

(19)日本国特許庁(JP)

(12)公開特許公報(A)

(11)公開番号
特開2023-35976
(P2023-35976A)

(43)公開日 令和5年3月13日(2023.3.13)

(51)国際特許分類	F I	テーマコード (参考)
G 0 6 N 20/00 (2019.01)	G 0 6 N 20/00	1 3 0 5 L 0 9 6
G 0 6 T 7/00 (2017.01)	G 0 6 T 7/00	3 5 0 C
G 0 6 V 10/82 (2022.01)	G 0 6 V 10/82	
G 0 6 V 10/778(2022.01)	G 0 6 V 10/778	

審査請求 未請求 請求項の数 15 O L 外国語出願 (全30頁)

(21)出願番号 特願2022-135601(P2022-135601)	(71)出願人 390023711
(22)出願日 令和4年8月29日(2022.8.29)	ローベルト ボツシュ ゲゼルシャフト
(31)優先権主張番号 21193766	ミット ベシユレンクテル ハフツング
(32)優先日 令和3年8月30日(2021.8.30)	ROBERT BOSCH GMBH
(33)優先権主張国・地域又は機関 欧州特許庁(EP)	ドイツ連邦共和国 シュツツトガルト (番地なし)
(特許庁注 : 以下のものは登録商標)	Stuttgart , Germany
1 . B L U E T O O T H	(74)代理人 100114890
2 . Z I G B E E	弁理士 アインゼル・フェリックス＝ラ
	インハルト
	(74)代理人 100098501
	弁理士 森田 拓
	(74)代理人 100116403
	弁理士 前川 純一
	(74)代理人 100134315

最終頁に続く

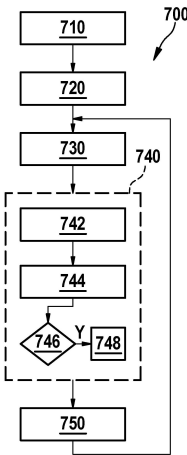
(54)【発明の名称】 画像分類器のオンザフライ校正

(57)【要約】 (修正有)

【課題】画像分類器によって出力されたクラス確率により形成される尤度推定値を改善する方法を提供する。

【解決手段】画像分類器を訓練するコンピュータ実装された方法700は、画像分類器の訓練可能パラメータと、複数の訓練画像及び対応するクラスラベルを含む訓練データセットとにアクセスすること710と、画像分類器を訓練する1つ以上の初期訓練ステップを実行して、訓練可能パラメータの現在のパラメータ値を取得すること720と、訓練データセットから訓練画像及び対応するクラスラベルを選択すること730と、訓練画像につき画像分類器を訓練するための現在のラベルベクトルを決定すること740と、現在のラベルベクトルを用いて、訓練画像につき画像分類器を訓練する現在の訓練ステップを実行すること750と、を含む。

【選択図】図7



【特許請求の範囲】**【請求項 1】**

画像分類器を訓練するコンピュータ実装された方法（700）であって、前記画像分類器は、入力画像に対する予測ベクトルを決定するように構成されており、前記予測ベクトルは、複数のそれぞれのクラスに対するクラス確率を含み、前記方法は、

- 前記画像分類器の訓練可能パラメータと、複数の訓練画像及び対応するクラスラベルを含む訓練データセットとにアクセスすること（710）と、

- 前記画像分類器を訓練する1つ以上の初期訓練ステップを実行して、前記訓練可能パラメータの現在のパラメータ値を取得すること（720）と、

- 前記訓練データセットから訓練画像及び対応するクラスラベルを選択すること（730）と、

- 前記現在のパラメータ値に従って前記画像分類器を前記訓練画像に適用して予測ベクトルを取得すること（742）と、前記予測ベクトルから前記複数のクラスのうちの予測クラスを推論すること（744）と、前記予測クラスと前記クラスラベルとを比較すること（746）と、前記予測クラスが前記クラスラベルに等しい場合、前記予測ベクトルに基づいて現在のラベルベクトルをソフトラベルとして決定すること（748）とによって、前記訓練画像につき前記画像分類器を訓練するための現在のラベルベクトルを決定すること（740）と、

- 前記現在のラベルベクトルを用いて、前記訓練画像につき前記画像分類器を訓練する現在の訓練ステップを実行すること（750）と、

を含む、方法（700）。

20

【請求項 2】

前記方法は、

- 以前の学習率に従って、及び、以前のラベルベクトルに従って、前記画像分類器の1つ以上の以前の訓練ステップを実行することと、

- 前記以前の学習率を減衰させることと、

- 減衰させた学習率及び前記以前のラベルベクトルに従って前記画像分類器を訓練する1つ以上のさらなる訓練ステップを実行して、前記訓練可能パラメータの更新済みのパラメータ値を取得することと、

- 前記更新済みのパラメータ値に従って前記画像分類器を適用することによって、前記訓練画像の前記現在のラベルベクトルを決定することと、

- 前記現在のラベルベクトル及び前記減衰させた学習率に従って前記現在の訓練ステップを実行することと、

を含む、

請求項 1 に記載の方法（700）。

30

【請求項 3】

前記1つ以上のさらなる訓練ステップは、1つの訓練エポックを形成する、請求項 2 に記載の方法（700）。

【請求項 4】

前記方法は、前記画像分類器からは独立して決定されたラベルベクトルを用いて前記1つ以上の初期訓練ステップを実行することを含む、

請求項 1 乃至 3 のいずれか一項に記載の方法（700）。

40

【請求項 5】

前記画像分類器は、前記入力画像のそれぞれの部分について前記複数のクラスのクラス確率を計算するように構成されたセマンティックセグメンテーションモデルであり、前記方法は、前記入力画像の前記それぞれの部分について現在のラベルベクトルを決定して使用することを含む、

請求項 1 乃至 4 のいずれか一項に記載の方法（700）。

【請求項 6】

前記入力画像は、ピクセル値、ボクセル値又はポイントクラウドによって表現される、

50

請求項 1 乃至 5 のいずれか一項に記載の方法（ 7 0 0 ）。

【請求項 7】

前記方法は、前記予測クラスが前記クラスラベルに等しくない場合、前記予測ベクトルからは独立して前記クラスラベルに基づいて前記現在のラベルベクトルを決定することをさらに含む、

請求項 1 乃至 6 のいずれか一項に記載の方法（ 7 0 0 ）。

【請求項 8】

前記予測ベクトルに基づいて前記現在のラベルベクトルを決定することは、前記予測ベクトルに事後校正モデルを適用することを含む、

請求項 1 乃至 7 のいずれか一項に記載の方法（ 7 0 0 ）。

10

【請求項 9】

前記方法は、前記画像分類器のために前記事後校正モデルを訓練することをさらに含む、

請求項 8 に記載の方法（ 7 0 0 ）。

【請求項 10】

訓練済みの画像分類器を使用するコンピュータ実装された方法（ 8 0 0 ）であって、前記方法は、

- 請求項 1 乃至 9 のいずれか一項に記載の方法に従って訓練された画像分類器の訓練済みのパラメータにアクセスすること（ 8 1 0 ）と、

- 入力画像を取得すること（ 8 2 0 ）と、

20

- 前記画像分類器を前記入力画像に適用して（ 8 3 0 ）、前記入力画像に対する予測ベクトルを決定し、前記予測ベクトルに基づいて予測を出力すること（ 8 3 5 ）、及び / 又は、前記画像分類器の一部を前記入力画像に適用して（ 8 4 0 ）、前記入力画像の特徴空間表現を取得し、前記特徴空間表現を出力すること（ 8 4 5 ）と、

を含む、方法（ 8 0 0 ）。

【請求項 11】

前記方法は、前記予測ベクトルから前記入力画像の分布外スコアを決定することを含む、

請求項 10 に記載の方法（ 8 0 0 ）。

【請求項 12】

30

前記入力画像は、コンピュータ制御システム及び / 又は前記コンピュータ制御システムの環境を表現しており、前記方法は、前記画像分類器の前記一部を適用して前記特徴空間表現を取得することと、前記特徴空間表現にさらなる訓練済みのモデルを適用して、前記コンピュータ制御システムを制御するための制御データを推論することと、を含む、

請求項 10 に記載の方法（ 8 0 0 ）。

【請求項 13】

画像分類器を訓練するためのシステム（ 1 0 0 ）であって、前記画像分類器は、入力画像に対する予測ベクトルを決定するように構成されており、前記予測ベクトルは、複数のそれぞれのクラスに対するクラス確率を含み、前記システムは、

- 複数の訓練画像及び対応するクラスラベルを含む訓練データセット（ 0 3 0 ）にアクセスし、前記画像分類器の訓練可能パラメータ（ 0 4 0 ）にアクセスするためのデータインタフェース（ 1 2 0 ）と、

40

- プロセッササブシステム（ 1 4 0 ）であって、

- 前記画像分類器を訓練する 1 つ以上の初期訓練ステップを実行して、前記訓練可能パラメータの現在のパラメータ値を取得することと、

- 前記訓練データセットから訓練画像及び対応するクラスラベルを選択することと、

- 前記現在のパラメータ値に従って前記画像分類器を前記訓練画像に適用して予測ベクトルを取得することと、前記予測ベクトルから前記複数のクラスのうちの予測クラスを推論することと、前記予測クラスと前記クラスラベルとを比較することと、前記予測クラスが前記クラスラベルに等しい場合、前記予測ベクトルに基づいて現在のラベルベクトル

50

をソフトラベルとして決定することとによって、前記訓練画像につき前記画像分類器を訓練するための現在のラベルベクトルを決定することと、

- 前記現在のラベルベクトルを用いて、前記訓練画像につき前記画像分類器を訓練する現在の訓練ステップを実行することと、

によって、前記画像分類器を訓練するように構成されたプロセッササブシステム（１４０）と、

を含む、システム（１００）。

【請求項１４】

画像分類器を使用するためのシステム（２００）であって、前記システムは、

- 請求項１乃至９のいずれか一項に記載の方法に従って訓練された画像分類器の訓練済みのパラメータ（０４２）を含む記憶装置（０２２）と、

- プロセッササブシステム（２４０）であって、

- 入力画像を取得することと、

- 前記画像分類器を前記入力画像に適用して前記入力画像に対する予測ベクトルを決定し、前記予測ベクトルに基づいて予測を出力すること、及び／又は、前記画像分類器の一部を前記入力画像に適用して前記入力画像の特徴空間表現を取得し、前記特徴空間表現を出力することと、

を行うように構成されているプロセッササブシステム（２４０）と、

を含む、システム（２００）。

【請求項１５】

一時的又は非一時的コンピュータ可読媒体（９００）であって、

- プロセッサシステムによって実行されるときに、請求項１乃至９のいずれか一項に記載のコンピュータ実装された方法を前記プロセッサシステムに実施させるための命令、及び／又は、

- プロセッサシステムによって実行されるときに、請求項１０乃至１２のいずれか一項に記載のコンピュータ実装された方法を前記プロセッサシステムに実施させるための命令、及び／又は、

- 請求項１乃至９のいずれか一項に記載のコンピュータ実装された方法に従って訓練された画像分類器の訓練済みのパラメータ

を表現するデータ（９１０）を含む、一時的又は非一時的コンピュータ可読媒体（９００）。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、画像分類器を訓練するコンピュータ実装された方法、及び、対応するシステムに関する。本発明はさらに、訓練済みの画像分類器を使用するコンピュータ実装された方法、及び、対応するシステムに関する。本発明はさらに、コンピュータ可読媒体に関する。

【背景技術】

【０００２】

発明の背景

深層ニューラルネットワークなどの機械学習可能な画像分類技術は、現在、コンピュータビジョンタスクにおける従来の方法を凌駕することができ、運転者支援システム及び医療撮像などの生活に不可欠なアプリケーションにおいて益々展開されている。この種のアプリケーションにおいては、画像分類器が正確な分類を出力することのみが重要となるわけではない（例えば、マルチクラス分類において画像の正確なクラスを予測すること、オブジェクト検出において画像内の正確なオブジェクトを検出すること、又は、画像セグメンテーションにおいて入力画像のそれぞれの部分の正確なクラスを予測することも重要となる）。また、画像分類器が尤度を正確に推定することができること、例えば、そのクラスが正しいことの尤度を正確に表現するクラス確率を出力することができることも重要で

10

20

30

40

50

ある。例えば、運転者支援システムは、これらの確率を使用して、環境の他の利用可能な測定値に関して分類の重要性を重み付けし、又は、画像分類器が不確実である場合にセーフモードに切り替えることができる。

