

AI 特許紹介(88)
AI 特許を学ぶ！究める！
～DreamBooth 特許～

2026 年 5 月 8 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 Google

出願日 2023 年 8 月 23 日

登録日 2026 年 2 月 17 日

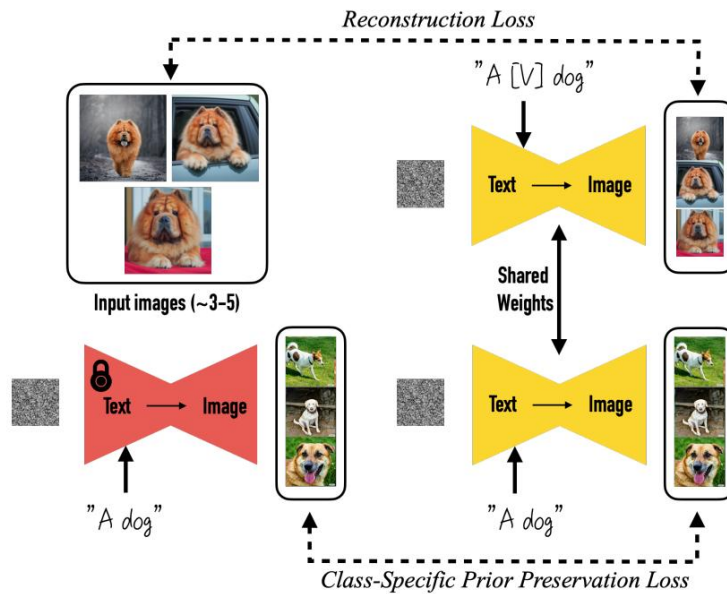
登録番号 US12555275

発明の名称 パーソナライズされたテキストから画像への拡散モデル

275 特許は、被写体の少数の画像と、被写体に固有の識別子を関連付けてテキストから画像へ変換を行う拡散モデルをファインチューニングすることにより、固有の識別子を使用して、異なるシーンに文脈化された被写体の新しい画像を合成することが可能な DreamBooth 技術に関する。

2.特許内容の説明

下記図は DreamBooth ファインチューニング時のワークフローである。



通常のファインチューニングに要する枚数よりはるかに少ない被写体の約 3~5 枚の画像を用意し、入力画像と、一意の識別子と被写体が属するクラス名 (例:A [V] dog) を含むテキストプロンプトをペアにして、テキストから画像への拡散モデルをファインチューニングする。オブジェクトクラスは、ランドマーク、景観または位置の特徴、乗り物、道具、食べ物、衣服、装置、動物、人間等である。また、識別子は、V 等の自然言語テキストのコーパスに現れる文字、サブワード、単語、句読点、数字、その他の記号である。

同時に、例えば"dog"等のクラス固有の事前保存損失を適用する。つまり、ファインチューニングの際、識別子 V を含むテキスト入力から画像への拡散モデルへの入力として与えられた場合には再構成損失を最適化し、識別子 V を含まないテキスト入力から画像への拡散モデルへの入力として与えられた場合にはクラス固有の事前分布保持損失を最適化する。これは、モデルがクラスに対して持つ意味的事前情報を活用し、テキストプロンプト (例:A dog) 内のクラス名を使用して、被写体のクラスに属する多様なインスタンスを生成するように促す。

下記図は、ファインチューニングされたテキストから画像への拡散モデルを用いて推論を行うためのアーキテクチャを示す図である。

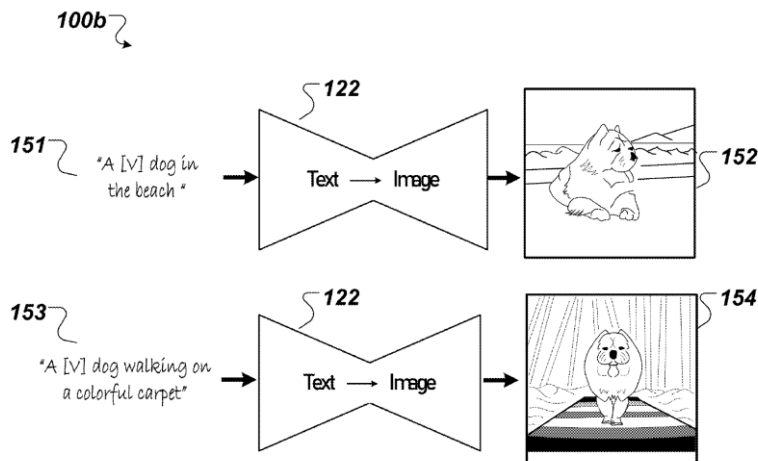


FIG. 1B

上記図は、それぞれ以下の形式の2つの新しいテキスト入力 151、153 を示している。

“a [V] [object class name] [context description]”

オブジェクトクラス名は「dog」である。テキスト入力 151 のコンテキスト説明は「in the beach」、テキスト入力 153 のコンテキスト説明は「walking on a colorful carpet」である。[V]は、カスタム画像データセットに含まれる各画像に現れる特定の対象インスタンスに対して生成された一意の識別子を表す。このデータセットは、ファインチューニング処理中に使用された。

