(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2017-4350
(P2017-4350A)
(43)公開日　平成29年1月5日(2017.1.5)

(51)Int.Cl.		F I			テーマコード（参考）
G06T 7/00	(2017.01)	G06T	7/00	350C	5L096

審査請求 未請求　請求項の数 9　ＯＬ　（全 25 頁）

(21)出願番号　特願2015-119147 (P2015-119147)
(22)出願日　平成27年6月12日 (2015.6.12)

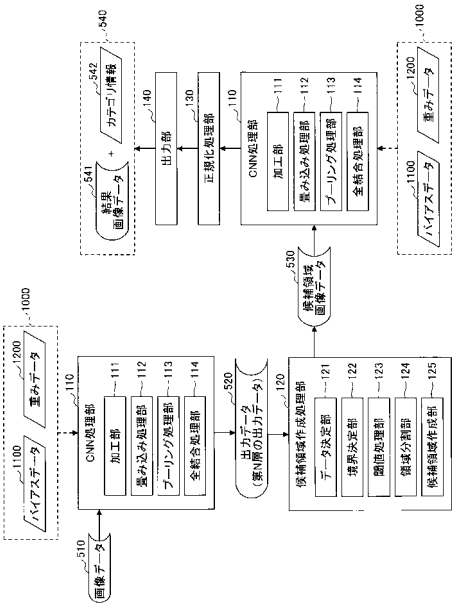
(71)出願人　000006747
株式会社リコー
東京都大田区中馬込１丁目３番６号
(74)代理人　100107766
弁理士　伊東　忠重
(74)代理人　100070150
弁理士　伊東　忠彦
(72)発明者　疋田　聡
東京都大田区中馬込１丁目３番６号　株式会社リコー内

Ｆターム(参考)　5L096 AA02　EA35　GA04　HA11　JA22

(54)【発明の名称】画像処理装置、画像処理方法、及びプログラム

(57)【要約】
【課題】認識処理の処理時間の削減を支援する。
【解決手段】画像データが示す画像において対象が含まれる第１の領域と、該対象が分類されるカテゴリとを認識する画像処理装置であって、畳み込みニューラルネットワークを用いて、入力された前記画像データが分類されるカテゴリを認識する認識手段と、前記認識手段における前記畳み込みニューラルネットワークの所定の層の出力結果を示す第１の出力データに基づいて、前記画像データが示す画像に含まれる１以上の候補領域を示す１以上の候補領域画像データを作成する候補領域作成手段とを有し、前記認識手段は、前記候補領域作成手段により作成された前記１以上の候補領域画像データがそれぞれ分類されるカテゴリを認識する。
【選択図】図２

本実施形態の画像処理装置の機能構成の一例を示す図



【特許請求の範囲】
【請求項１】
画像データが示す画像において対象が含まれる第１の領域と、該対象が分類されるカテゴリとを認識する画像処理装置であって、
畳み込みニューラルネットワークを用いて、入力された前記画像データが分類されるカテゴリを認識する認識手段と、
前記認識手段における前記畳み込みニューラルネットワークの所定の層の出力結果を示す第１の出力データに基づいて、前記画像データが示す画像に含まれる１以上の候補領域を示す１以上の候補領域画像データを作成する候補領域作成手段と
を有し、
前記認識手段は、
前記候補領域作成手段により作成された前記１以上の候補領域画像データがそれぞれ分類されるカテゴリを認識する、画像処理装置。
【請求項２】
前記第１の出力データは、前記畳み込みニューラルネットワークの前記所定の層のネットワークパラメータから特定されるフィルタ毎の第２の出力データを含み、
前記第１の出力データから所定の個数の前記第２の出力データを決定する決定手段を有し、
前記候補領域作成手段は、
前記決定手段で決定された前記第２の出力データに基づいて、前記１以上の候補領域データを作成する、請求項１記載の画像処理装置。
【請求項３】
前記決定手段は、
前記第２の出力データの代表データ値の昇順に、前記所定の個数の前記第２の出力データを決定する、請求項２記載の画像処理装置。」
【請求項４】
前記第２の出力データが示す画像を１以上の第２の領域に分割する分割手段を有し、
前記候補領域作成手段は、
前記分割手段により分割された前記１以上の第２の領域のそれぞれについて、該第２の領域を囲む最小の矩形領域を前記候補領域とする、請求項２又は３に記載の画像処理装置。
【請求項５】
前記分割手段は、
微分処理により前記１以上の第２の領域の境界を検出し、該検出された境界に基づいて分割する、請求項４記載の画像処理装置。
【請求項６】
前記分割手段は、
前記微分処理にSobelフィルタを用いる、請求項５記載の画像処理装置。
【請求項７】
所定の閾値以下のデータ値を削除する閾値手段を有し、
前記分割手段は、
前記閾値手段により所定の閾値以下のデータ値を削除した前記第２の出力データが示す画像を１以上の領域に分割する、請求項４ないし６のいずれか１項に記載の画像処理装置。
【請求項８】
画像データが示す画像において対象が含まれる第１の領域と、該対象が分類されるカテゴリとを認識する画像処理装置による画像処理方法であって、
畳み込みニューラルネットワークを用いて、入力された前記画像データが分類されるカテゴリを認識する認識手順と、
前記認識手順における前記畳み込みニューラルネットワークの所定の層の出力結果を示

す第１の出力データに基づいて、前記画像データが示す画像に含まれる１以上の候補領域を示す１以上の候補領域画像データを作成する候補領域作成手順と
を有し、
前記認識手順は、
前記候補領域作成手順により作成された前記１以上の候補領域画像データがそれぞれ分類されるカテゴリを認識する、画像処理方法。
【請求項９】
画像データが示す画像において対象が含まれる第１の領域と、該対象が分類されるカテゴリとを認識する画像処理装置を、
畳み込みニューラルネットワークを用いて、入力された前記画像データが分類されるカテゴリを認識する認識手段、
前記認識手段における前記畳み込みニューラルネットワークの所定の層の出力結果を示す第１の出力データに基づいて、前記画像データが示す画像に含まれる１以上の候補領域を示す１以上の候補領域画像データを作成する候補領域作成手段
として機能させ、
前記認識手段は、
前記候補領域作成手段により作成された前記１以上の候補領域画像データがそれぞれ分類されるカテゴリを認識する、プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、画像処理装置、画像処理方法、及びプログラムに関する。
【背景技術】
【０００２】
デジタルカメラや携帯情報端末等の機器において、撮影された画像中の被写体が属するカテゴリ（例えば、「人」、「動物」、「車」等）を分類する技術が知られている。
【０００３】
また、画像中において、被写体が占める領域と、当該被写体が分類されるカテゴリとを認識する技術が知られている（例えば特許文献１及び非特許文献１参照）。このような技術では、被写体が占める領域の候補である候補領域に対して、カテゴリを分類するための処理を行うことで、被写体が占める領域と、当該被写体が分類されるカテゴリとを認識する。
【発明の概要】
【発明が解決しようとする課題】
【０００４】
しかしながら上記の従来技術では、被写体が占める領域と、当該被写体が分類されるカテゴリとの認識処理に多くの時間を要する場合があった。例えば、候補領域の数が多い場合には、それぞれの候補領域に対してカテゴリを分類するための処理を行うため、認識処理に多くの時間を要することがある。
【０００５】
本発明の実施形態は、認識処理の処理時間の削減を支援することを目的とする。
【課題を解決するための手段】
【０００６】
上記目的を達成するため、本発明の実施の形態では、画像データが示す画像において対象が含まれる第１の領域と、該対象が分類されるカテゴリとを認識する画像処理装置であって、畳み込みニューラルネットワークを用いて、入力された前記画像データが分類されるカテゴリを認識する認識手段と、前記認識手段における前記畳み込みニューラルネットワークの所定の層の出力結果を示す第１の出力データに基づいて、前記画像データが示す画像に含まれる１以上の候補領域を示す１以上の候補領域画像データを作成する候補領域作成手段とを有し、前記認識手段は、前記候補領域作成手段により作成された前記１以上

の候補領域画像データがそれぞれ分類されるカテゴリを認識する。
【発明の効果】
【０００７】
本発明の実施形態によれば、認識処理の処理時間の削減を支援することができる。
【図面の簡単な説明】
【０００８】
【図１】本実施形態の画像処理装置のハードウェア構成の一例を示す図である。
【図２】本実施形態の画像処理装置の機能構成の一例を示す図である。
【図３】本実施形態の画像処理装置の認識処理のフローチャートの一例を示す図である。
【図４】本実施形態の畳み込みニューラルネットワーク処理のフローチャートの一例を示す図である。
【図５】本実施形態の入力画像データの加工処理の一例を示す図である。
【図６】本実施形態の第１層の畳み込み処理の一例を示す図である。
【図７】本実施形態の第１層のネットワークパラメータの一例を示す図である。
【図８】本実施形態の第１層のフィルタの一例を示す図である。
【図９】本実施形態の第１層のプーリング処理の一例を示す図である。
【図１０】本実施形態の第２層の畳み込み処理の一例を示す図である。
【図１１】本実施形態の第２層のネットワークパラメータの一例を示す図である。
【図１２】本実施形態の第２層のフィルタの一例を示す図である。
【図１３】本実施形態の候補領域の作成処理のフローチャートの一例を示す図である。
【図１４】本実施形態の微分処理の一例を示す図である。
【図１５】本実施形態の閾値処理の一例を示す図である。
【図１６】本実施形態の領域分割の一例を示す図である。
【図１７】本実施形態の最小矩形の一例を示す図である。
【図１８】本実施形態のカテゴリ分類処理のフローチャートの一例を示す図である。
【図１９】本実施形態の第３層の全結合処理の一例を示す図である。
【図２０】本実施形態の第３層のネットワークパラメータの一例を示す図である。
【図２１】本実施形態の正規化処理の一例を示す図である。
【発明を実施するための形態】
【０００９】
本実施形態は、画像データが示す画像において、当該画像の被写体を示す対象（例えば、人や物体等）を含む領域と、当該対象が分類されるカテゴリとを認識するものである。ここで、カテゴリとは、例えば、「人」、「動物」、「車」、「花」、「料理」等の対象が分類される種別のことである。
【００１０】
以降では、画像データに対して、上述した認識を行う処理（認識処理）を実行する画像処理装置１０について説明する。なお、本実施形態の画像処理装置１０は、例えば、デジタルカメラ、スマートフォン、タブレット端末、ゲーム機器、ノート型ＰＣ、デスクトップ型ＰＣ等である。
【００１１】
＜ハードウェア構成＞
まず、本実施形態の画像処理装置１０のハードウェア構成について、図１を参照しながら説明する。図１は、本実施形態の画像処理装置のハードウェア構成の一例を示す図である。
【００１２】
本実施形態の画像処理装置１０は、入力装置１１、表示装置１２、ＣＰＵ（Central Processing Unit）１３、及びＲＯＭ（Read Only Memory）１４を有する。また、本実施形態の画像処理装置１０は、ＲＡＭ（Random Access Memory）１５、インタフェース装置１６、記憶装置１７、及び撮像装置１８を有する。これら各ハードウェアは、バスBにより相互に接続されている。

