

明 細 書

発明の名称：画像生成装置、画像生成方法、及びプログラム

技術分野

[0001] 本発明は、画像生成装置、画像生成方法、及びプログラムに係り、特に、所望の固有特徴を有する画像を生成する画像生成装置、画像生成方法、及びプログラムに関する。

背景技術

[0002] 近年、デジタルカメラやスマートフォンの普及により画像の取得が容易となり、これらの画像に写った物体を機械で識別することで、工場での人の目視点検の代替や、小売り店舗での商品の欠品検知自動化等、様々な場面で人の作業効率化に役立ち始めている。

[0003] このように、画像に写った物体を機械で識別する、画像の識別技術への要望は高まっている。

[0004] 画像の識別技術に関して、近年、非特許文献1で開示されているような、CNN (Convolutional Neural Network) に基づく手法が多く開示されている。

[0005] CNNは、特徴を検出するフィルタをスライドさせながら入力画像に対して掛け合わせて作成した特徴マップを出力する畳み込み処理と、抽出した特徴を局所領域ごとにまとめ上げるプーリング処理を繰り返し行っている。

[0006] CNNが高い識別性能を発揮するためには、大量の学習データをCNNに入力し、これらを識別するようにフィルタを学習する必要がある。つまり、高精度な識別能力を持ったCNNを獲得するためには、大量の学習データが必要である。

[0007] 大量の学習データをすべて人手で用意するためには、多くのコストがかかる。具体的には、画像をカテゴリに分類する画像分類タスクのための学習データを用意するためには、カテゴリ毎に多くの画像が必要であり、例えば非特許文献2で開示されている画像分類タスク用公開データセット Image

Net を題材とした画像認識のコンペティション ILSVRC2012 で使用されたデータセットでは、1 カテゴリ当たり約 1, 200 枚、合計 1, 000 カテゴリ分の画像が用意されている。さらに、カテゴリが詳細（例えば、椅子カテゴリを、ソファ、ベンチ、ダイニングチェアカテゴリに分ける場合）になるほど、学習データの準備がより大変になる。

[0008] 上記の問題に対して、少量の画像データを準備し、これらを変換することで、画像データを拡張する方法がある。

[0009] 例えば、非特許文献 1 では、予め定めた画像の幾何学的な変換方法（クロップや回転など）を用いて画像データを拡張し、拡張した画像データセットで画像分類器の学習を行うことで画像分類精度の向上を確認している。

[0010] また、特許文献 1 や非特許文献 3 では、カテゴリに共通して存在する特徴（属性）に基づいて画像を変換する方法が提案されている。画像とその画像が持つ属性のペアデータを複数用意し、これを学習データとして画像生成装置を学習する。

[0011] この画像生成装置に、画像と変換したい属性とのペアを入力すると、変換したい属性を特徴として持つ画像が出力される。

先行技術文献

特許文献

[0012] 特許文献 1：特開 2018-55384 号公報

非特許文献

[0013] 非特許文献 1：C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich. "Going deeper with convolutions." In proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1-9.

非特許文献 2：O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. Berg, L. Fei-Fei. "ImageNet Large Scale Visual Recognition Challenge." In proc. International Journal of Computer Vision (IJCV), 2015, pp.211-252.

非特許文献3 : G. Lample, N. Zeghidour, N. Usunier, A. Bordes, L. Denoyer, M. Ranzato, “Fader Networks: Manipulating Images by Sliding Attributes.” In Proc. Of NIPS, 2017, pp.5963-5972.

発明の概要

発明が解決しようとする課題

- [0014] しかし、非特許文献1で得られる画像データセットは、画像の幾何学的な変換で得られる画像に限られるため、幾何学的な変換で得られない画像は正しく分類できない可能性がある、という問題があった。
- [0015] 例えば、色や柄の異なる画像は得られないため、少量の画像データセットに無い色や柄の画像を正しく分類できない可能性がある。
- [0016] また、非特許文献3では、属性に基づいて多様な変換が可能であるが、変換対象となる画像は、画像生成装置の学習に用いたカテゴリの物体に限られる。そのため、画像生成装置の学習に用いた学習データに無い未知のカテゴリの画像を変換する場合には、変換する場所が定まらず、望んだ第一の画像を得られない、という問題があった。
- [0017] 例えば、図8に示すように、カテゴリ「キャップ」が学習データに無い未知のカテゴリである場合、キャップのカテゴリに属する画像のどの位置を変換すべきか分からない。
- [0018] 本発明は上記の点に鑑みてなされたものであり、所望のカテゴリの画像であって、かつ、所望の固有特徴を有する画像を生成することができる画像生成装置、画像生成方法、及びプログラムを提供することを目的とする。

課題を解決するための手段

- [0019] 本発明に係る画像生成装置は、所望の固有特徴を有する第一の画像を生成する画像生成装置であって、前記第一の画像と同じカテゴリに属する第二の画像から得られた前記カテゴリに属する画像に共通する特徴であるカテゴリ特徴と、前記第一の画像と前記第二の画像とで異なる固有の特徴である固有特徴と、を関連付けることで前記第一の画像を生成する生成部を有し、前記固有特徴は、前記第二の画像を分割した分割領域の各々について、前記所望

の固有特徴が関連付けられたものである。

[0020] また、本発明に係る画像生成方法は、所望の特徴を有する第一の画像を生成する画像生成方法であって、生成部が、前記第一の画像と同じカテゴリに属する第二の画像から得られた前記カテゴリに属する画像に共通する特徴であるカテゴリ特徴と、前記第一の画像と前記第二の画像とで異なる固有の特徴である固有特徴と、を関連付けることで前記第一の画像を生成し、前記固有特徴は、前記第二の画像を分割した分割領域の各々について、前記所望の固有特徴が関連付けられたものである。

[0021] 本発明に係る画像生成装置及び画像生成方法によれば、生成部が、第一の画像と同じカテゴリに属する第二の画像から得られた当該カテゴリに属する画像に共通する特徴であるカテゴリ特徴と、第一の画像と第二の画像とで異なる固有の特徴である固有特徴と、を関連付けることで第一の画像を生成する。固有特徴は、第二の画像を分割した分割領域の各々について、所望の固有特徴が関連付けられたものである。

[0022] このように、第一の画像と同じカテゴリに属する第二の画像から得られた当該カテゴリに属する画像に共通する特徴であるカテゴリ特徴と、第一の画像と第二の画像とで異なる固有の特徴である固有特徴と、を関連付けることで第一の画像を生成することにより、所望のカテゴリの画像であって、かつ、所望の固有特徴を有する画像を生成することができる。

