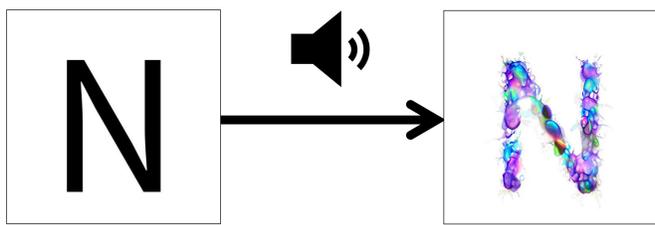


Introduction

- Models that embed sound and images into a shared space have emerged.
- We investigate what kinds of fonts can be created from sound.

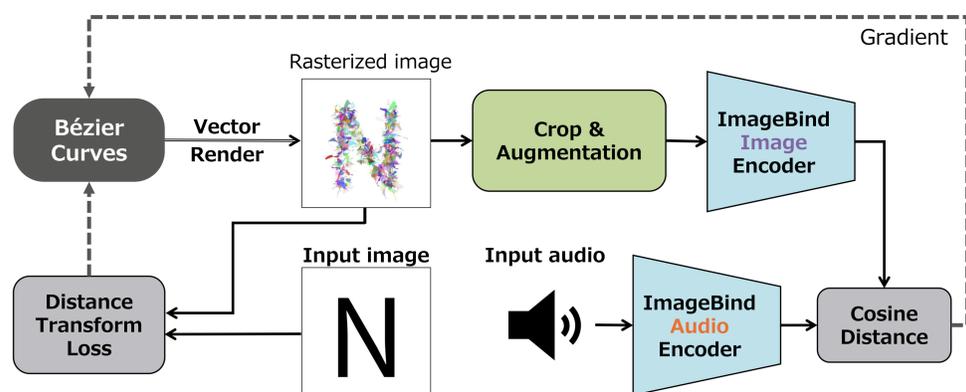
Objective

Font style transfer using sound



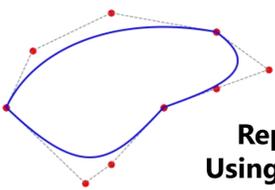
Method

- Based on our previous method[1] for font style transfer from text.
- Optimization of Bézier Curve Parameters (Control Points, colors)
- Rasterizing Bézier Curves Using a Differentiable Renderer[2]
- The loss function is computed using the pre-trained ImageBind[3], where audio and images are encoded and then the cosine distance is calculated.



Drawing Representation

Closed Bézier Curves



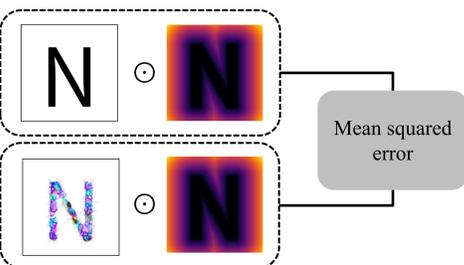
Representation of Characters Using Sets of Closed Bézier Curves

Enlarged view



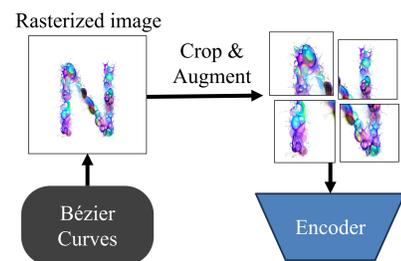
Loss function

- Using the Distance Transform Loss introduced by Atarsaikhan et al.[4] for style transfer to preserve the shape of characters.



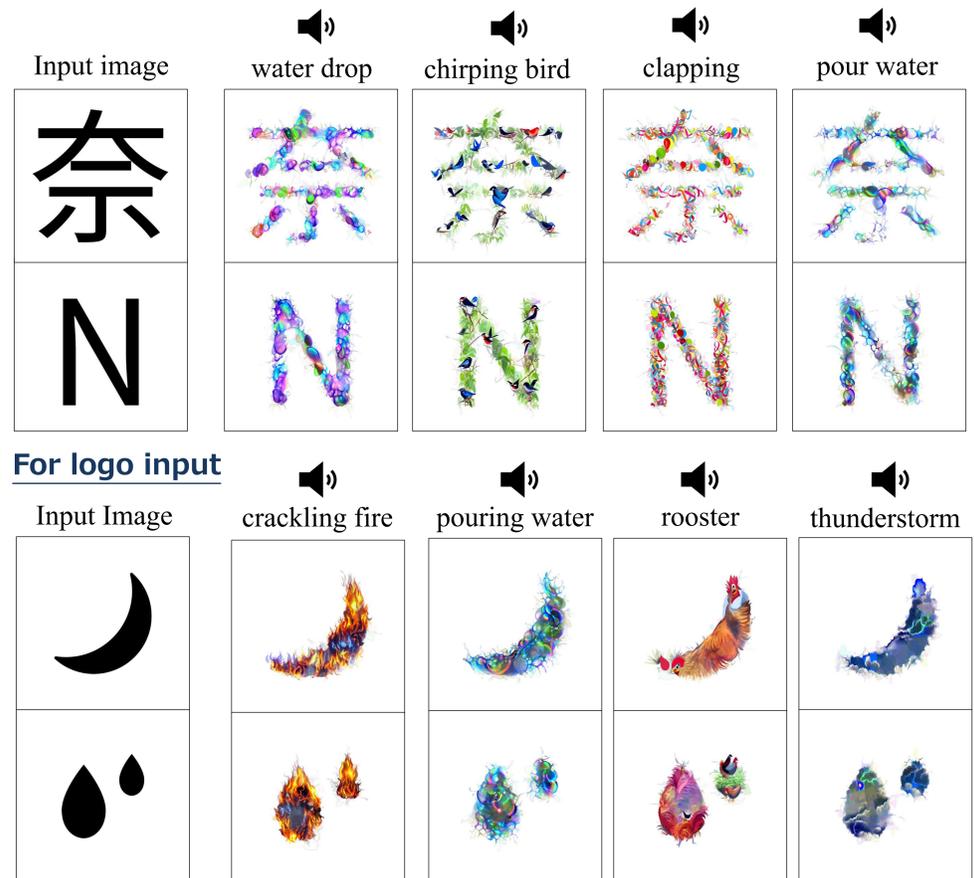
- Calculating the loss for randomly cropped patches and taking the average

