

特定の作家に酷似した  
顔アイコンを生成する  
拡散モデルの分析結果  
(プレプリント版)

尾崎安範



# アジェンダ

- 現象の発生
- 現象の具体的な再現方法
- 現象の分析のための実験
- 現象に対する考察
- 関連技術や先行事例との位置づけ

# 現象の発生

- 拡散モデルを使って、キメラみたいな新しい「いらすとや」のアイコンができないか、実験を繰り返していたところ、再現可能な方法を見つけてしまった



※これらはすべて拡散モデルを用いて作られた顔アイコン。キャラはいないし、作家も実在しない。

# 現象の具体的な再現方法

1. 拡散モデルのライブラリ[1] と、いらすとやの顔アイコンをダウンロードしてくる
2. 拡散モデルのサンプルコードの中にあるU-Netのパラメータ `dim_mults` を `(1,2,4,8)` を `(1,16)` などにする
3. サンプルコードを動かし、顔アイコンの生成モデルを作る
4. 作った生成モデルから出力される画像を選び出す

```
model = Unet(  
    dim = 64,  
    dim_mults = (1, 2, 4, 8)  
).cuda()
```



```
model = Unet(  
    dim = 64,  
    dim_mults = (1,16)  
).cuda()
```

[1] <https://github.com/lucidrains/denoising-diffusion-pytorch>

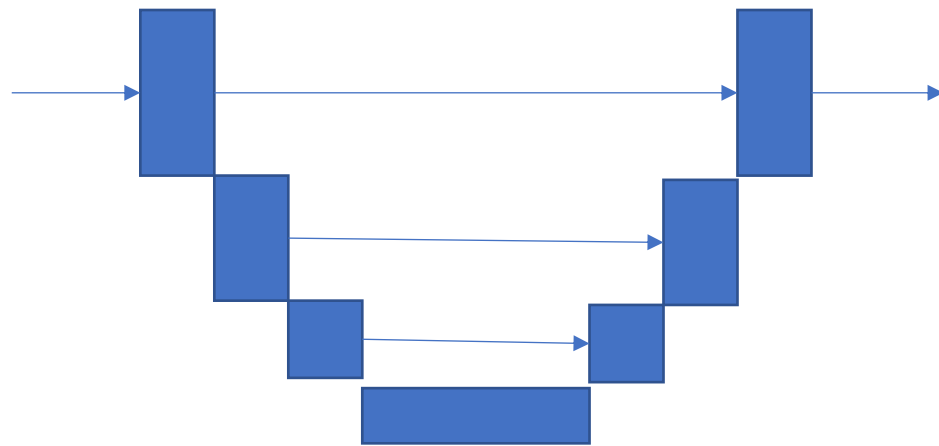
なぜこんなことをして発見したのか、  
拡散モデルの原理から説明する

# 拡散モデルの概要

- 拡散モデルとは、ガウシアンノイズで構成されるノイズ画像からノイズを取り除くことで画像を生成する技術
  - 厳密にはノイズ除去の最中にもノイズを付与している
- 画像を生成するとき、ノイズ除去方法として、**U-Net**を用いる

