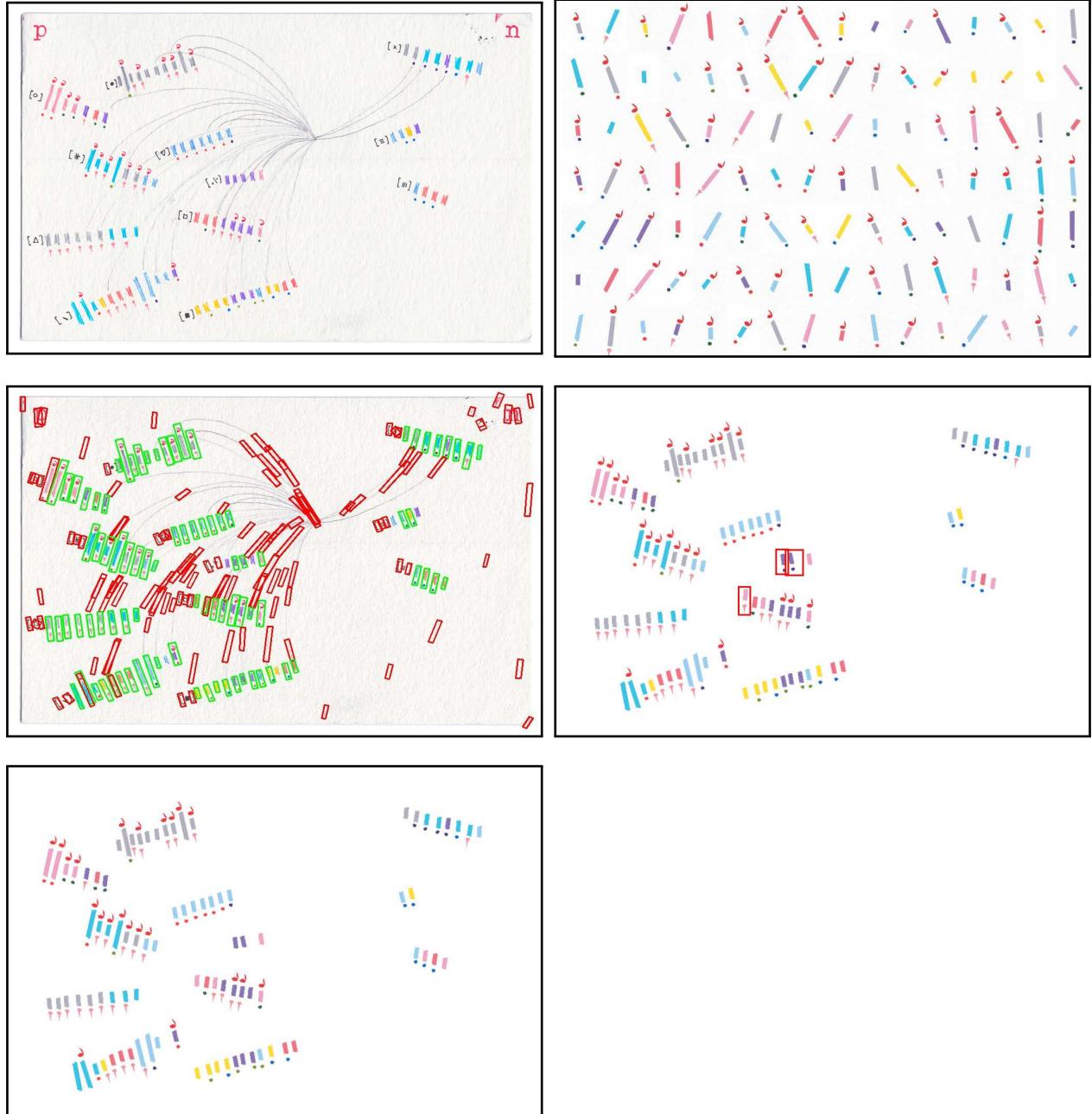


Figure 8: Results for the *Boyfriend* design. Input image (top left), sampled glyphs from the training dataset (top right), detected glyphs (center left, green boxes indicate True Positive detection and red boxes indicate False Positive detection), reconstruction from the estimated parameter values (center right, red outlines indicate glyphs for which one or several parameter values are erroneous), reconstruction from the ground truth parameter values (bottom left).



Figure 9: Results for the *Dog* design. Input image (top left), sampled glyphs from the training dataset (top right), detected glyphs (center left, green boxes indicate True Positive detection and red boxes indicate False Positive detection), reconstruction from the estimated parameter values (center right, red outlines indicate glyphs for which one or several parameter values are erroneous), reconstruction from the ground truth parameter values (bottom left).

