TEXT-BASED VECTOR SKETCH EDITING WITH IMAGE EDITING DIFFUSION PRIOR –SUPPLEMENTAL DOCUMENT–

Haoran Mo, Xusheng Lin, Chengying Gao and Ruomei Wang*

Sun Yat-sen University
{mohaor,linxsh8}@mail2.sysu.edu.cn, {mcsgcy,isswrm}@mail.sysu.edu.cn

1. IMPLEMENTATION DETAILS

1.1. Iterative Editing

We show the implementation details of iterative editing in Algorithm 1, following some formulations in Prompt-to-Prompt [1].

Algorithm 1 Algorithm for iterative editing.

Input: All prompts $\{P^1, P^2, ..., P^K\}$, a random seed ξ, and words $\{(w^1, w^2)^2, (w^2, w^3)^3, ..., (w^{K-1}, w^K)^K\}$ specifying the editing region for local editing.

Output: All images $x^1, x^2, ..., x^K$.

$z_T \sim N(0, I)$, a unit Gaussian random variable with ξ;
$(z_T^1, z_T^2, ..., z_T^K) \leftarrow z_T$;
for $t = T, T-1, ..., 1$ **do**
  $z_{t-1}^1, M_t^1 \leftarrow DM(z_t^1, P^1, t, \xi)$;
  for $k = 2, ..., K$ **do**
    $z_{t-1}^{k-1}, M_t^{k-1} \leftarrow DM(z_t^{k-1}, P^{k-1}, t, \xi)$;
    $M_t^k \leftarrow DM(z_t^k, P^k, t, \xi)$;
    $\widehat{M}_t^k \leftarrow Edit(M_t^{k-1}, M_t^k, t)$;
    $z_{t-1}^k \leftarrow DM(z_t^k, P^k, t, \xi)\{M \leftarrow \widehat{M}_t^k\}$;
    if local **then**
      $\alpha \leftarrow B(\overline{M}_{t,(w^{k-1})^k}^{k-1}) \cup B(\overline{M}_{t,(w^k)^k}^k)$;
      $z_{t-1}^k \leftarrow (1-\alpha) \odot z_{t-1}^{k-1} + \alpha \odot z_{t-1}^k$;
    end if
  end for
end for
$x^1, x^2, ..., x^K = Decode(z_0^1, z_0^2, ..., z_0^K)$.

1.2. Training Details

We choose the pre-trained Stable Diffusion v1.4 in our pipeline. When sampling the original and the edited images, we run 50 inference steps with a classifier-free guidance scale of 7.5. During the optimization of strokes, we train 1000 iterations for each example. We use 96 strokes each of which includes 4 control points to represent the vector sketches. The stroke width is defined with a fixed value 1.0. In the stroke-level local editing scheme, we adopt cross-attention maps of resolution 16×16 in up and down blocks in the diffusion model.

*The corresponding author is Chengying Gao.

2. MORE RESULTS

We show more results including:

- Comparisons in **Word Swap** mode: Fig. 1;
- Comparisons in **Prompt Refinement** mode: Fig. 2;
- **Attention Re-weighting** mode: Fig. 3;
- **Iterative editing**: Fig. 4 and 5.

3. REFERENCES

[1] Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-or, "Prompt-to-prompt image editing with cross-attention control," in *ICLR*, 2023.

Fig. 1. Comparisons with baseline methods in Word Swap mode.