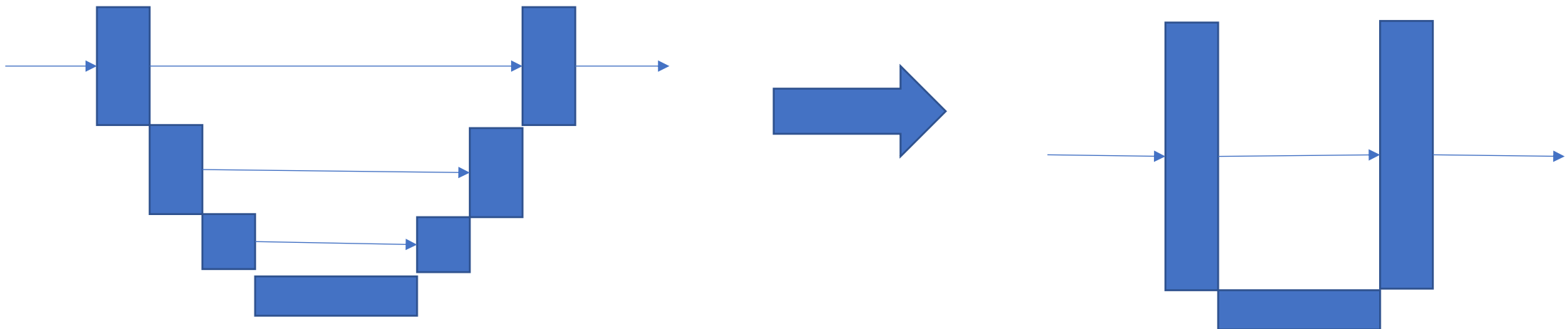
# U-Netの概要

- 線対称な砂時計型ニューラルネットワークであり、レイヤーごとのskip connectionを特徴とする。
- よくセマンティックセグメンテーションで用いられる



# パラメータを操作すると何が起きるのか

- 極端な砂時計型ニューラルネットワークが生成され、画像群から意味の単位を理解し、意味のある単位で分解し、単位のレイアウトがでたらめになるように画像を再構築する、との仮説を私は立てている
  - 「理解、分解、再構築は錬金術の基本」





# 現象の分析のための実験

- 目的

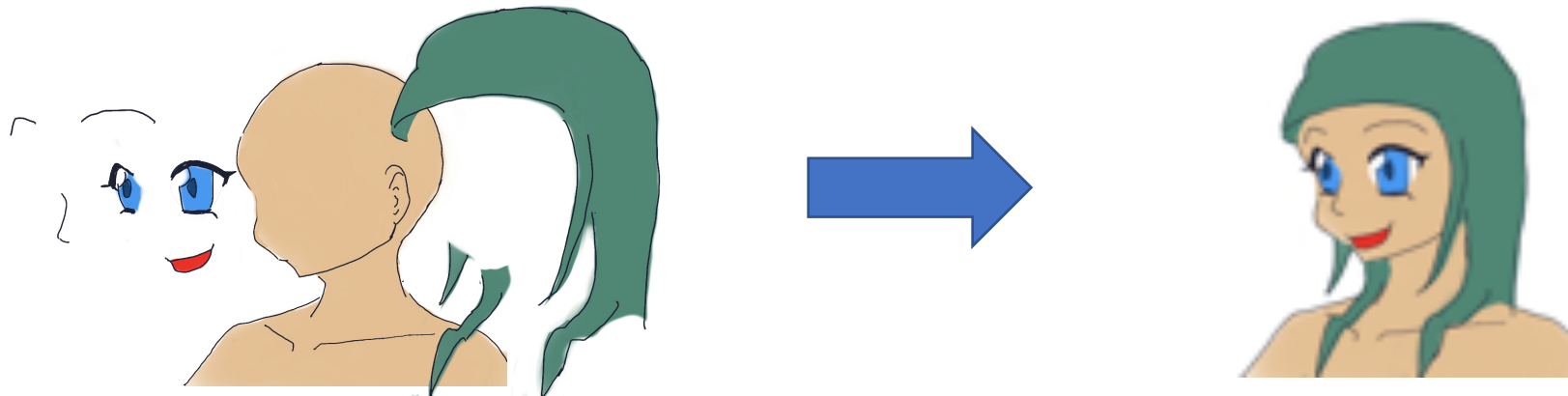
- “画像群から意味の単位を理解し、意味のある単位で分解し、単位のレイアウトがでたらめになるように画像を再構築する”との仮説を検証するため

- 手順

1. 意味のある顔のパーツごとに、レイヤーを分けた顔アイコンデータセットを自分で書いた絵で構築する
2. 顔のパーツをアルファブレンディングで合成してたくさんのパターンを作成する
3. 例の拡散モデルで学習し、顔アイコンを生成する
4. 仮説通りに再構築しているか定性的に確認する

