

(19)日本国特許庁(JP)

(12)特許公報(B1)

(11)特許番号
特許第7658024号
(P7658024)

(45)発行日 令和7年4月7日(2025.4.7)

(24)登録日 令和7年3月28日(2025.3.28)

(51)国際特許分類	F I	
G 0 6 T 1/40 (2006.01)	G 0 6 T 1/40	
G 0 6 T 7/00 (2017.01)	G 0 6 T 7/00	3 5 0 B
G 0 6 T 11/00 (2006.01)	G 0 6 T 11/00	
G 0 6 V 10/70 (2022.01)	G 0 6 V 10/70	

請求項の数 5 (全19頁)

(21)出願番号	特願2024-211898(P2024-211898)	(73)特許権者	394013002 三菱電機インフォメーションシステムズ株式会社 東京都港区三田三丁目5番19号
(22)出願日	令和6年12月5日(2024.12.5)	(74)代理人	110002491 弁理士法人クロスボーダー特許事務所
審査請求日	令和6年12月5日(2024.12.5)	(72)発明者	檜垣 宏策 東京都港区三田三丁目5番19号 三菱電機インフォメーションシステムズ株式会社内
早期審査対象出願		(72)発明者	安井 大希 東京都港区三田三丁目5番19号 三菱電機インフォメーションシステムズ株式会社内
		審査官	佐田 宏史

最終頁に続く

(54)【発明の名称】 学習データ生成装置、学習データ生成方法及び学習データ生成プログラム

(57)【特許請求の範囲】

【請求項1】

画像データを変換する画像変換モデルに対して、物体検知モデルに学習させる学習データである画像データを入力して、前記画像変換モデルによって前記画像データが変換された変換後画像を取得する画像変換部と、
前記画像変換部によって取得された前記変換後画像を前記学習データに加える学習データ追加部と
を備え、

前記画像変換モデルは、前記画像データの説明文を生成する文章生成モデルと、説明文を変換する文書変換モデルと、説明文から画像データを生成する画像生成モデルとを含み、前記画像変換部は、前記文章生成モデルに対して前記画像データと指定属性を示す属性情報とを入力するとともに、前記画像データに含まれる前記属性情報が示す前記指定属性に関する説明と、前記画像データの画角についての説明とを含む説明文の生成を指示して、前記文章生成モデルによって生成された前記指定属性に関する説明と前記画角についての説明とを含む前記画像データの説明文を取得し、前記文書変換モデルに対して取得された前記説明文を入力するとともに、前記指定属性及び前記画角については変更せず、他の部分を変更するように指示して、前記文書変換モデルによって変換された変換後文を取得し、前記画像生成モデルに対して変換された前記変換後文を入力して、前記画像生成モデルによって生成された画像データを前記変換後画像として取得する学習データ生成装置。

【請求項2】

前記学習データ生成装置は、さらに、
前記物体検知モデルに対して前記画像変換部によって取得された前記変換後画像を入力して、前記物体検知モデルで物体を検知した確信度が評価閾値以上であるか否かを判定するデータ判定部

を備え、

前記学習データ追加部は、前記データ判定部によって前記確信度が前記評価閾値未満であると判定された場合に、前記変換後画像を前記学習データに加える

請求項 1 に記載の学習データ生成装置。

【請求項 3】

前記データ判定部は、前記確信度が前記評価閾値未満である場合に、画像データに特定の属性の物体が存在するか否かを判定する属性判定モデルに対して、前記変換後画像と指定属性とを入力して、前記変換後画像に前記指定属性の物体が存在するか否かの判定結果を取得し、

10

前記学習データ追加部は、前記データ判定部によって前記変換後画像に前記指定属性の物体が存在するという判定結果が取得された場合に、前記変換後画像を前記学習データに加える

請求項 2 に記載の学習データ生成装置。

【請求項 4】

コンピュータが、画像データを変換する画像変換モデルに対して、物体検知モデルに学習させる学習データである画像データを入力して、前記画像変換モデルによって前記画像データが変換された変換後画像を取得し、

20

コンピュータが、前記変換後画像を前記学習データに加え、

前記画像変換モデルは、前記画像データの説明文を生成する文章生成モデルと、説明文を変換する文書変換モデルと、説明文から画像データを生成する画像生成モデルとを含み、コンピュータが、前記文章生成モデルに対して前記画像データと指定属性を示す属性情報とを入力するとともに、前記画像データに含まれる前記属性情報が示す前記指定属性に関する説明と、前記画像データの画角についての説明とを含む説明文の生成を指示して、前記文章生成モデルによって生成された前記指定属性に関する説明と前記画角についての説明とを含む前記画像データの説明文を取得し、前記文書変換モデルに対して取得された前記説明文を入力するとともに、前記指定属性及び前記画角については変更せず、他の部分を変更するように指示して、前記文書変換モデルによって変換された変換後文を取得し、前記画像生成モデルに対して変換された前記変換後文を入力して、前記画像生成モデルによって生成された画像データを前記変換後画像として取得する学習データ生成方法。

30

【請求項 5】

画像データを変換する画像変換モデルに対して、物体検知モデルに学習させる学習データである画像データを入力して、前記画像変換モデルによって前記画像データが変換された変換後画像を取得する画像変換処理と、

前記画像変換処理によって取得された前記変換後画像を前記学習データに加える学習データ追加処理と

を行う学習データ生成装置としてコンピュータを機能させ、

40

前記画像変換モデルは、前記画像データの説明文を生成する文章生成モデルと、説明文を変換する文書変換モデルと、説明文から画像データを生成する画像生成モデルとを含み、前記画像変換処理では、前記文章生成モデルに対して前記画像データと指定属性を示す属性情報とを入力するとともに、前記画像データに含まれる前記属性情報が示す前記指定属性に関する説明と、前記画像データの画角についての説明とを含む説明文の生成を指示して、前記文章生成モデルによって生成された前記指定属性に関する説明と前記画角についての説明とを含む前記画像データの説明文を取得し、前記文書変換モデルに対して取得された前記説明文を入力するとともに、前記指定属性及び前記画角については変更せず、他の部分を変更するように指示して、前記文書変換モデルによって変換された変換後文を取得し、前記画像生成モデルに対して変換された前記変換後文を入力して、前記画像生成モ

50

デルによって生成された画像データを前記変換後画像として取得する学習データ生成プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、物体検知モデルに対する学習データの生成技術に関する。

【背景技術】

【0002】

画像データから検知対象の属性の物体を検知する物体検知モデルがある。物体検知モデルの検知精度を高めるためには、様々な画像データを学習データとして物体検知モデルに学習させる必要がある。学習データの作成は人手で行われることが多く、作業負担が大きい。また、危険な状況等を学習させる場合には、危険な状況の画像データが必要になるが、人が実際に危険な状況を再現することは難しく、容易に学習データを作成することができない。

