

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

**特許第6647632号**  
(P6647632)

(45) 発行日 令和2年2月14日(2020.2.14)

(24) 登録日 令和2年1月17日(2020.1.17)

(51) Int. Cl.		F I			
<b>GO6N</b>	<b>20/00</b>	<b>(2019.01)</b>	GO6N	20/00	130
<b>GO6N</b>	<b>3/08</b>	<b>(2006.01)</b>	GO6N	3/08	
<b>GO6T</b>	<b>7/00</b>	<b>(2017.01)</b>	GO6T	7/00	350C

請求項の数 6 (全 12 頁)

<p>(21) 出願番号 特願2017-169856 (P2017-169856)</p> <p>(22) 出願日 平成29年9月4日(2017.9.4)</p> <p>(65) 公開番号 特開2019-46269 (P2019-46269A)</p> <p>(43) 公開日 平成31年3月22日(2019.3.22)</p> <p>審査請求日 令和1年8月5日(2019.8.5)</p> <p>早期審査対象出願</p>	<p>(73) 特許権者 517311242 株式会社SOAT 東京都千代田区神田平河町1</p> <p>(74) 代理人 100114971 弁理士 青木 修</p> <p>(72) 発明者 渡邊 誓旅 東京都千代田区神田平河町1 株式会社SOAT内</p> <p>審査官 多胡 滋</p>
--	--

最終頁に続く

(54) 【発明の名称】 機械学習用訓練データの生成

(57) 【特許請求の範囲】

【請求項1】

入力画像から出力画像を生成する機械学習システム用の訓練データを自動生成する訓練データ生成システムにおいて、

原サンプル画像に対応する原ラベル画像の類似画像を、擬似ラベル画像として生成する類似画像生成部と、

前記原ラベル画像から前記原サンプル画像への変換特性に従って前記擬似ラベル画像から類推される画像を、前記擬似ラベル画像に対応する擬似サンプル画像として生成する類推画像生成部とを備え、

前記擬似サンプル画像は、前記訓練データにおける前記機械学習システムの入力画像であり、前記擬似ラベル画像は、前記訓練データにおける前記機械学習システムの出力画像であること、

を特徴とする訓練データ生成システム。

【請求項2】

前記類似画像生成部は、1または複数の乱数値を入力値とし入力値に対応する出力画像を生成する画像生成エンジンを備え、複数の前記原ラベル画像を使用して前記画像生成エンジンの機械学習を行い、前記画像生成エンジンに別の乱数値を入力したときの出力画像を、前記原ラベル画像の類似画像として取得し、前記擬似ラベル画像とすることを特徴とする請求項1記載の訓練データ生成システム。

【請求項3】

10

20

前記類推画像生成部は、前記擬似ラベル画像の各画素を注目画素として、前記注目画素についての所定サイズの周辺領域の画素値分布に最も近い画素値分布を周辺領域に有する画素を前記原ラベル画像内で探索し、発見した画素の位置と同一の位置の前記原サンプル画像の画素値を、前記擬似サンプル画像において同一位置の画素にセットすることで、前記擬似サンプル画像を生成することを特徴とする請求項1または請求項2記載の訓練データ生成システム。

【請求項4】

前記原サンプル画像は、細胞組織画像であり、

前記原ラベル画像は、前記細胞組織画像内の細胞の境界線を示す画像を含むこと、

を特徴とする請求項1から請求項3のうちのいずれか1項記載の訓練データ生成システム。 10

【請求項5】

入力画像から出力画像を生成する機械学習システム用の訓練データを自動生成する訓練データ生成方法において、

原サンプル画像に対応する原ラベル画像の類似画像を、擬似ラベル画像として生成する類似画像生成ステップと、

前記原ラベル画像から前記原サンプル画像への変換特性に従って前記擬似ラベル画像から類推される画像を、前記擬似ラベル画像に対応する擬似サンプル画像として生成する類推画像生成ステップとを備え、

前記擬似サンプル画像は、前記訓練データにおける前記機械学習システムの入力画像であり、前記擬似ラベル画像は、前記訓練データにおける前記機械学習システムの出力画像であること、 20

を特徴とする訓練データ生成方法。

【請求項6】

コンピュータに、入力画像から出力画像を生成する機械学習システム用の訓練データを自動生成させる訓練データ生成プログラムであって、

前記コンピュータに、

原サンプル画像に対応する原ラベル画像の類似画像を、擬似ラベル画像として生成する類似画像生成ステップと、

前記原ラベル画像から前記原サンプル画像への変換特性に従って前記擬似ラベル画像から類推される画像を、前記擬似ラベル画像に対応する擬似サンプル画像として生成する類推画像生成ステップとを実行させ、 30