# 顔アイコンデータセット

- 顔のパーツをレイヤーごとに分離して保持して描かれている
  - 普通のイラストレーターはそうしている
- それらのパーツをアルファブレンディングで組み合わせて、顔アイコンを作成する
- パーツの直積分できるので簡単に増やせる
  - 実験時200枚ぐらい



# 実験結果

- おおよそレイヤーのとおりに分解され、再構築されていた
  - おそらくパーツの意味を理解していると思われる
    - ただし、教師なしであるため、パーツが中途半端に分解されている例もあったし、パーツごとで合成されているものもあった

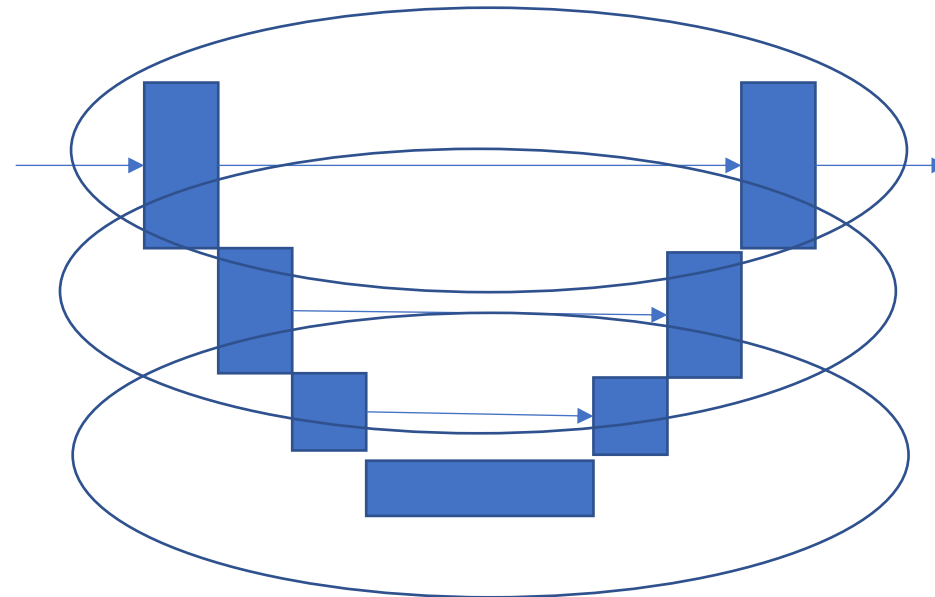


# 現象に対する考察: 拡散モデルについて

- 拡散モデルに使われているU-Netは  
ノイズ除去や  
セマンティックセグメンテーション、  
インペインティング、  
オートエンコーディングなどができる  
万能なTransformerである。
  - はっきりいって演算量がおかしい
    - 普通のTransformerの数十倍
- 拡散モデルはこのTransformerでぶっちゃけ支えられている

# 現象に対する考察: U-Netについて

- このU-Netはおそらく下の図のような働きを持っていて、今回、レイアウトを司る部分をロックアウトしたために、キメラが作成されるようになった



パーツという記号を  
ピクセルに変換する部分

パーツのレイアウトを  
司る部分

パーツの概念を  
司る部分

# 現象に対する考察: 応用可能性

- 活用した場合

- アーティストの創作を支援を量的にも質的にも支援ができる
  - 量的：微妙に違う創作物の生成
  - 質的：新しい組み合わせによるアイデアの創出

- 悪用した場合

- 他人の著作物から無断で二次創作物を作り、他人へ不利益を与えることが可能である
  - 著作権法違反の可能性がある
- すでに書かれている他人の署名から新しい署名の偽造が可能である
  - 有印私文書偽造罪などの可能性がある

# 関連技術や先行事例との位置づけ

- 特定作家に類似したイラストの生成技術
  - スタイルトランスファー
  - GAN
- 入力コストの低さと出力の品質にまさるものがない

手法名	入力のコストの低さ	出力の品質	制御可能性
スタイルトランスファー	×	○	◎
GAN (DCGAN, WGAN)	◎	×	○
拡散モデル	◎	○	△