

# 拡散モデルを用いたパッチ単位の任意スケール画像生成

荒川 深映<sup>1</sup>

Erik Härkönen<sup>2</sup>

綱島 秀樹<sup>1</sup>

堀田 大地<sup>3</sup>

森島 繁生<sup>4</sup>

<sup>1</sup>早稲田大学

<sup>2</sup>Aalto University

<sup>3</sup>東京大学

<sup>4</sup>早稲田大学理工学術院総合研究所

## はじめに

- 拡散モデル<sup>[1]</sup>は高品質・多様な画像の生成に成功
- ノイズ画像から繰り返しデノイズを行うことで画像を生成

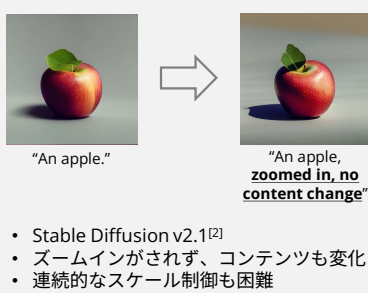


一方で、生成画像内のオブジェクトの**スケール**を制御することが困難

### 解像度とスケール



### 既存手法でスケールを変更



💡 拡散モデルは「**スケール**」と「**位置**」に対応した画像を生成可能か？

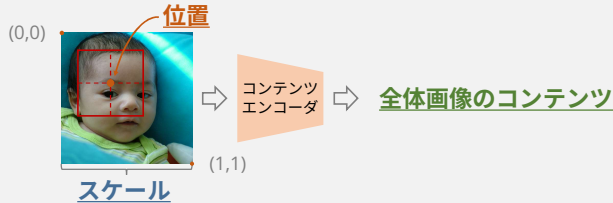
## 関連研究

- DiffCollage<sup>[3]</sup>
  - パッチをオーバーラップ
  - パッチ同士の接続関係の条件を基にデノイズ ➡ **スケール**の変更は不可
  - あらゆる解像度の画像を生成可能
- AnyresGAN<sup>[4]</sup>
  - 平行移動等価性を持つStyleGAN3<sup>[5]</sup>を任意解像度画像で学習
  - スケールと位置に対応する画像表現を獲得
  - 生成されたパッチをつなぎ合わせる⇒高解像度画像(> 8192)を生成可能

## 提案手法

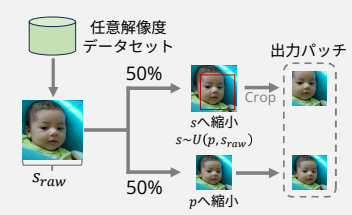
### ネットワーク

- ベースライン: Diffusion Autoencoder<sup>[6]</sup>
- AdaGNにより**位置** & **スケール** & **全体画像のコンテンツ**を条件付け

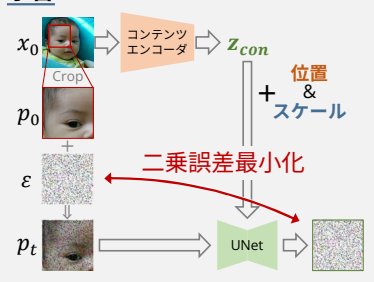


### 任意解像度学習<sup>[4]</sup>

- 画像を確率的に縮小
- パッチサイズ $p$ は常に一定

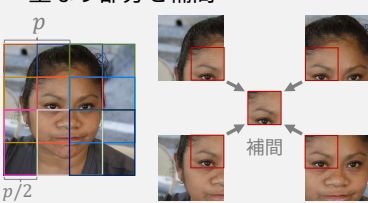


### 学習



### 任意解像度サンプリング

- パッチをつなぎ合わせる
- 重なりを持つパッチを複数生成
- 重なり部分を補間

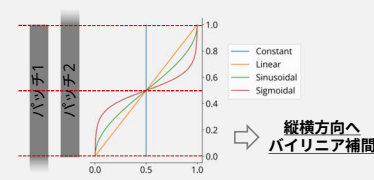


DDIM サンプリング方程式<sup>[7]</sup>

$$p_{t-1} = \sqrt{\alpha_{t-1}} \left( \frac{p_t - \sqrt{1 - \alpha_t} \epsilon_\theta(p_t)}{\sqrt{\alpha_t}} \right) + \sqrt{1 - \alpha_{t-1} - \sigma_t^2} \cdot \epsilon_\theta(p_t)$$