10

【0003】

特許文献1には、教師画像が不足しているときに、教師画像に対して拡大、縮小、移動、合成等の加工を加え、新たな教師画像を生成することが記載されている。

【先行技術文献】

【特許文献】

【0004】

【文献】特開2018-169672号公報

20

【発明の概要】

【発明が解決しようとする課題】

【0005】

特許文献1に記載された方法では、基本的に元の教師画像と類似した画像データしか新たな教師画像として生成することはできない。そのため、検知精度を高めるために有効な画像データを生成することが難しかった。

本開示は、人手によらず、検知精度を高めるために有効な画像データを生成可能にすることを目的とする。

【課題を解決するための手段】

【0006】

本開示に係る学習データ生成装置は、

画像データを変換する画像変換モデルに対して、物体検知モデルに学習させる学習データである画像データを入力して、前記画像変換モデルによって前記画像データが変換された変換後画像を取得する画像変換部と、

前記画像変換部によって取得された前記変換後画像を前記学習データに加える学習データ追加部とを備える。

30

【発明の効果】

【0007】

本開示では、画像変換モデルに対して学習データである画像データを入力して、画像変換モデルによって画像データが変換された変換後画像を学習データに加える。これにより、拡大、縮小、移動、合成等の単純な加工に限らず、背景の変更、周辺に存在する物体の変更、光の当たり方の変更といった様々な変更を加えた画像データを生成することが可能である。そのため、検知精度を高めるために有効な画像データを生成可能にすることが可能である。

40

【図面の簡単な説明】

【0008】

【図1】実施の形態1に係る学習データ生成装置10の構成図。

【図2】実施の形態1に係る学習データ生成装置10の処理のフローチャート。

50

【図3】実施の形態1に係る画像変換モデル41への入力の説明図。

【図4】実施の形態2に係る学習データ生成装置10の構成図。

【図5】実施の形態2に係る学習データ生成装置10の処理のフローチャート。

【図6】実施の形態2に係る文章生成モデル42への入力の説明図。

【図7】実施の形態2に係る文章生成モデル42からの出力の説明図。

【図8】実施の形態2に係る文書変換モデル43への入力の説明図。

【図9】実施の形態3に係る学習データ生成装置10の構成図。

【図10】実施の形態3に係る学習データ追加処理のフローチャート。

【図11】実施の形態3に係る属性判定モデル46への入力の説明図。

【図12】実施の形態4に係る学習データ生成装置10の構成図。

10

【図13】実施の形態4に係る学習データ生成装置10の処理のフローチャート。

【発明を実施するための形態】

【0009】

実施の形態1.

構成の説明

図1を参照して、実施の形態1に係る学習データ生成装置10の構成を説明する。

学習データ生成装置10は、コンピュータである。

学習データ生成装置10は、プロセッサ11と、メモリ12と、ストレージ13と、通信インタフェース14とのハードウェアを備える。プロセッサ11は、信号線を介して他のハードウェアと接続され、これら他のハードウェアを制御する。

20

【0010】

プロセッサ11は、プロセッシングを行うICである。ICはIntegrated Circuitの略である。プロセッサ11は、具体例としては、CPU、DSP、GPUである。CPUは、Central Processing Unitの略である。DSPは、Digital Signal Processorの略である。GPUは、Graphics Processing Unitの略である。

【0011】

メモリ12は、データを一時的に記憶する記憶装置である。メモリ12は、具体例としては、SRAM、DRAMである。SRAMは、Static Random Access Memoryの略である。DRAMは、Dynamic Random Access Memoryの略である。

30

【0012】

ストレージ13は、データを保管する記憶装置である。ストレージ13は、具体例としては、SSDである。SSDは、Solid State Driveの略である。また、ストレージ13は、SD(登録商標)メモ리카ード、CompactFlash(登録商標)、NANDフラッシュ、フレキシブルディスク、光ディスク、コンパクトディスク、Blu-ray(登録商標)ディスク、DVDといった可搬記録媒体であってもよい。SDは、Secure Digitalの略である。DVDは、Digital Versatile Diskの略である。

【0013】

通信インタフェース14は、外部の装置と通信するためのインタフェースである。通信インタフェース14は、具体例としては、Ethernet(登録商標)、USB、HDMI(登録商標)のポートである。USBは、Universal Serial Busの略である。HDMIは、High-Definition Multimedia Interfaceの略である。

40

【0014】

学習データ生成装置10は、機能構成要素として、画像変換部21と、学習データ追加部22とを備える。学習データ生成装置10の各機能構成要素の機能はソフトウェアにより実現される。

ストレージ13には、学習データ生成装置10の各機能構成要素の機能を実現するプロ

50

グラムが格納されている。このプログラムは、プロセッサ 11 によりメモリ 12 に読み込まれ、プロセッサ 11 によって実行される。これにより、学習データ生成装置 10 の各機能構成要素の機能が実現される。

ストレージ 13 には、学習データ 31 が記憶されている。学習データ 31 は、物体検知モデルに学習させるためのデータである。学習データ 31 は、複数の画像データを含む。物体検知モデルは、画像データから検知対象の属性の物体を検知する学習モデルである。物体検知モデルは、例えば、ディープラーニングを用いて構成されたモデルである。

【0015】

学習データ生成装置 10 は、通信インターフェース 14 を介して、画像変換モデル 41 と接続されている。

10

画像変換モデル 41 は、画像を変換する、いわゆる生成 AI である。AI は、Artificial Intelligence の略である。画像変換モデル 41 は、具体例としては、BERT、GPT 等のアルゴリズムを用いて構成されてもよい。BERT は、Bidirectional Encoder Representations from Transformers の略である。GPT は、Generative Pretrained Transformer の略である。学習モデル 112 は、これらのアルゴリズムを含む複数のアルゴリズムを組み合わせて構成されてもよい。

【0016】

ここでは、画像変換モデル 41 は、通信インターフェース 14 を介して接続されているとした。つまり、画像変換モデル 41 は学習データ生成装置 10 の外部にあるものとした。しかし、画像変換モデル 41 は学習データ生成装置 10 の内部にあってもよい。

20

【0017】

図 1 では、プロセッサ 11 は、1 つだけ示されていた。しかし、プロセッサ 11 は、複数であってもよく、複数のプロセッサ 11 が、各機能を実現するプログラムを連携して実行してもよい。