[0023] また、本発明に係る画像生成装置の記カテゴリ特徴は、前記第二の画像から前記固有特徴を除いて抽出されるように、かつ、所定の識別器に前記固有特徴を有すると識別されないように学習されていることができる。

[0024] また、本発明に係る画像生成装置の前記生成部は、前記所望の固有特徴と関連付けられた分割領域の位置情報を用いたマスクを、前記カテゴリ特徴にかけたものを変換し、前記変換により得られたデータを用いて前記第一の画像を生成することができる。

[0025] また、本発明に係る画像生成装置の前記生成部は、更に、前記所望の固有特徴から、前記所望の固有特徴が関連付けられた分割領域の位置情報を抑制

したデータと、前記カテゴリ特徴とを含むデータを変換し、前記変換により得られたデータを用いて前記第一の画像を生成することができる。

[0026] また、本発明に係る画像生成装置の前記生成部は、前記第二の画像を入力として前記カテゴリ特徴を抽出するエンコーダと、前記カテゴリ特徴と前記所望の固有特徴とを入力として前記第一の画像を生成するデコーダと、を更に含み、前記エンコーダ及びデコーダは、学習用の固有特徴と、前記学習用の固有特徴を有する学習用画像とのペアに基づいて、前記学習用画像を前記エンコーダに入力し、前記学習用の固有特徴を前記デコーダに入力したときに前記デコーダが、前記学習用画像を再構成し、かつ、前記カテゴリ特徴を入力とする所定の識別器により前記学習用の固有特徴を有すると識別されないように予め学習されていることができる。

[0027] また、本発明に係る画像生成装置の前記所定の識別器は、前記カテゴリ特徴を入力としたときに正しく前記固有特徴を有すると識別するように予め学習されていることができる。

[0028] 本発明に係るプログラムは、コンピュータを、上記の画像生成装置の各部として機能させるためのプログラムである。

発明の効果

[0029] 本発明の画像生成装置、画像生成方法、及びプログラムによれば、所望のカテゴリの画像であって、かつ、所望の固有特徴を有する画像を生成することができる。

図面の簡単な説明

[0030] [図1]本発明の実施の形態に係る画像生成装置の構成の一例を示すブロック図である。

[図2]本発明の実施の形態に係る画像生成装置のエンコーダ、デコーダ、及び識別器の関係を示すイメージ図である。

[図3]本発明の実施の形態に係る画像生成装置のデコーダの構成の一例を示すイメージ図である。

[図4]本発明の実施の形態に係る画像生成装置より生成される第一の画像の例

である。

[図5]本発明の実施の形態に係る画像生成装置の学習処理ルーチンを示すフローチャートである。

[図6]本発明の実施の形態に係る画像生成装置のデコード処理ルーチンを示すフローチャートである。

[図7]本発明の実施の形態に係る画像生成装置の画像生成処理ルーチンを示すフローチャートである。

[図8]本発明の課題を示すイメージ図である。

発明を実施するための形態

[0031] 以下、本発明の実施の形態について図面を用いて説明する。

[0032] <本発明の実施の形態に係る画像生成装置の構成>

図1を参照して、本発明の実施の形態に係る画像生成装置100の構成について説明する。図1は、本発明の実施の形態に係る画像生成装置の構成を示すブロック図である。

[0033] 画像生成装置100は、CPUと、RAMと、後述する画像生成処理ルーチンを実行するためのプログラムを記憶したROMとを備えたコンピュータで構成され、機能的には次に示すように構成されている。

[0034] 図1に示すように、本実施形態に係る画像生成装置100は、入力部1と、記憶部2と、生成部3と、パラメータ更新部4と、出力部5とを備えて構成される。以下、画像生成装置100の処理を、学習処理と、画像生成処理とに分けて各機能を説明する。

[0035] <<学習処理>>

入力部1は、学習用の固有特徴と、当該学習用の固有特徴を含む学習用画像との1以上のペアの入力を受け付ける。

[0036] 本実施形態において、カテゴリ特徴とは、所望のカテゴリに属する対象に共通する特徴である。例えば、帽子のカテゴリ「ハット」であれば、つばの部分が「丸い」といった特徴である。

[0037] また、固有特徴（属性）とは、所望のカテゴリに属する複数の対象に必ず

しも共通しない場合がある特徴である。例えば、帽子のカテゴリ「ハット」であれば、「青い」という特徴である。「青い」という特徴は、カテゴリ「ハット」に属する複数の対象に共通する場合もあれば、共通しない場合もある。

[0038] また、属性位置データとは、第一の画像（出力画像）における各位置の固有特徴（属性）を示すデータである。

[0039] 具体的には、学習用画像 x は、サイズが横幅 \times 縦幅 \times チャネル数のテンソルであり、ここでは、学習用画像 x の横幅を W 、縦幅を H 、チャネル数を D とする。また、学習用画像 x は、横幅と縦幅が等しいテンソル（すなわち、 $W=H$ ）であれば何でも良い。

[0040] また、テンソルの最左上手前の座標を $(0, 0, 0)$ とし、最左上手前から右に w 、下に h 進み、奥に d 枚目のチャネルに相当する座標を (w, h, d) と表記するものとする。

[0041] また、説明の簡単のため、各テンソルについて、横幅の次元を次元 1、縦幅の次元を次元 2、チャネル数の次元を次元 3 と記載する。すなわち、学習用画像 x の次元 1 のサイズは W 、次元 2 のサイズは H 、次元 3 のサイズは D となる。

[0042] 横幅と縦幅が等しくない画像 ($W \neq H$) から横幅と縦幅が等しい ($W=H$) 画像を作成する方法は、テンソルのサイズを変更する処理であれば何でも良く、例えばリサイズ処理、画像の一部を切り出すクロップ処理、画像の周囲に数値 0 や画像の端の画素を繰り返し追加するパディング処理、又は画像の端の画素を上下または左右で逆に追加するミラーリング処理等を行う。