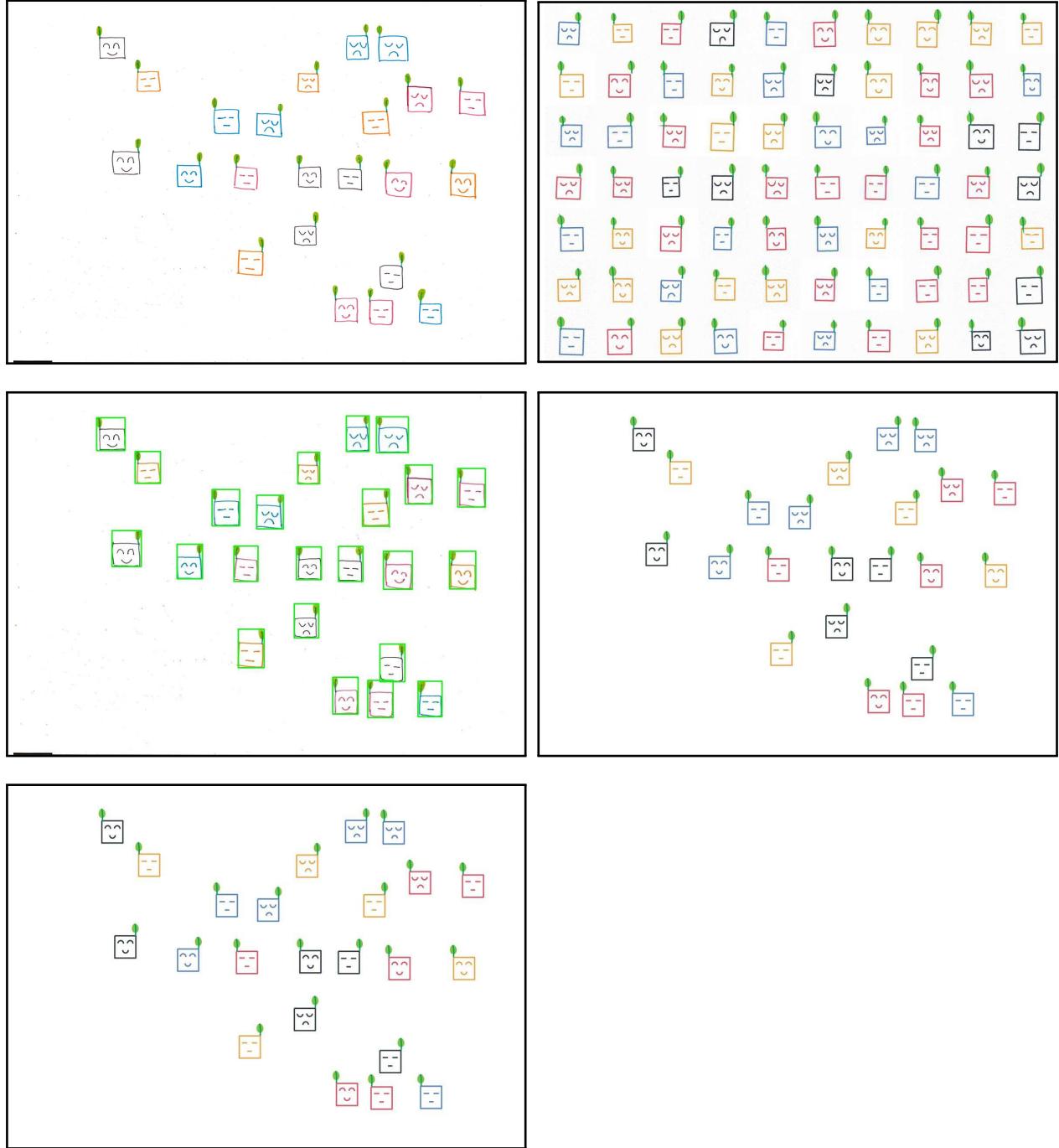


Figure 10: Results for the *Leaf* design. Input image (top left), sampled glyphs from the training dataset (top right), detected glyphs (center left, green boxes indicate True Positive detection and red boxes indicate False Positive detection), reconstruction from the estimated parameter values (center right, red outlines indicate glyphs for which one or several parameter values are erroneous), reconstruction from the ground truth parameter values (bottom left).

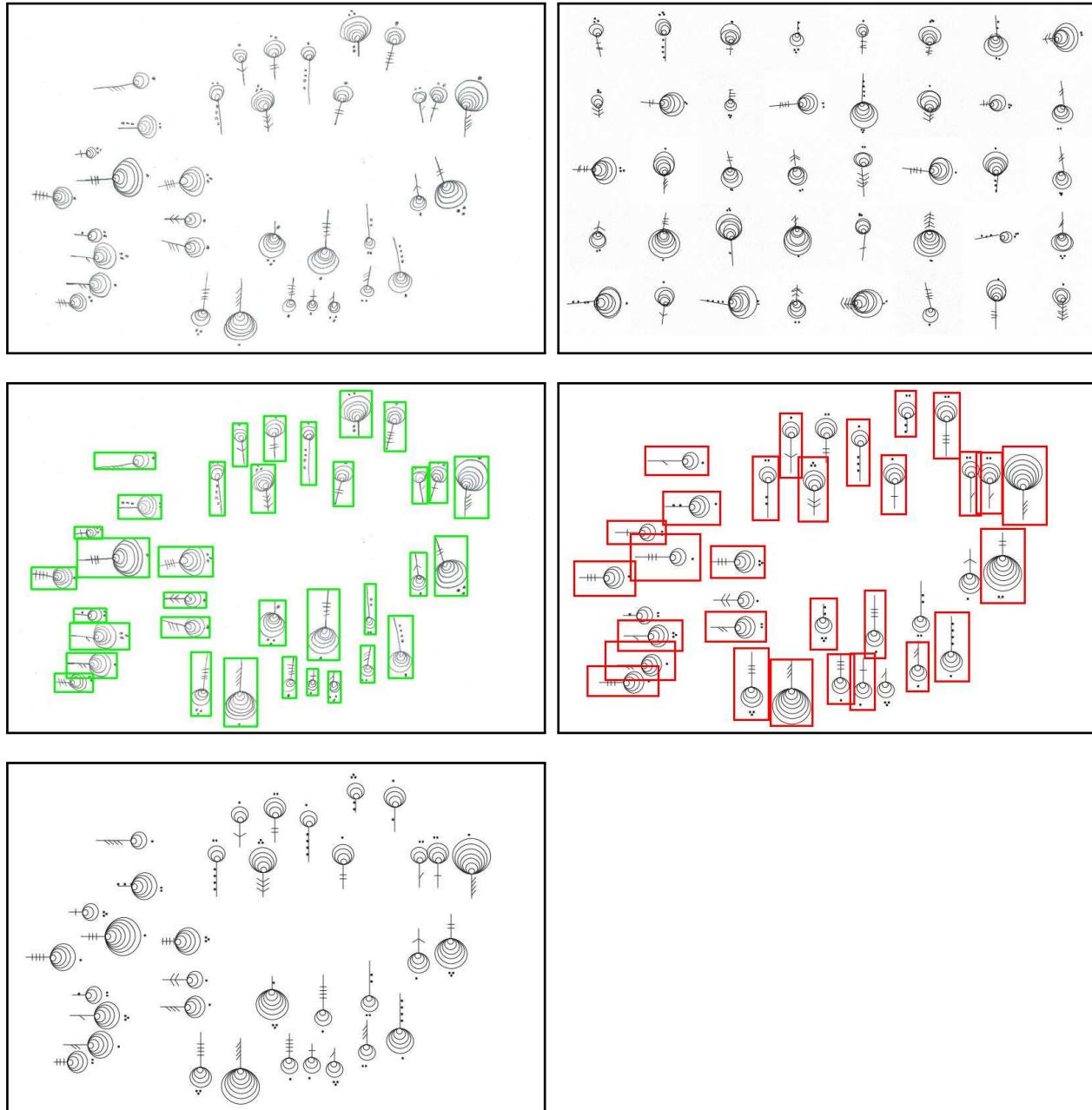


Figure 11: Results for the *Lollipop* design. Input image (top left), sampled glyphs from the training dataset (top right), detected glyphs (center left, green boxes indicate True Positive detection and red boxes indicate False Positive detection), reconstruction from the estimated parameter values (center right, red outlines indicate glyphs for which one or several parameter values are erroneous), reconstruction from the ground truth parameter values (bottom left).