【0018】

*** 動作の説明 ***

図 2 及び図 3 を参照して、実施の形態 1 に係る学習データ生成装置 10 の動作を説明する。

実施の形態 1 に係る学習データ生成装置 10 の動作手順は、実施の形態 1 に係る学習データ生成方法に相当する。また、実施の形態 1 に係る学習データ生成装置 10 の動作を実現するプログラムは、実施の形態 1 に係る学習データ生成プログラムに相当する。

30

【0019】

図 2 を参照して、実施の形態 1 に係る学習データ生成装置 10 の処理を説明する。

(ステップ S11: 入力処理)

画像変換部 21 は、学習データ 31 から 1 つの画像データを抽出する。画像変換部 21 は、画像変換モデル 41 に対して、抽出された画像データと、物体検知モデルの検知対象の属性である指定属性を示す属性情報とを入力する。物体検知モデルの検知対象の属性が複数存在する場合には、画像変換部 21 は、複数の属性のうち抽出した画像データに指定された属性情報を入力する。この際、画像変換部 21 は、画像データの変換規則を指示するプロンプトを画像変換モデル 41 に対して入力する。

40

ステップ S11 において抽出される 1 つの画像データは、物体検知モデルの学習前モデル評価段階において、特定の属性が検知できなかった画像あるいは、別の属性が検知された画像が望ましい。

【0020】

例えば、画像変換部 21 は、図 3 に示すように、画像データ及び属性情報とともに、画像データの変換規則を指示するプロンプトを画像変換モデル 41 に対して入力する。変換規則では、ランダムに変更するものと、絶対に変更しないものとが指定される。

ランダムに変更するものとしては、衣服の種類、性別、背景等が指定される。その他、光の当たり方、人数といった項目が状況に応じて指定されてもよい。

50

絶対に変更しないものとしては、属性情報に関する情報と、画角情報とが指定される。属性情報に関する情報とは、属性情報そのものだけでなく、属性情報に付随するものも含まれる。例えば、属性情報が白杖である場合には、属性情報に付随するものとは、白杖を右手で持っていること等である。図3のプロンプトでは、単に「\$ {属性情報}に関する情報」と指定しているが、より具体的に「\$ {属性情報}と、\$ {属性情報}を持つ手」のように指定してもよい。ここで、\$ {属性情報}は、入力情報で指定された属性情報（上記例では「白杖」）を表す。つまり、「\$ {属性情報}と、\$ {属性情報}を持つ手」は、「白杖と、白杖を持つ手」と読み替えられる。その他の\$ {***}も同様に、入力情報における***として指定された情報に置き換えられる。

【0021】

画像変換部21は、作成したい画像データが定まっている場合には、プロンプトとして作成したい画像データを詳細に指示してもよい。

【0022】

（ステップS12：画像変換処理）

画像変換モデル41は、ステップS21で入力された画像データを変換した変換後画像を指定枚数だけ生成する。そして、画像変換モデル41は、指定枚数の変換後画像を出力する。指定枚数は、例えば、学習データ31の画像データの数の1割程度の数である。

この際、画像変換モデル41は、プロンプトでの指示に従い、ランダムに変更するものとして指定されたものについては、変換後画像毎にランダムに変更する。また、画像変換モデル41は、絶対に変更しないものとして指定されたものについては変更しない。つまり、指定属性については、変更されず維持された変換後画像が生成される。ランダムに変更するものと、絶対に変更しないものとのいずれにも指定されていないものについては、任意に変更可能である。

画像変換部21は、画像変換モデル41によって出力された指定枚数の変換後画像を取得する。

【0023】

（ステップS13：学習データ追加処理）

学習データ追加部22は、ステップS12で画像変換部21によって取得された指定枚数の変換後画像を学習データ31に追加する。

図2で示した学習データ追加処理は、学習データに追加が必要な属性情報ごとに処理が実行される。

【0024】

実施の形態1の効果

以上のように、実施の形態1に係る学習データ生成装置10は、画像変換モデル41に対して学習データである画像データを入力して、画像変換モデル41によって画像データが変換された変換後画像を学習データ31に加える。これにより、背景の変更、周辺に存在する物体の変更、光の当たり方の変更といった様々な変更を加えた画像データを生成することが可能である。そのため、検知精度を高めるために有効な画像データを生成可能にすることが可能である。

【0025】

実施の形態1に係る学習データ生成装置10は、危険な状況のように収集が困難な学習データを作成することが可能である。また、実施の形態1に係る学習データ生成装置10は、収集が困難な学習データのバリエーションを増やすことが可能である。

【0026】

実施の形態2 .

実施の形態2は、画像データの説明文を生成させ、説明文を変換させ、変換された説明文から画像データを生成することにより、変換後画像を生成する点の実施の形態1と異なる。実施の形態2では、この異なる点を説明し、同一の点については説明を省略する。

【0027】

構成の説明

10

20

30

40

50

図4を参照して、実施の形態2に係る学習データ生成装置10の構成を説明する。

画像変換モデル41は、文章生成モデル42と、文書変換モデル43と、画像生成モデル44とを含む。

文章生成モデル42は、画像データの説明文を生成するモデルである。文書変換モデル43は、説明文を変換するモデルである。画像生成モデル44は、説明文から画像データを生成するモデルである。

【0028】

動作の説明

図5を参照して、実施の形態2に係る学習データ生成装置10の処理を説明する。

(ステップS21:画像入力処理)

画像変換部21は、学習データ31から1つの画像データを抽出する。画像変換部21は、文章生成モデル42に対して、抽出された画像データと、物体検知モデルの検知対象の属性である指定属性を示す属性情報とを入力する。物体検知モデルの検知対象の属性が複数存在する場合には、画像変換部21は、複数の属性のうち1つ以上の属性を示す属性情報を入力する。この際、画像変換部21は、説明文の生成規則を指示するプロンプトを文章生成モデル42に対して入力する。

【0029】

例えば、画像変換部21は、図6に示すように、画像データ及び属性情報とともに、説明文の生成規則を指示するプロンプトを文章生成モデル42に対して入力する。説明文の生成規則では、画角の情報と属性情報とについては必ず説明するように指定される。

【0030】

(ステップS22:説明文生成処理)

文章生成モデル42は、ステップS21で入力された画像データを説明した説明文を生成して出力する。この際、図7に示すように、文章生成モデル42は、プロンプトでの指示に従い、画角の情報と属性情報とについての説明を含む画像データ全体の説明文を生成する。そして、画像変換部21は、文章生成モデル42によって出力された説明文を取得する。

【0031】

(ステップS23:説明文入力処理)

画像変換部21は、文書変換モデル43に対して、ステップS22で取得された説明文と、物体検知モデルの検知対象の属性である指定属性を示す属性情報とを入力する。この際、画像変換部21は、説明文の変換規則を指示するプロンプトを画像変換モデル41に対して入力する。

