

AI 特許紹介(27)
AI 特許を学ぶ！究める！
～スタイル GAN～

2021 年 4 月 9 日
河野特許事務所
所長 弁理士 河野英仁

「AI 特許紹介」シリーズは、注目すべき AI 特許のポイントを紹介します。熾烈な競争となっている第 4 次産業革命下では AI 技術がキーとなり、この AI 技術・ソリューションを特許として適切に権利化しておくことが重要であることは言うまでもありません。

AI 技術は Google, Microsoft, Amazon を始めとした IT プラットフォーマ、研究機関及び大学から毎週のように新たな手法が提案されており、また AI 技術を活用した新たなソリューションも次々とリリースされています。

本稿では米国先進 IT 企業を中心に、これらの企業から出願された AI 特許に記載された AI テクノロジー・ソリューションのポイントをわかりやすく解説致します。

1.概要

特許権者 NVIDIA

出願日 2019 年 5 月 21 日

登録日 2020 年 5 月 14 日

登録番号 US2020/0151559

発明の名称 生成ニューラルネットワークのスタイルベースのアーキテクチャ

559 特許は、スタイル画像を生成するニューラルネットワークに関する。

2.特許内容の説明

図 1 A は、スタイルベースのジェネレータシステム 100 のブロック図を示している。

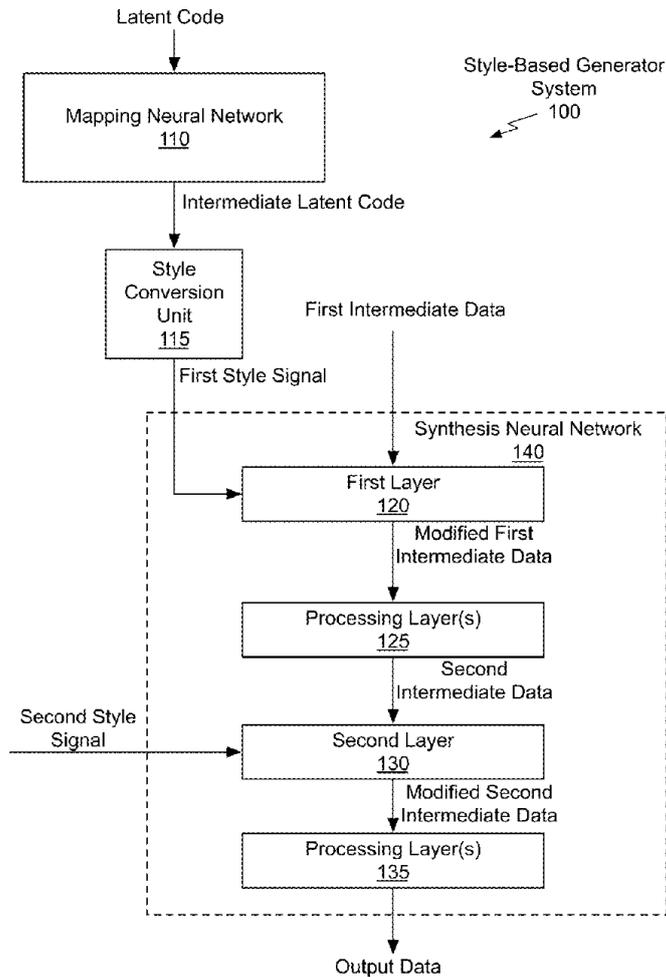


図 1 A

スタイルベースのジェネレータシステム 100 は、マッピングニューラルネットワーク 110、スタイル変換ユニット 115、および、合成ニューラルネットワーク 140 を含む。マッピングニューラルネットワーク 110 は潜在コードまたは追加の潜在コードから中間潜在コードを生成する。

入力潜在空間 Z に潜在コード z が与えられると、非線形マッピングネットワーク $f: Z \rightarrow W$ は中間潜在コード $w \in W$ を生成する。マッピング関数 f は、8 層 MLP (多層パーセプトロン、つまり、全結合層のみで構成されるニューラルネットワーク) を使用して実装される。

従来のジェネレータは、ジェネレータの入力層を介して潜在コードを供給するだけであるが、マッピングニューラルネットワーク 110 は、代わりに、入力潜在コード z を中間潜在空間 W にマッピングして、中間潜在コード w を生成する。

スタイル変換ユニット 115 は、中間潜在コード w を第 1 のスタイル信号に変換する。中間潜在コード w は、第 1 のスタイル信号および第 2 のスタイル信号を含む空間的に不変のスタイルに変換される。

