

[S1-P26] 潜在拡散モデルによる生成対象の配置制御による生成画像改善の試み

永井 大地¹, 根本 颯汰², 北田 俊輔², 彌富 仁^{1,2} ¹法政大学 理工学部, ²法政大学大学院 理工学研究科

{daichi.nagai.9s@stu., iyatomi@}hosei.ac.jp



Summary

拡散モデルの初期ノイズに注目した生成に有望な種を移植する Franken Latent Stein の提案

- 有望な種は“エッジ”が多数存在することが重要であることを特定し、その箇所を潜在空間へ移植
- Stable Diffusion (SD) をベースラインに提案手法を適用
→ プロンプトに忠実な画像生成を実現

Background

複雑なプロンプトに対する画像生成の困難性

- 😊 プロンプトを元にした画像生成の進歩
- 😞 複雑なプロンプトに対して生成が困難 [Chefer+ SIGGRAPH'23]
→ 物体や属性が正しく生成されない
- 😞 学習データの強化が有効であるがその用意やモデルの学習が高コスト
→ 追加学習不要な改善手法に注目



“photo of cat and dog in grassland”

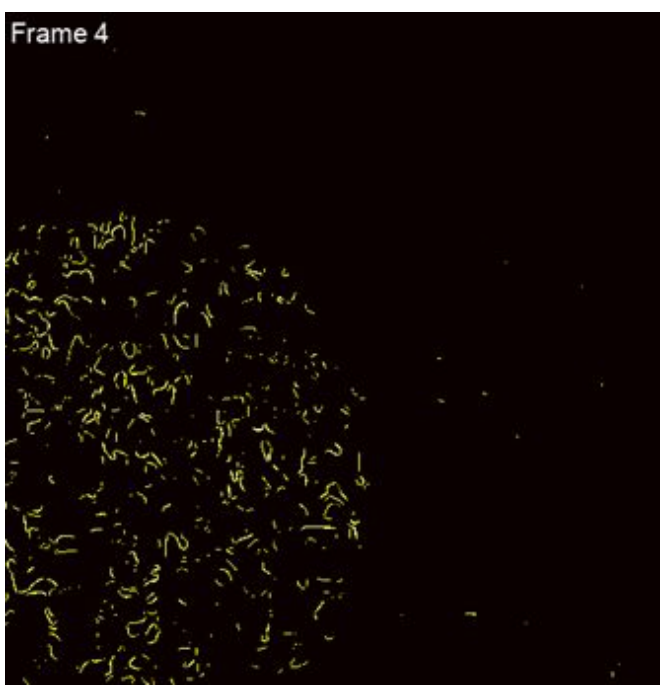
良いノイズの種からの画像生成

- 良い画像は良い初期ノイズから [Mao+ MM'23, ECCV'24]
→ 良い“種”を移植することで属性や配置を制御

Franken Latent Stein

事前分析

- 拡散過程初期のノイズ画像と生成画像のエッジには強い関係がある



左下以外のエッジを除去

良い種の取得

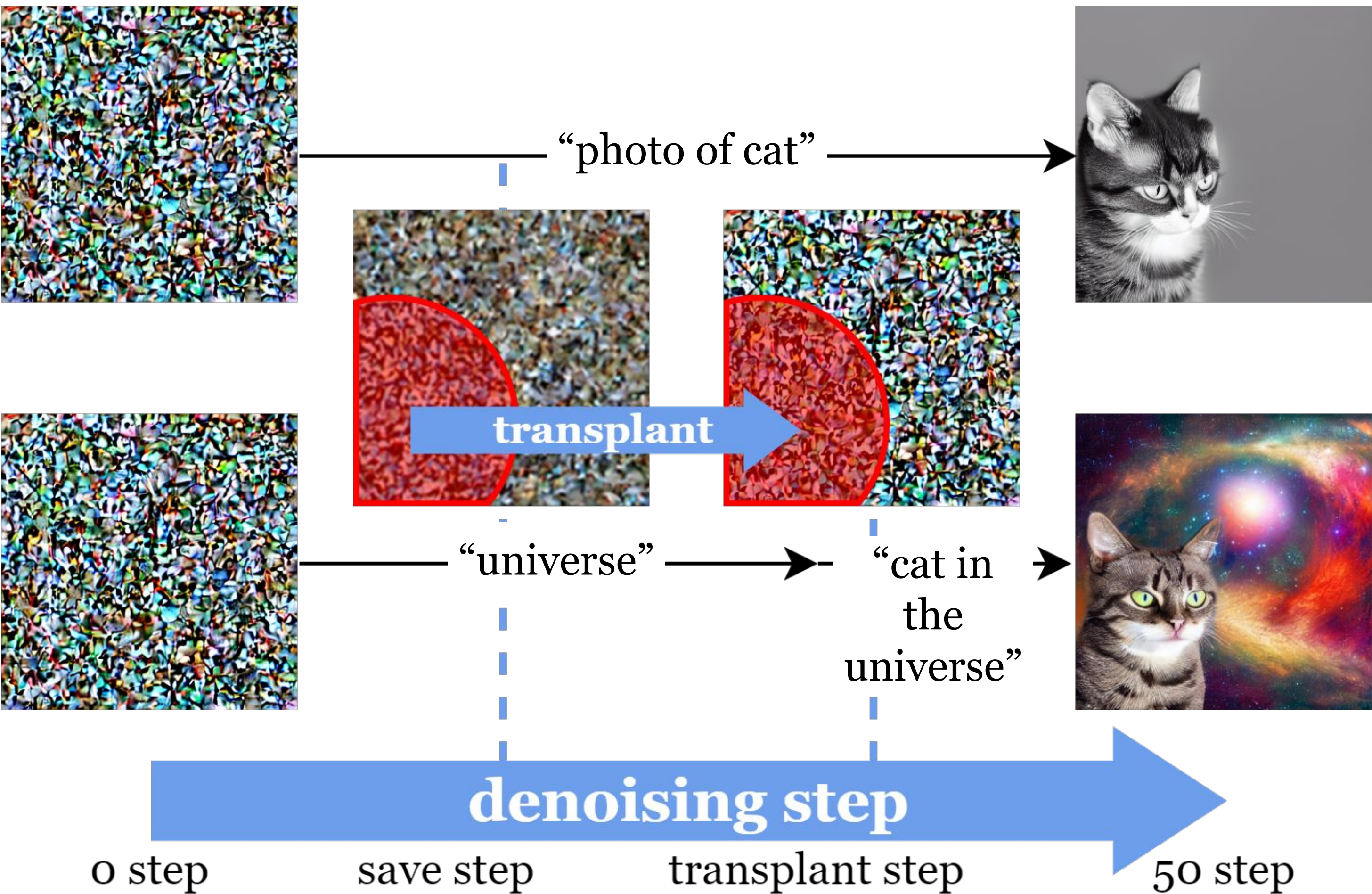
