

(19)日本国特許庁(JP)

(12)特許公報(B1)

(11)特許番号

特許第7112151号

(P7112151)

(45)発行日 令和4年8月3日(2022.8.3)

(24)登録日 令和4年7月26日(2022.7.26)

(51)国際特許分類

G 0 6 N 3/04 (2006.01)

F I

G 0 6 N 3/04

請求項の数 11 (全36頁)

(21)出願番号 特願2022-76897(P2022-76897)

(22)出願日 令和4年5月9日(2022.5.9)

審査請求日 令和4年5月10日(2022.5.10)

早期審査対象出願

(73)特許権者 518334141

E A G L Y S 株式会社

東京都渋谷区千駄ヶ谷 5-27-3 や

まとビル 7F

(74)代理人 100210815

弁理士 西田 聡子

(72)発明者 三原 健太郎

東京都渋谷区千駄ヶ谷 5-27-3 や

まとビル 7F E A G L Y S 株式会社内

審査官 杉浦 孝光

最終頁に続く

(54)【発明の名称】 機械学習装置、機械学習システム、機械学習方法、および機械学習プログラム

(57)【特許請求の範囲】

【請求項1】

画像データを取得するデータ取得部と、

前記画像データに基づいて、畳み込みニューラルネットワークを用いた計算処理を行う
機械学習計算部と、を備え、

前記畳み込みニューラルネットワークは複数の層から構成され、

前記機械学習計算部は、

準同型暗号による暗号実行領域において演算する暗号処理部と、

平文実行領域において演算する平文処理部と、

前記暗号処理部および前記平文処理部がアクセス可能な共有メモリと、

前記畳み込みニューラルネットワークの入力層から出力層に向けての途中の層である
第N層を指定する層指定情報を受け付ける受付部と、を有し、

前記暗号処理部は、前記入力層から前記第N層までの演算を実行し、

前記平文処理部は、第(N+1)層から前記出力層までの演算を実行し、

前記共有メモリは、前記暗号処理部および前記平文処理部が実行した演算の結果を記憶する機械学習装置。

【請求項2】

前記暗号処理部は、前記入力層から前記第N層までの演算の結果を前記共有メモリに記憶させ、

前記平文処理部は、前記共有メモリから、前記暗号処理部が実行した前記第N層までの

演算の結果を取得し、前記取得した演算の結果に基づいて、前記第（ $N + 1$ ）層から前記出力層までの演算を実行する、請求項 1 に記載の機械学習装置。

【請求項 3】

前記暗号処理部は、前記入力層から前記第 N 層までに関する第 1 のモデル情報を用いて、前記入力層から前記第 N 層までの前記演算を実行し、

前記平文処理部は、第（ $N + 1$ ）層から前記出力層までに関する第 2 のモデル情報を用いて、前記第（ $N + 1$ ）層から前記出力層までの演算を実行する、請求項 1 または請求項 2 に記載の機械学習装置。

【請求項 4】

画像データを取得するデータ取得部と、

前記画像データに基づいて、畳み込みニューラルネットワークを用いた計算処理を行う機械学習計算部と、を備え、

前記畳み込みニューラルネットワークは複数の層から構成され、

前記機械学習計算部は、

準同型暗号による暗号実行領域において演算する暗号処理部と、

平文実行領域において演算する平文処理部と、

前記畳み込みニューラルネットワークの入力層から出力層に向けての途中の層である第 N 層を指定する層指定情報を受け付ける受付部と、を有し、

前記暗号処理部は、前記入力層から前記第 N 層までの演算を実行し、

前記平文処理部は、第（ $N + 1$ ）層から前記出力層までの演算を実行し、

前記暗号処理部と、前記平文処理部とは、物理的に異なる実行環境において演算を実行する機械学習システム。

【請求項 5】

前記データ取得部および前記暗号処理部を含む暗号処理装置と、

前記平文処理部を含む平文処理装置と、をさらに備え、

前記暗号処理装置と、前記平文処理装置とは、ネットワークを介してデータを送受信する、請求項 4 に記載の機械学習システム。

【請求項 6】

前記暗号処理装置は、前記入力層から前記第 N 層までに関する第 1 のモデル情報を取得し、

前記平文処理装置は、前記第（ $N + 1$ ）層から前記出力層までに関する第 2 のモデル情報を取得し、

前記暗号処理部は、前記第 1 のモデル情報を用いて、前記入力層から前記第 N 層までの前記演算を実行し、

前記平文処理部は、前記第 2 のモデル情報を用いて、前記第（ $N + 1$ ）層から前記出力層までの演算を実行する、請求項 5 に記載の機械学習システム。

【請求項 7】

少なくとも 2 以上の前記暗号処理装置と、少なくとも 2 以上の前記平文処理装置とを備え、

前記暗号処理装置と前記平文処理装置とは、少なくとも一組のペアリングをされており、

前記機械学習計算部は、前記ペアリングされた前記暗号処理装置と前記平文処理装置とを用いて前記計算処理を実行する、

請求項 5 または請求項 6 に記載の機械学習システム。

【請求項 8】

前記ペアリングされていない前記暗号処理装置および前記平文処理装置は、前記計算処理とは異なる処理に関する通信を行う、請求項 7 に記載の機械学習システム。

【請求項 9】

画像データを取得するデータ取得部と、

前記画像データに基づいて、畳み込みニューラルネットワークを用いた計算処理を行う機械学習計算部と、を備え、

10

20

30

40

50

前記畳み込みニューラルネットワークは複数の層から構成され、
前記機械学習計算部は、
準同型暗号による暗号実行領域において演算する暗号処理部と、
平文実行領域において演算する平文処理部と、
前記暗号処理部および前記平文処理部がアクセス可能な共有メモリと、
前記畳み込みニューラルネットワークの入力層から出力層に向けての途中の層である
第 N 層を指定する層指定情報を受け付ける受付部と、を有し、
前記暗号処理部は、前記入力層から前記第 N 層までの演算を実行し、
前記平文処理部は、第 $(N + 1)$ 層から前記出力層までの演算を実行し、
前記共有メモリは、前記暗号処理部および前記平文処理部が実行した演算の結果を記憶する機械学習システム。

10

【請求項 10】

画像データを取得するステップと、
前記画像データに基づいて、畳み込みニューラルネットワークを用いた計算処理を行う
ステップと、を備え、
前記畳み込みニューラルネットワークは複数の層から構成され、
前記畳み込みニューラルネットワークを用いた計算処理を行うステップは、
前記畳み込みニューラルネットワークの入力層から出力層に向けての途中の層である
第 N 層を指定する層指定情報を受け付けるステップと、
準同型暗号による暗号実行領域において、前記入力層から前記第 N 層までの演算を実
行するステップと、
平文実行領域において、第 $(N + 1)$ 層から前記出力層までの演算を実行するステッ
プと、
前記暗号実行領域および前記平文実行領域からアクセス可能な共有メモリが、前記暗
号実行領域および前記平文実行領域における演算の結果を記憶するステップと、を有する
機械学習方法。

20

【請求項 11】

制御部と、記憶部と、を備えるコンピュータに実行させる機械学習プログラムであって、
前記制御部が、画像データを取得するステップと、
前記記憶部が、前記画像データを記憶するステップと、
前記制御部が、前記記憶部が記憶した前記画像データに基づいて、畳み込みニューラル
ネットワークを用いた計算処理を行うステップと、を備え、
前記畳み込みニューラルネットワークは複数の層から構成され、
前記畳み込みニューラルネットワークを用いた計算処理を行うステップは、
前記制御部が、前記畳み込みニューラルネットワークの入力層から出力層に向けての
途中の層である第 N 層を指定する層指定情報を受け付けるステップと、
前記制御部が、準同型暗号による暗号実行領域において、前記入力層から前記第 N 層
までの演算を実行するステップと、
前記制御部が、平文実行領域において、第 $(N + 1)$ 層から前記出力層までの演算を
実行するステップと、を有し、
前記記憶部は、前記暗号実行領域および前記平文実行領域からアクセス可能な共有メ
モリを含み、
前記共有メモリが、前記暗号実行領域および前記平文実行領域における演算の結果を
記憶するステップ、をさらに有する機械学習プログラム。

30

40

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、機械学習装置、機械学習システム、機械学習方法、および機械学習プログラ
ムに関する。

【背景技術】

50

【0002】

近年、畳み込みニューラルネットワークの学習モデルを用いた画像認識により、複雑なデータの分析を高精度に行うことが可能となっている。

【0003】

一方、分析対象が医療画像など個人のプライバシーに直結した機微データである場合は、プライバシー情報を秘匿（暗号化）したまま分析処理を行うことが求められる。

【0004】

例えば、特許文献1では、準同型暗号を用いたニューラルネットワークのシステムに関する技術が開示されている。

【先行技術文献】

【特許文献】

【0005】

【文献】特開2019-46460号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

準同型暗号を用いた畳み込みニューラルネットワークにより画像認識を行えば、プライバシー情報を暗号化したまま分析処理を行うことができる。しかしながら、一般的に、準同型暗号を用いた演算は、セキュリティレベルは高いものの、計算量が膨大になってしまい、実用的な計算コストで実行することが難しいという課題があった。

【0007】

そこで、本開示は、上記課題を解決すべくなされたものであって、その目的は、ユーザの求める実用的な計算コストとセキュリティレベルに応じた、畳み込みニューラルネットワークを用いた機械学習に関する技術を提供することである。

【課題を解決するための手段】

【0008】

上記目的を達成するため、本開示に係る機械学習装置は、画像データを取得するデータ取得部と、画像データに基づいて、畳み込みニューラルネットワークを用いた計算処理を行う機械学習計算部と、を備え、畳み込みニューラルネットワークは複数の層から構成され、機械学習計算部は、準同型暗号による暗号実行領域において演算する暗号処理部と、平文実行領域において演算する平文処理部と、暗号処理部および平文処理部がアクセス可能な共有メモリと、畳み込みニューラルネットワークの入力層から出力層に向けての途中の層である第N層を指定する層指定情報を受け付ける受付部と、を有し、暗号処理部は、入力層から第N層までの演算を実行し、平文処理部は、第(N+1)層から出力層までの演算を実行し、共有メモリは、暗号処理部および平文処理部が実行した演算の結果を記憶する。

【0009】

また、上記目的を達成するため、本開示に係る機械学習システムは、画像データを取得するデータ取得部と、画像データに基づいて、畳み込みニューラルネットワークを用いた計算処理を行う機械学習計算部と、を備え、畳み込みニューラルネットワークは複数の層から構成され、機械学習計算部は、準同型暗号による暗号実行領域において演算する暗号処理部と、平文実行領域において演算する平文処理部と、暗号処理部および平文処理部がアクセス可能な共有メモリと、畳み込みニューラルネットワークの入力層から出力層に向けての途中の層である第N層を指定する層指定情報を受け付ける受付部と、を有し、暗号処理部は、入力層から第N層までの演算を実行し、平文処理部は、第(N+1)層から出力層までの演算を実行し、共有メモリは、暗号処理部および平文処理部が実行した演算の結果を記憶する。

【0010】

また、上記目的を達成するため、本開示に係る機械学習方法は、画像データを取得するステップと、画像データに基づいて、畳み込みニューラルネットワークを用いた計算処理

10

20

30

40

50

を行うステップと、を備え、畳み込みニューラルネットワークは複数の層から構成され、畳み込みニューラルネットワークを用いた計算処理を行うステップは、畳み込みニューラルネットワークの入力層から出力層に向けての途中の層である第N層を指定する層指定情報を受け付けるステップと、準同型暗号による暗号実行領域において、入力層から第N層までの演算を実行するステップと、平文実行領域において、第(N+1)層から出力層までの演算を実行するステップと、暗号実行領域および平文実行領域からアクセス可能な共有メモリが、暗号実行領域および平文実行領域における演算の結果を記憶するステップと、を有する。

【0011】

また、上記目的を達成するため、本開示に係る機械学習プログラムは、制御部と、記憶部と、を備えるコンピュータに実行させる機械学習プログラムであって、制御部が、画像データを取得するステップと、記憶部が、画像データを記憶するステップと、制御部が、記憶部が記憶した画像データに基づいて、畳み込みニューラルネットワークを用いた計算処理を行うステップと、を備え、畳み込みニューラルネットワークは複数の層から構成され、畳み込みニューラルネットワークを用いた計算処理を行うステップは、制御部が、畳み込みニューラルネットワークの入力層から出力層に向けての途中の層である第N層を指定する層指定情報を受け付けるステップと、制御部が、準同型暗号による暗号実行領域において、入力層から第N層までの演算を実行するステップと、制御部が、平文実行領域において、第(N+1)層から出力層までの演算を実行するステップと、を有し、記憶部は、暗号実行領域および平文実行領域からアクセス可能な共有メモリを含み、共有メモリが、暗号実行領域および平文実行領域における演算の結果を記憶するステップ、をさらに有する。

【発明の効果】

【0012】

本開示によれば、ユーザの求める実用的な計算コストとセキュリティレベルに応じた畳み込みニューラルネットワークを用いた機械学習を実現することができる。

【図面の簡単な説明】

【0013】

【図1】機械学習システム1の全体図である。

【図2】本実施形態の機械学習装置100のハードウェア構成を示す図である。

【図3】本実施形態に係る機械学習装置100の機能的構成を示すブロック図である。

【図4】鍵格納部121の一例を示す図である。

【図5】学習モデル格納部122の一例を示す図である。

【図6】ユーザが画像データとラベルを入力するための画面の一例を示す図である。

【図7】一般的な畳み込みニューラルネットワークのモデル概念図である。

【図8】中間層の一例を示した図である。

【図9】機械学習装置100が実行する処理の一例を示すフローチャートである。

【図10】本実施形態に係る機械学習システム2を示す図である。

【図11】本実施形態における機械学習装置300の機能的構成を示すブロック図である。

【図12】提供部3324が提供する画面の一例を示す図である。

【図13】提供部3324が提供する画面の一例を示す図である。

【図14】機械学習装置300が実行する処理の一例を示すフローチャートである。

【図15】本実施形態に係る機械学習装置400の機能的構成を示すブロック図である。

【図16】共有メモリ431に関し、機能ブロック図とハードウェアとの対応関係を示す図である。

【図17】アカウント制御を用いて実行権限を分けた場合を示す図である。

【図18】機械学習装置400が実行する処理の一例を示すフローチャートである。

【図19】本実施形態に係る機械学習装置500の機能的構成を示すブロック図である。

【図20】機械学習システム3の全体図である。

【図21】本実施形態に係る暗号処理装置600の機能的構成を示すブロック図である。

10

20

30

40

50

【図 2 2】本実施形態に係る平文処理装置 7 0 0 の機能的構成を示すブロック図である。

【図 2 3】本実施形態に係るシステム管理装置 8 0 0 の機能的構成を示すブロック図である。

【図 2 4】物理的に異なる実行環境において畳み込みニューラルネットワークを用いた計算処理を行う場合の具体例を示す図である。

【図 2 5】機械学習システム 3 が実行する処理の一例を示すシーケンス図である。

【図 2 6】機械学習システム 4 の具体的構成を示す図である。

【発明を実施するための形態】

【0 0 1 4】

以下、本開示の実施形態について図面を参照して説明する。実施形態を説明する全図において、共通の構成要素には同一の符号を付し、繰り返しの説明を省略する。なお、以下の実施形態は、特許請求の範囲に記載された本開示の内容を不当に限定するものではない。また、実施形態に示される構成要素のすべてが、本開示の必須の構成要素であるとは限らない。