【 0 0 0 3 】

多くの画像分類器は、未知のデータに対しては、正確な尤度を推定することができないことが知られている。この現象は、誤校正と称されており、画像分類器の過大信頼 (over-confidence) につながることが多い。これは、高保証の設定にとって問題となる。例えば、自動運転車両が非常に高い信頼度で停止標識を速度制限なし標識であるとして認識した場合には、事故が起こり、人命が危険にさらされる可能性がある。予測の精度に見合った適当な信頼度を出力する分類器が望まれている。グラウンドトゥールースからの信頼度の偏差は、種々の指標、例えば、予測校正誤差及び負の対数尤度に従って測定することができる。

10

【 先行技術文献 】

【 非特許文献 】

【 0 0 0 4 】

【 非特許文献 1 】 “ Rethinking the Inception Architecture for Computer Vision ”、Szegedy 等著、Proceedings CVPR 2016 (<https://arxiv.org/abs/1512.00567> で入手可能であり、参照により本明細書に組み込まれる)

【 発明の概要 】

【 発明が解決しようとする課題 】

20

【 0 0 0 5 】

一般に、画像分類器の校正を改善するための 2 つのタイプの技術として、事後 (post-hoc) 校正及びオンザフライ (on-the-fly) 校正が知られている。事後校正においては、画像分類器自体は不変のままであるが、訓練済みの校正モデルがその出力に適用される。オンザフライ校正は、画像分類器が訓練される方法を変更するものである。既知のオンザフライ校正方法は、ラベル平滑化 (Label Smoothing) と称されており、“ Rethinking the Inception Architecture for Computer Vision ”、Szegedy 等著、Proceedings CVPR 2016 (<https://arxiv.org/abs/1512.00567> で入手可能であり、参照により本明細書に組み込まれる) に開示されている。クラスラベルの One-hot 符号化を表現するラベルベクトルを用いて画像分類器を訓練する代わりに、ラベル平滑化においては、ラベルベクトルの全ての非ターゲットクラスに、小さいながら、非ゼロの確率質量が割り当てられる。その結果、グラウンドトゥールース分布のスパイクが少なくなるため、評価サンプルに対してより低い信頼度を出力する画像分類器をもたらす。これにより、過大信頼の問題は低減するが、画像分類器の尤度推定値をさらに改善する必要がある。

30

【 課題を解決するための手段 】

【 0 0 0 6 】

発明の概要

本発明の第 1 の態様によれば、請求項 1 及び 13 によって特定されるように、画像分類器を訓練するコンピュータ実装された方法及び対応するシステムがそれぞれ提供される。本発明の他の態様によれば、請求項 10 及び 14 によって特定されるように、訓練済みの画像分類器を使用するコンピュータ実装された方法及び対応するシステムがそれぞれ提供される。本発明の一態様によれば、コンピュータ可読媒体は、請求項 15 によって特定されるように説明される。

40

【 0 0 0 7 】

種々の態様は、画像分類のためのオンザフライ校正技術、及び、かかるオンザフライ校正技術に従って訓練された画像分類器に関する。画像分類器は、入力画像についての予測ベクトルを決定するように構成することができる。予測ベクトルは、複数のそれぞれのクラスについて、例えば、画像分類器がマルチクラス分類器である場合には、画像全体について、又は、画像分類器がマルチクラスセマンティックセグメンテーションモデルである

50

場合には、入力画像のそれぞれの部分について、クラス確率を含み得る。提供する技術の目的は、画像分類器によって出力されたクラス確率により形成される尤度推定値を改善することである。

【 0 0 0 8 】

画像分類器は、それぞれの訓練ステップを実行することによって訓練することができる。かかる訓練ステップにおいては、その訓練可能パラメータの現在値に従って画像分類器を訓練画像に適用し、現在の予測ベクトルを得ることができる。訓練信号は、現在の予測ベクトルとラベルベクトルとを比較することによって導出され得る。従来技術においては、このラベルベクトルは、通常、既知のクラスラベルの One - h o t 符号化である。ラベル平滑化においては、この One - h o t 符号化が平滑化される。この訓練信号に基づいて、画像分類器の訓練可能パラメータは、例えば、勾配降下によって更新することができる。典型的には、画像分類器の訓練は、例えば、複数のかかる訓練ステップを含み、例えば、複数のエポックにわたって実行される。

10

【 0 0 0 9 】

本発明者らは、既知のラベル平滑化が機能する様態を研究し、その有効性を弱める複数の要因を見出した。特に、ラベル平滑化においては、訓練全体にわたって一定の平滑化係数が使用されており、これにより、誤較正を軽減する際の有効性が低下している。本発明者らは、ラベル平滑化は訓練の初期フェーズにおける過大信頼を低減することができるが、後期の訓練フェーズにおいては、以前の平滑化係数が大き過ぎて、モデルの現在状態に適合しないことを見出した。換言すれば、本発明者らは、ラベル平滑化が、訓練の後期フェーズにおいて一部の予測を事実上オーバーキルし、過小信頼 (underconfident) とすることを認識した。この新たな誤較正の導入に起因して、ラベル平滑化により、訓練の終了時にベースラインに対する優位性が縮小する可能性がある。

20

【 0 0 1 0 】

既知のラベル平滑化の他の態様は、非ターゲットクラスに対して一様分布を用いることである。非ターゲットクラスの相対誤差をより正確にすることによって、より効果的な訓練信号が得られるため、これは望ましくない。例えば、任意の「乗用車」の画像が「トラック」に誤分類される確率は、「エンジン」に誤分類される確率よりも大きいと予測され得るが、これは、「乗用車」が「エンジン」よりも多くの特徴を「トラック」と共有しているためである。本発明者らが認識したように、ラベル平滑化は、この相対誤差情報を効果的に除去してしまうため、最適を下回る訓練性能しか得られない。

30

【 0 0 1 1 】

既知のラベル平滑化のさらに他の態様は、クラス内の全てのサンプルに対して同一のラベルを用いることである。例えば、C I F A R - 1 0 0 データセットのクラス「赤ちゃん (baby)」においては、いくつかのサンプルが「女の子の赤ちゃん (baby girl)」であり、他のサンプルが「男の子の赤ちゃん (baby boy)」である。これらのクラスに共通する特徴を考慮すると、両者ともターゲットクラス「赤ちゃん」に加えて、前者のサンプルグループについて非ターゲットクラス「女の子」のより大きい確率質量が予想され、後者について非ターゲットクラス「男の子」のより大きい確率質量が予想され得ることは、妥当である。本発明者らは、ラベル平滑化はこの種の適合性を考慮に入れていないので、誤較正を低減する効果が低いことを認識した。

40

【 0 0 1 2 】

興味深いことに、本発明者らは、従来技術の画像分類器によって提供される尤度推定値を改善する効果的なオンザフライ較正技術を見出した。この技術によれば、画像分類器は、画像分類器が現在正しく分類している画像を訓練するために、画像分類器自体によって出力された予測ベクトルに基づいてソフトラベルとして決定されるラベルベクトルを用いて訓練される。ソフトラベルとは、One - h o t 符号化としても知られるハードラベルとは対照的に、複数の非ゼロ確率を含むラベルベクトルである。

【 0 0 1 3 】

概して、画像分類器の訓練中のある時点において、1つ以上の初期訓練ステップが実行

50

されており、その結果、画像分類器の訓練可能パラメータの現在のパラメータ値が得られる。画像分類器をこれらの現在のパラメータ値に従って使用して、較正されたラベルベクトルを決定することができる。即ち、現在のパラメータ値に従って画像分類器を訓練画像に適用して、訓練画像に対するクラス確率の予測ベクトルを得ることができる。予測クラスは、予測ベクトルから推論され、訓練データセットのグラウンドトゥールスクラスラベルと比較することができる。予測クラスがクラスラベルに等しい場合、ソフトラベルが予測ベクトルに基づいて決定され、後続の訓練ステップのために使用され得る。

【0014】

従って、提供される技術によれば、画像分類器の訓練ステップは、画像分類器の現在の状態に基づいて決定されたラベルベクトル、又は、少なくとも比較的最近のラベルベクトル、例えば、現在又は以前の訓練フェーズにおいて決定されたラベルベクトルを用いて実行され得る。特に、訓練中のある時点において、画像分類器は、同様のパラメータセットに従って使用されて、ラベルベクトルを決定することと、パラメータが例えば勾配降下によって更新される訓練ステップを実行することと、の両方を行うことができる。ラベルベクトルは、訓練中に複数回、例えば、固定数の訓練エポックが実行された後に、又は、訓練フェーズごとに1回更新され得る。これにより、画像クラスを訓練するために使用される現在のラベルベクトルを、画像分類器の時間的に較正された正確な予測に基づいて決定することができる。

10

【0015】

提供されるオンザフライ較正技術は、予測からのラベル平滑化(LSP: Label Smoothing from Predictions)と称され、複数の利点を有する。予測をソフトラベルとして用いることによって、訓練により、画像分類器が意味的に曖昧で不適当にラベル付けされたサンプルに対して高い信頼度を生成することを積極的に停止する。一方、予測が用いられるため、不必要な過小信頼は導入されない。結果として、提供される技術は、種々のデータセット及び画像分類器アーキテクチャにわたって最先端の技術を凌駕することがわかった。LSPはまた、異なる事後較正方法並びに広範囲の訓練技術及びスケジュールとの良好な互換性を示す。例えば、適当な訓練スケジュールが知られている画像分類器が与えられると、画像分類器は、ハイパーパラメータカスタマイゼーションをほとんど又は全く伴わずに、同様の訓練スケジュールを使用してLSPを用いて再訓練することができる。事後較正と比較して、提供される技術は、事後較正を訓練するための追加のホールドアウトセットの必要がないので、訓練データが少なくても済み、較正コンポーネントの追加によるオーバーヘッドを回避しつつ、有効性を改善することができる。ベイズモデルなどの完全確率モデルと比較して、本明細書において説明されるように訓練された画像分類器は、訓練及び使用がより効率的であるため、より大きくより複雑なデータに適用可能であり、典型的にはより高い精度も得られる。

20

30

【0016】

特に、本発明者らは、誤較正を引き起こし得る複数のタイプの訓練画像を特定した。これらのタイプのうちの1つ以上の訓練画像を含む訓練データセットに対して、提供される技術は特に効果的である。概して、これらのタイプの訓練画像は、意味的に曖昧であり、及び/又は、不適当にラベル付けされる。特に、以下の6つのタイプ、即ち、(a)複数のオブジェクト(例えば、列車及び森林が同一の画像内に共存するもの)、(b)2つのクラスに属する単一のオブジェクト(例えば、1つのオブジェクトが同時にカエデの木及び森林を含むもの)、(c)認識不可能な単一のオブジェクト(例えば、低分解能、劣悪な照明又は同様の特徵に起因して、カンガルーであるかリスであるかを区別することが困難であるもの)、(d)そのクラスにおいて非定型である単一のオブジェクト(例えば、訓練セット内のほとんどのナシは現実のオブジェクトのように見えるが、当該ナシは漫画のように見えるもの)、(e)隠された単一のオブジェクト(例えば、クロコダイルの頭が隠れているもの)、(f)不適当にラベル付けされた単一のオブジェクト(例えば、平野ではなく道路である方がはるかに適当なラベル)を特定した。従来技術の技術によれば、ハードラベルは、画像分類器がこれらのサンプルに対して100%の信頼度で予測を出

40

50

力することを促し、それらの固有の不確実性を無視する。結果として、訓練済みの画像分類器は、これらのサンプルに過剰適合し、誤較正される場合がある。対照的に、L S P は、較正されたソフトラベルを用いて画像分類器を訓練するため、修正されたグラウンドトゥールスにおける自然な不確実性を維持し、画像分類器の過大信頼を抑制する。

【 0 0 1 7 】

重要なことに、訓練に用いられるラベルベクトルは、画像分類器自体の予測に基づき得るものであり、画像分類器の訓練が進むにつれて更新することができる。これは、他の事前訓練済みの分類器による予測を用いた知識蒸留技術とは対照的である。かかる技術は、例えば、より高度なモデルを使用してより単純なモデルを訓練するためには有用であり得るが、提供される技術で行われるような過大信頼の低減に役立つものではない。