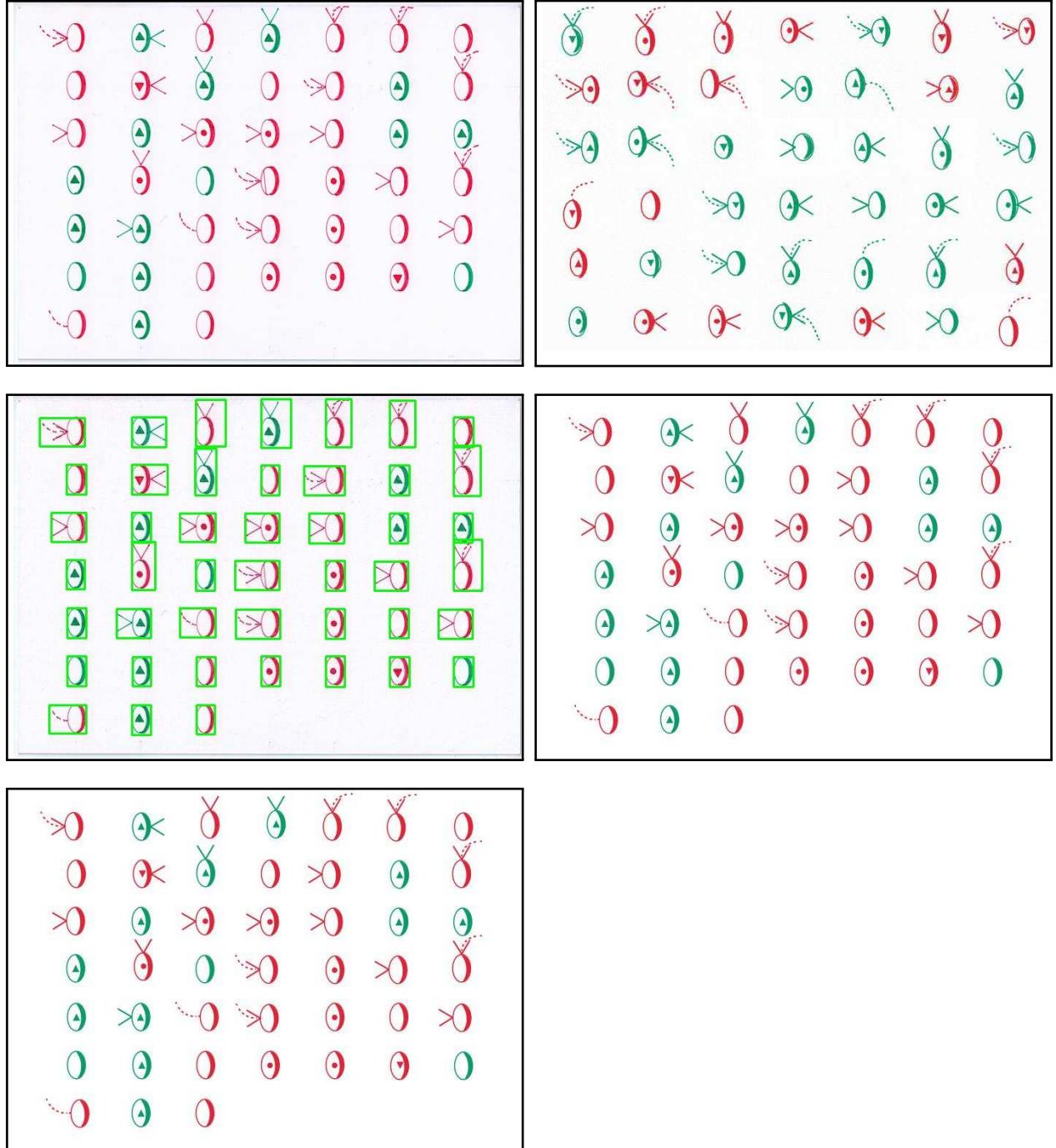


Figure 12: Results for the *Smell* design. Input image (top left), sampled glyphs from the training dataset (top right), detected glyphs (center left, green boxes indicate True Positive detection and red boxes indicate False Positive detection), reconstruction from the estimated parameter values (center right, red outlines indicate glyphs for which one or several parameter values are erroneous), reconstruction from the ground truth parameter values (bottom left).

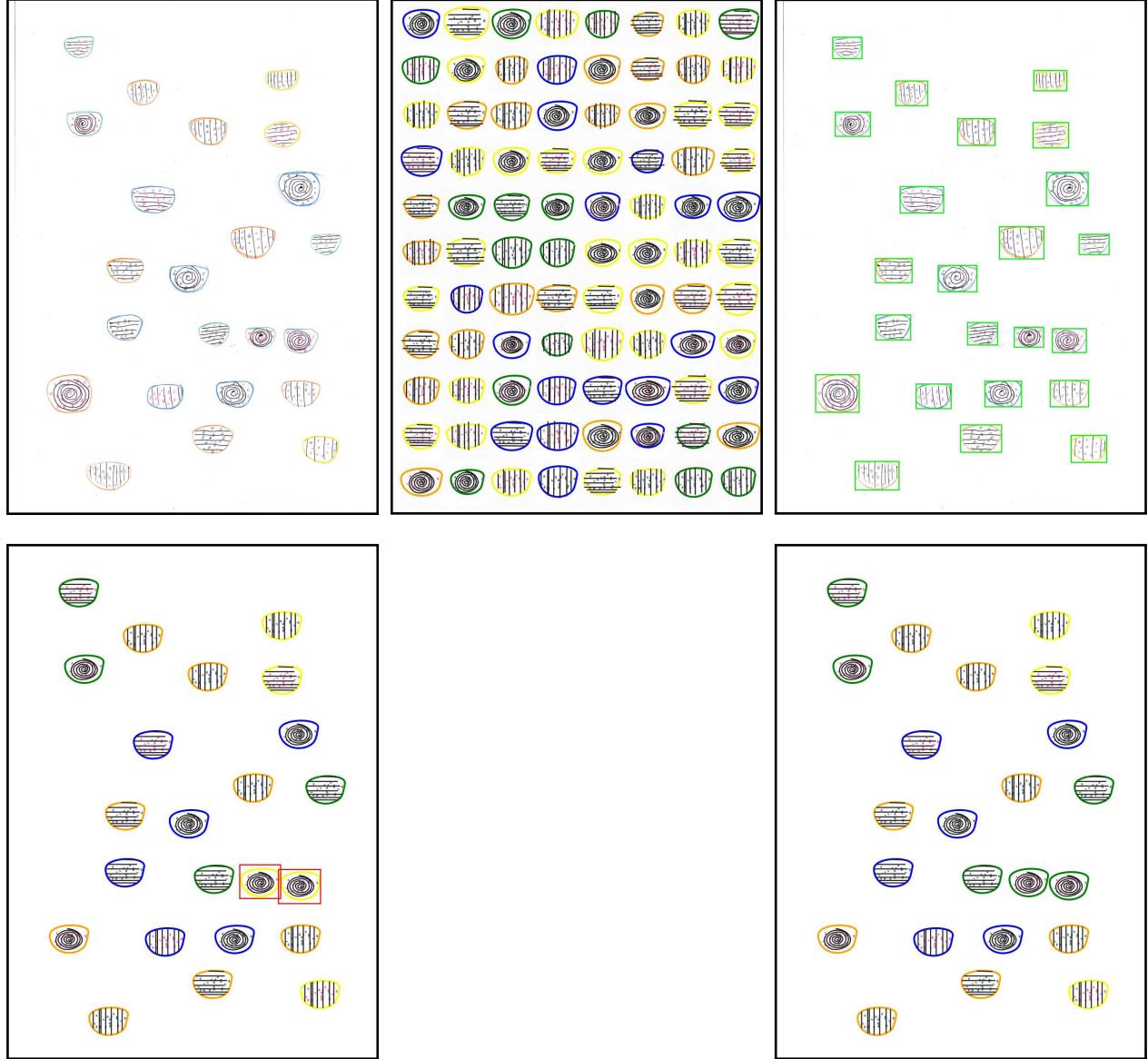


Figure 13: Results for the *Tree* design. Input image (top left), sampled glyphs from the training dataset (top center), detected glyphs (top right, green boxes indicate True Positive detection and red boxes indicate False Positive detection), reconstruction from the estimated parameter values (bottom left, red outlines indicate glyphs for which one or several parameter values are erroneous), reconstruction from the ground truth parameter values (bottom right).