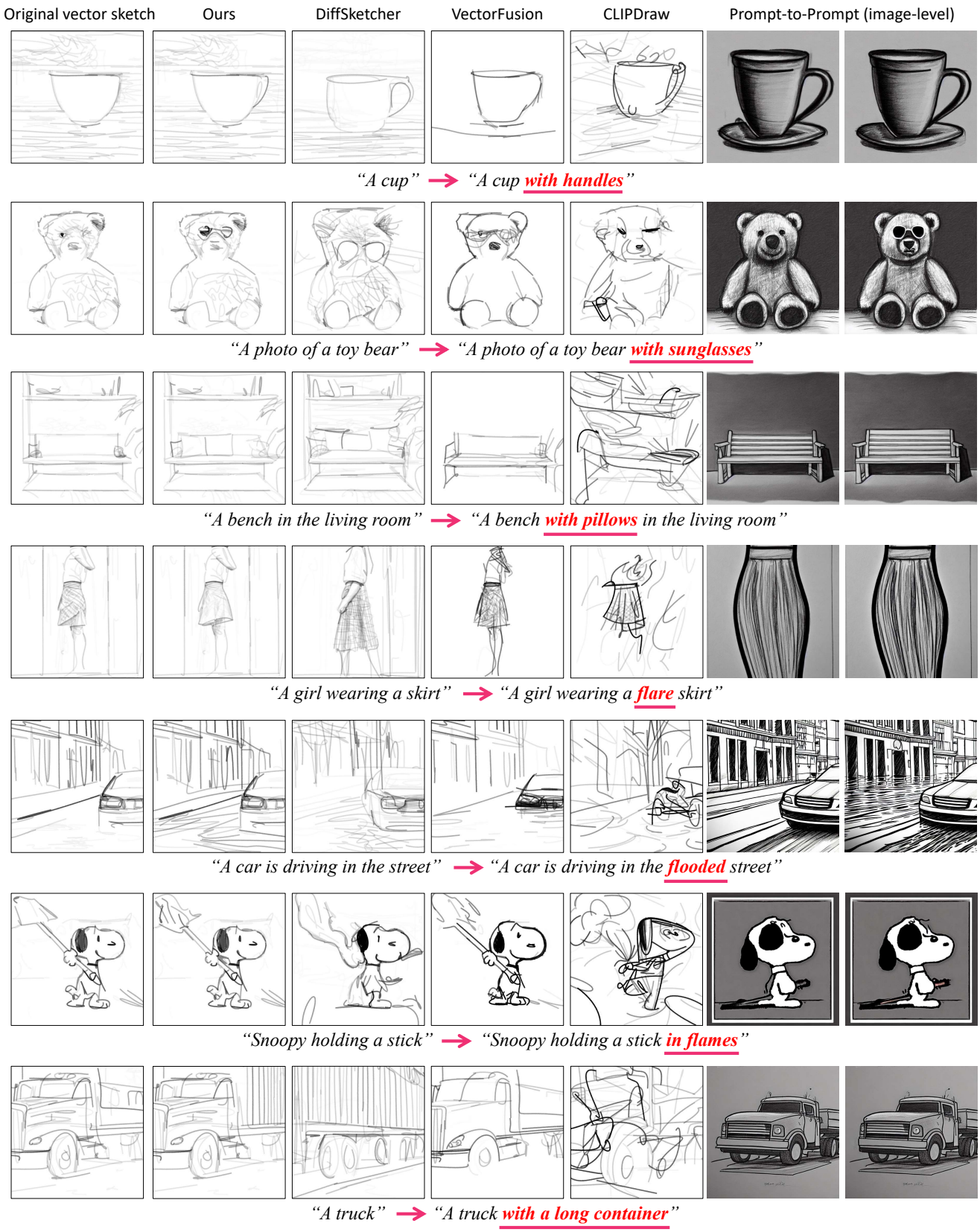

Fig. 2. Comparisons with baseline methods in Prompt Refinement mode.