例えば、画像変換部21は、図8に示すように、説明文及び属性情報とともに、説明文の変換規則を指示するプロンプトを文書変換モデル43に対して入力する。変換規則では、図3と同様に、ランダムに変更するものと、絶対に変更しない物とが指定される。また、図8では、説明文を英語に変換させている。これは、英語の説明文の方が、説明文から画像を生成する精度が高くなる傾向があるためである。

【0032】

(ステップS24:説明文変換処理)

文書変換モデル43は、ステップS23で入力された説明文を変換した変換後文を生成して出力する。そして、画像変換部21は、文書変換モデル43によって出力された変換後文を取得する。

この際、文書変換モデル43は、プロンプトでの指示に従い、ランダムに変更するものとして指定されたものについては、変換後文毎にランダムに変更する。また、文書変換モデル43は、絶対に変更しないものとして指定されたものについては変更しない。ランダムに変更するものと、絶対に変更しないものとのいずれにも指定されていないものについては、任意に変更可能である。また、文書変換モデル43は、説明文を英文に変換する。なお、ここでは、変換後文が画像生成用プロンプトとして生成される。

【0033】

10

20

30

40

50

(ステップ S 2 5 : 変換後文入力処理)

画像変換部 2 1 は、画像生成モデル 4 4 に対して、ステップ S 2 4 で取得された変換後文 (画像生成用プロンプト) を入力する。この際、画像変換部 2 1 は、変換後文を表す画像データを指定枚数だけ生成するように指示するプロンプトを画像生成モデル 4 4 に対して入力する。

【 0 0 3 4 】

(ステップ S 2 6 : 画像生成処理)

画像生成モデル 4 4 は、ステップ S 2 5 で入力された変換後文を表す画像データを指定枚数だけ生成する。画像生成モデル 4 4 は、変換後文を表しつつ、画像データ毎にランダムに変更を加えて画像データを生成する。そして、画像生成モデル 4 4 は、指定枚数の変換後画像を出力する。

10

画像変換部 2 1 は、画像生成モデル 4 4 によって出力された指定枚数の変換後画像を取得する。

【 0 0 3 5 】

(ステップ S 2 7 : 学習データ追加処理)

学習データ追加部 2 2 は、ステップ S 2 6 で画像変換部 2 1 によって取得された指定枚数の変換後画像を学習データ 3 1 に追加する。

【 0 0 3 6 】

*** 実施の形態 2 の効果 ***

以上のように、実施の形態 2 に係る学習データ生成装置 1 0 は、画像データの説明文を生成させ、説明文を変換させ、変換された説明文から画像データを生成することにより、変換後画像を生成する。説明文を変換した変換後文を生成することで、変換した内容を明確化することができる。

20

【 0 0 3 7 】

また、実施の形態 1 に係る学習データ生成装置 1 0 は、文章生成モデル 4 2 と文書変換モデル 4 3 と画像生成モデル 4 4 とを別の生成 AI とすることができる。これにより、文書変換モデル 4 3 を言語処理に特化したモデルとし、文章生成モデル 4 2 と画像生成モデル 4 4 とを画像処理に特化したモデルとするといったことが可能になる。その結果、処理速度を早くまた処理精度を高くすることが可能である。

【 0 0 3 8 】

*** 他の構成 ***

< 変形例 1 >

実施の形態 2 では、1 個の変換後文が生成され、1 個の変換後文から指定枚数の変換後画像が生成された。指定枚数の変換後文が生成され、1 個の変換後文から 1 個の変換後画像が生成されてもよい。また、M 個の変換後文が生成され、1 つの変換後文から N 個の変換後画像が生成されてもよい。このとき、M 及び N は 1 以上の整数であり、 $M \times N =$ 指定枚数である。

30

【 0 0 3 9 】

実施の形態 3 .

実施の形態 3 は、変換後画像を学習データ 3 1 に追加するか否かを判断する点が実施の形態 1 , 2 と異なる。実施の形態 3 では、この異なる点を説明し、同一の点については説明を省略する。

40

【 0 0 4 0 】

*** 構成の説明 ***

図 9 を参照して、実施の形態 3 に係る学習データ生成装置 1 0 の構成を説明する。

学習データ生成装置 1 0 は、通信インタフェース 1 4 を介して、物体検知モデル 4 5 と、属性判定モデル 4 6 と接続されている。

物体検知モデル 4 5 は、画像データから検知対象の属性の物体を検知するモデルであり、学習データ 3 1 を学習するモデルである。

属性判定モデル 4 6 は、画像データに指定された属性の物体が含まれるか否かを判定する

50

モデルである。属性判定モデル 4 6 は、画像変換モデル 4 1 と同様に、いわゆる生成 A I である。

【 0 0 4 1 】

*** 動作の説明 ***

図 1 0 を参照して、実施の形態 3 に係る学習データ追加処理（図 2 のステップ S 1 3、図 5 のステップ S 2 7）を説明する。

学習データ追加処理では、各変換後画像を対象として、以下の処理が実行される。

【 0 0 4 2 】

（ステップ S 3 1：第 1 入力処理）

学習データ追加部 2 2 は、物体検知モデル 4 5 に対して、対象の変換後画像と、変換後画像を生成する際に入力とした属性情報とを入力する。変換後画像を生成する際に入力とした属性情報とは、図 2 のステップ S 1 1 又は図 5 のステップ S 2 1 で入力した属性情報である。

10

【 0 0 4 3 】

（ステップ S 3 2：物体検知処理）

物体検知モデル 4 5 は、ステップ S 3 1 で入力された対象の変換後画像から、ステップ S 3 1 で入力された属性情報が示す属性の物体を検知する。物体検知モデル 4 5 は、検知結果と、検知の確信度とを出力する。

学習データ追加部 2 2 は、物体検知モデル 4 5 によって出力された検知結果及び確信度を取得する。

20

【 0 0 4 4 】

（ステップ S 3 3：確信度判定処理）

学習データ追加部 2 2 は、ステップ S 3 2 で取得された確信度が閾値以上であるか否かを判定する。

学習データ追加部 2 2 は、確信度が閾値以上である場合には、対象の変換後画像を学習データ 3 1 に追加せず、対象の変換後画像についての処理を終了する。一方、学習データ追加部 2 2 は、確信度が閾値未満である場合には、処理をステップ S 3 4 に進める。

【 0 0 4 5 】

（ステップ S 3 4：第 2 入力処理）

学習データ追加部 2 2 は、属性判定モデル 4 6 に対して、対象の変換後画像と、変換後画像を生成する際に入力とした属性情報とを入力する。この際、図 1 1 に示すように、学習データ追加部 2 2 は、入力された変換後画像に、入力された属性情報が示す属性の物体が存在するか否かを判定するように指示するプロンプトを属性判定モデル 4 6 に入力する。

