

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7371466号
(P7371466)

(45)発行日 令和5年10月31日(2023.10.31)

(24)登録日 令和5年10月23日(2023.10.23)

(51)国際特許分類 F I
G 0 6 T 7/00 (2017.01) G 0 6 T 7/00 3 5 0 C

請求項の数 3 (全18頁)

(21)出願番号	特願2019-218770(P2019-218770)	(73)特許権者	000006150 京セラドキュメントソリューションズ株式会社
(22)出願日	令和1年12月3日(2019.12.3)		
(65)公開番号	特開2021-89512(P2021-89512A)		
(43)公開日	令和3年6月10日(2021.6.10)	(74)代理人	110003443 弁理士法人T N Kアジア国際特許事務所
審査請求日	令和4年11月30日(2022.11.30)	(74)代理人	100129997 弁理士 田中 米藏
		(72)発明者	西田 篤志 大阪市中央区玉造1丁目2番28号 京セラドキュメントソリューションズ株式会社内
		審査官	片岡 利延

最終頁に続く

(54)【発明の名称】 画像処理装置

(57)【特許請求の範囲】

【請求項1】

生成部と、畳込みニューラルネットワーク部と、グラッドカム部と、記憶部と、比較部と、出力部と、補正部とを含む処理部を備え、

前記生成部は、処理対象とされる画像データから濃度マップを生成し、

前記畳込みニューラルネットワーク部は、前記濃度マップにフィルターをかけて特徴マップを生成し、当該特徴マップから分類データを生成する処理を行い、

前記グラッドカム部は、活性化関数を用いて前記特徴マップからアクティベーションマップを生成する処理を行い、当該アクティベーションマップをクラスタリングにより複数のグループに分割し、当該複数のグループについてそれぞれ補正関数を算出し、

前記記憶部は、教師データを記憶しており、

前記比較部は、前記分類データと前記教師データとを用いて、前記分類データについての第1損失関数を算出する処理を行い、

前記出力部は、前記比較部により算出された前記第1損失関数に前記補正関数をそれぞれ加算して、前記複数のグループ毎に第2損失関数を算出し、

前記補正部は、前記出力部が算出した前記各第2損失関数を合計した補正值を用いて、前記畳込みニューラルネットワーク部で用いる前記フィルターの重み付けを補正し、

前記補正部により作成された新規の前記フィルターを用いた前記畳込みニューラルネットワーク部による前記処理から、前記グラッドカム部、前記比較部、前記出力部、及び前記補正部による処理を繰り返すことで前記フィルターを補正して更新する、画像処理装置。

【請求項 2】

前記グラッドカム部は、前記複数のグループへの分割を、前記アクティベーションマップから、クラスタ毎に、濃度が濃い中心点を検出し、当該中心点から予め定められた一定の距離内に存在する各点の集合を1つのグループとすることにより行う請求項1に記載の画像処理装置。

【請求項 3】

前記処理部は、分類部を更に備え、

前記分類部は、前記更新されたフィルターを用いて前記畳込みニューラルネットワーク部により生成された分類データが示す各値の配列によって、画像の種別を判定する請求項1又は請求項2に記載の画像処理装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、画像処理装置に関する。

【背景技術】

【0002】

特許文献1には、オブジェクトを含む入力画像を取得し、入力画像から、背景画像を用いて変化する領域の画像である変化領域画像を抽出し、入力画像と変化領域画像とを結合して畳込み型ニューラルネットワークを利用することにより、R、G、B、Oの特徴画像を抽出し、当該特徴画像からオブジェクトの位置を検出する技術が記載されている。

【先行技術文献】

【特許文献】

【0003】

【文献】特開2017-191501公報

【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、特許文献1に記載された情報処理装置では、特徴画像から抽出されたオブジェクトの位置の正確さが低いという問題がある。例えば、ある画像の種別を判定させた場合、当該画像のどの特徴部分に注目してこの判定を行ったかを正確に示すことは困難である。また、判定に用いた上記特徴部分が単数ではなく複数である場合には、どの特徴部分に注目してこの判定を行ったかを正確に示すことは更に困難になる。

【0005】

本発明は上記課題に鑑みてなされたものであり、画像に含まれるオブジェクトに基づいて当該画像の種別を判定するときに、当該オブジェクトが単数及び複数のいずれの場合であっても、当該オブジェクトを的確に抽出して、当該画像の種別を判定する精度を高く保つことを目的とする。

【課題を解決するための手段】

【0006】

本発明の一局面に係る画像処理装置は、生成部と、畳込みニューラルネットワーク部と、グラッドカム部と、記憶部と、比較部と、出力部と、補正部とを含む処理部を備え、前記生成部は、処理対象とされる画像データから濃度マップを生成し、前記畳込みニューラルネットワーク部は、前記濃度マップにフィルターをかけて特徴マップを生成し、当該特徴マップから分類データを生成する処理を行い、前記グラッドカム部は、活性化関数を用いて前記特徴マップからアクティベーションマップを生成する処理を行い、当該アクティベーションマップをクラスタリングにより複数のグループに分割し、当該複数のグループについてそれぞれ補正関数を算出し、前記記憶部は、教師データを記憶しており、前記比較部は、前記分類データと前記教師データとを用いて、前記分類データについての第1損失関数を算出する処理を行い、前記出力部は、前記比較部により算出された前記第1損失関数に前記補正関数をそれぞれ加算して、前記複数のグループ毎に第2損失関数を算出し

10

20

30

40

50

、前記補正部は、前記出力部が算出した前記各第 2 損失関数を合計した補正值を用いて、前記畳込みニューラルネットワーク部で用いる前記フィルターの重み付けを補正し、前記補正部により作成された新規の前記フィルターを用いた前記畳込みニューラルネットワーク部による前記処理から、前記グラッドカム部、前記比較部、前記出力部、及び前記補正部による処理を繰り返すことで前記フィルターを補正して更新する、ものである。

【発明の効果】

【0007】

本発明によれば、画像に含まれるオブジェクトに基づいて当該画像の種別を判定するときに、当該オブジェクトが単数及び複数のいずれの場合であっても、当該オブジェクトを的確に抽出して、当該画像の種別を判定する精度を高く保つことができる。

10

【図面の簡単な説明】

【0008】

【図 1】本発明に係る画像処理装置の一実施形態に係る画像形成装置の内部構成を示す図である。

【図 2】画像形成装置の電氣的構成を示すブロック図である。

【図 3】(A)は原稿を示す図、(B)は原稿を読み取って得た画像データから作成した濃度マップを示す図、(C)は濃度マップに対して用いるフィルターを示す図である。

【図 4】(A)～(C)は畳込みニューラルネットワーク部による処理を説明する図である。

【図 5】(A)～(C)は畳込みニューラルネットワーク部による処理を説明する図である。

20

【図 6】(A)～(C)は畳込みニューラルネットワーク部による処理を説明する図である。

【図 7】活性化関数 ReLU を説明する図である。

【図 8】アクティベーションマップの一例を示す図である。

【図 9】(A)は教師データ、分類データ、及び差分データの一例を示す図、(B)は畳込みニューラルネットワーク部が用いるフィルターが補正されていく変遷を示す図である。

