(19)日本国特許庁(JP)　(12)特　許　公　報(B1)　(11)特許番号
特許第7698856号
(P7698856)
(45)発行日　令和7年6月26日(2025.6.26)　(24)登録日　令和7年6月18日(2025.6.18)

(51)Int.Cl.　FI
G06T 7/00 (2017.01)　G06T 7/00 350C

請求項の数 4 (全 31 頁)

(21)出願番号　特願2025-14514(P2025-14514)
(22)出願日　令和7年1月31日(2025.1.31)
審査請求日　令和7年2月17日(2025.2.17)

早期審査対象出願

(73)特許権者　517169746
ＣＴW株式会社
東京都港区六本木一丁目９番１０号
(74)代理人　100215027
弁理士　留場　恒光
(72)発明者　廖　威
東京都港区六本木一丁目９番１０号　ＣＴW株式会社内
(72)発明者　姜　暁嵐
東京都港区六本木一丁目９番１０号　ＣＴW株式会社内
(72)発明者　楊　銘
東京都港区六本木一丁目９番１０号　ＣＴW株式会社内

最終頁に続く

(54)【発明の名称】デフォルメキャラクタ生成プログラムおよびデフォルメキャラクタ生成システム

(57)【特許請求の範囲】
【請求項１】
コンピュータを、
キャラクタの画像である元キャラクタ画像を取得する元キャラクタ画像取得手段、
前記元キャラクタ画像について説明する文字情報を取得するキャラクタプロンプト取得手段、
出力するデフォルメキャラクタ画像の頭身比率を設定するユーザーインターフェースを表示してユーザの入力を受け付け、当該ユーザの入力による頭身比率の設定を取得する頭身比率取得手段、および、
前記元キャラクタ画像、前記文字情報および前記頭身比率を機械学習モデルへの入力として、前記元キャラクタ画像を前記頭身比率にデフォルメ化しているデフォルメキャラクタ画像を取得するデフォルメキャラクタ取得手段、
として機能させ、
前記機械学習モデルは、少なくともキャラクタ画像およびデフォルメ化しているキャラクタ画像を学習データとして学習し、
キャラクタ画像を入力として、当該キャラクタ画像をデフォルメ化しているデフォルメキャラクタ画像を出力するものであって、拡散モデルを備えることを特徴とする、デフォルメキャラクタ生成プログラム。
【請求項２】
さらに、前記機械学習モデルが備えるニューラルネットワークの少なくとも一部の層に

おいて、重みを変更して得る複数の出力画像の差分を取得して評価し、出力画像に影響の小さい層の重みの削減または削除を行う最適化手段を備えることを特徴とする、請求項１に記載のデフォルメキャラクタ生成プログラム。

【請求項３】

キャラクタの画像である元キャラクタ画像を取得する元キャラクタ画像取得部と、

前記元キャラクタ画像について説明する文字情報を取得するキャラクタプロンプト取得部と、

出力するデフォルメキャラクタ画像の頭身比率を設定するユーザーインターフェースを表示してユーザの入力を受け付け、当該ユーザの入力による頭身比率の設定を取得する頭身比率取得部と、

前記元キャラクタ画像、前記文字情報および前記頭身比率を機械学習モデルへの入力として、前記元キャラクタ画像を前記頭身比率にデフォルメ化しているデフォルメキャラクタ画像を取得するデフォルメキャラクタ取得部と、を備え、

前記機械学習モデルは、少なくともキャラクタ画像およびデフォルメ化しているキャラクタ画像を学習データとして学習し、

キャラクタ画像を入力として、当該キャラクタ画像をデフォルメ化しているデフォルメキャラクタ画像を出力するものであって、拡散モデルを備えることを特徴とする、デフォルメキャラクタ生成システム。

【請求項４】

プロセッサがキャラクタの画像である元キャラクタ画像を取得する元キャラクタ画像取得ステップ、

プロセッサが前記元キャラクタ画像について説明する文字情報を取得するキャラクタプロンプト取得ステップ、

プロセッサが出力するデフォルメキャラクタ画像の頭身比率を設定するユーザーインターフェースを表示部に表示させてユーザの入力を受け付け、当該ユーザの入力による頭身比率の設定を取得する頭身比率取得ステップ、および、

プロセッサが、前記元キャラクタ画像、前記文字情報および前記頭身比率を機械学習モデルへの入力として、前記元キャラクタ画像を前記頭身比率にデフォルメ化しているデフォルメキャラクタ画像を取得するデフォルメキャラクタ取得ステップ、

を備え、

前記機械学習モデルは、少なくともキャラクタ画像およびデフォルメ化しているキャラクタ画像を学習データとして学習し、

キャラクタ画像を入力として、当該キャラクタ画像をデフォルメ化しているデフォルメキャラクタ画像を出力するものであって、拡散モデルを備えることを特徴とする、デフォルメキャラクタ生成方法。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、デフォルメキャラクタ生成プログラムおよびデフォルメキャラクタ生成システムに関するものである。

【背景技術】

【０００２】

ゲームやアニメーション、教育、報道など様々な分野のコンテンツにおいて、例えば２頭身キャラクタなどのいわゆるデフォルメ化されたキャラクタが使用されている。

これらは視聴者に対して可愛らしい印象を与え、コンテンツに対し親しみを覚えやすくさせる効果などを有する。

【０００３】

オリジナルのキャラクタとして２頭身キャラクタが作成されることもあるが、元々８頭身などで描かれているキャラクタをデフォルメ化して２頭身キャラクタにすることなども行われる。

【０００４】
しかしながら、このようなデフォルメキャラクタを作成する場合、表情や服飾などの特徴的部分を残しつつ、不要な部分は削除するといった工夫が必要であり多くの労力が生じる。
【０００５】
例えば特許文献１には、キャラクタ画像を二頭身や三頭身のデフォルメ化したキャラクタに変換可能な画像変換装置が開示されている。
【先行技術文献】
【特許文献】
【０００６】
【特許文献１】特開２０２３－１５７３３４号公報
【非特許文献】
【０００７】
【非特許文献１】Miao Hua, Jiawei Liu, Fei Ding, Wei Liu, Jie Wu, Qian Heomiita、arXiv、“DreamTuner: Single Image is Enough for Subject-Driven Generation”、［online］、[2025年1月10日検索]、インターネット<URL: https://arxiv.org/html/2312.13691v1>
【０００８】
近年、例えばオリジナルとなる人物イラストなどを入力すると、表情や服装などが異なる別のイラストを生成することができるようになるなど、生成ＡＩなどを道具として活用することでイラストレーターなどの労力を軽減できるようになっている。
【０００９】
非特許文献１には、キャラクタ画像やポーズを入力として、好みのキャラクタ画像を出力させる技術について開示されている。
【００１０】
しかしながらその一方で、例えば８頭身のキャラクタを入力として、２頭身のキャラクタを生成ＡＩの出力とするような場合、いくつかの課題がある。
ひとつは、デフォルメ化に際しては、残す部分（強調する部分）と残さない部分（強調しない部分）を取捨選択しないといけない点である。
【００１１】
キャラクタの身体的特徴や服装などは多種多様である。例えば、髪型、目の色、服飾のデザイン、持ち物、装飾品のそれぞれに特徴のあるキャラクタが存在する。
デフォルメ化に際し、どの部分を残し、どの部分を削除するかはキャラクタにより異なる。
髪型に特徴のあるキャラクタをデフォルメ化したときにその髪型が変わっていたとすると、デフォルメ化としては適切でないと言える。
【００１２】
しかしながら、望み通りのデフォルメキャラクタを画像生成ＡＩの出力として得ようとすると、プロンプトの工夫に頼らざるを得ないなど、多大な労力が生じる問題がある。
【００１３】
もうひとつは、デフォルメキャラクタの頭身比率によって看者（そのデフォルメキャラクタを視る者）が受ける印象が大きく異なることから、デフォルメキャラクタの頭身比率を正しく制御する必要がある点である。
【発明の概要】
【発明が解決しようとする課題】
【００１４】
解決しようとする問題点は、通常のキャラクタからデフォルメ化キャラクタを生成するにあたり、デフォルメキャラクタ画像の頭身比率を制御することが困難な点である。
【課題を解決するための手段】
【００１５】

本発明は、ユーザの入力による頭身比率の設定を取得すること、元キャラクタ画像（例えば８頭身キャラクタ画像）を取得すること、そしてこれらの情報から、元キャラクタ画像を前記頭身比率にデフォルメ化しているデフォルメキャラクタ画像を取得することを最も主要な特徴とする。

【００１６】

非特許文献１には、デフォルメキャラクタに変換すること、または頭身比率を制御することについては記載されていない。

【００１７】

本発明は、上記課題を鑑みてなされたものであり、例えば以下の手段を採用している。

すなわち、　コンピュータを、キャラクタの画像である元キャラクタ画像を取得する元キャラクタ画像取得手段、

前記元キャラクタ画像について説明する文字情報を取得するキャラクタプロンプト取得手段、

ユーザの入力による頭身比率の設定を取得する頭身比率取得手段、および、

前記元キャラクタ画像、前記文字情報および前記頭身比率を機械学習モデルへの入力として、前記元キャラクタ画像を前記頭身比率にデフォルメ化しているデフォルメキャラクタ画像を取得するデフォルメキャラクタ取得手段、として機能させ、

前記機械学習モデルは、少なくともキャラクタ画像およびデフォルメ化しているキャラクタ画像を学習データとして学習し、

キャラクタ画像を入力として、当該キャラクタ画像をデフォルメ化しているデフォルメキャラクタ画像を出力することを特徴とする、デフォルメキャラクタ生成プログラムを提供する。

【発明の効果】

【００１８】

本発明のデフォルメキャラクタ生成プログラムは、ユーザの入力による頭身比率を反映しているデフォルメキャラクタ画像を出力できるという利点がある。またこのときユーザはＧＵＩによる簡単な操作で頭身比率を入力できるという利点もある。