30

【 0 0 4 6 】

（ステップ S 3 5：属性判定処理）

属性判定モデル 4 6 は、ステップ S 3 4 で入力された対象の変換後画像に、ステップ S 3 1 で入力された属性情報が示す属性の物体が存在するか否かを判定する。属性判定モデル 4 6 は、判定結果を出力する。

学習データ追加部 2 2 は、属性判定モデル 4 6 によって出力された判定結果を取得する。

【 0 0 4 7 】

（ステップ S 3 6：結果判定処理）

学習データ追加部 2 2 は、ステップ S 3 5 で取得された判定結果が、対象の変換後画像に物体が存在することを示す場合には、処理をステップ S 3 7 に進める。一方、学習データ追加部 2 2 は、ステップ S 3 5 で取得された判定結果が、対象の変換後画像に物体が存在しないことを示す場合には、対象の変換後画像を学習データ 3 1 に追加せず、対象の変換後画像についての処理を終了する。

40

【 0 0 4 8 】

（ステップ S 3 7：追加処理）

学習データ追加部 2 2 は、対象の変換後画像を、学習データ 3 1 に追加する。

【 0 0 4 9 】

50

実施の形態3の効果

実施の形態3に係る学習データ生成装置10は、物体検知モデル45及び属性判定モデル46を用いて変換後画像を学習データ31に追加するか否かを判定する。これにより、検知精度を劣化させてしまうような変換後画像、あるいは、学習が不要な変換後画像を学習データ31に追加してしまうことを防止可能である。

具体的にはステップS33の判定では、現在の物体検知モデルで検知できるものを学習データ31に追加しないことで物体検知モデルの過学習を防いでいる。また、ステップS36の判定では、変換後画像で学習したい属性情報を有していないものを学習データ31に追加しないことで、関係ない画像を学習データ31に入れないようにしている。

【0050】

他の構成

<変形例2>

実施の形態3では、物体検知モデル45及び属性判定モデル46を用いて登録可能になった変換後画像だけが学習データ31に追加された。そのため、指定枚数の変換後画像が生成されても、指定枚数の変換後画像が学習データ31に追加されるとは限らなかった。

そこで、学習データ追加部22は、少なくとも一部の変換後画像が学習データ31に追加されなかった場合には、処理を画像変換部21に戻して、不足数の変換後画像を再生成させてもよい。不足数とは、指定枚数から学習データ31に追加された変換後画像の数を減算した値である。つまり、学習データ追加部22は、不足数を新たな指定枚数として、処理を図2のステップS11又は図5のステップS21からやり直させるようにしてもよい。

【0051】

実施の形態4

実施の形態4は、物体検知モデル45の学習まで行う点が実施の形態3と異なる。実施の形態4では、この異なる点を説明し、同一の点については説明を省略する。

【0052】

構成の説明

図12を参照して、実施の形態4に係る学習データ生成装置10の構成を説明する。

学習データ生成装置10は、機能構成要素として、学習部23を備える点が、図9に示す学習データ生成装置10と異なる。学習部23は、画像変換部21及び学習データ追加部22と同様に、ソフトウェアによって実現される。

【0053】

動作の説明

図13を参照して、実施の形態4に係る学習データ生成装置10の処理を説明する。

(ステップS41：生成追加処理)

画像変換部21及び学習データ追加部22は、実施の形態3で説明した処理により、学習データ31に変換後画像を追加する。

【0054】

(ステップS42：学習処理)

学習部23は、ステップS41で変換後画像が追加された学習データ31を用いて、物体検知モデル45に学習させる。

【0055】

(ステップS43：モデル評価処理)

学習部23は、ステップS42で学習された物体検知モデル45の評価を行う。物体検知モデル45の評価は既存の技術を用いて行うことが可能である。

学習部23は、ステップS43で得られた評価が以前の評価よりも下がっている場合には、処理をステップS44に進める。学習部23は、ステップS43で得られた評価が以前の評価以上であるが、基準に達していない場合には、処理をステップS41に戻す。学習部23は、ステップS43で得られた評価が以前の評価以上であるが、基準に達している場合には、処理を終了する。

10

20

30

40

50

【0056】

(ステップS44:データ削除処理)

学習部23は、ステップS41で追加された変換後画像を学習データ31から削除する。また、学習部23は、物体検知モデル45を、ステップS42での学習前の状態に戻す。そして、学習部23は、処理をステップS41に戻す。

【0057】

実施の形態4の効果

以上のように、実施の形態4に係る学習データ生成装置10は、学習データ31の追加と物体検知モデル45の学習とを繰り返す。これにより、人手によることなく物体検知モデル45の検知精度を高めることが可能である。

10

【0058】

他の構成

<変形例3>

以上の実施の形態では、各機能構成要素がソフトウェアで実現された。しかし、変形例3として、各機能構成要素はハードウェアで実現されてもよい。この変形例3について、以上の実施の形態と異なる点を説明する。

【0059】

各機能構成要素がハードウェアで実現される場合には、学習データ生成装置10は、プロセッサ11とメモリ12とストレージ13とに代えて、電子回路を備える。電子回路は、各機能構成要素と、メモリ12と、ストレージ13との機能とを実現する専用の回路である。

20

【0060】

電子回路としては、単回路、複合回路、プログラム化したプロセッサ、並列プログラム化したプロセッサ、ロジックIC、GA、ASIC、FPGAが想定される。GAは、Gate Arrayの略である。ASICは、Application Specific Integrated Circuitの略である。FPGAは、Field-Programmable Gate Arrayの略である。

各機能構成要素を1つの電子回路で実現してもよいし、各機能構成要素を複数の電子回路に分散させて実現してもよい。

【0061】

<変形例4>

変形例4として、一部の各機能構成要素がハードウェアで実現され、他の各機能構成要素がソフトウェアで実現されてもよい。

30

【0062】

プロセッサ11とメモリ12とストレージ13と電子回路とを処理回路という。つまり、各機能構成要素の機能は、処理回路により実現される。

【0063】

また、以上の説明における「部」を、「回路」、「工程」、「手順」、「処理」又は「処理回路」に読み替えてもよい。

【0064】

以下、本開示の諸態様を付記としてまとめて記載する。

(付記1)

画像データを変換する画像変換モデルに対して、物体検知モデルに学習させる学習データである画像データを入力して、前記画像変換モデルによって前記画像データが変換された変換後画像を取得する画像変換部と、

前記画像変換部によって取得された前記変換後画像を前記学習データに加える学習データ追加部と
を備える学習データ生成装置。

(付記2)

