

(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2020-177582  
(P2020-177582A)

(43) 公開日 令和2年10月29日(2020.10.29)

(51) Int.Cl.  
G06N 3/08 (2006.01)

F I  
G06N 3/08

テーマコード (参考)

審査請求 未請求 請求項の数 12 O L (全 16 頁)

(21) 出願番号 特願2019-81223 (P2019-81223)  
(22) 出願日 平成31年4月22日 (2019.4.22)

(71) 出願人 000003078  
株式会社東芝  
東京都港区芝浦一丁目1番1号  
(74) 代理人 110002147  
特許業務法人酒井国際特許事務所  
(72) 発明者 新田 修平  
東京都港区芝浦一丁目1番1号 株式会社  
東芝内

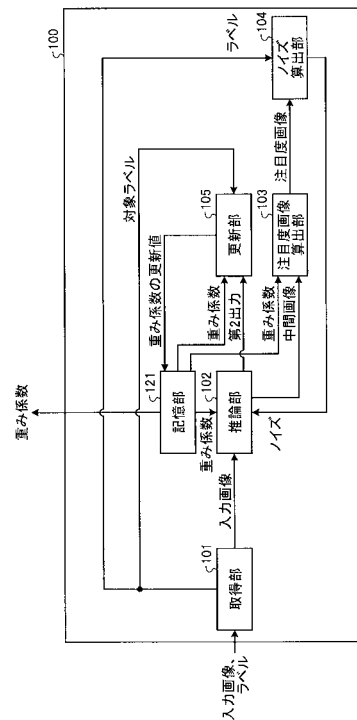
(54) 【発明の名称】 学習装置、学習方法、プログラムおよび認識装置

(57) 【要約】

【課題】ニューラルネットワークの過学習をより抑制する。

【解決手段】学習装置は、取得部と、推論部と、更新部と、を備える。取得部は、入力データと、入力データによる推論の正解を示す対象ラベルと、を取得する。推論部は、入力データをニューラルネットワークに入力したときの推論結果に対する入力データに含まれる複数の要素の寄与度に基づいたノイズを、入力データ、および、ニューラルネットワークの中間層のデータの少なくとも一方に付加し、入力データに対するニューラルネットワークによる推論を実行する。更新部は、ニューラルネットワークによる推論結果と、対象ラベルと、が一致するようにニューラルネットワークのパラメータを更新する。

【選択図】 図1



**【特許請求の範囲】****【請求項 1】**

入力データと、前記入力データによる推論の正解を示す対象ラベルと、を取得する取得部と、

前記入力データをニューラルネットワークに入力したときの推論結果に対する前記入力データに含まれる複数の要素の寄与度に基づいたノイズを、前記入力データ、および、前記ニューラルネットワークの中間層のデータの少なくとも一方に付加し、前記入力データに対する前記ニューラルネットワークによる推論を実行する推論部と、

前記ニューラルネットワークによる推論結果と、前記対象ラベルと、が一致するように前記ニューラルネットワークのパラメータを更新する更新部と、

を備える学習装置。

10

**【請求項 2】**

前記推論部は、前記対象ラベルと一致する推論結果に対する前記寄与度が閾値以上の要素に対して前記ノイズを付加する、

請求項 1 に記載の学習装置。

**【請求項 3】**

前記推論部は、前記ノイズを付加する要素の個数が上限値以下となるように、前記閾値を変更する、

請求項 2 に記載の学習装置。

**【請求項 4】**

前記推論部は、前記対象ラベルと一致しない推論結果に対する前記寄与度が閾値以上である要素に対して前記ノイズを付加する、

請求項 1 に記載の学習装置。

20

**【請求項 5】**

前記推論部は、前記対象ラベルと一致しない推論結果に対する前記寄与度が閾値より小さい要素に対して前記ノイズを付加する、

請求項 1 に記載の学習装置。

**【請求項 6】**

前記推論部は、前記対象ラベルと一致する推論結果に対する前記寄与度が第 1 閾値以上であり、かつ、前記対象ラベルと一致しない推論結果に対する前記寄与度が第 2 閾値以上である要素に対して前記ノイズを付加する、

請求項 1 に記載の学習装置。

30

**【請求項 7】**

前記推論部は、前記寄与度に基づく頻度で、前記ノイズを前記入力データおよび前記中間層のデータの少なくとも一方に付加する、

請求項 1 に記載の学習装置。

**【請求項 8】**

前記推論部は、

複数回実行される学習処理の各回で、前記入力データに対する前記ニューラルネットワークによる第 1 推論を実行し、前記第 1 推論の推論結果に対する前記入力データに含まれる複数の要素の寄与度に基づいたノイズを算出し、前記入力データ、および、前記中間層のデータの少なくとも一方に、算出したノイズを付加し、前記入力データに対する前記ニューラルネットワークによる第 2 推論を実行する、

請求項 1 に記載の学習装置。

40

**【請求項 9】**

前記推論部は、

複数回実行される学習処理の各回で、前記入力データおよび前記中間層のデータの少なくとも一方に前記ノイズを付加し、前記入力データに対する前記ニューラルネットワークによる推論を実行し、

複数回実行される学習処理のうち m 回目 ( m は 2 以上の整数 ) の学習処理で付加する前

50

記ノイズを、複数回実行される学習処理のうち (  $m - 1$  ) 回目の学習処理での前記ニューラルネットワークによる推論の推論結果に対する前記入力データに含まれる複数の要素の寄与度に基づいて算出する、

請求項 1 に記載の学習装置。

【請求項 1 0】

入力データと、前記入力データによる推論の正解を示す対象ラベルと、を取得する取得ステップと、

前記入力データをニューラルネットワークに入力したときの推論結果に対する前記入力データに含まれる複数の要素の寄与度に基づいたノイズを、前記入力データ、および、前記ニューラルネットワークの中間層のデータの少なくとも一方に付加し、前記入力データ 10

