Zero-shot Font Style Transfer with a Differentiable Renderer

Kota Izumi, Keiji Yanai

The University of Electro-Communications, Tokyo, Japan
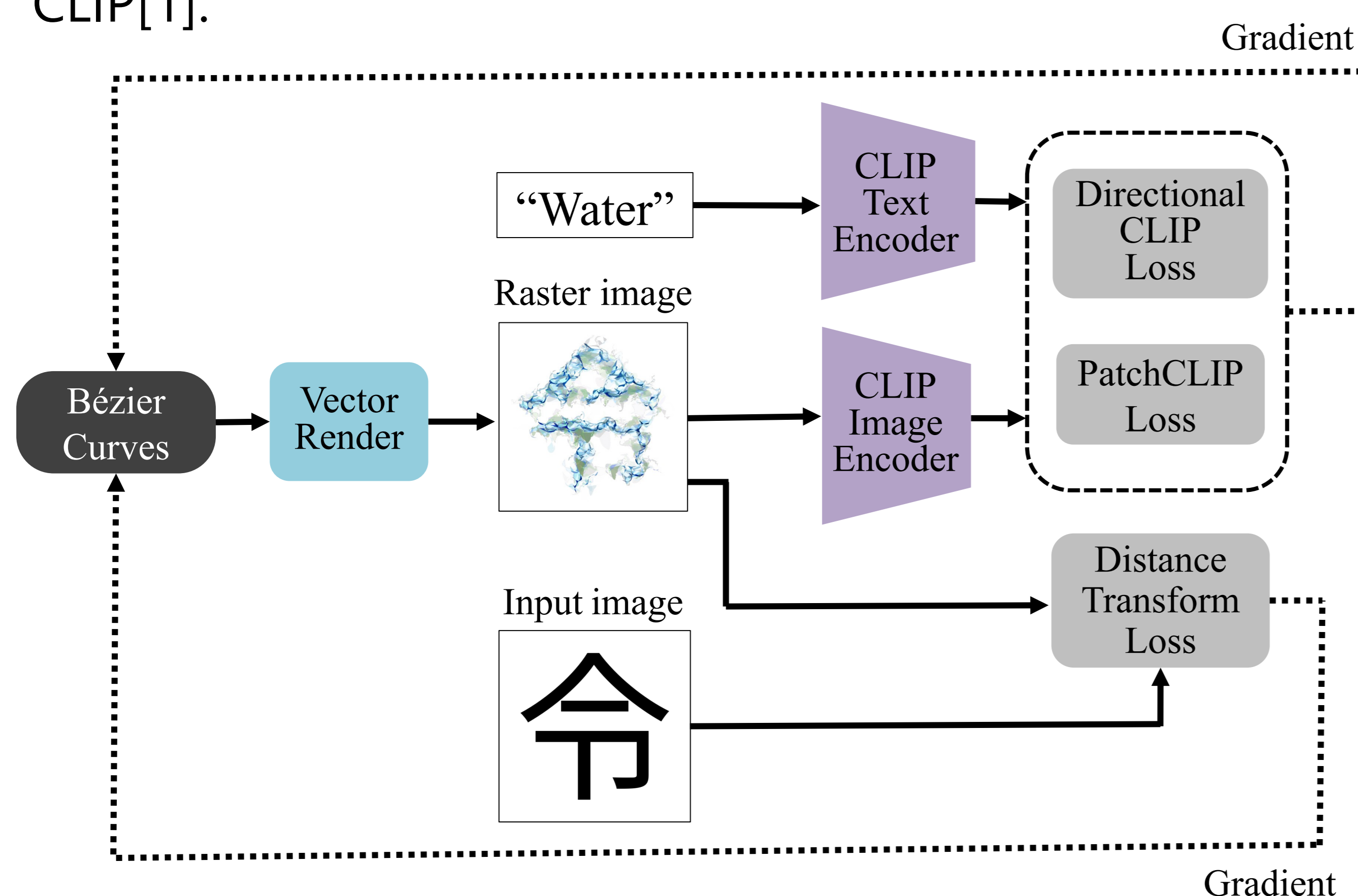
Introduction