10

【 0 0 1 8 】

本明細書において説明されるように訓練された画像分類器は、種々の方法により使用され得る。入力画像が与えられると、画像分類器は、入力画像についての予測ベクトルを決定するために適用することができ、予測は、予測ベクトルに基づいて出力され得る。予測は、例えば、クラスへの分類、特定のクラスについてのクラス確率、又は、クラス確率自体の予測ベクトルであり得る。提供される技術を使用することによって、例えば、より正確な予測クラス確率を出力することができるため、例えば制御システムにおいてより安全に使用することができる。また、画像分類器の分類精度が向上する可能性があること、特に、画像分類器が未知の例に対してより良く汎化することも見出された。従って、特定のクラスへの分類を単に出力する場合にも、説明されるように訓練された分類器は有利である。また、他の箇所でも説明されるように、分布外サンプルを認識する能力の改善も見出された。最後に、特徴抽出器としての画像分類器の性能も改善される。この目的のために、完全な画像分類器を入力画像に適用する代わりに、初期部分のみを適用することができる。例えば、畳み込み部分とこれに続くモデルの全結合層のサブセットとを適用することにより、入力画像の特徴空間表現を取得することができる。また、この特徴空間表現は、提供される技術の使用により、未知の例に対してより良く汎化されることがわかった。

20

【 0 0 1 9 】

任意選択的に、訓練画像の決定された現在のラベルベクトルは、訓練画像につき画像分類器を訓練する複数の訓練ステップにおいて用いることができる。例えば、ラベルベクトルは、訓練データセットの訓練画像に対して決定されるものとしてよく、その後、訓練の複数のエポック、例えば少なくとも 10 又は少なくとも 20 のエポックが、このラベルベクトルを用いて実行され得る。これにより、画像に対して実行される各訓練ステップについて画像のラベルベクトルを更新することと比較して、訓練の効率及び安定性が増加する。これにより、本質的に、ラベルベクトルが移動ターゲットとなって訓練が収束不可能になることが回避される。

30

【 0 0 2 0 】

任意選択的に、画像分類器を訓練する 1 つ以上の初期訓練ステップは、画像分類器からは独立して、例えばクラスラベルの One - h o t 符号化として決定されるラベルベクトルを用いて実行され得る。従って、訓練は、クラスラベルから導出されたラベルベクトルが用いられる第 1 の部分と、これに続く第 2 の部分とを含み得るものであり、当該第 2 の部分においては、正確な分類のために、予測ベクトルに基づくソフトラベルが用いられて、この第 2 の部分の間に任意選択的に更新される。第 1 の部分においては、訓練の第 2 の部分が特に効果的であり得るように、十分な精度及び偏りの除去が達成され得る。かかる初期訓練ステップを実行しない場合、訓練済みの画像分類器の性能が劣化する可能性がある。第 1 の部分が、例えば、学習率の第 1 の減衰の後、所定数（例えば、1 つ又は 2 つ）のエポックまで続くことが良好な選択である。

40

【 0 0 2 1 】

任意選択的に、訓練は、複数の訓練フェーズにおいて実行されるものとしてよく、かかる訓練フェーズは、学習率減衰から開始する。かかる訓練フェーズ中にラベルベクトルの更新は、複数の訓練ステップが実行された後にのみ、例えば、1 つの、2 つの又は他の固

50

定数のエポックが訓練フェーズに入った後にのみ行うことが有利であることがわかった。これにより、更新済みのラベルベクトルが学習率減衰に関連付けられた信頼度及び精度のブーストを使用することができるという利点が得られるが、これは、このブーストが典型的には複数の訓練ステップを経て行われるためである。次いで、残余の訓練フェーズにおいては、例えば、更新済みのラベルベクトル及び減衰させた学習率を使用することができる。

【 0 0 2 2 】

特に、画像分類器の1つ以上の以前の訓練ステップは、以前の学習率に従って、及び、以前のラベルベクトルに従って実行されるものとしてよく、その後、以前の学習率は減衰され得る。学習率減衰の後、減衰させた学習率及び以前のラベルベクトルを用いた画像分類器を訓練する1つ以上のさらなる訓練ステップを実行して、訓練可能パラメータの更新済みのパラメータ値を得ることができる。次いで、更新済みのパラメータ値に従って画像分類器を適用することによって、訓練画像のラベルベクトルを更新することができ、次いで、更新済みのラベルベクトルを用いて、フェーズのさらなる訓練ステップを実行することができる。例えば、ラベルベクトルは、各学習率減衰の後、1つおきの学習率減衰の後などに更新されるものとしてもよい。

10

【 0 0 2 3 】

任意選択的に、学習率の減衰とラベルベクトルの更新との間に実行される訓練ステップは、1つの訓練エポックを形成する。これは、前述のブーストが十分に効き、かつ、残余のフェーズにおいて更新済みのラベルを用いることができるため、ほとんどのアーキテクチャ及びデータセットにとって良好な選択であることがわかった。

20

【 0 0 2 4 】

任意選択的に、画像分類器によって訓練画像について予測されたクラスが、その訓練画像についてのクラスラベルに等しくない場合、現在のラベルベクトルは、予測ベクトルからは独立して、クラスラベルに基づいて決定され得る。例えば、ラベルベクトルは、クラスラベルの One - h o t 符号化に設定されるものとしてよい。ソフトラベルを用いることも可能であるが、例えば、従来技術のラベル平滑化に従って、ハード One - h o t 符号化を用いることが好ましいことがわかった。これは、モデルが画像を正しく分類することを促し、ソフトラベルを用いる必要がないように、誤った予測におけるモデルの信頼度を減少させる効果を既に有するものである。

30

【 0 0 2 5 】

任意選択的に、提供される技術は、事後校正モデルと組み合わせることができる。事後校正モデルを訓練中に適用して、予測ベクトルから現在のラベルベクトルを決定することができる。予測クラスは、元の予測ベクトル又は事後分類の結果のいずれかから推論することができる。事後分類は通常、予測クラスに影響を及ぼさないことから、この両方が可能となる。しかしながら、事後分類は、訓練ステップの一部として画像分類器を適用する場合には適用されない。使用時において、事後校正モデルは、画像分類器によって決定された予測ベクトルに適用され得る。興味深いことに、本明細書において説明するオンザフライ校正を事後校正方法と組み合わせることによって、特に、オンザフライ校正に基づく訓練中に既に事後校正を使用することによって、いずれかの技術を個別に適用するよりも全体的により良好に校正された尤度を取得することができることがわかった。例えば、温度スケールリング及び潜在ガウスプロセスなどのスケールリング方法、又は、ヒストグラムビンニング、特に I - M A X などのビンニング方法など、種々の事後校正方法がそれ自体知られており、本明細書において使用することができる。

40

【 0 0 2 6 】

任意選択的に、事後校正モデルは、例えば、事後校正モデルが画像分類器のパラメータ値の新しいセットに適用されるたびに、画像分類器のために訓練され得る。このように、訓練全体を通して、事後校正モデルは、画像分類器と共に進化することができる。

【 0 0 2 7 】

任意選択的に、使用時において、訓練済みの画像分類器を使用して、画像分類器によっ

50

て出力された予測ベクトルから入力画像の分布外スコアを決定することができる。例えば、分布外スコアは、予測ベクトルのエントロピーに基づいて決定されるものとしてもよい。興味深いことに、過大信頼が少ないという追加の利点として、提供される技術は、分布外サンプルを認識する画像分類器の能力を改善し得ることがわかった。多くの既知の画像分類器は、分布外サンプルに対して高い信頼度を出力する傾向がある。分布外スコアを決定することに関しても、提供される技術は、先行技術のオンザフライ較正方法よりも性能が優れていることが確認された。また、提供される技術に関して小さい分散が確認され、異なるデータ及びアーキテクチャにわたる堅牢性が示された。

【0028】

概して、提供される技術は、ビデオデータ、レーダデータ、LiDARデータ、超音波データ、運動データ、熱画像、又は、ソナーデータを含む、種々のタイプの画像データに適用される。画像データは、例えば、ピクセル値（例えば2D画像データの場合）、ボクセル値（例えば3D画像データの場合）、又は、点群（例えば2D又は3D画像データの場合）によって表現することができる。画像データはまた、ECG画像として表現されるECGデータ、スペクトログラムによって表現される音声データなど、種々のタイプのセンサデータを表現することができる。

【0029】

特に、画像分類器に入力される画像は、コンピュータ制御システム及び/又はその環境を表現することができ、画像分類器を使用して、コンピュータ制御システムを制御するための制御データを推論することができる。例えば、システムは、自律型又は半自律型車両などの車両、ロボット、製造機械、建物などであるものとしてよい。原則として、提供される技術は、他のタイプのデータ、例えば、センサ測定値の時系列などのセンサデータにも同様に適用され得る。

【0030】

画像分類器には、手元にあるデータの種類及びコンテンツの種類に応じて、種々のモデルアーキテクチャ及びパラメータを使用することができる。例えば、画像分類器は、深層ニューラルネットワーク、例えば、畳み込みニューラルネットワーク又は完全畳み込みニューラルネットワークであるものとしてよい。画像分類器は、入力画像のそれぞれの部分についてクラス確率を計算するように構成されたセマンティックセグメンテーションモデルであるものとしてもよい。

【0031】

任意選択的に、訓練済みの画像分類器の一部を適用することによって取得される特徴空間表現が、さらなる訓練済みのモデルへの入力として使用され得る。さらなる訓練済みのモデルは、例えば、コンピュータ制御システムを制御するための制御データを推論するために使用されるものとしてもよい。従って、画像分類器は、下流の他のタスクの基礎として使用され得る。このようにして、画像分類器は、特徴抽出器として使用されて、その知識を転移学習によってこうした他のタスクに付与することもできる。不確実性較正が画像分類器の分類タスクに対して改善されるので、下流のタスクにおける対応するより良好な予測不確実性を取得することができる。

【0032】

本発明の上述した実施形態、実現形態、及び/又は、任意の態様のうちの2つ以上が、有用であると考えられる任意の方法により組み合わせられ得ることが、当業者によって理解されるであろう。

【0033】

対応するコンピュータ実装された方法の説明された修正及び変形に対応する、任意のシステム及び/又は任意のコンピュータ可読媒体の修正及び変形は、本説明に基づいて当業者によって実行され得る。

【0034】

本発明のこれらの態様及び他の態様は、以下の説明において例として説明される実施形態を参照することにより、また、添付の図面を参照することにより、さらに解明され、明

10

20

30

40

50

らかになるであろう。

【図面の簡単な説明】

【 0 0 3 5 】

【図 1】分類器を訓練するためのシステムを示す図である。

【図 2】分類器を使用するためのシステムを示す図である。

【図 3】分類器の使用の詳細な例を示す図である。

【図 4】分類器を訓練する方法の詳細な例を示す図である。

【図 5 a】分類器を訓練する方法の詳細な例を示す図である。

【図 5 b】ラベル校正の詳細な例を示す図である。

【図 6】画像分類器の詳細な例を示す図である。

10

【図 7】分類器を訓練するコンピュータ実装された方法を示す図である。

【図 8】分類器を使用するコンピュータ実装された方法を示す図である。

【図 9】データを含むコンピュータ可読媒体を示す図である。

【 0 0 3 6 】

図面は単に概略的なものであり、縮尺通りに描かれていないことに留意されたい。図において、既に説明した要素に対応する要素は、同一の参照符号を有する場合がある。

【発明を実施するための形態】

【 0 0 3 7 】

実施形態の詳細な説明

図 1 は、分類器を訓練するためのシステム 100 を示している。分類器は、入力インスタンスの予測ベクトルを決定するように構成され得る。例えば、分類器は、入力画像に作用する画像分類器であるものとしてもよい。予測ベクトルは、複数のそれぞれのクラスに対するクラス確率を含み得る。

20

【 0 0 3 8 】

システム 100 は、データインタフェース 120 を含み得る。データインタフェース 120 は、複数の訓練インスタンス及び対応するクラスラベルを含む訓練データセット 030 にアクセスするためのものであり得る。例えば、訓練インスタンスの数は、少なくとも 1000 個、少なくとも 10000 個、又は、少なくとも 100000 個であるものとしてもよい。これに代えて又はこれに加えて、データインタフェース 120 は、分類器の訓練可能パラメータ 040 にアクセスするためのものであり得る。パラメータ 040 は、例えば、図 2 のシステム 200 又は図 3 のシステム 300 によって、本明細書において説明される方法に従って分類器を使用するためのものであり得る。

