微分可能レンダラーを用いた ロゴ画像生成

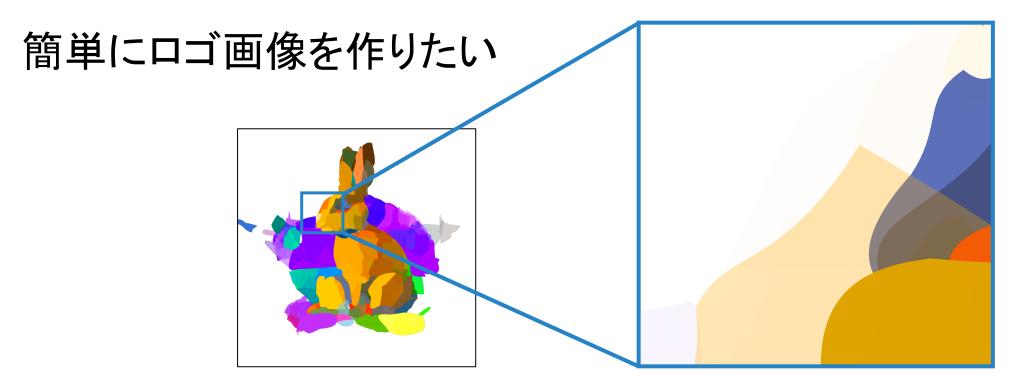
情報理工学域 I 類 メディア情報学プログラム 柳井研究室 2010697 山倉隆太

2024/2/9

研究背景

ベクター画像生成は、ロゴやアイコンなどによく使われる

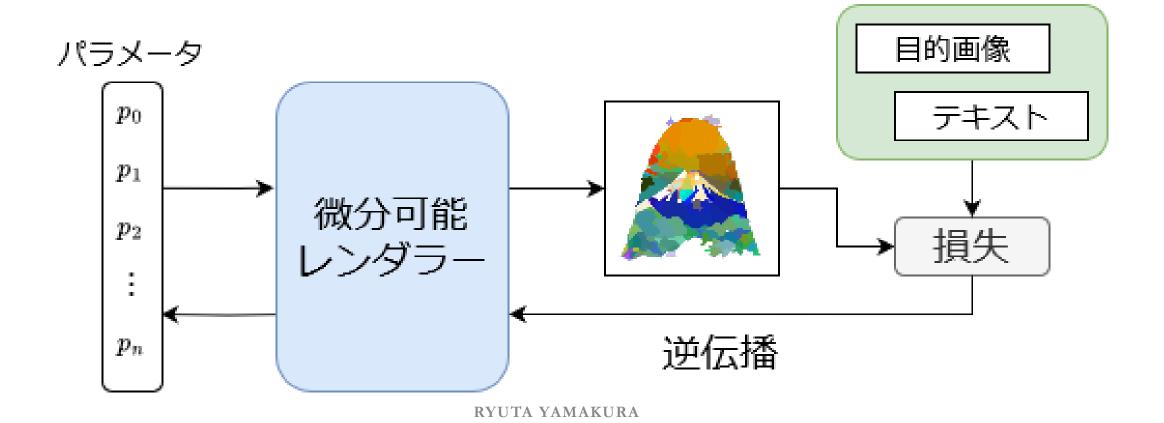
- ・拡大・縮小で画質劣化がない
- ・コンパクトなデータサイズ



ベクター画像の例

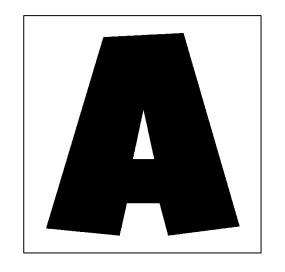
研究背景

・微分可能なレンダリング手法 [4] の登場 レンダリング工程で勾配が発生できるため、 勾配降下法によるベクターパラメータの調整が可能。

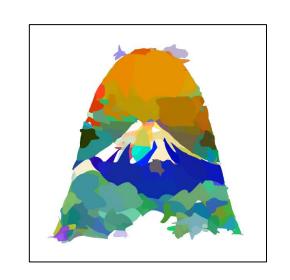


本研究の目的

入力画像の形状を維持して テキストプロンプトの内容を反映した ベクター形式ロゴ画像の生成



+ "Mt.Fuji" =



関連研究

- CLIPDraw [1]
 - ・微分可能レンダラーと大規模言語 画像モデルCLIP [7] を用いて、 テキストからのベクター画像描画を行った。



"A drawing of a cat".



"Horse eating a cupcake".



"A 3D rendering of a temple".