【図面の簡単な説明】

【００１９】

【図１】デフォルメキャラクタ生成プログラムＰ１による処理の概要を示す図である。

【図２】基本設定画面を示す図である。

【図３】元キャラクタ画像確認画面を示す図である。

【図４】デフォルメキャラクタ設定欄を示す図（拡大図）である。

【図５】生成画像表示画面を示す図である。

【図６】変換画像取得処理を示すフローチャートである。

【図７】複数の領域を認識するイメージを表す図である。

【図８】最適化処理を示すフローチャートである。

【図９】重みの削減等を説明するための図である。

【図１０】最適化処理前後のデフォルメキャラクタ画像を示す図である。

【図１１】デフォルメキャラクタ生成処理を（図１とは別の形で）示す図である。

【図１２】デフォルメキャラクタ生成システム１の概要を示すネットワーク構成図である。

【図１３】サーバ１０のハードウェア構成図である。

【発明を実施するための形態】

【００２０】

本発明の実施形態を図面に基づいて説明する。以下の各実施形態では、同一又は対応する部分については同一の符号を付して説明を適宜省略する場合がある。

また、以下に用いる図面は本実施形態を説明するために用いるものであり、実際の装置の構成やユーザーインターフェース（ＵＩ）、データ構成などとは異なる場合がある。

【００２１】

（実施形態の概要）
本実施形態の概要について、図１を用いて説明する。
図１は、本実施形態のデフォルメキャラクタ生成プログラムＰ１による処理の概要を示す図である。
【００２２】
デフォルメキャラクタ生成プログラムＰ１は、高頭身キャラクタ（図１では左上の５～６頭身キャラクタ）の画像を入力として、低頭身キャラクタ（図１では右下の２～３頭身キャラクタ）の画像を出力するプログラムである。
【００２３】
デフォルメキャラクタ生成プログラムＰ１は、（１）元キャラクタ画像を取得して各種情報（キャラクタプロンプト・キャラクタモデルなど）を取得する元キャラクタ画像分析手段（図１右上）、（２）参照画像を取得してコントロールモデルを取得する参照画像分析手段（図１左下）（第一の実施形態）、および（３）（１）および（２）で得られている情報をもとにデフォルメキャラクタを出力するデフォルメキャラクタ取得手段（図１右下）を備える。
【００２４】
また、デフォルメキャラクタ生成プログラムＰ１による処理を行うシステムを以下「デフォルメキャラクタ生成システム１」とする。
【００２５】
なお、頭身比率の制御は、ユーザが設定する頭身比率に対応する参照画像による制御のほか、ユーザが設定する頭身比率を文字情報（プロンプト）に変換して制御する方法がある。
以下、第一の実施形態では前者の説明を行い、第二の実施形態において後者の設雌尾を行う。
【００２６】
（実施形態の詳細）
以下、本実施形態に係るデフォルメキャラクタ生成システム１について、詳細を説明する。デフォルメキャラクタ生成システム１は、デフォルメキャラクタ生成プログラムＰ１を備えるコンピュータ（サーバ１０）を含み、ユーザに対し、入力に係るキャラクタ画像をデフォルメ化して出力するシステムをオンラインで提供する。
すなわちデフォルメキャラクタ生成システム１は、デフォルメキャラクタ生成プログラムＰ１による情報処理が、ハードウェア資源を用いて具体的に実現されるものである。
以下、デフォルメキャラクタ生成システム１を構成する、１．ユーザーインターフェース、２．プログラム処理、および３．ハードウェア構成、について順に説明する。
【００２７】
（用語の定義）
ここで、いくつか言葉の定義を行う。
「キャラクタ」とは、目視可能な外観を有するものであり、生物・非生物や実在・非実在を問わない。ここでは特に、全身のうち頭部とそれ以外を分けられるものを意味する。例えば創作アニメの人物キャラクタである（生物・非実在）。ここで非生物キャラクタとは、例えばロボットキャラクタなどを意味する。さらに、実在キャラクタとは、例えば芸能人をアニメ調に描いたものなどである。なお、例えばシーツを頭からすっぽり被った人物のように、首の場所が隠れるなどし、頭部が明確でない場合であっても、おおよその位置が推認できるものであればよい。また例えば、顔（目や口）のある部位を頭部としてもよい。
「キャラクタ画像」とは、キャラクタの画像である。イラスト画像か写真画像かを問わない。
「ポーズ」とは、キャラクタがとる姿勢を意味する。
「デフォルメ」とは、対象の形態を意図的に変えることをいう。例えば、キャラクタの頭身比率を変化させ、高頭身キャラクタから低頭身キャラクタにすることを指す。

「高頭身キャラクタ」「低頭身キャラクタ」はそれぞれ一般的に、高い頭身（例えば６～９頭身）のキャラクタおよび低い頭身（例えば２～５頭身）のキャラクタをいう。ただし、ここでいう高い・低いは絶対的なものではなく相対的なものである。デフォルメと言う場合、元キャラクタ画像（以下「入力画像」と称する場合がある。）が高い頭身、元キャラクタ画像をデフォルメ化している画像（以下「出力画像」と称する場合がある。）が低い頭身となる。例えば、４頭身キャラクタを２頭身キャラクタに変化させることもデフォルメと称する。なお一般に、低頭身キャラクタはちびキャラクタ（chibi character）などと呼ばれる。

「特徴量分離」（Feature Decoupling）とは、機械学習モデル、特に画像生成モデルにおいて、異なる特徴量（例えばポーズ（姿勢）、衣服、表情に関する特徴量など。）を独立して扱う手法である。特徴量は変数などともいう。

「ユーザ」は、デフォルメキャラクタ生成システム１を使用する者をいう。端末２０を操作し、個人と法人とを問わない。デフォルメキャラクタ生成システム１において、ユーザは高頭身キャラクタの画像を入力し、変換後の低頭身キャラクタを出力として取得する。

「取得」は、プロセッサがデータまたはプログラム等を入力部または外部端末などから取得することのほか、プロセッサがデータまたはプログラム等を作成・更新等することを含む。下記実施形態のキャラクタモデルなどを参照されたい。

【００２８】

以下に図面等で使用されている用語を説明する。

ただし以下は発明の理解を助けるためのものであり、本明細書の内容を限定するものではない。

「Stable Diffusion」は、テキスト等を画像に変換する機械学習モデルであり、いわゆる画像生成ＡＩの１つである。Stable Diffusionは拡散モデル（Diffusion Model）（特に、潜在拡散モデル（Latent Diffusion Model））を用いていることを特徴とする。また、Stable Diffusionは条件付けを行うテキストを入力するためのテキストエンコーダ（Text Encoder）を有する。テキストエンコーダはテキストをベクトルに変換するものであり、例えばCLIPという手法（大量の画像とテキストの組み合わせを学習し、画像とテキストの類似度を算出）で学習済みのTransformer（機械学習モデル）である。Stable Diffusion における処理は、VAE（次項）により画像情報から潜在変数を得る処理、および、VAEにより潜在変数から画像情報を得る処理を含む。VAEを用いることで計算量を抑えていることが特徴である。拡散モデルの具体的態様については下記の実施形態において説明する。

「VAE（Variational Auto-Encoder）」は、変分オートエンコーダ（変分自己符号化器）のことである。VAEはエンコーダー・デコーダ構造を有する生成モデルである。ただしVAEの語は、このような構造を有するニューラルネットワーク、このようなネットワーク構造、機械学習モデル、またはこれらを含む技術を指すこともある。VAEは、入力データを平均と標準偏差（または分散など）で表現するように学習する（平均ベクトル、分散ベクトルなど）。例えば、エンコーダは入力データを（統計分布のある一点となる）潜在変数に変換し、デコーダは潜在変数（統計分布から（ランダムに）サンプリングした一点）を復元することで新しいデータを生成する。また、学習時において誤差逆伝播（Backpropagation）のための工夫（Reparametrization trick）がなされている。

「Unet」は、CNN（Convultional Neural Network）をベースとするニューラルネットワークである。例えば畳み込み層を含むモジュールと、アテンション層を含むモジュールを含む。一般的には画像のセグメンテーションに使用されるが、ここでは拡散モデルの逆拡散過程、例えば潜在空間におけるノイズ除去に使用される。Unetは例えば、ノイズを入力として、除去すべきノイズを出力する。入力されているノイズから除去すべきノイズを除去することで、当該ノイズが除去されている画像が得られる。これを繰り返すことで、徐々に画像からノイズが除去されていく。

「ControlNet」は、Stable Diffusionの拡張機能であり、ユーザが指定したポーズや構図に基づいて（高品質な）画像を生成することを可能にする。ControlNetは、画像の構造

的特徴（ポーズ、線画、エッジなど）を条件（Condition）として与え、それを反映して画像を生成するための追加的なニューラルネットワークである。Unetのニューラルネットワークブロックを保持しつつ、ControlNet上のニューラルネットワークでユーザから与えられた追加条件を付加してUnetに反映させる。つまり、条件（Condition）（例えば構図データ）などが入力となり、Unet上のブロックを出力先とする。詳細は割愛するが、ゼロ畳み込みやスキップコネクションといった技術が組み込まれている。

「LoRA（Low-Rank Adaptation）」は、（事前）学習済み（機械学習）モデルのファインチューニングに関する技術である。なお、低消費電力長距離無線通信技術のLoRa（登録商標）とは異なる。

【００２９】

以下において、「○○」処理と記載している場合、コンピュータのプロセッサは、プログラム格納部に記憶されている「○○」プログラムに基づく処理を実行することを意味する。本段落において、「○○」の箇所には同じ語が入る。

すなわち、「○○」プログラムは、「○○」処理の実行により、コンピュータを「○○」手段として機能させるプログラムである。またこの際、当該プロセッサを備える制御部は、「○○」部（または「○○」装置）としても機能することを意味する。

この場合において、「○○」部は、「○○」プログラムに基づく「○○」処理を実行することを意味する。

また、（「○○」処理の）処理手順を示す場合は「○○」ステップと記載する。

【００３０】

例えば、デフォルメキャラクタ生成プログラムＰ１は、デフォルメキャラクタ生成処理の実行により、コンピュータをデフォルメキャラクタ生成手段として機能させるプログラムである。またこの際、プロセッサ１２２を備えるコンピュータの制御部１２は、デフォルメキャラクタ生成部（またはデフォルメキャラクタ生成装置）として機能する。

【００３１】

デフォルメキャラクタ生成システム１において、端末２０などの各端末（コンピュータ）はそれぞれプロセッサを備えるが、単にプロセッサという場合は、デフォルメキャラクタ生成プログラムＰ１により処理を行うプロセッサ、本実施形態ではサーバ１０のプロセッサ１２２、を指すものとする。

【００３２】

なお以下において、簡単のため、「サーバ１０のプロセッサ１２２が、端末からのリクエストを受けて、当該端末のブラウザに表示するためのデータを返す」ことを、「プロセッサ１２２が端末のブラウザに表示する（させる）」または「プロセッサ１２２が表示する（させる）」などと記載する場合がある。

また同様に、「サーバ１０のプロセッサ１２２が、記憶部１４のデータ記憶部１４ｂにデータを保存させる」ことを、「プロセッサが（データを）保存する（させる）」などと記載する場合がある。

【００３３】

（第一の実施形態）

１．ユーザーインターフェース（Ｕｓｅｒ　Ｉｎｔｅｒｆａｃｅ（ＵＩ））

まず、本実施形態のデフォルメキャラクタ生成システム１が端末２０に表示させるインターフェースについて、図を用いて説明する。

以降で説明するインターフェースは、プロセッサ１２２が端末２０のブラウザに表示させるものを簡略化したものである。

【００３４】

また、説明に必要な機能に関わるアイコン等のみ表示することとし、それ以外の公知のアイコンなどは省略する。例えば、直前に表示されていたページに戻るための戻るボタンなどは省略している。

【００３５】

図２は、基本設定画面を示す図である。

基本設定画面左側の画像アップロード欄ＵＩ－１１は、ユーザが入力画像である元キャラクタ画像をアップロードするためのＵＩであるとともに、アップロードされた元キャラクタ画像を表示する欄である。
ユーザはドラッグアンドドロップで元キャラクタ画像をアップロードする。
なお、アップロードの方法はこれに限らない。画像が格納されているフォルダを選択し、画像ファイルを選択できるようにしてもよい（不図示）。

【００３６】

デフォルメキャラクタ設定欄ＵＩ－１２は、ユーザが出力するデフォルメキャラクタ画像に関する設定を行うためのＵＩである。
デフォルメキャラクタ設定欄ＵＩ－１２は頭身比率設定ボタンＵＩ－１２１、参照デフォルメキャラクタ画像ＵＩ－１２２、およびデフォルメキャラクタ生成開始ボタンＵＩ－１２３を表示する。詳細は後述する。

【００３７】

図３は、元キャラクタ画像確認画面を示す図である。
ユーザが元キャラクタ画像を画像アップロード欄ＵＩ－１１にアップロードすると、プロセッサ１２２はその画像から後述するキャラクタプロンプト（文字情報）およびキャラクタモデル（機械学習モデル）を取得・作成し、元キャラクタ画像確認画面（ウィンドウ）ＵＩ－１３を表示する。
なお図中において、キャラクタモデルの作成については「トレーニング」という用語を用いている。

【００３８】

なお本実施形態において、プロセッサ１２２はアップロード後にトレーニング（学習）を行うかについての確認画面を表示し、ユーザがそこで表示されているＯＫボタンアイコン等を押下することによりパラメータ更新などによる学習を開始するが、ここでは省略している。

【００３９】

図３に示すように、プロセッサ１２２は元キャラクタ画像確認画面（ウィンドウ）ＵＩ－１３に画像情報ＵＩ－１３１（ファイル名、ユーザＩＤ、作成者、および作成日時など）、ユーザがアップロードしている元キャラクタ画像ＵＩ－１３２、ＯＫボタンアイコンＵＩ－１３３、再トレーニングアイコンＵＩ－１３４、および削除ボタンアイコンＵＩ－１３５を表示する。