【図 10】(A)～(C)は、教師データ、分類データ、及び差分データの一例を示す図である。

【図 11】(A)～(C)は、フィルターが補正される度にアクティベーションマップが変化していく変遷を示す図である。

30

【発明を実施するための形態】

【0009】

以下、本発明の一実施形態に係る画像処理装置及び画像処理方法について、図面を参照しながら説明する。なお、以下の説明において、同一又は近似する各部については同一の符号を付し、繰り返しの説明は省略する。

【0010】

図 1 を参照して、本発明の一実施形態に係る画像処理装置について説明する。図 1 は、本発明に係る画像処理装置の一実施形態に係る画像形成装置の内部構成を示す図である。

【0011】

画像形成装置 100 は、複写機、プリンター、及びファクシミリの機能を兼ね備えた複合機である。

40

【0012】

図 1 に示すように、画像形成装置 100 は、原稿搬送部 2 と、読取部 4 と、給送部 6 と、搬送部 8 と、画像形成部 10 と、排出部 12 と、処理部 20 とを備える。更に、画像形成装置 100 は、操作部 5 及びネットワークインターフェイス部 91 を備える(図 2)。

【0013】

原稿搬送部 2 は、トレイ 80 に配置された原稿 G を搬送する。原稿搬送部 2 は、ピックアップローラー 82 と、複数の搬送ローラー 84 とを含んでよい。また、原稿搬送部 2 の一例は、ADF (Auto Document Feeder) である。原稿 G は、紙、ま

50

たはプロジェクターに用いるプラスチックシートである。

【0014】

読取部4は、原稿Gの画像を読み取る。読取部4は、画像を読み取って画像データを生成する。読取部4は、光源86と、複数の反射ミラー88と、レンズ90と、撮像部92とを含む。読取部4の一例は、スキャナーである。

【0015】

光源86は、複数の発光素子を有する。発光素子は、例えば、発光ダイオード(Laser Emitting Diode: LED)である。光源86から出射された光は、原稿搬送部2を搬送される原稿Gによって反射した後、複数の反射ミラー88で反射されて、レンズ90を通り、撮像部92に到達する。

10

【0016】

撮像部92は、レンズ90から光を受け取る複数の受光素子を有している。撮像部92は、例えば、電荷結合素子(Charge Coupled Device: CCD)である。撮像部92は、撮像部92に到達した光からアナログ電気信号を生成する。その後、A/D変換部(図示せず)において、当該アナログ信号がデジタル信号に変換され、このデジタル信号により画像データが構成される。読取部4は、当該画像データを処理部20に出力する。

【0017】

給送部6は、複数のシートSを収容し、搬送部8へシートSを給送する。シートSは、例えば、紙製または合成樹脂製である。搬送部8は、複数の搬送ローラー対を含み、画像形成部10にシートSを搬送する。

20

【0018】

画像形成部10は、電子写真方式によってシートSにトナー像を形成する。具体的には、画像形成部10は、感光体ドラムと、帯電装置と、露光装置と、現像装置と、補給装置と、転写装置と、クリーニング装置と、除電装置とを含む。

【0019】

トナー像は、例えば、原稿Gの画像を示す。排出部12は、画像形成装置100の外部にシートSを排出する。

【0020】

次に、図2を参照して、本実施形態に係る画像形成装置100の電気的構成を説明する。図2は、本実施形態に係る画像形成装置100の電気的構成を示すブロック図である。図3(A)は原稿を示す図、(B)は原稿を読み取って得た画像データから作成した濃度マップを示す図、(C)は濃度マップに対して用いるフィルターを示す図である。

30

【0021】

処理部20は、プロセッサ、RAM(Random Access Memory)、ROM(Read Only Memory)、及び専用のハードウェア回路を含んで構成される。プロセッサは、例えばCPU(Central Processing Unit)、ASIC(Application Specific Integrated Circuit)、又はMPU(Micro Processing Unit)等である。処理部20は、生成部22と、畳込みニューラルネットワーク部24と、グラッドカム部26と、記憶部28と、比較部30と、出力部32と、補正部34と、分類部36とを備えている。

40

【0022】

処理部20は、画像形成装置100が備えるHDD111又は上記ROMに記憶されている制御プログラムに従った上記プロセッサによる動作により、生成部22と、畳込みニューラルネットワーク部24と、グラッドカム部26と、記憶部28と、比較部30と、出力部32と、補正部34と、分類部36として機能するものである。但し、当該生成部22～分類部36は、上記プロセッサによる制御プログラムに従った動作によらず、それぞれハードウェア回路により構成することも可能である。

【0023】

生成部22、畳込みニューラルネットワーク部24、グラッドカム部26、記憶部28、比較部30、出力部32、及び補正部34は、例えば読取部4による原稿読取で得られ

50

た画像データに対して、以下に示す処理を行う。例えば、原稿Gは、図3(A)に示すように、表題に法人名が記載された文書である。原稿Gの表題には、例えば、「A B C D株式会社」と記載された画像部分A1を有し、更に、文末に「A B C D株式会社」がもう一度記載された画像部分A2を有している。

【0024】

生成部22は、上記画像データから濃度マップ50(図3(B))を生成する。

【0025】

畳込みニューラルネットワーク部24は、濃度マップ50にフィルター52(図3(C))をかけて、第1特徴マップ54(図4(C))を生成する処理を行う。更に、畳込みニューラルネットワーク部24は、第1特徴マップ54から第2特徴マップ(図5(C))を生成する処理を行う。更に、畳込みニューラルネットワーク部24は、第2特徴マップから分類データを生成する処理を行う。

10

【0026】

グラッドカム部26は、Grad-cam処理を行うことにより、第2特徴マップ56からアクティベーションマップ66を生成する。

【0027】

更にグラッドカム部26は、生成したアクティベーションマップ66を複数のグループに分割する。グラッドカム部26は、アクティベーションマップ66を複数のグループに分割する処理を、クラスタリング(例えば、k-means)により行う。本実施形態では、グラッドカム部26は、アクティベーションマップ66に対するk-meansによるクラスタリングを行って、アクティベーションマップ66を複数に分割する例を説明する。例えば、グラッドカム部26は、アクティベーションマップ66から、クラスタ毎に、濃度が濃い中心点を検出し、当該中心点から予め定められた一定の距離内に存在する各点の集合を1つのグループとする分割処理を行う。

20

【0028】

なお、グラッドカム部26は、上記画像データに対して、上記グループ数が推定されるx-meansでのクラスタリングを行って、推定されたグループ数にアクティベーションマップ66を分割してもよい。