[0043] 固有特徴とは、同じカテゴリの画像で異なる固有の特徴であり、学習用画像を分割した分割領域の各々について、固有特徴が関連付けられたものである。

[0044] 本実施形態では、入力部 1 は、各位置についての、固有特徴としての属性を表す属性位置データ y の入力を受け付ける。属性位置データ y は、学習用

画像 x の変換後における各位置の属性を示すデータである。

- [0045] 属性とは、当該画像生成装置 100 で変換する予め定義した画像の固有特徴を表す言葉であれば何でも良く、例えば赤、青等の色、木製、ガラス等の素材、ドットやストライプ等の柄等の特徴を示す言葉である。
- [0046] また、各属性が特定可能な識別子が付与されているものとする。例えば、予め定義した属性が A 種類るとき、 0 以上 A 未満の自然数を付与する。また、学習用画像 x の変換後の各位置における各属性の有無を、属性位置データ y が示すものとする。
- [0047] 属性位置データ y は、属性が A 種類るとき、サイズが $M \times N \times A$ のテンソル Y であり、学習用画像 x のサイズが $W \times H \times D$ としたとき $1 \leq M \leq W$ 、 $1 \leq N \leq H$ となり、 $M = N$ であるものとする。
- [0048] 学習用画像 x について、横幅を M 個に、縦幅を N 個に分割しグリッドに分け、学習用画像 x の変換後の画像について、最左上から右に m 番目、下に n 番目のグリッドが持つ属性を特定する数値が a であるとき、テンソル Y の (m, n, a) の位置に 1 を配置する。
- [0049] 一方、当該グリッドが数値 a で特定される属性を持っていない場合、テンソル Y の (m, n, a) の位置に 0 を配置する。
- [0050] そして、入力部 1 は、受け付けた学習用画像 x と属性位置データ y との 1 以上のペアを、生成部 3 に渡す。
- [0051] 記憶部 2 には、学習用画像 x を入力として学習用画像 x と同一のカテゴリに属する画像に共通する特徴であるカテゴリ特徴としての潜在表現 $E(x)$ を抽出するエンコーダ、潜在表現 $E(x)$ と属性位置データ y とを入力として各位置に属性を有し、かつ、当該カテゴリに属する画像を生成するデコーダ、及び潜在表現 $E(x)$ を入力として各属性が含まれるか否かを識別する識別器が格納されている。
- [0052] 具体的には、エンコーダ、デコーダ、及び識別器は、それぞれニューラルネットワークであり、記憶部 2 には、各ニューラルネットワークのパラメータが格納されている。

[0053] 生成部3は、出力画像



と同じカテゴリに属する学習用画像から得られた潜在表現 $E(x)$ と、属性位置データ y と、を関連付けることで出力画像



を生成する。

[0054] 具体的には、生成部3は、まず、記憶部2からエンコーダ、デコーダ及び識別器の各パラメータを取得する。

[0055] 次に、生成部3は、学習用画像 x をエンコーダに入力して潜在表現 $E(x)$ を抽出し、抽出した潜在表現 $E(x)$ と属性位置データ y をデコーダに入力して、出力画像



を生成する。

[0056] 図2に、エンコーダ、デコーダ、及び識別器の関係を示す。

[0057] エンコーダは、学習用画像 x を入力とし、学習用画像 x から属性情報を除いたカテゴリ特徴を抽出するニューラルネットワークであれば何でも良い。以下、本実施形態では、カテゴリ特徴の例として、潜在表現 $E(x)$ を用いて説明する。

[0058] 例えば、非特許文献3のエンコーダを採用することができる。非特許文献3のエンコーダは、入力される画像のサイズが $256 \times 256 \times 3$ であるとき、出力される潜在表現 $E(x)$ のサイズが $2 \times 2 \times 512$ となるニューラルネットワークを用いている。

[0059] デコーダは、潜在表現 $E(x)$ と属性位置データ y を入力とし、学習用画像 x とサイズが同じ画像であり、かつ、属性位置データ y によって与えられた各位置の属性情報を持つ画像を生成するニューラルネットワークである。

[0060] 図3にデコーダの構成を示す。図3に示すように、デコーダは、ローカル

用潜在表現前処理、ローカル用属性位置データ前処理、ローカル用入力データ統合処理、ローカルデコーダの処理、グローバル用属性位置データ前処理、グローバル用入力データ統合処理、グローバルデコーダの処理、及び画像デコーダの処理の各処理を行う。

[0061] ローカルデコーダ、グローバルデコーダ、及び画像デコーダはそれぞれニューラルネットワークであり、デコーダでは、ローカルデコーダの出力であるテンソルと、グローバルデコーダの出力であるテンソルと、属性位置データ y とを次元 3 の方向に重ね合わせたテンソルを画像デコーダに入力し、出力画像



を生成する。以下、デコーダの各処理について説明する。

[0062] ローカルデコーダは、属性のある位置のみをフィルタリングするためのデコーダである。ローカルデコーダは、属性位置データ y をマスクとして使用し、属性のある位置のみに着目するように潜在表現 $E(x)$ を変換する。

[0063] 具体的には、ローカルデコーダは、入力となるテンソルのサイズの次元 1 及び次元 2 が属性位置データ y の次元 1 及び次元 2 のサイズと同じであり、次元 3 のサイズが潜在表現 $E(x)$ の次元 3 のサイズと同じであり、入力のテンソルと同じサイズのテンソルを出力するニューラルネットワークであれば何でも良い。

[0064] ローカルデコーダに属性位置データ y の次元 1 及び次元 2 のサイズと同じテンソルを入力するために、ローカル用潜在表現前処理により潜在表現 $E(x)$ のサイズを変形し、ローカル用属性位置データ前処理により属性位置データ y のサイズを変形し、各前処理の出力をローカル用入力データ統合処理により統合する。

[0065] 具体的には、ローカル用潜在表現前処理は、潜在表現 $E(x)$ のサイズの次元 1 及び次元 2 が属性位置データ y の次元 1 及び次元 2 のサイズと同じテンソルになるよう変形する処理である。

[0066] 例えば、ローカル用潜在表現前処理は、潜在表現 $E(x)$ のサイズが $2 \times$

2 × 5 1 2 であり、属性位置データ y のサイズが $16 \times 16 \times 11$ であるとき、ローカル用潜在表現前処理の出力を $16 \times 16 \times 5 1 2$ のテンソルとするために、潜在表現 $E(x)$ のテンソルを $1 \times 1 \times 5 1 2$ に変形した後に、次元 1 及び次元 2 の方向に 16×16 個複製し、 $16 \times 16 \times 5 1 2$ のテンソルを出力する処理を行う。

[0067] また、ローカル用属性位置データ前処理は、属性位置データ y の次元 3 のサイズが、潜在表現 $E(x)$ の次元 3 のサイズと同じテンソルになるよう変形する処理である。

[0068] 例えば、ローカル用属性位置データ前処理は、潜在表現 $E(x)$ のサイズが $2 \times 2 \times 5 1 2$ であり、属性位置データ y のサイズが $16 \times 16 \times 11$ であるとき、ローカル用属性位置データ前処理の出力を $16 \times 16 \times 5 1 2$ のテンソルとするために、属性位置データ y のテンソルを次元 3 の方向に足し合わせ、 $16 \times 16 \times 1$ のテンソルにした後、これを次元 3 の方向に $5 1 2$ 個複製し、 $16 \times 16 \times 5 1 2$ のテンソルを出力する処理を行う。