【００４０】

ユーザがＯＫボタンアイコンＵＩ－１３３を押下すると、プロセッサ１２２は基本設定画面の画像アップロード欄ＵＩ－１１に元キャラクタ画像を表示する（不図示）。
なおユーザが基本設定画面で当該元キャラクタ画像を選択（押下）すると、プロセッサ１２２は再度元キャラクタ画像確認画面（ウィンドウ）ＵＩ－１３を表示する。

【００４１】

元キャラクタ画像確認画面（ウィンドウ）ＵＩ－１３でユーザが再トレーニングアイコンＵＩ－１３４を押下すると、プロセッサ１２２は元キャラクタ画像を選択する画面を表示する（不図示）。ユーザがそこで元キャラクタ画像を選択して再トレーニングを実行させるボタンを押下すると、プロセッサ１２２はユーザの選択による当該元キャラクタ画像（入力画像）からキャラクタプロンプトおよびキャラクタモデルを（再度）作成する。
この際ユーザの任意により、または自動的に、パラメータが調整される。このほか、トレーニングの方法はユーザが調整することができる（不図示）。

【００４２】

ユーザが削除ボタンアイコンＵＩ－１３５を押下すると、プロセッサ１２２はアップロードされている元キャラクタ画像を削除する。ユーザは新しい元キャラクタ画像をアップロードすることができる。

【００４３】

図４は、デフォルメキャラクタ設定欄（ＵＩ－１２）を示す図（拡大図）である。

【００４４】
本実施形態において、参照デフォルメキャラクタ画像（参照されるデフォルメキャラクタの画像）はあらかじめ記憶部１４（データ格納部１４ｂ）に多数格納されており、ユーザはデフォルメキャラクタの向きやポーズなどを自由に選択することができる。
【００４５】
また、ユーザは自分の好みのデフォルメキャラクタをアップロードすることもできる。
例えば本実施形態において、デフォルメキャラクタ設定欄ＵＩ－１２に画像をドラッグアンドドロップすることで好みのキャラクタを参照画像とすることもできる。
【００４６】
頭身比率設定ボタンＵＩ－１２１は、出力画像の頭身比率を設定するボタンアイコンである。例えば、１：４のボタンを選択すると、プロセッサ１２２はデフォルメキャラクタ設定欄ＵＩ－１２に頭の高さ：全身の高さが１：４となる画像を出力させる（図２参照。また、図２・図４における参照デフォルメキャラクタ画像ＵＩ－１２２の影の部分参照）。
【００４７】
図４に示すように、頭身比率設定ボタンＵＩ－１２１のＡｌｌボタンは、示されている（図４であれば１：２．５～１：５）比率の画像について一通り表示するものである。
【００４８】
ユーザが参照デフォルメキャラクタ画像ＵＩ－１２２を選択して押下すると、プロセッサ１２２はチェック（レ点）を表示してその画像が選択されたことを表示する。
ユーザは複数の参照デフォルメキャラクタ画像を選択することができ、その場合プロセッサ１２２はその選択に応じて複数のデフォルメキャラクタを生成する。
つまりプロセッサ１２２は、一度に複数のデフォルメキャラクタ画像を出力することができる。
【００４９】
参照デフォルメキャラクタ画像ＵＩ－１２２は、出力する予定のデフォルメキャラクタのポーズと頭身比率の目安を表示するものである。この画像は多くのストックのなかからプロセッサ１２２が自動で選択する。
【００５０】
その一方でプロセッサ１２２はユーザの選択により、（立ち絵や座り絵といった）ポーズを選択するためのポーズ表示ウィンドウを表示し（不図示）、ユーザは当該ポーズ表示ウィンドウに一覧表示されるポーズ群から好みのポーズを選択することもできる。
【００５１】
つまりユーザは、頭身比率設定ボタンＵＩ－１２１を押下することで、または参照画像をアップロードすることで、簡単に頭身比率を入力することができる。
言い換えると、ユーザはＧＵＩによる簡単な操作で頭身比率を入力することができる。
【００５２】
ユーザがデフォルメキャラクタ生成開始ボタンＵＩ－１２３を押下すると、プロセッサ１２２はデフォルメキャラクタ生成に関する処理を開始する。処理については後述する。
【００５３】
図５は、生成画像表示画面を示す図である。
プロセッサ１２２がデフォルメキャラクタの画像（出力画像）を取得すると、生成画像表示画面の出力画像表示欄ＵＩ－１４に出力画像ＵＩ－１４１を表示する。またプロセッサはこれ以外にも、出力画像加工関連ボタンＵＩ－１４２、出力画像情報ＵＩ－１４３、および出力画像保存関連ボタンＵＩ－１４４を表示する。
【００５４】
出力画像加工関連ボタンＵＩ－１４２は、出力されている画像に対してさらに加工を行うか、画像を再生成させるためのボタンアイコンである。
【００５５】
図５に示す拡大率ボタンは表示されている拡大率を変化させるためのボタンアイコンで

ある。
色ボタンはキャラクタの肌の色や髪の色など、色の加工を行うためのボタンアイコンである。
パーツ分けボタンはデフォルメキャラクタのパーツ（例えば髪型、服装など）を他のパーツと分離して表示させるためのボタンアイコンである。ユーザはそのパーツについて別個独立に加工等をすることができる。
再生成ボタンはデフォルメキャラクタ画像を再生成するためのボタンアイコンである。機械学習モデルにおいて異なるパラメータが設定され（拡散モデルにおいてランダムで潜在変数がサンプリングされ）、パラメータに応じたデフォルメキャラクタ画像（通常異なるデフォルメキャラクタ画像）を生成することができる。つまり、ユーザはボタン一つで簡単に異なる画像を取得することができるという利点がある。

【００５６】
出力画像情報ＵＩ－１４３は画像に関する情報を表示する欄である。
スタイルは、参照画像のファイル名である。
バリエーション強度は、バリエーションを加えるためのパラメータである。バリエーション強度が大きいほど、元キャラクタ画像からの変化が大きい画像が出力画像として出力される。
本実施形態において、バリエーション強度はユーザによる設定が可能である（ＵＩは不図示）。

【００５７】
出力画像保存関連ボタンＵＩ－１４４は、出力している画像の保存等に関する操作ボタンアイコンである。ＤＬは出力している画像を端末（端末２０）にダウンロードするものであり、削除は出力している画像を削除するものであり、保存ボタンはサーバ（サーバ１０）上に出力している画像を（サーバ上などに）保存するものである。

【００５８】
以上のような構成により、ユーザはＧＵＩ（Graphical User Interface）による簡単な操作でデフォルメキャラクタ画像を生成することができる。
特に、ユーザは参考例としての参照デフォルメキャラクタ画像ＵＩ－１２２を見ながら頭身比率を選択できるため、出力される画像が予想しやすく、希望するデフォルメキャラクタ画像をより確実に得ることができる。
またこの際、プロンプトを細かく調整するなどの入力を行わなくてもＧＵＩで処理を実行できるという著しい効果がある。

【００５９】
２．プログラム処理
＜２－１．デフォルメキャラクタ生成処理＞
本実施形態のデフォルメキャラクタ生成システム１において行われるプログラム処理について説明する。

【００６０】
本実施形態において、プロセッサ１２２は、デフォルメキャラクタ生成プログラムＰ１に基づき、デフォルメキャラクタ生成処理を行う。
デフォルメキャラクタ生成プログラムＰ１は、少なくとも（１）元キャラクタ画像分析プログラムＰ１１、（２）参照画像分析プログラムＰ１２、（３）デフォルメキャラクタ取得プログラムＰ１３、および（４）最適化プログラムＰ１４を含み、プロセッサ１２２はこれらの各プログラムに基づいて、元キャラクタ画像分析処理、参照画像分析処理、デフォルメキャラクタ取得処理、および最適化処理をそれぞれ実行する。
また上記のほか、デフォルメキャラクタ生成処理は、ユーザに上述したＵＩをオンラインで提供するオンラインＵＩ提供処理を含む。提供するＵＩについては説明済みであるためここでは省略する。

【００６１】
（１）元キャラクタ画像分析処理は、元キャラクタ画像を取得して分析する処理であり

、
キャラクタの画像である元キャラクタ画像を取得する元キャラクタ画像取得処理、および前記元キャラクタ画像について説明する文字情報を取得するキャラクタプロンプト取得処理、を含む。
（２）参照画像分析処理は、デフォルメキャラクタの参照画像を取得して分析する処理であり、ユーザの入力による頭身比率の設定を取得する頭身比率取得処理、を含む。
（３）デフォルメキャラクタ取得処理は、
前記元キャラクタ画像、前記文字情報および前記頭身比率を機械学習モデルへの入力として、前記元キャラクタ画像を前記頭身比率にデフォルメ化しているデフォルメキャラクタ画像を取得するデフォルメキャラクタ取得処理、を含む。
ここで、前記機械学習モデルは、少なくともキャラクタ画像およびデフォルメ化しているキャラクタ画像を学習データとして学習し、
キャラクタ画像を入力として、当該キャラクタ画像をデフォルメ化しているデフォルメキャラクタ画像を出力することを特徴とする。
【００６２】
より詳細に説明すると、本実施形態において機械学習モデルとして拡散モデルと当該拡散モデルを制御する制御用モデルとを含み、
前記拡散モデルは、キャラクタ画像を説明する文字情報（プロンプト）などを入力として当該文字情報の特徴を有するキャラクタ画像を出力するための機械学習モデルであり、
前記制御用モデルは、デフォルメ化しているキャラクタ画像であるデフォルメキャラクタ画像を出力するように前記拡散モデルを制御する機械学習モデルである。
【００６３】
さらに補足すると、前記拡散モデルは、少なくともキャラクタ画像、当該キャラクタ画像に基づく情報（データ）（潜在変数）に加えるノイズ、当該ノイズを加えて得られる情報、を学習データとして学習する。
ここで、キャラクタ画像に基づく情報（データ）とは、キャラクタ画像から取得している複数の画像をVAEにより潜在変数化しているものなどである。
【００６４】
推論段階では、ノイズあり情報を入力として、当該ノイズあり情報から除去すべきノイズを出力する（通常この工程は繰り返される）。このことにより、ノイズあり情報からノイズが除去されている情報を出力する。
この結果、キャラクタ画像およびキャラクタ画像を説明する文字情報を入力として、当該文字情報の特徴を有するデフォルメキャラクタ画像を出力することができる。
【００６５】
ここでユーザの入力による頭身比率は、コントロールモデルまたは拡散モデルのどちらかに関わることで、デフォルメキャラクタ画像の頭身比率を制御する。
【００６６】
つまり、頭身比率がコントロールモデルを構成する場合もあれば、頭身比率がキャラクタ画像を説明する文字情報（プロンプト）となって拡散モデルへの入力になる場合もある。
ここでは前者について説明する。後者の形態については後述する。
【００６７】
上述した機械学習モデルは、少なくともキャラクタ画像、前記キャラクタ画像について説明する文字情報、前記キャラクタ画像をデフォルメ化しているデフォルメキャラクタ画像、および前記キャラクタ画像および前記デフォルメキャラクタ画像の頭身比率を学習データとして学習する機械学習モデルであって、
キャラクタの画像である元キャラクタ画像、前記元キャラクタ画像について説明する文字情報、およびユーザによる頭身比率の設定を入力として、前記元キャラクタ画像を前記頭身比率にデフォルメ化しているデフォルメキャラクタ画像を出力するデフォルメキャラクタ生成用の学習済み機械学習モデルであるといえる。