- A method of style transfer using only text is proposed.
- However, when input are text images, **font shape does not change significantly.**

Objective

Realize style transfer of font images without style image but only with input texts
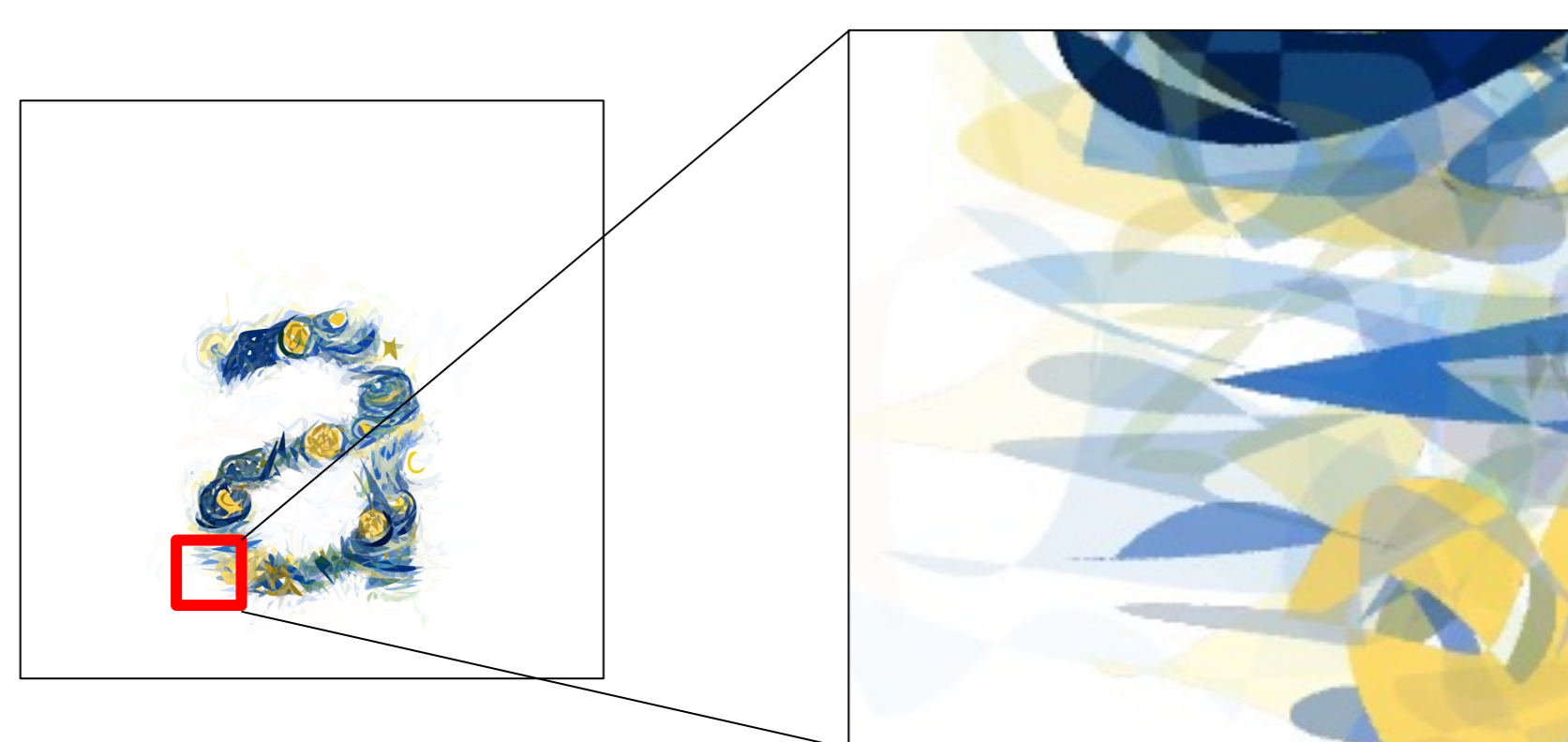
Method

- The parameters of the Bézier curves are optimized.
- The loss is calculated using image and text encoders of CLIP[1].