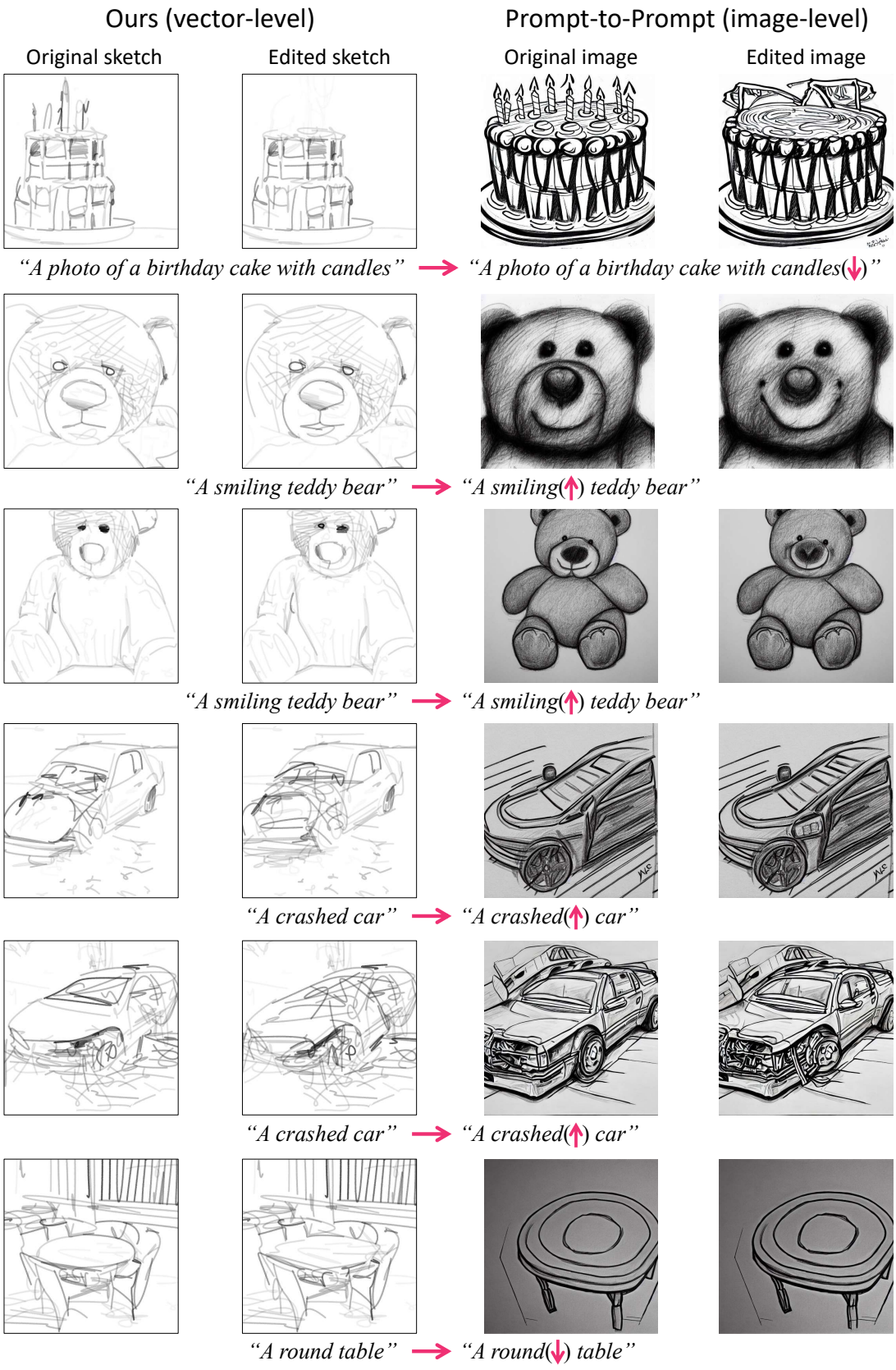

Fig. 3. Results in Attention Re-weighting mode.