【0029】

記憶部28は、例えばHDD又はメモリーであり、教師データ60を記憶している。本実施形態では、処理部20により、原稿G(図3(A))がA B C D株式会社宛に作成された文書であるのかを判定する場合を例にする。このため、例えば、「A B C D株式会社」の文字を示す見本画像に対して、生成部22及び畳込みニューラルネットワーク部24による処理を行って分類データを生成し、この生成された分類データを教師データとする。

30

【0030】

比較部30は、分類データ62と、教師データ60とを比較し、差分データ64を算出する。また、比較部30は、分類データ62と教師データ60を用いて第1損失関数(Loss function)を算出する。

【0031】

出力部32は、上記第1損失関数から更に第2損失関数(Loss function)を算出する。

40

【0032】

補正部34は、出力部32から出力されてくる上記第2損失関数を用いてフィルター52の重み付けを補正する。

【0033】

分類部36は、画像判定処理を行う。分類部36は、上記のように補正部34により補正されて更新された最新のフィルター52を用いて畳込みニューラルネットワーク部24により生成された分類データが示す各値の配列によって、画像の種別を判定する。

【0034】

操作部5は、ユーザーから各種操作指示の入力を受け付ける。

50

【 0 0 3 5 】

ネットワークインターフェイス部 9 1 は、図略の LAN チップなどの通信モジュールを備える通信インターフェイスである。ネットワークインターフェイス部 9 1 は、ローカルエリア内、又はインターネット上の外部装置と種々のデータの送受信を行う。

【 0 0 3 6 】

次に、図 2 に加えて更に図 3 乃至図 1 1 を参照して、本実施形態に係る画像形成装置 1 0 0 の具体例を説明する。

【 0 0 3 7 】

まず、図 3 (A) ~ (C) を参照して、畳込みニューラルネットワーク部 2 4 の処理の概略を説明する。畳込みニューラルネットワーク部 2 4 は、畳込み処理とプーリング処理とを繰り返し、分類データ (図 9 (A)) を得る処理を行う。

10

【 0 0 3 8 】

畳込みニューラルネットワーク部 2 4 は、生成部 2 2 により生成された濃度マップ 5 0 (図 3 (B)) に、予め記憶しているフィルター 5 2 (図 3 (C)) を掛け、図 4 (C) に示す第 1 特徴マップ 5 4 を生成し、更に第 1 特徴マップ 5 4 から第 2 特徴マップ 5 6 を生成する処理を行う。

【 0 0 3 9 】

続けて、畳込みニューラルネットワーク部 2 4 は、プーリング処理により、第 2 特徴マップ 5 6 から、予め定められたマトリクス (この実施形態では 2×2 を例にして説明する) ごとに代表値 5 6 C、代表値 5 6 D を抽出する (図 6 (A) (B))。畳込みニューラルネットワーク部 2 4 は、(i) 当該プーリング処理を繰り返して第 2 特徴マップ 5 6 を小さくし、更に、このように小さくした第 2 特徴マップ 5 6 を、一次的に並ぶ一列の予め定められた画素数のデータに展開する、(ii) 第 2 特徴マップ 5 6 を全結合層により、一次的に並ぶ一列の予め定められた画素数のデータに展開する、等の処理により、図 9 (A) に示す分類データ 6 2 を得る。分類データ 6 2 については、図 9 (A) を参照して後述する。

20

【 0 0 4 0 】

グラッドカム部 2 6 は、Grad - cam 処理を行うことにより、第 2 特徴マップ 5 6 からアクティベーションマップ 6 6 を生成する。更にグラッドカム部 2 6 は、生成したアクティベーションマップ 6 6 を複数のグループに分割する。この分割処理により、グラッドカム部 2 6 は、例えば、図 3 (A) に示した原稿 G の画像における画像部分 A 1 を含むグループと、画像部分 A 2 を含むグループとに、アクティベーションマップ 6 6 を分割する。

30

【 0 0 4 1 】

ここで、畳込みニューラルネットワーク部 2 4 による上記畳込み処理を更に詳細に説明する。読取部 4 (図 2) は原稿 G を読み取り、図 3 (A) に例を示す画像データを生成する。

【 0 0 4 2 】

そして、生成部 2 2 は、原稿 G 全体を示す画像について濃度マップ 5 0 (図 3 (B)) を生成する。濃度マップ 5 0 は、画素 $x 1 1$ 、画素 $x 1 2$ 、画素 $x 1 3$ 、 \dots 、画素 $x 2 1$ 、画素 $x 2 2$ 、画素 $x 2 3$ 、 \dots 、画素 $x 3 1$ 、画素 $x 3 2$ 、画素 $x 3 3$ 、 \dots 、画素 $x i j$ 、 \dots がマトリクス状に配列されたものである。それぞれの画素 $x i j$ には、濃度情報が付与されている。

40

【 0 0 4 3 】

なお、一例として、画素 $x i j$ の画素数は 2 桁であるが、2 桁に限られず、1 桁でもよく、3 桁以上であってもよい。また、以降、画素番号を特定する必要がない場合は、画素 $x i j$ と記載する。本実施形態においては、説明を簡略にするため、濃度マップ 5 0 は、画素 $x 1 1$ ~ 画素 $x 9 9$ の 9×9 のマトリクスとする。i および j は、正の整数である。フィルター 5 2 は、画素 $w 1 1$ 、画素 $w 1 2$ 、画素 $w 1 3$ 、画素 $w 2 1$ 、画素 $w 2 2$ 、画素 $w 2 3$ 、画素 $w 3 1$ 、画素 $w 3 2$ 、および画素 $w 3 3$ の画素 $w i j$ が例えば 3×3 のマトリクス状に配列されている。

50

【 0 0 4 4 】

まず、畳込みニューラルネットワーク部 2 4 は、濃度マップ 5 0 の畳込みを行う。具体的には、図 4 (A) に示すように、畳込みニューラルネットワーク部 2 4 は、フィルター 5 2 を濃度マップ 5 0 に掛け合わせる。まず、畳込みニューラルネットワーク部 2 4 は、フィルター 5 2 の画素 $w_{11} \sim w_{33}$ を、濃度マップ 5 0 の画素 $x_{11} \sim x_{33}$ に掛け合わせる。

【 0 0 4 5 】

続いて、畳込みニューラルネットワーク部 2 4 は、フィルター 5 2 を濃度マップ 5 0 で列番号が増える方向に 1 列ずらし、フィルター 5 2 の画素 $w_{11} \sim w_{33}$ を、濃度マップ 5 0 の画素 $x_{12} \sim x_{34}$ に掛け合わせる。以降、同様に、畳込みニューラルネットワーク部 2 4 は、フィルター 5 2 を濃度マップ 5 0 の画素 $x_{11} \sim x_{33}$ のマトリクスから画素 $x_{77} \sim x_{99}$ のマトリクスまで順次掛け合わせる。