【００６８】
本実施形態において、上記（１）から（３）の処理はさらに以下のように説明することができる。
（１）元キャラクタ画像分析処理は、元キャラクタ画像を取得して分析する処理であり、
キャラクタの画像である元キャラクタ画像を取得する元キャラクタ画像取得処理、
１枚の前記元キャラクタ画像から、少なくとも顔部分を含む複数の領域の画像情報を取得することにより、複数の特徴量を取得する複数特徴量取得処理、
複数の領域（の画像）に紐づく情報についてタグを付して管理するタグ管理処理、
前記複数の特徴量を、ポーズの特徴量とそれ以外の特徴量に分離（特徴量分離（Feature Decoupling））する特徴量分離処理、
前記複数の特徴量のうち、ポーズの特徴量以外の特徴量を含むコンセプトの特徴量について説明する文字情報を取得するキャラクタプロンプト取得処理、および、
少なくともコンセプトの特徴量を含むデータを学習データとして学習している（学習する）機械学習モデル（拡散モデル）を取得するキャラクタモデル取得処理、を含む。
【００６９】
（２）参照画像分析処理は、デフォルメキャラクタの参照画像を取得して分析する処理であり、
ユーザの入力による頭身比率の設定を取得する頭身比率取得処理（本実施形態では「デフォルメキャラクタの参照画像を取得する参照画像取得処理」と称する。）、および、
前記参照画像から、デフォルメキャラクタの形状の一部を示す線画情報を少なくとも含むコントロールモデルを取得するコントロールモデル取得処理、を含む。
つまり本実施形態においてプロセッサ１２２は、頭身比率を参照画像と言う形で取得し、参照画像からコントロールモデルを取得する。
【００７０】
（３）デフォルメキャラクタ取得処理は、
前記コントロールモデルにより制御されている前記機械学習モデル（拡散モデル）に対して前記文字情報（キャラクタプロンプト）を入力し、前記元キャラクタ画像をデフォルメ化しているデフォルメキャラクタ画像を出力として取得する処理である。
【００７１】
大きく分けると、画像（元キャラクタ画像・参照画像）を取得して分析等する処理（上記（１）および（２））と、画像生成によりデフォルメ画像を得る処理（上記（３））である。
言い換えると、画像生成によりデフォルメ画像を得る処理（上記（３））は、（上記（１）および（２）の）から得られるデータ・モデル等からデフォルメキャラクタを実際に生成する処理である。
【００７２】
図６は、デフォルメキャラクタ生成処理を示すフローチャートである。
プロセッサ１２２は、デフォルメキャラクタを生成する処理の開始の指示（例えば上記デフォルメキャラクタ設定欄ＵＩ－１２のデフォルメキャラクタ生成開始ボタンＵＩ－１２３の押下（図４参照））を受け付けることにより、デフォルメキャラクタ生成処理を開始する。
【００７３】
まず、（１）元キャラクタ画像分析処理について説明する。
プロセッサ１２２は、ユーザにより入力されている、デフォルメ化する対象である元キャラクタ画像を取得する（ステップ１・元キャラクタ画像取得処理）。
【００７４】
プロセッサ１２２は取得した画像から複数の特徴量を抽出し、取得する（ステップ２・複数特徴量取得処理）。
本実施形態において、プロセッサ１２２は、１枚の元キャラクタ画像から、例えば頭部

、上半身、または全身に係る部分など、複数の領域を認識してその画像情報を取得して複数の特徴量を取得する。

【００７５】

図７は複数の領域を認識するイメージを表す図である。

図７中の、スケールの異なる複数の矩形領域（四角形の線で囲まれた部分・バウンディングボックス）が複数の領域を示す。

【００７６】

本実施形態において、プロセッサ１２２は顔を含む画像を異なるスケールで取得（クロップ）している。

本実施形態において特に、顔を認識する顔認識（Face Detection）の処理を行っている。顔認識の方法は公知の方法が適宜用いられる。

なお、画像を（切り取って）取得する処理は一般的に「クロップ（crop）」または「クロッピング（cropping）」と呼ばれる。

【００７７】

図７に示す例の場合、複数の領域は、図７の矩形領域の面積が小さいものから順に、顔、頭部、上半身、頭頂より腰まで、膝より上（帽子を含む、以下同じ。）、ふくらはぎより上、くるぶしより上、踵より上、および爪先より上、に係る部分を含む。

【００７８】

スケールの異なる複数の領域の画像を取得（クロップ）することは、特徴量を取得すべき部位（位置）を明確にし、部位（位置）に応じた複数の特徴量を得ることを可能にする。例えば、顔とその他の部位とに関するデータを取得することなどができる。また、少数（１枚）の画像からデータ量を増やすという利点がある。

【００７９】

１枚の元キャラクタ画像から複数の領域を認識して画像を取得する方法として公知の物体検出技術を用いることができる。

【００８０】

なお物体検出技術のうち、大まかな物体の位置を特定したあとにその物体のクラスを識別するものは２段階モデルと呼ばれ、例えばR-CNN（Region Based Convolutional Neural Networks）やFPN（Feature Pyramid Networks）が挙げられる。また、物体の位置の特定とクラスの識別を同時に行うものは１段階モデルと呼ばれ、例えばＹＯＬＯ（You Only Look Once）やＳＳＤ（Single Shot multibox Detector）が挙げられる。このほか、カスケード分類器（Open CV）などを用いてもよい。

【００８１】

「複数の領域」は、元キャラクタ画像における認識単位（切り出し（クロッピング）単位）を示す。プロセッサ１２２は１つの領域について１つのクロッピング画像を得る。

「複数の特徴量」は、元キャラクタ画像から得られる複数の特徴量である。

例えば、クロッピングした画像をVAEの入力として得られる潜在変数である。

【００８２】

プロセッサ１２２は、領域（の画像）および／または特徴量ごとにタグを付して管理する（タグ管理処理）。

本実施形態において、クロッピングにより得ている画像とタグとが１対１で対応する。

本実施形態においてタグは、頭部、衣装、全身、顔、上半身、髪、装飾、およびポーズに関するものである（このあたりについては図１１も参照されたい）。

【００８３】

本実施形態においてプロセッサ１２２は、画像（ファイル）と、タグを含むテキスト（ファイル）とをセットで記憶部に記憶させることで管理する。

またプロセッサ１２２は、このデータ（テキスト）の一部を後述するキャラクタプロンプトに用いる。

【００８４】

以下にこのテキストについてタグごとに例示する。各タグは領域（の画像）および／ま

たは特徴量を説明する文字情報を含む。
なお、この領域（の画像）および／または特徴量を説明する文字情報を便宜上プロンプト要素またはタグ情報と称する場合がある。
・頭部（Head）：羽根付き帽子（hat feather）、海賊帽（pirate hat）、海軍帽（navy hat）
・衣装（Cloth）：ミニスカート（miniskirt）、制服（uniform）、ニーハイブーツ（knee boots）、ジャケット（jacket）
・全身（Full body）：全身（full body）、立っている（standing）、寝ている（lying）、座っている（sitting）
・顔：笑顔（smile）、閉じた口（closed mouth）、髭のある（beard）、赤面した（blush）
・上半身（Upper body）：上半身（upper body）
・髪（Hair）：長髪（long hair）、短髪（short hair）、１つの三つ編み（single braid）、ツインポニーテール（twin ponytails）
・装飾（Decoration）：ネックレス（necklace）、イヤリング（earrings）、包帯（bandage）、宝石（jewelry）
・ポーズ（Pose）：手を腰に回している（hand on hip）、片足を挙げている（leg lift）、膝を立てている（knee up）

【００８５】
この例ではタグやテキストは単語または熟語である。ただしこれに限るものではなく、文章などであってもよい。

【００８６】
なお一般に、画像データの入力を受け付け、その内容を要約する（説明する）テキストを出力する技術をImage Captioningという。Image Captioningを実行するための（マルチモーダル）モデルとしてCLIPやFlamingoなどが知られている。

【００８７】
このような技術はイメージｔｏプロンプト（Image to Prompt）と呼ばれ、画像からテキストを生成する技術であるが、これは画像中の文字情報を取得するＯＣＲ（Optical Character Recognition）とは異なる。

【００８８】
なおここでは画像からテキストに取得する方法について説明しているが、画像に基づく特徴量をテキストに変換する方法を用いてもよい。

【００８９】
またプロセッサ１２２は、タグに応じた特徴量を取得する。
例えばタグごと（つまりはクロッピングしている画像ごと）に特徴量を取得するようにしてもよいし、取得している複数の特徴量をタグに応じて分類してもよい。

【００９０】
なお、一つの領域（クロッピングにより得ている１枚の画像）から複数の特徴量を得てもよいし、複数の領域から１つの特徴量を得てもよい。
例えばプロセッサ１２２は、頭部の画像から頭部に係る特徴量と髪に係る特徴量を取得してもよい。

【００９１】
また図７に示すように、爪先より上の部分の画像はキャラクタの全身を含むため、ポーズの特徴量を含んでいる。
ただし、ポーズの特徴量は全身の画像がなければ取得できないというものではない。例えば上半身を含む画像は、キャラクタが手を腰に当てている、というポーズの特徴量を含み得る。

【００９２】
また、例えば顔の領域の画像があるからといって、プロセッサ１２２は上半身の領域の画像から顔の部分に係る特徴量を得ないというものではない。

つまりプロセッサ１２２が、ある領域（画像）からどのような情報を抽出するかについては制限されない。

【００９３】

本実施形態において、プロセッサ１２２が元キャラクタ画像の顔の領域の画像を取得し、顔の特徴量を取得することは必須である。

例えばデフォルメキャラクタ画像において特に顔部分が大きく表示されるように、顔の特徴量はデフォルメキャラクタの特徴量を決定づけるためである。

【００９４】

図６に戻り、プロセッサ１２２は、ステップ２で得ている複数の特徴量のうち、ポーズの特徴量と、ポーズの特徴量以外の特徴量を含むコンセプトの特徴量とを分離する（ステップ３・特徴量分離処理）。この分離を特徴量分離（Feature Decoupling）と称する。

【００９５】

特徴量分離を実現する方法として、生成モデルの潜在空間（Latent Space）を設計し、異なる特徴量を異なる次元で制御するように学習させることが挙げられる。

本実施形態において特に、ポーズの特徴量をそれ以外の特徴量と分けている。

特徴量分離により、各特徴量が他の特徴量に依存せずに処理されるため、モデルの柔軟性と精度が向上する。

【００９６】

本実施形態において、ポーズ以外の特徴・要素をコンセプトと表現する。例えばポーズの特徴量以外の特徴量をコンセプトの特徴量としている。

上記の例であれば、複数の特徴量のうち、頭、衣装、全身、装飾、顔、上半身、および髪に関する特徴量はコンセプトの特徴量となる。

つまり本実施形態において、複数の特徴量＝ポーズの特徴量＋コンセプトの特徴量である。

【００９７】

プロセッサ１２２は元キャラクタ画像から、当該元キャラクタ画像について説明している文字情報を取得する（ステップ４・キャラクタプロンプト取得処理）。

この元キャラクタ画像について説明している文字情報（プロンプト）を便宜上「キャラクタプロンプト」と称する。

【００９８】

プロセッサ１２２は特に、ポーズ以外の要素（画像および／または特徴量）について説明する文字情報を取得する。

本実施形態においてプロセッサ１２２は、画像のうちポーズ以外に係る部分、つまり頭部、衣装、全身、顔、上半身、髪、または装飾など（コンセプト）を説明する文字情報を取得する。

【００９９】

キャラクタプロンプトを例示すると以下のようになる。

「1girl, blue eyes, purple hair, long hair, police uniform, transparent background, open clothes, black skirt, mini skirt, high-heeled shoes, simple background, …（略）」

【０１００】

キャラクタプロンプトを取得して後述するデフォルメキャラクタ取得処理の入力に使用することにより、キャラクタの詳細な特徴を制御し、デフォルメキャラクタに反映することができるという利点がある。

【０１０１】

また一方で、プロセッサ１２２は、上述した複数の特徴量のうち、ポーズの特徴量以外の特徴量のうち少なくとも一部（ポーズの特徴量以外の特徴量を含むコンセプトの特徴量）を学習データとして学習している（学習する）機械学習モデル（拡散モデル）を取得する（ステップ５・キャラクタモデル取得処理）。