【0 0 1 5】

<本発明の概要>

近年、畳み込みニューラルネットワークを用いて高精度で画像認識を行うことが可能となり、複雑なデータの分析を行うケースが増加しており、画像認識の対象となる画像データに個人情報等の機微データが含まれる場合は、プライバシーへの配慮が求められる。そこで、畳み込みニューラルネットワークに係る計算過程を、暗号化したまま演算を行うことができる準同型暗号を用いて実行すれば、プライバシーを保護しつつ、画像データの分析処理を行うことができる。

【0 0 1 6】

しかしながら、一般的に、準同型暗号を用いた演算は、セキュリティレベルは高いものの、計算コストが膨大になってしまうという課題がある。そこで、本発明では、当該課題を解決すべく、畳み込みニューラルネットワークを用いた画像認識において、畳み込み演算のフィルタリング処理が繰り返される過程で、画像を目視で識別することが困難になっていくことに着目した。すなわち、本発明では、ユーザが予め定めた計算過程までを、準同型暗号による暗号実行領域において処理し、それ以降の計算過程は平文実行領域において処理を行う。

【0 0 1 7】

目視で画像を識別することが困難になる一定の計算過程以降では、平文実行領域において計算を行うことにより、計算過程の画像が第三者に流出したとしても、当該画像を目視で識別することが困難なため、暗号化されていなくても、機微データの内容を第三者に知られることを防ぐことができる。それと同時に、全ての計算過程を準同型暗号による暗号実行領域において実行しないため、計算コストを抑えることができる。

【0 0 1 8】

<実施形態 1>

(機械学習システム 1 の構成)

図 1 は、機械学習システム 1 の全体図である。図 1 に示すように、本実施形態に係る機械学習システム 1 は、機械学習装置 1 0 0 と、ユーザによって使用される一つ以上の端末装置 2 0 0 とを備える。

【0 0 1 9】

機械学習装置 1 0 0 と、端末装置 2 0 0 とは、ネットワーク NW を介して通信可能に接続される。ネットワーク NW は、WAN (Wide Area Network)、LAN (Local Area Network) 等から構成されるが、任意のネットワークから構成されてもよい。また、機械学習システム 1 を構成する装置等は、ネットワーク NW を介さず、ローカルに接続されてもよい。機械学習システム 1 は、ユーザからの入力に基づいて機械学習の学習モデルを構築し、又はユーザからの入力および学習モデルを用いて、機械学習による推論を行う。

【0 0 2 0】

10

20

30

40

50

端末装置 200 は、ユーザからの入力を受け付け、その入力内容を、ネットワーク NW を介して機械学習装置 100 に送信する。また、端末装置 200 は、機械学習装置 100 からネットワーク NW を介して送信されたデータを受信し、ユーザに提示する。

【0021】

(機械学習装置 100 のハードウェア構成)

図 2 は、本実施形態の機械学習装置 100 のハードウェア構成を示す図である。機械学習装置 100 は、汎用のコンピュータである。機械学習装置 100 は、例えば、サーバコンピュータ、据え置き型の PC (Personal Computer)、ラップトップ PC 等により実現される。

【0022】

図 2 に示すように、機械学習装置 100 は、プロセッサ 11 と、メモリ 12 と、ストレージ 13 と、通信 IF 14 と、入出力 IF 15 と、を備える。

【0023】

プロセッサ 11 は、プログラムに記述された命令セットを実行するためのハードウェアであり、演算装置、レジスタ、周辺回路などにより構成される。メモリ 12 は、プログラム、および、プログラム等で処理されるデータ等を一時的に記憶するためのものであり、例えば DRAM (Dynamic Random Access Memory) 等の揮発性のメモリにより実現される。メモリ 12 に記憶されるプログラムは、本発明に係る機械学習を実行するプログラム等である。ストレージ 13 は、データを保存するための記憶装置であり、例えばフラッシュメモリ、HDD (Hard Disc Drive)、SSD (Solid State Drive) により実現される。通信 IF 14 は、機械学習装置 100 が外部の装置と通信するため、信号を送受信するためのインタフェースである。入出力 IF 15 は、ユーザからの入力を受け付けるための入力装置、および、ユーザに対し情報を提示するための出力装置とのインタフェースとして機能する。

【0024】

なお、端末装置 200 においても、図 2 に示したハードウェア構成と同様の構成としてもよい。端末装置 200 の各構成要素の動作は、上述の機械学習装置 100 と同様に、メモリ 12 に記憶されたプログラムに従ったプロセッサ 11 により実現する。

【0025】

(機械学習装置 100 の機能的構成)

図 3 は、本実施形態に係る機械学習装置 100 の機能的構成を示すブロック図である。図 3 を参照して、実施形態 1 に係る機械学習装置 100 の機能的構成について説明する。

【0026】

機械学習装置 100 は、通信部 110 と、記憶部 120 と、制御部 130 とを備える。

【0027】

通信部 110 は、機械学習装置 100 が他の装置と通信するための処理を行う。ここでいう他の装置とは、ネットワーク NW で接続された PC でもよいし、スマートフォン、タブレットなどの端末である。また、ヘッドマウントディスプレイ等のウェアラブル端末、AR (Augmented Reality) / VR (Virtual Reality) / MR (Mixed Reality) 装置等であってもよい。また、他の装置は、入出力装置、例えば、フラッシュメモリや HDD、SSD などによりデータの入出力を行う装置であってもよい。

【0028】

通信部 110 は、また、ネットワーク NW を介した装置やローカルに接続された装置等と、セキュリティが確保されたセキュアな通信チャネルでデータを送受信する。セキュアな通信チャネルの構築、通信方法は、共通鍵 (セッション鍵など) や公開鍵等を用いた周知の技術であるため、説明を省略する。

【0029】

記憶部 120 は、例えば、RAM 等の揮発性のメモリ、フラッシュメモリ、HDD、SSD 等により構成され、機械学習装置 100 が利用するデータ、およびプログラムを記憶する。記憶部 120 は、例えば、後述する暗号処理部 1322 が暗号実行領域において演

10

20

30

40

50

算するため、データを暗号化し、又は暗号化されたデータを復号するための鍵を記憶する鍵格納部 121、機械学習装置 100 において使用する学習モデルを記憶する学習モデル格納部 122 を含む。なお、鍵は、機械学習装置 100 内で記憶するのではなく、機械学習装置 100 とネットワークを介して、又は接続される鍵管理装置（不図示）において記憶するようにしてもよい。また、機械学習装置 100 が、機械学習のために利用した画像データ等は、セキュリティ向上のため、使用後は記憶部 120 から削除することが好ましい。

【0030】

鍵格納部 121（後述する図 4 参照）は、暗号処理部 1322 が暗号実行領域において演算するために、データを暗号化し、又は暗号化されたデータを復号するための鍵を記憶する。鍵格納部 121 は、例えば、ID、鍵データや、鍵ファイルの保存されている場所のパスを記憶する。ID は、ユーザや、組織など任意のグループ毎に割り当てて鍵を利用する構成としてもよいし、セッション毎に割り当てて鍵を利用する構成としてもよい。

10

【0031】

学習モデル格納部 122（後述する図 5 参照）は、機械学習装置 100 において使用する学習モデルを記憶する。学習モデル格納部 122 は、学習モデルに加え、複数の学習モデルを区別するための ID、学習モデルが適用できる範囲などを示すタグ、画像診断において特定の部位の特定までも可能とする学習モデルか否かの区別、当該学習モデルを用いて計算を行う際の計算コスト（計算量、計算時間、費用、計算資源等を含む。）および計算速度、当該学習モデルを用いて推論を行った場合の精度などを関連付けて記憶してもよい。

20

【0032】

制御部 130 は、機械学習装置 100 のプロセッサが記憶部 120 に記憶されるプログラムを読み込み、プログラムに含まれる命令を実行することにより実現される。制御部 130 は、機械学習装置 100 の動作を制御する。具体的には、制御部 130 は、データ取得部 131 と、機械学習計算部 132 と、受付部 1321 と、暗号処理部 1322 と、平文処理部 1323 と、を含む。

【0033】

データ取得部 131 は、機械学習により学習モデルのパラメータを更新する場面（以下、特に言及する場合を除き、学習モデルを新たに生成する場合も含む。）においては、教師データとなる画像データおよび当該画像データに対するラベルを取得する。「ラベル」は、例えば、画像データが示す対象物の種別を示したり、画像データにおいて特定したい対象物の範囲（位置）を示したりするものを含む。また、データ取得部 131 は、学習モデルを用いて推論を行う場面においては、推論対象となる画像データを取得する。例えば、ユーザは、端末装置 200 等において画像データおよびラベルを入力し、データ取得部 131 は、ネットワーク NW を介して、ユーザが入力した画像データおよびラベルを取得する。

30

【0034】

機械学習計算部 132 は、データ取得部 131 において取得した画像データ等に基づいて、畳み込みニューラルネットワークを用いた計算処理を行う。この計算処理は、具体的には、学習モデルのパラメータを更新するための学習、もしくは、取得した画像データ等から学習モデルを用いて行う推論のための計算処理である。なお、畳み込みニューラルネットワークは複数の層から構成される。「層」については、図 7 および図 8（後述）において説明する。

40

【0035】

機械学習計算部 132 は、計算処理の過程において、予めユーザから暗号実行領域において演算する範囲を取得しておき、計算処理の過程の一部を準同型暗号による暗号実行領域において演算し、計算処理の過程の一部を平文実行領域において演算する。機械学習計算部 132 は、具体的には、受付部 1321 と、暗号処理部 1322 と、平文処理部 1323 と、を有する。

50

【0036】

受付部1321は、畳み込みニューラルネットワークの入力層から出力層に向けての途中の層である第N層（ただし、Nは自然数。）を指定する層指定情報を受け付ける。すなわち、層指定情報により第N層が指定されたとき、入力層から第N層までは暗号実行領域において演算し、第(N+1)層から出力層までの演算を平文実行領域において演算する。

【0037】

受付部1321は、第N層を指定する層指定情報を受け付けるが、ここで取得する層指定情報は、「N」を示す具体的な数字であって、畳み込みニューラルネットワークにおいて、中間層のどの層までを暗号実行領域において演算するかという範囲を示すものである。畳み込みニューラルネットワークは、一般的には、入力層（0層）、中間層（1層～M層）、出力層（M+1層）といった複数の層を有する。ここで、Mは任意の自然数である。このとき、ユーザは、中間層の第N層までを暗号実行領域において演算することを決め、受付部1321は、暗号実行領域において演算を行う境界の層となるNを取得する。

10

【0038】

また、層指定情報は、「N」を示す具体的な数字の他に、「N」を指定することができる情報であってもよい。例えば、ユーザから受け付けた計算コストの条件であってもよい。受付部1321は、層指定情報として、計算コストの条件を受け付けた場合は、その計算コストで実行することができる「N」を算出する。

【0039】

暗号処理部1322は、準同型暗号による暗号実行領域において演算する機能を有する。暗号処理部1322は、機械学習計算部132における計算処理の過程において、入力層（0層）から中間層のN層までの計算処理を準同型暗号による暗号実行領域において演算する。

20

【0040】

暗号処理部1322は、具体的には、鍵格納部121において管理する暗号化のための鍵（公開鍵）を用いて、データ取得部131が取得した画像データ等を暗号化する。このとき、暗号処理部1322は、暗号化の方法について、暗号化状態のまま演算が可能な準同型暗号方式（Paillier方式等、Lifted-Ellgama方式、Somewhat Homomorphic Encryption方式、Fully Homomorphic Encryption方式等）を用いる。

【0041】

暗号処理部1322は、データ取得部131が取得した画像データ等および機械学習に用いる学習モデルを暗号化し、暗号実行領域において演算する。

30

【0042】

暗号処理部1322は、入力層（0層）からユーザの指定した中間層のN層までの計算処理を暗号化実行領域において演算した後、鍵格納部121において管理する復号のための鍵（秘密鍵）を用いて、データを復号する。

【0043】

平文処理部1323は、平文実行領域において演算する機能を有する。平文処理部1323は、暗号処理部1322において演算されたN層までの演算結果を取得し、引き続き(N+1)層から出力層までの演算を平文実行領域において行う。これにより、推論または学習モデルのパラメータの更新を行う。

40

【0044】

なお、上述した機械学習装置100の機能の一部を、端末装置200が担ってもよい。例えば、予め、機械学習装置100と端末装置200とで、画像データ等を準同型暗号で暗号化するための鍵を共有しておき、端末装置200において、推論や機械学習の対象となる画像データを準同型暗号で暗号化するようにしてもよい。

【0045】

端末装置200は、準同型暗号で暗号化された画像データを機械学習装置100に送信し、機械学習装置100は、通信部110を介して受信した当該画像データを、記憶部120に格納する。そして、データ取得部131は、暗号化された画像データを記憶部12

50

0 から取得し、暗号処理部 1 3 2 2 へ送る。次いで、暗号処理部 1 3 2 2 は、暗号化された画像データについて演算を実行する。その際、上述したように暗号処理部 1 3 2 2 が演算結果を復号するのに代えて、平文処理部 1 3 2 3 が演算結果を復号するようにしてもよい。すなわち、暗号処理部 1 3 2 2 では、準同型暗号で暗号化されたデータのみ扱うようにする。このような構成をとることで、機械学習装置 1 0 0 においては、画像データが準同型暗号で暗号化された状態、もしくは、復号されても人間が画像を目視では識別困難な状態であるため、機械学習装置 1 0 0 は、よりセキュリティを高めて、計算処理を実行することができる。

【0 0 4 6】

(データベースの具体例)