30

【 0 0 3 9 】

例えば、図 1 にも示されているように、データインタフェース 120 は、データ記憶装置 021 からのデータ 030、040 にアクセスし得るデータ記憶装置インタフェース 120 によって構成されるものとしてもよい。例えば、データ記憶装置インタフェース 120 は、メモリインタフェース又は恒久的記憶装置インタフェース、例えば、ハードディスク又は SSD インタフェースであるものとしてもよいが、Bluetooth、Zigbee 若しくは Wi-Fi インタフェース、又は、イーサネット若しくは光ファイバインタフェースなどのパーソナル、ローカル又はワイドエリアネットワークインタフェースであるものとしてもよい。データ記憶装置 021 は、ハードドライブ又は SSD などのシステム 100 の内部データ記憶装置であるものとしてもよいが、外部データ記憶装置、例えばネットワークアクセス可能なデータ記憶装置であるものとしてもよい。いくつかの実施形態においては、データ 030、040 はそれぞれ、例えば、データ記憶装置インタフェース 120 の異なるサブシステムを介して、異なるデータ記憶装置からアクセスされるものとしてもよい。各サブシステムは、データ記憶装置インタフェース 120 について上述したようなタイプであるものとしてもよい。

40

【 0 0 4 0 】

システム 100 は、システム 100 の動作中に分類器を訓練するように構成され得るプロセッササブシステム 140 をさらに含み得る。プロセッササブシステム 140 による訓

50

練は、分類器を訓練する1つ以上の初期訓練ステップを実行して、訓練可能パラメータ040の現在のパラメータ値を取得することを含み得る。プロセッササブシステム140による訓練は、訓練データセット030から訓練インスタンス及び対応するクラスラベルを選択することを含み得る。プロセッササブシステム140による訓練は、訓練インスタンスに対して分類器を訓練するための現在のラベルベクトルを決定することを含み得る。現在のラベルを決定するために、プロセッササブシステム140は、現在のパラメータ値に従って分類器を訓練インスタンスに適用して予測ベクトルを取得することと、予測ベクトルから複数のクラスのうちの予測クラスを推論することと、予測クラスとクラスラベルとを比較し、予測クラスがクラスラベルに等しい場合、予測ベクトルに基づいて現在のラベルベクトルをソフトラベルとして決定することと、を行うことができる。プロセッササブシステム140による訓練は、現在のラベルベクトルを用いて訓練インスタンスに対して分類器を訓練する現在の訓練ステップを実行することを含み得る。

10

【0041】

システム100は、学習済みの（又は「訓練済みの」）モデルのパラメータを表現する訓練済みのデータ040を出力するための出力インタフェースをさらに含み得る。例えば、図1にも示されているように、出力インタフェースは、データインタフェース120によって構成されるものとしてよく、当該インタフェースは、これらの実施形態においては、入力/出力（「IO」）インタフェースであり、それを介して、訓練済みのパラメータ040がデータ記憶装置021に記憶されるものとしてもよい。例えば、「未訓練の」モデルの訓練可能パラメータは、訓練データ030に訓練を反映するために、訓練の間又は後に、ニューラルネットワークの重み及び他のタイプのパラメータなどの訓練済みのパラメータによって少なくとも部分的に置き換えられ得る。これはまた、未訓練のパラメータ及び訓練済みのパラメータの両方を表現するレコード040によって図1に示されている。他の実施形態においては、訓練済みのパラメータ040は、「未訓練の」パラメータとは別個に記憶されるものとして行うことができる。いくつかの実施形態においては、出力インタフェースは、データ記憶装置インタフェース120とは別個であるものとしてもよいが、概して、データ記憶装置インタフェース120について上述したようなタイプのものであるものとしてもよい。

20

【0042】

図2は、例えば、図1のデバイス100により、本明細書において説明されるような訓練済みの分類器を使用するためのシステム200を示している。システム200は、分類器の訓練済みのパラメータ040にアクセスするためのデータインタフェース220を含み得る。システムは、モデルを使用することに加えて、モデルを訓練することができ、例えば、図1のシステムと組み合わせることができる。

30

【0043】

例えば、図2にも示されているように、データインタフェースは、データ記憶装置022からのデータ040にアクセスし得るデータ記憶装置インタフェース220によって構成されるものとしてもよい。概して、データインタフェース220及びデータ記憶装置022は、データインタフェース120及びデータ記憶装置021について図1を参照して説明したものと同様のタイプとすることができる。記憶装置は、システム200の一部であり得る。この図には示されていないが、記憶装置022はまた、分類器が適用され得る入力インスタンスを含み得る。

40

【0044】

システム200は、システム200の動作中に入力インスタンス、例えば入力画像を取得するように構成され得るプロセッササブシステム240をさらに含み得る。プロセッササブシステム240は、分類器を入力インスタンスに適用して、入力インスタンスの予測ベクトルを決定し、予測ベクトルに基づいて予測を出力するようにさらに構成され得る。これに代えて又はこれに加えて、プロセッササブシステム240は、分類器の一部を入力インスタンスに適用して、入力インスタンスの特徴空間表現を取得し、特徴空間表現を出力するように構成され得る。

50

【 0 0 4 5 】

図 1 のプロセッササブシステム 1 4 0 と同様の考慮事項及び実装選択肢がプロセッササブシステム 2 4 0 に適用されることが理解されるであろう。特に断りのない限り、図 1 のシステム 1 0 0 の場合と同様の考慮事項及び実装選択肢がシステム 2 0 0 に概して適用され得ることがさらに理解されよう。

【 0 0 4 6 】

図 2 は、システム 2 0 0 の種々の任意選択的な構成要素をさらに示している。例えば、いくつかの実施形態においては、システム 2 0 0 は、環境 0 8 2 内のセンサ 0 7 2 によって取得されたセンサデータ 2 2 4 に直接アクセスするためのセンサインタフェース 2 6 0 を含み得る。センサは、環境 0 8 2 内に配置され得るが、例えば、量を遠隔で測定することができる場合は、環境 0 8 2 から遠隔に配置されるものとしてもよい。センサ 0 7 2 は、システム 2 0 0 の一部であるものとしてもよいが、必ずしもそうである必要はない。センサ 0 7 2 は、画像センサ、l i d a r センサ、レーダセンサ、圧力センサ、内容物温度センサなどの任意の適当な形態を有し得る。例えば、図に示されているように、センサ 0 7 2 は、システム 2 0 0 及び / 又はその環境を表現する入力画像 2 2 6 を取り込むカメラであるものとしてもよい。センサデータインタフェース 2 6 0 は、例えば、I 2 C 若しくは S P I データ通信に基づく低レベル通信インタフェース、又は、データインタフェース 2 2 0 について上述したようなタイプのデータ記憶装置インタフェースを含むがこれらに限定されるものではない、センサのタイプにタイプが対応する任意の適当な形態を有するものとしてもよい。

【 0 0 4 7 】

いくつかの実施形態においては、システム 2 0 0 は、環境 0 8 2 内のアクチュエータ（図示せず）に制御データ 2 2 6 を提供するためのアクチュエータインタフェース 2 8 0 を含み得る。かかる制御データ 2 2 6 は、プロセッササブシステム 2 4 0 によって生成されて、分類器を使用することに基づいて、例えば、分類器による予測に基づいて、又は、分類器の一部を適用することによって取得された特徴空間表現に基づいて、アクチュエータを制御し、さらなる訓練済みのモデルを特徴空間表現に適用して制御データ 2 2 6 を推論することができる。アクチュエータは、システム 2 0 0 の一部であり得る。例えば、アクチュエータは、電気式、油圧式、空気圧式、熱式、磁気式及び / 又は機械式アクチュエータであるものとしてもよい。具体的であるが非限定的な例としては、電気モータ、電気活性ポリマー、油圧シリンダ、圧電アクチュエータ、空気圧アクチュエータ、サーボ機構、ソレノイド、ステッパモータなどが挙げられる。かかるタイプの制御は、（半）自律車両について図 3 を参照して説明される。

【 0 0 4 8 】

他の実施形態（図 2 に図示せず）においては、システム 2 0 0 は、ディスプレイ、光源、ラウドスピーカ、振動モータなどのレンダリングデバイスへの出力インタフェースを含み得るものであり、これらを使用して、分類器の使用に基づいて生成され得る感覚的に知覚可能な出力信号が生成され得る。感覚的に知覚可能な出力信号は、分類器によって出力された分類、クラス確率、及び / 又は、特徴空間表現を直接示することができるが、例えば、システム 2 0 0 の誘導、ナビゲーション、又は、他のタイプの制御において使用するための、派生した感覚的に知覚可能な出力信号を表現することもできる。

【 0 0 4 9 】

概して、本明細書において説明される各システムは、図 1 のシステム 1 0 0 及び図 2 のシステム 2 0 0 を含むがこれらに限定されるものではなく、ワークステーション若しくはサーバなどの単一のデバイス又は装置として若しくはその中に具現化され得る。デバイスは、埋め込み型デバイスであり得る。デバイス又は装置は、適当なソフトウェアを実行する 1 つ以上のマイクロプロセッサを含み得る。例えば、それぞれのシステムのプロセッササブシステムは、単一の中央処理装置（C P U）によって具現化され得るが、かかる C P U 及び / 又は他のタイプの処理ユニットの組合せ又はシステムによって具現化されるものとしてもよい。ソフトウェアは、対応するメモリ、例えば、R A M などの揮発性メモリ又

10

20

30

40

50

はフラッシュなどの不揮発性メモリにダウンロード及び／又は記憶されるものとしてよい。代替的に、それぞれのシステムのプロセッササブシステムは、例えばフィールドプログラマブルゲートアレイ（FPGA）などのプログラマブルロジックの形態でデバイス又は装置に実装され得る。概して、それぞれのシステムの各機能ユニットは、回路の形態で実装され得る。それぞれのシステムはまた、例えば、分散ローカル又はクラウドに基づくサーバなどの異なるデバイス又は装置を伴う分散様式で実装されるものとしてもよい。いくつかの実施形態においては、システム200は、車両、ロボット又は同様の物理的エンティティの一部であるものとしてよく、及び／又は、物理的エンティティを制御するように構成された制御システムを表現するものとしてもよい。

【0050】

10

図3は、システム200が環境50内において動作する（半）自律車両62の制御システムであるように示されるという点において、上記の一例を示している。自律車両62は、自律運転システム又は運転支援システムを含み得るという点において、自律型であるものとしてよく、後者は、半自律システムとも称される。自律車両62は、例えば、車両62に組み込まれたビデオカメラ22から取得されたセンサデータに基づいて自律車両のステアリング及びブレーキを制御するためにシステム200を組み込んだものであってよい。例えば、システム200は、予期しない交通状況の場合、例えば、自律車両62が交通参加者と衝突することが予測される場合、（回生）制動を実行するように電気モータ42を制御し得る。システム200は、危険な状況に対してステアリング及び／又はブレーキを制御して、例えば交通参加者との衝突を回避することができる。システム200は、ビデオカメラによって提供される入力画像に画像分類器を適用することができる。システムは、例えば、画像分類器による分類が閾値を超えるクラス確率を有するか否か、又は、入力画像が閾値を超える分布外スコアを有するか否かを判定することができる。これに基づいて、システム200は、例えば、車両62を制御するために分類ベクトル若しくは予測ベクトルを用いること（例えば、クラス確率及び／又は分布外スコアが十分な信頼度を示す場合）、又は、代わりにフォールバック制御機構、例えば制動機構を使用すること（例えば、クラス確率及び／又は分布外スコアが不十分な信頼度を示す場合）を判定するものとしてもよい。

20

【0051】

次に、従来技術のオンザフライ較正方法のラベル平滑化（LS）の複数のボトルネックを評価し、説明する。ラベル平滑化においては、分類器自体による予測ベクトルではなく、グラウンドトゥールスクラスラベルのみを使用してラベルベクトルが決定される。WRN-28-10のデータセットCIFAR-100において、LSが、交差エントロピーOne-hotコーディングベースラインと比較される。平滑化係数として、0.1が使用される。複数のボトルネックが特定された。

30

【0052】

経時的な一定の平滑化係数：予測較正誤差（ECE）を使用して、モデルが十分に較正されている状態からどの程度逸脱しているかを経験的に調査することができる。しかしながら、ECEにおける絶対差計算においては、この情報は排除されるため、この場合、偏差の符号は未知となる。そこで、ECEの絶対差を総和に置き換えたFlatECEと称される代替指標を提案する。FlatECEは、

40

【数1】

$$\text{FlatECE} = \sum_{m=1}^M \frac{|B_m|}{N} (\text{Conf}_m - \text{Acc}_m)$$

として形式的に定義することができる。全てのピンが過大信頼である場合、例えば m , $\text{Conf}_m - \text{Acc}_m > 0$ である場合、FlatECEはECEと等価である。FlatECEは、いくつかのピンが過小信頼である場合にECEよりも小さくなり、全てのピンが過小信頼である場合に負のECEに等しくなる。本発明者らは、LS-0.1についてECE及びFlatECEを計算した。3つの学習率減衰においては、訓練は4つのフェ