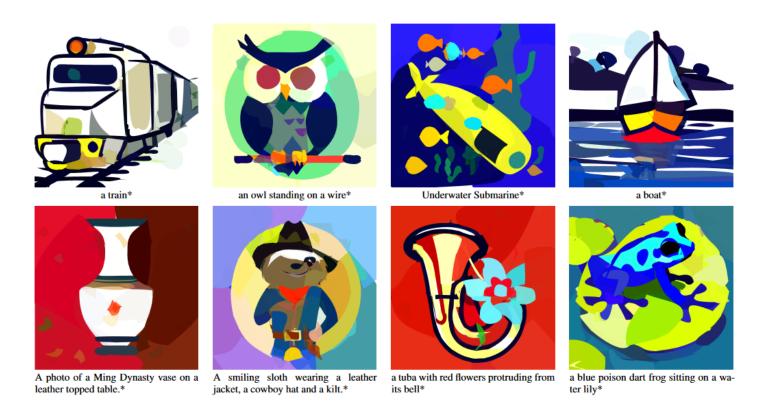
"Family vacation to Walt Disney World".



"Self".

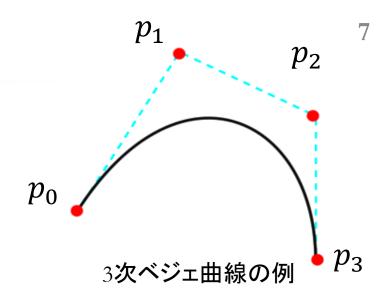
関連研究

- VectorFusion [3]
 - Stable Diffusionを用いて、Score Distillation Sampling Loss [6] を算出し、最適化することでより高品質なベクター画像生成を行った。