図 4 は、鍵格納部 1 2 1 の一例を示す図である。鍵格納部 1 2 1 は、例えば、ユーザ、セッションなど鍵の利用場面を識別するための「ID」、鍵（公開鍵「key 1」および秘密鍵「key 2」）の値を、関連付けて記憶する。図 4 において、「ID」が「0000001」の公開鍵は「0KJFDEMNF679FDDFHKL…」、秘密鍵は、「2SF5HLJHFSKLJADL7DJLH…」である。同様に、「ID」が「0000002」の公開鍵は「43JFLKEJFDHKLFS3JKLFN…」、秘密鍵は、「FDLK732492SFGKJHKSDH…」であり、「ID」が「0000003」の公開鍵は「JLKJ02SFHL89JLKHJKF897…」、秘密鍵は、「40389439LFKSJFLKJFSLH…」である。なお、図 4 は、鍵管理の一例であって、鍵の管理方法は、これらに限られない。例えば、鍵と復号権限を持つユーザとを関連付けて記憶してもよい。

【0 0 4 7】

図 5 は、学習モデル格納部 1 2 2 の一例を示す図である。学習モデル格納部 1 2 2 は、複数の学習モデルを区別するための「ID」、学習モデルが適用できる範囲などを示す「タグ」、画像診断において特定の部位の特定までも可能とする学習モデルか否かの区別する「部位特定」、当該学習モデルを用いて計算を行う際のコスト（計算量）を示す「コスト／MB」、および計算速度を示す「計算速度／Mb」、当該学習モデルを用いて推論を行った場合の精度を示す「精度」、学習モデルのファイル名を示す「ファイル」が関連付けて記憶されている。

【0 0 4 8】

図 5 において、「ID」が「0001」の学習モデルは、「タグ」欄が「CT、ガン、腎臓、医療、画像診断」であり、これらの用途に適用可能で、部位特定には対応していないことを示している。また、「コスト」は10円/Mb、「計算速度」は1 Mbあたり5'10、「精度」は80%で、学習モデル自身は「0001.json」というファイルに記憶されていることを示している。

【0 0 4 9】

また、「ID」が「0002」の学習モデルは、「タグ」欄が「CT、ガン、腎臓、医療、画像診断」であり、これらの用途に適用可能で、「部位特定」欄が「○」であることから、部位特定に対応していることを示している。また、「コスト」は45円/Mb、「計算速度」は1Mbあたり6'00、「精度」は85%で、学習モデル自身は「0002.json」というファイルに記憶されていることを示している。

【0 0 5 0】

「ID」が「0003」や「0004」についても同様であるので、説明を省略する。なお、学習モデル格納部 1 2 2 は、上述した項目に限らず、学習モデルを適用する組織（会社、部署など）、業界（医療、食品、ECなど）や、所有者等に関連付けて記憶されていてもよい。

【0 0 5 1】

(画像データおよびラベルの具体例)

図 6 は、ユーザが画像データとラベルを入力するための画面の一例を示す図である。データ取得部 1 3 1 は、ユーザから対象となる画像データの入力を受け付けると、当該画像データを読み込んだ上で、画面 4 0 1 に示すような画像 4 1 0 やファイル情報（ファイル名、ファイル生成日時等）をユーザが操作する端末装置 2 0 0 に表示させる。このとき、

10

20

30

40

50

データ取得部 131 は、ユーザからラベルも取得する。画面 401 において、ユーザは、入力した画像データが「正常」を示すデータか「異常」を示すデータかを選択し、ラベル付けを行う。また、ユーザは、異常部分を示す部位の特定まで行う学習モデルのラベル付けを行う場合には、当該部位の範囲を画面上で指定し、ラベル付けを行う。例えば、図 6 において、指定された部位特定の範囲は、画像 410 の斜線部分にて示されている。

【0052】

このようにして、データ取得部 131 は、画像データおよび当該画像データに対するラベルを関連付けた教師データを取得する。そして、機械学習計算部 132 は、教師データに基づいて、画像データに対するラベルを出力するように畳み込みニューラルネットワークの学習モデルのパラメータを生成（更新）する。ラベルの出力として、例えば、入力した画像データについて、ガンの診断を行う場合、正常を示す画像か異常を示す（ガンの疑い）画像であるかを出力する。また、ガンの疑いがあると出力するときは、その部位特定も出力するようにしてもよい。

【0053】

なお、推論を行うために画像データの入力のみを行う場合は、ラベル付けは不要であり、画像データの入力のみが行われる。

【0054】

（畳み込みニューラルネットワークの一例）

図 7 は、一般的な畳み込みニューラルネットワークのモデル概念図である。ニューラルネットワークは、入力層（0 層）、中間層（1 層から M 層）、出力層（M + 1 層）といった複数の層を有する。機械学習計算部 132 は、その計算処理過程において、受付部 1321 が暗号実行領域で演算する範囲の境界（第 N 層）を示す層指定情報を取得し、暗号処理部 1322 において暗号実行領域での演算を、平文処理部 1323 において平文実行領域での演算を行う。

【0055】

暗号処理部 1322 は、データ取得部 131 において取得した画像データ等と選択された学習モデルを、鍵格納部 121 に記憶する鍵（公開鍵）を用いて準同型暗号で暗号化し、推論または学習モデルのパラメータ更新のための演算を暗号実行領域において行う。具体的には、暗号処理部 1322 は、0 層から中間層の第 N 層までの演算処理を行う。また、暗号処理部 1322 は、暗号実行領域における演算結果を、鍵格納部 121 に記憶する鍵（暗号化した公開鍵と対になる秘密鍵）を用いて復号する。復号した結果は、平文処理部 1323 へ送ってもよい。なお、暗号実行領域における演算結果を記憶部 120 に格納する場合は、セキュリティ上、暗号化したまま格納し、平文処理部 1323 において、演算結果を復号することが好ましい。

【0056】

平文処理部 1323 は、暗号処理部 1322 において復号した画像データ等および学習モデルを用いて、第（N + 1）層から第（M + 1）層までの計算過程について、平文実行領域において演算処理を行う。

【0057】

図 8 は、中間層の一例を示した図である。畳み込みニューラルネットワークは、中間層として、一般的に、畳み込み演算を行うための畳み込み層、ReLU 層、プーリング層、全結合層などを有する。本発明において、「層」とは、畳み込み層、ReLU 層、プーリング層など各層単位を、1 層、2 層、3 層としてもよいし、畳み込み、ReLU、プーリングといった一連の処理を一つの処理単位として 1 層、2 層、3 層としてもよい。この場合、例えば、「畳み込み」、「ReLU」、「プーリング」の組み合わせを一層としてもよいし、「畳み込み」、「プーリング」の組み合わせや、「ReLU」、「プーリング」の組み合わせを一層としてもよい。なお、ReLU 層は、活性化層の一例であって他の活性化関数を用いてもよい。

【0058】

なお、畳み込みニューラルネットワークは、一般的に、全結合層を中間層の後半で用い

10

20

30

40

50

ることが多いが、全結合層を暗号化実行領域で演算してもよい。

【0059】

(処理の流れ)

図9は、機械学習装置100が実行する処理の一例を示すフローチャートである。

【0060】

制御部130は、画像データ等のデータ、又は画像データ等のデータおよびラベルを、通信部110を介して、端末装置200から取得する(ステップS101)。

【0061】

制御部130は、暗号実行領域で行う演算と平文実行領域で行う演算との境界となる、第N層を指定する層指定情報を、通信部110を介して、端末装置200から取得する(ステップS102)。

【0062】

制御部130は、取得した画像等のデータと適用する学習モデルを準同型暗号で暗号化する。そして、0層から第N層までを暗号化実行領域において演算し、その後復号する(ステップS103)。

【0063】

制御部130は、ステップS103において復号した演算結果に基づいて、第(N+1)層から出力層までを平文実行領域において演算し、推論又は学習モデルのパラメータの更新を行う。そして、制御部130は、演算結果を記憶部120に格納し、処理を終了する(ステップS104)。

【0064】

(効果の説明)

本実施形態に係る機械学習装置100は、機械学習における演算の一部を暗号化実行領域において、その他の部分を平文実行領域において演算することで、全てを暗号化実行領域において演算する場合と比較して、計算コストを削減し、現実的な時間で演算することを可能にする。

【0065】

畳み込みニューラルネットワークを用いた機械学習モデルでは、一般的に中間層の処理過程において、畳み込み演算のフィルタリング処理が繰り返されることにより、入力層に入力された画像データは、目視で識別が困難な状態となっていく。すなわち、畳み込み演算を繰り返した後の入力データは、人間によって識別することが困難であるから、暗号化していなくとも情報漏洩のリスクが低い。本実施形態に係る機械学習装置100は、この点に着目し、ユーザが指定した第N層までを暗号化実行領域において演算し、第(N+1)層以降を平文実行領域において演算することで、計算コストの削減と、セキュリティの確保の両立を実現することができる。

【0066】

<実施形態2>

実施形態1に係る機械学習システム1は、ユーザから層指定情報を取得した後、暗号実行領域および平文実行領域において計算処理を行う。これに対し、実施形態2に係る機械学習システム2は、ユーザが層指定情報を指定する際、入力画像を平文実行領域において計算したときに入力画像が畳み込み演算によりどのように変化するか、また、計算コストがどのように変化するかなどの情報をユーザに提供する。ユーザは、提供された情報を参照して、ユーザが求める計算コストやセキュリティレベルに応じた層指定情報を容易に指定することができるようになる。

【0067】

(機械学習システム2の構成)

図10は、本実施形態に係る機械学習システム2を示す図である。図10に示すように、本実施形態に係る機械学習システム2は、機械学習装置300と、ユーザによって使用される一つ以上の端末装置200と、を備える。機械学習装置300は、第1実施形態に係る機械学習装置100と比較して、提供部3324と、推定部3325と(図11にお

10

20

30

40

50

いて詳述。)をさらに備える。提供部3324は、ユーザが入力した画像データを平文実行領域において計算した場合にどのように変化するか、また計算コストなどの情報を示すことで、ユーザが暗号実行領域において演算を行う範囲の選択を補助する機能を備える。

【0068】

(機械学習装置300の機能的構成)

図11は、本実施形態における機械学習装置300の機能的構成を示すブロック図である。機械学習装置300は、通信部110と、記憶部120と、制御部330と、を備える。機械学習装置300における通信部110および記憶部120の機能構成は、機械学習装置100と同様であるため、説明を省略する。

【0069】

制御部330は、機械学習装置300のプロセッサが記憶部120に記憶されるプログラムを読み込み、プログラムに含まれる命令を実行することにより実現される。制御部330は、機械学習装置300の動作を制御する。具体的には、制御部330は、データ取得部131と、機械学習計算部332と、受付部1321と、暗号処理部1322と、平文処理部1323と、提供部3324と、推定部3325と、を含む。制御部330は、提供部3324および推定部3325を備えること以外は、機械学習装置100における制御部130と同様であるため、重複する説明を省略する。

【0070】

提供部3324は、データ取得部131において取得するユーザから入力される画像データについて、入力層から出力層までの演算を、平文処理部1323が実行した場合の当該演算に関する平文途中演算情報を提供する。平文途中演算情報とは、入力層から出力層までの途中過程における演算に関する情報であり、平文実行領域における、途中の層で入力又は出力されるデータ、途中の層までにかかる計算時間(および、ここから推定される推論又は学習に係る全時間)、途中の層までにかかる計算量(および、ここから推定される推論又は学習にかかる全計算量)、計算量から算出される料金などを含むものである。平文途中演算情報は、例えば、平文処理部1323が、ユーザから入力される一つの画像データについて入力層から出力層までの演算を実行し、当該演算に関する情報に基づいて算出することができる。平文途中演算情報は、例えば、通信部110を介して、ユーザが利用する端末装置200に送信されてもよいし、機械学習装置100における表示装置(不図示)に表示されてもよい。

【0071】

提供部3324は、例えば、平文途中演算情報として、入力層に入力される画像データと、入力層から出力層までの各層に入力されるデータとの情報の差を、ユーザが比較できる態様で提供してもよい。また、提供部3324は、必ずしもユーザが比較できる態様で提供せず、平文途中演算情報として、各層に入力される画像データを提供してもよい。

【0072】

提供部3324は、畳み込み演算が繰り返されることで目視による識別が困難になることを前提とし、各層に入力される画像データをユーザに提供することで、どの層に入力される画像データまで目視で識別可能かということに関する参考情報(情報の差)を提供し、ときには、比較を行うことにより、暗号化して機械学習を行う必要のある層を判断するための情報を提供する。例えば、サンプルの画像データについて、ある中間層である第1の層と、第1の層より後に処理が実行される第2の層とに入力されるデータのそれぞれをユーザに提示してもよい。これにより、サンプルの画像データが、層の処理によって、どのように変化し、情報の差がでるか、ユーザが比較でき、ユーザが層指定情報を指定するのを支援する。

【0073】

提供部3324は、入力層から出力層までの全ての入力画像データを提供しなくともよく、入力層から出力層までの複数の所定の層について、暗号処理部1322が演算を行った場合に入力される画像データと、平文処理部1323が演算を行った場合に入力される画像データとを、ユーザが比較できる態様で提供してもよい。このとき、複数の所定の層

10

20

30

40

50

の選び方については、5層ごとに1層を抽出するなど、周期性をもって選択してもよいし、ランダムに複数の層を抽出してもよい。

【0074】

提供部3324は、さらに、暗号処理部1322が演算を行った場合に、途中の層に入力される画像データに代えて、マスク画像を提供してもよい。すなわち、暗号処理部1322が演算を行った場合に入力される画像データは暗号化されているため、人にとって識別不能なデータであることから、あえてその画像データを提示する必要性に乏しく、代わりにマスク画像を提供することで、人にとって識別不能なデータであることを示してもよい。

【0075】

提供部3324は、例えば、入力層から所定の層までを暗号処理部1322が演算し、当該所定の層の次の層から出力層までを平文処理部1323が実行した場合の計算コストを、ユーザが比較できる態様で提供してもよい。提供部3324は、所定の層までを暗号実行領域において演算し、残りを平文実行領域において演算した場合の計算時間および／又は計算量などをユーザに提供することで、暗号化して機械学習を行う必要のある層を判断するための情報提供をすることができる。

【0076】

提供部3324は、推定部3325（後述）が算出する層指定推定情報を提供してもよい。層指定推定情報は、データ取得部131が取得した画像データに関連するデータと、当該画像データについて指定された層指定情報の相関関係に基づいて算出される情報である。層指定推定情報は、例えば、第N層を具体的に特定する数値であってもよいし、Nの範囲であってもよい。