前記擬似サンプル画像は、前記訓練データにおける前記機械学習システムの入力画像であり、前記擬似ラベル画像は、前記訓練データにおける前記機械学習システムの出力画像であること、

を特徴とする訓練データ生成プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、機械学習用訓練データの生成に関するものである。 40

【背景技術】

【0002】

ある画像処理方法では、ニューラルネットワークを使用して、生物学的組織サンプルの画像内の生物学的構成を自動検出している（例えば特許文献1参照）。

【0003】

一般的に、このような、画像が入力されるニューラルネットワークの訓練データとしては、サンプル画像を入手して、そのサンプル画像に対するラベルを手作業で付したものが使用される。

【先行技術文献】

【特許文献】

【 0 0 0 4 】

【特許文献1】特表2017-516992号公報

【発明の概要】

【発明が解決しようとする課題】

【 0 0 0 5 】

しかしながら、ディープニューラルネットワークのような機械学習システムでは多くの訓練データセット（つまり、多くの入力と出力との対）が必要になるため、適切な訓練データを作成するのに長い時間がかかってしまう。

【 0 0 0 6 】

原訓練データからの擬似的な訓練データセットを生成する方法としては、データ拡張がある。データ拡張では、原訓練データセットの原サンプル画像（機械学習用の入力画像）に対して、平行移動、鏡像反転、回転、所定の幾何学的変形、若干の色変更、ノイズ付加などを行うことで、新たな訓練データセットを生成している。通常、データ拡張では、原サンプル画像に対して付されているラベルが、そのまま、新たな訓練データセットでも使用される。ただし、機械学習システムの出力が画像（ラベル画像）である場合には、データ拡張では、原訓練データセットの原ラベル画像（機械画像用の出力画像）に、原サンプル画像の平行移動、鏡像反転、回転、および所定の幾何学的変形に合わせて、同様の平行移動、鏡像反転、回転、および所定の幾何学的変形を施した画像が、新たな訓練データセットのラベル画像として使用される。なお、若干の色変更、およびノイズ付加によるデータ拡張の場合、原ラベル画像と同一の画像が新たな訓練データセットのラベル画像として使用される。

【 0 0 0 7 】

しかし、データ拡張を使用した場合、原画像の訓練データセットのみに比べて過学習が発生しにくいものの、データ拡張で生成した新たな訓練データセットのサンプル画像およびラベル画像は、上述のような比較的単純な方法で加工されたものであり、実際に取得された原サンプル画像および原ラベル画像の特徴を多く残しているため、データ拡張で多くの訓練データセットを新たに生成したとしても、原サンプル画像および原ラベル画像の少なさに起因して、多くの訓練データセットで機械学習が行われるディープニューラルネットワークのような機械学習システムでは依然として過学習が発生しやすい。

【 0 0 0 8 】

本発明は、上記の問題に鑑みてなされたもので、短時間で多くの適切な訓練データセットを自動的に生成する訓練データ生成システム、訓練データ生成方法、および訓練データ生成プログラムを得ることを目的とする。

【課題を解決するための手段】

【 0 0 0 9 】

本発明に係る訓練データ生成システムは、入力画像から出力画像を生成する機械学習システム用の訓練データを自動生成する訓練データ生成システムであり、原サンプル画像に対応する原ラベル画像の類似画像を、擬似ラベル画像として生成する類似画像生成部と、原ラベル画像から原サンプル画像への変換特性に従って擬似ラベル画像から類推される画像を、擬似ラベル画像に対応する擬似サンプル画像として生成する類推画像生成部とを備える。ここで、擬似サンプル画像は、訓練データにおける機械学習システムの入力画像であり、擬似ラベル画像は、訓練データにおける機械学習システムの出力画像である。

【 0 0 1 0 】

本発明に係る訓練データ生成方法は、入力画像から出力画像を生成する機械学習システム用の訓練データを自動生成する訓練データ生成方法であり、原サンプル画像に対応する原ラベル画像の類似画像を、擬似ラベル画像として生成する類似画像生成ステップと、原ラベル画像から原サンプル画像への変換特性に従って擬似ラベル画像から類推される画像を、擬似ラベル画像に対応する擬似サンプル画像として生成する類推画像生成ステップとを備える。ここで、擬似サンプル画像は、訓練データにおける機械学習システムの入力画像であり、擬似ラベル画像は、訓練データにおける機械学習システムの出力画像である。