→ Transferring style to finer details



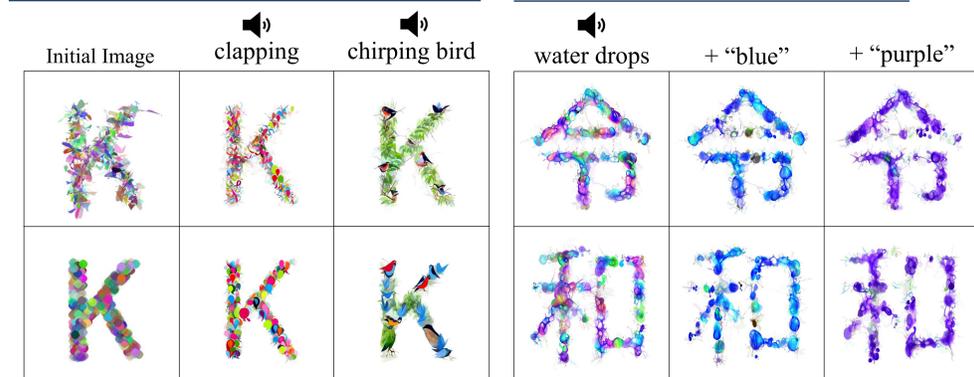
$$L_{total} = \lambda_{dist} L_{dist} + \lambda_{sim} L_{sim}$$

Results

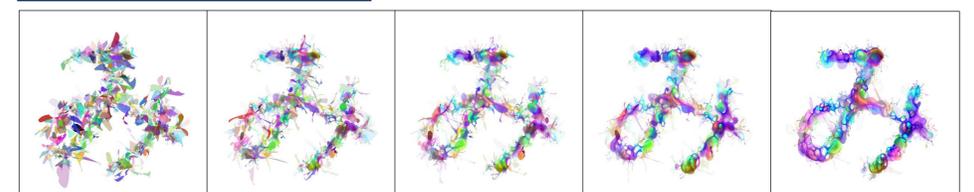


Initial configuration differences

Addition of text conditions

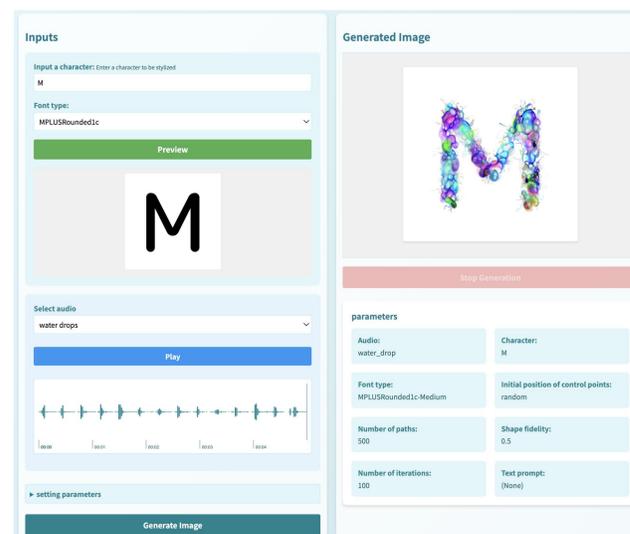


Optimization process



Demo

Style transfer of any character using sound



Website for
[viewing
generated
results](#)



[1] Izumi, K. and Yanai, K. "CLIPFontDraw: Stylizing Fonts With CLIP", IEEE Access, 2024.

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

[3] Girdhar, R., El-Nouby, A., Liu, Z., Singh, M., Alwala, K. V., Joulin, A. and Misra, I.: ImageBind: One Embedding Space To Bind Them All, CVPR2023.

[4] Atarsaikhan et al. Contained neural style transfer for decorated logo generation. DAS 2018.