[0069] ローカル用入力データ統合処理は、ローカル用潜在表現前処理の出力であるテンソルと、ローカル用属性位置データ前処理の出力であるテンソルを入力とし、これら入力された 2 つのテンソルと同じサイズのテンソルを出力する処理である。

[0070] 例えば、ローカル用入力データ統合処理は、入力された 2 つのテンソルを掛け合わせることにより、次元 1 及び次元 2 が属性位置データ y の次元 1 及び次元 2 のサイズと同じであり、かつ、次元 3 のサイズが潜在表現 $E(x)$ の次元 3 のサイズと同じであるテンソルを出力する処理を行う。

[0071] グローバルデコーダは画像全体としての構造を保持するためのデコーダである。グローバルデコーダは、属性位置データ y の位置情報を、事前に属性位置データ y の次元 1 及び次元 2 のサイズを縮小することで、抑制し、潜在表現 $E(x)$ と合わせて変換することで画像全体の構造を保つように変換する。

[0072] 具体的には、グローバルデコーダは、次元 1 及び次元 2 が潜在表現 $E(x)$

) の次元 1 及び次元 2 のサイズと同じテンソルを入力とし、次元 1 及び次元 2 のサイズが属性位置データ y の次元 1 及び次元 2 と同じサイズのテンソルを出力するニューラルネットワークであれば何でも良い。

[0073] グローバルデコーダに次元 1 及び次元 2 が潜在表現 $E(x)$ の次元 1 及び次元 2 のサイズと同じテンソルを入力するために、グローバル用属性位置データ前処理により属性位置データ y のサイズを変形し出力したテンソルと、潜在表現 $E(x)$ とをグローバル用入力データ統合処理により統合する処理である。

[0074] 具体的には、グローバル用属性位置データ前処理は、属性位置データ y の次元 1 及び次元 2 のサイズが、潜在表現 $E(x)$ の次元 1 及び次元 2 のサイズと同じテンソルになるよう変形する処理である。

[0075] 例えば、グローバル用属性位置データ前処理は、潜在表現 $E(x)$ のサイズが $2 \times 2 \times 5 \times 1 \times 2$ であり、属性位置データ y のサイズが $16 \times 16 \times 1 \times 1$ であるとき、畳み込みニューラルネットワークにより畳み込み処理を行い、 $2 \times 2 \times 5 \times 1 \times 2$ のテンソルを出力する処理を行う。

[0076] グローバル用入力データ統合処理は、潜在表現 $E(x)$ と、グローバル用属性位置データ前処理の出力であるテンソルを入力とし、次元 1 及び次元 2 のサイズが、これら入力された 2 つのテンソルの次元 1 及び次元 2 と同じサイズのテンソルを出力する処理である。

[0077] 例えば、グローバル用入力データ統合処理は、入力された 2 つのテンソルを次元 3 の方向に重ね合わせることで、次元 1 及び次元 2 が潜在表現の次元 1 及び次元 2 のサイズと同じであるテンソルを出力する処理を行う。

[0078] 画像デコーダは、ローカルデコーダの出力であるテンソルと、グローバルデコーダの出力であるテンソルと、属性位置データ y とを次元 3 の方向に重ね合わせたテンソルを入力として、出力画像



を生成する処理を行う。

[0079] 識別器は、画像から得られた潜在表現 $E(x)$ を入力としたときに当該画

像が有する属性を識別するニューラルネットワークである。

[0080] 例えば、識別器は、潜在表現 $E(x)$ のサイズが $2 \times 2 \times 512$ であり、属性数が 10 であるとき、入力として $2 \times 2 \times 512$ のテンソルを受け付け、出力が長さ 10 のベクトルとなるニューラルネットワークを用いることができる。

[0081] そして、生成部 3 は、学習用画像 x 、生成した出力画像



、及び属性位置データ y を、パラメータ更新部 4 に渡す。

[0082] パラメータ更新部 4 は、属性位置データ y と、属性位置データ y が表す各位置の属性を有する学習用画像 x とのペアに基づいて、学習用画像 x をエンコーダに入力し、属性位置データ y をデコーダに入力したときにデコーダが、学習用画像 x を再構成し、かつ、潜在表現 $E(x)$ を入力とする識別器により属性位置データ y が表す属性を有すると識別されないようにエンコーダ及びデコーダの各パラメータを更新し、潜在表現 $E(x)$ を入力としたときに属性位置データ y が表す属性を有すると正しく識別するように識別器のパラメータを更新する。

[0083] 具体的には、パラメータ更新部 4 は、まず、記憶部 2 からエンコーダ、デコーダ、及び識別器の各パラメータを取得する。

[0084] 次に、パラメータ更新部 4 は、下記の二種類の制約を満たすようニューラルネットワークであるエンコーダ、デコーダ、及び識別器の各パラメータを更新する。

[0085] 一つ目の制約は、生成した出力画像



が、学習用画像 x を再構成するようにエンコーダ及びデコーダの各パラメータを更新することである。

[0086] この一つ目の制約を満たすよう設定された学習方法であれば何でも良く、例えば非特許文献 3 では、学習用画像 x と生成した出力画像



との二乗誤差を計算し、これを小さくするようにエンコーダ及びデコーダの各パラメータを更新する。

[0087] 二つ目の制約は、学習用画像 x が入力されたエンコーダが属性情報を除くように潜在表現 $E(x)$ を抽出し、かつ、識別器は潜在表現 $E(x)$ から属性位置データ y が表す属性を有すると正しく識別するようにエンコーダ及び識別器の各パラメータを更新することである。

[0088] この二つ目の制約を満たすよう設定された学習方法であれば何でも良く、例えば、非特許文献3では、識別器が潜在表現 $E(x)$ から属性位置データ y が表す属性を有すると正しく識別する確率を小さくするようエンコーダのパラメータを更新することに対し、識別器が潜在表現 $E(x)$ から属性位置データ y が表す属性を有すると識別する確率が大きくなるように識別器のパラメータを更新する。

[0089] そして、パラメータ更新部4は、学習したエンコーダ、デコーダ、及び識別器の各パラメータを、記憶部2に格納する。

[0090] なお、学習処理では、入力された学習画像 x 及び属性位置データ y の1以上のペアの各々について、1つのペア毎にエンコーダ、デコーダ、及び識別器の各パラメータを学習してもよいし、バッチ処理等により複数同時、又は一括して学習してもよい。