## 【 0 0 1 1 】

本発明に係る訓練データ生成プログラムは、コンピュータに、入力画像から出力画像を生成する機械学習システム用の訓練データを自動生成させる訓練データ生成プログラムであり、コンピュータに、原サンプル画像に対応する原ラベル画像の類似画像を、擬似ラベル画像として生成する類似画像生成ステップと、原ラベル画像から原サンプル画像への変換特性に従って擬似ラベル画像から類推される画像を、擬似ラベル画像に対応する擬似サンプル画像として生成する類推画像生成ステップとを実行させる。ここで、擬似サンプル画像は、訓練データにおける機械学習システムの入力画像であり、擬似ラベル画像は、訓練データにおける機械学習システムの出力画像である。

## 【 発明の効果 】

## 【 0 0 1 2 】

本発明によれば、短時間で多くの適切な訓練データセットが自動的に生成される。

## 【 0 0 1 3 】

本発明の上記又は他の目的、特徴および優位性は、添付の図面とともに以下の詳細な説明から更に明らかになる。

## 【 図面の簡単な説明 】

## 【 0 0 1 4 】

【 図 1 】 図 1 は、本発明の実施の形態に係る訓練データ生成システムの構成を示すブロック図である。

【 図 2 】 図 2 は、入力画像から出力画像を生成する機械学習システムについて説明する図である。

【 図 3 】 図 3 は、図 1 に示すシステムにおける訓練データ生成方法について説明するフローチャートである。

【 図 4 】 図 4 は、図 1 に示すシステムにおける原訓練データセットと擬似訓練データセットとの対応関係を説明する図である。

【 図 5 】 図 5 は、原サンプル画像および原ラベル画像の具体例を説明する図である。

【 図 6 】 図 6 は、擬似ラベル画像の生成の具体例を説明する図である。

【 図 7 】 図 7 は、擬似サンプル画像の生成の具体例を説明する図である。

【 図 8 】 図 8 は、機械学習後の画像生成エンジンの出力画像の具体例を説明する図である。

。

## 【 発明を実施するための形態 】

## 【 0 0 1 5 】

以下、図に基づいて本発明の実施の形態を説明する。

## 【 0 0 1 6 】

図 1 は、本発明の実施の形態に係る訓練データ生成システムの構成を示すブロック図である。図 1 に示す訓練データ生成システムは、1台の情報処理装置で構成されているが、後述の各処理部を、互いにデータ通信可能な複数の情報処理装置に分散させてもよい。

## 【 0 0 1 7 】

この実施の形態に係る訓練データ生成システムは、ディープニューラルネットワークなどの画像生成エンジンの学習に使用する訓練データセットに使用可能な複数（多数）の擬似サンプル画像および擬似ラベル画像を、1または複数の原訓練データセット（原サンプル画像と原ラベル画像との対）から自動的に生成する。

## 【 0 0 1 8 】

図 2 は、入力画像から出力画像を生成する機械学習システムについて説明する図である。図 2 に示すように、この機械学習システムは、例えばディープニューラルネットワークなどの画像生成エンジンを使用して、入力画像から、その入力画像に対応する出力画像を自動的に生成するためのシステムである。

## 【 0 0 1 9 】

このようなシステムでは、ユーザーの用意する未知の入力画像から所望の出力画像が自動的に生成されるように、使用開始前に、あるいは、随時、既知の入力画像とその既知の

10

20

30

40

50

入力画像に対応する出力画像との対である訓練データセットに基づいて、ディープニューラルネットワークなどの画像生成エンジンの学習が実行される。

【 0 0 2 0 】

この実施の形態に係る訓練データ生成システムを使用することで、原訓練データセット（原サンプル画像と原ラベル画像との対）から多数の擬似訓練データセット（擬似サンプル画像と擬似ラベル画像との対）が得られ、それらの原訓練データセットおよび擬似訓練データセットを使用した画像生成エンジンの学習が適切に実行される。

【 0 0 2 1 】

図 1 に示す訓練データ生成システムは、記憶装置 1、通信装置 2、画像読取装置 3、および演算処理装置 4 を備える。

【 0 0 2 2 】

記憶装置 1 は、フラッシュメモリー、ハードディスクなどの不揮発性の記憶装置であって、各種データやプログラムを格納する。

【 0 0 2 3 】

ここでは、記憶装置 1 には、訓練データ生成プログラム 1 1 が格納されており、また、生成された訓練データが必要に応じて格納される。なお、訓練データ生成プログラム 1 1 は、C D (Compact Disk) などの可搬性のあるコンピュータ読み取り可能な記録媒体に格納されていてもよい。その場合、例えば、その記録媒体から記憶装置 1 へ訓練データ生成プログラム 1 1 がインストールされる。また、訓練データ生成プログラム 1 1 は、1 つのプログラムでも、複数のプログラムの集合体でもよい。