Predicted  $x_0$

### 重み付きバイリニア補間

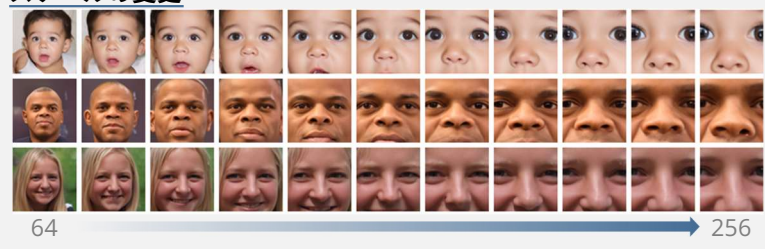


## 実験

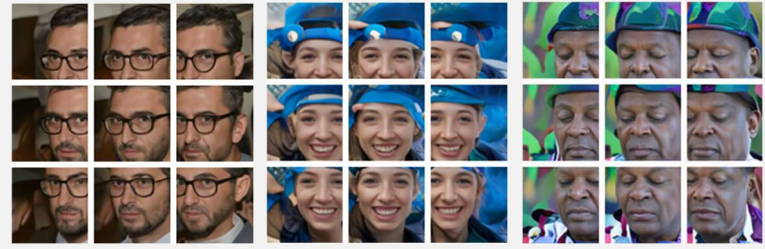
### 実験設定

- 顔画像を集めたFFHQ-256<sup>[8]</sup>データセット
- パッチサイズ $p$ : 64, 最大解像度: 256

### スケールの変更



### 位置の変更



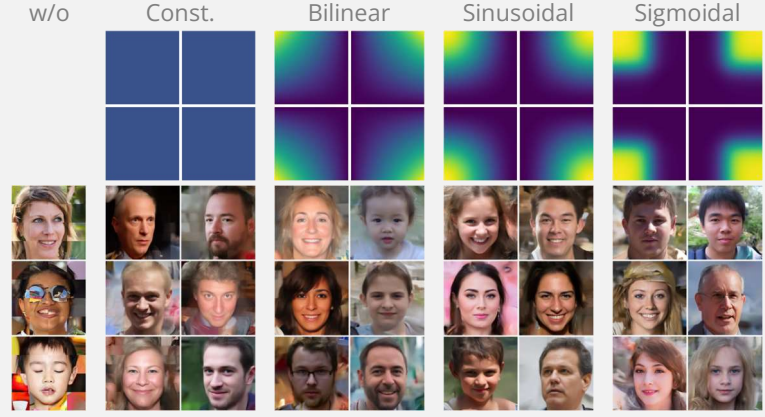
- 位置**と**全体画像のコンテンツ**が特にもつれている  
⇒ 解像度の変更を行った際のパッチ間の不整合につながる  
⇒ 平行移動の正確性をより重要視する損失関数が必要か

### 解像度の変更



- 高解像度の画像は、低解像度の画像と比べて**高周波な情報**を含む
- 高解像度の画像は、**背景と顔の境界付近に不自然さ**が残る

### 補間手法の比較



## 結論

- オブジェクトのスケールを制御可能な拡散モデルを提案
- 位置**と**スケール**に応じたパッチ生成により、あらゆる解像度の画像を生成
- サンプリング手法により、後処理的にパッチ間の不整合を修復  
⇒ パッチ間の不整合が起こる本質的な問題を解決する必要がある

### 引用文献

[1] Ho et al., "Denoising diffusion probabilistic models," NeurIPS, 2020.  
[2] Rombach et al., "High-Resolution Image Synthesis With Latent Diffusion Models," CVPR, 2022.  
[3] Zhang et al., "DiffCollage: Parallel Generation of Large Content with Diffusion Models," CVPR, 2023.  
[4] Chai et al., "Any-resolution Training for High-resolution Image Synthesis," ECCV, 2022.  
[5] Karras et al., "Alias-Free Generative Adversarial Networks," NeurIPS, 2019.  
[6] Preechakul et al., "Diffusion Autoencoders: Toward a Meaningful and Decodable Representation," CVPR, 2022.  
[7] Song et al., "Denoising Diffusion Implicit Models," ICLR, 2021.  
[8] Karras et al., "A Style-Based Generator Architecture for Generative Adversarial Networks," CVPR, 2019.