[0091] <<画像生成処理>>

次に、画像生成処理について説明する。画像生成装置100の画像生成処理では、固有特徴である属性位置データ y を有する第一の画像



を生成する。

[0092] なお、簡単のため、本実施形態では、第二の画像 x は、学習画像 x と同様のテンソルであるものとする。

[0093] 入力部1は、生成したい第一の画像



と同じカテゴリに属する第二の画像 x 、及び所望の固有特徴である属性位置データ y の入力を受け付ける。

- [0094] 具体的には、第二の画像 x は、サイズが横幅×縦幅×チャンネル数のテンソルであり、ここでは、第二の画像 x の横幅を W 、縦幅を H 、チャンネル数を D とする。また、第二の画像 x は、横幅と縦幅が等しいテンソル（すなわち、 $W=H$ ）であれば何でも良い。
- [0095] また、テンソルの最左上手前の座標を $(0, 0, 0)$ とし、最左上手前から右に w 、下に h 進み、奥に d 枚目のチャンネルに相当する座標を (w, h, d) と表記するものとする。
- [0096] また、説明の簡単のため、学習処理と同様に、各テンソルについて、横幅の次元を次元1、縦幅の次元を次元2、チャンネル数の次元を次元3と記載する。すなわち、第二の画像 x の次元1のサイズは W 、次元2のサイズは H 、次元3のサイズは D となる。
- [0097] 横幅と縦幅が等しくない画像 ($W \neq H$) から横幅と縦幅が等しい ($W=H$) 画像を作成する方法は、テンソルのサイズを変更する処理であれば何でも良く、例えばリサイズ処理、画像の一部を切り出すクロップ処理、画像の周囲に数値0や画像の端の画素を繰り返し追加するパディング処理、又は画像の端の画素を上下または左右で逆にして追加するミラーリング処理等を行う。
- [0098] そして、入力部1は、受け付けた第二の画像 x と属性位置データ y とを、生成部3に渡す。
- [0099] 記憶部2には、学習処理により学習されたエンコーダ、デコーダ、及び識別器の各パラメータが格納されている。
- [0100] 生成部3は、第一の画像



と同じカテゴリに属する第二の画像 x から得られた潜在表現 $E(x)$ と、属

性位置データ y と、を関連付けることで第一の画像



を生成する。

[0101] 具体的には、生成部3は、まず、記憶部2から学習済みのエンコーダ、デコーダ及び識別器の各パラメータを取得する。

[0102] 次に、生成部3は、第二の画像 x をエンコーダに入力して潜在表現 $E(x)$ を抽出し、抽出した潜在表現 $E(x)$ と属性位置データ y をデコーダに入力して、第一の画像



を生成する。

[0103] そして、生成部3は、生成した第一の画像



を出力部5に渡す。

[0104] 出力部5は、第一の画像



を出力する。

[0105] 図4に画像生成装置100により生成される第一の画像の例を示す。図4の例では、カテゴリ「帽子」に属する第二の画像 x と、対象物全体の位置が属性「黒色」を有することを表す属性位置データ y とから、帽子全体の位置が黒色となる第一の画像



を生成することが示されている。

[0106] これは、未知のカテゴリの第二の画像であっても、上記学習処理により学習済みのエンコーダにより属性情報を除いた当該カテゴリの潜在表現 $E(x)$ を抽出することができ、抽出した潜在表現 $E(x)$ に対して、所望の属性

位置データ y が表す属性を関連付けることができることによる。

[0107] 様々な属性位置データ y を本実施形態に係る画像生成装置 100 に入力することにより、第二の画像と同一のカテゴリであって、属性位置データ y が表す属性を有する複数の第一の画像

x

を生成することができる。

[0108] このように生成された複数の第一の画像

x

は、例えば物体検出器の学習用の画像に用いることができる。

[0109] <本発明の実施の形態に係る画像生成装置の作用>

図5は、本発明の実施の形態に係る学習処理ルーチンを示すフローチャートである。

[0110] 入力部1に、属性位置データ y が表す各位置の属性を有する学習用画像 x 及び当該属性位置データ y の1以上のペアが入力されると、画像生成装置100において、図5に示す学習処理ルーチンが実行される。

[0111] まず、ステップS100において、属性位置データ y と、学習用画像 x との1以上のペアの入力を受け付ける。

[0112] ステップS110において、生成部3は、学習用画像 x をエンコーダに入力して潜在表現 $E(x)$ を抽出する。

[0113] ステップS120において、上記ステップS110において抽出した潜在表現 $E(x)$ と属性位置データ y をデコーダに入力して、出力画像

x

を生成する。

[0114] ステップS130において、属性位置データ y と、属性位置データ y が表す各位置の属性を有する学習用画像 x とのペアに基づいて、学習用画像 x をエンコーダに入力し、属性位置データ y をデコーダに入力したときにデコー

だが、学習用画像 x を再構成し、かつ、潜在表現 $E(x)$ を入力とする識別器により属性位置データ y が表す属性を有すると識別されないようにエンコーダ及びデコーダの各パラメータを更新し、潜在表現 $E(x)$ を入力としたときに属性位置データ y が表す属性を有すると正しく識別するように識別器のパラメータを更新する。

[0115] 図6は、上記ステップS120におけるデコード処理ルーチンを示すフローチャートである。

[0116] ステップS121において、生成部3は、潜在表現 $E(x)$ のサイズの次元1及び次元2が属性位置データ y の次元1及び次元2のサイズと同じテンソルになるよう変形するローカル用潜在表現前処理を行う。

[0117] ステップS122において、生成部3は、属性位置データ y の次元3のサイズが、潜在表現 $E(x)$ の次元3のサイズと同じテンソルになるよう変形するローカル用属性位置データ前処理を行う。

[0118] ステップS123において、生成部3は、上記ステップS121により得られたテンソルと、上記ステップS122により得られたテンソルを入力とし、これら入力された2つのテンソルと同じサイズのテンソルを出力するローカル用入力データ統合処理を行う。

[0119] ステップS124において、生成部3は、上記ステップS124により得られたテンソルをローカルデコーダに入力し、属性位置データ y をマスクとして使用し、属性のある位置のみに着目するように潜在表現 $E(x)$ を変換する。

[0120] ステップS125において、生成部3は、属性位置データ y の次元1及び次元2のサイズが、カテゴリ特徴の次元1及び次元2のサイズと同じテンソルになるよう変形するグローバル用属性位置データ前処理を行う。

[0121] ステップS126において、生成部3は、潜在表現 $E(x)$ と、上記ステップS125により得られたテンソルを入力とし、次元1及び次元2のサイズが、これら入力された2つのテンソルの次元1及び次元2と同じサイズのテンソルを出力するグローバル用入力データ統合処理を行う。