この機械学習モデル（拡散モデル）を便宜上「キャラクタモデル」と称する。

【０１０２】
このキャラクタモデルの訓練構造は、拡散モデル（Diffusionモデル）と当該拡散モデルを制御するネットワーク（ControlNet）を含む。
キャラクタモデルは、ControlNetの重みは凍結して、Diffusionモデルのパラメータのみを更新して（学習して）得られるものである。
「（プロセッサ１２２は）キャラクタモデルを都度作成する」と表現する場合がある。
【０１０３】
ここで「拡散モデル（Diffusion Model）」の一般的な内容について説明する。
拡散モデルは、本実施形態でも用いているStable Diffusionに関する技術の中核となる機械学習モデルである。
拡散モデルは、ノイズ画像などからノイズを取り除いてノイズなし画像などを生成可能なモデルである。拡散モデルは学習段階において、元情報（画素、潜在空間など）にノイズを加えていく拡散過程と、ノイズあり情報からノイズを除去していく逆拡散過程を含む。
【０１０４】
拡散モデルは、画像などの元情報、加えていくノイズ、元情報からノイズを加えていくことで得るノイズあり情報（例えばノイズ画像など）、ノイズあり情報から除去すべきノイズ、および、ノイズあり情報からノイズを除去していくことで得る元情報、などを学習データとする。
【０１０５】
そして拡散モデル（に含まれるUnet）は、上記ノイズあり情報などを入力として、ノイズあり情報から取り除くべきノイズを学習する（学習段階）。
そして拡散モデル（に含まれるUnet）は、ノイズあり情報（ノイズ画像など）を入力として、当該ノイズあり情報から除去すべきノイズを出力する。この工程を繰り返す結果として、ノイズあり情報（ノイズ画像など）からノイズが除去されている情報（ノイズなし画像など）を出力することができる（推論段階）。
【０１０６】
本実施形態では、特徴量分離を行うことで得ているコンセプトの特徴量により、機械学習モデル（拡散モデル）のパラメータを更新する。
【０１０７】
なお、ポーズをVAEにより潜在変数に変換したものはレイアウトの維持などに使用している（ControlNet・図１中の「Freezed」の部分）。
【０１０８】
例えば本実施形態において、特徴量分離を行うことで得ているコンセプトの特徴量が上記元情報となる。
つまりプロセッサ１２２は、コンセプトの特徴量（潜在変数）にノイズを加えていく拡散工程およびノイズを除去する逆拡散工程を経る学習により機械学習モデル（拡散モデル）を作成し取得する。
【０１０９】
なおデフォルメキャラクタ画像を得られるのであれば、元キャラクタ画像から複数の画像を取得しない態様（態様１）や、複数の画像を取得した上でポーズの特徴量とそれ以外の特徴量とにデカップリング（Decoupling）をしない方法（態様２）であってもよいが、本実施態様の方法によると元キャラクタ画像の特徴を精度高く反映しているデフォルメキャラクタ画像を得られるという顕著な効果がある。
【０１１０】
小括すると、プロセッサ１２２は一枚の元キャラクタ画像から複数の特徴量を取得してポーズの特徴量とそれ以外の特徴量（コンセプト）に分離（デカップリング）する。またこのコンセプトを説明する文字情報（キャラクタプロンプト）を取得する。さらにコンセプトの特徴量により学習する機械学習モデル（拡散モデル（キャラクタモデル））を作成する。

【０１１１】
本実施形態のキャラクタモデルは、１枚の人物（キャラクタ）画像のみを使用して訓練される。この訓練を完了したモデルは、異なるポーズやスタイル（例えば、SD（スーパーデフォルメ）スタイルなど）を持つこの人物（キャラクタ）の特徴を生成するために使用することができる。
【０１１２】
具体的な訓練方法について以下にまとめる。プロセッサは、顔が含まれる画像であって異なるスケールでクロップ（crop, cropping）した画像を使用する。プロセッサはこの異なるスケールの画像からプロンプト（タグ情報）を抽出する。機械学習モデルの訓練構造はDiffusionモデルとControlNetを含み、ControlNetの重みは凍結して、プロセッサはDiffusionモデルのパラメータのみを更新する。パラメータ更新後のDiffusionモデルが、本実施形態のキャラクタモデルである。
【０１１３】
なお、既存の機械学習モデル（拡散モデル）を基礎として、組み込んで、または改良して使用してもよい。
例えば、（デフォルメキャラクタ画像を含む）キャラクタ画像にノイズを加えていくことなどによりすでに学習している機械学習モデル（拡散モデル）などを使用してもよい。
【０１１４】
つぎに、（２）参照画像分析処理について説明する。
まずプロセッサ１２２は、ユーザの入力による頭身比率の設定を取得する（頭身比率取得処理）。
本実施形態においてプロセッサ１２２は、デフォルメキャラクタの参照画像を取得すること（ステップ６・参照画像取得処理）により頭身比率の設定を取得する。
別の例については後述する第二の実施形態で説明する。
【０１１５】
本実施形態においてプロセッサ１２２は、まず前処理として参照画像のコンセプト情報を削除するとともに、デフォルメキャラクタの形状の一部を示す線画情報（Scribbleと呼ばれる線画情報）と、ポーズ情報（Pose）と、を取得する。
【０１１６】
図１左下に示すように、プロセッサ１２２は線画情報として少なくとも参照画像の頭部、手先部分、および足先部分の線画を取得する。
これにより出力するデフォルメキャラクタのポーズ（向きなど）や頭身比率などを制御することができる（後述）。
【０１１７】
また、プロセッサ１２２はポーズ情報として参照画像の関節位置情報を取得する。
【０１１８】
図１左下において、白黒反転している画像の上段が線画情報（Scribble）、下段がポーズ情報（Pose）のイメージ図である。
【０１１９】
本実施形態においてプロセッサ１２２は、取得している線画情報（Scribble）およびポーズ情報（Pose）から、制御用（機械学習）モデルを作成する。
この制御用（機械学習）モデルを便宜上「コントロールモデル」と称する。
【０１２０】
なお、上述した機械学習モデル（キャラクタモデル）と区別するために制御用モデルと表記している。
よって例えば、キャラクタモデルを第一の機械学習モデル、コントロールモデルを第二の機械学習モデルなどと称してもよい。
【０１２１】
このコントロールモデルは、拡散モデルを制御する追加的なニューラルネットワークである。

特に、入力画像のキャラクタなどを、意図したポーズなどに制御できるという優れた効果を有する。

本実施形態のコントロールモデルは、デフォルメキャラクタ画像のポーズおよび／または（頭身数に影響する）外形などを制御する。

【０１２２】

上述した処理は、Stable Diffusionのプラグインや拡張機能で実行することができる。例えばStable Diffusion Web UI・Automatic1111版のWeb UIに追加できる「ControlNet Extension」などを好適に用いることができる。

【０１２３】

本実施形態において、線画情報の取得には（ControlNet）ControlNet Scribble（ControlNetのモデルの一つ）を用いている。

また、ポーズ情報の取得には（ControlNet）Open Pose（ControlNetのモデルの一つ）を用いている。

【０１２４】

なお、Open Poseは姿勢推定手法である。キャラクタの画像にOpen Poseを適用することで、そのキャラクタのポーズを推定する。Open PoseはParts Affinity Fieldsと呼ばれる骨格間の位置関係を考慮した処理を導入している。

【０１２５】

ControlNetの各種「preprocessor」が入力画像を解析・変換（例：ポーズ検出、エッジ検出、線画化など）し、それをモデルの条件として使用する。つまりControlNetの機能として、画像を線画化（edge detection／scribble detection）する仕組みが用意されている場合もあり、ユーザはそれを活用して「Scribble（線画）」を生成することができる。

【０１２６】

小括すると本実施形態においてプロセッサ１２２は、頭身比率を参照画像と言う形で取得し、参照画像からコントロールモデルを取得する。参照画像からデフォルメキャラクタの形状の一部を示す線画情報を少なくとも取得して作成しているコントロールモデルを取得する（ステップ７・コントロールモデル取得処理）。

【０１２７】

より具体的には、参照画像におけるデフォルメキャラクタの（形状の）一部を示す線画情報およびポーズ情報を（より正確にはこれらをVAEにより潜在変数化したものを）学習データとして学習している、（上述した拡散モデルを制御するための追加的なニューラルネットワークである）制御用（機械学習）モデルを作成し取得する。

つまり、この線画情報およびポーズ情報は制御用モデルの入力（条件（Condition））になる。

【０１２８】

なお本実施形態において、このコントロールモデルは都度作成しているが、参照画像ごとにあらかじめ作成してサーバ１０の記憶部１４に記憶しているコントロールモデルを取得するものであってもよい。この場合、処理が早くなるという利点がある。

ただし、ユーザが新たな画像を参照画像としてアップロードしているときは、その新たな画像についてプロセッサ１２２はコントロールモデルを作成する。

【０１２９】

図６に戻り、最後に（３）デフォルメキャラクタ取得処理について説明する。

上述したように、プロセッサ１２２は、前記元キャラクタ画像、前記文字情報および前記頭身比率を機械学習モデルへの入力として、前記元キャラクタ画像を前記頭身比率にデフォルメ化しているデフォルメキャラクタ画像を取得する（ステップ８・デフォルメキャラクタ取得処理）。

【０１３０】

より具体的には、上述の制御用モデル（＝コントロールモデル）により制御されている拡散モデル（＝キャラクタモデル）に対して文字情報（＝キャラクタプロンプト）を入力することで、元キャラクタ画像をデフォルメ化しているデフォルメキャラクタ画像を生成

して出力として取得する。
【０１３１】
（不慣れなユーザなどが）プロンプトにより条件入力することなどに比べ、格段に精度が高いデフォルメキャラクタ画像が得られるという顕著な効果がある。
【０１３２】
そしてプロセッサ１２２は、取得しているデフォルメキャラクタ画像を端末２０の表示部２８４ａに表示させる。
ユーザにより複数の頭身比率が選択されているときは、それぞれの頭身比率に対応するデフォルメキャラクタ画像を出力する。
またＵＩの項で示したように、プロセッサ１２２は、ユーザによる入力を受け付けてデフォルメキャラクタ画像を再生成することなどができる。
【０１３３】
なお、本実施形態におけるデフォルメキャラクタ生成処理の効果を発揮できるものであれば、処理の順序は図６のものに限られない。例えば図６において、ステップ４、ステップ５、およびステップ６～７を逐次的に描画しているが、これらは並列的に処理してもよい。
または、ステップ４、ステップ２・３・５、およびステップ６～７を並列的に処理してもよい。
【０１３４】
＜２－２．最適化処理＞
最適化処理において、プロセッサ１２２は、上述した拡散モデル（キャラクタモデル）が備えるニューラルネットワーク（Unet）の少なくとも一部の層において、重みを変更して得る複数の出力画像の差分を取得して評価し、出力画像に影響の小さい層の重みの削減または削除を行う。
【０１３５】
プロセッサ１２２は、最適化プログラムＰ１４に基づき、最適化処理を行う。
すなわち、最適化プログラムＰ１４は、プロセッサ１２２による最適化処理の実行により、コンピュータを最適化手段として機能させる。
【０１３６】
図８は、最適化処理を示すフローチャートである。
プロセッサ１２２は、ユーザによるデフォルメ画像再生成の指示を受け付けることにより、またはあらかじめ設置している条件に基づき自動的に、最適化処理を開始する（図５・生成画像表示画面参照）。
本実施形態において、ユーザは画像の少なくとも一部を変化させている画像を再生成することができる。
【０１３７】
まずプロセッサ１２２は、（拡散モデルの）ニューラルネットワークにおける複数の層のそれぞれにおいて重みを１および０に設定した場合における、デフォルメキャラクタ画像の差分を取得する（ステップ２１）。
プロセッサ１２２は、影響が大きい層の重みを保持する一方で、影響が大きい層の重みを削減または削除（以下「削減等」とする。）する（ステップ２２）。
【０１３８】
図９は、重みの削減等を説明するための図である。
図９において、右上の角丸四角で囲まれている部分はニューラルネットワークの同じ層であることを示す。また、同じ角丸四角で囲まれている２体のキャラクタはそれぞれ、当該層において重みを０としている場合と１としている場合におけるデフォルメキャラクタ画像である。
図９中のＤｉｆｆは、重みが０のときのデフォルメキャラクタ画像と１のときのデフォルメキャラクタ画像との差分を示す。白抜きの部分が差分である。
【０１３９】

