

画像生成AIの新機能を試す

新連載

第1回

Stable Diffusionで生成した画像を整える リファイナ技術

湊 雄一郎



画像生成AIは、機械学習の技術を用いて、莫大な画像データセットを学習し、そこから新しい画像を作り出す技術です。作り出す画像はプロンプトと呼ばれるテキストや画像そのものをガイドとしてなるべく指示通りに画像を生成しようとします。このような新しい画像生成AIは今では気軽に個人でも利用できるようになりました。

本連載では、拡散モデルと呼ばれる現在画像生成で最も利用される機械学習モデルを利用し、画像生成の仕組みや、利用の仕方をご紹介します。また、今後、LoRA（簡単に追加学習させる技術）やLCM（生成時間短縮の技術）、モデル・マージ（異なるモデル同士を統合する技術）などのさらに画像生成AIの利用を広げる新しい拡張性のある技術についても触れていきます。

画像生成AIの基本

● 拡散モデル

以前はGAN (Generative Adversarial Networks) やVAE (Variational Autoencoder) と呼ばれる画像を生成するようなAIが有名でした。最新のトレンドとして拡散モデルが挙げられます。拡散モデルは、逆拡散過程と呼ばれる、ノイズの多い状態から少しずつノイズを取り除く作業を連続して行うことで、高品質な画像へと復元する仕組みを利用しています(図1)。

通常、このようなノイズの取り除き方を予想するのは難しいのですが、拡散モデルではその部分にニューラル・ネットワークを利用することで、高精度にノイズを取り除き、高品質な画像を生成することに成功しています。

● 拡散モデルで利用されるニューラル・ネットワーク

高品質な画像生成を実現するために、ノイズの除去にはニューラル・ネットワークと呼ばれるモデルが採用されています。拡散モデルでは画像に強いCNN (Convolutional Neural Network) ベースのU-Netと呼ばれるU形をしたニューラル・ネットワークが広く使われています。

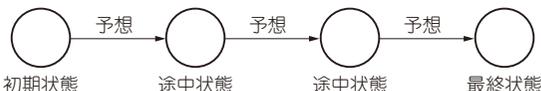


図1 拡散モデルのイメージ

ノイズの多い初期状態から、ノイズを取り除き、最終状態へ

ニューラル・ネットワークが拡散モデルで果たす役割は大きく分けて2つあります。1つがノイズの除去で、画像の特徴をつかむ特徴量を元に、どの状態にいるかのタイムスタンプも含めてノイズの取り除き方を決めていきます。そしてもう1つが、画像を生成するときに反映するテキスト・ベースの指示内容を適切に処理するための機構です。これらの機構をうまく複雑なニューラル・ネットワークのモデルに取り込んでいます。

LoRAやモデル・マージなど、今後紹介する予定の画像生成の拡張方法をより深く利用し理解するためにはこのようなニューラル・ネットワークの構造も把握しておくにより性能を引き出すことができます。

● U-Netの代わりに拡散トランスフォーマーが広がりつつある

最近ではU-Netの代わりにトランスフォーマー・モデルを利用する、拡散トランスフォーマー (Diffusion Transformers) の利用も広がっています。これまではU-Netが標準でしたが、OpenAIやStable Diffusionの最新モデルでは、相次いで拡散トランスフォーマーが採用されています。今後、拡散トランスフォーマーが採用されると、業界におけるツール類の大きな改良が必要になるため、注意深く推移を見届ける必要があるでしょう。

画像生成AIを試す準備

● 画像生成AIツール「Stable Diffusion」

ここでは、現在画像生成AIを手元のマシンで実行する際に最もポピュラーであるStable Diffusionを利用します。Stable DiffusionはStability AI社の提供するオープンソースの画像生成AIツールです。画像生成AIでは大量の画像を学習したモデルを準備するのが