[0122] ステップS 1 2 7において、生成部3は、上記ステップS 1 2 6により得られたテンソルをグローバルデコーダに入力し、属性位置データ y の位置情報を抑制し、潜在表現 $E(x)$ と合わせて変換することで画像全体の構造を保つように変換する。

[0123] ステップS 1 2 8において、生成部3は、上記ステップS 1 2 4によりデコードされたテンソルと、上記ステップS 1 2 7によりデコードされたテンソルと、属性位置データ y とを次元3の方向に重ね合わせたテンソルとを画像デコーダに入力して、出力画像



を生成する

[0124] 図7は、本発明の実施の形態に係る画像生成処理ルーチンを示すフローチャートである。なお、学習処理ルーチンと同様の処理については、同一の符号を付して詳細な説明は省略する。

[0125] 入力部1に第二の画像 x 及び属性位置データ y が入力されると、画像生成装置100において、図7に示す画像生成処理ルーチンが実行される。

[0126] まず、ステップS 2 0 0において、生成したい第一の画像



と同じカテゴリに属する第二の画像 x 、及び属性位置データ y の入力を受け付ける。

[0127] ステップS 2 3 0において、上記ステップS 1 2 0により得られた第一の画像



を出力する。なお、画像生成処理ではステップS 1 2 8において、生成部3は、第一の画像



を生成する。

- [0128] 以上説明したように、本発明の実施形態に係る画像生成装置によれば、第一の画像と同じカテゴリに属する第二の画像から得られた当該カテゴリに属する画像に共通する特徴であるカテゴリ特徴と、第一の画像と第二の画像とで異なる固有の特徴である固有特徴と、を関連付けることで第一の画像を生成することにより、所望のカテゴリの画像であって、かつ、所望の固有特徴を有する画像を生成することができる。
- [0129] なお、本発明は、上述した実施の形態に限定されるものではなく、この発明の要旨を逸脱しない範囲内で様々な変形や応用が可能である。
- [0130] 本実施形態では、学習処理と画像生成処理とを同一の画像生成装置100により行うこととしたが、別の装置で行うようにしてもよい。この場合、学習処理により学習済みのエンコーダ、デコーダ、及び識別器が格納された記憶部2を画像生成処理に用いればよい。
- [0131] また、本願明細書中において、プログラムが予めインストールされている実施形態として説明したが、当該プログラムを、コンピュータ読み取り可能な記録媒体に格納して提供することも可能である。

符号の説明

- [0132] 1 入力部
2 記憶部
3 生成部
4 パラメータ更新部
5 出力部
100 画像生成装置

請求の範囲

- [請求項1] 所望の固有特徴を有する第一の画像を生成する画像生成装置であつて、
- 前記第一の画像と同じカテゴリに属する第二の画像から得られた前記カテゴリに属する画像に共通する特徴であるカテゴリ特徴と、前記第一の画像と前記第二の画像とで異なる固有の特徴である固有特徴と、を関連付けることで前記第一の画像を生成する生成部を有し、
- 前記固有特徴は、前記第二の画像を分割した分割領域の各々について、前記所望の固有特徴が関連付けられたものである画像生成装置。
- [請求項2] 前記カテゴリ特徴は、前記第二の画像から前記固有特徴を除いて抽出されるように、かつ、所定の識別器に前記固有特徴を有すると識別されないように学習されている
- 請求項1記載の画像生成装置。
- [請求項3] 前記生成部は、
- 前記所望の固有特徴と関連付けられた分割領域の位置情報を用いたマスクを、前記カテゴリ特徴にかけたものを変換し、前記変換により得られたデータを用いて前記第一の画像を生成する
- 請求項1又は2記載の画像生成装置。
- [請求項4] 前記生成部は、
- 更に、前記所望の固有特徴から、前記所望の固有特徴が関連付けられた分割領域の位置情報を抑制したデータと、前記カテゴリ特徴とを含むデータを変換し、前記変換により得られたデータを用いて前記第一の画像を生成する
- 請求項3記載の画像生成装置。
- [請求項5] 前記生成部は、
- 前記第二の画像を入力として前記カテゴリ特徴を抽出するエンコー

ダと、

前記カテゴリ特徴と前記所望の固有特徴とを入力として前記第一の画像を生成するデコーダと、

を更に含み、

前記エンコーダ及びデコーダは、学習用の固有特徴と、前記学習用の固有特徴を有する学習用画像とのペアに基づいて、前記学習用画像を前記エンコーダに入力し、前記学習用の固有特徴を前記デコーダに入力したときに前記デコーダが、前記学習用画像を再構成し、かつ、前記カテゴリ特徴を入力とする所定の識別器により前記学習用の固有特徴を有すると識別されないように予め学習されている

請求項 1 記載の画像生成装置。

[請求項6] 前記所定の識別器は、前記カテゴリ特徴を入力としたときに正しく前記固有特徴を有すると識別するように予め学習されている

請求項 5 記載の画像生成装置。

[請求項7] 所望の特徴を有する第一の画像を生成する画像生成方法であって、生成部が、前記第一の画像と同じカテゴリに属する第二の画像から得られた前記カテゴリに属する画像に共通する特徴であるカテゴリ特徴と、前記第一の画像と前記第二の画像とで異なる固有の特徴である固有特徴と、を関連付けることで前記第一の画像を生成し、

