Stable DiffusionとControlNet

2023/05/09

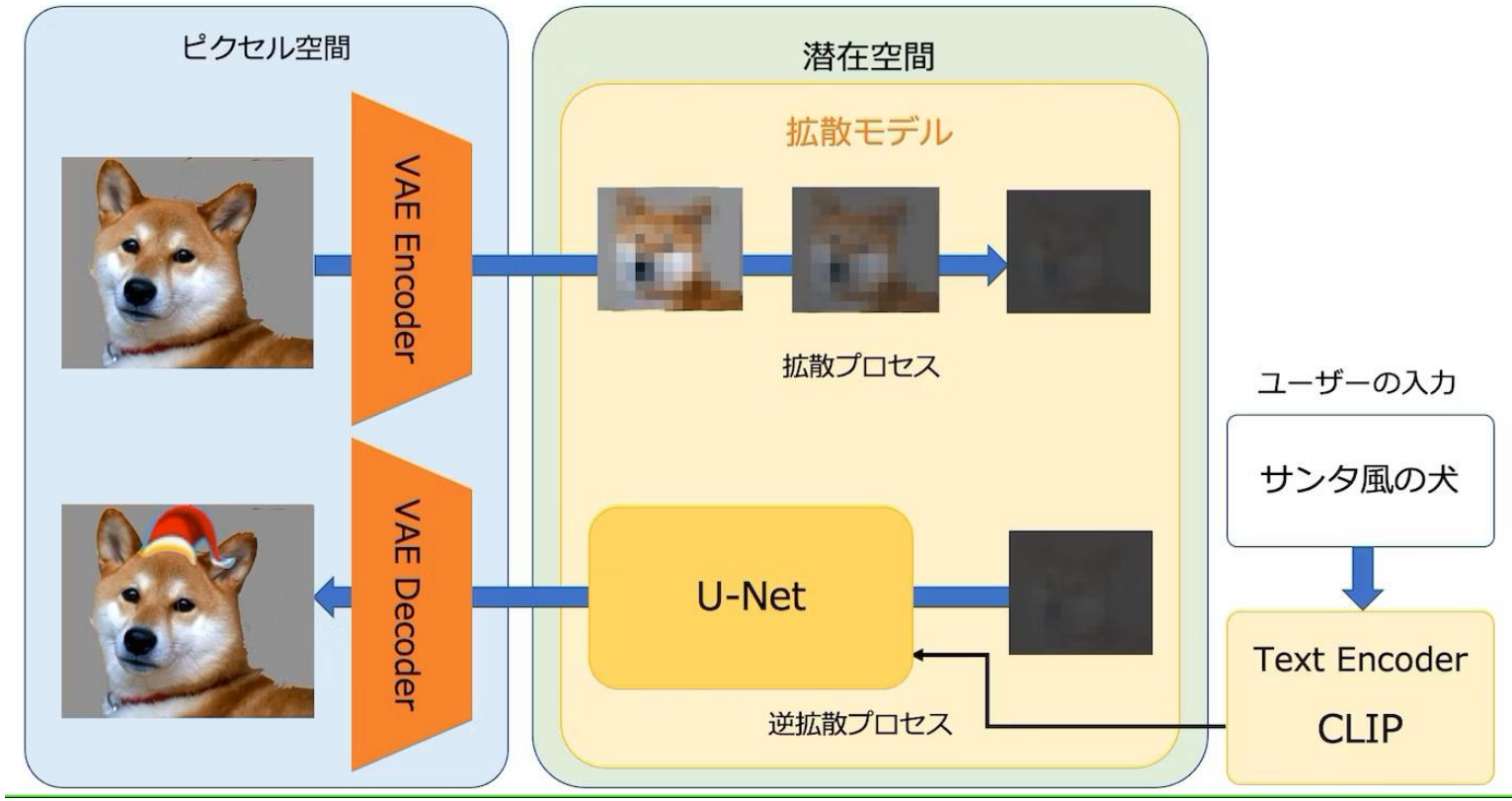
f19135　本田 涼大

Stable DiffusionとControlNet

- Stable Diffusion：テキストに沿った画像を生成するモデル
- ControlNetを使うとポーズが指定できる
- 研究テーマ：ControlNetのポーズ指定のところに任意の動画突っ込んでアニメっぽくしよう！

Stable Diffusion

- Diffusion Modelをベースとしたtext-to-imageの画像生成モデル
- VAE、Text Encoder、Diffusion Modelからなる



図で見てわかる！画像生成AI「Stable Diffusion」の仕組み - Qiita

VAE (Variational Autoencoder)