10

【 0 0 4 6 】

すなわち、畳込みニューラルネットワーク部 2 4 は、上記のようにフィルター 5 2 を濃度マップ 5 0 に掛け合わせることで、図 4 (B) に示す第 1 特徴データ $y_{ij} = x_{ij} \times w_{ij}$ の行列式を得る。 i および j は、正の整数である。具体的には、畳込みニューラルネットワーク部 2 4 は、フィルター 5 2 の画素 $w_{11} \sim w_{33}$ を濃度マップ 5 0 の画素 $x_{11} \sim x_{33}$ に掛け合わせて、第 1 特徴データ $y_{11} = x_{11}w_{11} + x_{12}w_{12} + x_{13}w_{13} + x_{21}w_{21} + x_{22}w_{22} + x_{23}w_{23} + x_{31}w_{31} + x_{32}w_{32} + x_{33}w_{33}$ を得る。次に、フィルター 5 2 の画素 $w_{11} \sim w_{33}$ を、濃度マップ 5 0 の画素 $x_{12} \sim x_{34}$ に掛け合わせて、第 1 特徴データ $y_{12} = x_{12}w_{11} + x_{13}w_{12} + x_{14}w_{13} + x_{22}w_{21} + x_{23}w_{22} + x_{24}w_{23} + x_{32}w_{31} + x_{33}w_{32} + x_{34}w_{33}$ を得る。以降、同様に、畳込みニューラルネットワーク部 2 4 は、フィルター 5 2 を濃度マップ 5 0 の画素 $x_{77} \sim x_{99}$ のマトリクスまで順次掛け合わせる。

20

【 0 0 4 7 】

そして、畳込みニューラルネットワーク部 2 4 は、濃度マップ 5 0 にフィルター 5 2 を掛け合わせて得た第 1 特徴データ y_{11} 、第 1 特徴データ y_{12} 、 \dots 、第 1 特徴データ y_{ij} 、 \dots 、第 1 特徴データ y_{77} によって構成されるマトリクス状に配置した第 1 特徴マップ 5 4 (図 4 (C)) を得る。これにより、畳込みニューラルネットワーク部 2 4 は、 9×9 の濃度マップ 5 0 を、 7×7 の第 1 特徴マップ 5 4 にする畳込み処理を行ったことになる。

30

【 0 0 4 8 】

更に図 5 (A) ~ (C) を参照して、畳込みニューラルネットワーク部 2 4 による畳込み処理の具体例を説明する。図 5 (A) に示す濃度マップ 5 0 は、図 3 (B) に示す濃度マップ 5 0 の具体例である。図 5 (A) に示すフィルター 5 2 は、図 3 (B) に示すフィルター 5 2 の具体例である。

【 0 0 4 9 】

ここでは、濃度マップ 5 0 を分かりやすく説明するために、濃度マップ 5 0 を構成する各画素の値を 2 値「 1 」及び「 - 1 」のいずれかで示した例を用いて説明する。また、フィルター 5 2 も、 2 値「 1 」及び「 - 1 」のいずれかで示した例を用いて説明する。

40

【 0 0 5 0 】

図 5 (A) に示す濃度マップ 5 0 は、画素 $x_{11} = - 1$ 、画素 $x_{12} = - 1$ 、画素 $x_{13} = - 1$ 、画素 $x_{21} = - 1$ 、画素 $x_{22} = 1$ 、画素 $x_{23} = 1$ 、画素 $x_{31} = - 1$ 、画素 $x_{32} = - 1$ 、および画素 $x_{33} = - 1 \dots$ とした例を示す。

【 0 0 5 1 】

図 5 (A) に示すフィルター 5 2 は、図 3 (C) に示すフィルター 5 2 の画素 w_{ij} を 2 値の「 1 」か「 - 1 」のいずれかで表す。一例として、画素 $w_{11} = 1$ 、画素 $w_{12} = - 1$ 、画素 $w_{13} = - 1$ 、画素 $w_{21} = - 1$ 、画素 $w_{22} = 1$ 、画素 $w_{23} = - 1$ 、画素 $w_{31} = - 1$ 、画素 $w_{32} = 1$ 、および画素 $w_{33} = - 1$ である。

50

【 0 0 5 2 】

畳込みニューラルネットワーク部 2 4 は、濃度マップ 5 0 の画素 x_{ij} とフィルター 5 2 の画素 w_{ij} とを掛け合わせ、図 4 (B) において説明したように、第 1 特徴データ $y_{ij} = x_{ij} \times w_{ij}$ を得る。具体的には、第 1 特徴データ $y_{11} = -1 \times 1 + -1 \times -1 + -1 \times -1 + -1 \times -1 + 1 \times 1 + 1 \times -1 + -1 \times -1 + -1 \times 1 + -1 \times -1 = 3$ 、第 1 特徴データ $y_{12} = -1 \times 1 + -1 \times -1 + -1 \times -1 + 1 \times -1 + 1 \times 1 + 1 \times -1 + -1 \times -1 + -1 \times 1 + -1 \times -1 = 1$ 、 \dots である。

【 0 0 5 3 】

畳込みニューラルネットワーク部 2 4 は、第 1 特徴データ $y_{11} = 3$ 、第 1 特徴データ $y_{12} = 1$ 、 \dots 、および第 1 特徴データ $y_{77} = -3$ をマトリクス状に並べ、図 5 (B) に示すように、第 1 特徴マップ 5 4 を生成する。

10

【 0 0 5 4 】

さらに、畳込みニューラルネットワーク部 2 4 は、図 5 (A) に示す例ではフィルター 5 2 が 3×3 の 9 個のマトリクスからなることから、第 1 特徴マップ 5 4 を構成する各第 1 特徴データ y_{ij} を例えば $1/9$ の値に変換して、第 2 特徴マップ 5 6 を生成する。つまり、第 2 特徴マップ 5 6 が第 2 特徴データ z_{ij} により構成されるとすると、第 2 特徴データ $z_{ij} =$ 第 1 特徴データ $y_{ij} \times 1/9$ である。

【 0 0 5 5 】

具体的には、第 2 特徴データ $z_{11} = 3 \times 1/9 = 0.33$ 、第 2 特徴データ $z_{12} = 1 \times 1/9 = 0.11$ 、 \dots 、第 2 特徴データ $z_{77} = -3 \times 1/9 = -0.33$ である。畳込みニューラルネットワーク部 2 4 は、これらの第 2 特徴データ z_{ij} をマトリクス状に並べ、図 5 (C) に示すように、第 2 特徴マップ 5 6 を生成する。以上が畳込みニューラルネットワーク部による畳込み処理である。

20

【 0 0 5 6 】

次に、畳込みニューラルネットワーク部 2 4 によるプーリング処理を説明する。図 6 (A) に示すように、畳込みニューラルネットワーク部 2 4 は、第 2 特徴マップ 5 6 を複数の特徴マトリクス 5 6 A、特徴マトリクス 5 6 B、 \dots 、に分割し、それぞれの特徴マトリクス 5 6 A、特徴マトリクス 5 6 B、 \dots 、から代表値 5 6 C、代表値 5 6 D、 \dots 、を抽出する。