に対する前記ニューラルネットワークによる推論を実行する推論ステップと、

前記ニューラルネットワークによる推論結果と、前記対象ラベルと、が一致するように前記ニューラルネットワークのパラメータを更新する更新ステップと、

を含む学習方法。

【請求項 1 1】

コンピュータを、

入力データと、前記入力データによる推論の正解を示す対象ラベルと、を取得する取得部と、

前記入力データをニューラルネットワークに入力したときの推論結果に対する前記入力データに含まれる複数の要素の寄与度に基づいたノイズを、前記入力データ、および、前記ニューラルネットワークの中間層のデータの少なくとも一方に付加し、前記入力データ 20

に対する前記ニューラルネットワークによる推論を実行する推論部と、

前記ニューラルネットワークによる推論結果と、前記対象ラベルと、が一致するように前記ニューラルネットワークのパラメータを更新する更新部と、

【請求項 1 2】

請求項 1 に記載の学習装置により学習された前記ニューラルネットワークを用いて、入力データの認識を実行する認識部と、  
前記認識部による認識結果を出力する出力制御部と、  
を備える認識装置。 30

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本発明の実施形態は、学習装置、学習方法、プログラムおよび認識装置に関する。

【背景技術】

【0 0 0 2】

ニューラルネットワークの過学習を抑制する方法として、中間層のノード値の一部をランダムにゼロにしながら最適化する方法、および、入力データの一部をランダムに遮蔽しながら最適化する方法が提案されている。

【先行技術文献】 40

【特許文献】

【0 0 0 3】

【特許文献 1】国際公開第 2 0 1 4 / 1 0 5 8 6 6 号

【非特許文献】

【0 0 0 4】

【非特許文献 1】G. Ghiasi, T. Lin, Q. V. Le, "DropBlock: A regularization method for convolutional networks," in arXiv:1810.12890v1, 2018.

【非特許文献 2】T. DeVries and G. W. Taylor, "Improved Regularization of Convolutional Neural Networks with Cutout", in arXiv:1708.04552, 2017

.

10

20

30

40

50

【非特許文献3】Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, “Random Erasing Data Augmentation”, in arXiv:1708.04896, 2017.

【非特許文献4】B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, “Learning Deep Features for Discriminative Localization”, In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2921-2929, 2016.

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかしながら、従来技術では、ノイズを付加する頻度、データ中でノイズを付加する位置（要素）、ノイズのサイズ、および、学習進捗に合わせたスケジュールなどの、ノイズの付加方法を調整するパラメータの設定が難しく、過学習を十分に抑制できない場合があった。

【課題を解決するための手段】

【0006】

実施形態の学習装置は、取得部と、推論部と、更新部と、を備える。取得部は、入力データと、入力データによる推論の正解を示す対象ラベルと、を取得する。推論部は、入力データをニューラルネットワークに入力したときの推論結果に対する入力データに含まれる複数の要素の寄与度に基づいたノイズを、入力データ、および、ニューラルネットワークの中間層のデータの少なくとも一方に付加し、入力データに対するニューラルネットワークによる推論を実行する。更新部は、ニューラルネットワークによる推論結果と、対象ラベルと、が一致するようにニューラルネットワークのパラメータを更新する。

【図面の簡単な説明】

【0007】

【図1】第1の実施形態にかかる学習装置のブロック図。

【図2】第1の実施形態における更新処理のフローチャート。

【図3】入力画像の一例を示す図。

【図4】注目度画像の一例を示す図。

【図5】ノイズおよびノイズを付加した画像の例を示す図。

【図6】変形例における更新処理のフローチャート。

【図7】第2の実施形態にかかる情報処理システムのブロック図。

【図8】第1または第2の実施形態にかかる装置のハードウェア構成図。

【発明を実施するための形態】

【0008】

以下に添付図面を参照して、この発明にかかる学習装置の好適な実施形態を詳細に説明する。

【0009】

以下では、画像を入力データとして推論を行うニューラルネットワークを学習する学習装置を例に説明する。より具体的には、入力された画像（2次元画像）を、数字の0～9のいずれかに分類する10クラスの画像分類問題を例に説明する。適用可能な処理はこれに限られるものではなく、その他の推論処理（認識処理、分類処理を含む）にも適用できる。例えば、1クラスの分類処理（異常検出など）などの10クラス以外の分類処理にも適用できる。また、3次元画像を用いた推論、音声データからの推論対象（特定の話者など）の推論、および、テキストデータを用いた推論などにも適用できる。

【0010】

（第1の実施形態）

第1の実施形態は、ニューラルネットワークの推論（認識）の寄与度に基づいて、付加するノイズのパラメータを決定する。これにより、付加するノイズのパラメータが、学習の進捗および認識率に対して適応的に決定されるため、過学習をより安定して抑制できる。また、このように学習されたニューラルネットワークを用いれば、認識率を向上させる

10

20

30

40

50

ことができる。

【0011】

図1は、第1の実施形態にかかる学習装置100を示すブロック図である。学習装置100は、取得部101と、推論部102と、注目度画像算出部103と、ノイズ算出部104と、更新部105と、記憶部121と、を備える。

【0012】

記憶部121は、学習装置100による各種処理に用いる各種データを記憶する。例えば記憶部121は、学習するニューラルネットワークを定めるパラメータ、および、学習のための訓練データを記憶する。ニューラルネットワークを定めるパラメータは、例えば重み係数およびバイアスである。

10

【0013】

記憶部121は、フラッシュメモリ、メモリカード、RAM (Random Access Memory)、HDD (Hard Disk Drive)、および、光ディスクなどの一般的に利用されているあらゆる記憶媒体により構成することができる。

【0014】

取得部101は、学習装置100による各種処理で用いる各種情報を取得する。例えば取得部101は、入力画像(入力データの一例)、および、入力画像による推論の正解を示す対象ラベルを含む訓練データを取得する。取得部101が訓練データを取得する方法はどのような方法であってもよい。取得部101は、例えば、記憶部121に記憶された訓練データを読み出すことにより取得してもよいし、学習装置100の外部のサーバ装置などから訓練データを受信することにより取得してもよい。