【0077】

推定部3325は、データ取得部が取得した画像データに関連するデータと、当該画像データについて指定された層指定情報との相関関係を学習する。画像データに関連するデータとは、画像データを取り扱う業界（医療、食品、EC等）、内容（診断、マーケティング等）、組織、ユーザなどである。推定部3325は、何層目まで暗号化するかを、画像データに関連するデータと相関関係を学習する。そして、推定部3325は、学習した結果、入力される画像データに関連するデータと相関関係にある層指定情報を、層指定推定情報として算出する。上述したように、提供部3324が、推定部3325が算出した層指定推定情報を、平文途中演算情報とともにユーザに提供することで、層指定情報を指定する際の目安となる。これにより、ユーザは層指定情報を容易に指定することができるようになる。

【0078】

（提供部3324が提供する平文途中演算情報の具体例）

図12は、提供部3324が提供する画面の一例を示す図である。画面402には、平文処理部1323が演算を実行した場合に、計算過程5層の各層に入力される画像データ（「牛」を示す画像）を示しており、ユーザが暗号実行領域において演算を行う範囲を「選択中」であることを示すアイコン412が表示されている。

【0079】

図12において、「Input」の部分には、ユーザが入力した入力画像が表示される。また、画面402には、平文途中演算情報として、複数の層（計算過程1～5）における入力画像が、表示されている。一般的に、畳み込みニューラルネットワークにおいて、畳み込み演算のフィルタリング処理が繰り返される過程で、入力画像は目視で識別しにくくなるため、深い層における入力画像ほど、識別が困難な画像となっている。例えば、画面402の入力画像(Input)は、牛であり、Inputと計算過程1で示される画像は、目視では、殆ど区別がつかない。一方、計算過程3～5で示される画像は、牛の輪郭がなく、目視では、画像が示す対象物を識別することが困難であることが分かる。

【0080】

また、画面402には、複数の層に入力される画像データだけではなく、途中の層まで

10

20

30

40

50

を暗号実行領域において演算し、それ以降を平文実行領域において演算した場合の計算時間（「Time」）や費用（「Cost」）も合わせて示されている。例えば、計算過程1に対して、「Time」は「5'36」であり、「Cost」は「\$4」である。すなわち、計算過程1までを暗号実行領域において演算し、それ以降を平文実行領域において演算した場合の計算時間は「5'36」、費用は「\$4」であることを示している。畳み込みニューラルネットワークにおける処理の層が深くなる（計算過程1から5へ進む）につれて、暗号実行領域において演算する計算量が増加していくため、計算時間や費用も増加していく。なお、図12では、計算コストとして、計算時間や費用が提示されているが、演算にかかる計算量や、計算資源なども計算コストとしてユーザに提示してもよい。

【0081】

また、ユーザが、ボタン422をクリック等することで、上述の「牛」の入力画像ではなく、別の画像や、プラットフォーム側で予め用意したサンプル画像、情報量の差が示されるイラスト等の画像を入力画像として変更できるようにしてもよい。この場合、これらのサンプル画像に基づいて、平文途中演算情報が提供される。

【0082】

図13は、提供部3324が提供する画面の一例を示す図である。画面403の領域425には、ユーザが暗号実行領域において演算を行う範囲として「計算過程2」を選択したことが示されている。また、ユーザは、「決定」ボタン413をクリック等することにより「計算過程2」を層指定情報として入力できる。なお、層の指定は、例えば、各計算過程において示されている画像データや、当該画像データの下方に示されているアイコン（図13では鍵アイコン）をクリック等することにより行われてもよいし、キーボードや音声により行われてもよい。

【0083】

図13では、ユーザが選択した層である計算過程2までの層における画像データをマスク画像で示している。ある層までを暗号実行領域において演算する場合、その層までに入力される画像データは暗号化されるため、人による識別は不可能であり、その暗号化された画像データを敢えてユーザに提示する必要性は乏しい。マスク画像によって代替して表示を行うことにより、暗号実行領域における演算を減らすことができる。しかしながら、ユーザが演算結果をイメージしやすくするために、暗号実行領域における演算結果を表示することを妨げるものではない。

【0084】

（処理の流れ）

図14は、機械学習装置300が実行する処理の一例を示すフローチャートである。

【0085】

制御部130は、画像データ等のデータ、又は画像データ等のデータおよびラベルを、通信部110を介して、端末装置200から取得する（ステップS201）。

【0086】

制御部130は、入力層から出力層までの各層において、平文処理部1323が実行した場合の当該演算に関する平文途中演算情報を、通信部110を介して、端末装置200に提供する（ステップS202）。

【0087】

制御部130は、暗号実行領域で行う演算と平文実行領域で行う演算との境界となる、第N層を指定する層指定情報を、通信部110を介して、端末装置200から取得する（ステップS203）。

【0088】

制御部130は、取得した画像等のデータと適用する学習モデルを準同型暗号で暗号化する。そして、0層から第N層までを暗号化実行領域において演算し、その後復号する（ステップS204）。

【0089】

制御部130は、ステップS204において復号した演算結果に基づいて、第（N+1

10

20

30

40

50

）層から出力層までを平文実行領域において演算し、推論又は学習モデルのパラメータの更新を行う。そして、制御部 130 は、演算結果を記憶部 120 に格納し、処理を終了する（ステップ S205）。

【0090】

（効果の説明）

本実施形態に係る機械学習装置 300 は、機械学習における演算の一部を暗号実行領域において、その他の部分を平文実行領域において演算することを前提として、入力画像を平文実行領域で処理した場合の、平文途中演算情報をユーザに提供する。平文途中演算情報を参照することにより、ユーザは、ユーザが求める実用的な計算コストや、セキュリティレベルに応じた、境界となる層の選択が可能となる。

10

【0091】

＜実施形態 3＞

実施形態 1 および実施形態 2 は、画像を目視で識別することが困難になるまで、畳み込みニューラルネットワークを用いた計算処理の一部を暗号実行領域において行う。これにより、計算過程の画像が第三者に流出したとしても、当該画像を秘匿することができる。一方で、例えば、暗号処理部と平文処理部とが同じ権限で実行可能だとすると、悪意のある第三者が当該権限を奪取したときに、暗号処理部における計算過程の画像に対し、復号のための鍵を用いて入力画像の復元を試みることのおそれがある。

【0092】

そこで、実施形態 3 に係る機械学習装置は、暗号実行領域において演算する暗号処理部と平文実行領域において演算する平文処理部との権限を分けることで、セキュリティを高める。仮に、悪意ある第三者が暗号処理部の権限を奪取しても、平文処理部の権限を有していなければ、復号のための鍵にアクセスできないため、暗号化された画像を復号し、入力画像を認識することは困難である。また、悪意ある第三者が平文処理部の権限を奪取しても、畳み込みニューラルネットワークにおける計算過程は、暗号処理部における処理を経て、平文処理部における処理へと進んでいるため、入力画像の復元はより困難である。

20

【0093】

上記構成を実現するために、本実施形態に係る機械学習装置では、実施形態 1 または 2 で述べた機械学習システム 1 または 2 で行う処理を 1 つのコンピュータ上で実行する際に、暗号処理部が実行する処理と、平文処理部が実行する処理との権限を分け、暗号処理部から平文処理部へのデータの受け渡しは共有メモリを介して行う。これにより、ある処理の権限を奪取した悪意ある第三者による他の処理へのアクセスを防止し、悪意ある第三者にハッキングされた場合のセキュリティを高めることができる。

30

【0094】

また、近年、畳み込みニューラルネットワークを用いた画像処理においては、モデル情報および計算途中の画像から元の画像に復元を試みる研究もされている。このため、たとえば、暗号化された画像や、目視で識別困難な画像であったとしても、計算途中の画像等の流出リスクを減らすことは重要である。本実施形態に係る機械学習装置は、上述のように、暗号処理部から平文処理部へのデータの受け渡しは共有メモリを介して行うため、計算途中における画像流出の可能性を低減させることも可能であり、よりセキュリティに配慮した構成となっている。

40

【0095】

（機械学習装置 400 の機能的構成）

図 15 は、本実施形態に係る機械学習装置 400 の機能的構成を示すブロック図である。図 15 を参照して、本実施形態に係る機械学習装置 400 の機能的構成について説明する。なお、実施形態 1 と共通の構成要素には同一の符号を付し、繰り返しの説明を省略する。

【0096】

機械学習装置 400 は、通信部 110 と、記憶部 120 と、制御部 430 と、を備え、制御部 430 は、データ取得部 131 と、共有メモリ 431 と、機械学習部 432 とを含

50

む。機械学習部 4 3 2 は、受付部 1 3 2 1 と、暗号処理部 4 3 2 2 と、平文処理部 4 3 2 3 とを有する。

【0097】

共有メモリ 4 3 1 は、例えば、揮発性メモリなどであり、暗号処理部 4 3 2 2（後述）および平文処理部 4 3 2 3（後述）がアクセス可能である。暗号処理部 4 3 2 2 および平文処理部 4 3 2 3 は、実行した演算の結果（データ）を共有メモリ 4 3 1 に記憶させる。また、暗号処理部 4 3 2 2 および平文処理部 4 3 2 3 が行う演算に必要な鍵やモデル情報などを、データ取得部 1 3 1 が記憶部 1 2 0 から読み出し、共有メモリ 4 3 1 に記憶させてもよい。これにより、機械学習装置 4 0 0 の暗号処理部 4 3 2 2 と平文処理部 4 3 2 3 は、共有メモリ 4 2 3 を介して、計算過程のデータを共有することができる。

10

【0098】

暗号処理部 4 3 2 2 は、まず、データ取得部 1 3 1 が取得した画像データ等を、鍵格納部 1 2 1 において管理する暗号化のための鍵（公開鍵）を用いて準同型暗号による暗号化を行う。そして、暗号化した画像データ等と、学習モデル格納部 1 2 2 の学習モデルを用いて、入力層（第 0 層）からユーザの指定した中間層の第 N 層までの計算処理を暗号化実行領域において演算する。なお、データ取得部 1 3 1 が、準同型暗号によって暗号化された画像データを取得していた場合は、暗号処理部 4 3 2 2 による上記暗号化は省略する。

【0099】

また、当該演算の前に、データ取得部 1 3 1 が、鍵格納部 1 2 1 から暗号化のための鍵（公開鍵）、および学習モデル格納部 1 2 2 から入力層から第 N 層までにに関するモデル情報（第 1 のモデル情報）を読み出し、共有メモリ 4 3 1 に記憶させておいてもよい。暗号処理部 4 3 2 2 は、共有メモリ 4 3 1 から鍵やモデル情報を読み出して、演算に用いることにより、処理速度を速めることができる。

20

【0100】

次に、暗号処理部 4 3 2 2 は、入力層（第 0 層）から第 N 層までの演算の結果を、共有メモリ 4 3 1 に記憶させる。このとき、暗号処理部 4 3 2 2 は、第 N 層までの演算の結果だけでなく、入力層（第 0 層）から第 N 層までの各層における計算過程の全てのデータ又は、計算過程の一部のデータを共有メモリ 4 3 1 に記憶させてもよい。この際、演算の結果を一時的に外部の記憶装置に出力した後、共有メモリ 4 3 1 に記憶させるようにしてもよい。なお、暗号処理部 4 3 2 2 の演算終了後、暗号処理部 4 3 2 2 が演算で用いた第 1 のモデル情報は、共有メモリ 4 3 1 から削除することが望ましい。上述したように、モデル情報および計算途中の画像から元の画像を復元することができる。このため、悪意ある第三者が平文処理部の権限を奪取した場合、モデル情報と計算途中の画像から、元の入力画像の復元を試みるおそれがある。したがって、元の入力画像への復元を困難にするため、モデル情報の一部である第 1 のモデル情報は、当該演算結果が復号できる者にはアクセスできないようにしておくことが重要である。

30

【0101】

平文処理部 4 3 2 3 は、共有メモリ 4 3 1 から、暗号処理部 4 3 2 2 が実行した第 N 層までの演算の結果を取得し、取得した演算の結果に基づいて、第 (N + 1) 層から出力層までの演算を行う。このとき、平文処理部 4 3 2 3 は、学習モデル格納部 1 2 2 の学習モデルを用いて、第 (N + 1) 層から出力層までの演算を行う。また、平文処理部 4 3 2 3 は、共有メモリ 4 3 1 から取得した暗号処理部 4 3 2 2 が実行した第 N 層までの演算結果を、鍵格納部 1 2 1 において管理する復号のための鍵（秘密鍵）を用いて復号する。

40

【0102】

なお、平文処理部 4 3 2 3 が、暗号処理部 4 3 2 2 の演算結果を取得する前に、データ取得部 1 3 1 が、鍵格納部 1 2 1 から復号のための鍵（秘密鍵）、および学習モデル格納部 1 2 2 から第 (N + 1) 層から出力層までにに関するモデル情報（第 2 のモデル情報）を読み出し、共有メモリ 4 3 1 に記憶させておいてもよい。平文処理部 4 3 2 3 は、共有メモリ 4 3 1 から鍵やモデル情報を読み出して、演算に用いることにより、処理速度を速めることができる。

50

【0103】

また、暗復号で用いる鍵は、処理の都度、または一定期間ごとに使い捨て等して、変更することが望ましい。これにより、元の入力画像や計算途中の画像の復号を困難にし、セキュリティを高めることができる。

【0104】

平文処理部4323は、第2のモデル情報を用いて、第(N+1)層から出力層までの演算を平文実行領域において行い、これにより推論または学習モデルのパラメータの更新を行う。

【0105】

(共有メモリを用いた処理の具体例)

図16は、共有メモリ431に関し、機能ブロック図とハードウェアとの対応関係を示す図である。共有メモリ431は、暗号処理部4322および平文処理部4323のいずれからもアクセス可能な記憶領域である。暗号処理部4322および平文処理部4323は、プロセッサの機能として処理を実行する。また、共有メモリ431は、鍵格納部121において管理する鍵や、暗号処理部4322が実行した演算の結果および平文処理部4323が実行した演算の結果を記憶する。

【0106】

暗号処理部4322は、演算の結果を共有メモリ431に記憶させる。平文処理部4323は、共有メモリ431から、暗号処理部4322が実行した演算の結果を取得する。このとき、平文処理部4323は、鍵格納部121から復号のための鍵(秘密鍵)を取得し、暗号処理部4322が実行した演算の結果について復号する。