Font Representation

→ Closed Bézier curves[2].

- Each curve is represented by a line thickness, position of control points, color, and opacity.

Directional CLIP loss L_{dir} & Patch CLIP loss L_{patch}
→ Force the font texture match a prompt text.

Distance Transform Loss $L_{distance}$
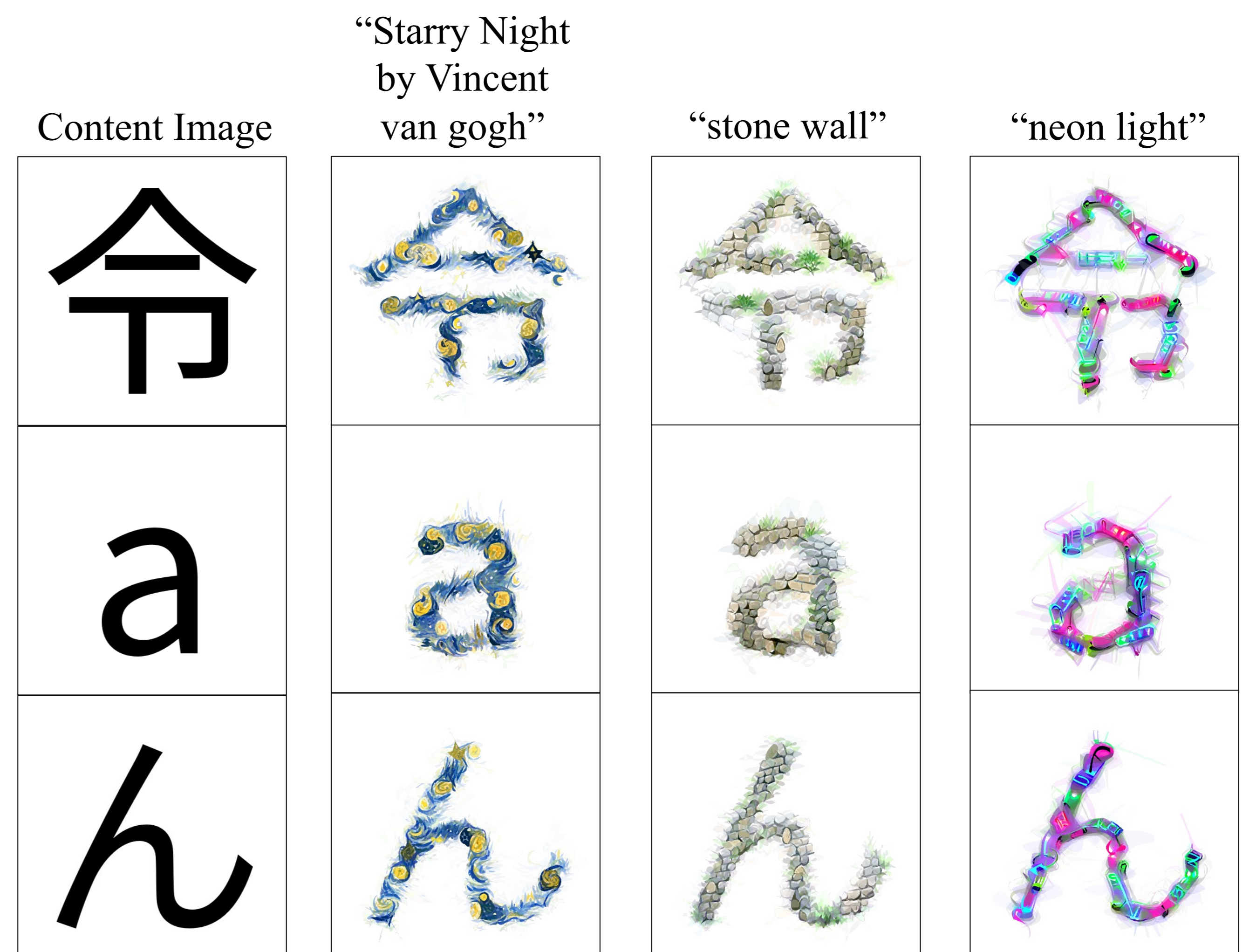→ Preserve the shape of the input fonts.