【 0 0 2 4 】

通信装置 2 は、ネットワークインターフェイス、周辺機器インターフェイス、モデムなどのデータ通信可能な装置であって、必要に応じて、他の装置とデータ通信を行う。画像読取装置 3 は、入力画像としての原画像を含む写真などの物理媒体の原稿の画像を光学的に読み取り、原画像の画像データを生成する。通信装置 2 および画像読取装置 3 は、原訓練データセットの取得に使用される。

【 0 0 2 5 】

演算処理装置 4 は、C P U (Central Processing Unit)、R O M (Read Only Memory)、R A M (Random Access Memory) などを備えるコンピュータであって、プログラムを、R O M、記憶装置 1 などから R A M にロードし C P U で実行することで、各種処理部として動作する。

【 0 0 2 6 】

ここで、訓練データ生成プログラム 1 1 を実行することで、演算処理装置 4 は、原データ取得部 2 1、類似画像生成部 2 2、および類推画像生成部 2 3 として動作する。

【 0 0 2 7 】

原データ取得部 2 1 は、1 または複数の原訓練データセットを取得する。1 つの原訓練データセットは、入力画像から出力画像を生成する機械学習システム用の訓練データとして使用可能な、1 つの原サンプル画像と、その原サンプル画像に対応する原ラベル画像とを含む。原サンプル画像は、機械学習時に入力画像として使用可能な、実際に得られた写真などの画像であり、原ラベル画像は、機械学習時に出力画像として使用可能な、原サンプル画像に対して手作業などで生成された適正なラベル画像である。

【 0 0 2 8 】

原データ取得部 2 1 は、記憶装置 1 に格納されている原訓練データセットを読み出したり、通信装置 2 を使用して、外部の装置から原訓練データセットを受信したり、画像読取装置 3 を使用して、画像読取装置 3 により読み取られた画像を原訓練データセットとして取得したりする。なお、通信装置 2 および画像読取装置 3 を使用して取得された原訓練データセットは、記憶装置 1 に格納されるようにしてもよい。

【 0 0 2 9 】

類似画像生成部 2 2 は、取得された原ラベル画像の類似画像を、擬似ラベル画像として生成する。ここで、擬似ラベル画像は、機械学習時に機械学習システムの出力画像として

10

20

30

40

50

使用されるものである。

【 0 0 3 0 】

この実施の形態では、類似画像生成部 2 2 は、所定範囲（例えば - 1 ~ + 1）の値を有する 1 または複数の乱数値（乱数ベクトル）を入力値とし入力値に対応する画像を生成する画像生成エンジンを備え、複数の原ラベル画像を使用してその画像生成エンジンの機械学習を行い、その画像生成エンジンに別の乱数値を入力したときの出力画像を類似画像（つまり、擬似ラベル画像）とする。例えば、類似画像生成部 2 2 は、D C G A N S（Deep Convolutional Generative Adversarial Networks）などの G A N（Generative Adversarial Network）に基づいて、原ラベル画像の類似画像を、擬似ラベル画像として生成する。この類似画像の生成方法は、通常データ拡張（平行移動、鏡像反転、回転変動、幾何学的変形、濃度変更、色の変動、ランダムノイズの付加など）とは異なるものである。つまり、通常データ拡張は、何らかの規則に従って原画像を変換したものであるか、原画像を多少不規則に変化させたものであり、データ拡張で得られる画像には、1 枚の原画像に固有の規則性がある程度残されるが、この生成方法では、複数枚の原画像（原ラベル画像）の規則性が機械学習で抽出され、類似画像に反映されるため、個々の原画像に固有の規則性が、生成された類似画像に残りにくくなっている。

10

【 0 0 3 1 】

類似画像生成部 2 2 は、G A N で類似ラベル画像を生成する場合、ジェネレータおよびディスクリミネータを備え、ディスクリミネータおよびジェネレータの機械学習を行った後、ジェネレータで、乱数ベクトルから擬似ラベル画像を生成する。ディスクリミネータは、原ラベル画像とジェネレータにより生成された画像とを互いに区別するニューラルネットワークであり、ジェネレータは、乱数ベクトルから画像を生成するニューラルネットワークである。ディスクリミネータおよびジェネレータの機械学習では、( a ) 原ラベル画像に基づいて、原ラベル画像が原ラベル画像であると識別されるように、ディスクリミネータが機械学習され、( b ) 任意にサンプリングされた乱数ベクトルに対してジェネレータにより生成された画像について、当該画像が原ラベル画像であると識別されるようにジェネレータが機械学習されるとともに、当該画像がジェネレータにより生成された画像であると識別されるようにジェネレータが機械学習される。