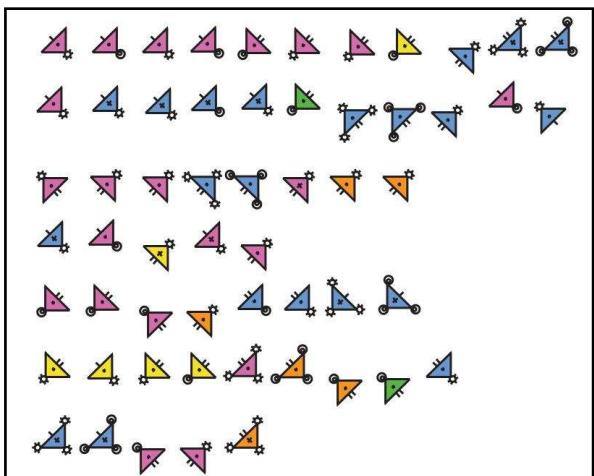
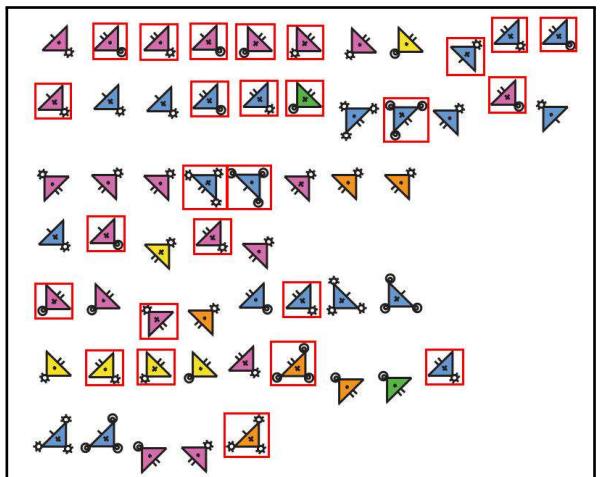
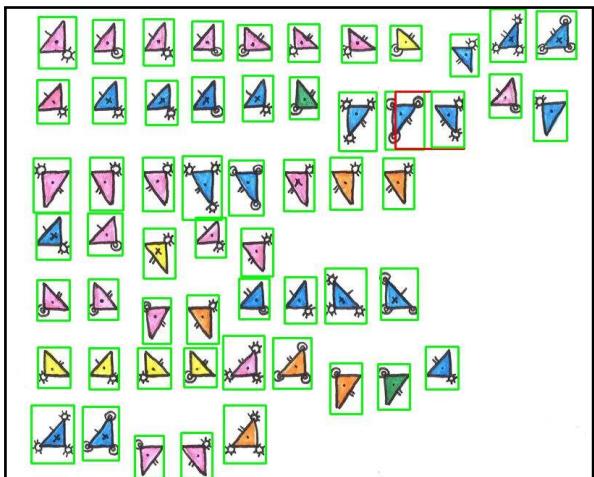
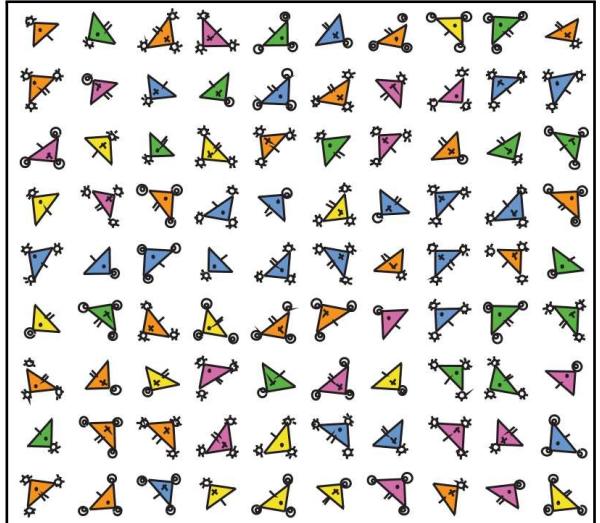
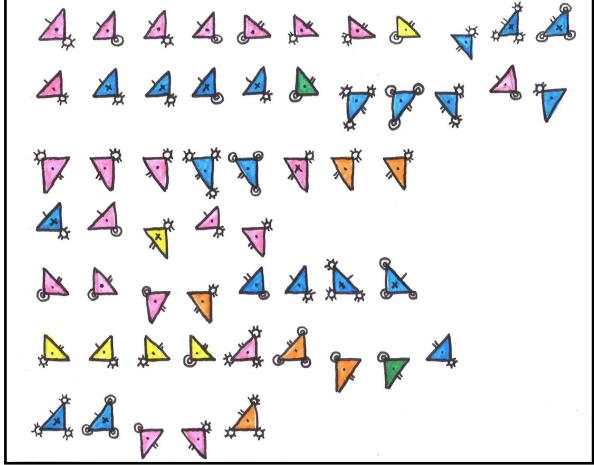


Figure 14: Results for the *Triangle* design. Input image (top left), sampled glyphs from the training dataset (top right), detected glyphs (center left, green boxes indicate True Positive detection and red boxes indicate False Positive detection), reconstruction from the estimated parameter values (center right, red outlines indicate glyphs for which one or several parameter values are erroneous), reconstruction from the ground truth parameter values (bottom left).

