

人気の Stable Diffusion を Google Colab で体験

荻野 聖也

表1 主な画像生成AIの特徴

	クオリティ	多様性	速さ
Diffusionモデル	○	○	×
GAN	○	×	○
VAE	×	○	○

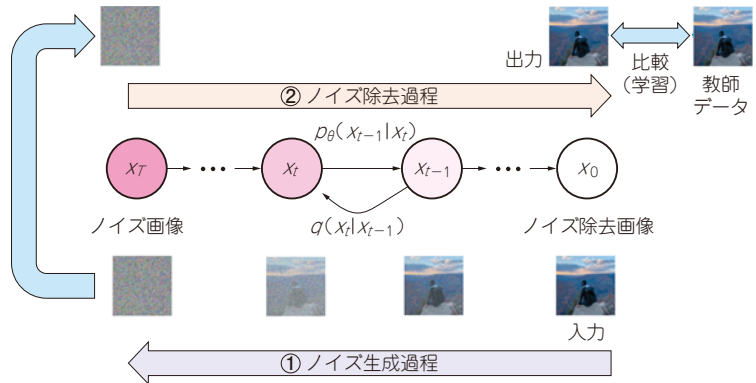


図1 Diffusion Modelの学習の仕組み
ノイズの付与とノイズの除去を学習に組み込んでいる

昨今、生成AIというワードが世の中のトレンドになってきていますが、画像生成の技術も最近大幅に発展してきています。Stable DiffusionやDALL・E3のような画像を生成するAIが登場し、現在の最新モデルなどを使えば、生成した画像が実際に写真として撮った画像か分からないレベルまで進化してきています。

本稿ではその礎になっている基礎技術をイメージと数学で解説しつつ、実際にプログラムを書いて試してみます。

画像生成AI… 一番人気はDiffusionモデル

表1に、生成モデルとして、現在特にメジャーなものを示します。

生成AIとして特に人気のある、Stable DiffusionはDiffusionモデルの一種です。Diffusionモデルとは、一言でいうと複数ステップにわたってノイズから画像を復元するようなモデルのことで、

GAN (Generative Adversarial Network) は、名前の「Adversarial (敵対的)」という言葉が示すように、生成した画像と本物の画像を比較し、生成したデータが本物のデータと間違えるように学習を進めていく方法で

す。日本語では敵対的生成ネットワークと呼ばれます。

VAE (Variational Autoencoder) は画像を潜在変数に落とし込むエンコーダと潜在変数から画像を再構成するデコーダを学習することで、学習済みのデコーダを使って画像を生成するモデルです。

表1のようにそれぞれのモデルで得意・不得意があるので、実際に生成モデルを作る際は、どういうモデルを作りたいか、などで決めることをお勧めします。

昨今では、生成するクオリティやモデルの守備範囲などから、Diffusionモデルが最も人気であるといえます。今回は、そのDiffusionモデルについて解説します。

Diffusionモデルの仕組み

● 全体イメージ①：学習過程

Diffusionモデルの学習過程のイメージを図1に示します。学習時は、ある画像からガウシアン・ノイズを付与された画像をステップごとに作っていき、最終的にはノイズそのもののようなモノを作ります。一方でこの過程は前のステップとノイズの四則演算で表せるので、ノイズからノイズを除去し「それっぽい画像」を表現することも可能です。学習ではこの「それっぽ