ある層において重みを１および０に設定して得たそれぞれの画像の差分があまり大きくない場合、その層は全体の結果に対して大きな影響を与えていないことを意味する。
つまり各層でこの差分を評価することで、当該各層の全体に対する影響度を把握することができる。そしてプロセッサ１２２は、この結果を受けて影響の少ない層の重みを削除等する。

【０１４０】

例えば、図９の（４つの角丸四角で表す層の）一番右の層では、重みの違い（重み１および０）により生成されるそれぞれの画像の差が大きい（差分が大きい）。このことから、この層の重みの変化により生成される画像の変化が大きいこと、つまりこの層の重みが画像生成に大きく影響することがわかる。
一方で、左から２番目の層は差分が小さいことから、この層が生成画像に与える影響は小さいと考えられる。

【０１４１】

この設定は学習済みモデルのファインチューニング方法（LoRA）におけるパラメータ層を最適化し、最終的に高品質な結果を生成するという効果を奏する。

【０１４２】

図８のフローチャートに戻る。
この影響が大きい層、つまり鍵となる特徴に関する層を（Unetの）ニューラルネットワークの出力層に配置し、また、誤差逆伝播によりニューラルネットワークの層の重みを更新する（ファインチューニング）（ステップ２３）。
ファインチューニングにより、元キャラクタ画像の特徴を適切に再現しているデフォルメキャラクタ画像が得られる。

【０１４３】

最適化処理により得られた最適な画像の例を示す。
図１０は、最適化処理前後のデフォルメキャラクタ画像を示す図である。
図１０に示すように、最適化後の画像は最適化前の画像より元キャラクタ画像を忠実に再現している（特に腰のあたりの四角で囲っている部分）。

【０１４４】

ここで、本実施形態で使用している具体的な方法について説明する。
上記の処理を行う方法として例えばLoRA Block Weightを用いることができる。LoRa Block Weightにより、LoRAの適用度を調整することができる。

【０１４５】

なおLoRA（Low-Rank Adaptation）は、学習済みモデルに対してファインチューニングを行う方法である。
そのなかで、コンセプトを対象にチューニングを行うもの、生成したい画像のコンセプトを反映させるもの、またはキャラクタに特定のコンセプトを適用させるものをConcept LoRAと称する。Concept LoRAを用いることで、上記のような調整を行うことができる。

【０１４６】

以上まとめると、元キャラクタ画像分析処理により、１枚の元キャラクタ画像の複数の領域から複数の特徴量を取得し、デフォルメキャラクタに元キャラクタ画像の特徴を多く反映できるようにしている。ここではキャラクタプロンプトやキャラクタモデルを作成（取得）する。
また、参照画像分析処理により、プロセッサ１２２は参照画像という形で頭身比率を取得し、参照画像から線図やポーズを抽出する処理などを経て、キャラクタモデルを制御するためのコントロールモデルを取得する。
そしてデフォルメキャラクタ取得処理において、コントロールモデルにコントロール（制御）されているキャラクタモデルに対してキャラクタプロンプトを入力することで、元キャラクタ画像に忠実なデフォルメキャラクタを生成することができる。その精度の高さは図面で示すとおりである。
そして最適化処理により、さらに高品質なデフォルメ画像を得ることができる。

【０１４７】
また、上述する拡散モデル（キャラクタモデル）は、キャラクタの画像である元キャラクタ画像から当該元キャラクタ画像をデフォルメ化しているデフォルメキャラクタ画像を取得する学習済みモデルであって、
前記元キャラクタ画像から複数の特徴量を取得し、当該複数の特徴量のうちポーズの特徴量以外の特徴量を含むコンセプトの特徴量、前記コンセプトの特徴量に加えるノイズ、およびノイズを加えることで得るノイズあり情報、を少なくとも含むデータを学習データとして、ノイズあり情報から取り除くべきノイズを学習し、
ノイズあり情報を入力として、当該ノイズあり情報から除去すべきノイズを出力することを特徴とし、
さらに、参照画像におけるデフォルメキャラクタの（形状の）一部を示す線画情報を少なくとも１つの学習データとして学習している制御用モデル（＝コントロールモデル）により、出力するデフォルメキャラクタ画像におけるポーズが制御されることを特徴とする、学習済みモデルである。
【０１４８】
図１１は、上述したデフォルメキャラクタ生成処理を（図１とは別の形で）示す図である。
キャラクタの画像（元キャラクタ画像）からデフォルメキャラクタ画像を得るほか、デフォルメキャラクタ画像を最適化する処理を示す。
【０１４９】
３．ハードウェア構成
図１２はデフォルメキャラクタ生成システム１の概要を示すネットワーク構成図である。
図１２に示すように、本実施形態におけるデフォルメキャラクタ生成システム１は、システムサーバ１０（サーバ１０）、および（ユーザ）端末２０を備える。また、これらの各装置は、ネットワークＮを介して接続されている。ネットワークＮは例えばインターネットなどである。
サーバ１０には、デフォルメキャラクタ生成プログラムＰ１を含み、本実施形態に係るデフォルメキャラクタ生成システム１を動作させるためのソフトウェア（アプリケーションソフトウェア）がインストールされており、当該ソフトウェアの機能により、各種処理が実行される。
以下、各ハードウェアについて説明する。
【０１５０】
＜サーバ１０＞
サーバ１０は、デフォルメキャラクタ生成プログラムＰ１を実行するためのコンピュータである。またサーバ１０は、ユーザの用に供するユーザーインターフェースを提供するウェブサイトWを表示する。
【０１５１】
図１２においてサーバ１０は１台のみ図示しているが、数は１台に限られるものではなく、複数のサーバにより実現してもよい。例えばウェブサーバと機械学習用サーバを別にすることなどが挙げられる。
複数のサーバは負荷分散や可用性の観点から用いることも考えられる。
また、画像生成に関する処理を、他のサーバに任せるものまたは他のサーバと協力して行うものであってもよい。
このほか、サーバ１０はクラウドサービス事業者のコンピュータを利用するものであってもよいし、ユーザがコンピュータを用意してもよい。
【０１５２】
図１３は、サーバ１０のハードウェア構成図である。
図１３に示すように、サーバ１０は、制御部１２、記憶部１４、および通信制御部１６を備える。また制御部１２は、プロセッサ１２２、ＲＯＭ１２４、ＲＡＭ１２６、計時部

１２８を備える。それぞれの基本的な機能については後でまとめて説明する。
【０１５３】
プロセッサ１２２を備える制御部１２は、サーバ１０においてデフォルメキャラクタ生成部としても機能する（不図示）。デフォルメキャラクタ生成部は、デフォルメキャラクタ生成プログラムＰ１を実行してデフォルメキャラクタ生成処理を行う。本実施形態において、プロセッサ１２２はＣＰＵ（Central Processing Unit）である。
【０１５４】
また、一のプログラムは、別のプログラムを含んでいてもよい。例えば本実施形態において、デフォルメキャラクタ生成プログラムＰ１は、元キャラクタ画像分析プログラムＰ１１や参照画像分析プログラムＰ１２などを含む。
【０１５５】
図１３に示すように、記憶部１４は、プログラム格納部１４ａとデータ格納部１４ｂを備え、各種処理に必要なプログラムやデータを備える。
例えば、プログラム格納部１４ａには、本実施形態に係るデフォルメキャラクタ生成プログラムＰ１のほか、サーバ１０に接続されている機器を制御するための制御プログラム、例えば通信制御部１６を制御する通信制御プログラムなどが格納されている。
【０１５６】
図１３に示すように、通信制御部１６は、サーバ１０と、外部にある端末、例えば、後述する端末２０などとの間で通信を行う装置である。通信制御部１６は図１２に示すように、サーバ１０をネットワークＮに接続する。
【０１５７】
上記のほか、サーバ１０は、命令やデータの入力を行うための入力部（例えばキーボード）や、情報を何らかの形で出力するための出力部（例えば音声出力装置）などを備えていてもよい（不図示）。また、本実施形態の用途のために追加的に必要な装置や、本実施形態の用途について利便性を向上させるための装置を備えていてもよい。
【０１５８】
＜端末２０＞
端末２０は、ユーザがデフォルメキャラクタ生成システム１を利用するための情報処理装置である。ユーザは、端末２０を用いてサーバ１０にアクセスすることにより、デフォルメキャラクタ生成システム１を利用する。
端末２０は、制御部２２、記憶部２４、通信制御部２６、および入出力部２８を備える。また制御部は、プロセッサ、ＲＯＭ、ＲＡＭ、および計時部を備える。
入出力部２８は入力部２８２および出力部２８４を備え、出力部２８４は表示部２８４ａを備える。
説明済みの内容と重複するものや、後述する基本的な機能に関するものは説明を省略する。
【０１５９】
本実施形態において、端末２０はデスクトップＰＣである。ただし、端末２０はこれに限られるものではなく、スマートフォンやタブレットなどの携帯型端末であってもよい。
【０１６０】
なお本実施形態において、端末２０は特定のアプリケーションプログラムをインストールする必要はなく、ウェブサイトWにアクセスすることで各種処理が実行可能であり、デフォルメキャラクタ生成システム１を利用することができる。
【０１６１】
（コンピュータの基本的機能に係る説明）
以下、制御部（プロセッサ、ＲＯＭ、ＲＡＭ、計時部）、記憶部、通信制御部、入力部、および出力部について説明する。
なお、本実施形態のいずれの端末においても、機能部間の接続態様（ネットワークトポロジ）は特に限定されない。例えばバス型であってもよいし、スター型、メッシュ型などであってもよい。

【０１６２】
プロセッサは、ＲＯＭや記憶部などに記憶されたプログラムに従って、情報処理や各種装置の制御を行う。本実施形態において、プロセッサはＣＰＵ（Central Processing Unit）である。
【０１６３】
ただし、プロセッサはＣＰＵに限られない。ＣＰＵ、ＤＳＰ（Digital Signal Unit）、ＧＰＵ（Graphics Processing Unit）、ＧＰＧＰＵ（General Purpose computing on GPU）、ＴＰＵ（Tensor Processing Unit）、またはＡＳＩＣ（Application Specific Integrated Circuit）など、各種プロセッサを単独で、あるいは組み合わせて用いてもよい。
例えば、ＣＰＵとＧＰＵを統合したプロセッサはＡＰＵ（Accelerated Processing Unit）などと呼ばれるが、このようなプロセッサを用いてもよい。
【０１６４】
ＲＯＭは、プロセッサが各種制御や演算を行うための各種プログラムやデータがあらかじめ格納された、リードオンリーメモリである。
【０１６５】
ＲＡＭは、プロセッサにワーキングメモリとして使用されるランダムアクセスメモリである。このＲＡＭには、本実施形態の各種処理を行うための各種エリアが確保可能になっている。
【０１６６】
計時部は、時間情報の取得などに係る計時処理を行う。コンピュータが通信制御部を備える場合は、ＮＴＰ（ネットワーク・タイム・プロトコル）により外部から時間情報を取得してもよい。
【０１６７】
記憶部は、プログラムやデータなどの情報を記憶するための装置である。記憶部はストレージとも称する。記憶部は内蔵型か、外付型かを問わない。
【０１６８】
記憶部は、データの読み書きが可能な記憶媒体と、当該記憶媒体に読み書きするドライブとを含む。
記憶媒体は例えば、内蔵型や外付型があり、ＨＤ（ハードディスク）、ＣＤ－ＲＯＭ、フラッシュメモリなどが挙げられる。
ドライブは例えば、ＨＤＤ（ハードディスクドライブ）、ＳＳＤ（ソリッドステートドライブ）などが挙げられる。
【０１６９】
記憶部は、機能部としてプログラム格納部とデータ格納部を備える。
プログラム格納部には、各種機器を制御するための制御プログラム、例えば通信を制御する通信制御プログラムなどが格納されている。
【０１７０】
通信制御部は、端末などの間で通信を行うための装置である。通信制御部は、当該通信制御部を備える端末をネットワークＮに接続する。
【０１７１】
通信制御部の通信方式は公知の方式であり、機器に応じて有線による方式や無線による方式が適用される。
例えば、端末がデスクトップＰＣであれば有線、無線の両方の場合が考えられ、また、端末がスマートフォンであれば、無線による通信方式が考えられる。
【０１７２】
有線であれば、例えばＩＥＥＥ８０２．３（例えばバス型やスター型の有線ＬＡＮ）で規定される通信方式を好適に用いることができるが、それ以外にも、ＩＥＥＥ８０２．５（例えばリング型の有線ＬＡＮ）で規定される通信方式などを用いてもよい。
【０１７３】