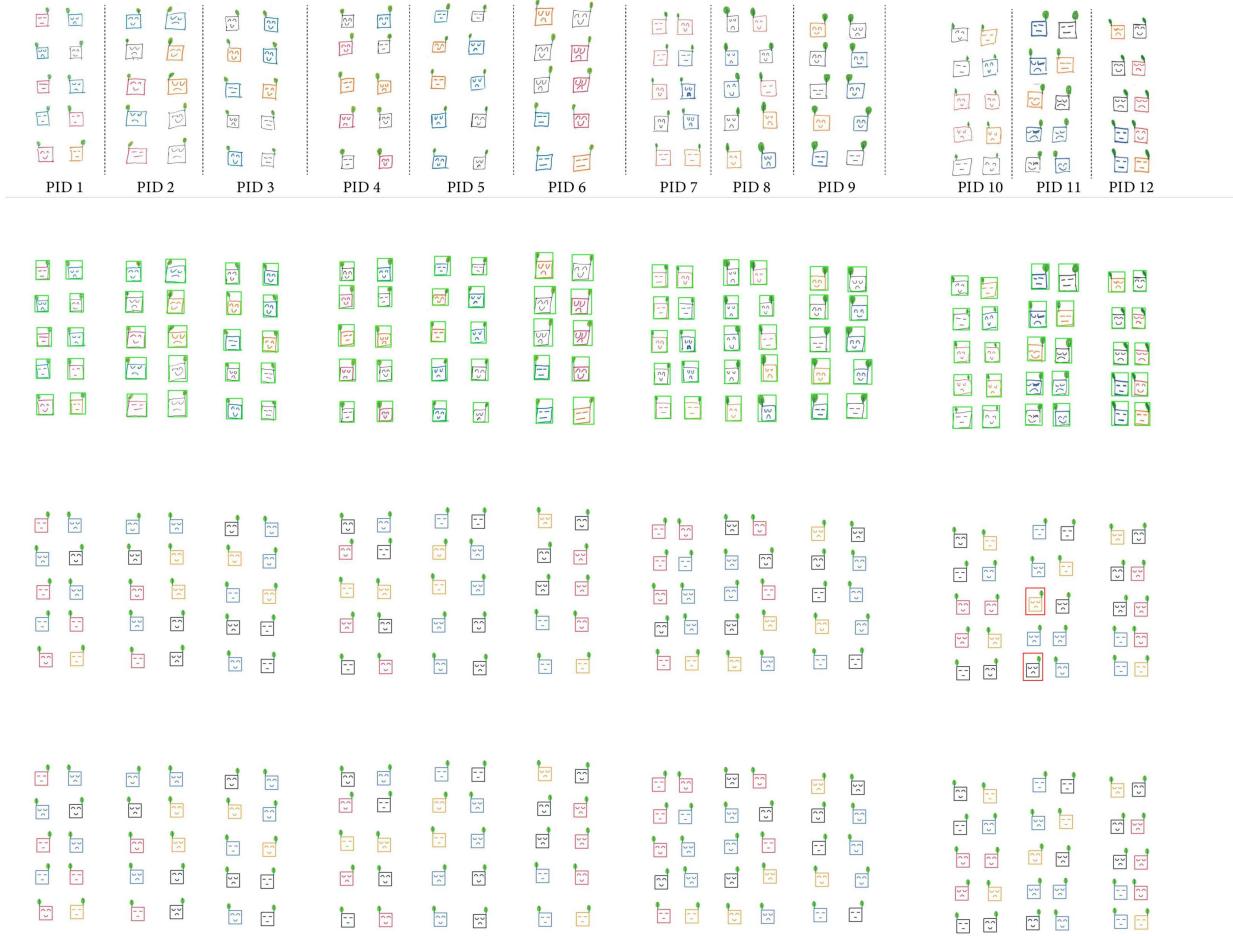


Figure 15: Results for the *Leaf* design from 12 participants. Input image (top, drawings from 12 participants and PID indicates participant ID), detected glyphs (second row, green boxes indicate True Positive detection and red boxes indicate False Positive detection), reconstruction from the estimated parameter values (third row, red outlines indicate glyphs for which one or several parameter values are erroneous), reconstruction from the ground truth parameter values (bottom).

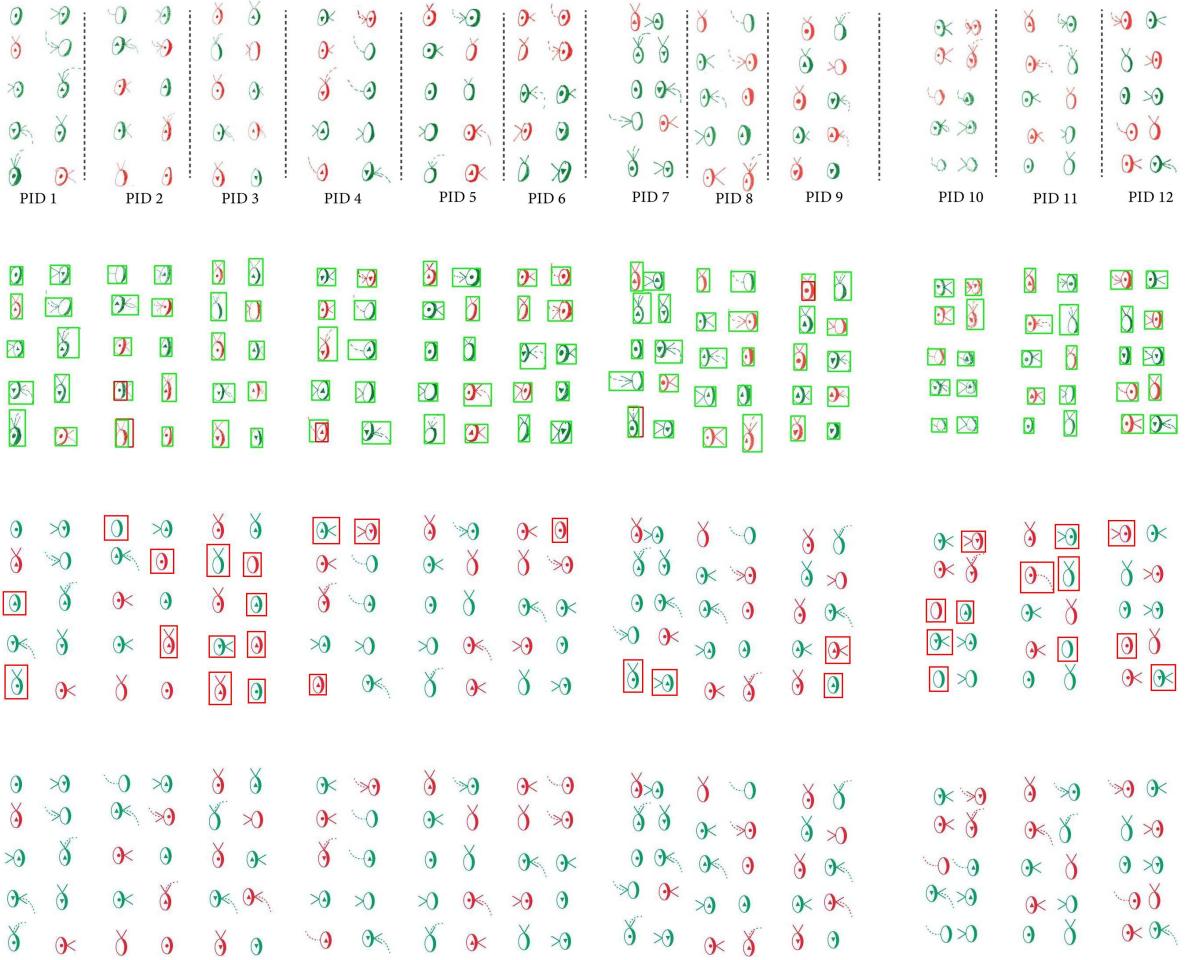


Figure 16: Results for the *Smell* design from 12 participants. Input image (top, drawings from 12 participants and PID indicates participant ID), detected glyphs (second row, green boxes indicate True Positive detection and red boxes indicate False Positive detection), reconstruction from the estimated parameter values (third row, red outlines indicate glyphs for which one or several parameter values are erroneous), reconstruction from the ground truth parameter values (bottom).



Figure 17: Results for the *Book* design from 12 participants. Input image (top, drawings from 12 participants and PID indicates participant ID), detected glyphs (second row, green boxes indicate True Positive detection and red boxes indicate False Positive detection), reconstruction from the estimated parameter values (third row, red outlines indicate glyphs for which one or several parameter values are erroneous), reconstruction from the ground truth parameter values (bottom).