【００１３】
入力装置１１は、キーボードやマウス、タッチパネル、各種ボタン等を含み、画像処理装置１０に各種信号を入力するのに用いられる。表示装置１２は、ディスプレイ等を含み、各種の処理結果を表示する。特に、表示装置１２には、本実施形態の認識処理の処理結果が表示される。すなわち、表示装置１２には、入力された画像データが示す画像において、被写体等の対象が含まれる領域と、当該対象が分類されるカテゴリと示す処理結果が表示される。
【００１４】
ＣＰＵ１３は、例えば記憶装置１７やＲＯＭ１４等からプログラムやデータをＲＡＭ１５上に読み出して、各種処理を実行する演算装置である。ＲＯＭ１４は、電源を切ってもデータを保持することができる不揮発性の半導体メモリである。ＲＡＭ１５は、プログラムやデータを一時保存することができる揮発性の半導体メモリである。
【００１５】
インタフェース装置１６は、外部装置とのインタフェースである。外部装置には、例えば、ＣＤ（Compact Disk）やＤＶＤ（Digital Versatile Disk）、ＳＤメモリカード(SD memory card）、ＵＳＢメモリ（Universal Serial Bus memory）等の記録媒体がある。画像処理装置１０は、インタフェース装置１６を介して、本実施形態の認識処理の処理対象となる画像データを記録媒体から読み取ることができる。
【００１６】
記憶装置１７は、プログラムやデータを格納しているＨＤＤ（Hard Disk Drive）やＳＳＤ（Solid State Drive）等の不揮発性のメモリである。記憶装置１７に格納されるプログラムやデータには、本実施形態の認識処理を実行する画像処理プログラム２０がある。また、本実施形態の認識処理の処理対象となる画像データが格納されても良い。
【００１７】
撮像装置１８は、カメラ等であり、本実施形態の認識処理の処理対象となる画像データを作成する。
【００１８】
本実施形態の画像処理装置１０は、上記ハードウェア構成により後述する各種処理を実現することができる。
【００１９】
＜機能構成＞
次に、本実施形態の画像処理装置１０の機能構成について、図２を参照しながら説明する。図２は、本実施形態の画像処理装置の機能構成の一例を示す図である。
【００２０】
本実施形態の画像処理装置１０は、ＣＮＮ処理部１１０、候補領域作成処理部１２０、正規化処理部１３０、及び出力部１４０を有する。これら各部は、画像処理装置１０にインストールされた画像処理プログラム２０が、ＣＰＵ１３に実行させる処理により実現される。
【００２１】
ＣＮＮ処理部１１０は、ネットワークパラメータ１０００に基づいて、畳み込みニューラルネットワーク（ＣＮＮ：Convolutional Neural Network）処理を行う。畳み込みニューラルネットワークは、一般に、ｎを３以上の任意の自然数として、畳み込み処理及びプーリング処理を行う第１層～第ｎ－２層と、畳み込み処理を行う第ｎ－１層と、全結合処理を行う第ｎ層とを含む。
【００２２】
ここで、ネットワークパラメータ１０００は、教師あり学習の手法により、学習データに基づいて畳み込みニューラルネットワークの各層毎に予め学習されたデータである。教師あり学習の手法には、例えば誤差逆伝播法（Backpropagation）を用いれば良い。
【００２３】
このようなネットワークパラメータ１０００は、例えば記憶装置１７等に格納され、バ

イアスデータ１１００及び重みデータ１２００が含まれる。なお、以降では、第ｎ層のネットワークパラメータ１０００を「ネットワークパラメータ１０００－ｎ」と表す。したがって、第ｎ層のバイアスデータ１１００及び重みデータ１２００はそれぞれ「バイアスデータ１１００－ｎ」及び「重みデータ１２００－ｎ」と表される。ネットワークパラメータ１０００の詳細については後述する。

【００２４】

ＣＮＮ処理部１１０は、入力画像を示す画像データ５１０に対して、畳み込みニューラルネットワーク処理を行い、予め設定された第Ｎ層における畳み込み処理の処理結果を示す出力データ５２０を出力する。

【００２５】

ここで、本実施形態では、Ｎ＝２であるものとして説明する。Ｎ＝２の場合、出力データ５２０は、例えば、２８×２８×６４チャンネルの画像データとして表すことができる。換言すれば、出力データ５２０は、６４個の２８×２８チャンネルの画像データの集合として表すことができる。なお、Ｎの値は、画像処理プログラム２０の設計者等により予め設定される。Ｎの値は、例えば、２～２０程度が好ましい。

【００２６】

また、ＣＮＮ処理部１１０は、後述する候補領域作成処理部１２０により作成された候補領域画像データ５３０に対して、畳み込みニューラルネットワーク処理を行い、出力結果を正規化処理部１３０に出力する。

【００２７】

さらに、ＣＮＮ処理部１１０は、加工部１１１、畳み込み処理部１１２、プーリング処理部１１３、及び全結合処理部１１４を有する。加工部１１１は、ＣＮＮ処理部１１０に入力された画像データの加工処理を行う。畳み込み処理部１１２は、畳み込みニューラルネットワークの各層において畳み込み処理を行う。プーリング処理部１１３は、畳み込みニューラルネットワークの各層においてプーリング処理を行う。全結合処理部１１４は、全結合処理を行う。

【００２８】

ここで、ＣＮＮ処理部１１０は、全結合処理部１１４をカテゴリの組毎に有しているものとする。カテゴリの組とは、カテゴリと、当該カテゴリ以外を示すカテゴリとのペアである。具体的には、カテゴリの組は、「人」「人以外」、「車」「車以外」、「動物」「動物以外」等の、あるカテゴリと、当該カテゴリ以外を示すカテゴリとのペアである。なお、以降では、複数の全結合処理部１１４を区別して表す場合は、「全結合処理部１１４－１」、「全結合処理部１１４－２」等と表す。

【００２９】

候補領域作成処理部１２０は、出力データ５２０に基づいて、１以上の候補領域画像データ５３０を作成する。候補領域画像データ１２０とは、画像データ５１０が示す画像において、対象が含まれる領域の候補を示すデータである。なお、以降では、複数の候補領域画像データ５３０を区別して表す場合は、「候補領域画像データ５３０－１」、「候補領域画像データ５３０－２」等と表す。

【００３０】

ここで、候補領域作成処理部１２０は、データ決定部１２１、境界決定部１２２、閾値処理部１２３、領域分割部１２４、及び候補領域作成部１２５を有する。

【００３１】

データ決定部１２１は、例えば６４個の２８×２８チャンネルのデータとして表される出力データ５２０から所定のＭ個の２８×２８チャンネルのデータを決定する。ここで、Ｍの値は、画像処理プログラム２０の設計者等により予め設定される。Ｍの値は、例えば、３～２０程度が好ましい。

【００３２】

境界決定部１２２は、データ決定部１２１により決定されたそれぞれのデータに対して、所定の微分処理を行い、領域分割部１２４が分割する領域の境界を決定する。

【００３３】
閾値処理部１２３は、閾値処理を行う。閾値処理とは、予め設定された閾値以下のデータを削除（すなわち、「０」とする）する処理である。なお、このような閾値は、画像処理プログラム２０の設計者等により予め設定される。閾値の値は、例えば、１０～５０程度が好ましい。
【００３４】
領域分割部１２４は、境界決定部１２２により決定された境界に基づいて、データ決定部１２１により決定されたデータが示す画像を、複数の領域に分割する。
【００３５】
候補領域作成部１２５は、領域分割部１２４により分割された複数の領域に基づいて、候補領域を作成し、作成した候補領域を示す候補領域画像データ５３０を出力する。
【００３６】
例えば、候補領域作成部１２５は、領域分割部１２４により分割された複数の領域のうちの一の領域に基づいて、候補領域画像データ５３０－１を出力する。同様に、候補領域作成部１２５は、領域分割部１２４により分割された複数の領域のうちの他の領域に基づいて、候補領域画像データ５３０－２を出力する。
【００３７】
このように、本実施形態の候補領域作成処理部１２０は、出力データ５２０に基づいて候補領域画像データ５３０を作成する。これにより、本実施形態では、認識処理の精度の低下を防ぎつつ、候補領域を削減させることができる。したがって、本実施形態では、認識処理の処理時間を削減させることができる。
【００３８】
正規化処理部１３０は、ＣＮＮ処理部１１０による処理結果を正規化する。ＣＮＮ処理部１１０の各全結合処理部１１４による処理結果を比較することができる。以降では、正規化処理部１３０により正規化された、全結合処理部１１４の処理結果を「確信度」と表す。
【００３９】
例えば、カテゴリの組「人」「人以外」に対応する全結合処理部１１４の確信度は、ＣＮＮ処理部１１０に入力された画像データが示す画像が、カテゴリ「人」に分類される度合いを示す第１の値と、カテゴリ「人以外」に分類される度合いを示す第２の値との組で表される。
【００４０】
同様に、カテゴリの組「車」「車以外」に対応する全結合処理部１１４の確信度は、ＣＮＮ処理部１１０に入力された画像データが示す画像が、カテゴリ「車」に分類される度合いを示す第１の値と、カテゴリ「人以外」に分類される度合いを示す第２の値との組で表される。
【００４１】
出力部１４０は、認識結果５４０を出力する。ここで、認識結果５４０には、候補領域画像データ５３０から選択された結果画像データ５４１と、当該結果画像データ５４１のカテゴリを示すカテゴリ情報５４２とが含まれる。なお、出力部１４０は、候補領域画像データ５３０の確信度に基づいて、当該候補領域画像データ５３０から結果画像データ５４１を選択するとともに、当該結果画像データ５４１のカテゴリを決定してカテゴリ情報５４２を作成する。
【００４２】
これにより、画像データ５１０が示す画像において、対象が含まれる領域の画像と、当該対象が分類されるカテゴリとが出力される。
【００４３】
＜処理の詳細＞
次に、本実施形態の画像処理装置１０の認識処理の詳細について、図３を参照しながら説明する。図３は、本実施形態の画像処理装置の認識処理のフローチャートの一例を示す

