## A. Code and Full Supplementary Material

Our code and the full supplementary material are available at https://github.com/xiaorongjun000/Self-Rectification.

## B. Comparison with Image Inpainting

Inpainting addresses a different problem, lacking the ability to control the content generated in unknown regions. Specifically, it does not fill missing regions with source texture patterns in a coherent manner, as demonstrated in Fig. 13 left, where a SOTA diffusion based inpainting method, RePaint [25], is applied to our problem.

## C. Comparison with Mask-guided Synthesis using GCD Loss

Although GCD loss can be used to control the spatial distribution of texture elements, it requires accurate annotated masks (label maps, in fact). Otherwise, its control may fail, as shown in Fig. 13 (right).

## D. Multimodal Outputs

Fig. 14 presents results where we randomly draw different sizes of patches from the source to fill the canvas, leading to significantly variant backgrounds, yet the user edits are all well-preserved.

## E. Limitations

Fig. 15 shows two examples we collected from our full supplementary material. The cause is attributed to the lack of suitable source patterns in these two references (see in the full supplementary material.) that can perfectly complete the gap regions, leading to deviations from the user intent.

## F. More Results of Guided Texture Synthesis

Fig. 16 show additional examples for guided texture synthesis using color layouts. More results can be seen in our full supplementary material.

## G. Algorithm

We apply our method on a pre-trained Stable Diffusion (SD) model, which contains an encoder $\mathcal{E}$, a decoder $\mathcal{D}$, and a noise predictor $\epsilon_\theta$. The full pipeline of our method is depicted by Algorithms 1 to 3. Note the functions Invert(*) and Sample (*) refer to a DDIM inversion step and a DDIM sampling step, respectively, and Att(*) denotes the self-attention mechanism in Stable Diffusion.


Figure 13. Results from RePaint [25] (left), and mask-guided GCD loss [49] (right), where the source and target masks are displayed in the top-right, respectively.

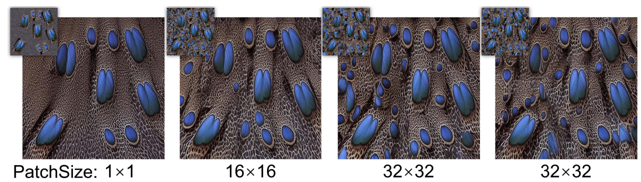


Figure 14. Multimodal outputs by differently filling the canvas.

Figure 15. Limitations and failure cases of our method.

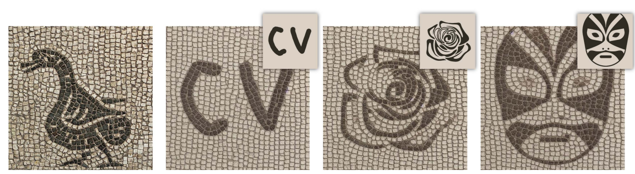


Figure 16. More results of guided synthesis with color layouts.

**Algorithm 1** Overall Framework

**Input:** Reference texture $I^R$
**Output:** Output texture $I^*$
$I^{tar} \leftarrow$ USER_EDIT$(I^R)$
$I^{IR} \leftarrow I^{tar}$
$z_T^{tar} \leftarrow$ StruPreserving_Inversion$(I^{tar}, I^{IR})$
$I^*_{coarse} \leftarrow$ FineTexture_Sampling$(z_T^{tar}, I^R)$
$z_T^* \leftarrow$ StruPreserving_Inversion$(I^*_{coarse}, I^{IR})$
$I^* \leftarrow$ FineTexture_Sampling$(z_T^*, I^R)$
**Return** $I^*$

**Algorithm 2** Structure-preserving Inversion

**Input:** A target image $I^{tar}$, an inversion reference $I^{IR}$
**Output:** Inversion code $z_{\mathrm{T}}^{tar}$

$z_0^{IR} \leftarrow \mathcal{E}(I^{IR})$
$\{z_0^{IR}, z_1^{IR}, \ldots, z_{\mathrm{T}}^{IR}\} \leftarrow$ DDIM_INVERSION$(z_0^{IR})$
$z_0^{tar} \leftarrow \mathcal{E}(I^{tar})$
**for** $t = 0, 1, \ldots \mathrm{P} - 1$ **do**
  $\{Q_{\mathrm{T}-t}^{IR}, K_{\mathrm{T}-t}^{IR}, V_{\mathrm{T}-t}^{IR}\} \leftarrow \epsilon_\theta(z_{\mathrm{T}-t}^{IR}, t)$
  $\{Q_t^{tar}, K_t^{tar}, V_t^{tar}\} \leftarrow \epsilon_\theta(z_t^{tar}, t)$
  $\epsilon = \epsilon_\theta(z_t^{tar}, t) \sim \mathrm{Att}(Q_t^{tar}, K_{\mathrm{T}-t}^{IR}, V_{\mathrm{T}-t}^{IR}))$
  $z_{t+1}^{tar} \leftarrow \mathrm{Invert}(z_t^{tar}, \epsilon, t)$
**end for**
**for** $t = \mathrm{P}, \mathrm{P}+1, \ldots, \mathrm{T} - 1$ **do**
  $\{Q_t^{tar}, K_t^{tar}, V_t^{tar}\} \leftarrow \epsilon_\theta(z_t^{tar}, t)$
  $\epsilon = \epsilon_\theta(z_t^{tar}, t) \sim \mathrm{Att}(Q_t^{tar}, K_t^{tar}, V_t^{tar})$
  $z_{t+1}^{tar} \leftarrow \mathrm{Invert}(z_t^{tar}, \epsilon, t)$
**end for**
**Return** $z_{\mathrm{T}}^{tar}$

**Algorithm 3** Fine-texture Sampling

**Input:** A start code $z_{\mathrm{T}}^{*}$, a reference texture $I^{R}$
**Output:** Output texture $I^*$

$z_0^{R} \leftarrow \mathcal{E}(I^{R})$
$\{z_0^{R}, z_1^{R}, \ldots, z_{\mathrm{T}}^{R}\} \leftarrow$ DDIM_INVERSION$(z_0^{R})$
**for** $t = \mathrm{T}, \mathrm{T}-1, \ldots, \mathrm{T} - \mathrm{S} - 1$ **do**
  $\{Q_t^{*}, K_t^{*}, V_t^{*}\} \leftarrow \epsilon_\theta(z_t^{*}, t)$
  $\epsilon = \epsilon_\theta(z_t^{*}, t) \sim \mathrm{Att}(Q_t^{*}, K_t^{*}, V_t^{*})$
  $z_{t-1}^{*} \leftarrow \mathrm{Sample}(z_t^{*}, \epsilon, t)$
**end for**
**for** $t = \mathrm{T}-\mathrm{S}, \mathrm{T}-\mathrm{S}+1, \ldots, 1$ **do**
  $\{Q_t^{R}, K_t^{R}, V_t^{R}\} \leftarrow \epsilon_\theta(z_t^{R}, t)$
  $\{Q_t^{*}, K_t^{*}, V_t^{*}\} \leftarrow \epsilon_\theta(z_t^{*}, t)$
  $\epsilon = \epsilon_\theta(z_t^{*}, t) \sim \mathrm{Att}(Q_t^{*}, K_t^{R}, V_t^{R})$
  $z_{t-1}^{*} \leftarrow \mathrm{Sample}(z_t^{*}, \epsilon, t)$
**end for**
$I^* \leftarrow \mathcal{D}(z_0^*)$
**Return** $I^*$