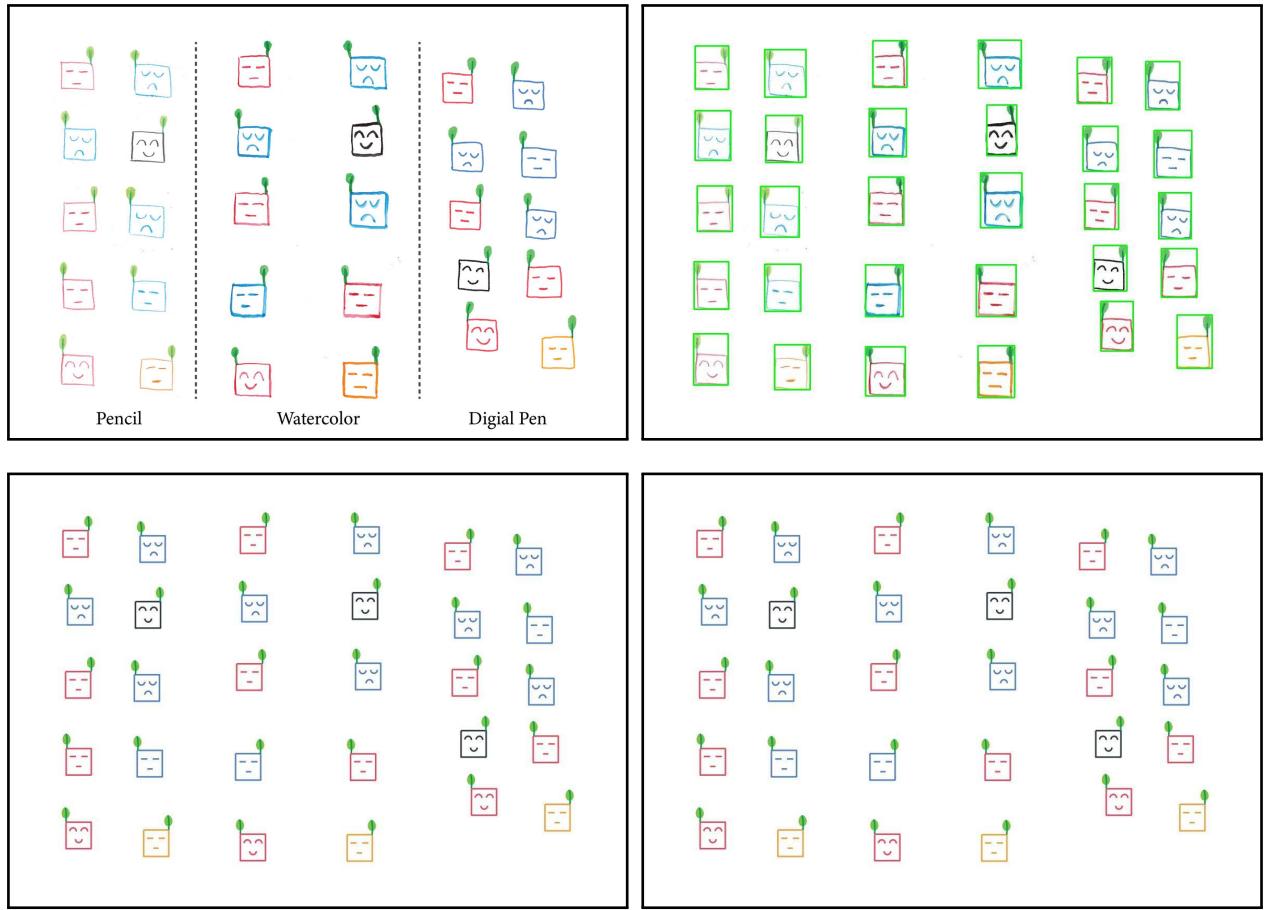


Figure 18: Results for the *Leaf* design from different drawing tools. Input image (top left, glyphs drawn by pencil, watercolor and digital pen), detected glyphs (top right, green boxes indicate True Positive detection and red boxes indicate False Positive detection), reconstruction from the estimated parameter values (bottom left, red outlines indicate glyphs for which one or several parameter values are erroneous), reconstruction from the ground truth parameter values (bottom right).