図である。
【００４４】
画像処理装置１０は、画像データ５１０を入力する（ステップＳ３１）。画像処理装置１０は、例えば、記憶装置１７に格納されている画像データ５１０を入力しても良いし、撮像装置１８により生成された画像データ５１０を入力しても良い。また、画像処理装置１０は、例えば、ネットワーク経由でダウンロードした画像データ５１０を入力しても良い。
【００４５】
画像処理装置１０は、ＣＮＮ処理部１１０により、入力された画像データ５１０に対して、予め設定された第Ｎ層の畳み込み処理までの畳み込みニューラルネットワーク処理を行う（ステップＳ３２）。この畳み込みニューラルネットワーク処理についての詳細については、後述する。ここでは、本ステップの畳み込みニューラルネットワーク処理において、第Ｎ層の畳み込み処理の処理結果を示す出力データ５２０が得られたものとして説明を続ける。
【００４６】
なお、上述したように、Ｎ＝３である場合、出力データ５２０は、例えば６４個の２８×２８チャンネルのデータとして表される。
【００４７】
画像処理装置１０は、候補領域作成処理部１２０により、出力データ５２０を入力して候補領域の作成処理を行う（ステップＳ３３）。この候補領域の作成処理において、候補領域作成処理部１２０は、出力データ５２０に基づいて、１以上の候補領域画像データ５３０を作成する。この候補領域の作成処理の詳細については、後述する。ここでは、本ステップの候補領域の作成処理において、１以上の候補領域画像データ５３０が得られたものとして説明を続ける。
【００４８】
画像処理装置１０は、ＣＮＮ処理部１１０及び正規化処理部１３０により、一の候補領域画像データ５３０を入力し、当該一の候補領域画像データ５３０のカテゴリを分類するカテゴリ分類処理を行う（ステップＳ３４）。このカテゴリ分類処理により、入力された一の候補領域画像データ５３０の確信度が得られる。このカテゴリ分類処理の詳細については、後述する。ここでは、本ステップのカテゴリ分類処理において、一の候補領域画像データ５３０の確信度が得られたものとして説明を続ける。
【００４９】
画像処理装置１０は、ＣＮＮ処理部１１０及び正規化処理部１３０により、すべての候補領域画像データ５３０の確信度が得られたか否かを判定する（ステップＳ３５）。確信度が得られていない（すなわち、カテゴリ分類処理を行っていない）候補領域画像データ５３０が存在する場合には、ステップＳ３４に戻る。すなわち、画像処理装置１０は、候補領域画像データ５３０－１、候補領域画像データ５３０－２、・・・等に対して、それぞれの確信度を順に取得する。
【００５０】
一方、すべての候補領域画像データ５３０の確信度が得られた場合には、ステップＳ３６に進む。
【００５１】
画像処理装置１０は、出力部１４０により、得られた確信度に基づいて候補領域画像データ５３０から結果画像データ５４１を選択するとともに、当該結果画像データ５４１のカテゴリを決定してカテゴリ情報５４２を作成する。（ステップＳ３６）。すなわち、出力部１４０は、認識結果５４０を決定する。
【００５２】
出力部１４０は、すべての候補領域画像データ５３０を結果画像データ５４１と選択しても良いし、候補領域画像データ５３０のうちの一部を結果画像データ５４１と選択しても良い。

【００５３】
また、出力部１４０は、例えば、候補領域画像データ５３０が示す画像のうち、一部が重畳している画像が存在する場合に、当該重畳している画像が示す候補領域画像データ５３０のうち、最も確信度が高い候補領域画像データ５３０を結果画像データ５４１と選択しても良い。より具体的には、例えば、候補領域画像データ５３０－１が示す第１の画像と、候補領域画像データ５３０－２が示す第２の画像と、候補領域画像データ５３０－３が示す第３の画像とが、少なくとも一部の領域において重畳しているものとする。この場合、第１の画像の確信度の第１の値と、第２の画像の確信度の第１の値と、第３の画像の確信度の第１の値とを比較し、最も値が高い画像を示す候補領域画像データ５３０を結果画像データ５４１と選択すれば良い。

【００５４】
なお、ステップＳ３６において、出力部１４０は、２以上の認識結果５４０を決定しても良い。すなわち、出力部１４０は、候補領域画像データ５３０から２以上の結果画像データ５４１を選択するとともに、当該２以上の結果画像データ５４１のそれぞれのカテゴリ情報５４２を作成しても良い。これにより、例えば、画像データ５１０が示す画像において、複数の対象（例えば、「人」と「車」等）が写っている場合にも、それぞれの対象が含まれる領域の画像と、それぞれの対象が分類されるカテゴリとを決定することができる。

【００５５】
画像処理装置１０は、出力部１４０により、決定された認識結果５４０を出力する（ステップＳ３７）。このとき、出力部１４０は、例えば表示装置１２に認識結果５４０を出力すれば良い。これにより、画像データ５１０が示す画像において、対象が含まれる領域の画像と、当該対象が分類されるカテゴリとが表示装置１２に表示される。

【００５６】
次に、図３のステップＳ３２の畳み込みニューラルネットワーク処理について、図４を参照しながら説明する。図４は、本実施形態の畳み込みニューラルネットワーク処理のフローチャートの一例を示す図である。

【００５７】
加工部１１１は、入力された画像データ５１０の加工処理を行う（ステップＳ４１）。この加工処理は、入力された画像データ５１０を、畳み込み処理部１１２が処理可能な形式とするための処理である。

【００５８】
ここで、加工処理について、図５を参照しながら説明する。図５は、本実施形態の入力画像データの加工処理の一例を示す図である。なお、入力された画像データ５１０の色空間がＲＧＢ色空間である（すなわち、画像データ５１０の色チャンネルが３チャンネルである）ものとして説明する。ただし、画像データ５１０の色空間は、ＲＧＢ色空間に限られず、例えば、ＣＭＫ色空間、ＨＳＶ色空間、ＨＬＳ色空間等であっても良い。

【００５９】
Ｓｔｅｐ４１１）加工部１１１は、入力された画像データ５１０を６４×６４（ピクセル）となるように縮小する、このとき、加工部１１１は、画像データ５１０の長辺が６４（ピクセル）となるように縮小を行う。また、加工部１１１は、短辺が縮小された結果６４（ピクセル）に満たない部分については値０（すなわち、ＲＧＢの各色成分の値が０）でパディングして６４（ピクセル）とする。なお、画像データ５１０を縮小するためのアルゴリズムには、例えば、バイリニア法を用いれば良い。

【００６０】
Ｓｔｅｐ４１２）加工部１１１は、ＳｔｅｐＳ４１１で得られた６４×６４の画像データの各画素値から、所定の値を減算した画像データを生成する。

【００６１】
ここで、所定の値は、各学習データに含まれる画像データ（以降、「学習画像データ」という）の各画素値の平均値である。すなわち、学習画像データの画素位置（ｉ，ｊ）に

おける各学習画像データの画素値の平均値をＭ（ｉ，ｊ）とした場合、上記のＳｔｅｐ４１１において得られた６４×６４の画像データの各画素位置（ｉ，ｊ）の画素値からＭ（ｉ，ｊ）を減算する。ここで、ｉ，ｊ＝１，・・・，６４である。

【００６２】

Ｓｔｅｐ４１３）加工部１１１は、Ｓｔｅｐ４１２で得られた画像データの中心の５６×５６（ピクセル）の画像データ以外を０クリアする。換言すれば、Ｓｔｅｐ４１２において得られた画像データの周辺４ピクセル分を０クリアする。なお、図５において、網掛け部分が０クリアした部分である。

【００６３】

そして、加工部１１１は、図５のＳｔｅｐ４１３で得られた６４×６４（ピクセル）の画像データ（この画像データを「画像データ５１１」とする。）を畳み込み処理部１１２に出力する。

【００６４】

ＣＮＮ処理部１１０は、畳み込みニューラルネットワークの層を示す変数ｎを１とする（ステップＳ４２）。

【００６５】

畳み込み処理部１１２は、画像データ５１１を入力して、第１層の畳み込み処理を行う（ステップＳ４３）。

【００６６】

ここで、第１層の畳み込み処理について、図６を参照しながら説明する。図６は、本実施形態の第１層の畳み込み処理の一例を示す図である。

【００６７】

Ｓｔｅｐ４３１）畳み込み処理部１１２は、画像データ５１１を入力する。ここで、入力した画像データ５１１の色空間はＲＧＢ色空間であるため、色チャンネルは６４×６４×３チャンネルである。

【００６８】

Ｓｔｅｐ４３２）畳み込み処理部１１２は、重みデータ１２００－１からフィルタを生成し、画像データ５１１の中心の５６×５６（ピクセル）の部分に対して、生成したフィルタを用いてフィルタ処理を行う。ここで、重みデータ１２００－１のデータ構成及び当該重みデータ１２００－１から生成されるフィルタ$1300f_j-1$（ｊ＝１，・・・，６４）のデータ構成について説明する。

【００６９】

図７（ｂ）は、第１層の重みデータ１２００－１の一例を示す図である。図７（ｂ）に示すように、第１層の重みデータ１２００－１は、７５×６４の行列で表される。なお、重みデータ１２００－１の各値w_1（ｉ，ｊ）は、上述したように、学習データに基づいて予め学習された値である。