20

【0015】

推論部102は、入力画像に対するニューラルネットワークによる推論を実行する。例えば推論部102は、入力画像をニューラルネットワークに入力し、ニューラルネットワークから出力される推論結果を得る。

【0016】

注目度画像算出部103は、入力画像をニューラルネットワークに入力したときの入力画像の推論に寄与した領域を示す注目度画像を生成する。注目度画像は、領域(画素など)ごとの推論時に注目(着目)した度合い(注目度)を示す画像と解釈することができる。注目度画像は、顕著性マップ(saliency map)などと呼ばれる場合がある。注目度画像の生成方法の詳細は後述する。

30

【0017】

ノイズ算出部104は、入力画像と、ニューラルネットワークの中間層のデータ(中間層データ)と、のうち少なくとも一方に付加するノイズを算出する。例えばノイズ算出部104は、対象ラベルと注目度画像とからノイズを算出する。なお、中間層データにノイズを付加することは、中間層データに対応する重み係数(中間層データに乘じられる重み係数)にノイズを付加することと等価である。

【0018】

推論部102は、ノイズ算出部104により算出されたノイズを、入力画像および中間層データ(重み係数)の少なくとも一方に付加し、入力データに対するニューラルネットワークによる推論をさらに実行する。

40

【0019】

更新部105は、ニューラルネットワークのパラメータを更新する。例えば更新部105は、入力画像および中間層データの少なくとも一方にノイズを付加した後のニューラルネットワークによる推論結果と、対象ラベルと、が一致するように、誤差逆伝搬法によりニューラルネットワークの重み係数を更新する。パラメータの更新方法は誤差逆伝播法に限られるものではなく、どのような方法であってもよい。

【0020】

上記各部(取得部101、推論部102、注目度画像算出部103、ノイズ算出部104、更新部105)は、例えば、1または複数のプロセッサにより実現される。例えば上

50

記各部は、CPU (Central Processing Unit) などのプロセッサにプログラムを実行させること、すなわちソフトウェアにより実現してもよい。上記各部は、専用のIC (Integrated Circuit) などのプロセッサ、すなわちハードウェアにより実現してもよい。上記各部は、ソフトウェアおよびハードウェアを併用して実現してもよい。複数のプロセッサを用いる場合、各プロセッサは、各部のうち1つを実現してもよいし、各部のうち2以上を実現してもよい。

【0021】

次に、このように構成された第1の実施形態にかかる学習装置100による更新処理について説明する。図2は、第1の実施形態における更新処理の一例を示すフローチャートである。

【0022】

まず、取得部101は、訓練データである入力画像およびラベルを取得する(ステップS101)。以下では、入力画像を $x_i$  ( $i = 1, 2, \dots, N$ )、ラベルを $t_i$ と表す。なお、 $i$ は訓練データを識別する識別情報(通し番号など)を表し、 $N$ は訓練データの数を表す。

【0023】

図3は、入力画像の一例を示す図である。入力画像 $x_i$ は、例えば横幅 $W$ 、縦幅 $H$ の画素集合であり、 $W \times H$ 次元のベクトルとする。図3に示すように、入力画像301は、 $x_i(u, v)$ と表される。 $u$ は水平方向の画素位置を表し、 $v$ は垂直方向の画素位置を表す。

【0024】

ラベル $t_i$ は、入力画像 $x_i$ による推論の正解に該当する要素が1となり、それ以外の要素が0となる10次元ベクトルである。図3のように、正解が「5」である入力画像 $x_i$ の場合は、ラベル $t_i$ は、 $(0, 0, 0, 0, 0, 1, 0, 0, 0, 0)^T$ のように表される。以下ではラベル $t_i$ のうち、正解を示す要素を対象ラベルといい、それ以外の要素を非対象ラベルという場合がある。

【0025】

図2に戻り、推論部102は、入力画像 $x_i$ をニューラルネットワークに入力し、この入力に対するニューラルネットワークの出力 $y_{1i}$ (第1出力)を取得する(ステップS102)。出力 $y_{1i}$ は、以下の(1)式により表される。

$$y_{1i} = f(w, x_i) \quad \dots (1)$$

【0026】

$f$ は、重み係数集合 $w$ を保持するニューラルネットワークの関数である。例えば関数 $f$ は、畳み込み層、全結合層、正規化層、および、プーリング層などの複数の層の処理を繰り返して、ラベル $t_i$ と同じ次元数の10次元ベクトルを出力する。なお、関数 $f$ の出力ベクトルは、ソフトマックス処理後の値であって、すべて要素が非負の値であり、かつ、要素の総和が1に正規化されているものとする。

【0027】

次に、注目度画像算出部103は、中間層データ(中間画像)およびパラメータ(重み係数)などに基づいて注目度画像を生成する(ステップS103)。注目度画像 $s_i$ は、例えば以下の(2)式により生成される。

$$s_i = g(w, x_i) \quad \dots (2)$$

【0028】

$g$ は、ニューラルネットワークの重み係数集合 $w$ 、および、入力画像 $x_i$ に基づいて注目度画像を出力する関数である。注目度画像は、例えば非特許文献4に示されるクラス・アクティベーション・マッピング(Class Activation Mapping)などを用いることができる。クラス・アクティベーション・マッピングを用いる場合は、入力画像 $x_i$ を推論したときのニューラルネットワークの中間画像を、対象ラベルに関連する重み係数集合 $w$ の値で線形合成する関数が、関数 $g$ として使用して生成できる。中間画像は、ニューラルネットワークの中間層の出力に相当する画像である。

10

20

30

40

50

## 【0029】

図4は、注目度画像の一例を示す図である。図4は、正解が「9」である入力画像401に対する推論結果をもとに算出される注目度画像402の例を示す。注目度画像402の上部の数値(0~9)は、注目度画像402の10個の要素それぞれに対応する推論結果を表す。注目度画像402は、推論の寄与度に応じた画素値を有する画像である。図4では、ハッチングの種類の違いにより画素値(寄与度)の違いを表しているが、例えば非特許文献4と同様に、色により画素値(寄与度)を表すように構成してもよい。