50

ーズに分割される。フェーズ 1 及びフェーズ 2 において、 $LS - 0.1$ は、ベースラインよりもはるかに低い ECE を有し、これは、分類器の過大信頼を相殺する $LS - 0.1$ の有効性を示している。しかし、フェーズ 3 及びフェーズ 4 において、ベースラインに対する $LS - 0.1$ の ECE の利点は、著しく減少する。 $LS - 0.1$ において、 $Flat ECE$ 及び ECE の曲線が徐々に分割されることから、フェーズ 3 及び 4 における $LS - 0.1$ のほとんどの校正誤差が、実際には過小信頼の成長に由来することが示される。早期停止は汎化と校正との間のトレードオフであることから、訓練を停止して現在のチェックポイントを取得することは、実現可能な選択肢ではない。この時点において、平滑化係数 0.1 はオーバーキルであり、過小信頼を引き起こすと結論付けられる。これが調整された場合には、動的な減少が生じ、完全な訓練後のより良好な校正性能を期待することができる。

【0053】

種々のクラス及びインスタンスにわたる一定の平滑化係数： LS においては、全ての非ターゲットクラスについて均一に分布した確率が仮定される。この仮定に対しては、欠点を探ることができる。一部の非ターゲットクラスの確率が他の非ターゲットクラスより高いことは、直観的に理解することができるものである。例えば、 $CIFAR - 10$ において乗用車とラベル付けされた画像に対しては、特定の画像を見ることなく、より大きい非ターゲット確率質量が、猫又は犬ではなくトラックに割り当てられるはずである。なぜならば、トラックは、猫又は犬よりも乗用車に似た外観を有するからである。

【0054】

LS における他の仮定は、クラス内の全てのサンプルが同一の平滑化ラベルであるということである。これは、同一の特徴がクラス内の異なるサンプルにわたって共有されると仮定されることを意味する。これに対しても、反例を挙げることができる。例えば、赤ちゃんとして注釈が付けられた 2 つの画像を検討する。一方の写真では、赤ちゃんは、左側の赤ちゃんよりもわずかに「女性」らしく見え、より多くの「女性」の特徴を有しているかもしれない。この場合、非ターゲットクラス「女の子」には、他の写真よりも大きい確率的重みが与えられるはずである。

【0055】

図 4 は、分類器を訓練する方法の詳細であるが非限定的な例を示している。例示的な例として、画像分類器が使用される。

【0056】

画像分類器 $IC420$ は、入力画像に適用することができる。例えば、図は、訓練データセットの訓練画像 $TI(431)$ に適用されている画像分類器 IC を示している。概して、入力画像は、種々の異なる方法によって、例えば、ピクセル値（例えば、少なくとも 128 又は少なくとも 1024 ピクセル）、ボクセル値（例えば、少なくとも 128 又は少なくとも 1024 ボクセル）、又は、ポイントクラウド（例えば、少なくとも 1024 又は少なくとも 4096 ポイントのポイントクラウド）によって、表現することができる。画像は、例えば、白黒画像、グレースケール画像、又は、カラー画像であるものとしてよい。

【0057】

画像分類器 IC は、入力画像についての予測ベクトルを決定するように構成することができる。予測ベクトルは、複数のそれぞれのクラスに対するクラス確率を含み得る。例えば、図は、訓練画像 $TI(431)$ に適用されて、クラス確率 $CP1(421)$, ..., $CPn(422)$ を取得する画像分類器 IC を示している。例えば、クラスの数、少なくとも 3、少なくとも 10、又は、少なくとも 20 であるものとしてよい。訓練画像は、訓練データセット、例えば、訓練画像 TI が割り当てられる単一のクラスに従って、グラウンドトゥールスクラスラベル $CL(432)$ を有し得る。この図には示されていないが、画像分類器 IC は、入力画像のそれぞれの部分、例えばピクセルに対するクラス確率 $CP1$, ..., CPn のセットを出力するように構成されたセマンティックセグメンテーションモデルであるものとしてもよく、この場合、訓練データセットは、訓練画像 TI に対して

、訓練画像 T I のそれぞれの部分にクラスを割り当てるそれぞれのクラスラベル C L を含むものとしてもよい。

【 0 0 5 8 】

画像分類器 I C は、適用される入力画像に適した種々の既知のモデルアーキテクチャを有し得る。画像分類器は、訓練可能パラメータ P A R S (4 4 0) のセットによってパラメータ化され得る。訓練可能パラメータの数は、例えば、少なくとも 1 0 0 0 個、少なくとも 1 0 0 0 0 個、又は、少なくとも 1 0 0 0 0 0 個であるものとしてもよい。画像分類器 I C は、例えば、深層ニューラルネットワーク又は畳み込みニューラルネットワークなどのニューラルネットワークであるものとしてよい。この場合、パラメータのセットは、例えば、ニューラルネットワークのノードの重みを含み得る。かかる画像分類器の例は、
図 6 に関して説明される。

10

【 0 0 5 9 】

図に示されているように、画像分類器 I C は、訓練動作である「訓練」4 6 0 において訓練することができる。「訓練」は、本明細書に記載されるように決定されたラベルベクトルを用いることを除いて、従来技術から知られている訓練であり得る。例えば、訓練には、手元のデータに対して良好に機能することが知られているハイパーパラメータ及び/又は訓練スケジュールを使用することができる。概して、訓練には、画像分類器 I C がそれぞれの訓練画像 T I に適用されるそれぞれの訓練ステップが含まれ得る。かかる訓練ステップにおいては、その訓練可能パラメータ P A R S の現在の値に従って画像分類器 I C を訓練画像 T I に適用し、現在の予測ベクトル C P 1 , … , C P n を得ることができる。「訓練」は、現在の予測ベクトル C P i を訓練画像 T I のラベルベクトル L V 1 (4 5 1) , … , L V n (4 5 2) と比較することによって訓練信号を導出し、訓練信号に基づいてパラメータ P A R S を更新することができる。

20

【 0 0 6 0 】

一例においては、画像分類器 I C は、例えば K > 2 クラスのマルチクラス分類器であるものとしてよい。ここで、

【数 2】

$$D = \{(x_i, y_i)\}_{i=1}^N$$

30

を、N 個のサンプルから構成される訓練データセットとする。ここで、サンプル i について、訓練画像 T I , x i は、画像分類器 I C の入力であり、y i = [y i , 1 , … , y i , j , … , y i , K] [0 , 1] K であり、

【数 3】

$$\sum_{j=1}^K y_{i,j} = 1$$

は、ラベルベクトル L V i である。サンプル i の最も有望なクラスインデックスは、s i = a r g m a x 1 j K y i , j である。例えば、ラベルベクトルは、クラスラベルのハードラベル、換言すれば、O n e - h o t 符号化であるものとしてよく、

40

【数 4】

$$y_{i,j=s_i} = 1 \text{ かつ } y_{i,j \neq s_i} = 0$$

である。マルチクラス分類器 I C は、

【数 5】

$$\sum_{j=1}^K q_{i,j} = 1$$

となるように、x i を確率的予測ベクトル C P i q i = [q i , 1 , … , q i , j , … , q

50

$i, k] [0, 1]^K$ に整合させることができる。この場合に画像分類器 IC を訓練することは、交差エントロピー損失を最小化するために最適化を実行することを含むものとしてよく、例えば、

【数 6】

$$L_{CE} = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^K y_{i,j} \log q_{i,j}$$

であり、ここで、 n は最適化反復におけるサンプルの数である。

【0061】

例えば、「訓練」は、確率的勾配降下法などの確率的手法を使用して、例えば、「“Adam: A Method for Stochastic Optimization”、Kingma and Ba 著（参照により本明細書に組み込まれ、<https://arxiv.org/abs/1412.6980>で入手可能）」に開示されているアダムオプティマイザを使用して実行され得る。周知のように、かかる最適化方法は、ヒューリスティックであるものとしてよく、及び/又は、局所最適に到達するものとしてもよい。訓練は、インスタンスごとに、又は、バッチで、例えば、最大若しくは少なくとも 64 個、又は、最大若しくは少なくとも 256 個のインスタンスで、実行され得る。

【0062】

さらに、「訓練」によって使用されるラベルベクトル LVi を決定するラベル校正動作 LC (450) が示されている。興味深いことに、本明細書において説明する実施形態においては、訓練画像 TI のラベルベクトル LVi は、パラメータ PARS の現在のセットに従って画像分類器 IC を訓練画像 TI に適用することによって取得された予測ベクトル CPi に基づいて決定され得る。

【0063】

典型的には、画像分類器 IC を訓練する 1 つ以上の初期訓練ステップ中に、ラベルベクトル LVi は、画像分類器 IC によって決定される分類又はクラス確率からは独立して、動作 LC によって決定される。例えば、ラベルベクトル LVi は、クラスラベル LC の One-hot 符号化などに対して設定され得る。

【0064】

訓練中のある時点において、ラベル校正動作 LC は、訓練画像 TI のラベルベクトル LVi を更新することができる。これは、典型的には、訓練画像につき実行される訓練ステップごとに行われるのではなく、より少ない回数、例えば、学習率の減衰ごとに 1 回行われる。ラベルベクトル LVi が更新され得る場合の詳細な例は、図 5 に関して提供される。通常、ラベル校正動作 LC は、訓練データセット内の全ての画像のラベルベクトルを更新するが、これは必須ではない。

【0065】

訓練画像 TI のラベルベクトル LVi を更新するために、その現在のパラメータ値に従って画像分類器 IC を使用して推論を実行して、訓練画像 TI の予測ベクトル CPi を取得することができる。この予測ベクトルから、訓練画像 TI についての予測クラスを推論することができる。予測は、その正確度に従って分割することができる。誤分類された予測には、そのソフトラベルに対する手がかりはない。ラベルベクトル LVi は、この場合クラスラベル CL のためのハードラベルを表現する One-hot コーディングラベルに設定され得るが、そうでない場合には、例えばクラスラベルに対応する平滑化された One-hot ラベルとして、予測ベクトルから独立して決定されるものとしてもよい。

【0066】

一方、正しく分類されたサンプルについては、予測ベクトル CPi に基づいてラベルベクトル LVi を設定することができる。特に、1 つの可能性として、ラベルベクトル LVi を予測ベクトル CPi に等しく設定することが考えられる。

【0067】

また、図に示されているように、校正性能をさらに改善するための方策として、事後較

正モデル $PHC(470)$ を予測ベクトル CP_i に適用し、校正フレームワークにおけるソフトラベル LV_i として、直接予測 CP_i ではなく、事後校正予測 $CC1(471)$, ... , $CCn(472)$ を使用することが挙げられる。

【0068】

原則として、任意の事後校正器 PHC (post-hoc calibrator) を使用することができる。事後校正器は、例えば、スケーリング事後校正器であり得る。かかる方法により、連続校正関数を使用して、画像分類器の未加工の出力、例えばロジットを拡大又は縮小することができる。例えば、スケーリング事後校正器は、温度スケーリング (TS) 又は潜在ガウス過程 (GP) であり得る。温度スケーリングにおいては、ロジットは、温度係数 T によって除算することができ、当該温度係数 T は、検証セットにつき NLL を最適化することによって取得することができる。潜在ガウス過程 (GP) は、単一の温度因子ではなく、連続曲線を学習することができる。負の対数尤度 (NLL) 指標の場合、 GP は、ほとんどの状況において TS よりも性能が優れていることが実験的に見出された。予測校正誤差 (ECE) 指標については、限られた校正サンプルが利用可能である場合、 TS が GP よりも良好に機能することがわかった。その他の場合においては、構成可能なパラメータの数が多い GP がより良好に機能することがわかった。

10

【0069】

事後校正器 PHC は、ビンニング事後校正器でもあり得る。ビンニング事後校正器は、離散ビンニング関数を介してロジットを調整することができる。例えば、ヒストグラムビンニング、特に、 $I-MAX$ ヒストグラムビンニング方法などのビンエッジを最適化する方法が使用されるものとしてもよい。 $I-MAX$ はさらに、校正セットからのクラスごとの予測をマージして、サンプリングの効率を高めることができる。概して、スケーリング方法は、予測の精度を維持することができるが、 ECE 評価のビン数が変化した場合、 ECE に対して一貫した性能を有することができない。ビンニング方法は、定量化による精度の損失にもかかわらず、 ECE に対してより一貫した性能を有することができる。