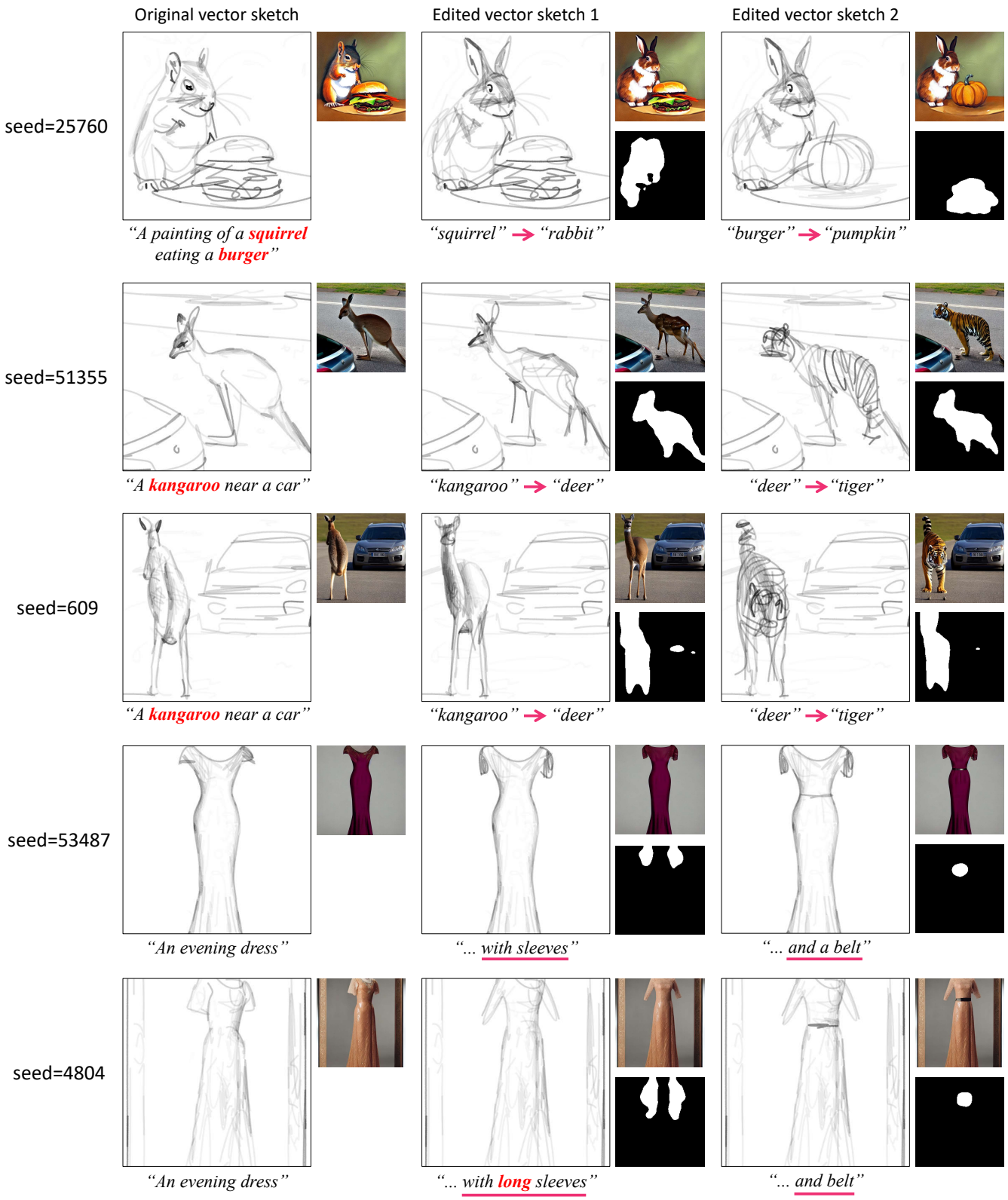

Fig. 4. Results of iterative editing.

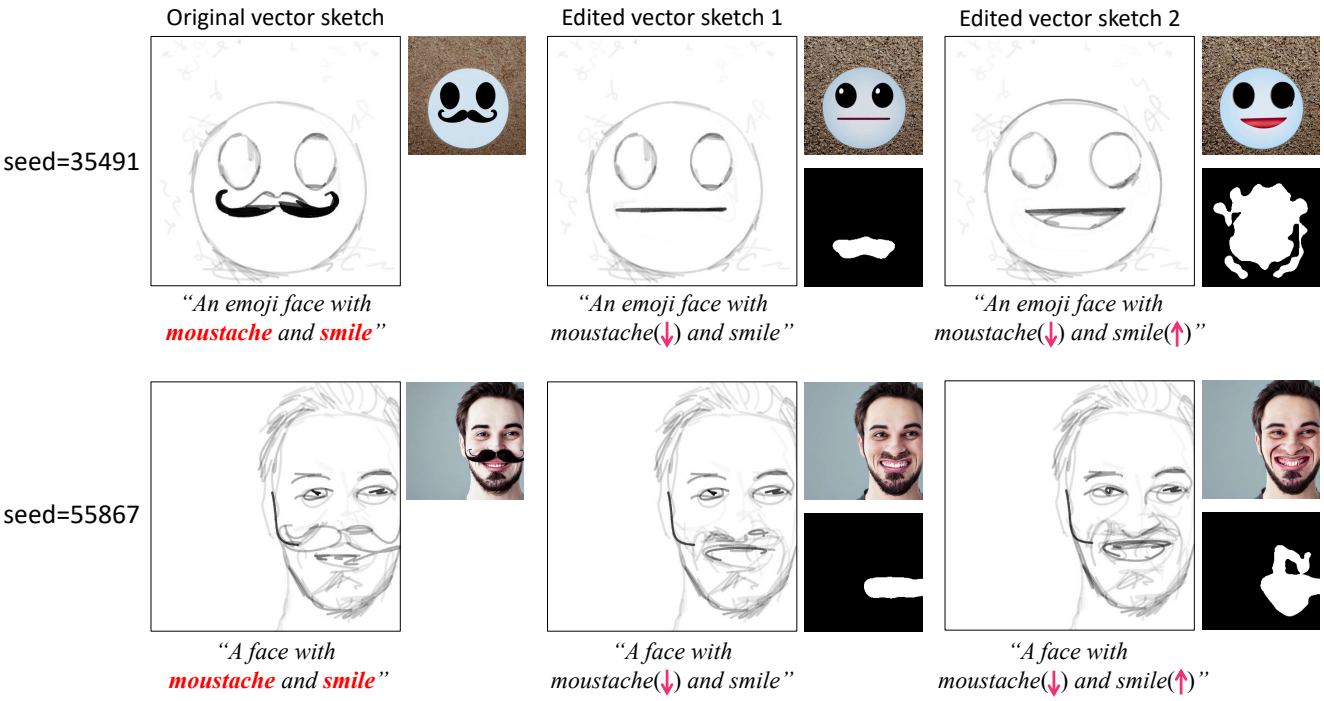

Fig. 5. Results of iterative editing.