## 【0030】

図4の例では、「9」に対応する注目度画像の画素411、および、「4」に対応する注目度画像の一部の画素412が、寄与度が大きい画素となっている。

10

## 【0031】

図4に示す注目度画像は一例であり、その他のどのような注目度画像を用いてもよい。多くの注目度画像の算出方法は、図4の例と同様に、出力 $y_{1i}$ のそれぞれの要素ごとに寄与度を算出することが可能である。算出方法によって注目度画像の縦幅と横幅は異なるが、本実施形態では入力画像 $x_i$ と同じ縦幅および横幅で注目度画像を算出することとする。すなわち、本実施形態では、注目度画像 $s_i$ は、 $(W \times H \times C)$ 次元ベクトルとなる。Cは、分類するクラスの個数に相当する。本実施形態では $C = 10$ である。

## 【0032】

以上のように、注目度画像は、入力画像をニューラルネットワークに入力したときの入力画像の推論に寄与した領域を示す画像である。推論に寄与した領域は、例えば、その領域の値が変わることによる推論結果への影響が大きい領域、出力のうち活性がある部分からニューラルネットワークの経路を逆にたどって得られる領域、などを含む。

20

## 【0033】

図2に戻り、ノイズ算出部104は、入力画像 $x_i$ および重み係数集合 $w$ (中間層データ)に付加するノイズを、ラベル $t_i$ と注目度画像 $s_i$ に基づいて算出する(ステップS104)。

## 【0034】

以下では、対象ラベル $c$ に対応する注目度画像 $s_i$ に基づいて、入力画像 $x_i$ にノイズを付加する方法を説明する。ノイズ算出部104は、以下の(3)式により付加するノイズ $n_i$ を算出する。Tは、閾値である。 $s_i(u, v, c)$ は、水平方向の画素位置 $u$ 、垂直方向の画素位置 $v$ 、対象ラベル $c$ に対応する注目度画像の画素値を表す。

30

## 【0035】

$$\begin{aligned} n_i(u, v) &= 0 && (s_i(u, v, c) \geq T \text{ のとき}), \\ n_i(u, v) &= 1 && (s_i(u, v, c) < T \text{ のとき}) \quad \dots (3) \end{aligned}$$

## 【0036】

後述するように、このノイズ $n_i$ は、入力画像 $x_i$ の画素(要素)ごとに画素値に乘じられることにより、入力画像 $x_i$ に付加される。従って(3)式のノイズは、寄与度が閾値T以上である画素をゼロにしながらか最適化するためのノイズであると解釈することができる。

## 【0037】

(3)式では、入力画像 $x_i$ の画素のうち、値が「0」の要素が乘じられた画素の画素値がゼロになり、値が「1」の要素が乘じられた画素の画素値は、元の値が維持される。従って、値が「0」の要素に対応する画素(要素)に対して、ノイズが付加されると解釈することもできる。

40

## 【0038】

なお、(3)式に示すように、ノイズは、画素ごとの寄与度と閾値との比較結果に応じて求められる。ノイズを求めることができる情報(寄与度など)が得られれば、ステップS103で注目度画像を生成する必要はない。

## 【0039】

次に、推論部102は、入力画像 $x_i$ にノイズ $n_i$ を付加した画像をニューラルネット

50

ワークに入力し、この入力に対する出力  $y_{2i}$  (第2出力) を取得する (ステップ S 1 0 5)。出力  $y_{2i}$  は、以下の (4) 式により表される。記号「 $\cdot$ 」は、ベクトルの要素ごとの積を算出するアダマール積を表す。

$$y_{2i} = f(w, x_i \cdot n_i) \quad \cdots (4)$$

【0040】

図5は、ノイズ、および、ノイズを付加した画像 ( $x_i \cdot n_i$ ) の例を示す図である。図5の入力画像501は、正解が「9」である入力画像である。注目度画像502は、入力画像501に対する推論結果をもとに算出される注目度画像の例である。入力画像501は、数字「9」の右上部が切れた画像となっている。このため、図4の入力画像401と比較して、入力画像501は、数字の「4」または「7」であると誤って推論される可能性が高い画像である。すなわち、図5の例では、「4」に対応する注目度画像の画素511、「7」に対応する注目度画像の一部の画素512、および、「9」に対応する注目度画像の一部の画素513が、寄与度が大きい画素となっている。

10

【0041】

本実施形態では、図5に示すように、正解である「9」に対応する注目度画像に対して現時点で推論に寄与している画素の画素値をゼロとするように、入力画像501にノイズ503が付加され、新しい画像504が得られる。これにより、ニューラルネットワークは、画素値がゼロに変更された領域 (注目度画像への寄与度が大きい領域) 以外の領域を含む新しい画像から正しく推論が実行されるように学習が進められる。このため、入力画像のより広い範囲から特徴が抽出できるようになり、過学習の抑制が期待できる。

20

【0042】

また、学習初期の段階では注目度画像の画素値は小さくランダムに近いが、学習が進むにつれて局所的な領域で画素値が大きくなる傾向にある。このため、学習が進むにつれて、寄与度が大きい領域、すなわち、画素値がゼロにされる領域の位置およびサイズが変更される。これは、閾値  $T$  を調整するのみで、ノイズを付加する頻度、ノイズを付加する位置 (要素)、および、ノイズのサイズが容易に調整可能となることを意味する。

【0043】

図2に戻り、更新部105は、ラベル  $t_i$  および出力  $y_{2i}$  に基づいて重み係数集合  $w$  の値を更新する (ステップ S 1 0 6)。例えば更新部105は、対象ラベルと出力  $y_{2i}$  との一致度が高いほど小さくなるようにロス  $L$  を算出し、誤差逆伝搬法などの方法により、ロス  $L$  を最小化するように、重み係数集合  $w$  を更新する。更新部105は、例えば、以下の (5) 式によりロス  $L$  を算出する。