前記固有特徴は、前記第二の画像を分割した分割領域の各々について、前記所望の固有特徴が関連付けられたものである

画像生成方法。

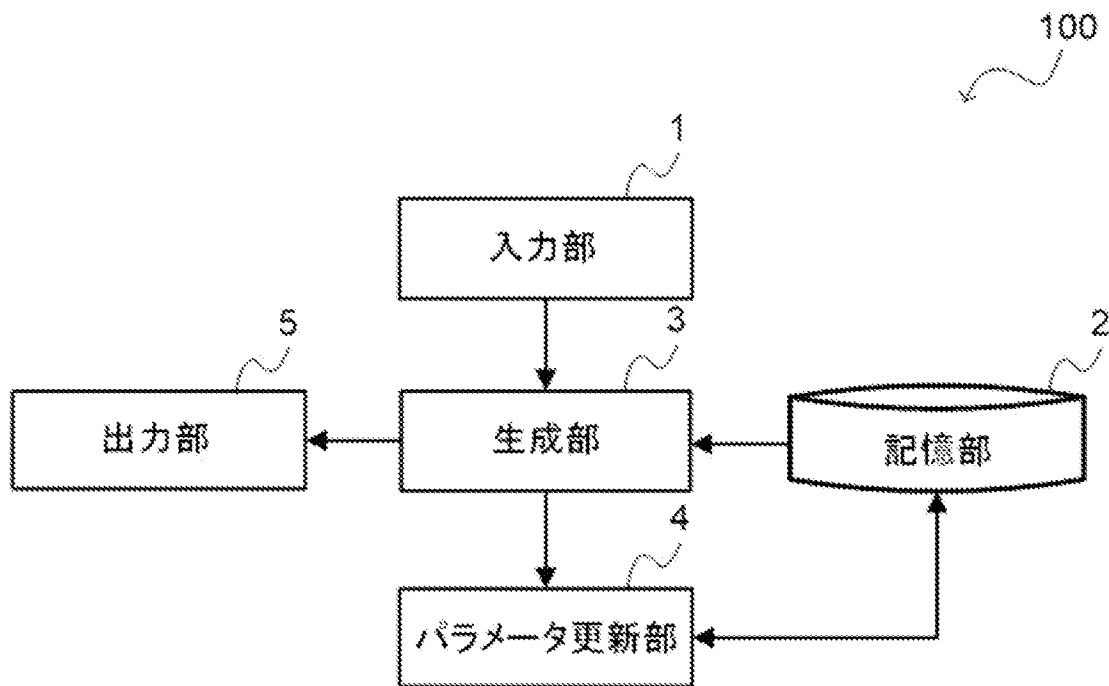
[請求項8] コンピュータを、請求項 1 乃至 6 の何れか 1 項記載の画像生成装置の各部として機能させるためのプログラム。

要 約 書

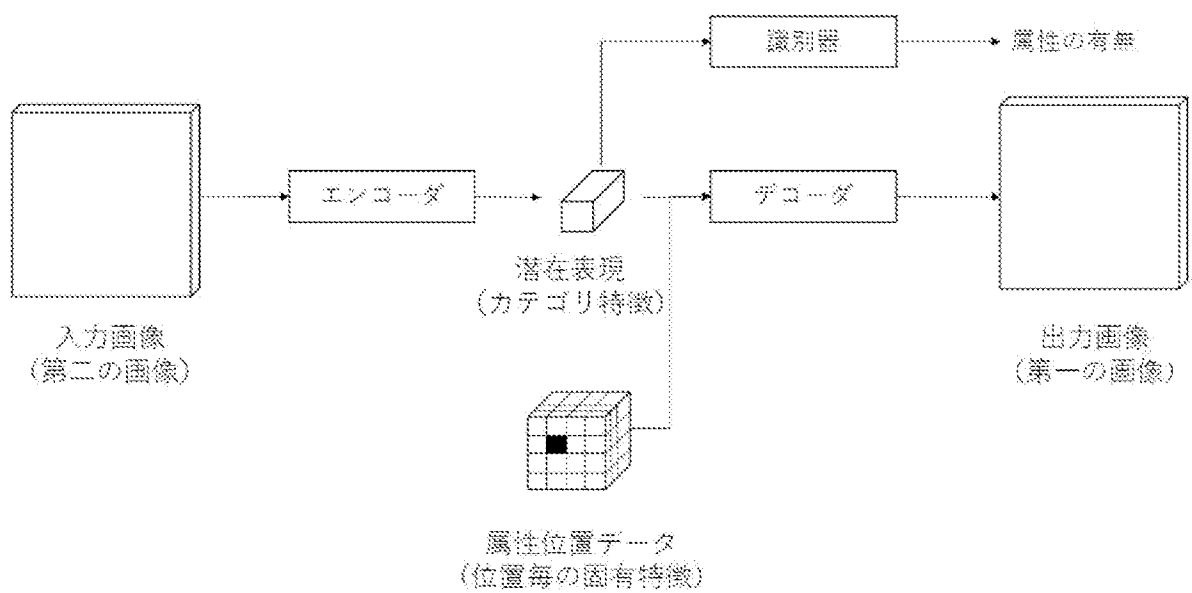
所望のカテゴリの画像であって、かつ、所望の固有特徴を有する画像を生成することができるようにする。

生成部3は、第一の画像と同じカテゴリに属する第二の画像から得られた当該カテゴリに属する画像に共通する特徴であるカテゴリ特徴と、第一の画像と第二の画像とで異なる固有の特徴である固有特徴と、を関連付けることで第一の画像を生成する。

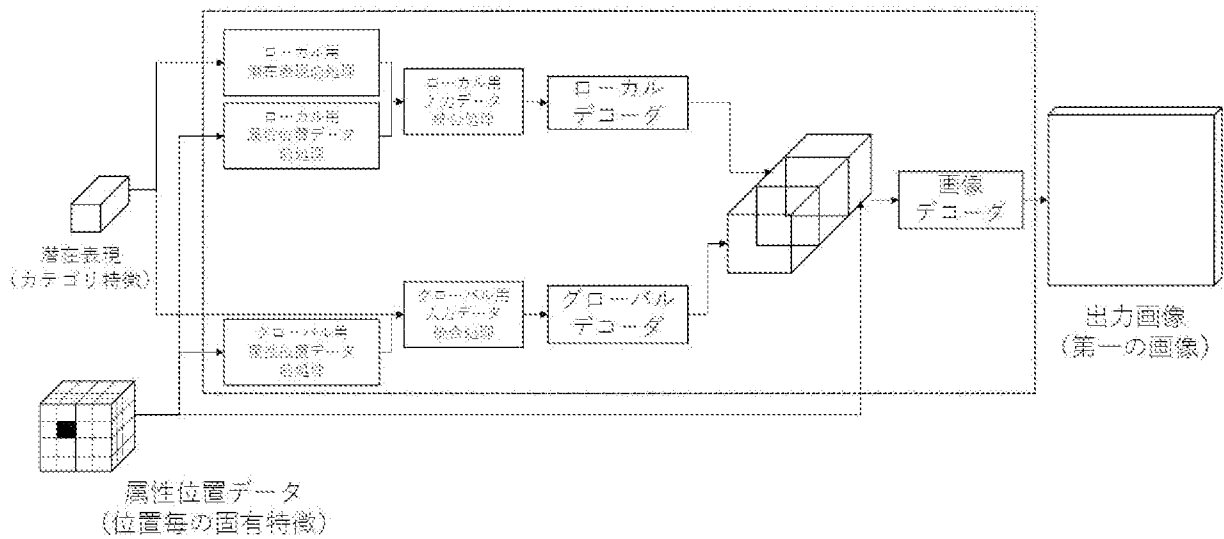
[図1]



