第2章 2005年のオート・エンコーダから最新ディフュージョン・モデルまで

進化の過程を実体験！歴代の画像生成AI

長澤 和樹

表1　主要な生成モデル

発表年	モデル	概　要
2005	オート・エンコーダ(1)	• 入力データを圧縮した後，元のサイズに復元させるモデル • 復元したデータを元の入力データに適合するように学習することで，効率良いデータの次元圧縮器として利用された • 学習の過程で元のデータとの適合を行うため，生成器としても利用が可能．学習時の特徴を利用した復元によるデータ生成が行われる
2013	Variational オート・エンコーダ(4)	• オート・エンコーダのエンコーダ出力パラメータを潜在確率変数とすることにより，潜在特徴空間を（ある程度）活用しやすくすることを目的としたモデル • 2つのデータの中間のデータを生成するなど，潜在特徴空間を利用した生成が可能
2014	Generative Adversarial Network（GAN）(3)	• 入力データの真偽を判定する識別器および，ランダム・ノイズからデータの生成を行う生成器の2つのモデルを並べて学習させることによって，生成器に高精度な生成を行わせることを目指したモデル
2015	アテンション(5)	• 「重要な特徴がどの位置にあるのか」を学習する仕組み • シンプルな構造ながらも翻訳を始めとした時系列構造を持つデータの学習において精度向上に貢献した • 2017年提案の論文「Attention Is All You Need」(6)にてトランスフォーマの構造の一部として使用されている
2015	ディフュージョン(8)	• 大元(8)は非平衡熱力学の考えを元にした生成モデル • 画像においては，画像にノイズを加えていく順処理および，画像からノイズを除いていく逆処理を考え，学習により逆処理を獲得することで，完全なランダム・ノイズから画像を生成する(9)
2016	スタイル変換（Style transfer）(2)	• コンテンツ画像およびスタイル画像の2枚を用意し，スタイル画像の風合いでコンテンツ画像を描画する • 一般的なディープ・ラーニング学習とは異なり，モデルは調整せず入力データを調整することで与えられた2枚の画像の中間の状態に近づける
2017	トランスフォーマ(6)	• アテンション構造を組み合わせることにより構築された，大量のデータセットの学習に長けたモデル．翻訳をはじめとした時系列データの学習はもちろん，入出力の形状が異なるマルチモーダルの学習にも発展している • 画像を扱う課題においてはビジョン・トランスフォーマとして，トランスフォーマ・モデルが活用されている(7)

本記事では，生成モデルの基となっているオート・エンコーダ，GAN，アテンション，トランスフォーマ，ディフュージョンなど，主要な画像生成ディープ・ラーニング・モデルを中心にそれらの技術的なつながりを紹介し，どのように発展していったのかを探ります．

画像を扱う課題として，ディープ・ラーニングが流行し始めた当初（2012年）は，「この画像に写っているのは猫」のような認識の課題のみでした．しかし現在は，より複雑な課題に対応できるようになり，「この画像のこの場所に猫が写っている」のような位置までを含めた検出の課題や，「画像をピクセルごとに色分けしたとき，猫の写っている領域はこの範囲」のように画像に写る物体の領域分割の課題にまで対応が可能となりました．また，自然言語における翻訳や文章生成の課題に対してもディープ・ラーニングを用いることで「少し前の情報を考慮しながら推論する」という時系列の特徴を取り扱うモデルも増えていきました．

最近では，これらの仕組みを組み合わせ，応用してさまざまなデータの生成を行う分野も発展を遂げています．特に文字列（プロンプト）を入力にして画像を生成するマルチモーダルなモデルをはじめとした生成モデルは賛否両論のさまざまな驚きを生んでいます．

● 生成AIのここ10年の進化を俯瞰

本記事では次のモデルを扱います．それぞれの発表年や概要を**表1**に示します注1．

開発環境のセットアップは以下のサポート・ページをご覧ください．
https://interface.cqpub.co.jp/202408_tryai/