【００７０】

次に、重みデータ１２００－１から生成されるフィルタ$1300f_j-1$（ｊ＝１，・・・，６４）について説明する。図８は、本実施形態の第１層のフィルタの一例を示す図である。

【００７１】

図８に示すように、各フィルタ$1300f_j-1$（ｊ＝１，・・・，６４）は、５×５の行列の３つの組で表される。換言すれば、各フィルタ$1300f_j-1$（ｊ＝１，・・・，６４）は、５×５×３で表される。

【００７２】

ここで、重みデータ１２００－１のw_1（１，１）～w_1（２５，１）、w_1（２６，１）～w_1（５０，１）、及びw_1（５１，１）～w_1（７５，１）からフィルタ$1300f_1-1$が生成される。同様に、重みデータ１２００－１のw_1（１，２）～w_1（２５，２）、w_1（２６，２）～w_1（５０，２）、及びw_1（５１，２）～w_1（７５，２）からフィルタ$1300f_2-1$が生成される。ｊ＝３，・・・，６４の場合も同様で

ある。
【００７３】
以上のように生成された各フィルタ１３００ｆ$_{j}$－１（ｊ＝１，・・・，６４）を用いて、畳み込み処理部１１２は、画像データ５１１に対してフィルタ処理を行う。畳み込み処理部１１２は、例えば以下のようにしてフィルタ処理を行う。
【００７４】
（１）画像データ５１１の中心５６×５６×３の部分に対してフィルタ１３００ｆ$_{1}$－１をかける（すなわち、画像データ５１１とフィルタ１３００ｆ$_{1}$－１の対応する値の乗算を行う）。
【００７５】
これは、例えば、Ｒチャンネルを固定し、フィルタ１３００ｆ$_{1}$－１のＲチャンネル用フィルタの中心を、画像データ５１１のＲチャンネルの５６×５６の部分に対して、左上から５ずつ右にずらしながら行う。そして、フィルタ１３００ｆ$_{1}$－１のＲチャンネル用フィルタの中心が画像データ５１１のＲチャンネルの５６×５６の部分の右端まで辿り着いたら、当該Ｒチャンネル用フィルタの中心を下に５ずらして、再度、左端から行えば良い。
【００７６】
（２）次に、画像データ５１１のＧチャンネルに対しても、上記（１）と同様の方法でフィルタ１３００ｆ$_{1}$－１のＧチャンネル用フィルタをかける。画像データ５１１のＢチャンネルに対しても同様である。
【００７７】
（３）フィルタ１３００ｆ$_{2}$－１～フィルタ１３００ｆ$_{64}$－１についても、上記と同様に、画像データ５１１のＲＧＢの各チャンネルに対してフィルタ処理を順に行う。
【００７８】
以上のフィルタ処理により、画像データ５１１から６４×６４×３×６４チャンネルの画像データが生成される。
【００７９】
Ｓｔｅｐ４３３）畳み込み処理部１１２は、Ｓｔｅｐ４３２で得られた６４×６４×３×６４チャンネルの画像データの各ＲＧＢ成分を加算する。この結果、６４×６４×６４チャンネルの画像データが得られる。
【００８０】
Ｓｔｅｐ４３４）畳み込み処理部１１２は、Ｓｔｅｐ４３３で得られた６４×６４×６４チャンネルの画像データの各画素値に対して、バイアスデータ１１００－１を加算する。
【００８１】
ここで、図７（ａ）は、第１層のバイアスデータ１１００－１の一例を示す図である。図７（ａ）に示すように、バイアスデータ１１００－１は、１×６４の行列により表される。そこで、畳み込み処理部１１２は、１つめの６４×６４チャンネルの画像データの各画素値に対してバイアスデータ１１００－１のデータ値ｂ$_{1}$（１）を加算する。同様に、２つ目の６４×６４チャンネルの画像データの各画素値に対してバイアスデータ１１００－１のデータ値ｂ$_{1}$（２）を加算する。以降、同様に、６４個すべての６４×６４チャンネルの画像データの各画素値に対して、それぞれ、バイアスデータ１１００－１のデータ値を加算する。
【００８２】
Ｓｔｅｐ４３５）畳み込み処理部１１２は、Ｓｔｅｐ４３４で得られた６４×６４×６４チャンネルの画像データに対して、所定の活性化関数を適用して出力画像データを得る。所定の活性化関数としては、例えば、任意の画素値ｘに対して、ｆ（ｘ）＝ｍａｘ（０，ｘ）で定義される関数が挙げられる。
【００８３】
そして、６４×６４×６４チャンネルの画像データに対して、活性化関数を適用した後

、ステップＳ４１の加工処理において０クリアした部分は取り除き、画像データの中心の５６×５６部分をプーリング処理部１１３に出力する。したがって、第１層において、畳み込み処理部１１２がプーリング処理部１１３に出力する画像データの色チャンネルは、５６×５６×６４である。このようにして得られた５６×５６×６４チャンネルの画像データを「画像データ５１２」と表す。なお、ステップＳ４１の加工処理において０クリアした部分は、Ｓｔｅｐ４３３又はＳｔｅｐ４３４で取り除いても良い。

【００８４】

プーリング処理部１１３は、画像データ５１２を入力して、第１層のプーリング処理を行う（ステップＳ４４）。

【００８５】

ここで、第１層のプーリング処理について、図９を参照しながら説明する。図９は、本実施形態の第１層のプーリング処理の一例を示す図である。

【００８６】

Ｓｔｅｐ４４１）プーリング処理部１１３は、５６×５６×６４チャンネルの画像データ５１２を入力する。

【００８７】

Ｓｔｅｐ４４２）プーリング処理部１１３は、画像データ５１２の３×３の領域内の最大値を出力する処理を繰り返し行い、２８×２８×６４の画像データ（この画像データを以降「画像データ５１３」とする）を生成する。これは、例えば、以下のようにして行う。

【００８８】

（１）画像データ５１３の１つの５６×５６の画像データ（１つのチャンネルを固定した５６×５６の画像データ）について、左上を中心とした３×３の領域における画素値の最大値を得る。そして、この最大値を、画像データ５１３の画素位置（１，１）の画素値とする。

【００８９】

（２）次に、３×３の領域を右に２ずつ移動させながら、それぞれの領域内における画素値の最大値を得て、それぞれ、画像データ５１３の画素位置（１，２）～（１，２８）の画素値とする。

【００９０】

（３）続いて、３×３の領域の中心を下に２移動させ、左端から同様に２ずつ領域の中心を移動させながら、それぞれの領域内における画素値の最大値を得て、それぞれ、画像データ５１３の画素位置（２，１）～（２，２８）の画素値とする。以降、同様に、（３，１）～（２８，２８）の画素値を得る。

【００９１】

（４）上記の（１）～（３）を、すべての５６×５６の画像データについて行う。すなわち、上記の（１）～（３）を、６４個の５６×５６の画像データについて行う。

【００９２】

Ｓｔｅｐ４４３）プーリング処理部１１３は、画像データ５１３を第２層の畳み込み処理部１１２に出力する。

【００９３】

次に、ＣＮＮ処理部１１０は、畳み込みニューラルネットワークの層を示す変数ｎに１を加算する（ステップＳ４５）。

【００９４】

次に、ＣＮＮ処理部１１０は、変数ｎが、予め設定されたＮと等しいか否かを判定する（ステップＳ４６）。変数ｎがＮと等しい場合、ＣＮＮ処理部１１０は、ステップＳ４７に進む。

【００９５】

一方、変数ｎがＮと等しくない場合（すなわち、変数ｎがＮより小さい場合）、ＣＮＮ処理部１１０は、ステップＳ４３に戻る。すなわち、この場合、ＣＮＮ処理部１１０は、

畳み込みニューラルネットワークの次の層の畳み込み処理及びプーリング処理を行う。
【００９６】
本実施形態では、Ｎ＝２であるため、ＣＮＮ処理部１１０は、ステップＳ４７に進むものとする。
【００９７】
畳み込み処理部１１２は、画像データ５１３を入力して、第２層の畳み込み処理を行う（ステップＳ４７）。
【００９８】
ここで、第２層の畳み込み処理について、図１０を参照しながら説明する。図１０は、本実施形態の第２層の畳み込み処理の一例を示す図である。なお、第２層の畳み込み処理は、第１層の畳み込み処理と各データのチャンネル数が異なること以外は同様である。より一般には、第ｎ層の畳み込み処理は、他の層の畳み込み処理と各データのチャンネル数が異なること以外は同様である。
【００９９】
Ｓｔｅｐ４７１）畳み込み処理部１１２は、画像データ５１３を入力する。ここで、入力した画像データ５１３の色チャンネルは、上述した通り、２８×２８×６４チャンネルである。
【０１００】
Ｓｔｅｐ４７２）畳み込み処理部１１２は、重みデータ１２００－２からフィルタを生成し、画像データ５１３に対して、生成したフィルタを用いてフィルタ処理を行う。ここで、重みデータ１２００－２のデータ構成及び当該重みデータ１２００－２から生成されるフィルタ１３００ｆ$_{j}$－２（ｊ＝１，・・・，６４）のデータ構成について説明する。
【０１０１】
図１１（ｂ）は、第２層の重みデータ１２００－２の一例を示す図である。図１１（ｂ）に示すように、第２層の重みデータ１２００－２は、１６００×６４の行列で表される。なお、重みデータ１２００－２の各値ｗ$_{2}$（ｉ，ｊ）は、上述したように、学習データに基づいて予め学習された値である。
【０１０２】
次に、重みデータ１２００－２から生成されるフィルタ１３００ｆ$_{j}$－２（ｊ＝１，・・・，６４）について説明する。図１２は、本実施形態の第２層のフィルタの一例を示す図である。
【０１０３】
図１２に示すように、各フィルタ１３００ｆ$_{j}$－２（ｊ＝１，・・・，６４）は、５×５の行列の６４個の組で表される。換言すれば、各フィルタ１３００ｆ$_{j}$－２（ｊ＝１，・・・，６４）は、５×５×６４で表される。
【０１０４】
ここで、重みデータ１２００－２のｗ$_{2}$（１，１）～ｗ$_{2}$（２５，１）、・・・、ｗ$_{2}$（１５７６，１）～ｗ$_{2}$（１６００，１）からフィルタ１３００ｆ$_{1}$－２が生成される。同様に、重みデータ１２００－２のｗ$_{2}$（１，２）～ｗ$_{2}$（２５，２）、・・・、ｗ$_{2}$（１５７６，２）～ｗ$_{2}$（１６００，２）からフィルタ１３００ｆ$_{2}$－２が生成される。ｊ＝３，・・・，６４の場合も同様である。
【０１０５】
以上のように生成された各フィルタ１３００ｆ$_{j}$－２（ｊ＝１，・・・，６４）を用いて、畳み込み処理部１１２は、画像データ５１３に対してフィルタ処理を行う。畳み込み処理部１１２は、例えば以下のようにしてフィルタ処理を行う。
【０１０６】
（１）画像データ５１３に対してフィルタ１３００ｆ$_{1}$－２をかける（すなわち、画像データ５１３とフィルタ１３００ｆ$_{1}$－２の対応する値の乗算を行う）。
【０１０７】
これは、例えば、１つのチャンネルを固定し、フィルタ１３００ｆ$_{1}$－２の中心を、画