30

$$L = - \sum_i (t_i \cdot \ln(y_{2i})) \quad \cdots (5)$$

【0044】

次に、更新部105は、学習が終了したか否かを判定する (ステップ S 1 0 7)。更新部105は、例えば、出力  $y_{1i}$  または出力  $y_{2i}$  と対象ラベルとの一致度、ロス  $L$  の絶対値、ロス  $L$  の減少幅、および、重み係数の更新回数 (ステップ S 1 0 2 からステップ S 1 0 6 までの学習処理の回数) などにより、学習の終了を判定する。学習を継続する場合は (ステップ S 1 0 7 : No)、ステップ S 1 0 2 に戻り、処理が繰り返される。学習が終了したと判定された場合 (ステップ S 1 0 7 : Yes)、更新処理を終了する。

40

【0045】

以上のように、本実施形態は、ニューラルネットワークの学習時に入力画像に付加するノイズの頻度、位置、サイズなどの調整パラメータを注目度画像に基づいて調整する。このため、適応的な正則化が実現でき、過学習をより抑制することが可能となる。

【0046】

(変形例1)

上記実施形態では、通常は複数回実行される学習処理の各回で、入力画像に対するニューラルネットワークによる推論 (第1推論、ステップ S 1 0 2) と、ノイズを付加した入力画像に対するニューラルネットワークによる推論 (第2推論、ステップ S 1 0 3 ~ ステップ S 1 0 5) と、が実行される。すなわち、繰り返される学習処理の各回で、ニューラ

50

ルネットワークによる推論が2回実行される。

【0047】

学習装置100が、学習処理の各回で1回の推論を実行するように構成してもよい。例えば学習装置100は、学習処理のうち $m$ 回目 ( $m$ は2以上の整数)の学習処理で付加するノイズを、( $m-1$ )回目の学習処理でのニューラルネットワークによる推論の推論結果に対する寄与度に基づいて算出する。

【0048】

図6は、このように構成される本変形例における更新処理の一例を示すフローチャートである。ステップS201は、図2のステップS101と同様であるため説明を省略する。

10

【0049】

次に、推論部102は、入力画像 $x_i$ にノイズ $n_i$ を付加した画像をニューラルネットワークに入力し、この入力に対する出力 $y_{2i}$ を取得する(ステップS202)。初回の学習処理では、例えばノイズを付加しない画像、または、ランダムに選択した画素の画素値をゼロにするようなノイズを加えた画像を用いる。

【0050】

次に更新部105は、ラベル $t_i$ および出力 $y_{2i}$ に基づいて重み係数集合 $w$ の値を更新する(ステップS203)。この処理は、図2のステップS106と同様である。次に更新部105は、学習が終了したか否かを判定する(ステップS204)。この処理は、図2のステップS107と同様である。

20

【0051】

次に、注目度画像算出部103は、中間層データ(中間画像)およびパラメータ(重み係数)などに基づいて注目度画像を生成する(ステップS205)。本変形例では、注目度画像算出部103は、ノイズが付加された入力画像の推論時の中間層データなどから、注目度画像を生成する。

【0052】

次にノイズ算出部104は、次回の学習処理の入力画像 $x_i$ および重み係数集合 $w$ (中間層データ)の少なくとも一方に付加するノイズを、対象ラベルと注目度画像 $s_i$ に基づいて算出する(ステップS206)。この処理は、図2のステップS104と同様である。

30

【0053】

ステップS206でノイズを算出した後、ステップS202に戻り、次回の学習処理が繰り返される。推論部102は、このときに入力画像に付加するノイズとして、ステップS206で算出されたノイズを用いる。すなわち推論部102は、( $m-1$ )回目の学習処理で算出されたノイズ $n_i$ を、 $m$ 回目の学習処理で入力画像 $x_i$ に付加して推論を行う。

【0054】

このような構成により、ニューラルネットワークによる推論の回数を減らし、演算の負荷を削減することが可能となる。

【0055】

(変形例2)

上記のように、ノイズは、中間層データに付加されてもよい。本変形例では、中間層データにノイズを付加する方法について説明する。

40

【0056】

上記のように畳み込み層を含むようなニューラルネットワークを用い、畳み込み層のデータにノイズを付加する場合、推論部102は、中間層の画像サイズに合うように、注目度画像のサイズを変更(正規化)する。画像サイズの変更方法は、縦および横の画素数が一致するように画像を縮小または拡大する方法などの、従来から用いられているどのような方法を適用してもよい。

【0057】

50

(3)式のように、画素(要素)ごとに寄与度に応じたノイズが付加されるようにノイズが算出される場合、中間層データは、注目度画像の画素の位置に対応する要素の位置を特定可能なデータである必要がある。例えば、上記のように畳み込み層、全結合層、正規化層、および、プーリング層などの複数の層を含む畳み込みニューラルネットワーク(CNN: Convolutional Neural Network)を対象とする場合、畳み込み層およびプーリング層などが、位置を特定可能な中間層データとして用いることができる。なお適用可能なニューラルネットワークはCNNに限られるものではない。

【0058】

全結合層では画素(要素)の位置の情報が(全結合により)失われるため、例えば寄与度に対応する画素(要素)を定めることができない。このため全結合層のデータにノイズを付加する場合、ノイズを付加する位置(要素)ではなく、ノイズを付加する頻度が寄与度に応じて制御される。例えばノイズ算出部104は、ノイズを付加する頻度を、注目度画像に基づき算出される値で決定する。

10

【0059】

例えばノイズ算出部104は、注目度画像の画素値の平均値、最頻値、および、最大値を算出する。また、ノイズ算出部104は、付加するノイズを算出する。例えばノイズ算出部104は、要素の一部をランダムにゼロにするようなノイズを算出する。