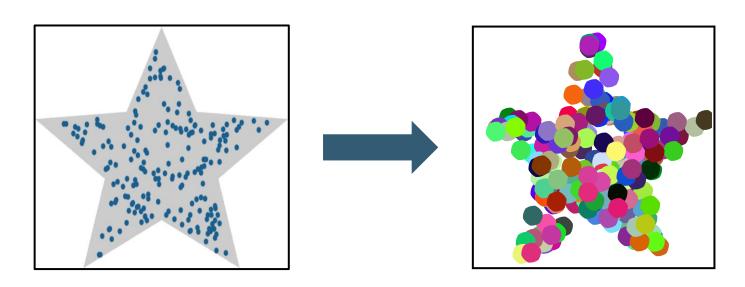


提案手法 – 表現形式と初期化

- ベクター画像を閉じたベジェ曲線で表現する
 - 各曲線はパラメータとして、制御点座標と色を持つ。



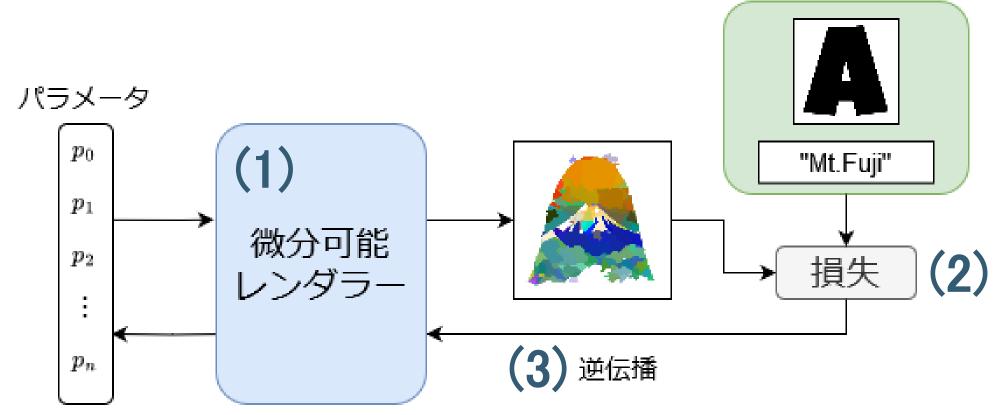
・入力画像の領域内のランダムな座標を中心として、円形に初期化



繰り返し

提案手法 – 概要

- (1) 微分可能レンダラーを用いてラスタライズ
- (2) ラスター画像を用いて損失を計算
- (3) 損失を逆伝播してパラメータを更新

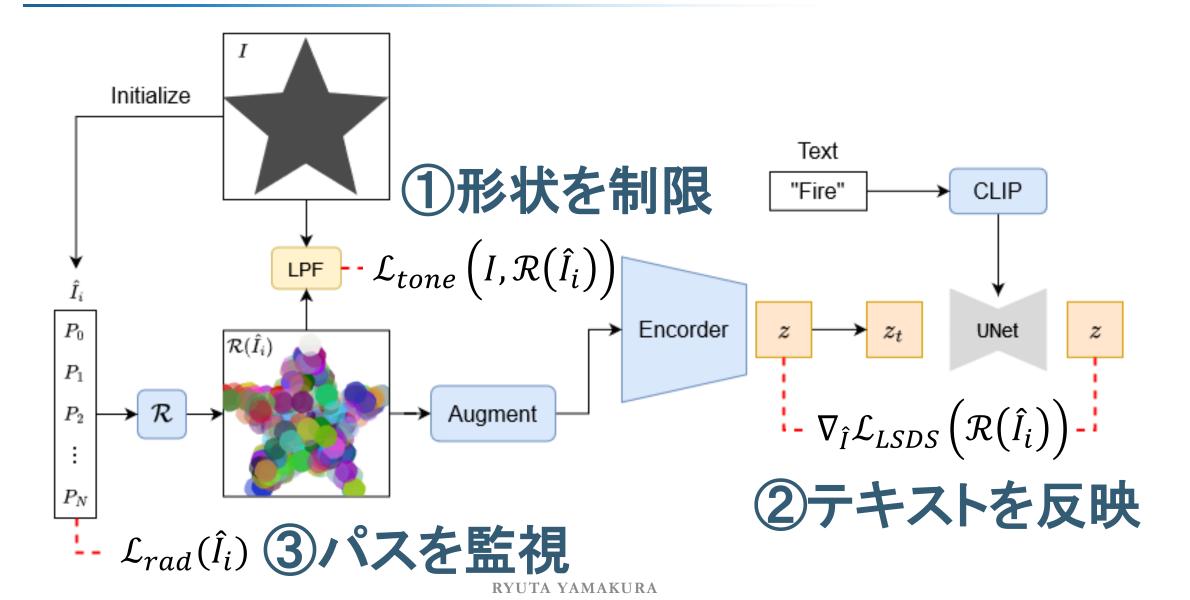


入力画像の形状を維持して テキストプロンプトの内容を反映した ベクター形式ロゴ画像の生成

- ① 入力画像の形に形状を制限
- ② テキストの内容をコンテンツに反映

+

③ ベジェ曲線の形を監視してきれいな出力画像



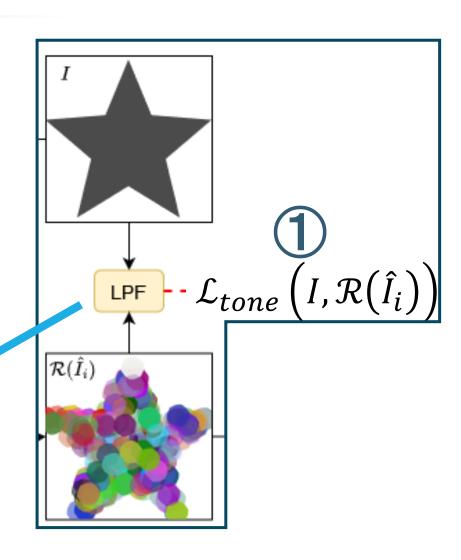
Tone Loss [2]
 入力画像の形状に形を制限する

$$\mathcal{L}_{tone} = \left\| LPF(I) - LPF\left(\mathcal{R}(\hat{I})_{binary}\right) \right\|_{2}^{2}$$

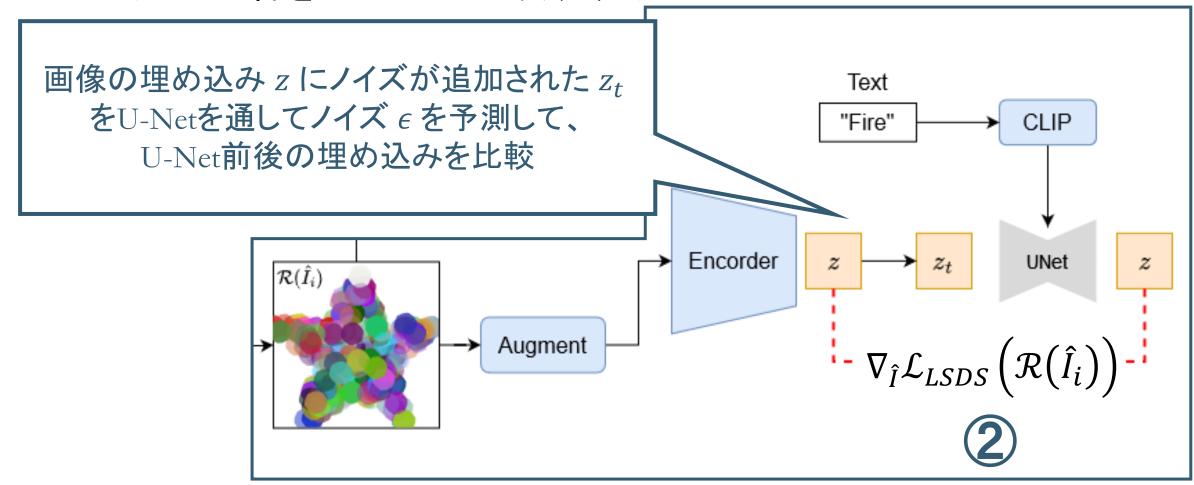
・入力画像と出力画像をローパスフィルタに かけた後 L2 距離を測った損失



ローパスフィルタを適用した例([2]より引用)