- 指定領域外のエッジを除去しエッジの分布を指定領域に限定
→ 生成対象を指定領域に誘導



生成画像

種の移植

- 指定領域内に成長した種を生成
 - 20ステップ以内のものを取得
- 生成した種を他の初期ノイズの潜在表現へ移植
 - 数ステップ逆拡散過程を経た状態
 - 生成対象の配置を制御可能・無地の背景を回避



Experiments & Results

実験設定

- ベースラインモデル: SD v1.5
- full プロンプト: “*obj₁* (and *obj₂*) in *background*”
- 手順と使用したプロンプト
 - “photo of *obj₁*” にて *obj₁* の種を生成
 - “(*obj₂* in) *background*” で生成を開始した潜在表現の左下に種を移植し生成継続、プロンプトを full に変更
 - “cat in the beach”
 - “cat and motorcycle in snow field”
 - “dog and red car in grassland”
 - “zebra and bus in city”

obj₁
obj₂
background

評価方法

- 各プロンプトで生成した画像各 100 枚に対し、YOLOv8 [Varghese+ ADICS'24] にて生成対象を認識できるか確認、生成対象が1 つの場合、その位置も確認
- 各画像の審美性スコア (LAION Aesthetics) を算出

実験結果

生成対象が 1 つの場合の結果

	existence ↑	position ↑	aesthetics score ↑
SD 1.5 (baseline)	0.95	(0.35)	6.43
SD 1.5 (proposed)	0.71	0.99	6.57