像データ５１３の２８×２８の部分の左上から５ずつ右にずらしながら行う。そして、フィルタ１３００ｆ$_1$－２の中心が画像データ５１３の２８×２８の部分の右端まで辿り着いたら、フィルタ１３００ｆ$_1$－２の中心を下に５ずらして、再度、左端から行えば良い。

【０１０８】

（２）次に、画像データ５１３の他のチャンネルに対しても、上記（１）と同様の方法でフィルタ１３００ｆ$_1$－２をかける。この処理をすべてのチャンネル１～６４に対して繰り返す。

【０１０９】

（３）フィルタ１３００ｆ$_2$－２～フィルタ１３００ｆ$_{64}$－２についても、上記と同様に、１～６４のチャンネル毎に、画像データ５１３の２８×２８の部分に対して、フィルタ処理を順に行う。

【０１１０】

以上のフィルタ処理により、画像データ５１３から２８×２８×６４×６４チャンネルの画像データが生成される。

【０１１１】

Ｓｔｅｐ４７３）畳み込み処理部１１２は、Ｓｔｅｐ４７２で得られた画像データの２８×２８の部分について、各画素値を１～６４チャンネルのそれぞれについて加算する。この結果、２８×２８×６４チャンネルの画像データが得られる。

【０１１２】

Ｓｔｅｐ４７４）畳み込み処理部１１２は、Ｓｔｅｐ４７３で得られた２８×２８×６４チャンネルの画像データの各画素値に対して、バイアスデータ１１００－２を加算する。

【０１１３】

ここで、図１１（ａ）は、第２層のバイアスデータ１１００－２の一例を示す図である。図１１（ａ）に示すように、バイアスデータ１１００－２は、１×６４の行列により表される。そこで、畳み込み処理部１１２は、１つめの２８×２８チャンネルの画像データの各画素値に対してバイアスデータ１１００－２のデータ値ｂ$_2$（１）を加算する。同様に、２つ目の２８×２８チャンネルの画像データの各画素値に対してバイアスデータ１１００－２のデータ値ｂ$_2$（２）を加算する。以降、同様に、６４個すべての２８×２８チャンネルの画像データの各画素値に対して、それぞれ、バイアスデータ１１００－２のデータ値を加算する。

【０１１４】

Ｓｔｅｐ４７５）畳み込み処理部１１２は、Ｓｔｅｐ４７４で得られた２８×２８×６４チャンネルの画像データに対して、所定の活性化関数を適用して出力画像データを得る。所定の活性化関数としては、例えば、任意の画素値ｘに対して、ｆ（ｘ）＝ｍａｘ（０，ｘ）で定義される関数が挙げられる。このようにして得られた出力画像データが、出力データ５２０である。このように本実施形態の出力データ５２０は、２８×２８×６４チャンネルの画像データである。

【０１１５】

なお、上記の説明で示されるように、出力データ５２０は、フィルタ１３００ｆ$_j$－２の各ｊ（ｊ＝１，・・・，６４）に対応する２８×２８の画像データ（出力データ）の集合と言うことができる。すなわち、出力データ５２０には、フィルタ１３００ｆ$_1$－２に対応する２８×２８の出力データ５２０－１，・・・、フィルタ１３００ｆ$_{64}$－２に対応する２８×２８の出力データ５２０－６４が含まれる。

【０１１６】

次に、図３のステップＳ３３の候補領域の作成処理について、図１３を参照しながら説明する。図１３は、本実施形態の候補領域の作成処理のフローチャートの一例を示す図である。

【０１１７】

候補領域作成処理部１２０のデータ決定部１２１は、出力データ５２０に含まれる出力データ５２０－１，・・・，出力データ５２０－６４のそれぞれについて代表値a_1，・・・a_{64}を決定する（ステップＳ１３１）。

【０１１８】

ここで、代表値a_1，・・・a_{64}としては、出力データ５２０－１，・・・，出力データ５２０－６４それぞれのデータ値の最大値とすれば良い。例えば、出力データ５２０－１に含まれるデータ値の最大値を代表値a_1とすれば良い。他の出力データ５２０－２，・・・，出力データ５２０－６４についても同様である。ただし、代表値a_1，・・・a_{64}は、最大値に限られず、例えば、平均値等を用いても良い。

【０１１９】

候補領域作成処理部１２０のデータ決定部１２１は、代表値a_1，・・・a_{64}に基づいて、出力データ５２０－１，・・・，出力データ５２０－６４から所定のＭ個のデータを決定する（ステップＳ１３２）。ここで、データ決定部１２１は、代表値a_1，・・・a_{64}の値が大きい順に（昇順に）、上位Ｍ個の代表値に対応する出力データを決定すれば良い。

【０１２０】

以降では、Ｍ＝３として、データ決定部１２１により、出力データ５２０－２、出力データ５２０－４３、及び出力データ５２０－４７が決定されたものとする。

【０１２１】

なお、Ｍの値を大きくすることで、認識処理の精度を向上させることができるが、処理速度は低下する。一方で、Ｍの値を小さくすることで、認識処理の精度は低下するものの処理速度が向上する。したがって、Ｍは、画像処理プログラム２０の設計者等により、認識対象の画像データ５１０の性質や、認識処理に求められる精度等に応じて適切な値が予め設定される。

【０１２２】

候補領域作成処理部１２０は、データ決定部１２１により決定されたＭ個の出力データ５２０のうち一の出力データを取得する（ステップＳ１３３）。すなわち、本実施形態では、データ決定部１２１は、出力データ５２０－２、出力データ５２０－４３、及び出力データ５２０－４７から一の出力データを取得する。以降では、候補領域作成処理部１２０は、出力データ５２０－２を取得したものとして説明する。

【０１２３】

候補領域作成処理部１２０の境界決定部１２２は、取得された出力データ５２０－２について、微分処理を行って、領域分割部１２４により分割される領域の境界を決定する（ステップＳ１３４）。

【０１２４】

ここで、境界決定部１２２により決定される領域の境界について、図１４を参照しながら説明する。図１４は、本実施形態の微分処理の一例を示す図である。

【０１２５】

図１４では、一例として、出力データ５２０－２について、微分処理を行った場合を示している。図１４に示すように、境界決定部１２２により微分処理を行い、微分値が負から正に変わる部分を、出力データ５２０－１の出力値の谷間として検出する。そして、境界決定部１２２は、検出された出力値の谷間を、境界Ｄ１及び境界Ｄ２として決定する。ここで、微分処理には、例えばSobelフィルタを用いれば良い。

【０１２６】

候補領域作成処理部１２０の閾値処理部１２３は、閾値処理を行う（ステップＳ１３５）。すなわち、閾値処理部１２３は、予め設定された閾値（例えば、閾値＝３０）以下のデータを削除する。

【０１２７】

ここで、閾値処理部１２３による閾値処理について、図１５を参照しながら説明する。図１５は、本実施形態の閾値処理の一例を示す図である。図１５では、一例として、出力

データ５２０－２に対して閾値処理を行った場合を示している。図１５に示すように、閾値処理部１２３は、閾値処理を行って所定の閾値以下のデータ値を削除することにより、出力データ５２０－２から出力データ５２１－２を作成する。なお、図１５に示す出力データ５２１において、網掛けで示した部分がデータ値を削除した部分である。

【０１２８】

候補領域作成処理部１２０の領域分割部１２４は、境界決定部１２２により決定された境界に基づいて、ステップＳ１３３で取得された一の出力データが示す画像を複数の領域に分割する（ステップＳ１３６）。

【０１２９】

ここで、領域分割部１２４により分割される領域について、図１６を参照しながら説明する。図１６は、本実施形態の領域分割の一例を示す図である。図１６では、出力データ５２１－２が示す画像を境界Ｄ１及び境界Ｄ２に基づいて分割した例を示している。図１６に示すように、出力データ５２１－２が示す画像は、境界Ｄ１及び境界Ｄ２に基づいて、領域Ｓ１、領域Ｓ２、領域Ｓ３、及び領域Ｓ４に分割される。

【０１３０】

候補領域作成処理部１２０の候補領域作成部１２５は、領域分割部１２４により分割された領域Ｓ１～Ｓ４について、各領域を含む最小矩形を特定し、当該特定された最小矩形に基づいて候補領域を示す候補領域画像データ５３０を作成する（ステップＳ１３７）。

【０１３１】

ここで、一例として、領域Ｓ１を囲む最小矩形Ｂ１を図１７に示す。このように最小矩形とは、領域分割部１２４により分割された領域された領域に外接する矩形のことである。したがって、候補領域作成部１２５は、各領域Ｓ１～Ｓ４について、それぞれ最小矩形を特定する。

【０１３２】

そして、候補領域作成部１２５は、画像データ５１０が示す画像において、当該特定された最小矩形によって囲まれる領域と対応する領域を候補領域として候補領域画像データ５３０を作成する。このとき、候補領域作成部１２５は、画像データ５１０が示す画像において、最小矩形によって囲まれる領域と対応する領域を、当該画像データ５１０の解像度を考慮した上で候補領域として候補領域画像データ５３０を作成する。