従来のスタイル変換手法とは対照的に、空間的に不変なスタイルは、サンプル画像からではなく、ベクトル、つまり中間潜在コード w から計算される。合成ニューラルネットワーク 140 は、スタイル信号に従って第 1 の中間データ（例えば、特徴マップとして符号化された学習定数）を処理し、例えば、第 1 の中間データの密度を 4×4 から 8×8 に増加させ、出力データ密度に達するまで継続する。

スタイル変換ユニット 115 は、アフィン変換を実行する。スタイル変換ユニット 115 は、合成ニューラルネットワーク 140 のトレーニング中にアフィン変換を学習するようにトレーニングされる。

第 1 のスタイル信号は、合成ニューラルネットワーク 140 の第 1 の層 120 での動作を制御して、修正された第 1 の中間データを生成する。第 1 のスタイルの信号は、合成ネットワーク 140 の第 1 の層 120 内の適応インスタンス正規化(AdaIN)動作を制御する。

AdaIN 操作は、コンテンツ特徴マップのセットとスタイル信号とを受信し、コンテンツ特徴マップの 1 次統計（つまり、「スタイル」）を変更して、スタイル信号によって定義された 1 次統計と一致させる。

第 1 の層 120 によって出力された修正された第 1 の中間データは、処理層 125 によって処理され、第 2 の中間データを生成する。処理層 125 は、 3×3 畳み込み層と、それに続く、図 1A に示されていない追加のスタイル信号を受信する AdaIn 動作とを含む。

第 2 の中間データは、合成ニューラルネットワーク 140 の第 2 の層 130 に入力される。第 2 のスタイルの信号は、修正された第 2 の中間データを生成するために、第 2 の層 130 での動作を制御する。第 1 のスタイル信号は、第 1 の中間データにエンコードされた第 1 の属性を変更し、第 2 のスタイル信号は、第 1 の中間データおよび第 2 の中間データにエンコードされた第 2 の属性を変更する。

例えば、第 1 の中間データは、第 2 の中間データと比較して粗いデータであり、第 1 のスタイルは、第 1 の層 120 で粗い特徴マップに変換され、第 2 のスタイルは、第 2 の

層 130 でより高密度の特徴マップに変換される。

第 2 の層 130 は、第 2 の中間データをアップサンプリングし、 3×3 の畳み込み層とそれに続く AdaIN 操作を含む。第 2 のスタイルの信号は、合成ネットワーク 140 の第 2 の層 130 内の AdaIN 動作を制御する。

第 2 の層 130 によって出力された修正された第 2 の中間データは、処理層 135 によって処理され、第 2 の中間データに対応するコンテンツを含む出力データを生成する。

処理層 135 は、 3×3 畳み込み層を含む。出力データは、第 1 のスケールに対応する第 1 の属性および第 2 のスケールに対応する第 2 の属性を含む画像であり、第 1 のスケールは第 2 のスケールと比較して粗い。第 1 のスケールは、第 1 の層 120 によって処理される特徴マップのスケールに対応し、第 2 のスケールは、第 2 の層 130 によって処理される特徴マップのスケールに対応する。したがって、第 1 のスタイル信号は第 1 のスケールで第 1 の属性を変更し、第 2 のスタイル信号は第 2 のスケールで第 2 の属性を変更する。

3.クレーム

559 特許のクレーム 1 は以下の通りである。

1. コンピュータで実行される方法において、

マッピングニューラルネットワークによって、中間潜在空間で定義された中間潜在コードを生成するために、入力空間で定義された潜在コードを処理し、

中間潜在コードを第 1 スタイルの信号に変換し、

第 1 のスタイル信号に従って第 1 の中間データを修正し、修正された第 1 の中間データを生成するために、合成ニューラルネットワークの第 1 の層に第 1 のスタイル信号を適用し、