前記画像変換部は、前記画像データに指定属性が含まれる状態を維持するように前記画

40

50

像変換モデルに対して指示して、前記画像データを変換させる
付記 1 に記載の学習データ生成装置。

(付記 3)

前記画像変換部は、前記画像データにおける指定属性に関する部分については変更せず、他の部分を変更するように前記画像変換モデルに対して指示して、前記画像データを変換させる

付記 1 又は 2 に記載の学習データ生成装置。

(付記 4)

前記画像変換モデルは、画像データの説明文を生成する文章生成モデルと、説明文を変換する文書変換モデルと、説明文から画像データを生成する画像生成モデルとを含み、

前記画像変換部は、前記文章生成モデルに対して前記画像データを入力して、前記文章生成モデルによって生成された前記画像データの説明文を取得し、前記文書変換モデルに対して取得された前記説明文を入力して、前記文書変換モデルによって変換された変換後文を取得し、前記画像生成モデルに対して変換された前記変換後文を入力して、前記画像生成モデルによって生成された画像データを前記変換後画像として取得する

付記 1 又は 2 に記載の学習データ生成装置。

(付記 5)

前記画像変換部は、前記説明文における指定属性に関する部分については変更せず、他の部分を変更するように前記文書変換モデルに対して指示して、前記説明文を変換させる
付記 4 に記載の学習データ生成装置。

(付記 6)

前記学習データ生成装置は、さらに、

前記物体検知モデルに対して前記画像変換部によって取得された前記変換後画像を入力して、前記物体検知モデルで物体を検知した確信度が評価閾値以上であるか否かを判定するデータ判定部

を備え、

前記学習データ追加部は、前記データ判定部によって前記確信度が前記評価閾値未満であると判定された場合に、前記変換後画像を前記学習データに加える

付記 1 から 5 までのいずれか 1 項に記載の学習データ生成装置。

(付記 7)

前記データ判定部は、前記確信度が前記評価閾値未満である場合に、画像データに特定の属性の物体が存在するか否かを判定する属性判定モデルに対して、前記変換後画像と指定属性とを入力して、前記変換後画像に前記指定属性の物体が存在するか否かの判定結果を取得し、

前記学習データ追加部は、前記データ判定部によって前記変換後画像に前記指定属性の物体が存在するという判定結果が取得された場合に、前記変換後画像を前記学習データに加える

付記 6 に記載の学習データ生成装置。

(付記 8)

コンピュータが、画像データを変換する画像変換モデルに対して、物体検知モデルに学習させる学習データである画像データを入力して、前記画像変換モデルによって前記画像データが変換された変換後画像を取得し、

コンピュータが、前記変換後画像を前記学習データに加える学習データ生成方法。

(付記 9)

画像データを変換する画像変換モデルに対して、物体検知モデルに学習させる学習データである画像データを入力して、前記画像変換モデルによって前記画像データが変換された変換後画像を取得する画像変換処理と、

前記画像変換処理によって取得された前記変換後画像を前記学習データに加える学習データ追加処理と

を行う学習データ生成装置としてコンピュータを機能させる学習データ生成プログラム。

10

20

30

40

50

【0065】

以上、本開示の実施の形態及び変形例について説明した。これらの実施の形態及び変形例のうち、いくつかを組み合わせる実施してもよい。また、いずれか1つ又はいくつかを部分的に実施してもよい。なお、本開示は、以上の実施の形態及び変形例に限定されるものではなく、必要に応じて種々の変更が可能である。

【符号の説明】

【0066】

10 学習データ生成装置、11 プロセッサ、12 メモリ、13 ストレージ、14 通信インタフェース、21 画像変換部、22 学習データ追加部、23 学習部、31 学習データ、41 画像変換モデル、42 文章生成モデル、43 文書変換モデル、44 画像生成モデル、45 物体検知モデル、46 属性判定モデル。