【 0 0 5 7 】

具体的には、畳込みニューラルネットワーク部 2 4 は、第 2 特徴マップ 5 6 を特徴マトリクス 5 6 A、5 6 B、 \dots 、の 2×2 のマトリクスに分割する。ただし、2 列に満たない場合は、 1×2 のマトリクスに分割する。そして、畳込みニューラルネットワーク部 2 4 は、特徴マトリクス 5 6 A から、特徴マトリクス 5 6 A の代表値 5 6 C を抽出する。ここでは、畳込みニューラルネットワーク部 2 4 は、代表値として、特徴マトリクス 5 6 A の最大値を抽出するものとする。例えば、代表値 5 6 C は、特徴マトリクス 5 6 A の最大値 0.33 である。但し、畳込みニューラルネットワーク部 2 4 は、代表値を、最大値ではなく、例えば、平均値又は中央値として抽出するようにしてもよい (本願明細書の全編に亘って同様) 。

30

【 0 0 5 8 】

続けて、畳込みニューラルネットワーク部 2 4 は、特徴マトリクス 5 6 B を構成する各値の中から、特徴マトリクス 5 6 B の代表値 5 6 D を抽出する。ここでは、畳込みニューラルネットワーク部 2 4 は、代表値として、特徴マトリクス 5 6 B の最大値を抽出するものとする。例えば、代表値 5 6 D は、特徴マトリクス 5 6 B の最大値 0.11 である。

40

【 0 0 5 9 】

畳込みニューラルネットワーク部 2 4 は、上記のようにして抽出した各代表値 (代表値 5 6 C、代表値 5 6 D、 \dots) を配置したマトリクスを生成する。以上がプーリング処理である。

【 0 0 6 0 】

更に、(i) 畳込みニューラルネットワーク部 2 4 は、予め定められた数の各代表値から構

50

成されるマトリクス（図6（B）に、予め定められた数を3×3としたマトリクス58の例を示す）が得られるまで、畳込み処理とプーリング処理を更に繰り返す。畳込みニューラルネットワーク部24は、当該マトリクスをなす各代表値を一次元的に展開して、図9（A）に一例を示す分類データ62を生成する。(ii)或いは、畳込みニューラルネットワーク部24は、(i)の場合よりも大きな予め定められた第2の数の各代表値から構成されるマトリクスが得られるまで畳込み処理とプーリング処理を更に繰り返し、当該マトリクスが得られた時点で全結合層の処理を行って、上記予め定められた数の各代表値から構成されるマトリクス（図6（B））を作成し、このマトリクスをなす各代表値を一次元的に展開して、図9（A）に一例を示す分類データ62を生成する。以上が、畳込みニューラルネットワーク部24による処理である。

10

【0061】

次に、図2に加え、図7、図8および図11を参照して、グラッドカム処理について説明する。グラッドカム部26は、グラッドカム処理（GradCAM: Gradient-weighted Class Activation Mapping）を行う。グラッドカム部26は、グラッドカム処理として、畳込みニューラルネットワーク部24により生成された第2特徴マップ56に、活性化関数ReLU（Rectified Linear Unit）を適用する処理を行って、アクティベーションマップ66（図8）を生成する。

【0062】

活性化関数ReLUは、図7に示すように、0未満の出力値を全て0にする関数である。すなわち、活性化関数ReLUは、ある閾値以上の部分だけを意味のある情報とする処理、すなわち、当該ある閾値以上の部分を特徴部分として強調する処理である。活性化関数ReLUは、横軸に y_{ij} をとり、縦軸に $f(y_{ij})$ をとる。 $y_{ij} < 0$ のとき、 $f(y_{ij}) = 0$ であり、 $y_{ij} \geq 0$ のとき、 $f(y_{ij}) = y_{ij}$ である。グラッドカム部26は、第2特徴マップ56の各値に活性化関数ReLUを適用して、適用後の各値からなるマトリクス状のアクティベーションマップ66を生成する。

20

【0063】

図8は、アクティベーションマップ66の一例を示す図である。アクティベーションマップ66は、グラッドカム部26により第2特徴マップ56を構成する各第2特徴データ z_{ij} を更に強調して示されることにより、原稿Gを読み取った文書データにおける特徴部分が強調されたものとなる。

30

【0064】

アクティベーションマップ66は、上記原稿Gの会社名の記載部分の画像を含め、原稿Gの会社名の記載部分以外の部分も強調表示している。そのため、補正部34は、アクティベーションマップ66が、画像種別判定の根拠となるべき画像部分、例えば、原稿Gの会社名の記載部分の画像部分のみを的確に強調表示したものとなるように（全体画像において、どの画像部分が「会社名の記載部分」を示すかを明確にするために）、畳込みニューラルネットワーク部24が用いる上記フィルター52の重み付けである画素 w_{ij} を補正する。

【0065】

ここで、グラッドカム部26は、上述した分割処理を行って、アクティベーションマップ66を複数のグループに分割する。そして、グラッドカム部26は、当該複数のグループについてそれぞれ補正関数 $f(\text{GradCAM})$ を算出する。補正関数 $f(\text{GradCAM})$ は、数式(1)で表される。

40

【数1】

$$f(\text{GradCAM}) = \sum \text{Act}(r)(r - c)^2$$

・・・(1)

50

【 0 0 6 6 】

数式(1)の補正関数 f (G r a d C A M) では、上記グループを構成する画素のうち特定の画素を画素 c とし、上記グループを構成する画素であって画素 c 以外の全ての画素をそれぞれ画素 r とする。補正関数 f は、アクティベーション関数 A c t (r) と、画素 c と各画素 r との距離の 2 乗を表す距離関数 (r - c) ^ 2 との積で表される。すなわち、補正関数 f は、上記グループの画素 c と全ての画素 r との各組み合わせについて、アクティベーション関数 A c t (r) と距離関数 (r - c) ^ 2 との積をそれぞれ算出し、これら算出した全ての積の和を更に算出したものである。グラッドカム部 2 6 は、分割した上記複数のグループ毎に補正関数 f を算出する。

【 0 0 6 7 】

ここでは、一例として、上記グループの最大濃度の画素を画素 c とし、任意の画素を画素 r とする。例えば、アクティベーション関数 A c t (r) は、画素 r におけるアクティベーションマップ 6 6 (上記グループ) の反応の大きさを表わす。

【 0 0 6 8 】

例えば、アクティベーション関数 A c t (r) として、画素 r の濃度を用いる。すなわち、画素 r の濃度の高低に応じてアクティベーション関数 A c t (r) の値は変化する。また、画素 c と画素 r との距離が大きいほど距離関数 (r - c) ^ 2 の値は大きく、画素 c と画素 r との距離が小さいほど距離関数 (r - c) ^ 2 の値は小さくなる。