【０１３３】

候補領域作成処理部１２０は、ステップＳ１３２で決定されたすべての出力データに対して、候補領域画像データ５３０を作成したか否かを判定する（ステップＳ１３８）。すなわち、候補領域作成処理部１２０は、出力データ５２０－２、出力データ５２０－４３、及び出力データ５２０－４７に対して、ステップＳ１３３～ステップＳ１３８の処理が実行されたか否かを判定する。

【０１３４】

ステップＳ１３２で決定されたすべての出力データに対して、候補領域画像データ５３０が作成された場合、候補領域作成処理部１２０は、処理を終了させる。一方、ステップＳ１３２で決定された出力データのうち、候補領域画像データ５３０が作成されていない出力データがある場合、候補領域作成処理部１２０は、ステップＳ１３３に戻る。

【０１３５】

これにより、本実施形態の画像処理装置１０では、入力された画像データ５１０が示す画像において、対象が含まれる領域の候補である候補領域を示す候補領域画像データ５３０が作成される。しかも、本実施形態の画像処理装置１０では、畳み込みニューラルネットワークの第Ｎ層における出力データ５２０を用いて、候補領域画像データ５３０が作成される。このため、本実施形態の画像処理装置１０では、認識処理の精度の低下を防ぎつつ、候補領域を削減させることができる。

【０１３６】

次に、図３のステップＳ３４のカテゴリ分類処理について、図１８を参照しながら説明する。図１８は、本実施形態のカテゴリ分類処理のフローチャートの一例を示す図である

。
【０１３７】
ＣＮＮ処理部１１０は、１以上の候補領域画像データ５３０から一の候補領域画像データ５３０を入力し、入力された候補領域画像データ５３０に対して、畳み込みニューラルネットワーク処理を行う（ステップＳ１８１）。すなわち、ＣＮＮ処理部１１０は、入力された候補領域画像データ５３０に対して、図４で示した畳み込みニューラルネットワーク処理を行う。
【０１３８】
なお、ステップＳ１８１において、ＣＮＮ処理部１１０は、予め設定された第Ｎ層までの畳み込みニュールラルネットワーク処理を行っても良いし、Ｎより大きい任意の自然数をＬとして、第Ｌ層までの畳み込みニューラルネットワーク処理を行っても良い。
【０１３９】
ここでは、ステップＳ１８１において、ＣＮＮ処理部１１０は、第Ｎ層までの畳み込みニューラルネットワーク処理を行ったものとして説明する。したがって、ステップＳ１８１の処理結果として、ＣＮＮ処理部１１０の畳み込み処理部１１２は、出力データ５２０と同じデータ構成である２８×２８×６４チャンネルの出力データ５３１を全結合処理部１１４に出力する。
【０１４０】
次に、ＣＮＮ処理部１１０の全結合処理部１１４は、出力データ５３１を入力して、全結合処理を行う。なお、全結合処理部１１４は、上述したように、カテゴリの組毎に存在する。したがって、各全結合処理部１１４は、それぞれ、出力データ５３１を入力する。
【０１４１】
例えば、カテゴリ数が「人」、「動物」、「車」の３つである場合、全結合処理部１１４は、カテゴリの組「人」「人以外」に対応する全結合処理部１１４－１、カテゴリの組「動物」「動物以外」に対応する全結合処理部１１４－２、及びカテゴリの組「車」「車以外」に対応する全結合処理部１１４－３の３つが存在する。
【０１４２】
ここで、全結合処理について、図１９を参照しながら説明する。図１９は、本実施形態の第３層の全結合処理の一例を示す図である。
【０１４３】
Ｓｔｅｐ１８２１）全結合処理部１１４は、出力データ５３１を入力する。ここで、入力した出力データ５３１の色チャンネルは、上述したように、２８×２８×６４である。
【０１４４】
Ｓｔｅｐ１８２２）全結合処理部１１４は、出力データ５３１の各データ値をベクトル値に変換する。すなわち、２８×２８×６４チャンネルの出力データ５３１の各データ値を５０１７６行１列のベクトル値に変換する。ここで、ベクトル値の各成分の値をx_1，・・・，x_{50176}とする。
【０１４５】
Ｓｔｅｐ１８２３）全結合処理部１１４は、それぞれ、バイアスデータ１１００－３及び重みデータ１２００－３を用いて、積和演算を行う。
【０１４６】
ここで、バイアスデータ１１００－３及び重みデータ１２００－３について、図２０を参照しながら説明する。図２０は、本実施形態の第３層のネットワークパラメータの一例を示す図である。
【０１４７】
図２０（ａ）は、第３層のバイアスデータ１１００－３の一例を示す図である。図２０（ａ）に示すように、第３層のバイアスデータ１１００－３は、カテゴリ毎のバイアスデータ$1100-3_1$，バイアスデータ$1100-3_2$，・・・を含む。また、カテゴリ毎のバイアスデータ$1100-3_k$は、１行２列のベクトル値である。なお、ベクトルの各成分の値b_3（ｋ，ｊ）は、上述したように、学習データに基づいて予め学習された値で

ある。
【０１４８】
ここで、ｋは、カテゴリを示す数値であるとする。例えば、ｋ＝１のときカテゴリ「人」を示し、ｋ＝２のときカテゴリ「動物」を示し、ｋ＝３のときカテゴリ「車」を示す等である。また、ｊは、カテゴリに分類されるか否かを示す数値である。例えば、ｊ＝１のときは該当のカテゴリに分類される場合を示し、ｊ＝２のときは該当のカテゴリに分類されない場合（すなわち、該当のカテゴリ以外のカテゴリに分類される場合）を示す。
【０１４９】
図２０（ｂ）は、第３層の重みデータ１２００－３の一例を示す図である。図２０（ｂ）に示すように、第３層の重みデータ１２００－３は、カテゴリ毎の重みデータ１２００－３_{1}，重みデータ１２００－３_{2}，・・・を含む。また、カテゴリ毎の重みデータ１２００－３_{k}は、５０１７６行２列の行列である。なお、この行列の各成分の値$ｗ_{3}$（ｉ，ｊ，ｋ）は、上述したように、学習データに基づいて予め学習された値である。
【０１５０】
図１９の説明に戻り、全結合処理部１１４は、それぞれ以下の積和演算を行う。すなわち、カテゴリｋに対して、全結合処理部１１４－ｋは、以下の積和演算を行う。
【０１５１】
【数１】

$$y_j(k)=b_3(k,j)+\sum_{i=1}^{50176}w_3(i,j,k)x_i$$

ここで、ｊ及びｋの意味は上述した通りである。
【０１５２】
Ｓｔｅｐ１８２４）全結合処理部１１４は、Ｓｔｅｐ１８２３で得られた２×１×｜ｋ｜のデータを正規化処理部１３０に出力する。なお、｜ｋ｜は、カテゴリ数である。
【０１５３】
なお、上記の積和演算の結果が、入力された候補領域画像データ５３０がカテゴリｋに分類される場合（ｊ＝１の場合）の算出結果と、当該候補領域画像データ５３０がカテゴリｋ以外のカテゴリに分類される場合（ｊ＝２の場合）の算出結果である。
【０１５４】
これにより、候補領域画像データ５３０が、あるカテゴリｋに分類されるか否かを数値として判定することができる。例えば、あるカテゴリｋについて、$ｙ_{1}$（ｋ）の値が０．７、$ｙ_{2}$（ｋ）の値が０．３である場合、当該候補領域画像データ５３０は、カテゴリｋに分類される場合が高いと判定することができる。換言すれば、あるカテゴリｋについて、$ｙ_{1}$（ｋ）の値が$ｙ_{2}$（ｋ）の値より高い場合、入力された候補領域画像データ５３０はカテゴリｋに分類される可能性が高いといえる。
【０１５５】
ただし、上記の算出結果では、各全結合処理部１１４の出力結果同士の比較ができない場合があるため、次のステップＳ１８３において正規化処理を行う。
【０１５６】
正規化処理部１３０は、全結合処理部１１４により出力された２×１×｜ｋ｜のデータを入力して、正規化処理を行う（ステップＳ１８３）。
【０１５７】
ここで、正規化処理について、図２１を参照しながら説明する。図２１は、本実施形態の正規化処理の一例を示す図である。
【０１５８】

Ｓｔｅｐ１８３１）正規化処理部１３０は、全結合処理部１１４により出力された２×１×｜ｋ｜のデータを入力する。

【０１５９】

Ｓｔｅｐ１８３２）正規化処理部１３０は、（y_1（ｋ），y_2（ｋ））について、カテゴリ毎に以下の式により正規化を行う。

【０１６０】

【数２】

$$z_j(k)=\frac{\exp(y_j(k))}{\exp(y_1(k))+\exp(y_2(k))}$$

このようにして得られた２×１×｜ｋ｜が確信度である、このように正規化処理を行うことにより、すべてのカテゴリにおける確信度は０以上１以下の値に正規化される。このため、異なるカテゴリ同士の確信度を比較することが可能となる。例えば、ｋ＝１をカテゴリ「人」、ｋ＝２をカテゴリ「動物」とした場合において、z_1（１）＝０．８，z_2（１）＝０．２，z_1（２）＝０．６，z_2（２）＝０．４であるとき、入力された候補領域画像データ５３０は、カテゴリ「人」に分類される可能性が高いと言える。

【０１６１】

Ｓｔｅｐ１８３３）正規化処理部１３０は、各カテゴリの確信度を出力部１４０に出力する。

【０１６２】

以上により、本実施形態の画像処理装置１０では、入力された画像データが示す画像において、被写体等を示す対象が含まれる領域の候補となる候補領域画像データを作成する。しかも、本実施形態の画像処理装置１０では、畳み込みニューラルネットワークの予め設定された層の出力結果に基づいて、候補領域画像データを作成することにより、認識処理の精度の低下を防ぎつつ、候補領域画像データの数の削減を図ることができる。