無線であれば、例えばＩＥＥＥ８０２．１１（例えばＷｉ－Ｆｉ）で規定される通信方式を好適に用いることができるが、それ以外にも、ＩＥＥＥ８０２．１５（例えばブルートゥース（登録商標）、ＢＬＥ（ブルートゥース（登録商標）ローエナジー）など）、ＩＥＥＥ８０２．１６（例えばＷｉＭＡＸ）、ＺｉｇＢｅｅ（登録商標）、９２０ＭＨｚ帯無線（Ｗｉ－ＳＵＮなど）、または赤外線通信などの光通信で規定される通信方式などを用いてもよい。

【０１７４】

入力部および出力部は、それぞれ端末に対する入力と出力を担う装置である。入力部および出力部をあわせて入出力部と称する場合がある。

入力部はユーザからの入力を受け付ける装置である。このような入力部として例えば、キーボード、ポインティングデバイスとしてのマウス、トラックパッド、タブレット、またはタッチパネルなどが挙げられる。

【０１７５】

端末がタブレットやスマートフォンなどであって、入力部がタッチパネルの場合、入力部はタッチスクリーンなど、画像などを表示する表示部の表面に配置される。この場合、入力部は、表示部に表示される各種操作アイコンに対応したユーザのタッチ位置を特定し、ユーザによる入力を受け付ける。

【０１７６】

出力部は例えば、画像や音声、帳票などを出力するための装置である。

出力部として例えば、タッチスクリーンやディスプレイ（液晶ディスプレイや有機ＥＬディスプレイ）などの表示装置（表示部）や、スピーカなどの音声出力装置、プリンタなどの帳票出力装置が挙げられる。

【０１７７】

以上のような構成により、ユーザはオンラインで用意に所望のデフォルメキャラクタを取得することができる。

【０１７８】

（第二の実施形態）

上述の実施形態においてプロセッサ１２２は、デフォルメキャラクタの参照画像を取得すること（ステップ６・参照画像取得処理）により頭身比率の設定を取得していた。

一方、本実施形態においてプロセッサ１２２は、デフォルメキャラクタ設定欄（ＵＩ－１２）への入力を受け付けること、つまりユーザの入力による頭身比率の設定を取得（頭身比率取得処理）したあと、

当該頭身比率を、機械学習モデル（拡散モデル）に対するプロンプト（文字情報・キャラクタプロンプト）に追加する（「プロンプト追加処理」とする。）ことで、出力されるデフォルメキャラクタ画像の頭身比率を制御する。

【０１７９】

つまり、拡散モデルを制御するコントロールモデルの形で頭身比率を制御するのではなく、拡散モデルに対する入力（テキストエンコーダによるテキスト入力）として頭身比率を制御する。

【０１８０】

第一の実施形態では参照画像分析処理があり、参照画像分析処理がユーザの入力による頭身比率の設定を取得する頭身比率取得処理を含んでいた。しかし本実施形態では、（頭身比率取得処理以外の）参照画像分析処理による処理の別手段として、上述の頭身比率取得処理とプロンプト追加処理がある。

【０１８１】

また、本実施形態において、ユーザにより参照画像がアップロードされた場合、プロセッサ１２２はその参照画像の頭部と全身とを認識し、それぞれの大きさ（長さ）を計測することで頭身比率を算出する（「画像頭身比率分析処理」とする。）。

【０１８２】

なお、プロセッサ１２２が頭身比率取得処理、プロンプト追加処理、画像頭身比率分析

処理を実行するためのプログラムをそれぞれ頭身比率取得プログラム、プロンプト追加プログラム、画像頭身比率分析プログラムとする。

【０１８３】

本実施形態の処理により、元キャラクタ画像の特徴を精密に再現するデフォルメキャラクタ画像を取得することができる。

なお第一の実施形態の態様は、本実施形態の態様よりも元キャラクタ画像の特徴をより精密に再現するなどの利点が確認された。

【０１８４】

（変形例）

本発明は上述の実施形態に限られるものではなく、本発明の趣旨を逸脱しない範囲において、上述の実施形態に種々の変更を加えたものを含む。

【０１８５】

例えば上述の実施形態では、パラメータの更新により機械学習モデル（拡散モデル）を都度作成していたが、機械学習モデルの学習や作成はこれに限るものではない。例えば、パラメータの更新に限らず、モデル構造を変更するようなものや、モデル構造から作成するようなものであってもよい。

ただし、パラメータの更新は処理が早くより好ましい。

【０１８６】

また上述したように、すでに作成されている機械学習モデル（拡散モデル）を取得するようにしてもよい。

また、作成したキャラクタモデルを、（既存の・他の）Stable Diffusionの拡散モデルに統合する（組み込む）ようにしてもよい。

ただし、元キャラクタ画像を取得して分析し、拡散モデルのパラメータを更新して新たな拡散モデルとすることで、デフォルメキャラクタ画像における元キャラクタ画像の再現度は著しく向上する。

【０１８７】

上述の実施形態において、機械学習モデル（キャラクタモデル）は拡散モデルとコントロールモデルを含むものであったが、これに限られない。まず、拡散モデルのみで構成されていてもよい。また、本発明の目的を達するものであれば、拡散モデル以外の機械学習モデルで構成してもよい。

ただし、拡散モデルとコントロールモデルを含む態様は、元キャラクタ画像の特徴量を精度高く再現しているデフォルメキャラクタ画像を生成することができる。

【０１８８】

特徴量分離（Feature Decoupling）を実現するために、以下のアプローチが用いられ得る。上記の実施形態では次の（１）を用いている。

（１）潜在空間の分離

生成モデルの潜在空間（Latent Space）を設計し、異なる次元が異なる特徴量を制御するように学習させる。例えば、各次元が「髪型」、「顔（表情）」、または「ポーズ」などをそれぞれ担当する。

（２）アーキテクチャ設計

生成ネットワーク（GANやDiffusion Modelなど）の層やモジュールを設計し、特定の層が特定の特徴量を処理するようにする。例えばスタイル変換のための「マルチスケール特徴量抽出」や「レイヤーレベルでのスタイル適用」などである。

（３）学習の正則化（Regularization）

特徴量間の独立性を保つために、学習プロセス中に特徴量間の相互作用を抑制するペナルティを導入する。

（４）モジュール間の条件付け

特定の条件（テキスト、属性、スケッチ）に基づいて生成プロセスを調整することで、特定の特徴量を強調する。

【０１８９】

本実施形態を含む本発明の態様は、換言すると以下の特徴量を備える。下記は本願出願時における特許請求の範囲と対応する。ただし、出願後における特許請求の範囲の補正により、当該補正後の特許請求の範囲の記載とは異なる場合がある。

（１）第１の態様では、コンピュータを、キャラクタの画像である元キャラクタ画像を取得する元キャラクタ画像取得手段、前記元キャラクタ画像について説明する文字情報を取得するキャラクタプロンプト取得手段、ユーザの入力による頭身比率の設定を取得する頭身比率取得手段、および、前記元キャラクタ画像、前記文字情報および前記頭身比率を機械学習モデルへの入力として、前記元キャラクタ画像を前記頭身比率にデフォルメ化しているデフォルメキャラクタ画像を取得するデフォルメキャラクタ取得手段、として機能させ、前記機械学習モデルは、少なくともキャラクタ画像およびデフォルメ化しているキャラクタ画像を学習データとして学習し、キャラクタ画像を入力として、当該キャラクタ画像をデフォルメ化しているデフォルメキャラクタ画像を出力することを特徴とする、デフォルメキャラクタ生成プログラムを提供する。

（２）第２の態様では、さらに、前記機械学習モデルが備えるニューラルネットワークの少なくとも一部の層において、重みを変更して得る複数の出力画像の差分を取得して評価し、出力画像に影響の小さい層の重みの削減または削除を行う最適化手段を備えることを特徴とする、第１の態様に記載のデフォルメキャラクタ生成プログラムを提供する。

この場合、元キャラクタ画像の再現度がさらに高いデフォルメキャラクタ画像を得ることができる。

（３）第３の態様では、キャラクタの画像である元キャラクタ画像を取得する元キャラクタ画像取得部と、前記元キャラクタ画像について説明する文字情報を取得するキャラクタプロンプト取得部と、ユーザの入力による頭身比率の設定を取得する頭身比率取得部と、前記元キャラクタ画像、前記文字情報および前記頭身比率を機械学習モデルへの入力として、前記元キャラクタ画像を前記頭身比率にデフォルメ化しているデフォルメキャラクタ画像を取得するデフォルメキャラクタ取得部と、を備え、前記機械学習モデルは、少なくともキャラクタ画像およびデフォルメ化しているキャラクタ画像を学習データとして学習し、キャラクタ画像を入力として、当該キャラクタ画像をデフォルメ化しているデフォルメキャラクタ画像を出力することを特徴とする、デフォルメキャラクタ生成システムを提供する。

（４）第４の態様では、プロセッサがキャラクタの画像である元キャラクタ画像を取得する元キャラクタ画像取得ステップ、プロセッサが前記元キャラクタ画像について説明する文字情報を取得するキャラクタプロンプト取得ステップ、プロセッサがユーザの入力による頭身比率の設定を取得する頭身比率取得ステップ、および、プロセッサが、前記元キャラクタ画像、前記文字情報および前記頭身比率を機械学習モデルへの入力として、前記元キャラクタ画像を前記頭身比率にデフォルメ化しているデフォルメキャラクタ画像を取得するデフォルメキャラクタ取得ステップ、を備え、前記機械学習モデルは、少なくともキャラクタ画像およびデフォルメ化しているキャラクタ画像を学習データとして学習し、キャラクタ画像を入力として、当該キャラクタ画像をデフォルメ化しているデフォルメキャラクタ画像を出力することを特徴とする、デフォルメキャラクタ生成方法を提供する。

（５）第５の態様では、少なくともキャラクタ画像、前記キャラクタ画像について説明する文字情報、前記キャラクタ画像をデフォルメ化しているデフォルメキャラクタ画像、および前記キャラクタ画像および前記デフォルメキャラクタ画像の頭身比率を学習データとして学習する機械学習モデルであって、キャラクタの画像である元キャラクタ画像、前記元キャラクタ画像について説明する文字情報、およびユーザによる頭身比率の設定を入力として、前記元キャラクタ画像を前記頭身比率にデフォルメ化しているデフォルメキャラクタ画像を出力するデフォルメキャラクタ生成用の学習済み機械学習モデルを提供する。