【 0 0 6 9 】

続いて、図 9 を参照して、記憶部 2 8 、比較部 3 0 を説明する。

【 0 0 7 0 】

上述したように、記憶部 2 8 は、教師データ 6 0 (図 9 (A)) を記憶している。教師データ 6 0 は、図 9 (A) に示すように、上記分類データと同じ数の値が一次的に並べられた数列からなる。記憶部 2 8 は、原稿の種別を判別するための注目画像となる各画像に対応する教師データ 6 0 を記憶している。

【 0 0 7 1 】

比較部 3 0 は、畳込みニューラルネットワーク部 2 4 から取得した分類データ 6 2 と、記憶部 2 8 が記憶する教師データ 6 0 とを比較し、差分データ 6 4 を算出する。更に、比較部 3 0 は、上記分類データ 6 2 と上記教師データ 6 0 を用いて下記数式 (2) により第 1 損失関数 (L o s s f u n c t i o n) を算出する。

【 数 2 】

$$Loss\ function = - \sum_{x \in X} p(x) \log q(x)$$

・・・ (2)

【 0 0 7 2 】

そして、出力部 3 2 は、グラッドカム部 2 6 から補正関数 f (G r a d C A M) (数式 1) の値を取得し、比較部 3 0 によって算出された第 1 損失関数 (数式 (2)) と補正関数 f (G r a d C A M) の値との和を第 2 損失関数 (L o s s f u n c t i o n) として算出する。すなわち、第 2 損失関数は、数式 (3) で表される。出力部 3 2 は、第 2 損失関数を、上記各グループの補正関数 f をそれぞれ用いて当該グループ毎に算出する。

【 数 3 】

$$Loss\ function = - \sum_{x \in X} p(x) \log q(x) + f(GradCAM)$$

・・・ (3)

【 0 0 7 3 】

補正部 3 4 は、出力部 3 2 が出力した上記各グループの第 2 損失関数を合計したものを

補正值とし、この補正值を用いて、畳込みニューラルネットワーク部 24 で用いる上記フィルター 52 の重み付けを補正する。これにより、補正部 34 は、該補正後のフィルター 52 を作成する。この第 2 損失関数は、アクティベーションマップ 66 において、上記判定の根拠とすべき原稿 G の会社名の記載部分の画像を的確に強調したものに基づくものであるため、補正部 34 が、第 2 損失関数を用いてフィルター 52 の重み付けを補正すると、畳込みニューラルネットワーク部 24 により将来生成される分類データ 62 (図 9 (A)) が教師データ 60 に近づくよう、フィルター 52 (図 9 (B)) の画素 w_{ij} の数値を補正することになる。このようにフィルター 52 を補正したフィルターをフィルター 52 B (図 9 (B)) とする。

【0074】

具体的には、図 10 (A) に示すように、比較部 30 は、分類データ 62 A を取得すると、教師データ 60 A と分類データ 62 A との差分をとって、差分データ 64 A を算出する。また、比較部 30 は、教師データ 60 A と分類データ 62 A から第 1 損失関数 (数式 2) を算出する。出力部 32 は、上記グループ毎に、第 1 損失関数と、上記グループ別の補正関数 f の値との和を算出し、この和を第 2 損失関数 (数式 3) とする。補正部 34 は、各グループの第 2 損失関数を合計して上記補正值を算出する。補正部 34 は、当該補正值に基づき (例えばフィルター 52 に補正值を乗算して)、図 9 (B) に示すように、畳込みニューラルネットワーク部 24 のフィルター 52 を、例えば、フィルター 52 B のように補正する。

【0075】

このようにフィルター 52 がフィルター 52 B に補正されると、畳込みニューラルネットワーク部 24 が、補正後のフィルター 52 B を用いて、新たな分類データ (一例を図 10 (B) に分類データ 62 B として示す) を生成し、グラッドカム部 26 は新たな補正関数 f を上記グループ毎に算出する。そして、比較部 30 は、教師データ 60 B と分類データ 62 B とを比較して、差分データ 64 B を出力する。このとき、分類データ 62 B は、分類データ 62 A と比較して、教師データ 60 B に近似することになる。更に比較部 30 は、上記分類データ 62 B と教師データ 60 B を用いて第 1 損失関数を算出し、出力部 32 は、グラッドカム部 26 が上記グループ毎に算出した上記新たな補正関数 f を用いて、新たな第 2 損失関数を上記グループ毎に算出する。補正部 34 は、グループ毎の新たな第 2 損失関数を合計した補正值を用いてフィルター 52 B を補正してフィルター 52 C を作成する。

【0076】

畳込みニューラルネットワーク部 24 は、補正後の各フィルター 52 C を用いてそれぞれに、更に新しい分類データ (一例を図 10 (C) に分類データ 62 C として示す) を生成する。グラッドカム部 26 は更に新たな補正関数 f を上記グループ毎に算出する。そして、比較部 30 は、教師データ 60 C と分類データ 62 C とを比較して、差分データ 64 C を出力する。分類データ 64 C は、分類データ 62 B と比較して、さらに教師データ 60 C に近似する (図 10 (C) には、教師データ 60 C と分類データ 62 C とが一致する例を示している)。図 10 (C) に示す例では、差分データ 64 C は、0 のデータのみをもつことになる。

【0077】

図 11 (A) ~ (C) は、上述のように畳込みニューラルネットワーク部 24 から補正部 34 までによるフィルター補正処理、及び補正された新たなフィルターを用いた畳込みニューラルネットワーク部 24 及びグラッドカム部 26 によるアクティベーションマップ生成処理を上記グループ毎に繰り返すことで、グラッドカム部 26 が生成する各グループのアクティベーションマップ 66 が補正されていく様子を示す図である。

【0078】

例えば、図 11 (A) に示す、上記複数のグループのうちの 1 つのグループについての第 1 アクティベーションマップ 66 A は、原稿 G の会社名の記載部分の画像のみでなく、原稿 G の全体に亘って高い濃度を示しているが、上記のようにしてフィルターが補正され

10

20

30

40

50

、畳込みニューラルネットワーク部 2 4 が、濃度マップ 5 0 に新たなフィルター 5 2 B を用いて新たな第 1 特徴マップ 5 4 及び新たな第 2 特徴マップ 5 6 を生成し、グラッドカム部 2 6 が新たな第 2 特徴マップ 5 6 から第 2 アクティベーションマップ 6 6 B を生成すると、第 2 アクティベーションマップ 6 6 B では、図 1 1 (B) に示すように、第 1 アクティベーションマップ 6 6 A と比較して、表題部分からの距離が大きい画素の濃度が大きく減少する。