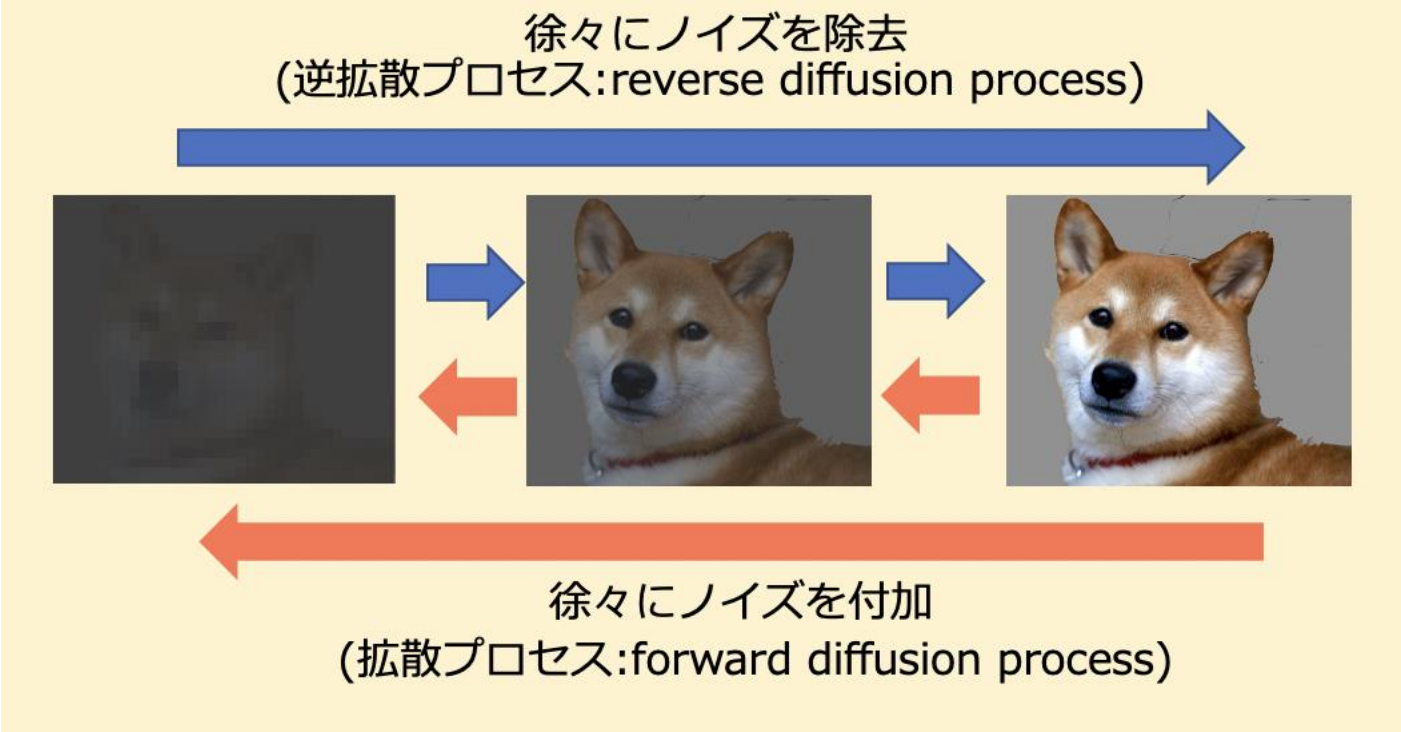
- 512x512のピクセル表現を8x8の潜在表現に変換
- 潜在表現はU-Netの入力として使用
- 推論時にはデコーダのみ使用

変分オートエンコーダ (VAE) – 【AI・機械学習用語集】

Diffusion Model

- ノイズからデータへの変換（逆拡散プロセス）を学習
  - 出力するノイズと正解ノイズの2乗誤差が最小化するように学習

- 潜在空間で動作するものをLatent Diffusion Modelという



[図で見てわかる！画像生成AI「Stable Diffusion」の仕組み - Qiita]

CLIP

- OpenAIが開発した画像分類モデル
- 学習サンプルが多く、汎用性が高いらしい
- Stable Diffusionでは訓練済みテキストエンコーダCLIPTextModelを使用

【日本語CLIP基礎】画像とテキストの類似度計算、画像やテキストの埋め込み計算、類似画像検索 - Qiita

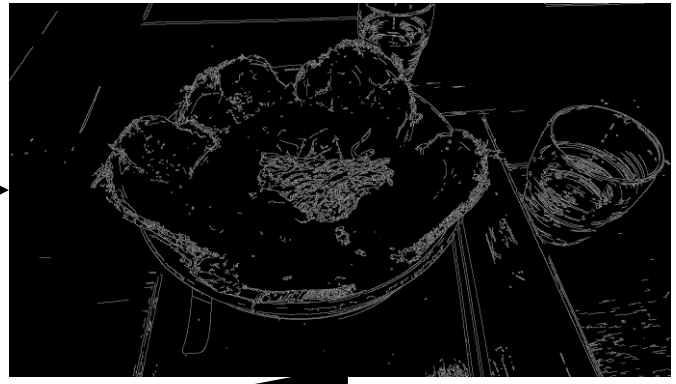
ControlNet

- Stable Diffusionで出力される画像のポーズを指定
- 画像の輪郭を取得し、色塗りをさせる
- Diffusion ModelのU-Netのデコーダ部分にControlNetのU-Netをぶちこむ

新時代の pix2pix？ ControlNet 解説

ControlNet

輪郭を抽出



着色

「ramen」

着色

「pasta」

これから

- Diffusion Modelの深堀り
  - 実はよく分かってない
- 動画を扱う方法の模索
  - 皆目見当がついてない
- 使用するモデルの検討とか
  - アニメ風にしたい