② Score Distillation Sampling (SDS) Loss [3] テキスト内容をコンテンツに反映する

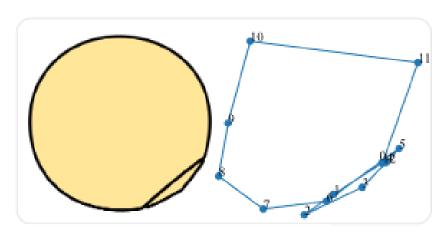


③ Radiation Loss ベジェ曲線パスの形状を安定させる

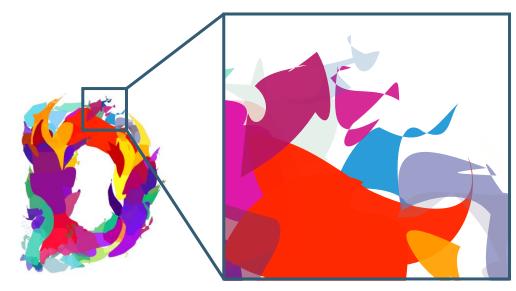
□自己交差問題 [5]

閉じたベジェ曲線において、パスが自分自身と交差することで

不自然な形状が生じる

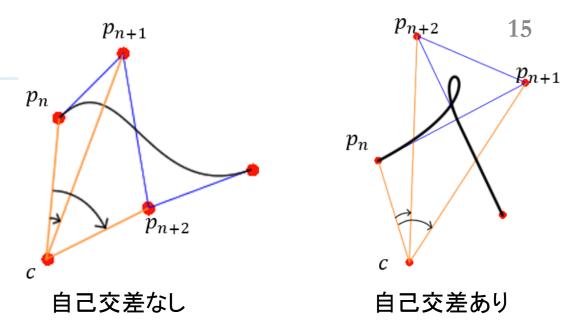


自己交差問題の例([5]より引用)



自己交差によるアーティファクトの例

③ Radiation Loss ベジェ曲線パスの形状を安定させる



自己交差問題

→全ての制御点が回転方向に対して順番に配置されていればいい

$$\mathcal{L}_{rad} = \sum_{n} \text{ReLU}(\angle p_n c p_{n+1} - \angle p_n c p_{n+2})$$

次の制御点の回転角が前の回転角を越したとき損失発生

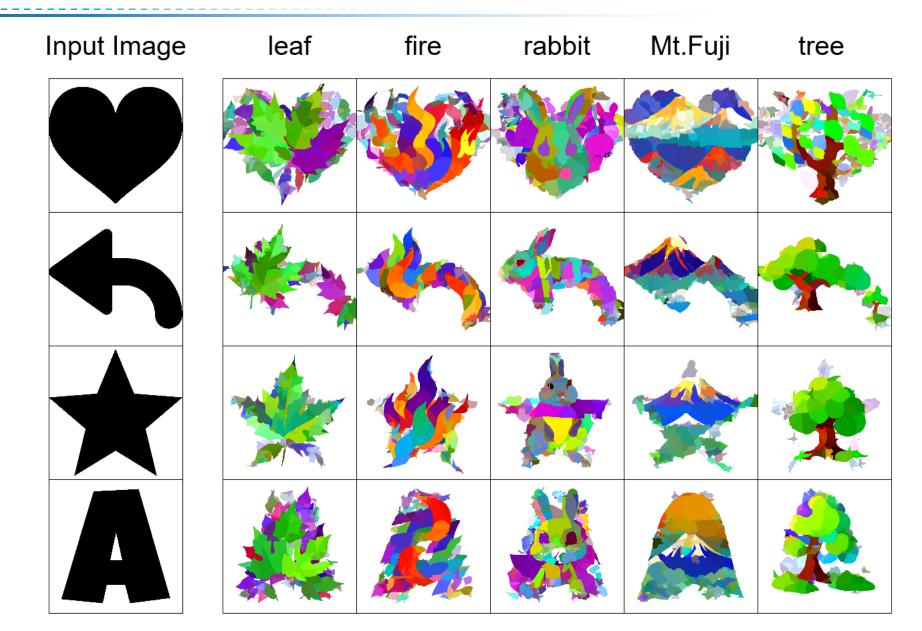
- 1 Tone Loss
- 2 SDS Loss
- 3 Radiation Loss

を合わせて Total Lossとする。

$$\mathcal{L}_{total} = \lambda_{tone} \mathcal{L}_{tone} + \lambda_{rad} \mathcal{L}_{rad} + \lambda_{LSDS} \mathcal{L}_{LSDS}$$

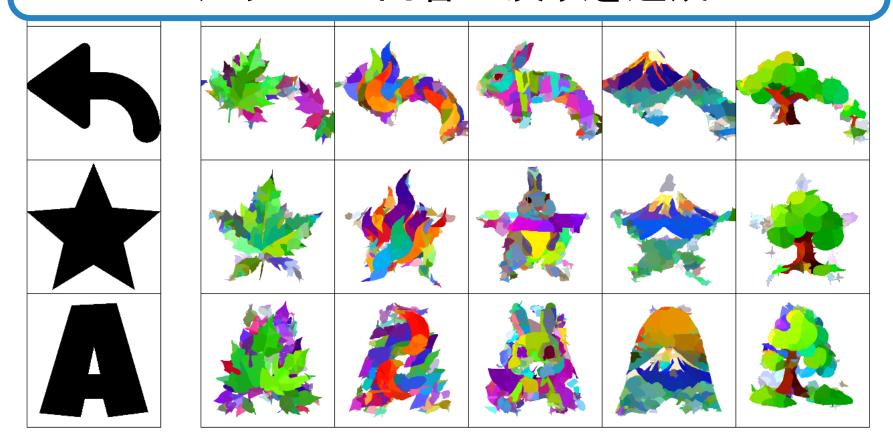
実験 – 設定

- •入出力:600×600
- 更新回数:1000回
- ベジェ曲線の本数:200
- $\lambda_{tone} = 200$, $\lambda_{rad} = 1$, $\lambda_{LSDS} = 1$
- 入力テキスト: "a logo of {コンセプト}. minimal flat 2d vector. lineal color. trending on artstation"
- ・800回目の更新で不要なパスを削除