【 0 0 7 9 】

そして更に、畳込みニューラルネットワーク部 2 4 から補正部 3 4 までによるフィルター補正処理でフィルター 5 2 B がフィルター 5 2 C に補正され、補正された更に新たなフィルターを用いた畳込みニューラルネットワーク部 2 4 及びグラッドカム部 2 6 によるアクティベーションマップ生成処理により、第 3 アクティベーションマップ 6 6 C が生成されると、第 3 アクティベーションマップ 6 6 C では、例えば、図 1 1 (C) に示すように、第 2 アクティベーションマップ 6 6 B と比較して、表題部分からの距離が大きい画素の濃度が更に大きく減少する。図 1 1 (C) には、第 2 アクティベーションマップ 6 6 B と比較して、原稿 G の会社名の記載部分の画像を示す画素からの距離が大きい画素の濃度はほぼ 0 に近くなり、原稿 G の会社名の記載部分の画像の濃度がより強調された例を示している。

10

【 0 0 8 0 】

このように、畳込みニューラルネットワーク部 2 4 から補正部 3 4 までによるフィルター補正処理を繰り返し、例えば、教師データと分類データとが一致して差分データが 0 になった時点（或いは、教師データと分類データとが予め定められた範囲内まで近似した時点）で当該繰り返しを終了し、この時点で補正されている最新のフィルターを、畳込みニューラルネットワーク部 2 4 が上記畳込み処理に用いるフィルターとして更新する。以上は、画像種別判定の根拠とすべき画像を抽出するための学習処理である。

20

【 0 0 8 1 】

また、画像種別判定の処理においては、分類部 3 6 は、上記のように補正されて更新されている最新のフィルター 5 2 を用いて畳込みニューラルネットワーク部 2 4 により原稿画像（読取部 4 による原稿読取で得た画像）から生成された分類データが示す各値の配列によって、画像の種別を判定する。例えば、分類部 3 6 は、（ A ）分類データの第 1 番目のみが 0 よりも大きな値を示し、他の各値が可全て 0 の場合は、画像種別判定の根拠とすべき画像部分が「 A B C D 会社」であるとして「 A B C D 会社宛書類」と判定し、（ B ）分類データの第 2 番目のみが 0 よりも大きな値を示し、他の各値が可全て 0 の場合は、画像種別判定の根拠とすべき画像部分が「 E F G H 会社」であるとして「 E F G H 会社宛書類」と判定する、等のようにして画像の種別判定を行う。

30

【 0 0 8 2 】

従って、本実施形態に示した学習処理を終えれば、この後に畳込みニューラルネットワーク部 2 4 により生成された分類データは教師データに一致又は近似するため、画像の種別を判定させるとき、画像種別判定の根拠とすべき画像部分としてのオブジェクトが単数ではなく複数存在する場合であっても、当該各オブジェクトを的確に抽出して、当該画像の種別を判定する精度を高く保つことができる。

40

【 0 0 8 3 】

また、本実施形態では、画像種別判定の処理を行う度に、上記フィルター 5 2 の補正処理までの学習処理も行うようにすれば、画像形成装置 1 0 0 による原稿 G の読み取りの件数、および画像処理の件数が増えるに従って、畳込みニューラルネットワーク部 2 4 のフィルターがより好適に補正されていくため、画像の種別の判定精度を向上させることができる。

【 0 0 8 4 】

また、図 1 乃至図 1 1 を用いて説明した上記実施形態の構成及び処理は、本発明の一実施形態に過ぎず、本発明を当該構成及び処理に限定する趣旨ではない。

【 0 0 8 5 】

50

また、上記実施形態では、本発明に係る画像処理の一実施形態を、画像形成装置（複合機）に適用する例を説明しているが、これは一例に過ぎず、本発明に係る画像処理を、他の電子装置、例えば、医療機器、パーソナルコンピュータ、携帯電話、スマートフォン、タブレット、ハブ装置、サーバー装置に適用するものとしてもよい。

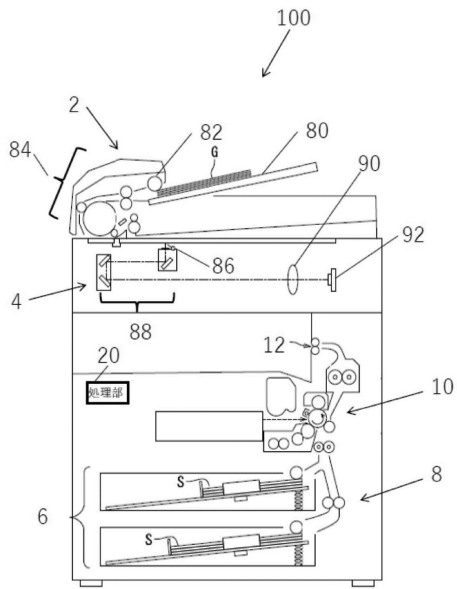
【符号の説明】

【 0 0 8 6 】

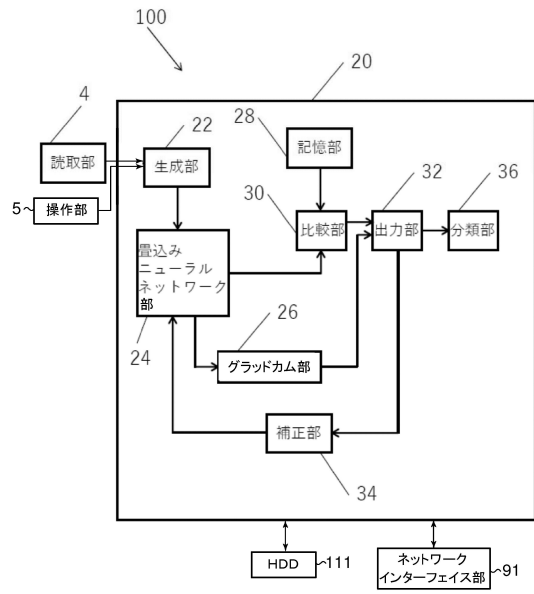
1 0 0	画像形成装置	
4	読取部	
2 0	処理部	
2 2	生成部	10
2 4	畳込みニューラルネットワーク部	
2 6	グラッドカム部	
2 8	記憶部	
3 0	比較部	
3 2	出力部	
3 4	補正部	
3 6	分類部	
		20
		30
		40
		50