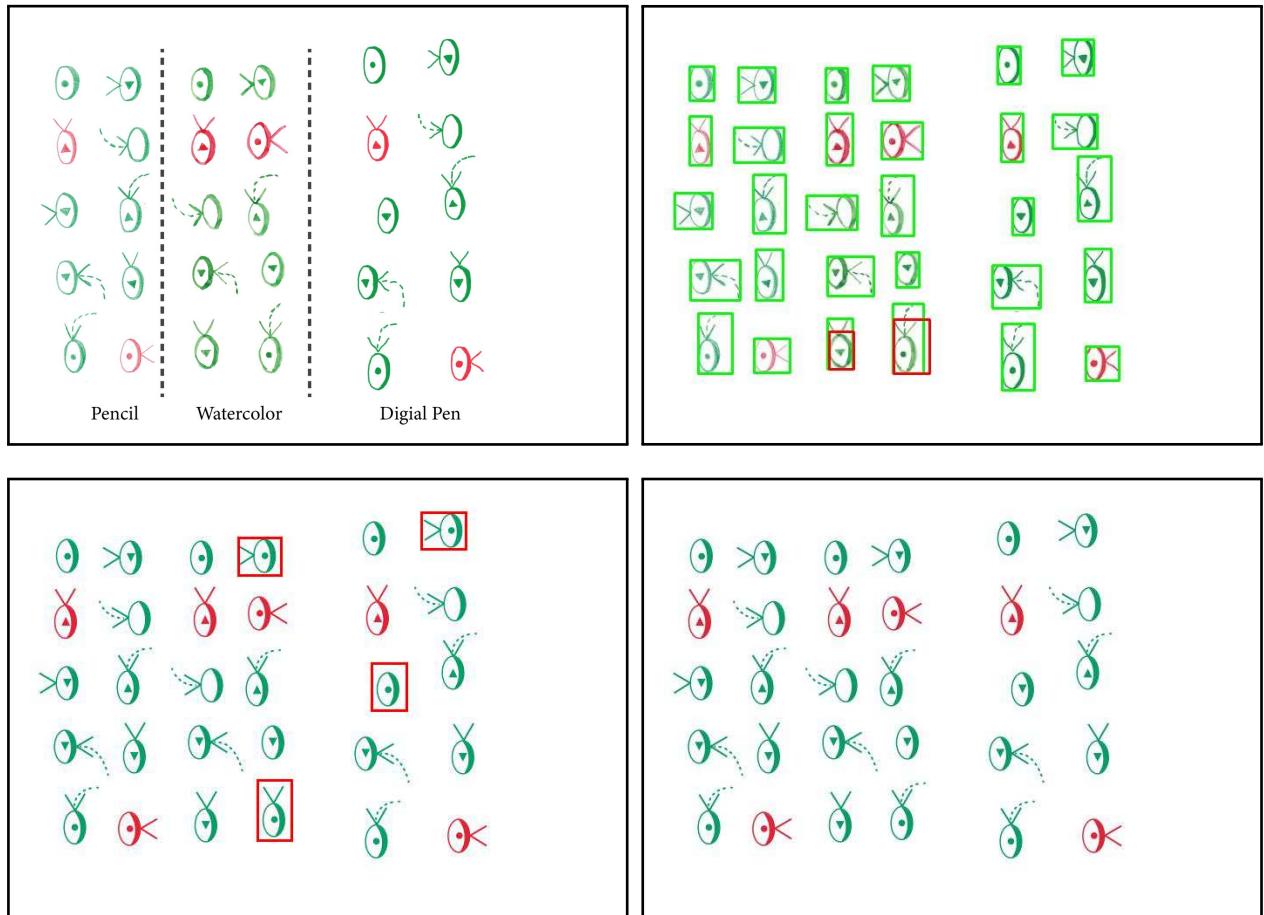


Figure 19: Results for the *Smell* design from different drawing tools. Input image (top left, glyphs drawn by pencil, watercolor and digital pen), detected glyphs (top right, green boxes indicate True Positive detection and red boxes indicate False Positive detection), reconstruction from the estimated parameter values (bottom left, red outlines indicate glyphs for which one or several parameter values are erroneous), reconstruction from the ground truth parameter values (bottom right).

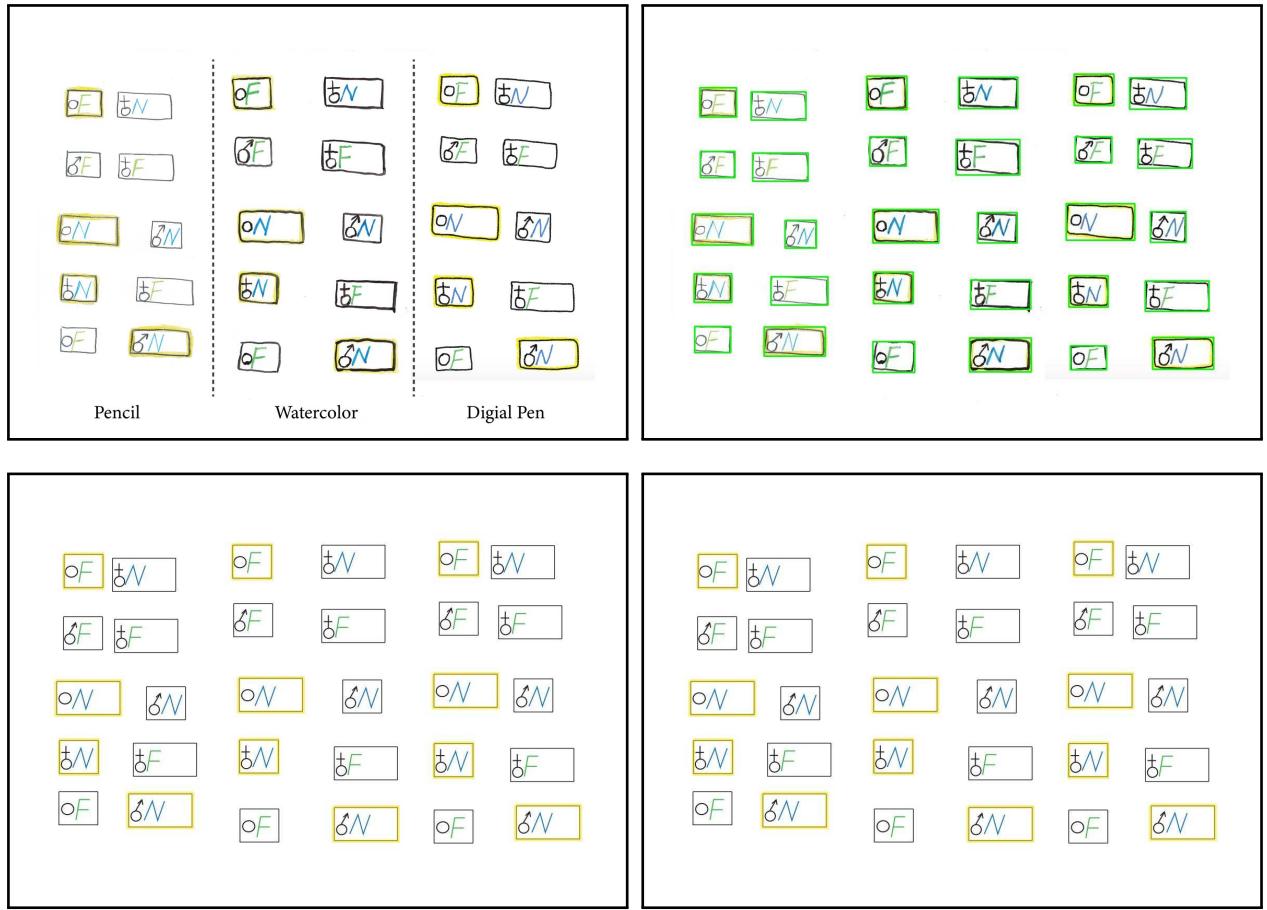


Figure 20: Results for the *Book* design from different drawing tools. Input image (top left, glyphs drawn by pencil, watercolor and digital pen), detected glyphs (top right, green boxes indicate True Positive detection and red boxes indicate False Positive detection), reconstruction from the estimated parameter values (bottom left, red outlines indicate glyphs for which one or several parameter values are erroneous), reconstruction from the ground truth parameter values (bottom right).