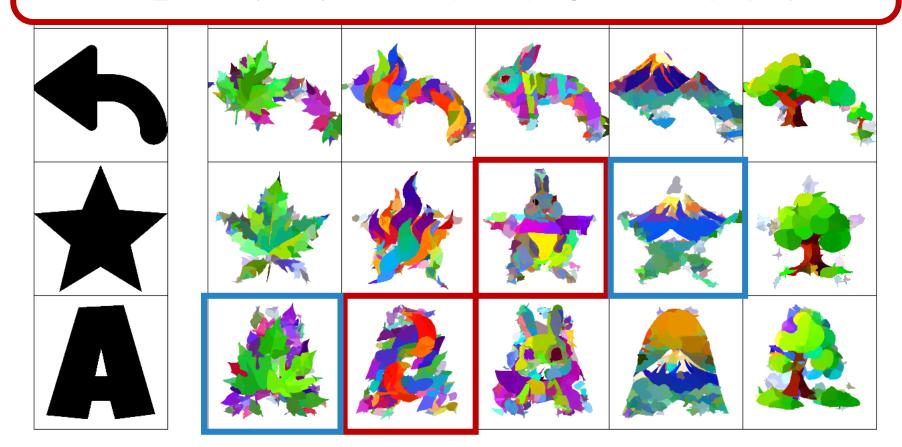
Input Image leaf fire rabbit Mt.Fuji tree

入力画像の形状を維持すること、 テキストの内容の反映を達成



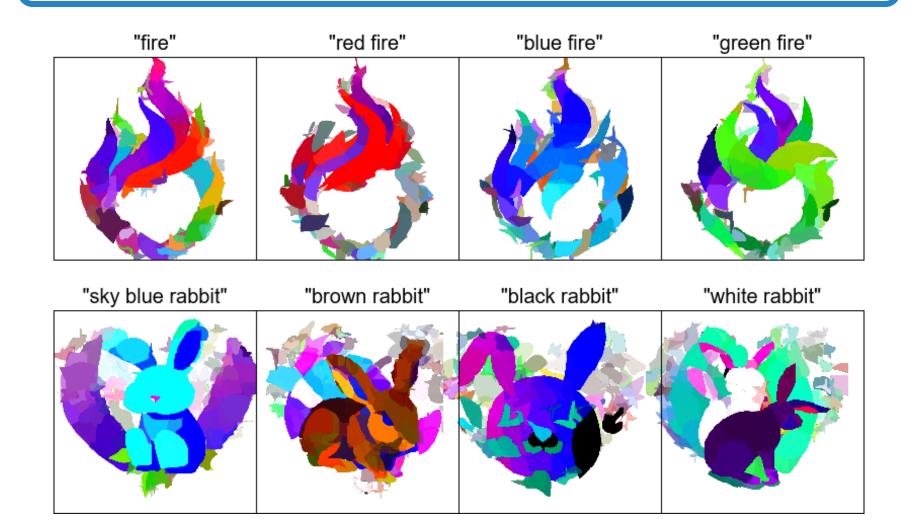
Input Image leaf fire rabbit Mt.Fuji tree