【図面】

【図 1】



【図 2】



10

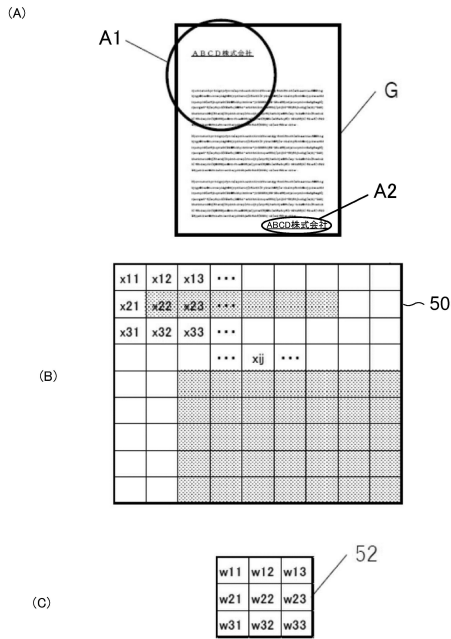
20

30

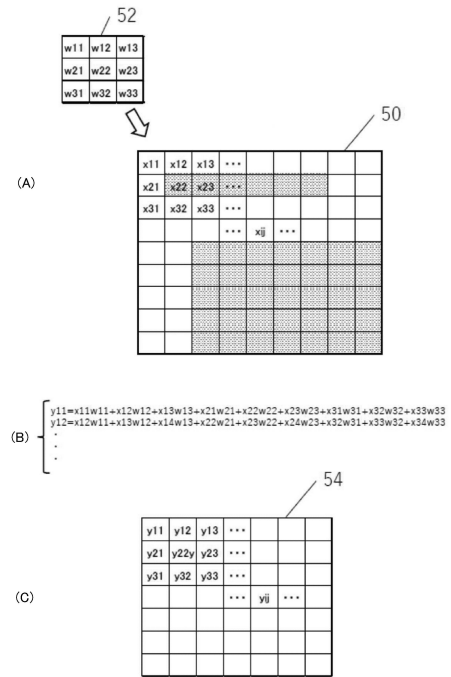
40

50

【 図 3 】



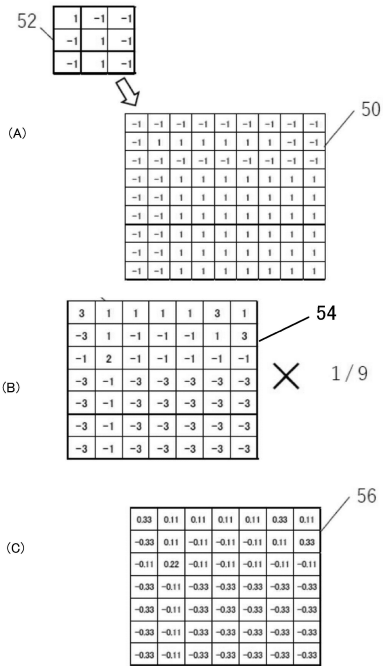
【 図 4 】



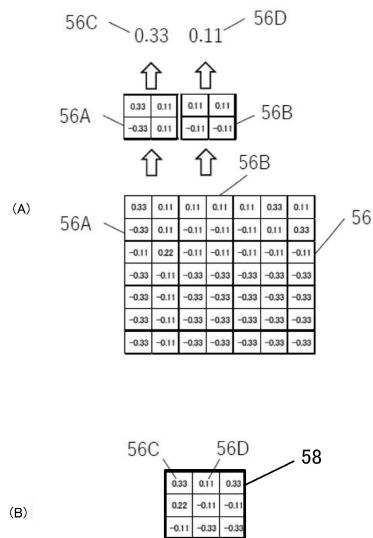
10

20

【 図 5 】



【 図 6 】

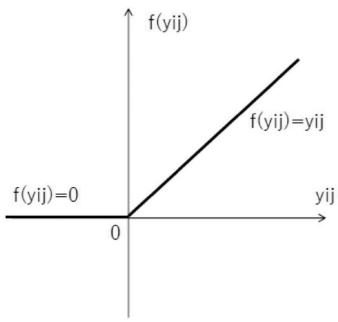


30

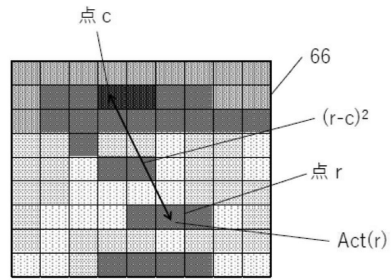
40

50

【 図 7 】



【 図 8 】

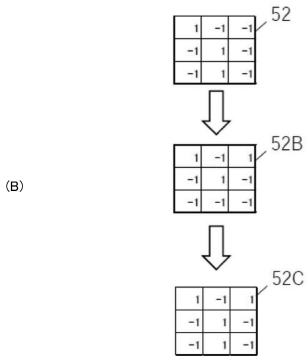


10

【 図 9 】

(A)

(教師データ)	0	1	0	0	0	0	0	0	0	0	0	60
(分類データ)	0.5	0.5	0	0	0	0	0	0	0	0	0	62
(差分データ)	-0.5	0.5	0	0	0	0	0	0	0	0	0	64



【 図 10 】

(A)

(教師データ)	0	1	0	0	0	0	0	0	0	0	0	60,60A
(分類データ)	0.5	0.5	0	0	0	0	0	0	0	0	0	62,62A
(差分データ)	-0.5	0.5	0	0	0	0	0	0	0	0	0	64,64A

20

(B)

(教師データ)	0	1	0	0	0	0	0	0	0	0	0	60,60B
(分類データ)	0.2	0.8	0	0	0	0	0	0	0	0	0	62,62B
(差分データ)	-0.2	0.2	0	0	0	0	0	0	0	0	0	64,64B

(C)

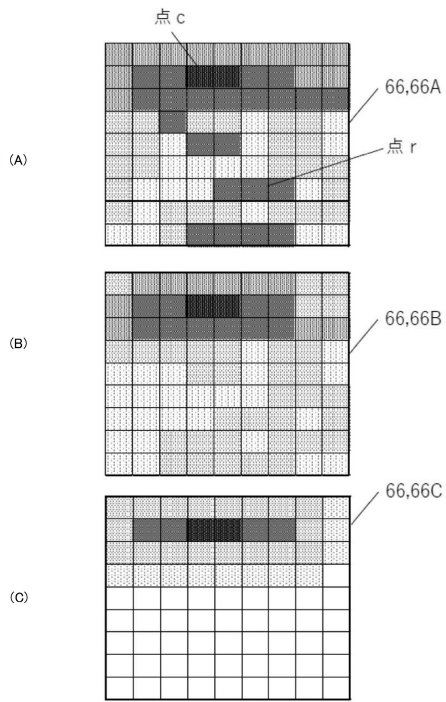
(教師データ)	0	1	0	0	0	0	0	0	0	0	0	60,60C
(分類データ)	0	1	0	0	0	0	0	0	0	0	0	62,62C
(差分データ)	0	0	0	0	0	0	0	0	0	0	0	64,64C

30

40

50

【 図 1 1 】



10

20

30

40

50

フロントページの続き

- (56)参考文献 特開 2019 - 133320 (JP, A)
特開 2020 - 184305 (JP, A)
米国特許出願公開第 2018 / 0260793 (US, A1)
中国特許出願公開第 110458221 (CN, A)
- (58)調査した分野 (Int.Cl., DB名)
G06T 7/00