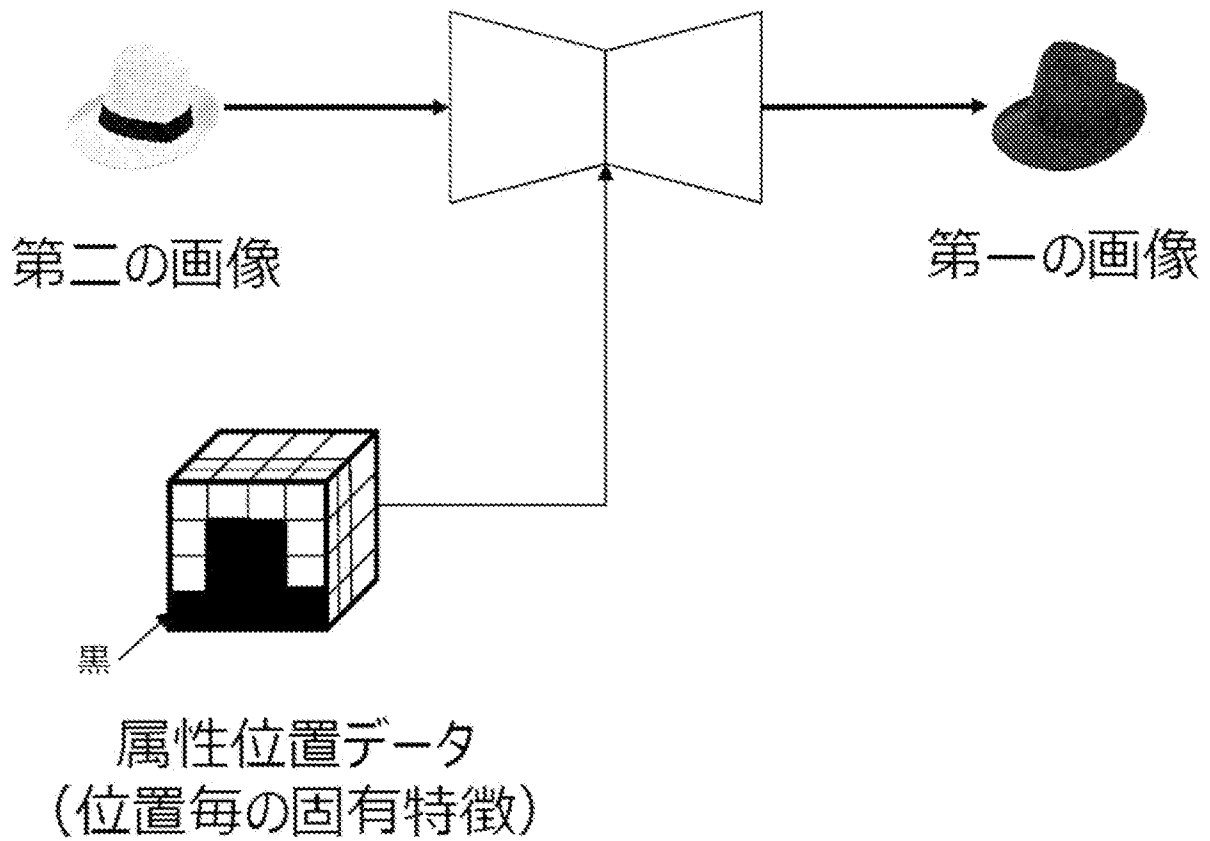
[図2]



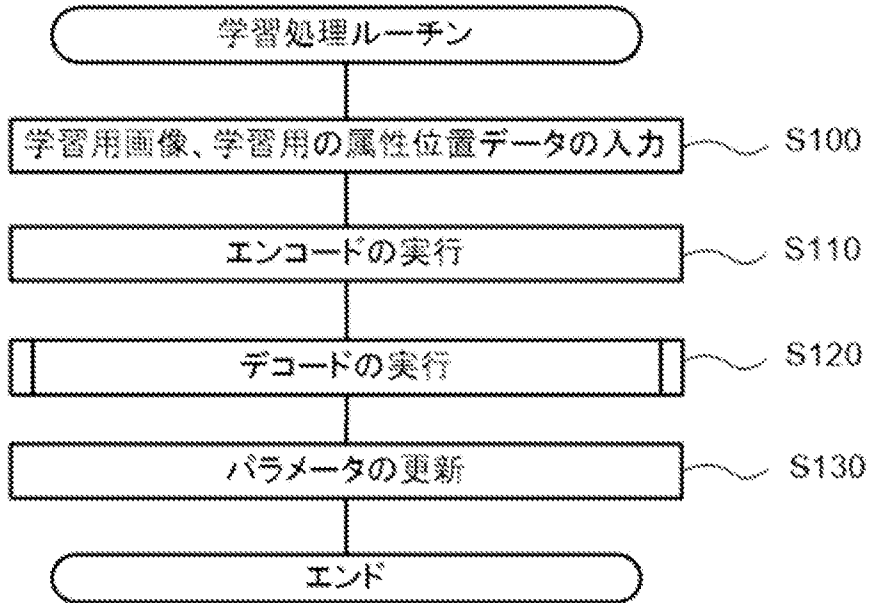
[図3]



[図4]



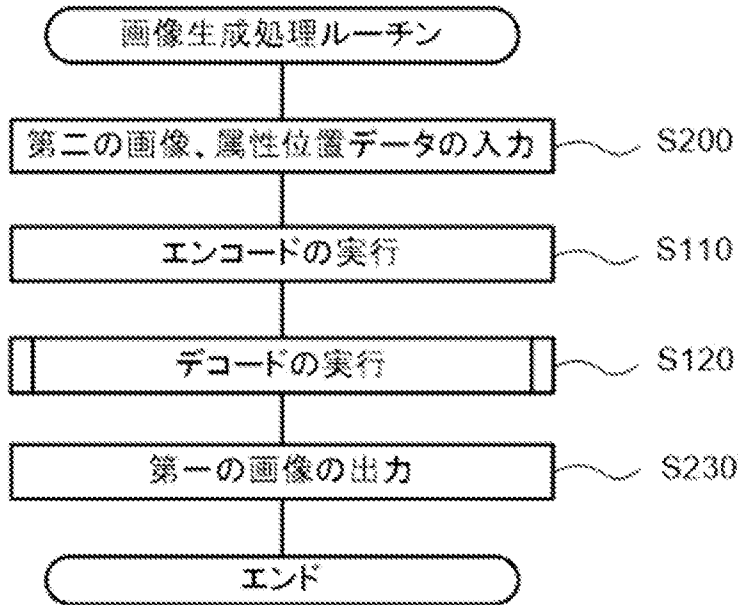
[図5]



[図6]



[図7]



[図8]

どこを変換すべきか分からない