20

【 0 0 3 2 】

なお、類似画像生成部 2 2 は、原ラベル画像で機械学習される別の画像生成エンジンで、原ラベル画像の類似画像を擬似ラベル画像として生成するようにしてもよい。

30

【 0 0 3 3 】

また、類似画像生成部 2 2 は、生成した擬似ラベル画像に対して、ノイズ除去処理を行うようにしてもよい。そのノイズ除去処理は、例えば、Non local Means Denoising algorithm などに従って実行される。

【 0 0 3 4 】

さらに、必要に応じて、類似画像生成部 2 2 は、生成した擬似ラベル画像に対して、通常データ拡張（平行移動、鏡像反転、回転変動、幾何学的変形、濃度変更、色の変動、ランダムノイズの付加など）を行って、擬似ラベル画像の数を増加させるようにしてもよい。

40

【 0 0 3 5 】

類推画像生成部 2 3 は、原ラベル画像から原サンプル画像への変換特性に従って擬似ラベル画像から類推される画像（類推画像）を、その擬似ラベル画像に対応する擬似サンプル画像として生成する。ここで、擬似サンプル画像は、機械学習時に機械学習システムの入力画像として使用されるものである。

【 0 0 3 6 】

例えば、類推画像生成部 2 3 は、擬似ラベル画像の各画素を注目画素として、注目画素の周辺領域（注目画素を含む所定サイズのウィンドウ）の画素値分布に最も近い画素値分布を周辺領域に有する画素を原ラベル画像内で探索し、発見した画素の位置と同一の位置の原サンプル画像の画素値を、擬似サンプル画像において同一位置の画素にセットするこ

50

とで、擬似サンプル画像を生成する。

【0037】

例えば、類推画像生成部23は、イメージアナロジー法 (Image analogies) に基づいて、原ラベル画像から原サンプル画像への変換特性に従って、擬似ラベル画像から類推される画像を、擬似サンプル画像として生成する。

【0038】

原訓練データセットが複数ある場合、類推画像生成部23は、複数の原訓練データセットから、1つの原訓練データセットを選択し、選択した原訓練データセットにおける原ラベル画像から原サンプル画像への変換特性に従って、擬似ラベル画像から類推される画像を、擬似サンプル画像として生成する。その際、原訓練データセットは、例えばランダムに、あるいは、所定の規則に従って選択される。

【0039】

また、原訓練データセットが複数ある場合、類推画像生成部23は、複数の原訓練データセットのそれぞれについて、選択した原訓練データセットにおける原ラベル画像から原サンプル画像への変換特性に従って、擬似ラベル画像から類推される画像を、擬似サンプル画像候補として生成し、複数の擬似サンプル画像候補から1つの擬似サンプル画像を選択したり、複数の擬似サンプル画像候補に対して所定の画像処理を行って1つの擬似サンプル画像を生成するようにしてもよい。複数の擬似サンプル画像候補から1つの擬似サンプル画像を選択する場合、例えばランダムに、あるいは、所定の規則に従って擬似サンプル画像が選択される。

【0040】

次に、上記システムの動作について説明する。図3は、図1に示すシステムにおける訓練データ生成方法について説明するフローチャートである。図4は、図1に示すシステムにおける原訓練データセットと擬似訓練データセットとの対応関係を説明する図である。

【0041】

まず、原データ取得部21が、上述のようにして1または複数の原訓練データセット (互いに対応する原サンプル画像と原ラベル画像との対) を取得する (ステップS1)。

【0042】

その後、類似画像生成部22は、取得された原訓練データセットにおける原ラベル画像から、上述のようにして、複数 (例えば数十万~数千万) の擬似ラベル画像を生成する (ステップS2)。例えば、類似画像生成部22は、GANを使用する場合、互いに異なる複数の乱数ベクトルを生成し、その乱数ベクトルから、乱数ベクトルと同数の擬似ラベル画像を得る。

【0043】

そして、類推画像生成部23は、取得された原訓練データセットにおける原サンプル画像および原ラベル画像の対から、上述のようにして、各擬似ラベル画像に対応する擬似サンプル画像を生成する (ステップS3)。平行移動などの従来のデータ拡張では、原サンプル画像を加工して、擬似的なサンプル画像を取得しているが、本実施の形態においては、原サンプル画像および原ラベル画像の両方の関係性に基づいて、擬似ラベル画像から、擬似サンプル画像が生成される。