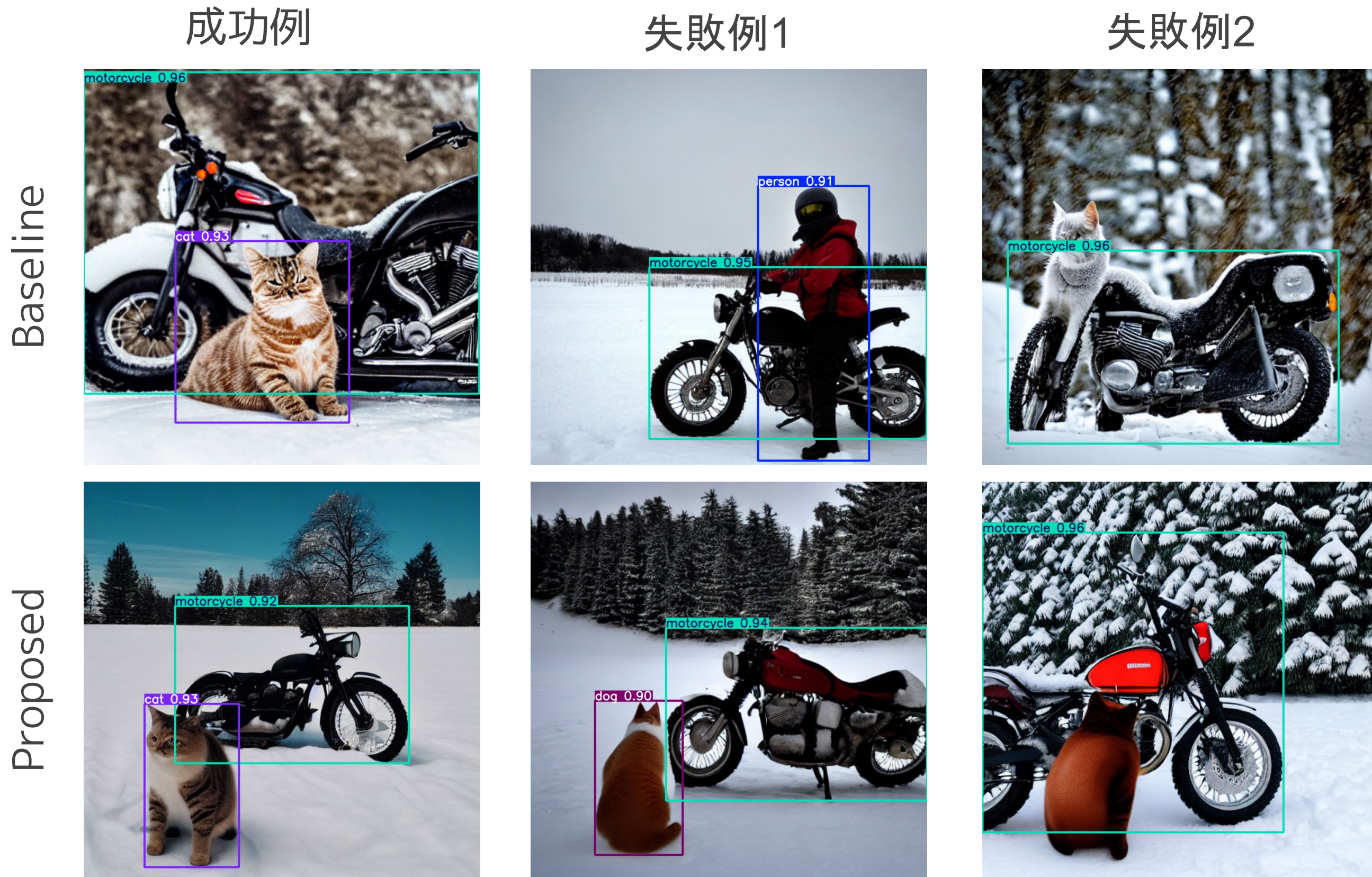
existence : *obj₁* が存在する割合
position : *obj₁* が存在する時、指定した位置に生成される割合
aesthetics score : 画像の美しさを表す指標

生成対象が 2 つの場合の結果 (平均)

	1-object ↓	2-object ↑	aesthetics score ↑
SD 1.5 (baseline)	0.56	0.43	6.55
SD 1.5 (proposed)	0.31	0.68	6.70

1-object : *obj₁* と *obj₂* のいずれか1つのみが生成される割合
2-object : *obj₁* と *obj₂* が両方とも生成される割合
aesthetics score : 画像の美しさを表す指標

各モデルによる1つ目のプロンプトの生成例



Discussion & Future Work

- 移植した種がどれほど対象物体の情報を持つか要調査
- エッジの少ない物体の誘導が難しい場合がある
- 移植タイミング等のハイパーパラメータが生成対象物体にどれほど依存するか要検討
- Diffusion Transformer ベースの画像生成モデル SD 3 や FLUX.1 に対する有効性の確認