10

20

30

40

50

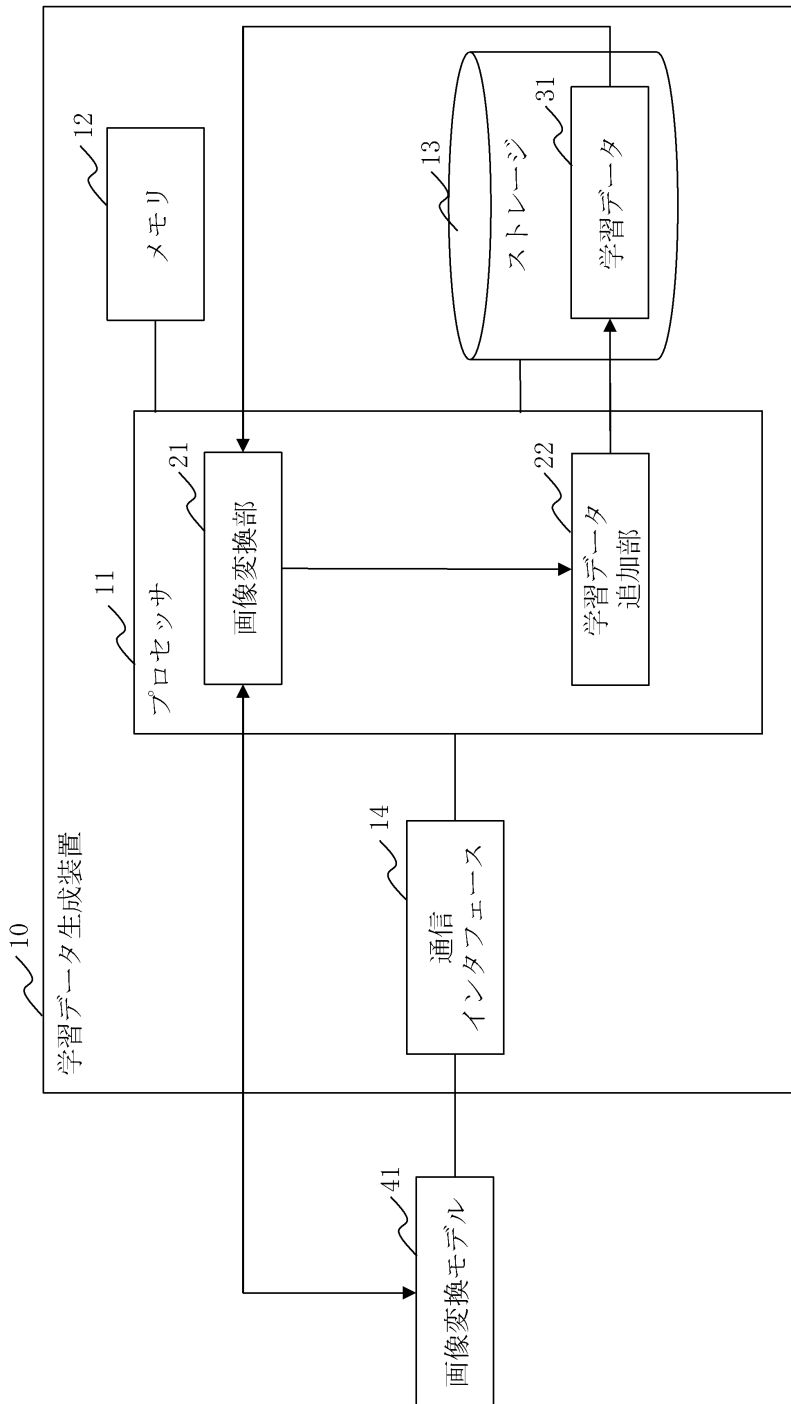
【要約】

【課題】人手によらず、検知精度を高めるために有効な画像データを生成可能にする。

【解決手段】画像変換部 21 は、画像データを変換する画像変換モデル 41 に対して、画像データから検知対象の属性の物体を検知する物体検知モデルに学習させる学習データ 31 である画像データを入力して、画像変換モデル 41 によって画像データが変換された変換後画像を取得する。学習データ追加部 22 は、画像変換部 21 によって取得された変換後画像を学習データ 31 に加える。

【選択図】図 1

図1



10

20

30

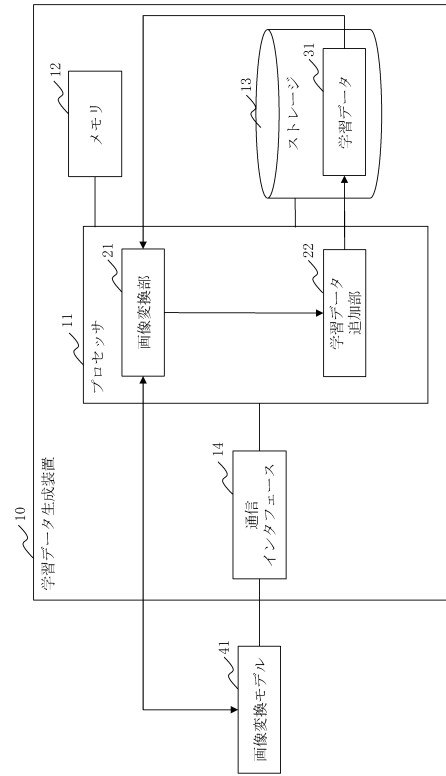
40

50

【図面】

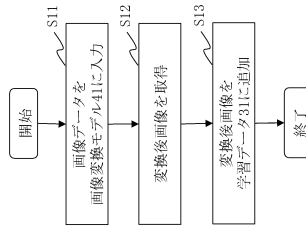
【図 1】

図1



【図 2】

図2

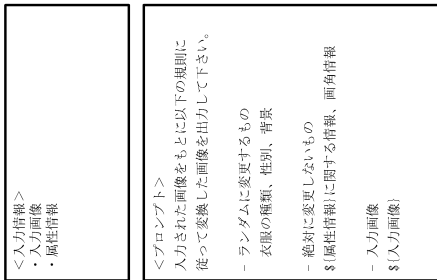


10

20

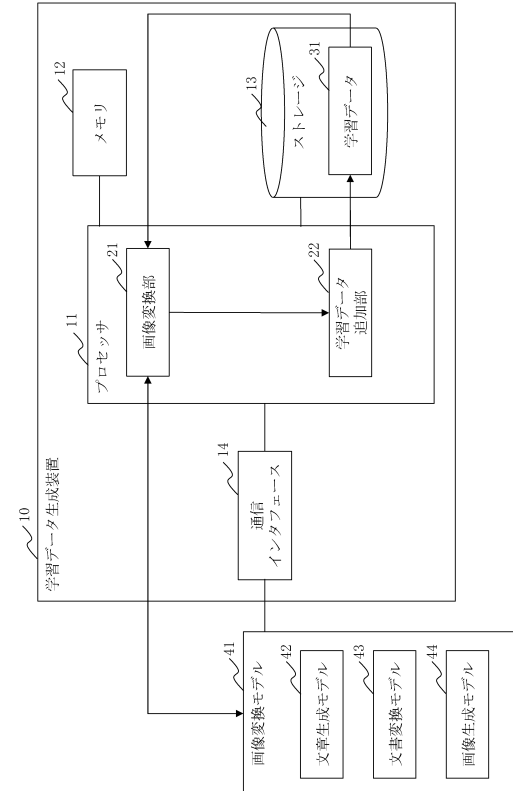
【図 3】

図3



【図 4】

図4



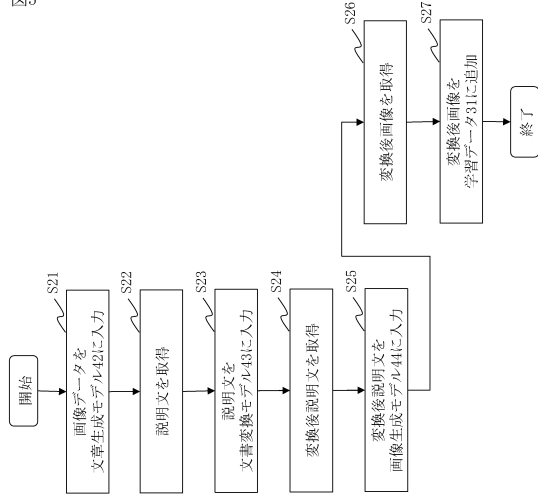
30

40

50

【 図 5 】

図5



【 図 7 】

図7

<出力情報> ・説明文 (状況を説明した文章)	<回答> 画像には、男性と女性が歩道を横断している様子が写っています。男性は赤いボロシャツを着て、ベージュのショートパンツを履いています。彼はサングラスをかけています。女性は青い花柄のシャツを着て、黒いパンツを履いています。彼女はサングラスをかけていて、白い杖を持っています。男性は女性より少し前に歩いています。写真画像の画角は、2人の人物が中心に配置され、彼らの足元から腰までの部分に焦点を当てています。歩道と道路が背景に見えています。道路の向こう側には、建物、木、駐車中の車が写っています。写真女性が行っている白い杖は、視覚障害者のための歩行補助具です。この杖は、障害物や危険を検知し、歩行者を安全に導くために使用されます。
----------------------------	---

【 図 6 】

図6

<入力情報> ・入力画像 ・属性情報	あなたは優れた画像分析者です。画像の状況を詳しく説明してください。また、画角の情報と\$[属性情報]も必ず説明してください。
--------------------------	--