【0044】

これにより、互いに対応する擬似サンプル画像および擬似ラベル画像の対 (つまり、擬似訓練データセット) が多数生成され、その後、例えば、記憶装置1に格納されたり、通信装置2で送信されたりする。そして、擬似訓練データセット (あるいは、原訓練データセットおよび擬似訓練データセット) に基づいて、多数の訓練データセットに基づく、ディープニューラルネットワークなどの画像生成エンジンの機械学習が行われる。

【0045】

ここで、原サンプル画像を細胞組織画像とし、原ラベル画像を細胞組織画像内の細胞の境界線を示す画像とした場合の具体例について説明する。

【0046】

10

20

30

40

50

図5は、原サンプル画像および原ラベル画像の具体例を説明する図である。

【0047】

図5(A)は、12枚の原サンプル画像としての細胞組織画像(グレースケール画像)を示しており、図5(B)は、12枚の原ラベル画像としての、図5(A)に示す細胞組織画像にそれぞれ対応する細胞境界をエッジ(黒白の境界)で示している2値画像である。

【0048】

このような、原ラベル画像は、原サンプル画像から手作業などで作成される。通常、多数(例えば10万枚)の原サンプル画像を入手することは困難であるとともに、原サンプル画像から原ラベル画像を作成する手間もかかるため、原訓練データセットのみで、有効な機械学習を行えるような数の訓練データセットを用意することは困難である。

【0049】

図6は、擬似ラベル画像の生成の具体例を説明する図である。

【0050】

図6(A)は、図5(B)に示した12枚の原ラベル画像を示している。図6(B)は、図6(A)に示す12枚の原ラベル画像から生成された6枚の擬似ラベル画像を示している。なお、ここでは、6枚の擬似ラベル画像を示しているが、上述のように、類似画像生成処理においては、画像生成エンジンの入力値(上述の乱数、つまり実数値)を変更することで、上限数なく多数の擬似ラベル画像が生成可能であるので、例えば、互いに異なる10万枚、1000万枚などの多数の擬似ラベル画像が生成可能である。図6(C)は、図6(B)に示す6枚の擬似ラベル画像に対してNon local Means Denoising algorithmに従ってノイズ除去処理を施したものである。なお、ノイズ除去処理は必要に応じて適用すればよく、別のアルゴリズムを使用してもよい。

【0051】

図7は、擬似サンプル画像の生成の具体例を説明する図である。図5に示す原訓練データセットのうちの一つに基づいて、図6(C)に示す擬似ラベル画像から、図7に示すような、擬似サンプル画像が得られた。

【0052】

次に、このようにして生成された擬似訓練データセットを使用した機械学習後の画像生成エンジンの出力画像の具体例について説明する。図8は、機械学習後の画像生成エンジンの出力画像の具体例を説明する図である。

【0053】

図8(A)は、テスト用細胞組織画像(つまり、訓練データセットには含まれていないもの)を示している。

【0054】

図8(B)は、3つの原訓練データセットに基づき生成された、上述のテスト用細胞組織画像に対応するラベル画像を示している。具体的には、3つの原訓練データセットに対して通常のデータ拡張を行って得られた訓練データセットで機械学習した画像生成エンジンで、図8(A)に示す細胞組織画像に対して、図8(B)に示すラベル画像が生成された。

【0055】

図8(C)は、12個の原訓練データセットに基づき生成された、上述のテスト用細胞組織画像に対応するラベル画像を示している。具体的には、12個の原訓練データセットに対して通常のデータ拡張を行って得られた訓練データセットで機械学習した画像生成エンジンで、図8(A)に示す細胞組織画像に対して、図8(C)に示すラベル画像が生成された。

【0056】

図8(D)は、12個の原訓練データセットおよび12個の擬似訓練データセットに基づき生成された、上述のテスト用細胞組織画像に対応するラベル画像を示している。具体的には、これらの訓練データセットに対して通常のデータ拡張を行って得られた訓練デー

10

20

30

40

50



タセットで機械学習した画像生成エンジンで、図 8 ( A ) に示す細胞組織画像に対して、図 8 ( D ) に示すラベル画像が生成された。

【 0 0 5 7 】

ラベル画像内での個々の細胞に対応する領域の分離状況などから、図 8 ( B ) および図 8 ( C ) に比べ、図 8 ( D ) に示すラベル画像が適切であることは明らかである。このように、通常の方法より、この実施の形態に係る訓練データ生成方法により訓練データを生成したほうが、適切なラベル画像が得られる。