【０１６３】

したがって、本実施形態の画像処理装置１０は、入力された画像データが示す画像において、対象が含まれる領域と、当該対象が分類されるカテゴリとを識別する識別処理の処理時間を削減することができる。

【０１６４】

本発明は、具体的に開示された上記の実施形態に限定されるものではなく、特許請求の範囲から逸脱することなく、種々の変形や変更が可能である。

【符号の説明】

【０１６５】

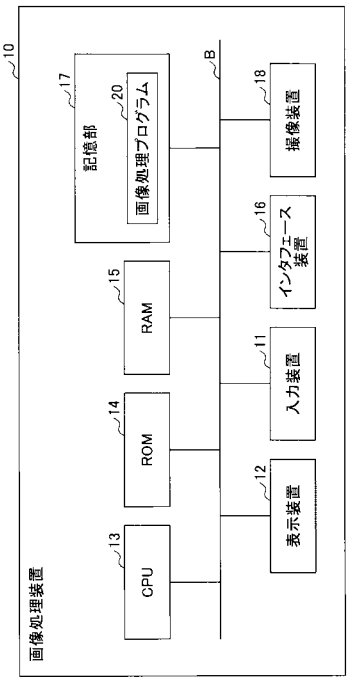
１０　画像処理装置
２０　画像処理プログラム
１１０　ＣＮＮ処理部
１１１　加工部
１１２　畳み込み処理部
１１３　プーリング処理部
１１４　全結合処理部
１２０　候補領域作成処理部
１２１　データ決定部
１２２　境界決定部
１２３　閾値処理部
１２４　領域分割部

１２５　候補領域作成部
１３０　正規化処理部
１４０　出力部

【先行技術文献】
【特許文献】
【０１６６】
【特許文献１】特許第４３２２９１３号公報
【非特許文献】
【０１６７】
【非特許文献１】Rich feature hierarchies for accurate object detection and semantic segmentation. Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik. CVPR 2014.