Total Loss

$$L_{total} = \lambda_d L_{dir} + \lambda_p L_{patch} + \lambda_{distance} L_{distance} + \lambda_{tv} L_{tv}$$
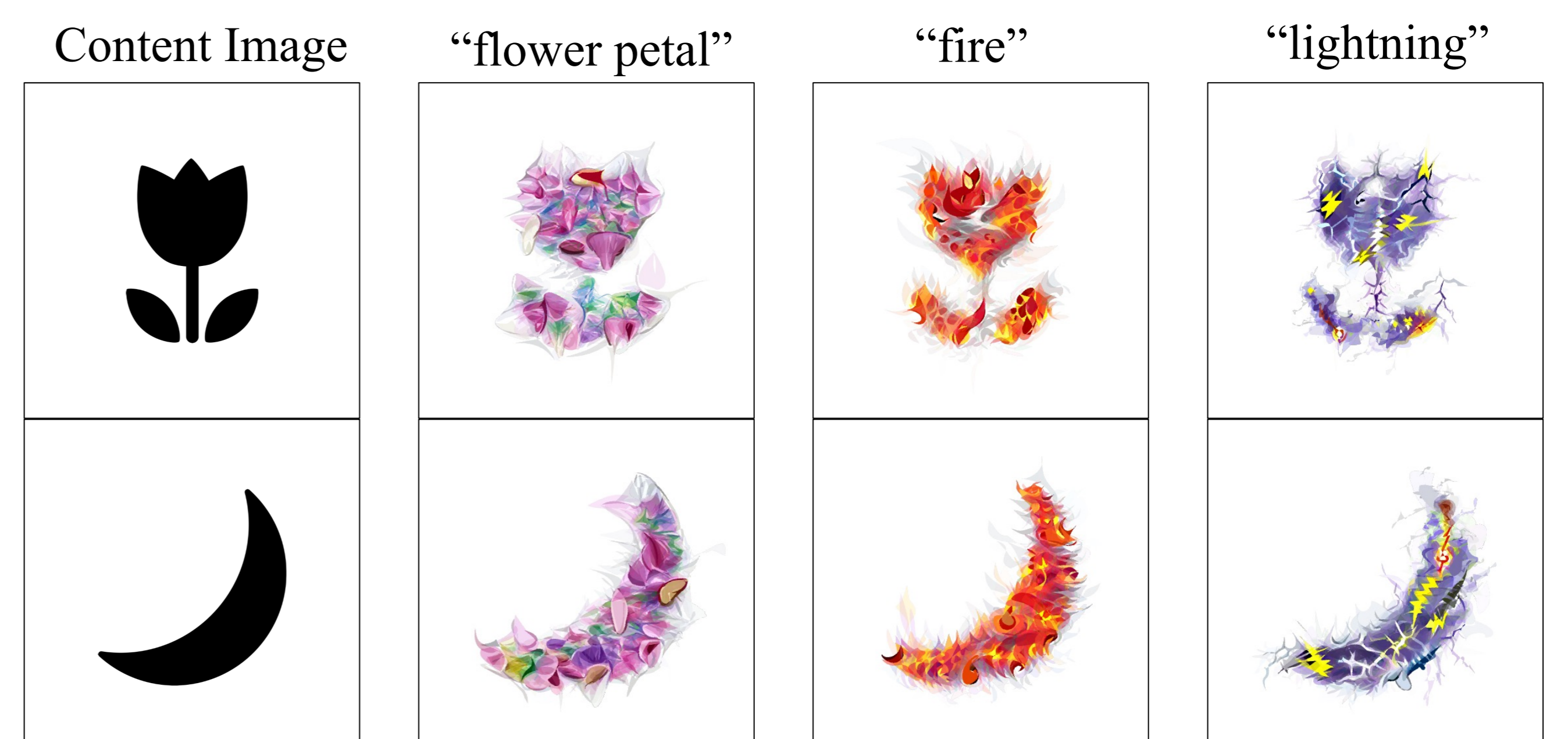
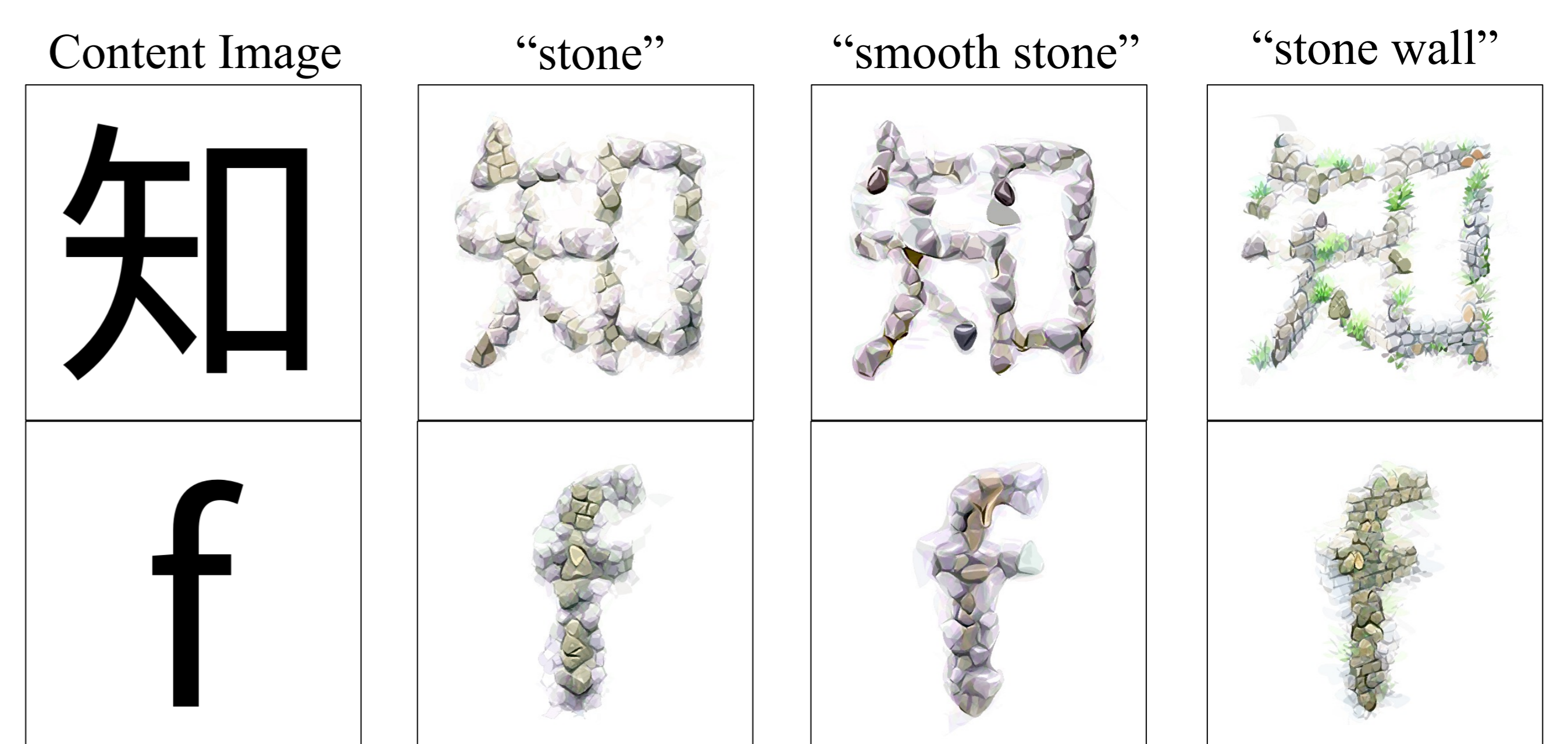L_{tv} → Total variation regularization loss