20

【0070】

事後校正モデル PHC は、好ましくは、その現在のパラメータ $PARS$ に従って、例えば、更新済みのパラメータ $PARS$ に従って適用されるたびに、画像分類器 IC のために訓練される。事後校正モデル PHC の訓練は、典型的には、画像分類器 IC を訓練するために使用される訓練データセットとは別個のホールドアウトセット、例えば検証セットを使用するものである。

30

【0071】

従って、ラベル校正 LC によって決定されるラベルベクトル LV_i は、a) 校正された正確な予測であり、かつ、b) 誤ったサンプルに対するクラスに基づく、例えば $One-hot$ コーディングのラベルであり得る。決定されたラベル LV_i を使用して、後続の訓練ステップ、典型的には訓練画像 TI ごとの複数の訓練ステップのために画像分類器を訓練することができる。この後、ここで後続の訓練ステップの結果として取得される訓練済みのパラメータ $PARS$ を用いて、更新済みのラベルベクトル LV_i を決定するプロセスが再び実行されるものとしてもよい。従って、訓練ステップである「訓練」及びラベル校正 LC は、訓練スケジュールの終了まで交互に行われ、その結果、ネイティブに良好に校正された画像分類器 IC が得られ、追加の校正層を必要とせずにこれをいくつかの実施形態において使用することができる。

40

【0072】

図5aは、分類器、例えば画像分類器を訓練する方法の詳細であるが非限定的な例を示している。この例は、図4の例に基づいており、訓練中のどの時点においてラベル校正 $LC(550)$ が適用され得るかを示している。図4に関しても説明したように、ラベル校正 LC は、典型的には訓練データセットの全ての訓練画像について、ラベルベクトルを決定することができる。最初に、訓練中に使用されるラベルベクトルは、例えば、クラスラベルの $One-hot$ 符号化として、訓練中の画像分類器から独立して決定することができる。訓練中のある時点において、ラベル校正 LC は、画像分類器によって決定された予

50

測ベクトルに基づいてソフトラベルとして正しく分類された画像のラベルベクトルを設定することによって、その訓練済みのパラメータの現在の値に従って画像分類器を使用することができる。誤って分類された画像については、ラベルベクトルは、予測ベクトルからは独立して、例えばハードラベルとして決定され得る。

【0073】

この例においては、画像分類器は、マルチステップスケジューラを用いた確率的勾配降下 (SGD) によって訓練される。かかる訓練は、例えば深層学習においてそれ自体周知である。マルチステップスケジューリングとは、訓練中に複数回減衰する学習率を使用することを指す。学習率減衰と学習率減衰との間の期間は、訓練フェーズと称され得る。訓練の開始561時には、学習率は比較的大きくなり得る。典型的には、精度及び信頼度は最初に増加し、次いで、あるレベルで停滞する。複数の訓練ステップ、例えば固定数のエポックの後、学習率を低下させる学習率減衰564を実行し得る。これは、典型的には、数回の反復後に、精度及び信頼度の上昇をもたらし、次いで、新たなより高いレベルで停滞することができ、以降同様となる。

10

【0074】

図に示すように、訓練は、第1の学習率減衰と第2の学習率減衰との間の点において、ラベル校正LCによって校正されたラベルを用いて開始することができる。第1の学習率減衰を待機することにより、典型的に、訓練精度が十分に高く、予測があまり偏らないという利点がある。学習率減衰の後、1つ以上の訓練ステップを実行することができ、当該1つ以上の訓練ステップにおいては、減衰させた学習率を使用しながらも、以前のラベルベクトルが用いられる。例えば、固定数の訓練エポックΔT(565)を実行して、その後ラベル校正LCを続けて実行することができる。ラベル校正の後、訓練フェーズのさらなる訓練ステップが、更新済みのラベルベクトル及び減衰させた学習率、例えば複数のエポックに従って実行され得る。

20

【0075】

訓練フェーズの開始時の、以前のラベルベクトルが用いられる時間は、学習率減衰後に数回の反復を待機した後にのみ精度及び信頼度の上昇が生じる典型的な訓練効果を指して、待機フェーズと称されることがある。学習率減衰後のエポックTからの予測は、次の学習率減衰前の将来の予測と概ね一致するため、既知のラベル平滑化におけるような余分な誤校正も回避される。ほとんどのアーキテクチャ及びデータセットにおいては、 $T = 1$ が良好な選択であることがわかっている。

30

【0076】

また、後続の学習率減衰の後、ラベル校正LCが適用されるものとしてもよく、また、この場合、典型的には、同一の数の訓練ステップ、例えば、ΔTEポックを含む待機フェーズが適用されるものとしてもよい。例えば、ラベル校正は、各学習率減衰の後、訓練フェーズごとに1回実行され得る。しかしながら、変形形態も可能であり、例えば、ラベル校正を、1つおきの訓練フェーズにおいて実行するものとしてもよいし、第1の学習率減衰などの代わりに第2の学習率減衰の後に最初に実行するものとしてもよい。ラベル校正LCは、訓練569の終わりに達するまで、訓練全体を通して実行することができる。例えば、画像分類器を使用することに基づいて実行されるラベル校正の数は、多くとも3つ若しくは少なくとも3つ、又は、多くとも5つ若しくは少なくとも5つであるものとしてもよい。待機時間、例えば、エポックの数ΔT、及び/又は、ラベル校正頻度を、ハイパーパラメータとして調整することが可能である。

40

【0077】

図5bは、ラベル校正LC(550)の詳細であるが非限定的な例を示している。このラベル校正LCは、画像分類器を使用して画像を訓練するためのラベルベクトルを決定するために、図4及び/又は図5aと組み合わせて使用することができる。

【0078】

ラベル校正LCは、まず、現在のパラメータ値に従って訓練画像に画像分類器を適用して、それぞれの予測ベクトル552を取得することができる。図は、かかる3つの予測[

50

0.7, 0.2, 0.1]、[0.1, 0.8, 0.1]及び[0.5, 0.3, 0.2]を示している。予測ベクトルから予測クラスを推論し、予測クラスとクラスラベルとを比較することによって、比較動作CP(557)において予測を2つのグループに分割することができる。この例においては、予測は、2つのグループ、即ち、正確な予測553[0.7, 0.2, 0.1]及び[0.1, 0.8, 0.1]と不正確な予測555[0.5, 0.3, 0.2]とに分割される。

【0079】

予測クラスがクラスラベルに等しくない画像については、ラベルベクトルは、例えば、One-hot符号化として、予測ベクトルからは独立してクラスラベルに基づいて決定されるものとしてもよい。この例においては、不正確な予測555[0.5, 0.3, 0.2]が画像分類器を誤って導く可能性が高いので、このサンプルの元のハードラベル[0.0, 0.0, 1.0]が取られている。

10

【0080】

予測クラスがクラスラベルに等しい画像について、ラベルベクトルは、予測ベクトル553に基づいてソフトラベルとして決定され得る。この例においては、分類器に追加の正則化を適用するために、温度スケールリング(TS)又は潜在ガウス過程(GP)などの事後較正モジュールPHC(570)によって正確な予測が較正される。例示として、図においては、較正されたソフトラベル554[0.5, 0.3, 0.2]及び[0.2, 0.6, 0.2]が返されている。事後較正PHCの導入により、画像分類器に対する負のフィードバックが可能となる。例えば、TSが統合される場合、画像分類器が極めて過大信頼であるとき、温度係数 $T > 1$ が返され、ソフトラベルがより平滑化されるため、より強力な正則化が生成される場合がある。過大信頼は、将来の訓練においてさらに軽減され得るものであり、逆もまた同様である。代替案としては、事後較正によって較正されない予測を用いること、換言すれば、アイデンティティ較正を適用することが挙げられる。

20

【0081】

正しく分類されたサンプルの較正済み予測554及び誤分類されたサンプルのモデル非依存ラベル555は、将来の訓練のために、新しいソフトラベル556[0.5, 0.3, 0.2]、[0.2, 0.6, 0.2]及び[0.0, 0.0, 1.0]として組み合わせることができる(CB(552))。

【0082】

図6は、本明細書において説明する技術と共に使用するための分類器IC(040)の詳細であるが非限定的な例を示している。この例においては、分類器は、ニューラルネットワーク分類器、特に深層ニューラルネットワークである。図示のプロセスは、説明のために、畳み込みニューラルネットワーク(CNN)における入力画像の分類を、高レベルでかつ大幅に簡略化して記載している。当業者が理解するように、深層ニューラルネットワークなどの他のタイプのニューラルネットワークが使用されるものとしてもよく、以下の例は非限定的であり、単なる例示である。多くの場合、CNNは、概念的に2つの構成要素、即ち、少なくとも1つの畳み込み層を含む畳み込み部分と、少なくとも1つの高密度層(dense layer)を含む高密度層(例えば、全結合(fully-connected))部分とに分割することができる。ただし、高密度層を有している必要はない。例えば、訓練済みの分類器TCは、例えば、いかなる全結合層も含まない完全畳み込みニューラルネットワークであるものとしてもよい。

30

40

【0083】

入力画像II(610)は、画像分類器ICに、より具体的にはフィルタの第1の層CL1(612-1)(又はいくつかのケースでは単一のフィルタ)に入力される。フィルタの各々は、入力画像II(610)に対して畳み込みを行い、共に活性化ボリューム又は「潜在表現」を生成することができる。活性化ボリュームの各「スライス」は、当該フィルタの1つの出力であり得る。潜在表現は、入力画像IIとは異なる空間解像度を有し得る。次いで、潜在表現は、後続の層CL2(612-2)への入力として使用され得る。次いで、CL2のフィルタは、第1のレイヤCL1によって出力された潜在表現に対し

50

て畳み込みを行い、第2の潜在表現を生成することができる。このプロセスは、さらなる層612に対して継続することができる。

【0084】

CNNの畳み込み部分全体を通して、潜在表現の空間次元は、潜在表現ごとに異なり、又は、入力画像の空間次元と異なり得るが、入力画像に対する潜在表現の空間関係は保存されることに留意されたい。

【0085】

CNNの畳み込み部分の最後の層 CL_k ($612 - k$) (最後の畳み込み層と称されることが多いが、いくつかの実施形態においては、畳み込み部分の最後の層 CL_k は、上述したように空間関係を保存するという条件のもとにあれば必ずしも畳み込み層でなくてもよい)の後、結果として生じる潜在表現 LR_k (622)は、高密度層 DL_1 ($614 - 1$)のうちの第1の層に入力される。高密度層614は、各全結合層のニューロンが他の層の全てのニューロンに結合されている全結合層であるものとしてもよい。CNNの高密度部分を通じた後、それぞれのクラスに対するクラス確率 CP_1 (621), ..., CP_n (622)が出力され得る。

【0086】

種々の実施形態においては、予測ベクトルを決定することによって訓練済みの画像分類器を使用する代わりに、訓練済みの画像分類器の一部のみを入力画像に適用して入力画像の特徴空間表現を取得することにより、訓練済みの画像分類器を使用することも可能である。この図の例示的な画像分類器ICにおいては、特徴空間表現は、画像分類器の高密度層 DL_i の出力であり得る。この特徴空間表現は、種々の上流用途を有することができ、例えば、さらなる訓練済みのモデルを特徴空間表現に適用して、例えば、コンピュータ制御システムを制御するための制御データを推論することができる。

【0087】

種々の実施形態においては、画像分類器によって出力された予測ベクトル、例えば画像分類器ICのクラス確率 CP_i を使用して、モデルが適用される入力画像についての分布外スコアを決定することができる。過大信頼が少ないという追加の利点として、本明細書において説明されるように訓練された画像分類器は、分布外サンプルを認識する能力が改善されている可能性がある。既知の画像分類器は、分布外サンプルに対して高い信頼度を出力する傾向がある。従って、既知の画像分類器の予測ベクトルに基づいて分布外スコアを決定することと比較して、本明細書において説明されるような訓練済みの分類器の使用することにより、スコアの精度を高めることができる。例えば、予測のエントロピーを使用して、外れ値サンプルが検出され得る。これは、図に示すCNNだけでなく、本明細書において説明するように訓練された任意の画像分類器に対して行うことができる。

【0088】

分布外スコアを決定するための画像分類器の使用は、分布内サンプル及び分布外サンプルが異なるラベルを有するテストセットを使用して評価することができる。本発明者らは、外れ値サンプルを検出するために、予測のエントロピーをAUCにおける閾値として用いて、かかる評価を行った。分布内データとしてCIFAR-10を用い、分布外データとしてストリートビュー家屋番号 (SVHN: Street View House Number) 及びCIFAR-10-C Gaussian Noise Severity 5を使用した。提供される技術は、複数の試験されたデータセット - ネットワーク対にわたって、他のオンザフライ較正方法よりも性能が優れていることがわかった。また、他の較正方法と比較して変動性が小さく、異なるデータ及びアーキテクチャに対する堅牢性を示している。