上記例では、各新規テキスト入力には同じ固有識別子[V]が含まれている。固有識別子[V]が新規テキスト入力の一部として含まれている場合、ファインチューニング処理の結果として学習された固有識別子から特定の対象インスタンスへのマッピングにより、パーソナライズされたテキスト画像変換モデル 122 は、新規テキスト入力を処理して、固有識別子によって識別されるオブジェクトクラスの特定の対象インスタンスをそれぞれ描写する新しい画像 152、154 を生成する。

上記図では、パーソナライズされたテキスト画像変換モデル 122 が、同じ皮膚/毛色、同じ耳の形、同じ大きさの特定の犬を異なる背景で写した2つの新しい画像 152、154 を生成することを示している。

一方、テキスト入力に固有識別子[V]が含まれていない場合、例えば、テキスト入力が次の形式である場合(オブジェクトクラス名は“dog”)、
 “a [object class name] [context description]”パーソナライズされたテキスト画像変換モデル 122 によって生成される対応する画像は、プロンプトテキストで指定された実際のコンテキスト説明に応じて、皮膚/毛皮の色、耳の形、大きさ、背景が異なる様々な

犬を表示する。

このモデル構成により、従来は不可能だった様々な技術的ユースケースが可能になる。例えば、パーソナライズされたテキストから画像へのモデル 122 を実装した画像合成システムは、例えばユーザから、対象となるインスタンス（特定の人、特定の動物、特定の車、特定のボートなど）を定義する一意の識別子を含むテキストプロンプトを受け取り、そのモデルを使用して、対象となるインスタンスを多様なコンテキスト、ポーズ、視点、照明条件でそれぞれ示す、任意の数の高忠実度合成画像（例えば、フォトリアリスティックな画像）を生成することができる。

3.クレーム

275 特許のクレーム 1 は以下の通りである。

1. コンピュータ実装方法において、

各対象インスタンスを描写する複数の画像からカスタム画像データセットを生成し、前記カスタム画像データセットの生成は、以下のステップを含む：

前記複数の画像のそれぞれについて、画像に描写されている対象インスタンスのオブジェクトクラスを決定し、

トークンの語彙から、前記トークンの所定の頻度に従って、前記複数の画像に描写されている対象インスタンスに対応する 1 つ以上のトークンのシーケンスを決定し、

前記 1 つ以上のトークンのシーケンスから、前記複数の画像に描写されている対象インスタンスの一意の識別子を生成し、

カスタム画像データセット上でテキスト画像変換モデルをトレーニングし、テキスト入力として固有識別子が与えられた場合に、テキスト画像変換モデルが対象クラスの対象インスタンスをそれぞれ描写する画像を生成するように、テキスト画像変換モデルのパラメータ値を調整し、カスタム画像データセット上でのテキスト画像変換モデルのトレーニングは、以下のステップを含む：

第 1 のトレーニング出力画像を生成するために、第 1 のトレーニングモデル入力を、事前学習済みのパラメータ値に従ってテキスト画像変換モデルのインスタンスを用いて処理し、ここで、第 1 のトレーニングモデル入力は、ノイズと対象クラスを指定する条件付け入力とを含み、

第 2 のトレーニング出力画像を生成するために、第 2 のトレーニングモデル入力を、テキスト画像変換モデルの現在のパラメータ値に従ってテキスト画像変換モデルを用いて処理し、ここで、第 2 のトレーニングモデル入力は、ノイズと対象クラスを指定する条件付け入力とを含み、

第 3 のトレーニング出力画像を生成するために、テキスト画像変換モデルを用い

て、現在のパラメータ値に従って第3のトレーニングモデル入力を処理し、ここで、第3のトレーニングモデル入力は、ノイズと、対象クラスを指定し、固有識別子を含む条件付け入力とを含み、

第1のトレーニング出力画像と第2のトレーニング出力画像との差を測定するクラス固有の事前保存損失項を含む目的関数の勾配から、テキスト画像変換モデルの現在のパラメータ値の更新を決定する。

4. 本特許に関連する論文

本特許に関する論文 “DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation”¹が、Google の Nataniel Ruiz 氏らにより公表されている。

下記図は、画像変換結果を示す説明図である。



被写体を数枚(通常3~5枚)撮影するだけで(左)、AI搭載フォトブース「DreamBooth」は、テキストプロンプトの指示に従い、さまざまな状況下で被写体を捉えた無数の画像を生成できる(右)。生成された画像は、被写体の主要な視覚的特徴を忠実に再現しながら、環境との自然な相互作用、斬新な表現、照明条件の変化など、多様な表情を捉えている。

下記図は DreamBooth と Textual Inversion との比較を示す図である。

¹ Nataniel Ruiz, et al. “DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation” arXiv:2309.16797v1 [cs.CL] 28 Sep 2023



4つの入力画像（上段）に対して、DreamBooth Imagen（2段目）、DreamBooth Stable Diffusion（3段目）、Textual Inversion（下段）を比較する。出力画像は、次のプロンプト（左から右）を使用して作成された。

「雪の中の [V] 花瓶」、「ビーチの [V] 花瓶」、「ジャングルの [V] 花瓶」、「背景にエッフェル塔がある [V] 花瓶」

DreamBooth は、被写体とプロンプトの両方の忠実度において優れている。DreamBooth のプロジェクトページは以下のとおりである。

プロジェクトページ：

<https://dreambooth.github.io/>

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、生成 AI ビジネスコース修了。

[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 3](#)」、「[ブロックチェーン 3.0](#)(共著)」がある。