テキストから直感的には<mark>想起されない配色</mark>が見られ、 形状を満たすように不自然な背景パスが存在する



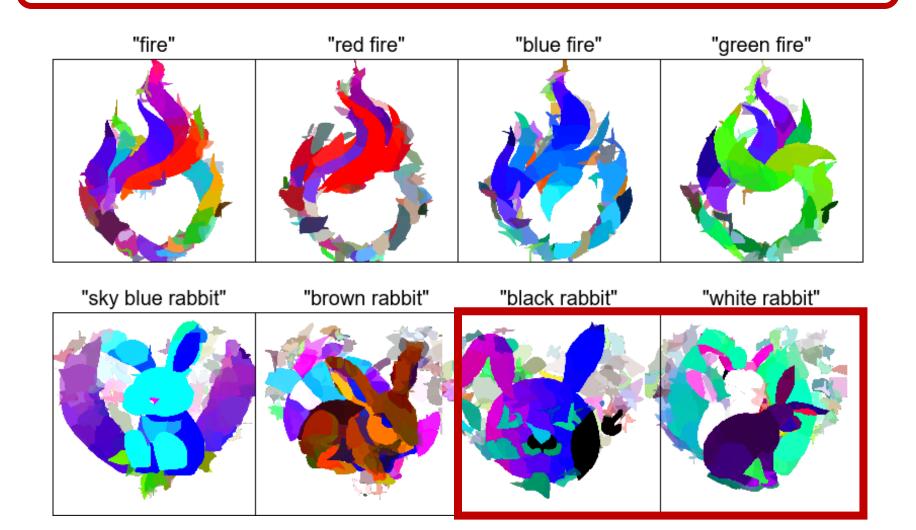
実験 - テキストによる操作

テキストによって色の操作が可能



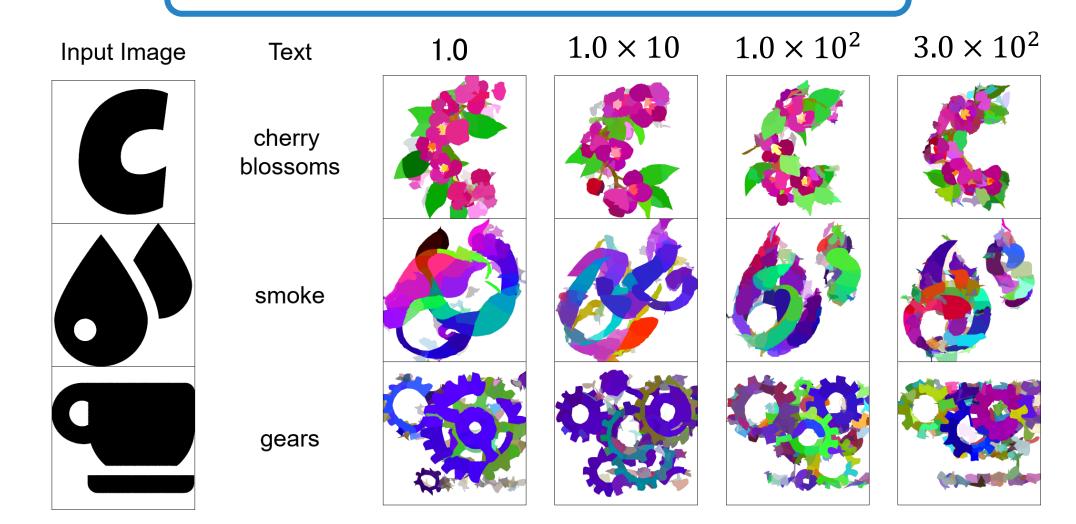
実験 - テキストによる操作

"black", "white"などの無彩色の反応が悪い



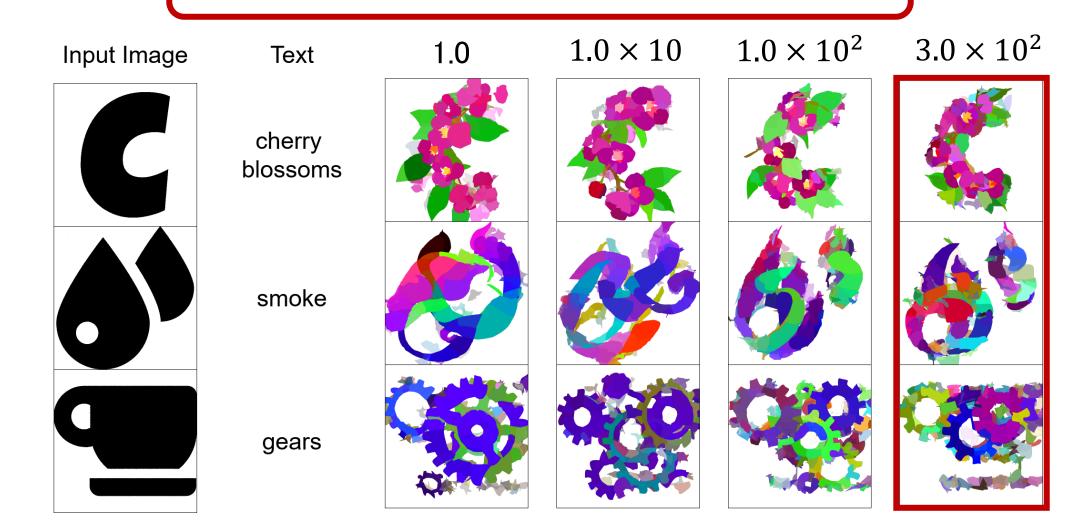
実験 – Tone Loss の重みを変化させた場合

大きいほど入力画像の形状に近づく

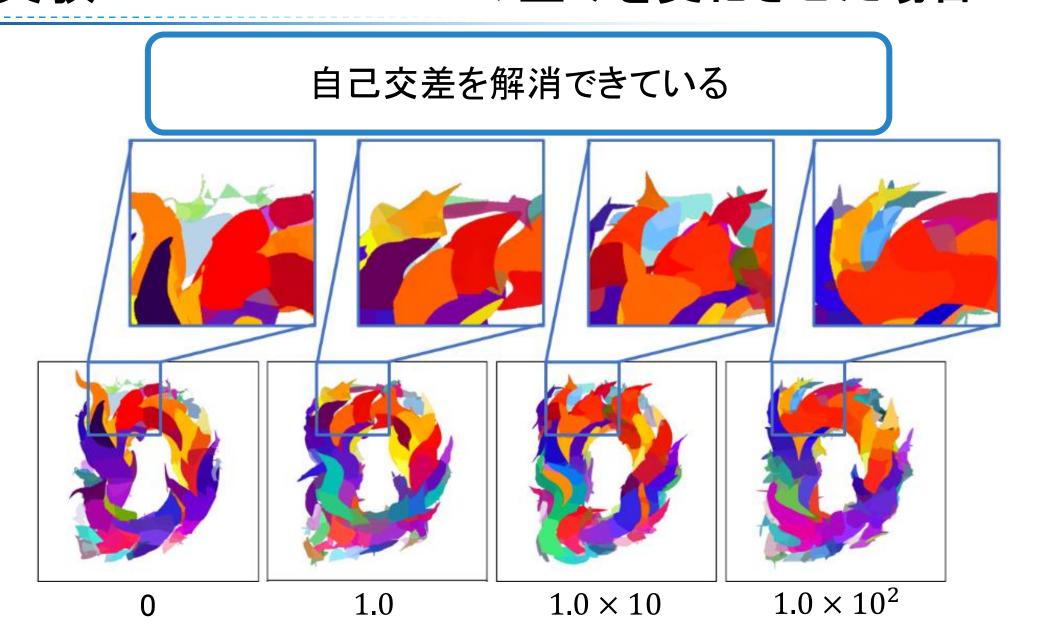


実験 – Tone Loss の重みを変化させた場合

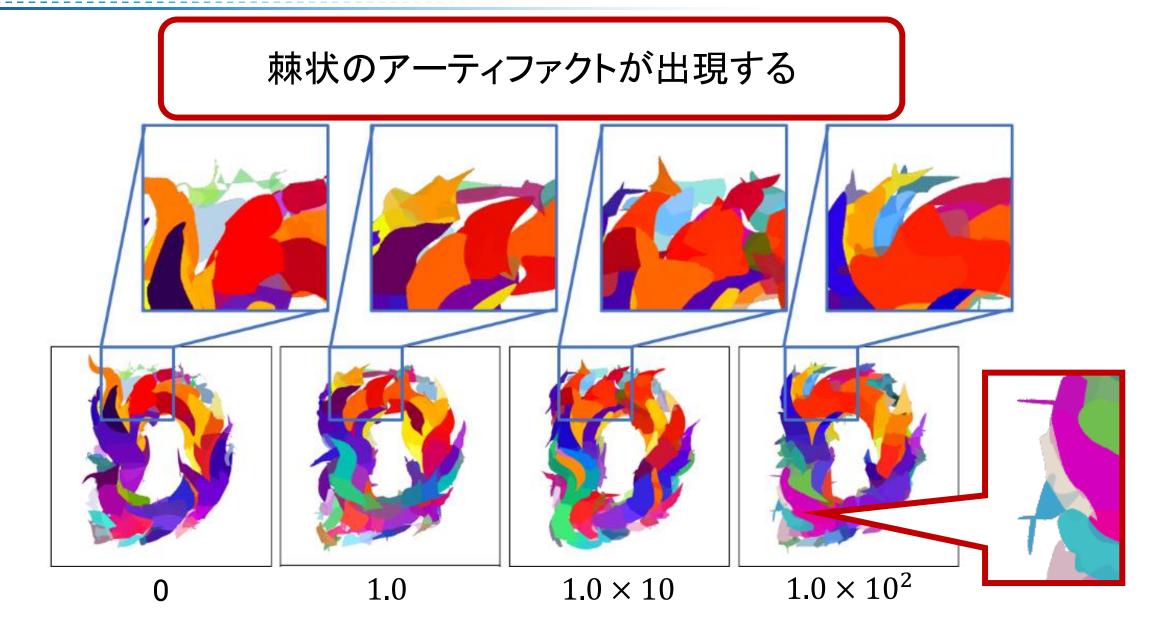
形状が優先されパスが不自然になる



実験 – Radiation Loss の重みを変化させた場合



実験 – Radiation Loss の重みを変化させた場合





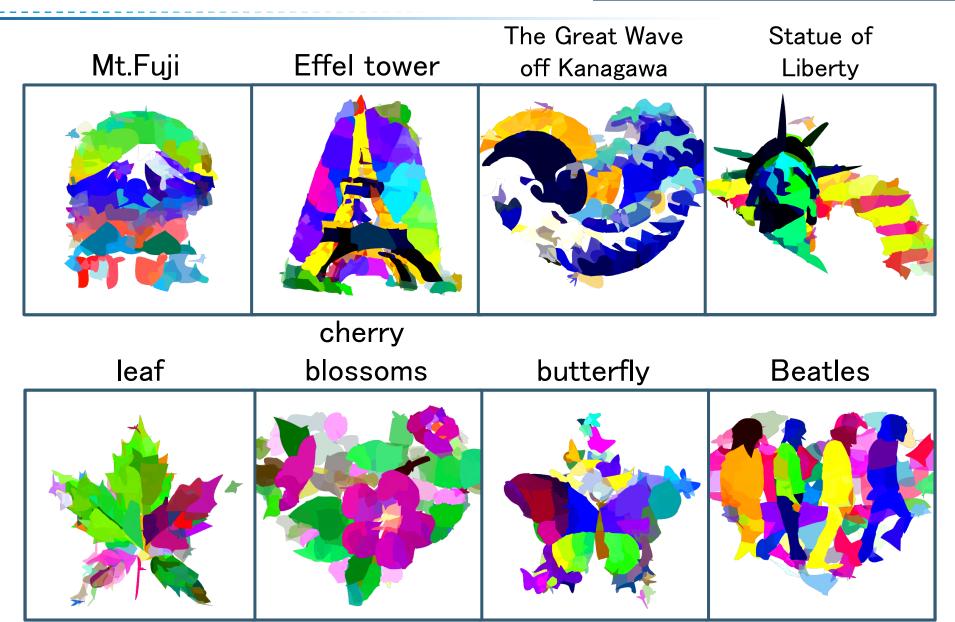
- テキスト+画像入力によるロゴ画像生成
 - ・テキストプロンプトによる入力画像の形状を維持したベクター形式ロゴ画像を、微分可能レンダラーを用いることによって生成する手法を提案