【0107】

図17は、アカウント制御を用いて実行権限を分けた場合を示す図である。1台のコンピュータ内で暗号処理部4322と平文処理部4323の実行権限を分ける場合、例えば、2つのアカウント(アカウント1, アカウント2)を用意し、それぞれのアカウントに実行権限を1つずつ割り当てることで実現できる。具体的には、図17において示すように、暗号処理部4322の実行権限をアカウント1に付与し、平文処理部4323の実行権限をアカウント2に与える。なお、アカウント1およびアカウント2には、共有メモリ431へのアクセス権限(データの読み出しと書き込みを行う権限)が与えられている。

【0108】

機械学習装置400では、アカウントを分離することにより、1台のコンピュータ内で、暗号処理部4322の実行権限と、平文処理部4323の実行権限とを分けることができる。これにより、例えば、アカウント1は、平文処理部4323の実行権限がないため、復号鍵を取得する権限はない。すなわち、悪意ある第三者が、暗号処理部4322の権限を有するアカウント1を奪取しても、アカウント1は鍵格納部121から復号鍵を取得できないため、不正に取得された画像等のデータの復号は困難である。また、悪意ある第三者が、平文処理部4323の権限を有するアカウント2を奪取しても、畳み込みニューラルネットワークの計算過程が進んでいるため、もとの入力画像の復元は困難である。

【0109】

なお、入力画像の復元を困難にするため、暗号処理部4322の演算終了後、第1のモデル情報は、共有メモリ431から削除することが望ましい。上述したように、モデル情報、および計算途中の画像から元の画像を復元することができる。このため、悪意ある第三者が平文処理部の権限を有するアカウント2を奪取した場合、元の入力画像の復元を試みるおそれがある。したがって、第1のモデル情報は、アカウント2がアクセスできないようにすることにより、よりセキュリティが高まる。

【0110】

さらに、機械学習装置400は、暗号処理部4322と平文処理部4323が共通してアクセス可能な共有メモリ431を用いることで、計算過程のデータの流出を防ぐことができる。また、悪意のある第三者が計算過程の演算の結果を遡って取得することを防止することもできる。すなわち、磁気ディスク装置などの記憶装置を介して情報を共有すると

10

20

30

40

50

、悪意のある第三者にアカウントが乗っ取られた際に過去の演算の結果を取得されるリスクが発生するが、揮発性メモリなどを共有メモリ431として用いることで、暗号処理部4322および平文処理部4323の演算の結果は一次的に記憶されるのみとなり、悪意のある第三者が過去の演算の結果を不正に取得することを予防できる。

【0111】

(処理の流れ)

図18は、機械学習装置400が実行する処理の一例を示すフローチャートである。

【0112】

制御部430は、画像データ等のデータ、又は画像データ等のデータおよびラベルを取得する(ステップS301)。

【0113】

制御部430は、暗号実行領域で行う演算と平文実行領域で行う演算との境界となる第N層を指定する層指定情報を取得する(ステップS302)。

【0114】

制御部430は、取得した画像等のデータと適用する学習モデルを準同型暗号で暗号化する。そして、入力層(第0層)から第N層まで暗号化実行領域において演算し、共有メモリ431に記憶する(ステップS303)。

【0115】

制御部430は、ステップS303において行った演算の結果を共有メモリ431から取得し、当該演算の結果に基づいて、復号鍵を用いて演算の結果を復号する。そして、復号した演算の結果に基づいて、第(N+1)層から出力層までを平文実行領域において演算し、推論又は学習モデルのパラメータの更新を行う。次いで、制御部430は、出力層までの演算の結果を出力し、処理を終了する(ステップS304)。

【0116】

(効果の説明)

本実施形態に係る機械学習装置400は、暗号処理部4322が実行する処理と平文処理部4323が実行する処理の権限を分け、暗号処理部4322から平文処理部4323への情報の受け渡しは共有メモリ431を介して行う。これにより、暗号処理部4322又は平文処理部4323の権限を奪取した悪意ある第三者が、他の処理にアクセスすることを防止し、セキュリティを高めることができる。

【0117】

また、機械学習装置400は、暗号処理部4322と平文処理部4323の両方がアクセス可能な共有メモリ431を用いることで、演算の結果の共有が可能となる。揮発性メモリなどに構成した共有メモリ431を用いることで、演算の結果を短期的に記憶し、過去のデータを上書きすることで、悪意ある第三者が不正なアクセスを試みた場合にも、元の画像等の復元につながるデータの取得を困難にし、セキュリティを高めることができる。

【0118】

また、入力画像の復元を困難にするため、暗号処理部4322の演算終了後、第1のモデル情報を、共有メモリ431から削除することで、第1のモデル情報、および平文処理部4323において復号される第N層までの演算結果に基づいて、元の入力画像に復元することを防ぐことができ、よりセキュリティを高めることができる。

【0119】

<実施形態4>

実施形態4は、実施形態3の変形例である。本実施形態に係る機械学習装置は、実施形態3に係る機械学習装置と比較して、学習モデルに対するアクセス権限を分ける。本実施形態に係る機械学習装置は、暗号処理部の実行権限と平文処理部の実行権限を分けるだけでなく、学習モデルのモデル情報へのアクセス権限も分けることで、悪意のある第三者に一部の権限を奪取されても、元の画像等の復元をより困難にすることを可能とする。

【0120】

図19は、本実施形態に係る機械学習装置500の機能的構成を示すブロック図である

10

20

30

40

50

。以下、図 19 を参照して、実施形態 4 に係る機械学習装置 500 の機能的構成について説明する。なお、実施形態 3 と共通の構成要素には同一の符号を付し、繰り返しの説明を省略する。

【0121】

機械学習装置 500 は、通信部 110 と、記憶部 520 と、制御部 530 と、を備える。記憶部 520 は、鍵格納部 121 と、学習モデル格納部 522 と、を含み、制御部 530 は、データ取得部 531 と、共有メモリ 532 と、機械学習計算部 533 と、を含む。機械学習計算部 533 は、受付部 1321 と、暗号処理部 5322 と、平文処理部 5323 と、を有する。

【0122】

学習モデル格納部 522 は、入力層から第 N 層までに係る第 1 のモデル情報と、第 (N + 1) 層から出力層までに係る第 2 のモデル情報とを記憶する。第 1 のモデル情報は、学習モデルのうち、暗号処理部 5322 における演算のために必要となる学習モデルの情報である。第 2 のモデル情報は、学習モデルの情報のうち、平文処理部 5323 における演算のために必要となる学習モデルの情報である。

【0123】

第 1 のモデル情報と、第 2 のモデル情報は、例えば、同じ記憶領域に格納され、アクセス権限が設定される。第 1 のモデル情報は、暗号処理部の実行権限を有する者がアクセス可能であり、第 2 のモデル情報は、平文処理部の実行権限を有する者がアクセス可能である。また、第 1 のモデル情報と、第 2 のモデル情報とを異なる記憶領域に格納してもよい。この場合、第 1 のモデル情報は、暗号処理部の実行権限を有する者がアクセス可能であり、第 2 のモデル情報は、平文処理部の実行権限を有する者がアクセス可能である。

【0124】

暗号処理部 5322 は、データ取得部 531 を介して取得した画像等のデータと第 1 のモデル情報を取得し、暗号実行領域における入力層から第 N 層までの演算を行う。そして、暗号処理部 5322 は、演算結果を共有メモリ 532 に記憶させる。

【0125】

平文処理部 5323 は、共有メモリ 532 から、暗号処理部 5322 が実行した第 N 層までの演算の結果を取得する。そして、データ取得部 521 を介して取得した第 2 のモデル情報を用いて、平文実行領域における第 (N + 1) 層から出力層までの演算を行い、これにより推論又は学習モデルのパラメータの更新を行う。この際、平文処理部 5323 は、鍵格納部 121 において管理する復号のための鍵（秘密鍵）を用いて第 N 層までの演算の結果を復号する。

【0126】

なお、各モデル情報は、機械学習装置 500 以外の装置が記憶していてもよく、データ取得部 531 は、各モデル情報を、機械学習装置 500 以外の装置等から別々に取得してもよい。

【0127】

(効果の説明)

本実施形態に係る機械学習装置 500 は、暗号処理部 5322 が実行する処理と平文処理部 5323 が実行する処理の権限を分けるだけでなく、学習モデルへのアクセス権限も分ける。これにより、悪意ある第三者が実行権限の一部を奪取したとしても、実施形態 3 に係る機械学習装置 400 よりも、取得できる情報を減らすことができる。したがって、元の画像等を復元することをより困難にし、セキュリティを高めることができる。

【0128】

<実施形態 5>

実施形態 3 に係る機械学習装置は、1 つのコンピュータにおいて、暗号処理部と平文処理部の実行権限を分けることで、悪意ある第三者が一部の権限を奪取しても、元の画像等の復元を困難にし、セキュリティを高めた。実施形態 5 に係る機械学習システムは、実施形態 1 又は 2 で述べた機械学習システムにおける暗号処理部の処理と平文処理部の処理を

10

20

30

40

50

異なる端末装置で実行させるなど、暗号処理部と平文処理部とを物理的に異なる実行環境において演算を実行することで、悪意ある第三者が一部の実行環境の権限を奪取しても元の画像等の復元を困難にし、セキュリティを高める。物理的に異なる実行環境において演算することで、1つのコンピュータにおいて実行権限を分けるよりも、実行権限の分離をより確実なものにし、セキュリティを高めることができる。

【0129】

(機械学習システム3の構成)

図20は、機械学習システム3の全体図である。図20に示すように、本実施形態に係る機械学習システム3は、暗号実行領域の演算を行う暗号処理装置600と、平文実行領域の演算を行う平文処理装置700と、その余の機械学習のための処理を行うシステム管理装置800とを備える。

10

【0130】

暗号処理装置600と、平文処理装置700と、システム管理装置800とは、ネットワークNWを介してデータを送受信することができる。

【0131】

機械学習システム3の機能は、実施形態1に係る機械学習装置100の機能を、暗号処理装置600と、平文処理装置700と、システム管理装置800とに分ける。暗号処理装置600と、平文処理装置700と、システム管理装置800とがネットワークNWを介して通信することで、機械学習装置100の機能を3つに分割した装置により実現するものである。

20

【0132】

機械学習システム3の機能は、具体的には、以下の構成で実現できる。例えば、実施形態1に係る機械学習装置100の記憶部120における鍵格納部121と学習モデル格納部122の機能は、システム管理装置800に機能的に配置する。機械学習装置100の制御部130におけるデータ取得部131と機械学習計算部132の暗号処理部1322の機能は、暗号処理装置600に機能的に配置する。機械学習装置100の制御部130における機械学習計算部132の平文処理部1323の機能は、平文処理装置700に機能的に配置する。機械学習装置100の制御部130における機械学習計算部132の受付部1321の機能は、システム管理装置800に機能的に配置する。

【0133】

ただし、前述の例はあくまで一例であり、機械学習システム3の機能は、少なくとも暗号処理装置600に暗号処理部1322の機能が配置され、平文処理装置700に平文処理部1323の機能が配置されればよい。その他の機能は、暗号処理装置600、平文処理装置700、システム管理装置800のいずれかに配置されてもよい。

30

【0134】

また、例えば、実施形態1に係る機械学習装置100の記憶部120における鍵格納部121と学習モデル格納部122の機能は、暗号処理装置600および平文処理装置700の双方に機能的に配置してもよい。

【0135】

また、例えば、実施形態1に係る機械学習装置100の制御部130におけるデータ取得部131の機能は、システム管理装置800に機能的に配置されてもよい。機械学習装置100の受付部1321の機能は、暗号処理装置600に機能的に配置されてもよい。

40

【0136】

(暗号処理装置600の機能的構成)

図21は、本実施形態に係る暗号処理装置600の機能的構成を示すブロック図である。以下、図21を参照して、本実施形態に係る暗号処理装置600の機能的構成について説明する。

【0137】

暗号処理装置600は、通信部610と、記憶部620と、制御部630とを備える。

【0138】

50

通信部 610 は、暗号処理装置 600 が他の装置と通信するための処理を行う。通信部 610 は、他の装置との通信は、セキュアな通信チャネルを構築する。

【0139】

記憶部 620 は、例えば、RAM 等の揮発性のメモリ、フラッシュメモリ、HDD、SSD 等により構成され、暗号処理装置 600 が利用するデータ、およびプログラムを記憶する。

【0140】

制御部 630 は、暗号処理装置 600 のプロセッサが記憶部 620 に記憶されるプログラムを読み込み、プログラムに含まれる命令を実行することにより実現される。制御部 630 は、暗号処理装置 600 の動作を制御する。具体的には、制御部 630 は、データ取得部 631 と、機械学習計算部 632 と、を含む。

10

【0141】

データ取得部 631 は、ユーザが入力した画像データ等とラベルを取得する。

【0142】

機械学習計算部 632 は、データ取得部 631 において取得した画像データ等に基づいて、畳み込みニューラルネットワークを用いた計算処理を行う。この計算処理は、具体的には、学習モデルのパラメータを更新するための学習、もしくは、取得した画像データ等から学習モデルを用いて行う推論のための計算処理である。なお、畳み込みニューラルネットワークは複数の層から構成される。

【0143】

20

機械学習計算部 632 は、暗号処理部 6322 を有する。暗号処理部 6322 は、通信部 610 を介してシステム管理装置 800 の鍵格納部 821（後述）から取得した暗号化のための鍵を用いて、準同型暗号により画像データ等を暗号化する。なお、データ取得部 631 が、準同型暗号により暗号化された画像データを取得していた場合は、当該暗号化を省略する。

【0144】

暗号処理部 6322 は、通信部 610 を介して、システム管理装置 800 の学習モデル格納部 822（後述）から取得したモデル情報と、通信部 610 を介して取得したシステム管理装置 800 の受付部 831（後述）が取得した層指定情報を用いて、入力層から第 N 層までの計算処理を準同型暗号による暗号実行領域において演算する。

30

【0145】

暗号処理部 6322 は、通信部 610 を介して、暗号実行領域における演算の結果を平文処理装置 700 に送信する。平文処理装置 700 は、通信部 710 を介して、暗号実行領域における演算の結果を取得する。

【0146】

（平文処理装置 700 の機能的構成）

図 22 は、本実施形態に係る平文処理装置 700 の機能的構成を示すブロック図である。以下、図 22 を参照して、本実施形態に係る平文処理装置 700 の機能的構成について説明する。

【0147】

40

平文処理装置 700 は、通信部 710 と、記憶部 720 と、制御部 730 とを備える。

【0148】

通信部 710 は、平文処理装置 700 が他の装置と通信するための処理を行う。通信部 710 は、他の装置との通信は、セキュアな通信チャネルを構築する。

【0149】

記憶部 720 は、例えば、RAM 等の揮発性のメモリ、フラッシュメモリ、HDD、SSD 等により構成され、平文処理装置 700 が利用するデータ、およびプログラムを記憶する。