【0089】

次に、いくつかの実現形態の詳細及び評価結果について説明する。分類器の信頼度が精度からどの程度逸脱するかを測定するために広く使用される指標が、予測較正誤差 (ECE) である。ECEは、トップ1の予測確率をサンプルの信頼度とみなして、信頼度ダイアグラムの挙動を模倣する。サンプルはソートされ、その信頼度に従ってM個のピンに分割される。ECEは、次式

10

20

30

40

50

【数 7】

$$ECE = \sum_{m=1}^M \frac{|B_m|}{N} |\text{Conf}_m - \text{Acc}_m|$$

に従って、精度と平均信頼度との間の加重平均差を要約する。ここで、 Conf_m 及び Acc_m は、ビン m の平均信頼度及び精度である。 $|B_m|$ は、対応するビン中のサンプルの数を指す。 ECE は通常、静的ビンニング方式を使用するが、ほとんどの予測がヒストグラムの右側に高密度に位置するので、動的ビンニング方式を使用することも可能である。

【0090】

ビンの数 M は重要なハイパーパラメータである。ビンが少な過ぎると、多くの情報が平滑化され得るので、校正誤差を過小評価する（偏りが生じる）可能性がある。ビンが多過ぎると、指標は、高い分散の影響を受ける可能性がある。従来、15個のビンが使用されるが、最近の研究においては、100個のビンを使用した結果も報告されている。本発明者らは、100ビンの静的ビンニングを使用して評価を行ったが、これは、100ビンでは、異なるビンニングスキームが異なる方法のランキングを変更しないことがわかったからである。

【0091】

ECE は、簡便で一般的である。しかしながら、 ECE の他の問題は、その偏り - 分散のトレードオフに加えて、 ECE が 0 になるような些細なケースが常に存在することである。これは、分類器は完全に ECE 校正され得るが、任意の入力に対して一定の固定確率を出力する可能性があることを意味しており、望ましいものではない。負の対数尤度 (NLL) を指標として使用することによって、これらの問題を回避することが文献において示唆されている。 NLL は、汎化及び校正を考慮に入れる。2つのモデルが同様の精度を有する場合、より低い NLL を有するモデルでは、誤校正を受けることが少ない。

【0092】

評価のために、ソフトウェア `PyTorchImageClassification` を使用して、提案された技術を実装する。全てのネットワークアーキテクチャについて、標準的なデータ拡張（ランダムクロッピング及びランダムフリッピング）及び `He / Xavier` 初期化が行われる。異なるシード (`seed`) を用いて実験を 5 回繰り返す。訓練ハイパーパラメータは、元のモデルを訓練するために使用されるものと同様である。0.9 のネステロフのモメンタムを有する `SGD` オプティマイザが、ドロップアウトを使用せずに用いられる。モデルは、完全な訓練（早期停止なし）後に評価される。

【0093】

`WRN - 28 - 10` は、第 1 の 60 エポックについて 0.1、次の 60 エポックについて 0.02、次の 40 エポックについて 0.004、最後の 40 エポックについて 0.0008 の学習率で、200 エポックについて訓練される。訓練バッチサイズを 128 に設定し、重み減衰は 5×10^{-4} である。

【0094】

`DenseNet - BC - 100` は、300 エポックについてバッチサイズ 64 を使用して訓練される。初期学習率は 0.1 であり、エポック 150 及び 225 でそれぞれ 90% 減少する。わずかに異なる重み減衰が使用され、この場合、 10^{-4} である。実験においては、`DenseNet` の圧縮比は 0.5 であり、成長率は $k = 12$ である。

【0095】

`ResNeXt - 29` は、`DenseNet - BC - 100` と同様の学習率スケジュールを使用し、ベースチャンネル及び初期チャンネルは 64 に設定される。重み減衰は 5×10^{-4} であり、バッチサイズは 128 に設定される。

【0096】

全てのデータセットについて、2% の訓練サンプルを検証（校正）のために分割し、この場合、`CIFAR - 10 / CIFAR - 100` については 1000 個のサンプル、`Tiny ImageNet` については 2000 個のサンプルである。`Tiny ImageNet`

tにおける画像は、CIFARのために設計されたモデルが効率的に使用され得るように、 32×32 にダウンサンプリングされる。

【0097】

実験においては、 $\Delta T = 1$ エポックの待機フェーズが使用され、ラベル較正の数は、特定のアーキテクチャの訓練スケジュールに依存する。WRNは3回の学習率減衰を有し、3回のラベル較正が実行される。対照的に、DenseNet及びResNeXtは、2つの学習率減衰のみを有する。これに対応して、ラベル較正は、これらのネットワークにおいて2回実行される。

【0098】

提案された技術の有効性を評価するために、予備実験をWRNについてCIFAR-100で行った。これらの技術は、事後較正なし(LSPアイデンティティ)及び温度スケールリングあり(LSP-TS)で適用され、交差エントロピー(CE)One-hotコーディングベースラインと比較された。性能は、トップ1の精度(Acc@1)、ECE及びNLLに関して評価される。

【0099】

LSPアイデンティティは、ECE及びNLLに関してベースラインよりも大幅に性能が優れていることがわかる。ソフトラベルの追加の利点として、LSPアイデンティティもベースラインよりも正確である。事後較正を適用する有利な負のフィードバックにより、LSP-TSは、全ての指標に対してLSPアイデンティティよりも良好に機能することがわかった。この理由から、LSP-TSは、以下の実験において標準LSP設定として使用され、LSPとも称される。さらなる実験において、LSPは、交差エントロピーベースラインだけでなく、エントロピーペナルティ、ラベル平滑化及び焦点損失較正よりも性能が優れていることがわかった。LSPは、WRN及びResNeXtに対するソフトラベルから余分な精度ブーストを取得することがわかった。LSPの精度は、実験を通して競争力があり、他の方法がわずかにより高い精度を提供する場合においても、LSPは、はるかにより良好なNLLを提供し、LSPが他の方法よりも誤較正による影響はるかに少ないことを示している。

【0100】

提供される訓練方法は、少なくとも2つのハイパーパラメータ、即ち、待機フェーズのエポックの数 ΔT 、及び、ラベル較正の頻度に従って構成することができる。WRN-28-10に対するCIFAR-100上の最良のハイパーパラメータを網羅的に検索する実験を行った。

【0101】

待機フェーズ：WRN-28-10の訓練スケジュールには3つの学習率減衰が存在する。この実験においては、較正頻度は3に固定され、ラベル較正は、各学習率減衰の後に ΔT エポックずつ実行される。検索は、最良の $\Delta T \in \{-20, -10, -1, 0, +1, +10, +20\}$ について実行される。この検索は、 $\Delta T = 1$ が汎化及び較正の両方に対して最良に機能することを示している。

【0102】

頻度：別の実験を実行して、異なる頻度の結果を決定した。頻度 $\{1, 2, 3\}$ の実験においては、 $\Delta T = 1$ が固定され、異なる学習段階におけるラベル較正の影響が調査される。結果は、学習率減衰ごとに少なくとも1回ラベル較正を実行することが非常に有益であることを示している。WRN-28-10の場合、頻度=3は概して最良の性能を有し、これは概して良好な設定であると予想される。頻度 $\{4, 6, 199\}$ の実験について、第1の学習率減衰(頻度=4)の前、各学習段階の中間(頻度=6)、又は、各エポックの後(頻度=199)の追加のラベル較正が有益であるかどうか調査される。第1の学習率減衰の前に追加のラベル較正を実行することにより、約1%低い精度及びこれよりさらに劣悪なNLLを犠牲にして、ECEが改善することがわかった。このデータセットにおいては、頻度=6及び頻度=200について改善は見出されなかった。従って、ラベル較正は、多くても1回の学習率減衰で実行することが好ましい。

【0103】

実験は、全体として、各学習率減衰の1エポック後にラベル較正を実行することが最も有益であることを示している。

【0104】

さらなる実験においては、温度平滑化を用いた訓練を、事後較正器としての潜在ガウス過程を用いた訓練と比較した。これらの実験において、LSP-GPは、NLLを改善するが、ECEをわずかに劣化させる。これは、より多くの較正サンプルを必要とするGPの高い複雑性に起因し得る。概して、GP事後較正器は、提案された訓練技術と組み合わせで良好に機能し、他の較正方法と良好な互換性を有することを示している。

【0105】

事後較正のための複数の選択を比較した：なし(none)、温度平滑化、潜在ガウス過程、及び、I-MAX。Tiny ImageNet上のWRN-28-10について、並びに、CIFAR-100上のWRN-28-10、DenseNet-BC-100及びResNeXt-29について実験を行った。提案されたオンザフライ較正技術を用いずに事後較正を使用するだけでは、画像分類器の精度を改善することができず又はむしろ劣化させることがわかった。一方、提案されたオンザフライ較正を事後較正と組み合わせることにより、より良好な汎化が得られることがわかった。また、ほとんどの事後較正方法がNLL及びECEに関して誤較正を軽減することもわかった。提供される技術を使用して画像分類器が先験的に正規化される場合、誤較正はさらに低減される。I-MAXの場合、提供される技術は、ECEに関して較正を改善しないことがわかった。これは、

20

【0106】

別の実験として、提案された技術と自己蒸留との比較を行った。自己蒸留は、1つのモデルからのソフトラベルを用いて、別のモデルの性能を改善するものである。これは、提供される技術が訓練されるモデル自体のソフトラベルを用いることとは対照的である。次いで、興味深いことに、提供される技術は、自己蒸留よりも良好に機能することがわかった。具体的には、提供される技術を教師なし知識蒸留(TfKD-self:Teacher-free Knowledge Distillation)と称される技術と比較する実験を行った。比較は、Acc@1、NLL及びECEに関してCIFAR-10/100で行った。パラメータは、 $\tau = 20$ 及び $\alpha = 0.95$ を用いた。他のハイパーパラメータは、他の実験と同様である。

【0107】

提供される技術は、セットアップ全体にわたってはるかに良好なNLL及びECEを提供し、誤較正の防止において大きい改善を示唆することがわかった。このことは、大部分のデータセット-モデル対に対してより低い精度を有することを犠牲にしている。これは、TfKD-selfが、誤較正を防止するためにではなく、汎化のために設計されているという事実に対応している。

【0108】

TfKD-selfが提供される技術よりも較正が劣っている複数の理由を特定した。まず、TfKD-selfでは、ハードラベルを有する交差エントロピー損失が全てのサンプルに適用される。提供される技術では、ラベルベクトルを更新する際に、誤って分類されたサンプルに対してハードラベルが用いられるのみである。TfKD-selfにおける交差エントロピーの前方の係数 $1 - \alpha$ は通常小さいが、選択なしでOne-hotコーディングラベルを用いることは、依然として過大信頼のリスクを増加させる。過大信頼を回避するために α が1に設定された場合には、訓練精度が100%ではない重度に正則化された教師モデルは、誤ったソフトラベルを提供して、生徒モデルの性能を劣化させる

10

20

30

40

50

可能性がある。本明細書において説明しているサンプル選択のメカニズムにおいては、かかるトレードオフに対峙する必要はない。

【0109】

さらに、T f K D - s e l fにおける教師モデルは、事前訓練済みのモデルである。ハードラベルによって訓練された事前訓練モデルの予測は、ハードラベルに比較的近く、概して過大信頼である。T f K D - s e l fは、教師の出力と生徒の出力との間の一貫性を強制するので、生徒モデルの過大信頼が予想され得る。対照的に、提供される技術は、モデルを訓練する異なる学習フェーズにおいて、校正済み予測をソフトラベルとして用いることができ、これは、良好に校正されたモデルを模倣することと等価である。

【0110】

さらに、T f K D - s e l fでは、温度スケーリングは、完全な模倣のために教師モデル及び生徒モデルからのロジットに同時に適用されるが、提供される技術は、教師モデル（以前の予測）からのロジットに事後校正（例えば、温度スケーリング）を任意選択的に適用するのみである。これにより、生徒モデルの出力が教師モデルの出力から逸脱することが可能となり、訓練のための追加の校正の自由度が増す。