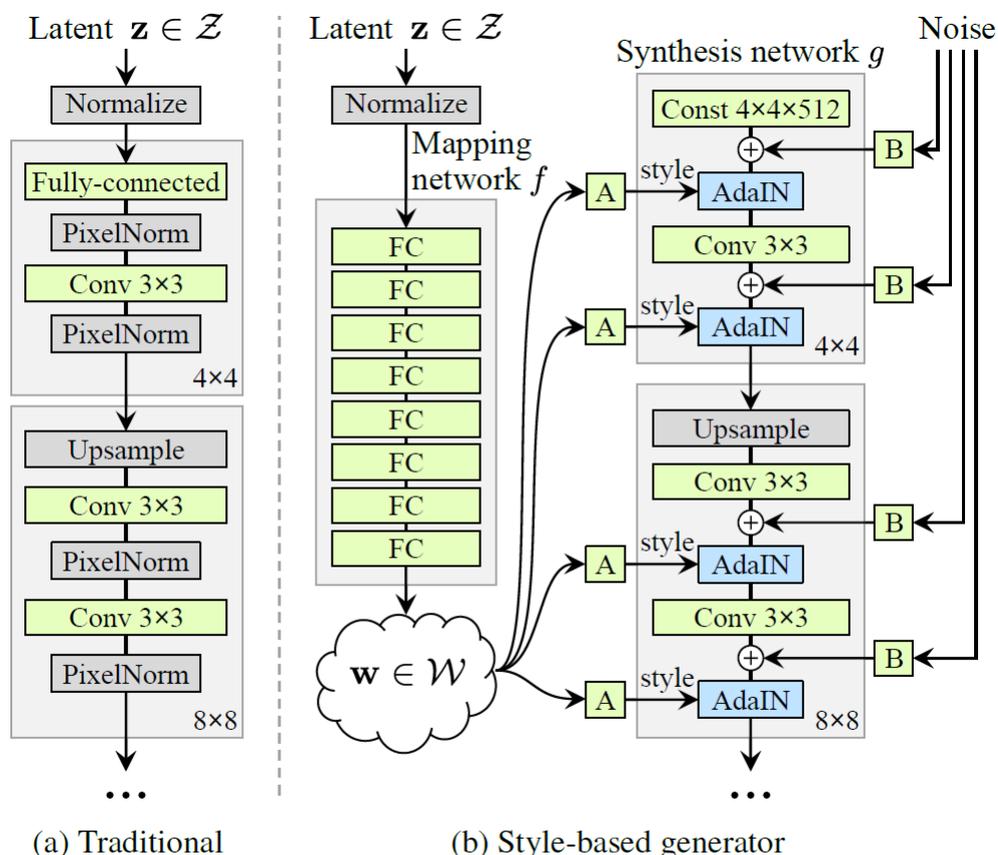
第 2 の中間データを生成するために、修正された第 1 の中間データを処理し、

第 2 のスタイル信号に従って第 2 の中間データを修正し、修正された第 2 の中間データを生成するために、合成ニューラルネットワークの第 2 の層に第 2 のスタイル信号を適用し、