【0060】

推論部102は、算出された値が閾値T以上である場合に、全結合層のデータにノイズを付加する。

20

【0061】

ノイズ算出部104は、画素値の平均値、最頻値、および、最大値を算出するときに、注目度画像の各画素値をそのまま使用するのではなく、例えば、画像の中央を基準としたガウス分布に従う重みを加えた画素値を使用してもよい。例えばノイズ算出部104は、注目度画像の中央で最大値となるようなガウス分布に従う重みを各画素値に乗じて加重平均を算出してもよい。

【0062】

ノイズ算出部104は、注目度画像の全画素の画素値を用いる代わりに、一部の画素の画素値のみを用いてもよい。例えば入力画像ごとに、注目すべき領域の位置を示すデータ(注目位置データ)が定められている場合、ノイズ算出部104は、この注目位置データが示す領域に対応する画素の画素値を用いて平均値等を算出してもよい。

30

【0063】

入力画像および中間層データの両方にノイズを付加する場合は、上記実施形態と本変形例を組み合わせ適用すればよい。

【0064】

(変形例3)

ノイズは、上記(3)式のように算出されるノイズに限られるものではない。例えば、ノイズ算出部104は、ゼロの代わりに、要素(画素)ごとに、または、全要素一括でランダム値を算出してもよい。ノイズ算出部104は、ゼロの代わりに、周辺画素の画素値の平均値を算出してもよい。

40

【0065】

ノイズの付加方法は、上記(4)式のように入力画像とのアダマール積により付加する方法に限られるものではない。より一般的には、推論部102は、以下の(6)式のようにノイズを付加してもよい。 $x_i'$ は、例えば固定の画素値を有する画像(ベタ画像)、または、周辺画素の画素値の平均値を各画素値とする画像などである。

$$n_i \cdot x_i + (1 - n_i) \cdot x_i' \quad \dots (6)$$

【0066】

なお、 $x_i'$ の画素値がすべて0かつ、 $n_i$ が0または1の場合、(6)式は、(4)式のアダマール積( $x_i \cdot n_i$ )と等価となる。 $x_i'$ がベタ画像の場合、(6)式は、このベタ画像を入力画像にブレンドするようにノイズを付加する式となる。 $x_i'$ が周辺

50

画素の画素値の平均値を各画素値とする画像の場合、(6)式は、入力画像をぼかすようにノイズを付加する式となる。なお、 $n_i$ は1か0の2通りではなく、例えば0.5などの中間値を用いることもできる。その場合、 $n_i$ は注目度画像の画素値を閾値Tで二値化するのではなく、例えば、 $n_i = \text{Clip}(s_i(u, v, c) / T, \text{MIN} = 0, \text{MAX} = 1)$ などとしてもよい。

【0067】

(変形例4)

閾値Tは、固定値であってもよいし、変更される値であってもよい。例えば推論部102は、ノイズを付加する要素(画素)の個数が上限値以下となるように、閾値Tを変更してもよい。上限値は、例えば、入力画像の画素数(または、中間層データの要素数)に予め定められた割合を乗じた値として決定されてもよい。

10

【0068】

(変形例5)

上記実施形態では、ラベル $t_i$ のうち対象ラベルcに対応する注目度画像 $s_i$ に基づいてノイズが付加された。ラベル $t_i$ のうち対象ラベルc以外の非対象ラベル $c'$ に対応する注目度画像 $s_i$ に基づいてノイズが付加されてもよい。例えば、ノイズ算出部104は、以下の(7)式により付加するノイズ $n_i$ を算出してもよい。 $T'$ は、非対象ラベルの注目度画像 $s_i$ に対して定められる閾値である。 $s_i(u, v, c')$ は、水平方向の画素位置u、垂直方向の画素位置v、非対象ラベル $c'$ に対応する注目度画像の画素値を表す。

20

【0069】

$$\begin{aligned} n_i(u, v) &= 0 && (\text{Max}(s_i(u, v, c')) < T' \text{ のとき}), \\ n_i(u, v) &= 1 && (\text{それ以外}) \quad \dots (7) \end{aligned}$$

【0070】

(7)式は、誤って注目している要素を考慮しないようにノイズを付加する式と解釈することができる。

【0071】

ノイズ算出部104は、以下の(8)式により付加するノイズ $n_i$ を算出してもよい。

【0072】

$$\begin{aligned} n_i(u, v) &= 0 && (s_i(u, v, c') < T' \text{ のとき}), \\ n_i(u, v) &= 1 && (s_i(u, v, c') \geq T' \text{ のとき}) \quad \dots (8) \end{aligned}$$

30

【0073】

(8)式は、誤って注目している要素を用いて正しく推論が実行されるように学習するためのノイズを付加する式と解釈することができる。

【0074】

ノイズ算出部104は、対象ラベルcに対応する注目度画像と、非対象ラベル $c'$ に対応する注目度画像と、の両方を用いてノイズを付加してもよい。例えば、ノイズ算出部104は、以下の(9)式により付加するノイズ $n_i$ を算出してもよい。

【0075】

$$\begin{aligned} n_i(u, v) &= 0 && (s_i(u, v, c) \geq T, \text{かつ}, s_i(u, v, c') < T' \text{ のとき}), \\ n_i(u, v) &= 1 && (\text{それ以外}) \quad \dots (9) \end{aligned}$$

40

【0076】

(9)式は、対象ラベルおよび非対象ラベルの両方で注目されている領域(共通する画像特徴)を除外し、この領域以外の領域に注目するようにノイズを付加する式と解釈することができる。

【0077】

例えば数字の「1」、「7」、「9」などは、下部(上下方向に延びる線)に相当する画素群が類似するが、それ以外の部分の画素群は類似しない場合が多い。また、トラックと乗用車とを分類するニューラルネットワークを学習する例では、車両の前部(ヘッドラ

50

イト、フロントガラス、運転席などを含む部分)に相当する画素群が類似するが、それ以外の部分の画素群は類似しない場合が多い。本変形例を適用すれば、類似しない部分の情報から正しく推論が実行されるように学習を進めることが可能となる。