【産業上の利用可能性】

【０１９０】

可愛らしいキャラクタはオンライン・オフラインを問わず産業（特に娯楽産業）上の至

ろところでニーズがあることから、このようなキャラクタを生成可能なプログラムはニーズに即したキャラクタの提供を可能にする。

【符号の説明】

【０１９１】

１　デフォルメキャラクタ生成システム
　１０　サーバ
　１２　制御部
　１２２　プロセッサ
　１２４　ＲＯＭ
　１２６　ＲＡＭ
　１２８　計時部
　１４　記憶部
　１４ａ　プログラム格納部
　１４ｂ　データ格納部
　１６　通信制御部
　１８　入出力部
　２０　端末
　２２　制御部
　２４　記憶部
　２６　通信制御部
　２８　入出力部
　２８２　入力部
　２８４　出力部
　２８４ａ　表示部
　ＵＩ－１１　アップロード欄
　ＵＩ－１２　デフォルメキャラクタ設定欄
　ＵＩ－１２１　頭身比率設定ボタン
　ＵＩ－１２２　参照デフォルメキャラクタ画像
　ＵＩ－１２３　デフォルメキャラクタ生成開始ボタン
　ＵＩ－１３　元キャラクタ画像確認画面（ウィンドウ）
　ＵＩ－１３１　画像情報
　ＵＩ－１３２　元キャラクタ画像
　ＵＩ－１３３　ＯＫボタンアイコン
　ＵＩ－１４　出力画像表示欄
　ＵＩ－１４１　出力画像
　ＵＩ－１４２　出力画像加工関連ボタン
　ＵＩ－１４３　出力画像情報
　ＵＩ－１４４　出力画像保存関連ボタン
　Ｐ１　デフォルメキャラクタ生成プログラム
　Ｐ１１　元キャラクタ画像分析プログラム
　Ｐ１２　参照画像分析プログラム
　Ｐ１３　デフォルメキャラクタ取得プログラム
　Ｐ１４　最適化プログラム

【要約】

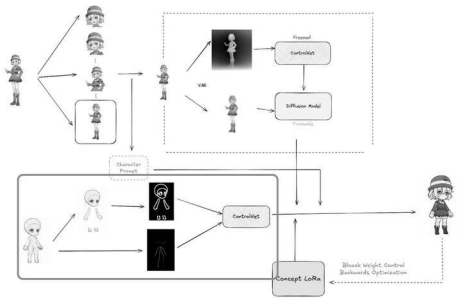
【課題】必要な特徴を残しながら、通常のキャラクタからデフォルメ化キャラクタを生成する。

【解決手段】デフォルメキャラクタ生成プログラムＰ１は、（１）元キャラクタ画像分析処理、（２）参照画像分析処理、および（３）デフォルメキャラクタ取得処理により、元キャラクタ画像からデフォルメ化キャラクタを生成する。ユーザの入力による頭身比率の設定を取得すること、元キャラクタ画像（例えば８頭身キャラクタ画像）を取得すること

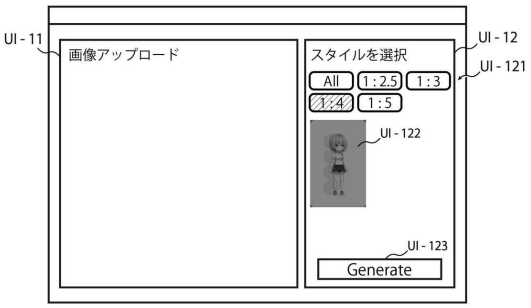
、そしてこれらの情報から、元キャラクタ画像を前記頭身比率にデフォルメ化しているデフォルメキャラクタ画像を取得することを最も主要な特徴とする。

【選択図】図１

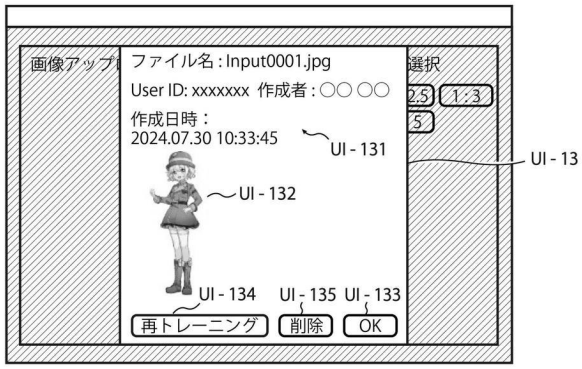
【図１】



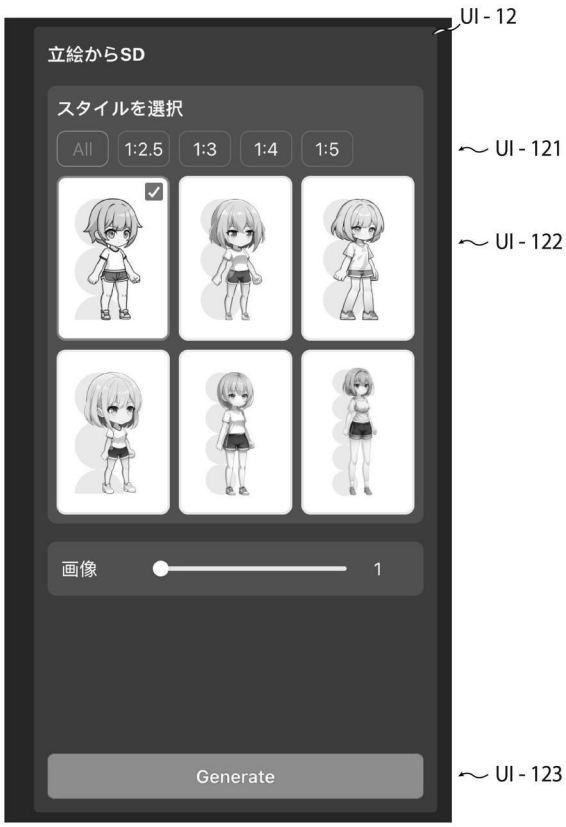
【図２】



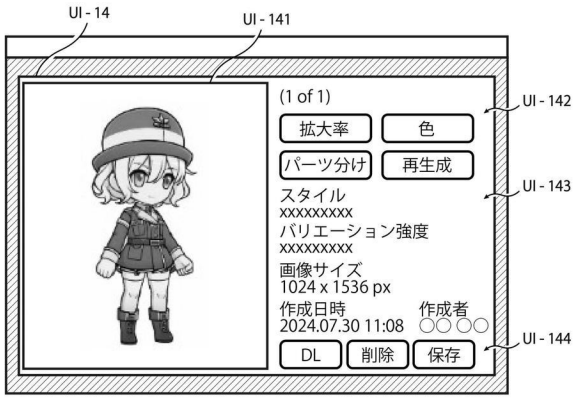
【図３】



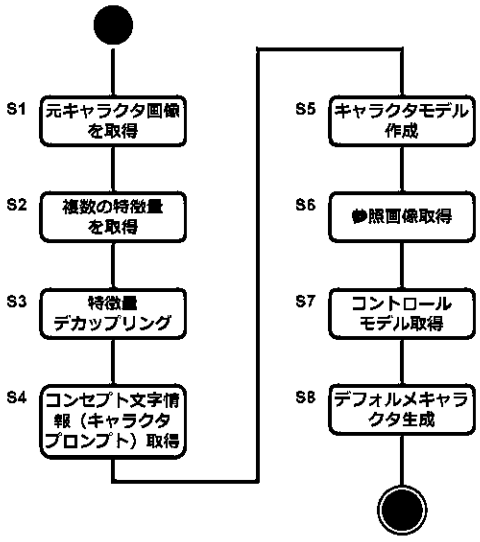
【図４】


UI - 12
立絵からSD
スタイルを選択
All
1:2.5
1:3
1:4
1:5
UI - 121
UI - 122
画像
1
Generate
UI - 123

【図５】

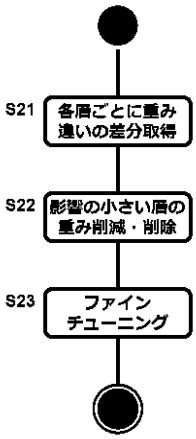

UI - 14
UI - 141
(1 of 1)
拡大率
色
パーツ分け
再生成
UI - 142
スタイル
xxxxxxxxx
バリエーション強度
xxxxxxxxx
画像サイズ
1024 x 1536 px
作成日時
2024.07.30 11:08
作成者
UI - 143
DL
削除
保存
UI - 144

【図６】

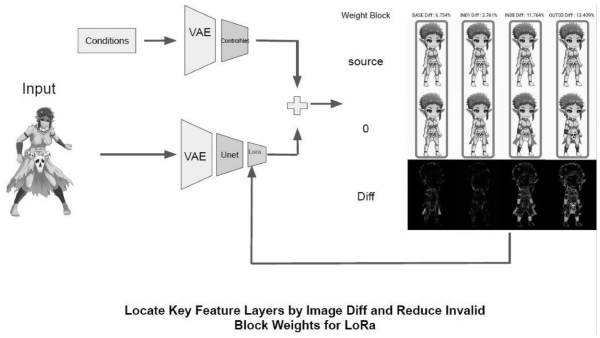

S1
元キャラクタ画像を取得
S2
複数の特徴量を取得
S3
特徴量デカップリング
S4
コンセプト文字情報（キャラクタプロンプト）取得
S5
キャラクタモデル作成
S6
参照画像取得
S7
コントロールモデル取得
S8
デフォルメキャラクタ生成

【図７】

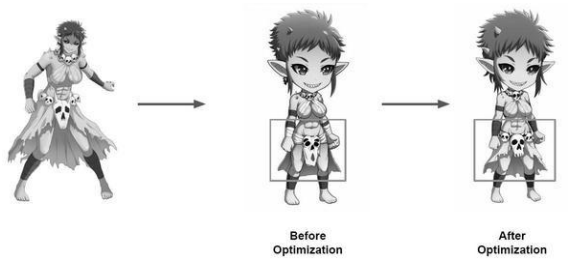
【図８】


S21
各層ごとに重み違いの差分取得
S22
影響の小さい層の重み削減・削除
S23
ファインチューニング

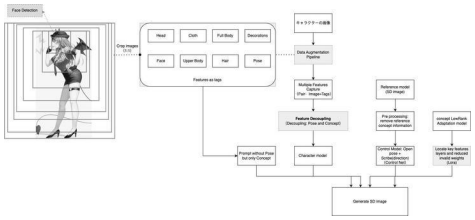
【図９】



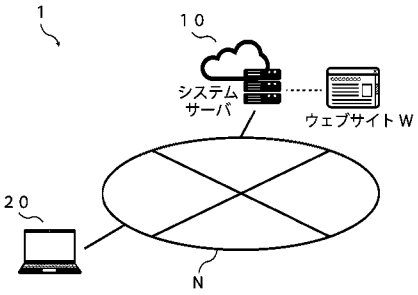
【図１０】



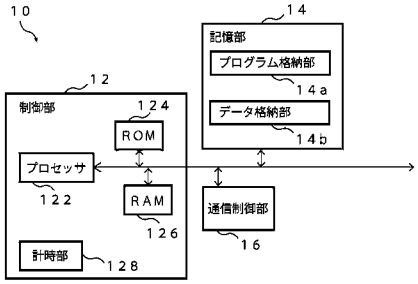
【図１１】



【図１２】



【図１３】



10

フロントページの続き

審査官　鈴木　肇

(56)参考文献　特開２０２３－１６９０４８（ＪＰ，Ａ）
特開２０２３－１５７３３４（ＪＰ，Ａ）
国際公開第２０２４／１８５０５４（WO，Ａ１）
特開２０２１－１０５９５０（ＪＰ，Ａ）
特開２０２１－０４７７１１（ＪＰ，Ａ）
国際公開第２０２４／１６１８３９（WO，Ａ１）
澁谷美樹，鈴木優，描いたイラストのデフォルメ具合を自動で調節するツール，インタラクション２０１８論文集 ［ｏｎｌｉｎｅ］，日本，情報処理学会，2018年，p994-997，URL:http://www.interaction-ipsj.org/proceedings/2018/data/pdf/3B33.pdf
Miao Hua, Jiawei Liu, Fei Ding, Wei Liu, Jie Wu, Qian He, DreamTuner: Single Image is Enough for Subject-Driven Generation, arXiv, [online], 2023年12月21日, [2025年3月25日検索], インターネット<URL: https://arxiv.org/pdf/2312.13691v1>, DOI: 10.48550/arXiv.2312.13691

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｔ　　１／００　－１９／２０