第 2 の中間データに対応するコンテンツを含む出力データを生成するために、修正された第 2 の中間データを処理する。

4. スタイル GAN に関する論文

本特許に関連する論文¹「生成的敵対的ネットワークのためのスタイルベースのジェネレータアーキテクチャ」が Tero Karras 氏らにより発表されている。



(a) Traditional

(b) Style-based generator

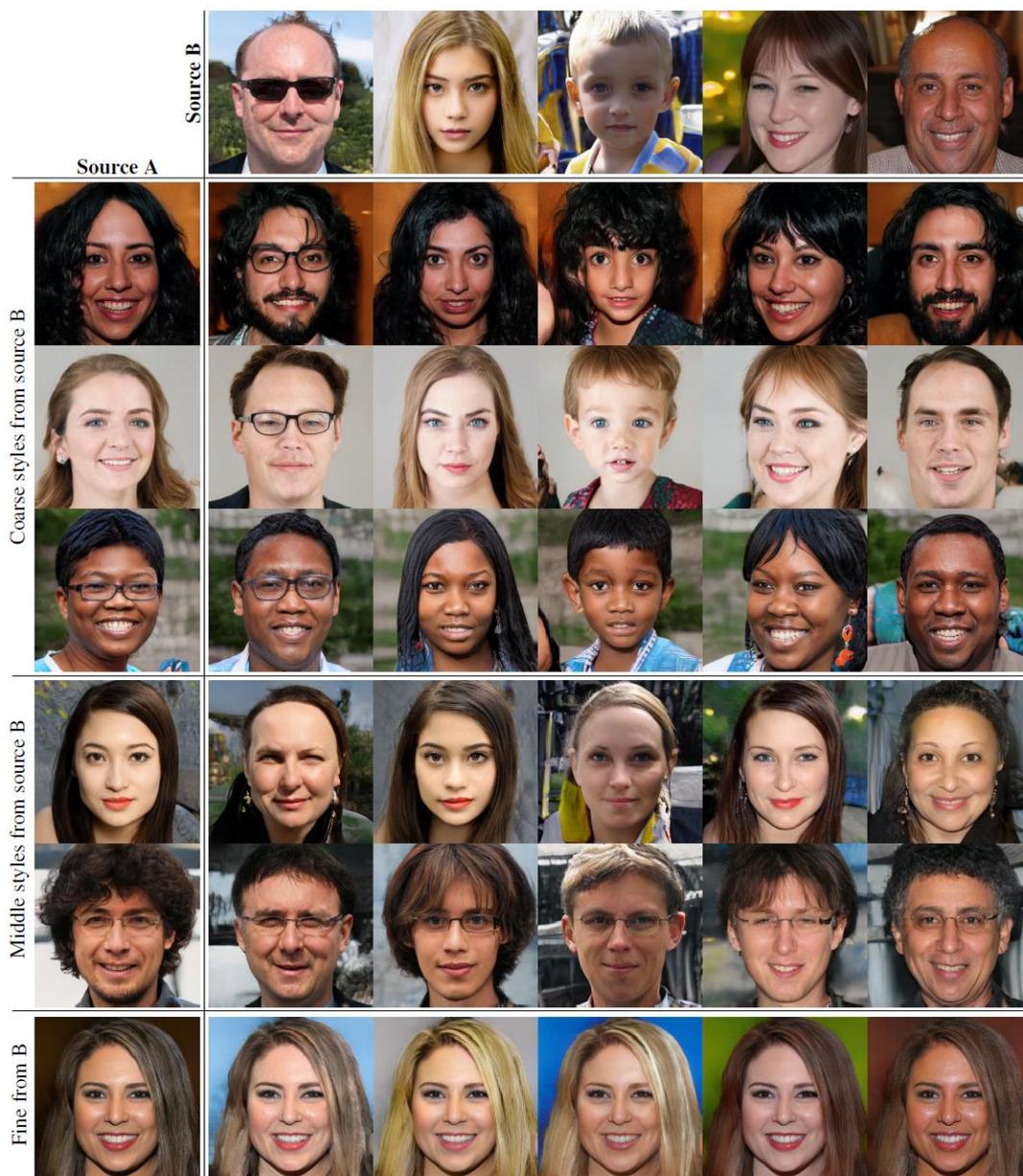
上記図は、従来のネットワーク構成、及び、本論文のスタイルベース生成器のネットワーク構成を示すブロック図である。

従来のジェネレータは入力レイヤーのみを介して潜在コードをフィードするが、本論文の方式では、最初に入力を中間潜在空間 W にマップし、次に各畳み込みレイヤーで適応インスタンス正規化 (AdaIN) を介してジェネレータを制御する。「A」は学習されたアフィン変換を表す。

また、ガウスノイズが、各畳み込みの後、非線形性を評価する前に追加される。「B」は学習されたチャンネルごとのスケール係数をノイズ入力に適用する。マッピングネットワーク f は 8 層で構成され、合成ネットワーク g は 18 層で構成される (解像度ごとに 2 つ $4^2 \sim 1024^2$)。最後のレイヤーの出力は、個別の 1×1 畳み込みを使用して

¹ Tero Karras , Samuli Laine , Timo Aila “A Style-Based Generator Architecture for Generative Adversarial Networks” arXiv:1812.04948v3 [cs.NE] 29 Mar 2019

RGBに変換される。



上記図は、スタイルベースのジェネレータシステムによって生成された画像を示す。粗い空間解像度 ($4^2 \sim 8^2$) に対応するスタイルをコピーすると、ソース B のポーズ、一般的なヘアスタイル、顔の形、眼鏡などの高レベルのアスペクトがもたらされるが、すべての色 (目、髪、ライティング) とより細かい顔の特徴は A に似ている。

代わりに、B から中解像度 ($16^2 \sim 32^2$) のスタイルをコピーすると、A のポーズ、一般的な顔の形、眼鏡は保持されたまま、B からの小規模な顔の特徴、髪型、目の開閉が

継承される。最後に、B からファインスタイル (64² - 1024²) をコピーすると、主に配色と微細構造がもたらされる。

以上

著者紹介

河野英仁

河野特許事務所、所長弁理士。立命館大学情報システム学博士前期課程修了、米国フランクリンピアースローセンター知的財産権法修士修了、中国清華大学法学院知的財産夏季セミナー修了、MIT(マサチューセッツ工科大学)コンピュータ科学・AI 研究所 AI コース修了。

[AI 特許コンサルティング](#)、[医療 AI 特許コンサルティング](#)の他、米国・中国特許の権利化・侵害訴訟を専門としている。著書に「世界のソフトウェア特許(共著)」、「FinTech 特許入門」、「[AI/IoT 特許入門 2.0](#)」、「[ブロックチェーン 3.0\(共著\)](#)」がある。