【0078】

(第2の実施形態)

第2の実施形態では、学習装置で学習したニューラルネットワークを用いて認識処理を実行する認識装置を備える情報処理システムについて説明する。

【0079】

図7は、第2の実施形態にかかる情報処理システムの構成の一例を示すブロック図である。図7に示すように、情報処理システムは、学習装置100-2と、認識装置200-2と、ネットワーク300-2により接続された構成となっている。

10

【0080】

ネットワーク300-2は、インターネットなどであるが、その他のどのような形態のネットワークであってもよい。例えば、ネットワーク300-2は、有線ネットワークおよび無線ネットワークのいずれであってもよい。

【0081】

学習装置100-2は、出力制御部106-2が追加されたことが、第1の実施形態の学習装置100と異なっている。その他の構成は第1の実施形態の学習装置100と同様であるため、同一の符号を付し説明を省略する。

【0082】

出力制御部106-2は、学習装置100-2からの各種情報の出力を制御する。例えば出力制御部106-2は、更新部105により更新されたニューラルネットワークのパラメータを認識装置200-2に出力する。

20

【0083】

認識装置200-2は、通信制御部201と、取得部202と、認識部203と、出力制御部204と、記憶部221と、を備えている。

【0084】

記憶部221は、認識装置200-2で実行される各種処理で用いる各種データを記憶する。例えば記憶部221は、学習装置100-2から送信されたニューラルネットワークのパラメータを記憶する。記憶部221は、フラッシュメモリ、メモリカード、RAM、HDD、および、光ディスクなどの一般的に利用されているあらゆる記憶媒体により構成することができる。

30

【0085】

通信制御部201は、学習装置100-2などの外部装置との間の通信を制御する。例えば通信制御部201は、学習装置100-2から、学習済みのニューラルネットワークを定めるパラメータの情報などを受信する。

【0086】

取得部202は、認識装置200-2による各種処理に用いる各種データを取得する。例えば取得部202は、認識処理の対象となる入力データ(画像など)を取得する。

【0087】

認識部203は、取得された画像をニューラルネットワークに入力することにより、画像を認識する。

40

【0088】

出力制御部204は、認識装置200-2からの各種情報の出力を制御する。例えば出力制御部204は、認識部203による認識結果を出力する。出力方法は、どのような方法であってもよいが、例えば、表示装置に表示する方法、外部のサーバ装置などに送信する方法などを用いることができる。

【0089】

なお、図7に示す情報処理システムの構成は一例であり、これに限られるものではない。例えば、認識装置200-2の各部と学習装置100-2の各部とを1つの装置(認識

50

装置)内に備える構成としてもよい。この場合、共通化できる機能(例えば、取得部101と取得部202、記憶部121と記憶部221)は共通化してもよい。また認識装置200-2は、物理的に1つの装置によって構成されてもよいし、物理的に複数の装置によって構成されてもよい。例えば認識装置200-2は、クラウド環境上で構築されてもよい。

#### 【0090】

認識処理は、どのようなデータを用いて、どのような対象を認識する処理であってもよい。例えば認識処理は、会議室内で集音された音声から、特定の話者の音声を認識する処理であってもよい。また認識処理は、撮像装置(カメラ)により撮影された画像から、撮影した対象物の異常を検出(認識)する処理であってもよい。このような認識処理は、例えば、工場内の機器を画像により外観検査し異常を検出する処理、および、医療用画像から異常(患部)を検出する処理などに適用できる。

10

#### 【0091】

以上説明したとおり、第1および第2の実施形態によれば、ニューラルネットワークの過学習をより抑制することが可能となる。

#### 【0092】

次に、第1または第2の実施形態にかかる各装置(学習装置、認識装置)のハードウェア構成について図8を用いて説明する。図8は、第1または第2の実施形態にかかる装置のハードウェア構成例を示す説明図である。

#### 【0093】

第1または第2の実施形態にかかる装置は、CPU(Central Processing Unit)51などの制御装置と、ROM(Read Only Memory)52やRAM(Random Access Memory)53などの記憶装置と、ネットワークに接続して通信を行う通信I/F54と、各部を接続するバス61を備えている。

20

#### 【0094】

第1または第2の実施形態にかかる装置で実行されるプログラムは、ROM52等に予め組み込まれて提供される。

#### 【0095】

第1または第2の実施形態にかかる装置で実行されるプログラムは、インストール可能な形式または実行可能な形式のファイルでCD-ROM(Compact Disk Read Only Memory)、フレキシブルディスク(FD)、CD-R(Compact Disk Recordable)、DVD(Digital Versatile Disk)等のコンピュータで読み取り可能な記録媒体に記録してコンピュータプログラムプロダクトとして提供されるように構成してもよい。

30

#### 【0096】

さらに、第1または第2の実施形態にかかる装置で実行されるプログラムを、インターネット等のネットワークに接続されたコンピュータ上に格納し、ネットワーク経由でダウンロードさせることにより提供するように構成してもよい。また、第1または第2の実施形態にかかる装置で実行されるプログラムをインターネット等のネットワーク経由で提供または配布するように構成してもよい。

#### 【0097】

第1または第2の実施形態にかかる装置で実行されるプログラムは、コンピュータを上述した装置の各部として機能させる。このコンピュータは、CPU51がコンピュータ読取可能な記憶媒体からプログラムを主記憶装置上に読み出して実行することができる。

40

#### 【0098】

本発明のいくつかの実施形態を説明したが、これらの実施形態は、例として提示したものであり、発明の範囲を限定することは意図していない。これら新規な実施形態は、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ともに、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。