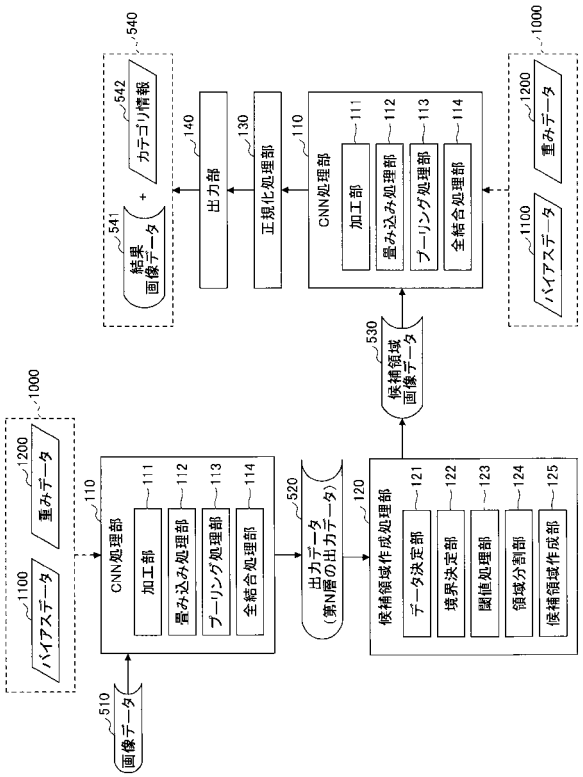
【図１】

本実施形態の画像処理装置のハードウェア構成の一例を示す図



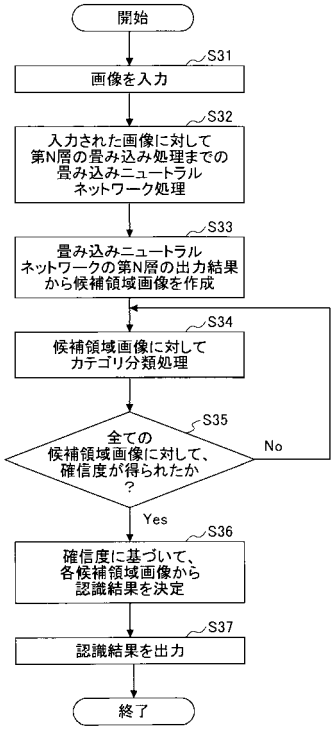
【図２】

本実施形態の画像処理装置の機能構成の一例を示す図



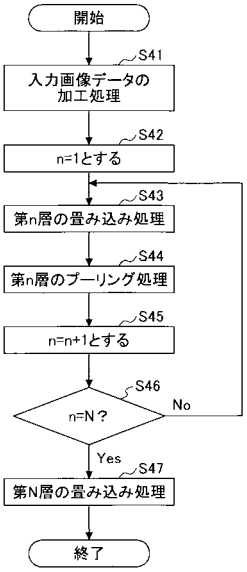
【図３】

本実施形態の画像処理装置の認識処理のフローチャートの一例を示す図



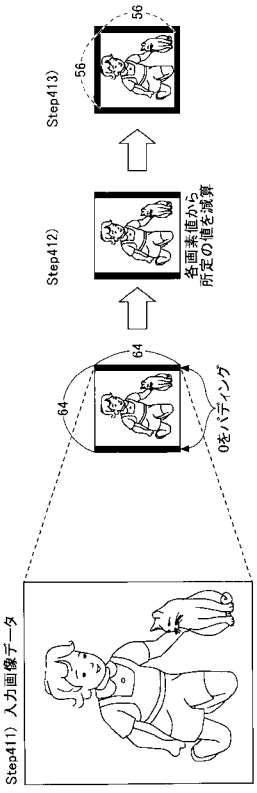
【図４】

本実施形態の
畳み込みニューラルネットワーク処理のフローチャートの一例を示す図



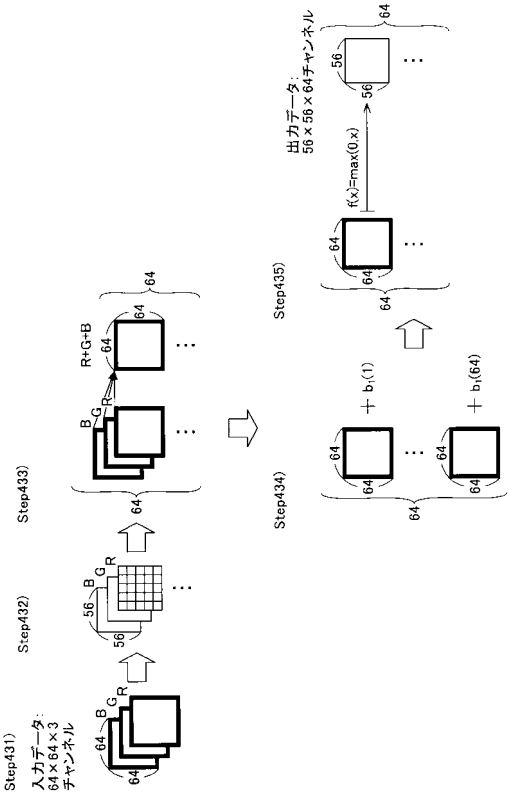
【図５】

本実施形態の入力画像データの加工処理の一例を示す図



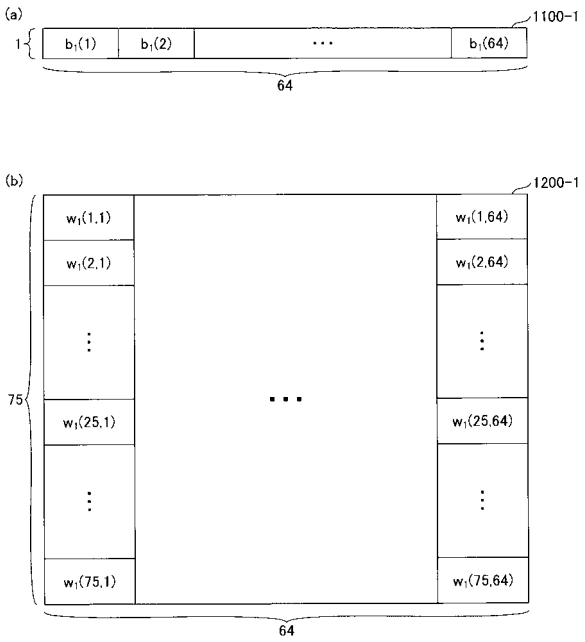
【図６】

本実施形態の第１層の畳み込み処理の一例を示す図



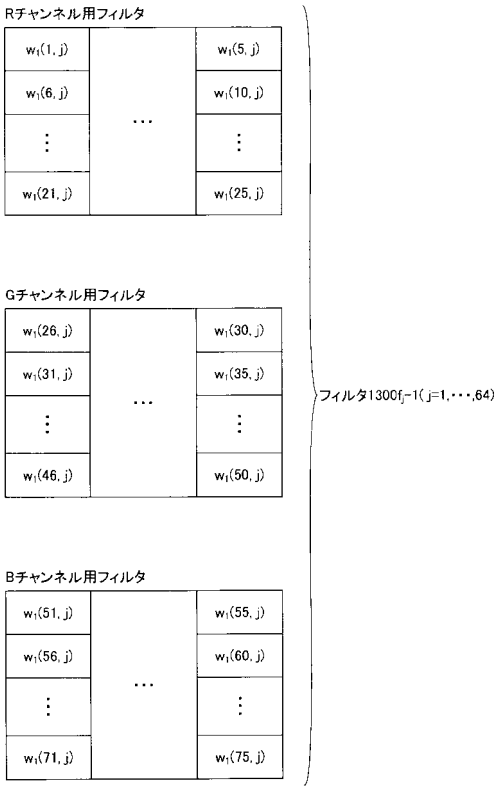
【図７】

本実施形態の第１層のネットワークパラメータの一例を示す図



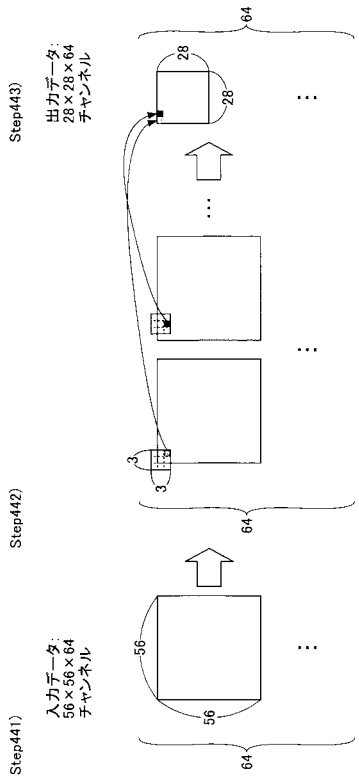
【図８】

本実施形態の第１層のフィルタの一例を示す図



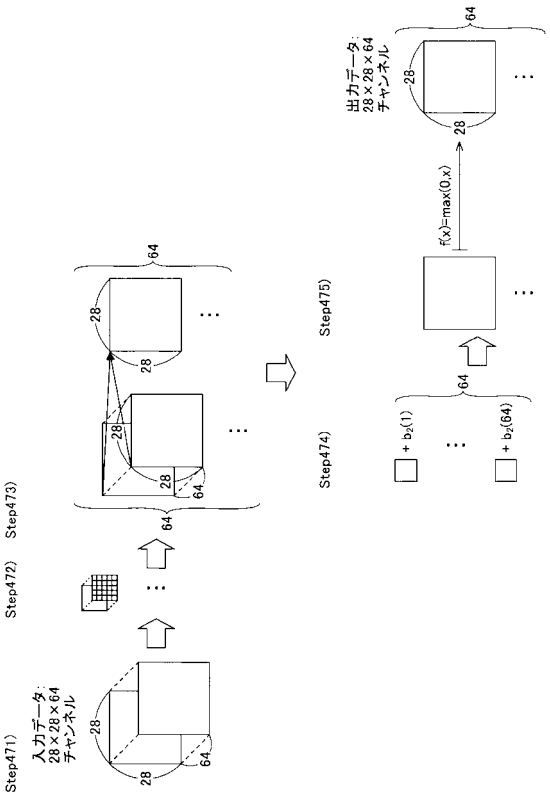
【図９】

本実施形態の第１層のプーリング処理の一例を示す図



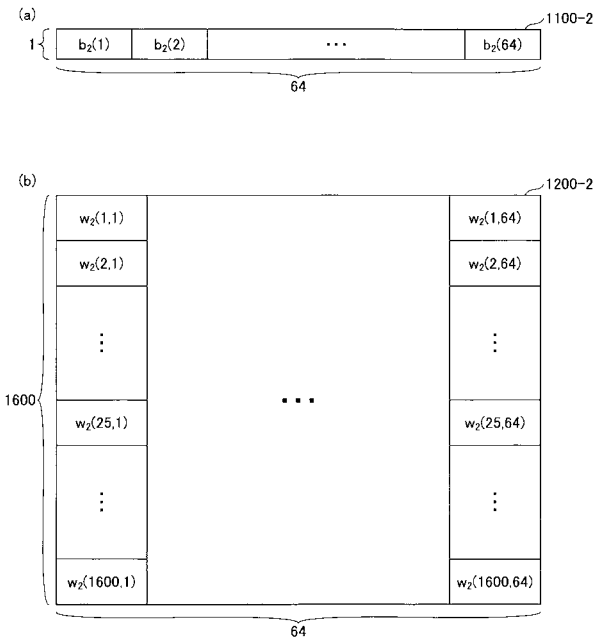
【図１０】

本実施形態の第２層の畳み込み処理の一例を示す図



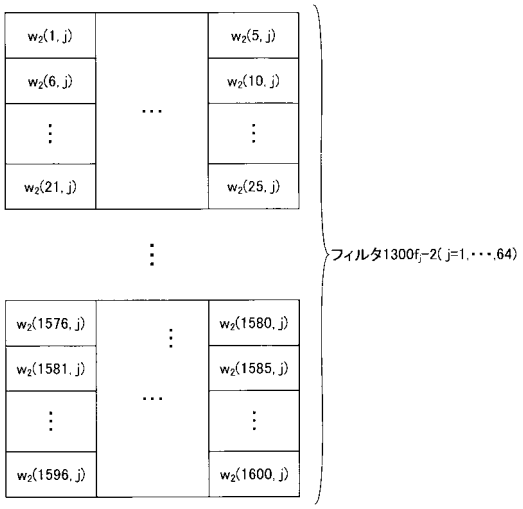
【図１１】

本実施形態の第２層のネットワークパラメータの一例を示す図



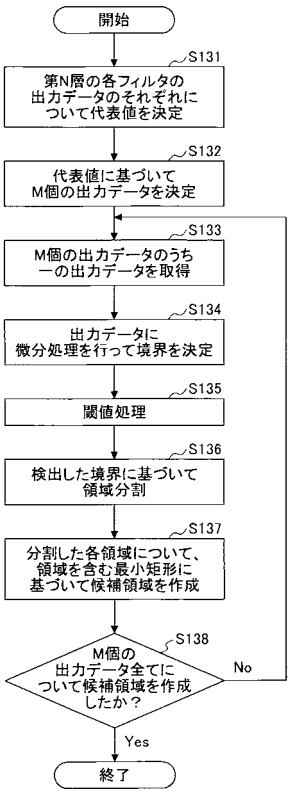
【図１２】

本実施形態の第２層のフィルタの一例を示す図



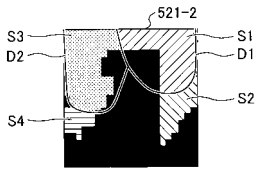
【図１３】

本実施形態の候補領域の作成処理のフローチャートの一例を示す図



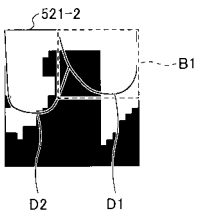
【図１６】

本実施形態の領域分割の一例を示す図



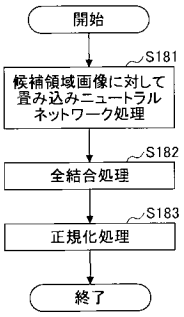
【図１７】

本実施形態の最小矩形の一例を示す図



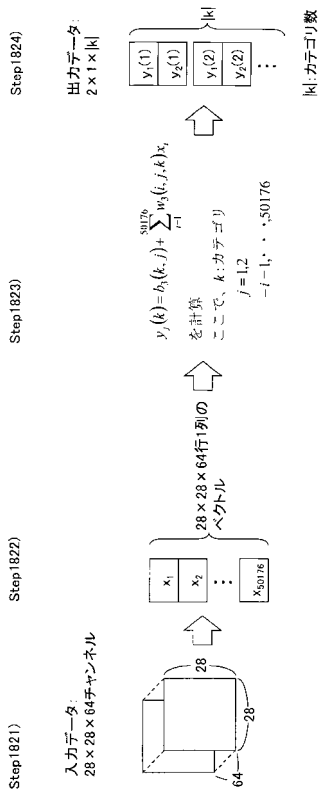
【図１８】

本実施形態のカテゴリ分類処理のフローチャートの一例を示す図



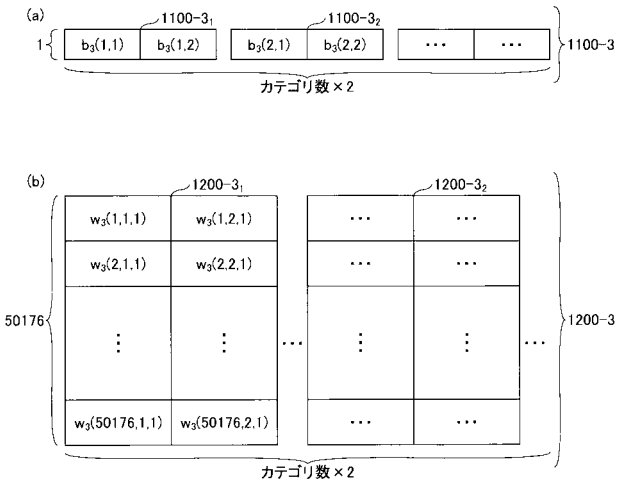
【図１９】

本実施形態の第3層の全結合処理の一例を示す図



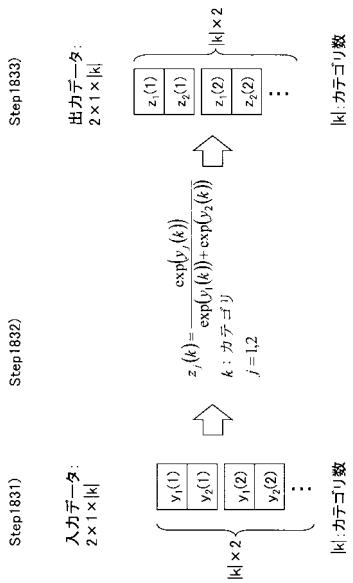
【図２０】

本実施形態の第3層のネットワークパラメータの一例を示す図



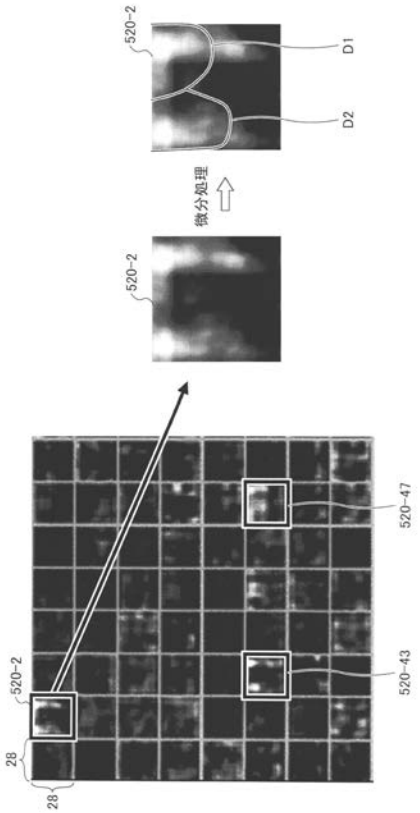
【図２１】

本実施形態の正規化処理の一例を示す図



【図１４】

本実施形態の微分処理の一例を示す図



【図１５】

本実施形態の閾値処理の一例を示す図