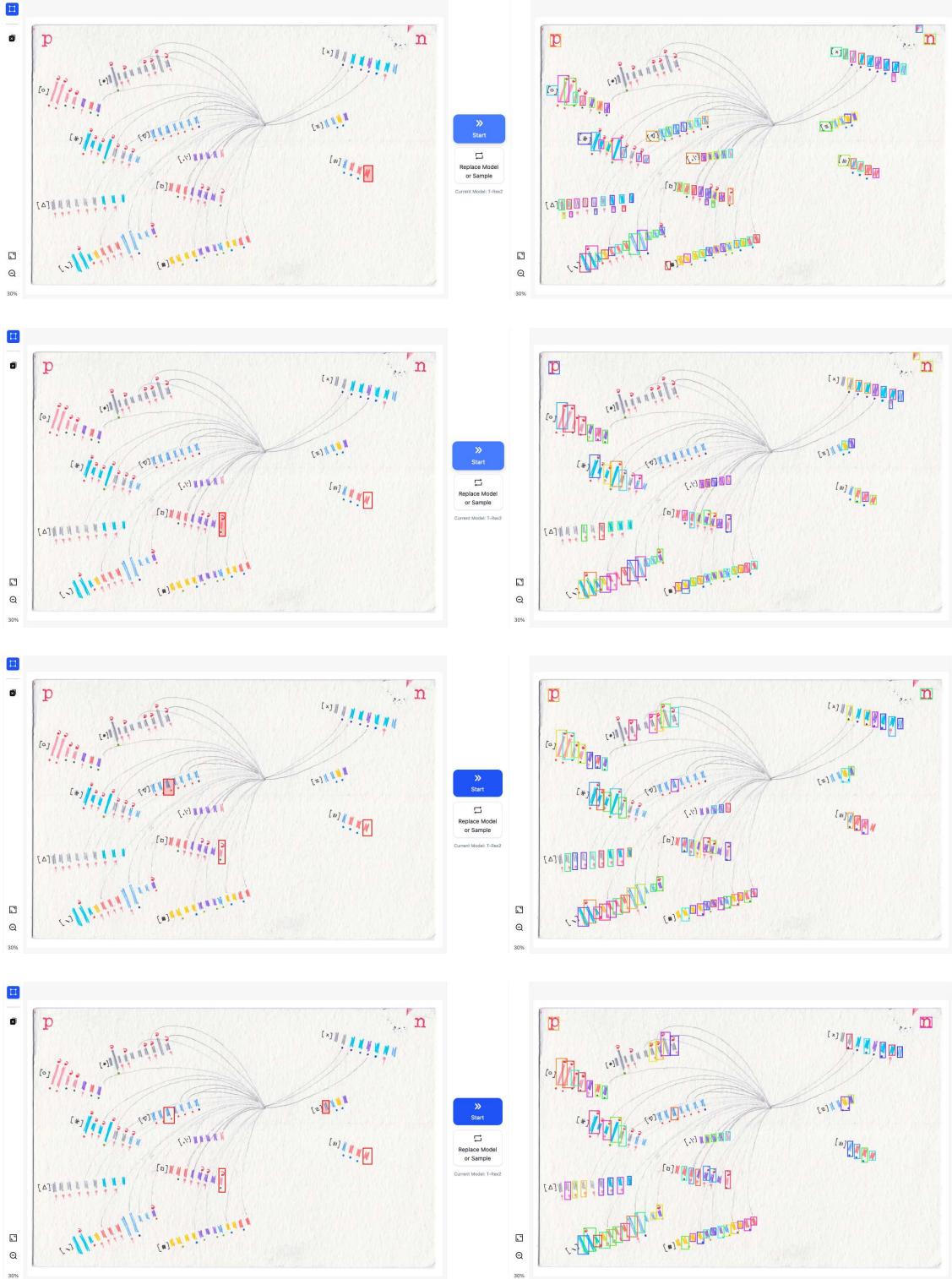


Figure 21: Bounding box detection results from T-REX (Jiang et al., 2024) of the *Boyfriend* design. The first row shows the results after specifying one bounding box, where the input box is highlighted in red in the left figure, and the detected boxes are displayed as colorful outlines in the right figure. The subsequent rows illustrate the results after specifying two bounding boxes (second row), three bounding boxes (third row), and four bounding boxes (fourth row). Despite more examples, the detector fails to locate even subtle variants of the glyphs.

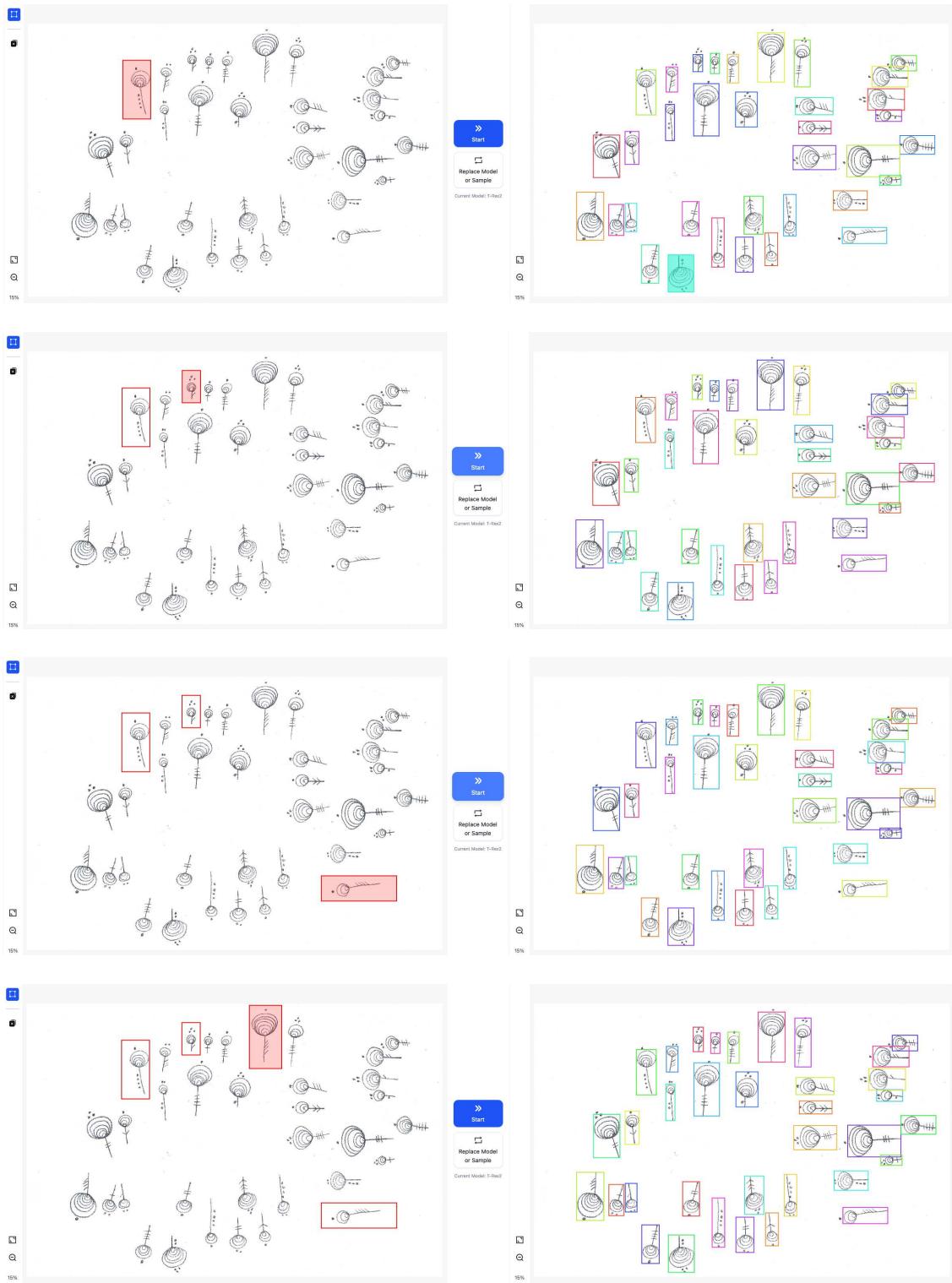


Figure 22: Bounding box detection results from T-REX (Jiang et al., 2024) of the *Lollipop* design. The first row shows the results after specifying one bounding box, where the input box is highlighted in red in the left figure, and the detected boxes are displayed as colorful outlines in the right figure. The subsequent rows illustrate the results after specifying two bounding boxes (second row), three bounding boxes (third row), and four bounding boxes (fourth row). While most glyphs are detected, the bounding boxes do not always capture the dots at the top of the glyphs.