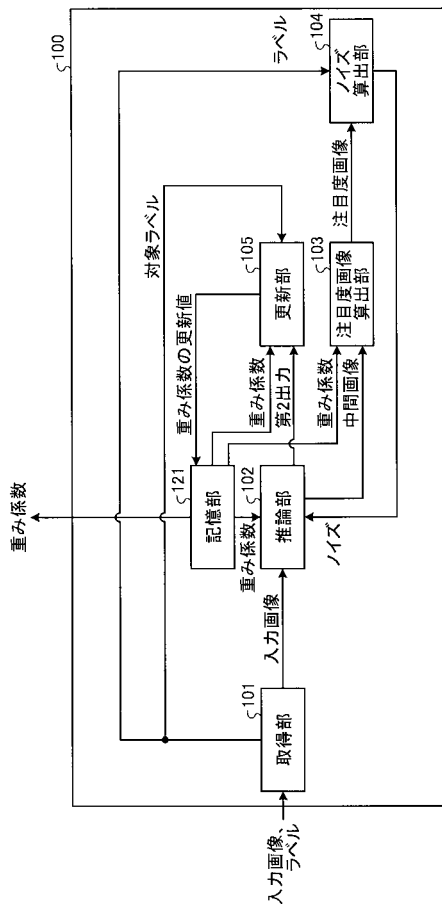
50

【符号の説明】

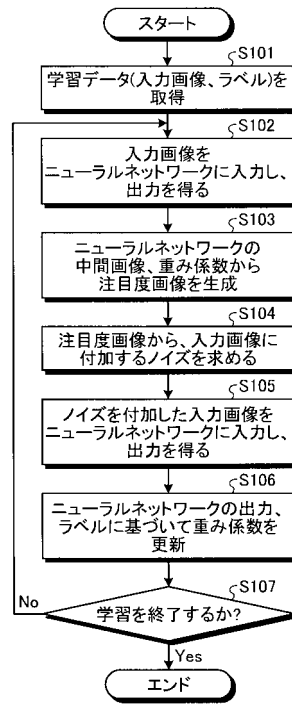
【0099】

- 100、100-2 学習装置
- 101 取得部
- 102 推論部
- 103 注目度画像算出部
- 104 ノイズ算出部
- 105 更新部
- 106-2 出力制御部
- 121 記憶部
- 200-2 認識装置
- 201 通信制御部
- 202 取得部
- 203 認識部
- 204 出力制御部
- 221 記憶部
- 300-2 ネットワーク

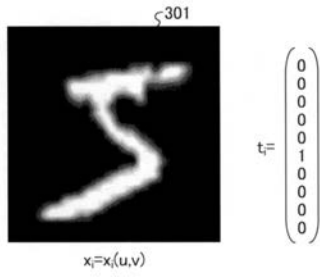
【図1】



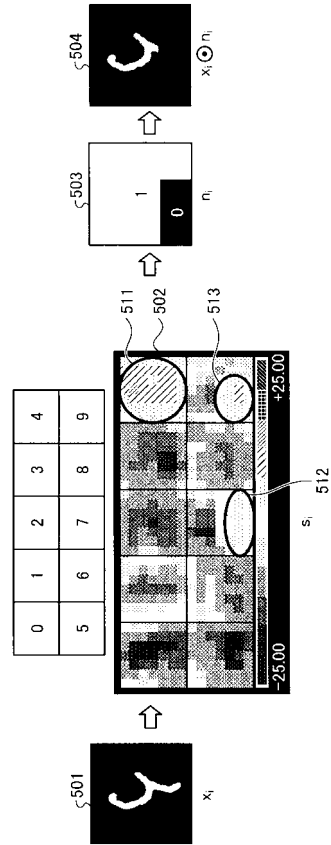
【図2】



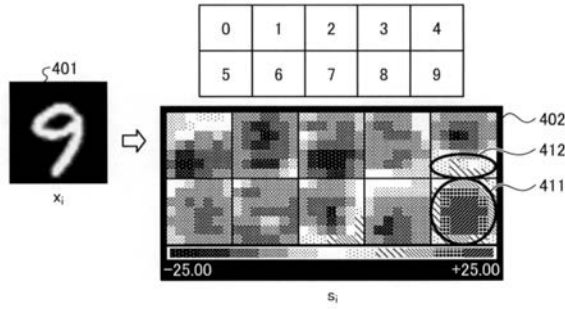
【 図 3 】



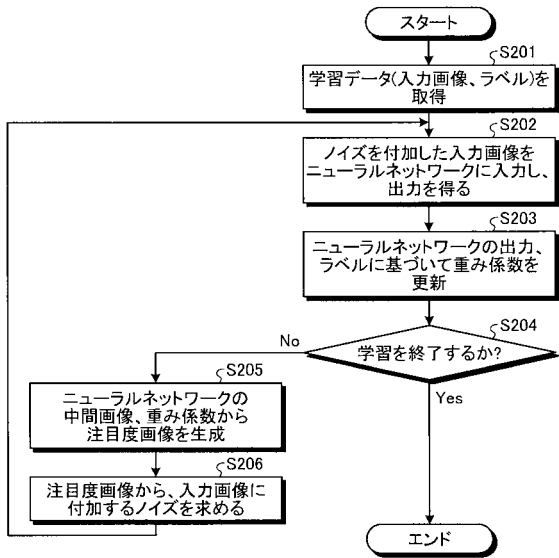
【 図 5 】



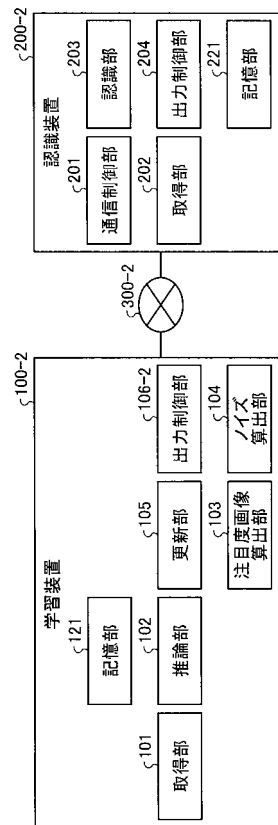
【 図 4 】



【 図 6 】



【 図 7 】



【 図 8 】