【0150】

制御部 730 は、平文処理装置 700 のプロセッサが記憶部 720 に記憶されるプログ

50

ラムを読み込み、プログラムに含まれる命令を実行することにより実現される。制御部 730 は、平文処理装置 700 の動作を制御する。具体的には、制御部 730 は、機械学習計算部 732 を含む。

【0151】

機械学習計算部 732 は、通信部 710 を介して、暗号処理装置 600 から取得した暗号実行領域における演算の結果に基づき、畳み込みニューラルネットワークを用いた計算処理を行う。この計算処理は、具体的には、学習モデルのパラメータを更新するための学習、もしくは、取得した画像データ等から学習モデルを用いて行う推論のための計算処理である。なお、畳み込みニューラルネットワークは複数の層から構成される。

【0152】

機械学習計算部 732 は、平文処理部 7323 を有する。平文処理部 7323 は、通信部 710 を介してシステム管理装置 800 の鍵格納部 821（後述）から取得した復号のための鍵を用いて、暗号実行領域における演算の結果を復号する。

【0153】

平文処理部 7323 は、通信部 710 を介してシステム管理装置 800 の学習モデル格納部 822（後述）から取得したモデル情報と、通信部 710 を介して取得したシステム管理装置 800 の受付部 831（後述）が取得した層指定情報を用いて、第（N+1）層から出力層までの計算処理を平文実行領域において演算する。これにより、推論または学習モデルのパラメータの更新を行う。

【0154】

平文処理部 7323 は、推論または学習モデルのパラメータの更新を行い、通信部 710 を介して、当該学習モデルをシステム管理装置 800 に送信し、システム管理装置 800 の学習モデル格納部 822（後述）に記憶されている学習モデルを更新する。

【0155】

（システム管理装置 800 の機能的構成）

図 23 は、本実施形態に係るシステム管理装置 800 の機能的構成を示すブロック図である。以下、図 23 を参照して、本実施形態に係るシステム管理装置 800 の機能的構成について説明する。

【0156】

システム管理装置 800 は、通信部 810 と、記憶部 820 と、制御部 830 とを備える。

【0157】

通信部 810 は、システム管理装置 800 が他の装置と通信するための処理を行う。通信部 810 は、他の装置との通信は、セキュアな通信チャネルを構築する。

【0158】

記憶部 820 は、例えば、RAM 等の揮発性のメモリ、フラッシュメモリ、HDD、SSD 等により構成され、システム管理装置 800 が利用するデータ、およびプログラムを記憶する。記憶部 820 は、鍵格納部 821 と、学習モデル格納部 822 と、を含む。

鍵格納部 821 は、データを暗号化し、又は暗号化されたデータを復号するための鍵を記憶する。学習モデル格納部 822 は、機械学習システム 3 において使用する学習モデルを記憶する。

【0159】

制御部 830 は、システム管理装置 800 のプロセッサが記憶部 820 に記憶されるプログラムを読み込み、プログラムに含まれる命令を実行することにより実現される。制御部 830 は、システム管理装置 800 の動作を制御する。制御部 830 は、受付部 831 を含む。

【0160】

受付部 831 は、畳み込みニューラルネットワークの入力層から出力層に向けての途中の層である第 N 層（ただし、N は自然数。）を指定する層指定情報を受け付ける。

【0161】

10

20

30

40

50

システム管理装置 800 は、通信部 810 を介して、暗号処理装置 600 または／および平文処理装置 700 に対して、層指定情報を送信する。

【0162】

(物理的に異なる実行環境において演算を行う場合の具体例)

図 24 は、物理的に異なる実行環境において畳み込みニューラルネットワークを用いた計算処理を行う場合の具体例を示す図である。

【0163】

本実施形態に係る機械学習システム 3 のシステムでは、ニューラルネットワークを用いた計算処理を、準同型暗号による暗号実行領域と、平文実行領域に分けて行う。そして、暗号実行領域における処理を暗号処理装置 600 において行い、平文実行領域における処理を平文処理装置 700 において行うことで、物理的に異なる実行環境において計算処理を行う。

10

【0164】

図 24 に示すように、暗号処理装置 600 において、暗号処理部 6322 が、データ取得部 631 (図 21 参照) が取得した画像データ等を暗号化し、暗号実行領域において、入力層 (第 0 層) から第 N 層までの演算を実行する。暗号処理装置 600 は、入力層から第 N 層までの演算の結果を、ネットワークを介して平文処理装置 700 に送信する。平文処理装置 700 において、平文処理部 7323 は、暗号処理装置 600 から送信された演算の結果を復号した上で、平文実行領域において、第 (N+1) 層から出力層までの演算を行う。このように、機械学習システム 3 は、暗号処理部 6322 と平文処理部 7323 とを異なる実行環境において計算処理させることで、悪意のある第三者に、計算処理に係る装置がハッキングされた際のリスクを軽減し、セキュリティを高めることができる。

20

【0165】

(処理の流れ)

図 25 は、機械学習システム 3 が実行する処理の一例を示すシーケンス図である。なお、ステップの順番は図 25 で示す例に限られない。

【0166】

暗号処理装置 600 は、画像データ等のデータ、又は画像データ等のデータおよびラベルを取得する (ステップ S401)。

【0167】

システム管理装置 800 は、暗号実行領域で行う演算と平文実行領域で行う演算との境界となる第 N 層を指定する層指定情報を取得する (ステップ S402)。

30

【0168】

システム管理装置 800 は、暗号処理装置 600 に対して、層指定情報、学習モデル、暗号化のための鍵 (暗号鍵) を送信する (ステップ S403)。

【0169】

暗号処理装置 600 は、取得した画像等のデータと適用する学習モデルを準同型暗号で暗号化する。そして、第 0 層から第 N 層まで暗号化実行領域において演算する (ステップ S404)。

【0170】

暗号処理装置 600 は、暗号実行領域における演算結果を平文処理装置 700 に送信する (ステップ S405)。

40

【0171】

システム管理装置 800 は、層指定情報、学習モデル、復号のための鍵 (秘密鍵) を平文処理装置 700 に送信する (ステップ S406)。なお、復号のための鍵は、セキュリティが確保された通信路において送信される他に、予め、平文処理装置 700 とシステム管理装置 800 とが共有しておいてもよい。

【0172】

平文処理装置 700 は、暗号処理装置 600 が行った演算の結果に基づいて、復号鍵を用いて演算結果を復号する。そして、復号した演算の結果に基づいて、第 (N+1) 層が

50

ら出力層までを平文実行領域において演算し、推論又は学習モデルのパラメータの更新を行う（ステップS407）。

【0173】

平文処理装置700は更新された学習モデルをシステム管理装置800に送信し、システム管理装置800は記憶している学習モデルのパラメータを更新する（ステップS408）。

【0174】

（効果の説明）

本実施形態に係る機械学習システム3は、暗号処理部6322が実行する処理と、平文処理部7323が実行する処理とを、物理的な実行環境に分けて行う。これにより、暗号処理部6322又は平文処理部7323の実行環境をハッキングした悪意ある第三者が、他の処理にアクセスすることを防止し、元の画像等を取得することを予防してセキュリティを高めることができる。機械学習システム3は、物理的な実行環境を分けることにより、1つのコンピュータ上で実行権限を分けるよりも、他の処理へのアクセスを困難にし、セキュリティを高めることができる。

【0175】

<実施形態6>

実施形態6は、実施形態5の変形例である。実施形態5に係る機械学習システム3は、暗号処理装置600と平文処理装置700のいずれにおいても、システム管理装置800の学習モデル格納部822に記憶された学習モデルを取得し、それぞれ畳み込みニューラルネットワークのための計算を行った。実施形態6に係る機械学習システム3は、暗号処理装置600は、学習モデルのうち入力層から第N層までにに関する情報（「第1のモデル情報」）を取得して演算を行い、平文処理装置700は、学習モデルのうち第（N+1）層から出力層までにに関する情報（「第2のモデル情報」）を取得して演算を行う。また、第1のモデル情報および第2のモデル情報は、システム管理装置800からではなく、信頼できるサーバ等から別々に取得するようにしてもよい。

【0176】

暗号処理装置600の暗号処理部6322は、第1のモデル情報を取得し、データ取得部631が取得した画像データ等と第1のモデル情報を用いて、入力層から第N層までの暗号実行領域における演算を行う。

【0177】

平文処理装置700の平文処理部7323は、第2のモデル情報を取得し、暗号処理部6322が行った演算の結果と、第2のモデル情報を用いて、第（N+1）層から出力層までの平文実行領域における演算を行い、推論またはモデル情報のパラメータの更新を行う。

【0178】

また、実施形態6における機械学習システム3は、前述のように、暗号処理部6322と平文処理部7323の両方において、それぞれの演算に必要な学習モデルの情報を取得するのではなく、層指定情報として指定されるNについて、一般的に用いられる値がある程度決まっている場合には、予め暗号処理装置600と平文処理装置700に、第1のモデル情報と第2のモデル情報を記憶しておいてもよい。

【0179】

すなわち、層指定情報として、例えば、 $N=5$ がよく用いられる場合には、暗号処理装置600の記憶部620に第1のモデル情報として、予め、入力層から第5層までのモデル情報を記憶し、平文処理装置700の記憶部720に第2のモデル情報として、第6層から出力層までのモデル情報を記憶してもよい。

【0180】

（効果の説明）

実施形態6における機械学習システム3は、実施形態5のように、暗号処理部6322と平文処理部7323の物理的な実行環境を分けるだけでなく、各処理部は、学習モデル

10

20

30

40

50

の全てのパラメータではなく、自己の演算に必要な学習モデルのパラメータのみを取得することで、悪意のある第三者がハッキングした際に、当該第三者が取得する情報をできる限り少なくすることにより、元の画像等を復元することを困難にし、セキュリティを高めることができる。

【0181】

<実施形態7>

実施形態7は、実施形態5の変形例である。実施形態7に係る機械学習システム4は、少なくとも2以上の暗号処理装置と、少なくとも2以上の平文処理装置とを備える。また、暗号処理装置と平文処理装置とは、少なくとも一組のペアリングがされており、ペアリングされた暗号処理装置と平文処理装置とを用いて畳み込みニューラルネットワークの入力層から出力層までの計算処理を実行する。

10

【0182】

また、実施形態7に係る機械学習システムは、ペアリングされていない暗号処理装置と平文処理装置は、畳み込みニューラルネットワークの入力層から出力層までの計算処理とは異なる処理に関する通信を行う。

【0183】

実施形態7に係る機械学習システムは、ペアリングされている暗号処理装置と平文処理装置の間で通信するだけでなく、ペアリングされていない暗号処理装置と平文処理装置の間でも通信を行うことにより、悪意のある第三者が、畳み込みニューラルネットワークの入力層から出力層までの一つの計算処理がどの装置により行われているか判別することを困難にさせる。すなわち、ペアリングされた暗号処理装置と平文処理装置との間でのみ通信を行えば、悪意のある第三者は、通信が行われていることを感知すれば、一つの計算処理においてどの装置がペアリングされているか容易に判別可能である。一方で、ペアリングされていない暗号処理装置と平文処理装置との間でも入力層から出力層までの計算処理とは異なる処理に関する通信（ダミーの通信でもよい）を行うことで、ペアリングされている端末を判別することを困難にさせる。

20

【0184】

（具体例）

図26は、機械学習システム4の具体的構成を示す図である。機械学習システム4は、2以上の暗号処理装置600-1～600-P（Pは2以上の整数。以下、特に、区別する場合を除き、「暗号処理装置600」と記載する。）と、2以上の平文処理装置700-1～700-Q（Qは2以上の整数。以下、特に区別する場合を除き、「平文処理装置700」と記載する。）を含み、これらはネットワークを介して接続されている。例えば、暗号処理装置600-1と平文処理装置700-1、暗号処理装置600-2と平文処理装置700-2、暗号処理装置600-Pと平文処理装置700-Qがペアリングされているとき、これらは、一つの畳み込みニューラルネットワークの入力層から出力層までの計算処理を実行する。一方で、ペアリングされていない暗号処理装置600と平文処理装置700は、畳み込みニューラルネットワークの入力層から出力層までの計算処理とは無関係の通信、すなわちダミーの通信を行う。これにより、悪意のある第三者がどの装置間でペアリングされているか感知することを防止し、ハッキングなどをさせにくくするシステムとすることが可能となる。

30

【0185】

また、望ましくは、ペアリングされた装置間とペアリングされていない装置間において、同程度の通信が行われるとよい。これにより悪意のある第三者がペアリングされている装置の組合せを感知することをより難しくすることが可能となる。なお、ペアリングは、図示しない管理装置等により管理され、一定期間ごとに変更してもよい。ペアリングがランダムに変更されることにより、ハッキングの可能性をさらに低減させることができる。

40

【0186】

（効果の説明）

実施形態7に係る機械学習システム4は、複数の暗号処理装置600と複数の平文処理

50

装置 700 を備え、一つの畳み込みニューラルネットワークの入力層から出力層までの計算処理を実行するためにペアリングされた暗号処理装置 600 と平文処理装置 700 の組合せと、ペアリングされていない組合せを含む。これにより、悪意のある第三者が、どの装置がペアリングされているか感知することを困難にし、元の画像等の入手や復元を困難にし、セキュリティを高めることが可能となる。

【0187】

以上で実施形態の説明を終了するが、上記実施形態は一例に過ぎない。そのため、機械学習装置 100、300、400、500、暗号処理装置 600、平文処理装置 700、システム管理装置 800 の具体的な構成、処理内容等は上記実施形態で説明したものに限られない。

【0188】

また、本開示に係る機械学習装置は、上記装置によらず、例えば、コンピュータがプログラムを実行することで、その機能を実現してもよい。情報提供システムの機能を実現するためのプログラムは、USB (Universal Serial Bus) メモリ、CD-ROM (Compact Disc-Read Only Memory)、DVD (Digital Versatile Disc)、HDD (Hard Disc Drive)、SSD (Solid State Drive) 等のコンピュータ読み取り可能な記録媒体に記憶されてもよいし、ネットワークを介してコンピュータにダウンロードされてもよい。

【0189】

以上、本開示の好ましい実施形態について説明したが、上記実施形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で種々の省略、置き換え、変更を行うことができる。これら実施形態やその変形は、発明の範囲や要旨に含まれると同様に、特許請求の範囲に記載された発明とその均等の範囲に含まれるものとする。