λ → Loss weight

Result

- Use pretrained CLIP model
- Training time → **100s** per image on a RTX2080Ti
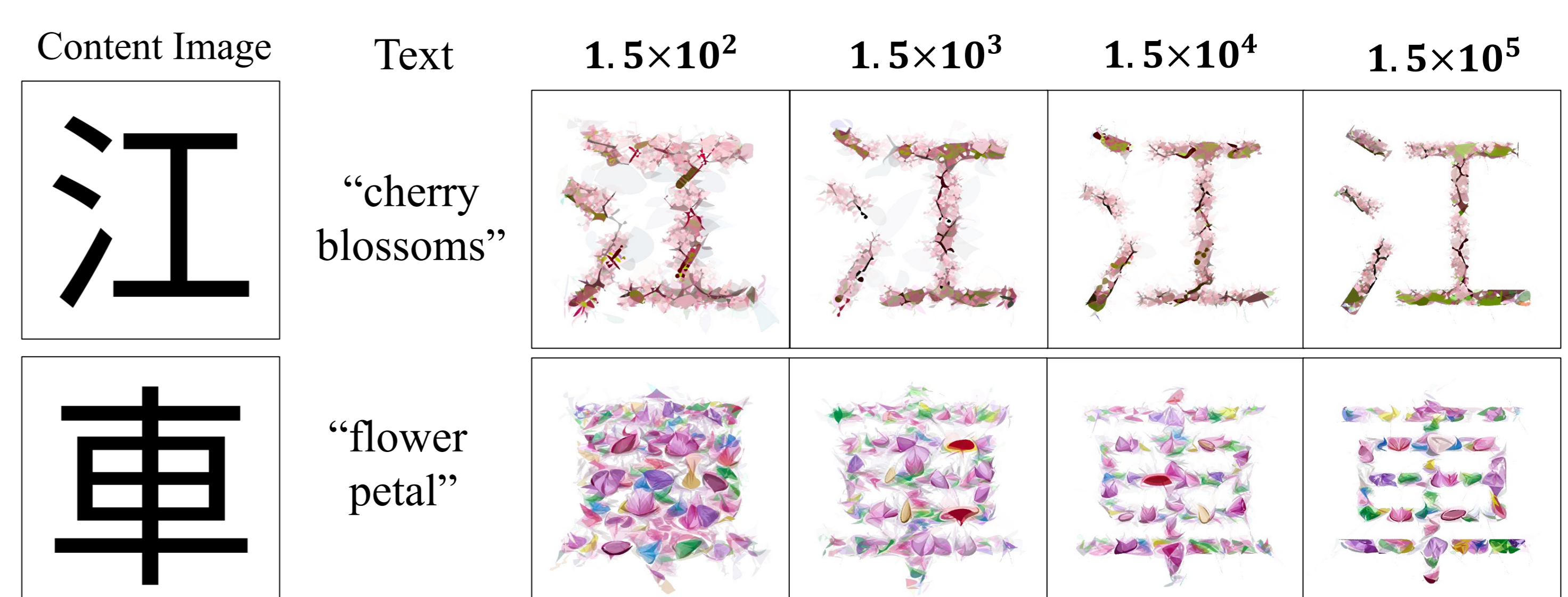

For simple logos

Effect of adjectives

Weight change of Distance Transform Loss

Increase the weight → characters becomes thin

[1] Radford, Alec, et al. "Learning transferable visual models from natural language supervision." International Conference on Machine Learning. PMLR, 2021.

[2] Li, Tzu-Mao, et al. "Differentiable Vector Graphics Rasterization for Editing and Learning." ACM Trans. Graph, 2020.