【 0 0 5 8 】

以上のように、上記実施の形態によれば、類似画像生成部 2 2 は、原サンプル画像に対応する原ラベル画像の類似画像を、擬似ラベル画像として生成する。類推画像生成部 2 3 は、原ラベル画像から原サンプル画像への変換特性に従って擬似ラベル画像から類推される画像を、擬似ラベル画像に対応する擬似サンプル画像として生成する。ここで、擬似サンプル画像は、訓練データにおける機械学習システムの入力画像であり、擬似ラベル画像は、訓練データにおける機械学習システムの出力画像である。

【 0 0 5 9 】

これにより、原訓練データセットの原サンプル画像および原ラベル画像とはそれぞれ異なる新規な擬似サンプル画像および擬似ラベル画像の訓練データセットが自動的に多数生成できる。つまり、短時間で多くの適切な訓練データセットが自動的に生成される。

【 0 0 6 0 】

なお、上述の実施の形態に対する様々な変更および修正については、当業者には明らかである。そのような変更および修正は、その主題の趣旨および範囲から離れることなく、かつ、意図された利点を弱めることなく行われてもよい。つまり、そのような変更および修正が請求の範囲に含まれることを意図している。

【 0 0 6 1 】

例えば、上記実施の形態において、ラベル画像は、境界を示すものに限らず、特定の細胞（癌細胞など）を示すものとしてもよい。また、上記実施の形態において、原サンプル画像は、細胞組織画像以外の生体画像でもよいし、その他の撮影画像でもよい。

【 0 0 6 2 】

また、上記実施の形態において、上述の画像生成エンジンは、ディープニューラルネットワーク（つまり、隠れ層が 2 以上の階層型ニューラルネットワーク）に限定されない。

【 産業上の利用可能性 】

【 0 0 6 3 】

本発明は、例えば、機械学習システム用の訓練データの自動生成に適用可能である。

【 符号の説明 】

【 0 0 6 4 】

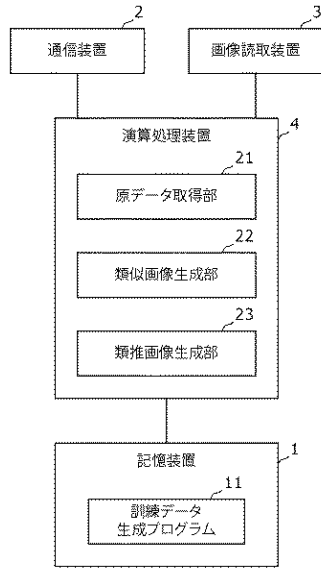
- 1 1 訓練データ生成プログラム
- 2 2 類似画像生成部
- 2 3 類推画像生成部