【符号の説明】

【0190】

1, 2, 3, 4 機械学習システム、11 プロセッサ、12 メモリ、13 ストレージ、14 通信IF、15 入出力IF、100, 300, 400, 500 機械学習装置、110, 610, 710, 810 通信部、120, 420, 620, 720, 820 記憶部、121, 821 鍵格納部、122, 522, 822 学習モデル格納部、431, 532 共有メモリ、130, 430, 530, 630, 730, 830 制御部、131, 531, 631 データ取得部、132, 332, 432, 533, 632, 732 機械学習計算部、200 端末装置、600 暗号処理装置、700 平文処理装置、800 システム管理装置、401, 402, 403 画面、410 画像、412 アイコン、413 ボタン、1321, 831 受付部、1322, 4322, 5322, 6322 暗号処理部、1323, 4323, 5323, 7323 平文処理部、3324 提供部、3325 推定部。

10

20

30

40

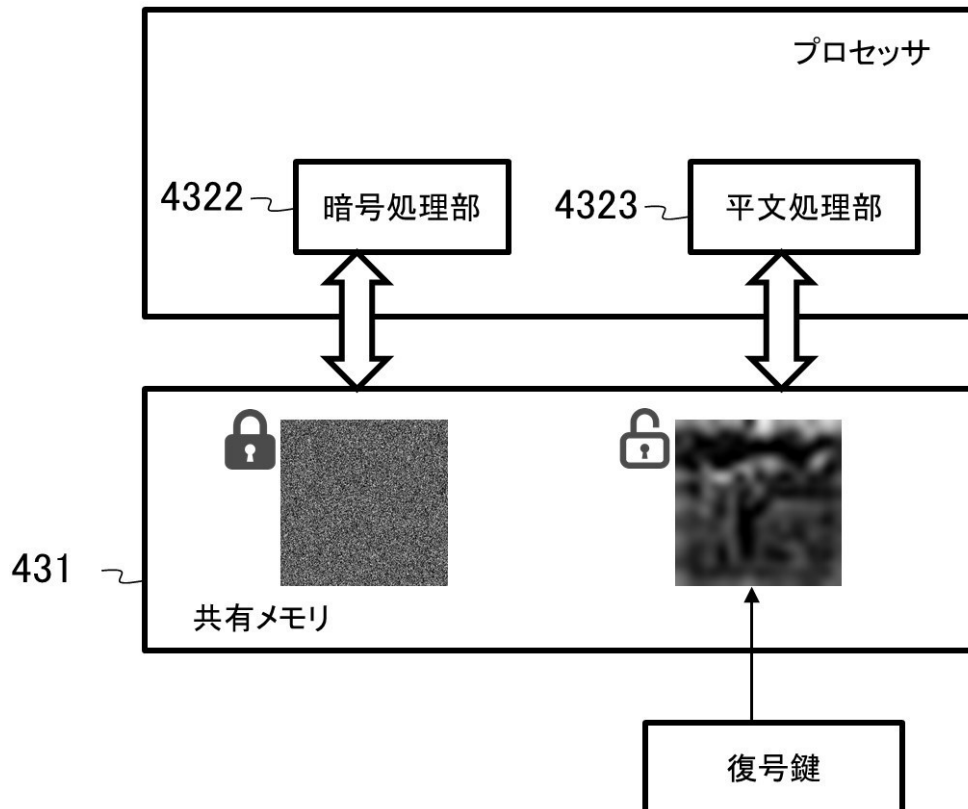
50

【要約】

【課題】セキュリティを考慮した、畳み込みニューラルネットワークを用いた機械学習に関する技術を提供する。

【解決手段】機械学習装置は、画像データを取得するデータ取得部と、画像データに基づいて、畳み込みニューラルネットワークを用いた計算処理を行う機械学習計算部とを備え、畳み込みニューラルネットワークは複数の層から構成され、機械学習計算部は、準同型暗号による暗号実行領域において演算する暗号処理部と、平文実行領域において演算する平文処理部と、暗号処理部および平文処理部がアクセス可能な共有メモリと、畳み込みニューラルネットワークの入力層から出力層に向けての途中の層である第N層を指定する層指定情報を受け付ける受付部とを有し、暗号処理部は、入力層から第N層までの演算を実行し、平文処理部は、第(N+1)層から出力層までの演算を実行し、共有メモリは暗号処理部および平文処理部が実行した演算の結果を記憶する。

【選択図】図16



10

20

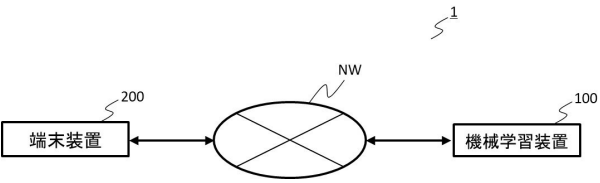
30

40

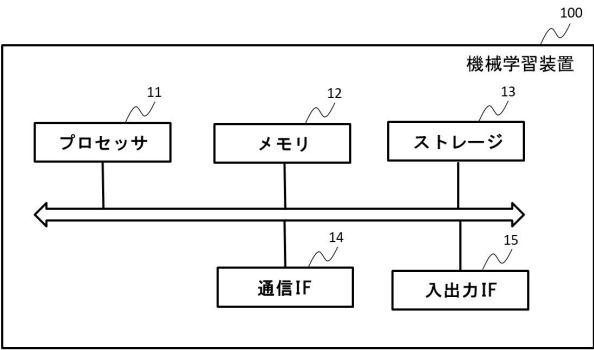
50

【図面】

【図 1】

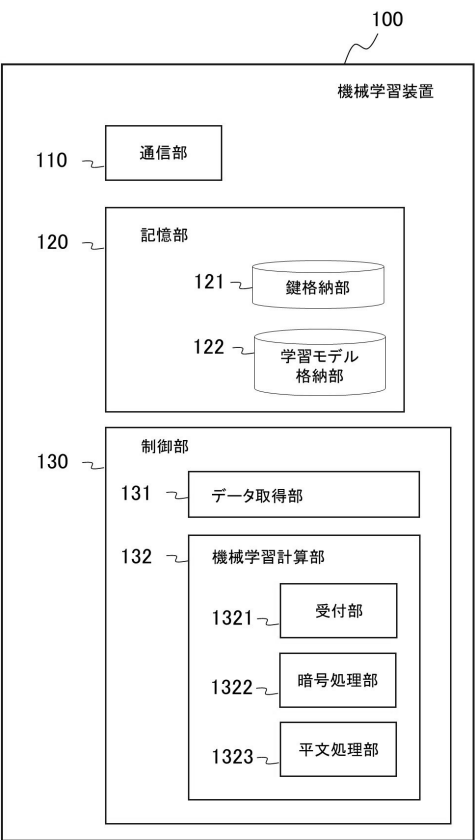


【図 2】



10

【図 3】



【図 4】

Figure 4 shows a table (121) with three columns: ID, key1, and key2.

ID	key1	key2
000001	0KJFDEMNF5679FDDFHKL...	2SFSLJHFSKLJADL7DJLH...
000002	43JFLKEJFDHKLFS3JKLFN...	FDLK732492SFGKJHKSDH...
000003	JLKJ02SFHL89JLKHJKF897...	40389439LFKSJFLKJFSLH...

20

30

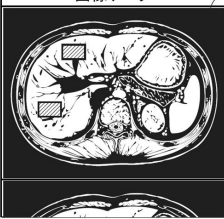
40

50

【図 5】

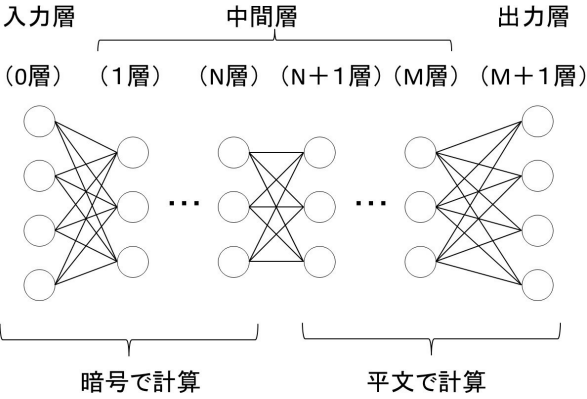
ID	タグ	部位 特定	122				ファイル
			コスト/Mb (円)	計算速度 /Mb	精度		
0001	CT、ガン、腎臓、医療、画像診断		10	5' 10	80%		0001.json
0002	CT、ガン、腎臓、医療、画像診断	○	45	6' 00	85%		0002.json
0003	CT、ガン、腎臓、医療、画像診断		20	3' 30	90%		0003.json
0004	CT、ガン、肝臓、医療、画像診断		15	4' 30	75%		0004.json

【図 6】

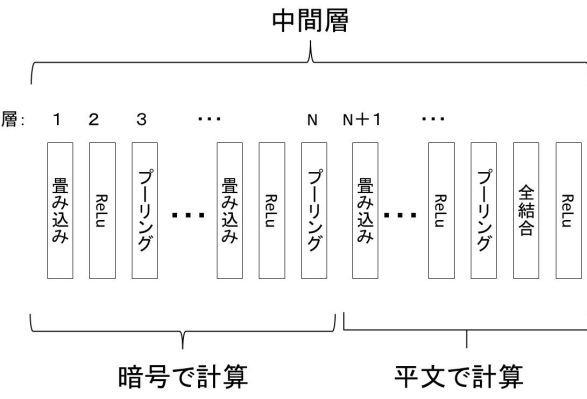
401 データラベルの指定			
部位特定する場合：ドラッグで範囲を選択			
画像データ	ファイル情報	正常	異常
	20210315_Kidney_0 2021/3/15 11:25		✓