 - 過不足するパス、予期されない配色、棘状のアーティファクトなどの問題

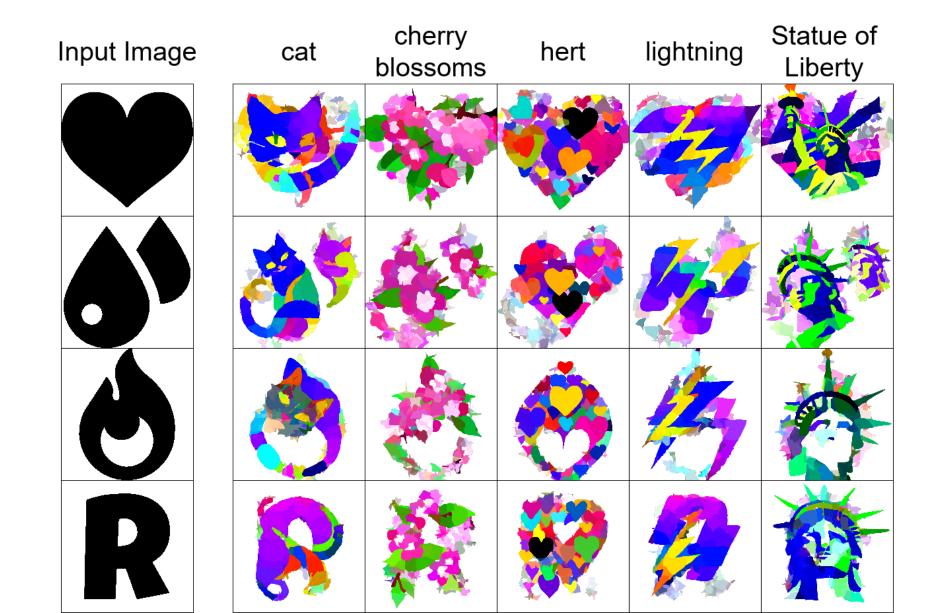
- ・ 今後の展望
 - ・以上の問題の改善に向けて、 拡散モデルに形状画像を条件として用いる方法を考える。



 0
 500
 1000

 繰り返し回数
 1000





参考文献

- [1] Kevin Frans, Lisa Soros, and Olaf Witkowski. CLIPDraw: Exploring text-to-drawing synthesis through language-image encoders. In Advances in Neural Information Processing Systems, Vol. 35, pp. 5207-5218, 2022.
- [2] Shir Iluz, Yael Vinker, Amir Hertz, Daniel Berio, Daniel Cohen-Or, and Ariel Shamir. Word-As-Image for Semantic Typography. ACM Transactions on Graphics, Vol. 42, No. 4, pp. 1–11, 2023.
- [3] Ajay Jain, Amber Xie, and Pieter Abbeel. VectorFusion: Text-to-SVG by Abstracting Pixel-Based Diffusion Models. In 2023 IEEE/CVF Conference on ComputerVision and Pattern Recognition, pp. 1911-1920, 2023.
- [4] T Li, M Luk´a´c, Gharbi M, and Jonathan Ragan-Kelley. Differentiable vector graphics rasterization for editing and learning. ACM Trans. Graph. (Proc. SIGGRAPHAsia), Vol. 39, No. 6, pp. 193:1-193:15, 2020.
- [5] Xu Ma, Yuqian Zhou, Xingqian Xu, Bin Sun, Valerii Filev, Nikita Orlov, YunFu, and Humphrey Shi. Towards layer—wise image vectorization. Procs. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16314–16323, 2022.