10

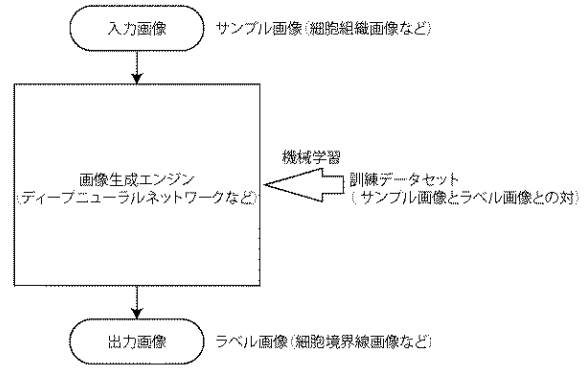
20

30

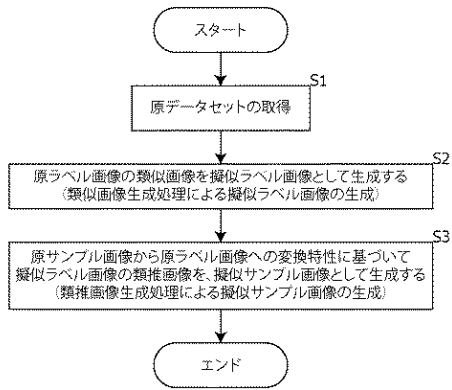
【図 1】



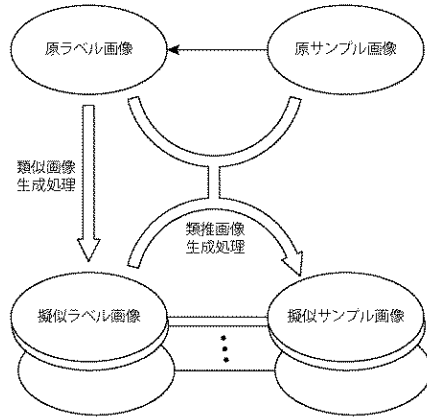
【図 2】



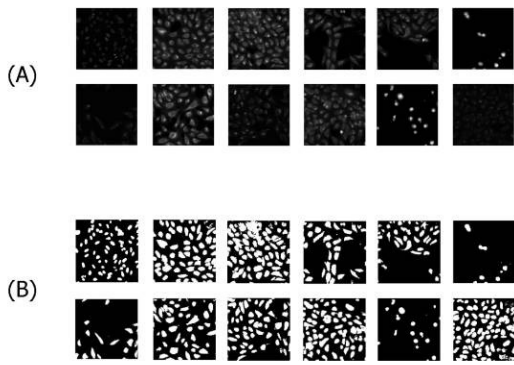
【図 3】



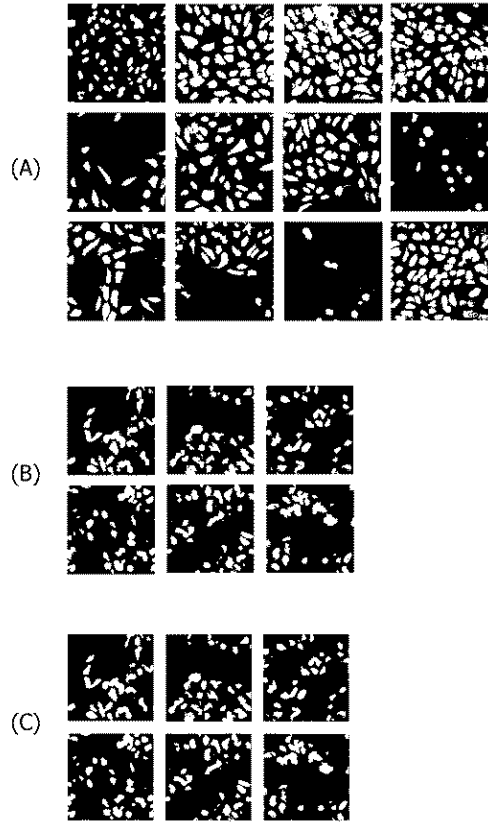
【図 4】



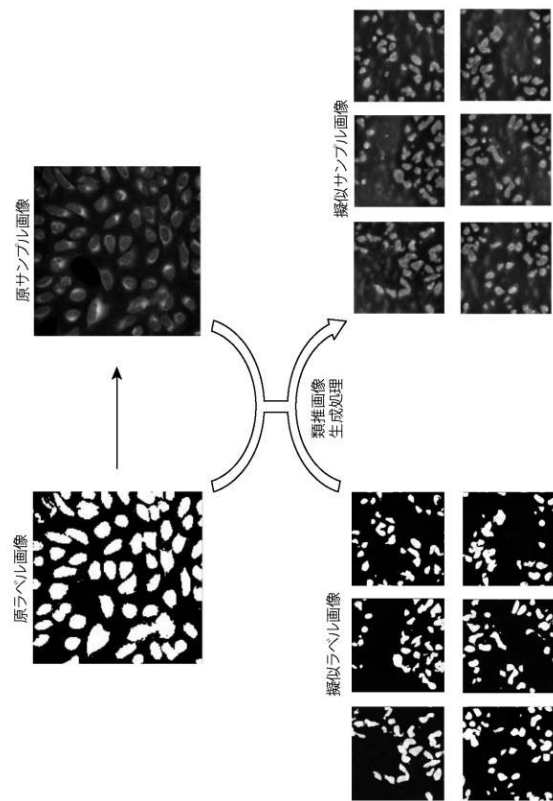
【 図 5 】



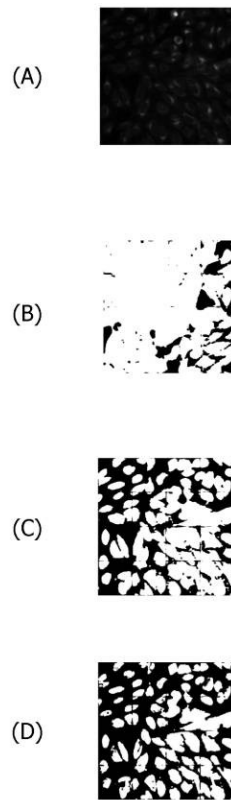
【 図 6 】



【 図 7 】



【 図 8 】



---

フロントページの続き

(56)参考文献 LIU, Shuangting ほか, PIXEL LEVEL DATA AUGMENTATION FOR SEMANTIC IMAGE SEGMENTATION USING GENERATIVE ADVERSARIAL NETWORKS, [online], 2019年11月26日, URL, <https://arxiv.org/pdf/1811.00174.pdf>

(58)調査した分野(Int.Cl., DB名)

G 0 6 N	2 0 / 0 0
G 0 6 N	3 / 0 0
G 0 6 T	7 / 0 0