10

【図 7】



【図 8】



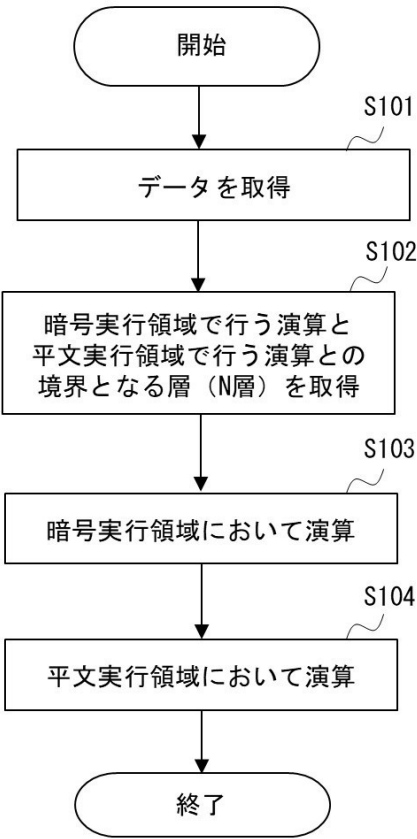
20

30

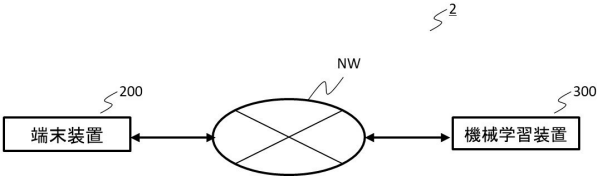
40

50

【図 9】



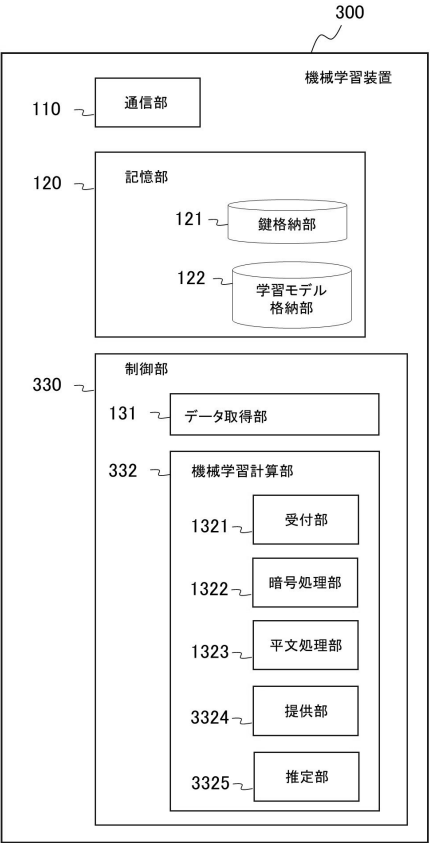
【図 10】



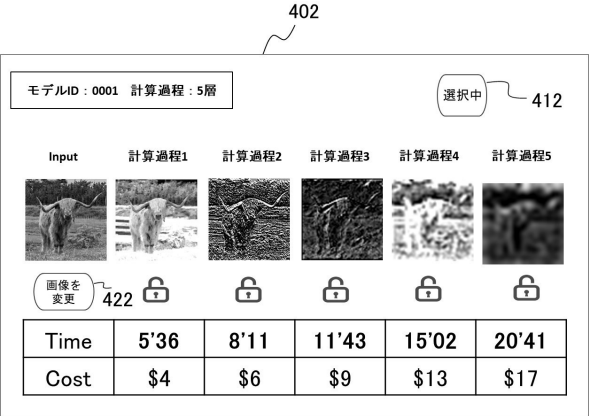
10

20

【図 11】



【図 12】

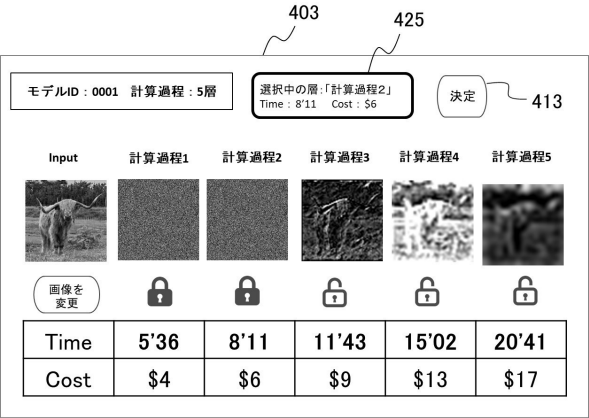


30

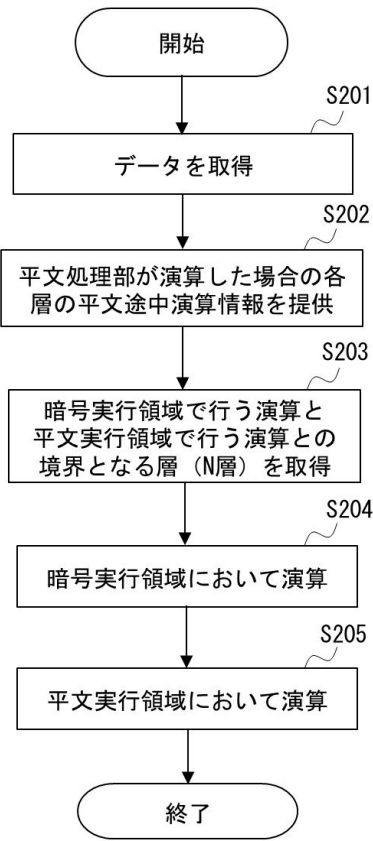
40

50

【図 1 3】



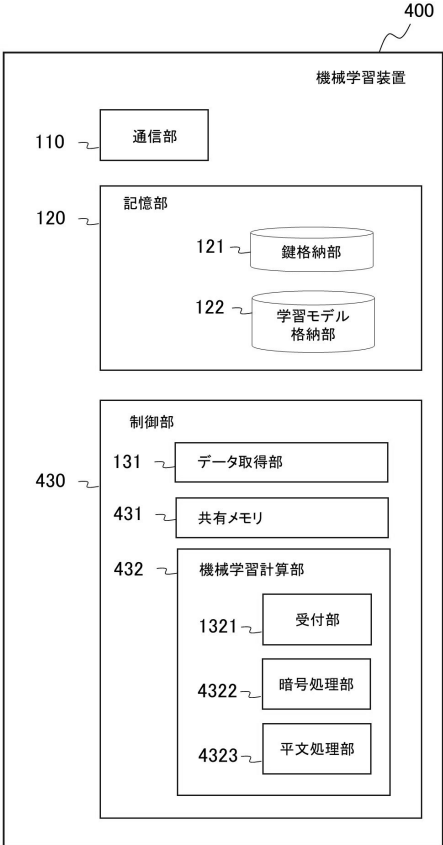
【図 1 4】



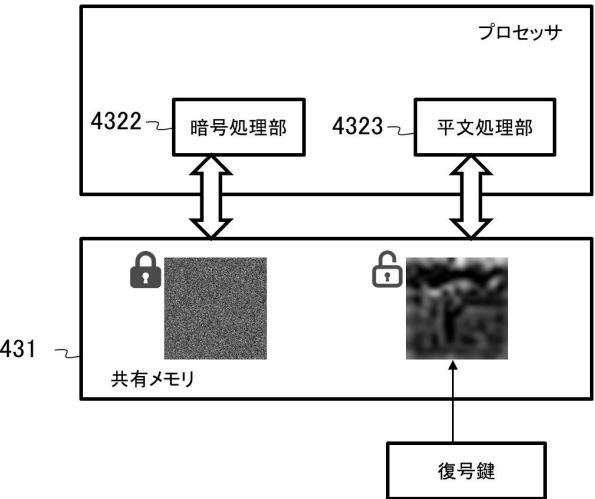
10

20

【図 1 5】



【図 1 6】

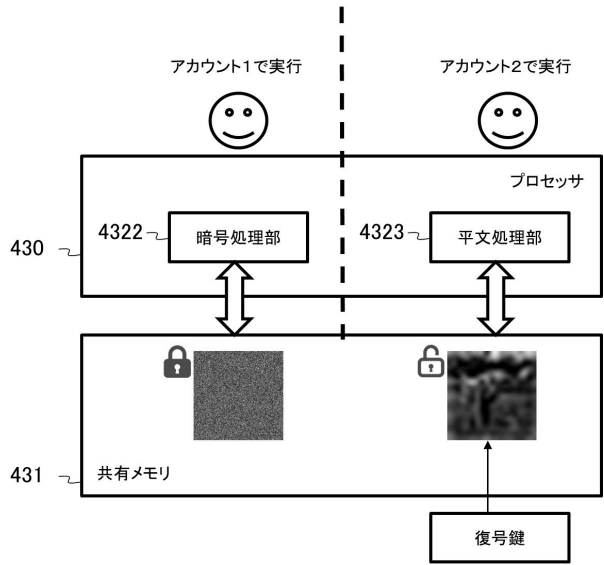


30

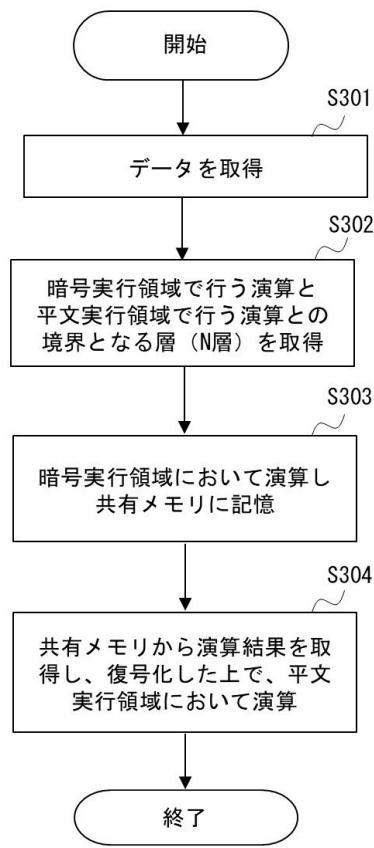
40

50

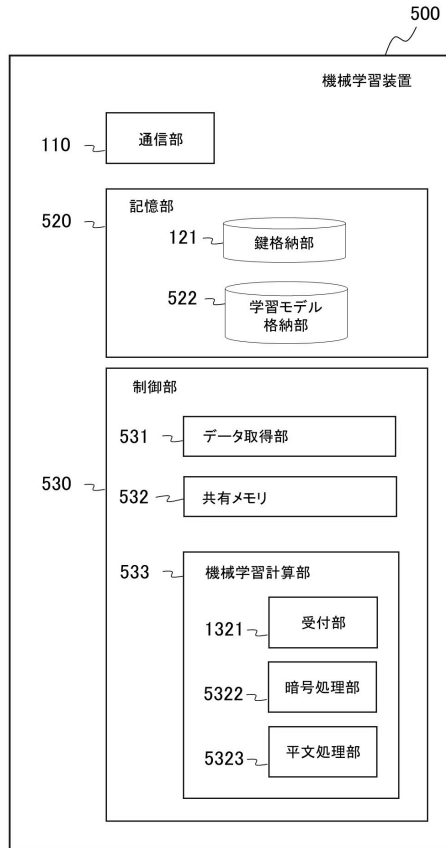
【図 17】



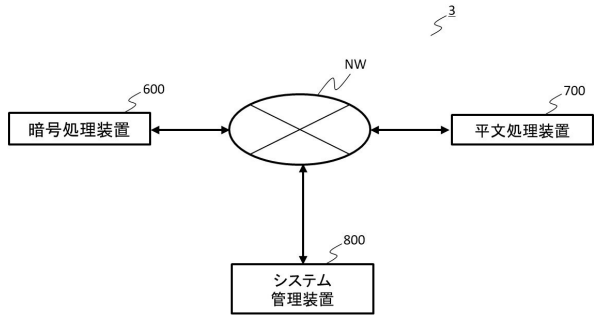
【図 18】



【図 19】



【図 20】



10

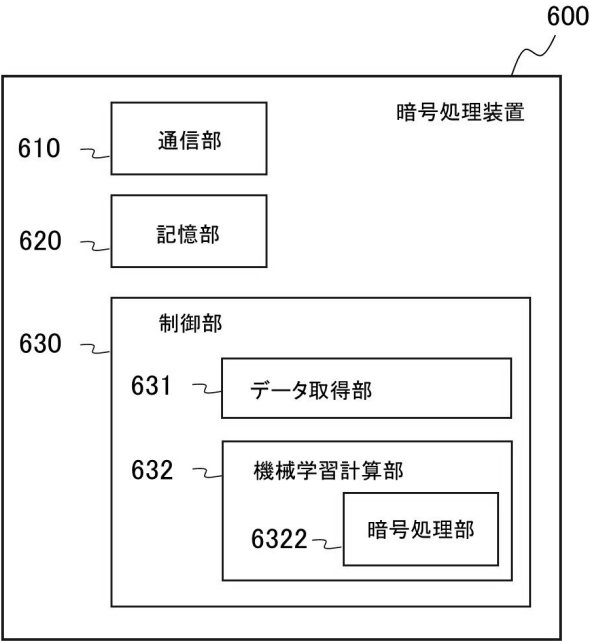
20

30

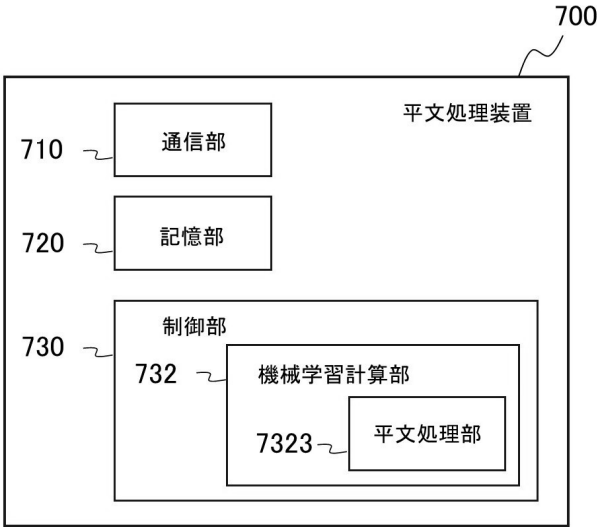
40

50

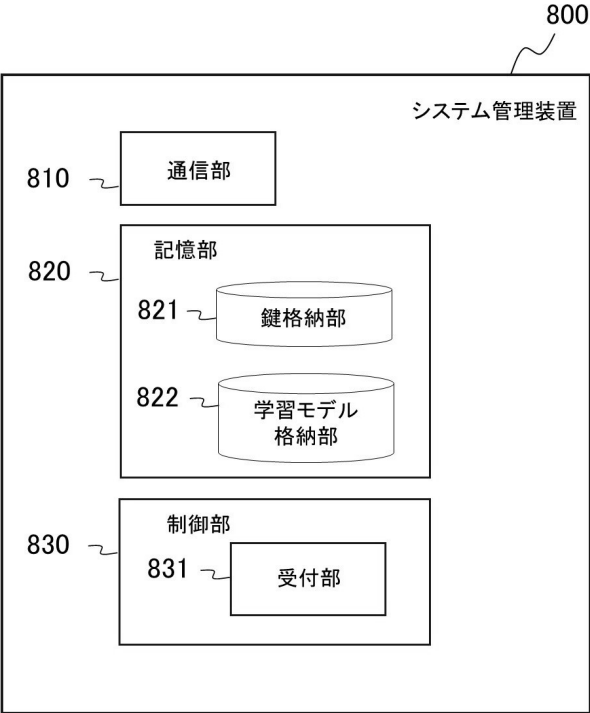
【図 2 1】



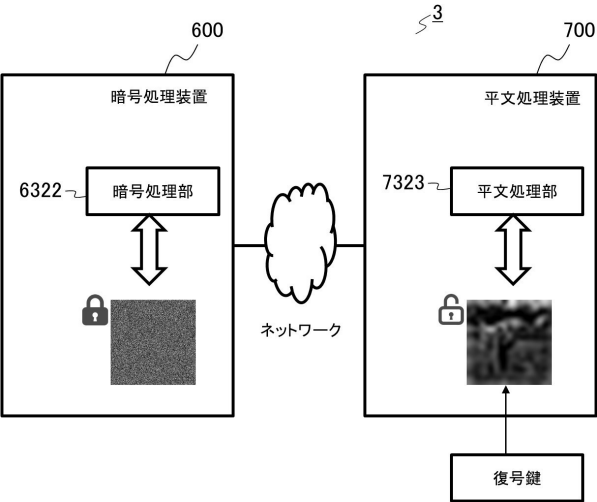
【図 2 2】



【図 2 3】



【図 2 4】



10

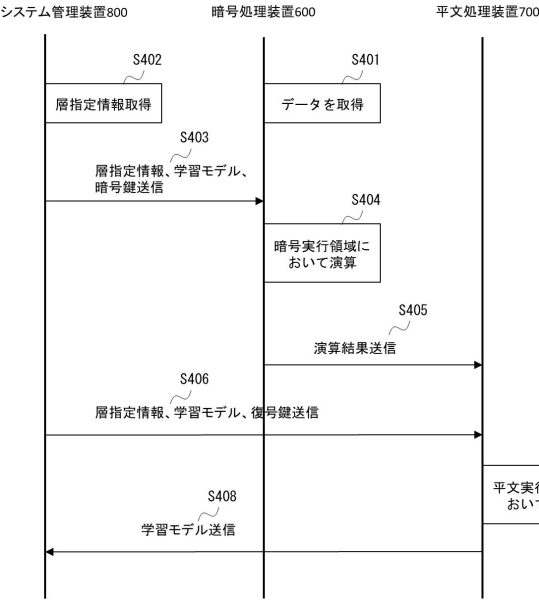
20

30

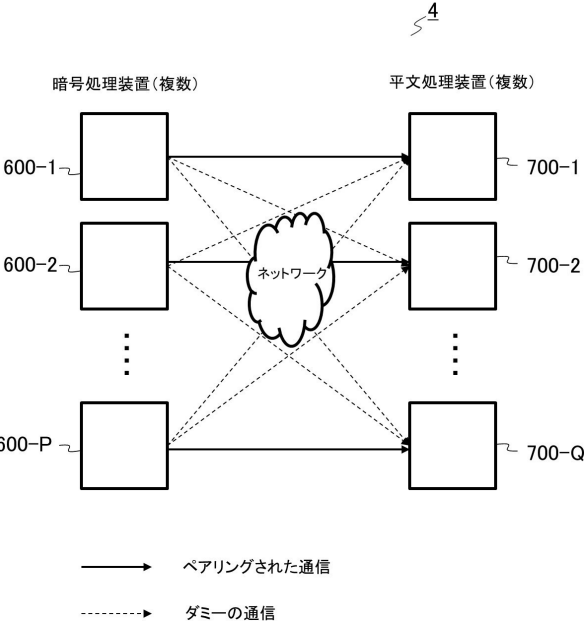
40

50

【図 2 5】



【図 2 6】



10

20

30

40

50

フロントページの続き

- (56)参考文献 特開 2021-64008 (JP, A)
中国特許第 108021819 (CN, B)
吉田康太、外 1 名、深層学習モデルで安全に推論するためのモデルパラメータ暗号化の検討、2021 年 暗号と情報セキュリティシンポジウム予稿集、2021 年 01 月
- (58)調査した分野 (Int.Cl., DB 名)
G06N 3/00-99/00
G06F 21/00-21/88
G09C 1/00-5/00
G06T 7/00