10

【 図 8 】

図8

<入力情報> ・説明文 (状況を説明した文章) ・属性情報	<プロンプト> 入力された文章について、以下の規則にしたがって文章を変換して、画像生成に使用するプロンプトを英語で作成してください。 - ランダムに変更するもの 衣服の種類、性別、背景 - 絶対に変更しないもの \$[属性情報]に関する情報、画角情報 - 入力文章 \$[説明文]
-------------------------------------	---

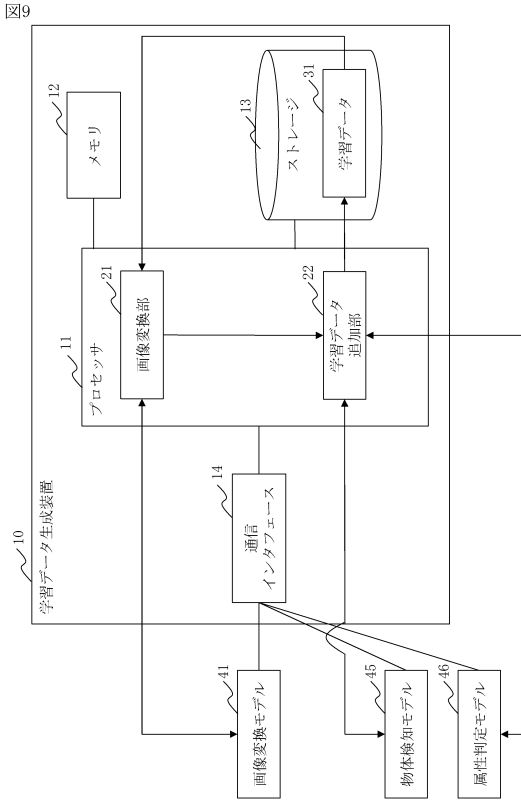
20

30

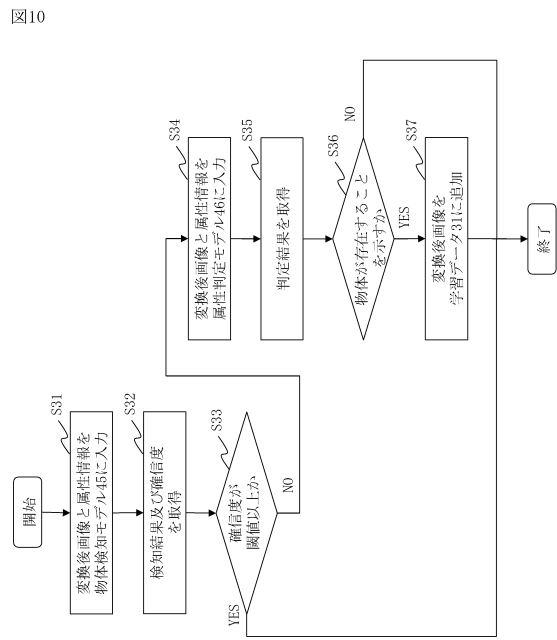
40

50

【 図 9 】



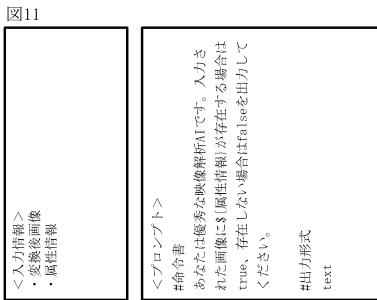
【 図 10 】



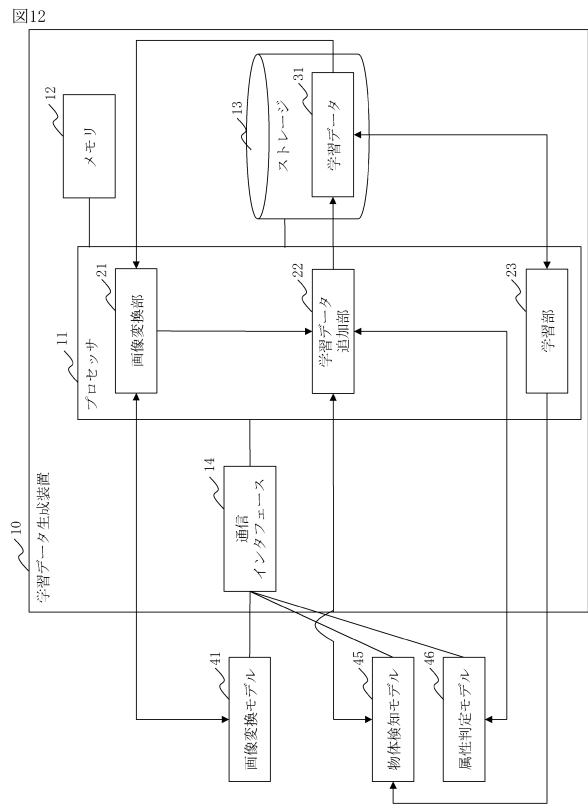
10

20

【 図 11 】



【 図 12 】



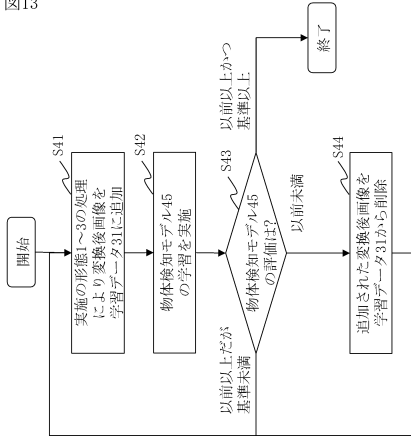
30

40

50

【 図 13 】

図13



10

20

30

40

50

 フロントページの続き

(56)参考文献

特開 2016-076073 (JP, A)

特開 2018-181157 (JP, A)

特開 2019-212073 (JP, A)

特開 2023-044190 (JP, A)

特開 2024-211898 (JP, A)

Bolong Liu et al. , "ACIGS: An automated large-scale crops image generation system based on large visual language multi-modal models" , 2023 20th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON) , 米国 , IEEE , 2023年10月23日 , pp.7-13

Yutong Xie et al. , "PairAug: What Can Augmented Image-Text Pairs Do for Radiology?" , 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) , 米国 , IEEE , 2024年06月16日 , pp.11652-11661

Qiufeng Wu et al. , "DCGAN-Based Data Augmentation for Tomato Leaf Disease Identification" , IEEE Access , 米国 , IEEE , 2020年05月25日 , Vol.8 , pp.98716-98728

(58)調査した分野 (Int.Cl. , DB名)

G06T 1/00 , 1/40 , 7/00 - 7/90 , 11/00

G06V 10/00 - 10/98