【0111】

図7は、分類器、例えば画像分類器を訓練するコンピュータ実装された方法700のブロック図を示している。分類器は、入力インスタンス、例えば、入力画像の予測ベクトルを決定するように構成され得る。予測ベクトルは、複数のそれぞれのクラスに対するクラス確率を含み得る。方法700は、図1のシステム100の動作に対応し得る。しかしながら、これは、方法700が他のシステム、装置又はデバイスを使用して実行されるものとしてもよいという点において、限定ではない。

【0112】

方法700は、“ACCESS CLASSIFIER, TRAINING DATA”と題する動作において、分類器の訓練可能パラメータと、複数の訓練インスタンス及び対応するクラスラベルを含む訓練データセットとにアクセスすること(710)を含み得る。

【0113】

方法700は、分類器を訓練することを含み得る。訓練は、“INITIAL TRAINING STEPS”と題する動作において、分類器を訓練する1つ以上の初期訓練ステップを実行して、訓練可能パラメータの現在のパラメータ値を取得すること(720)を含み得る。訓練は、“SELECT TRAINING DATA”と題する動作において、訓練インスタンス、例えば、訓練画像及び訓練データセットからの対応するクラスラベルを選択すること(730)を含み得る。

【0114】

訓練は、“DETERMINE LABEL VECTOR”と題する動作において、訓練インスタンスに対して分類器を訓練するための現在のラベルベクトルを決定すること(740)を含み得る。決定すること(740)は、“APPLY CLASSIFIER”と題する動作において、現在のパラメータ値に従って分類器を訓練インスタンスに適用して予測ベクトルを取得すること(742)を含み得る。決定すること(740)は、“INFERR CLASS”と題する動作において、予測ベクトルから複数のクラスのうちの予測クラスを推論すること(744)を含み得る。決定すること(740)は、“COMPARE INFERRED TO CLASS LABEL”と題する動作において、予測クラスとクラスラベルとを比較すること(746)を含み得る。決定すること(740)は、“DETERMINE LABEL VECTOR BASED ON PREDICTION”と題する動作において、予測クラスがクラスラベルに等しい場合、予測ベクトルに基づいて現在のラベルベクトルをソフトラベルとして決定すること(748)を含み得る。

【0115】

訓練は、“TRAINING STEP”と題する動作において、現在のラベルベクトル

10

20

30

40

50

を用いて訓練インスタンスに対して分類器を訓練する現在の訓練ステップを実行すること(750)を含み得る。

【0116】

図8は、訓練済みの分類器、例えば画像分類器を使用するコンピュータ実装された方法800のブロック図を示している。方法800は、図2のシステム200の動作に対応し得る。しかしながら、これは、方法800が他のシステム、装置又はデバイスを使用して実行されるものとしてもよいという点において、限定ではない。

【0117】

方法800は、“ACCESS CLASSIFIER”と題する動作において、分類器の訓練済みのパラメータにアクセスすること(810)を含み得る。分類器は、本明細書 10
に説明される方法、例えば、図7の方法700に従って以前に訓練されているものとしてよい。

【0118】

方法800は、“OBTAIN INSTANCE”と題する動作において、入力インスタンス、例えば入力画像を取得すること(820)を含み得る。

【0119】

方法800は、“APPLY CLASSIFIER”と題する動作において、分類器を入力インスタンスに適用して(830)、入力インスタンスの予測ベクトルを決定することを含み得る。方法800は、“OUTPUT PREDICTION”と題する動作において、予測ベクトルに基づいて予測を出力すること(835)をさらに含み得る。 20

【0120】

適用すること(830)及び出力すること(835)に代えて又はこれに加えて、方法800は、“APPLY CLASSIFIER PART”と題する動作において、分類器の一部を入力インスタンスに適用して(840)、入力インスタンスの特徴空間表現を取得することを含み得る。方法800は、“OUTPUT FEATURE SPACE REPRESENTATION”と題する動作において、特徴空間表現を出力すること(845)をさらに含み得る。

【0121】

概して、図7の方法700及び図8の方法800の動作は、適用可能な場合、例えば入力/出力関係によって必要とされる特定の順序に従って、任意の適当な順序で、例えば、 30
連続して、同時に、又は、これらの組合せにより実行され得ることが理解されよう。方法の一部又は全部を組み合わせることができ、例えば、訓練済みの分類器を使用する方法800は、この分類器が方法700に従って訓練されることに続いて適用されるものとしてもよい。

【0122】

方法は、コンピュータ実装された方法として、専用ハードウェアとして、又は、これら両方の組合せとして、コンピュータ上に実装され得る。また、図9に示されているように、コンピュータのための命令、例えば実行可能コードは、例えば、一連の機械可読物理マーク910の形態で、及び/又は、例えば、種々の電氣的、磁氣的若しくは光学的な特性若しくは値を有する一連の要素として、コンピュータ可読媒体900上に記憶されている 40
ものとしてよい。媒体900は、一時的又は非一時的であり得る。コンピュータ可読媒体の例として、メモリデバイス、光記憶デバイス、集積回路、サーバ、オンラインソフトウェアなどが挙げられる。図9は、光ディスク900を示している。代替的に、コンピュータ可読媒体900は、本明細書において説明されるような訓練済みの分類器の訓練済みのパラメータを表現するデータ910を含み得る。

【0123】

実施例、実施形態又は任意選択的な特徴は、非限定的なものとして示されているか否かにかかわらず、特許請求される本発明を限定するものとして理解されるべきではない。

【0124】

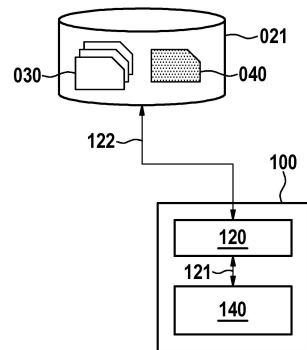
上述した実施形態は、本発明を限定するのではなく例示するものであり、当業者は、添 50

付の特許請求の範囲から逸脱することなく、多くの代替実施形態を設計することができることに留意されたい。特許請求の範囲において、括弧内に置かれた任意の参照符号は、特許請求の範囲を限定するものとして解釈されてはならない。動詞「含む (comprise)」及びその活用形の使用は、請求項に記載されたもの以外の要素又は段階の存在を排除するものではない。要素に先行する冠詞“a”又は“an”は、複数のかかる要素の存在を除外しない。要素のリスト又はグループに先行するときの「～のうちの少なくとも１つ」などの表現は、リスト又はグループからの要素の全て又は任意のサブセットの選択を表す。例えば、「A、B及びCのうちの少なくとも１つ」という表現は、Aのみ、Bのみ、Cのみ、A及びBの両方、A及びCの両方、B及びCの両方、又は、A、B及びCの全てを含むものとして理解されるべきである。本発明は、複数の別個の要素を含むハードウェアによって、及び、適当にプログラムされたコンピュータによって実装することができる。複数の手段を列挙するデバイスの請求項において、これらの手段の複数の、ハードウェアの全く同一の構成要素によって具現化されるものとしてよい。特定の手段が相互に異なる従属請求項に記載されているという単なる事実は、これらの手段の組合せを有利に使用することができないことを示すものではない。

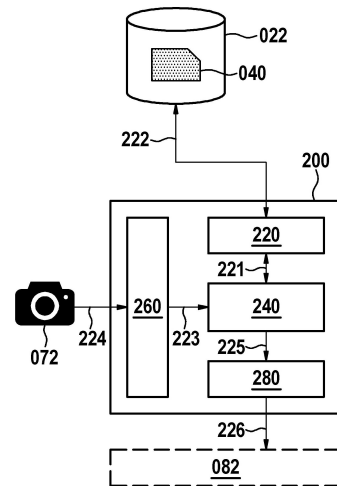
10

【図面】

【図 1】



【図 2】



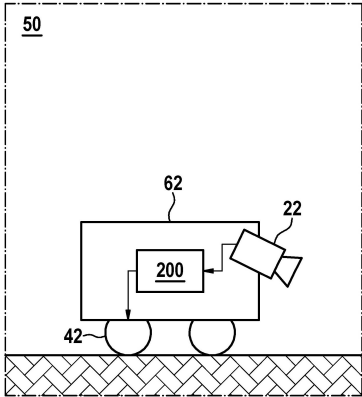
20

30

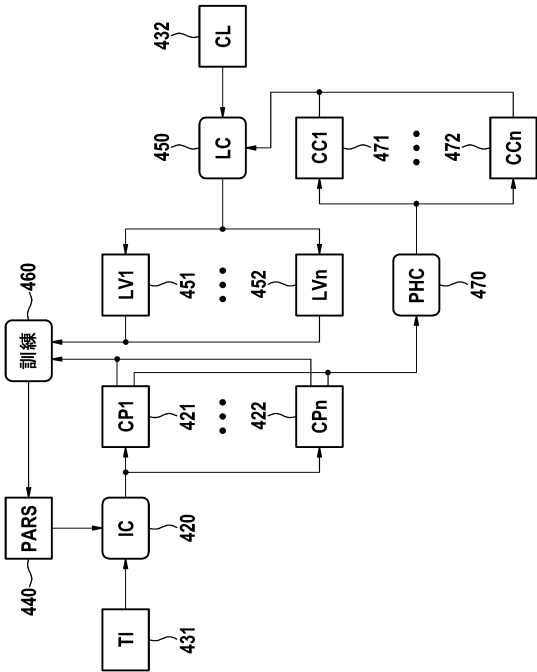
40

50

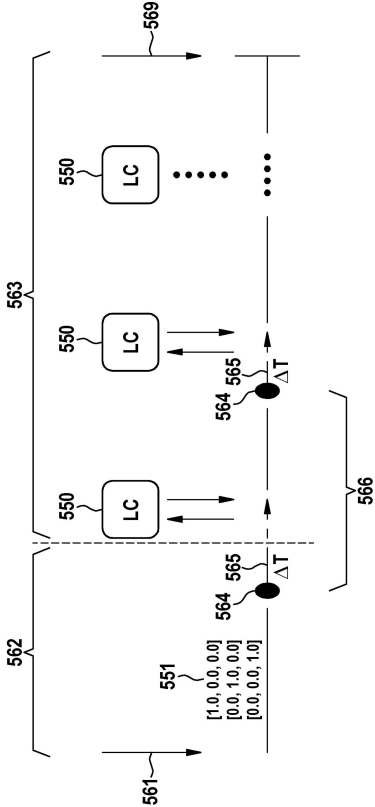
【 図 3 】



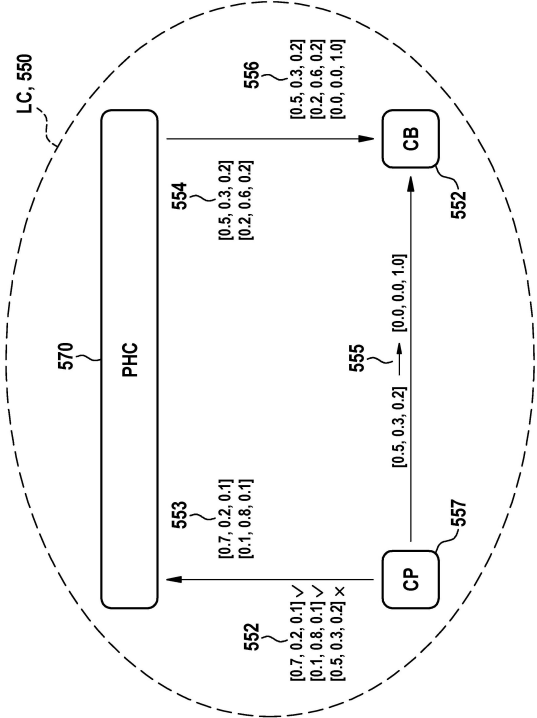
【 図 4 】



【 図 5 a 】



【 図 5 b 】



10

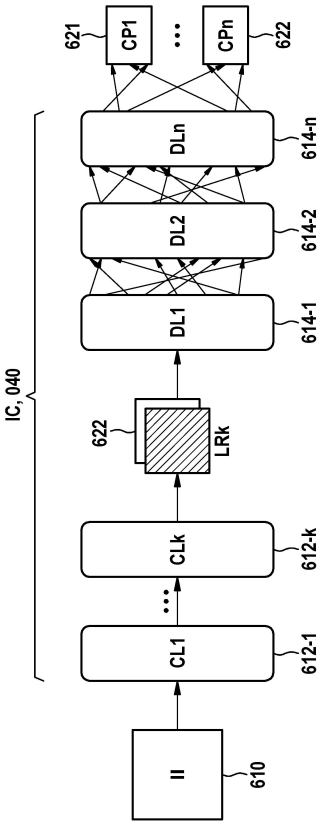
20

30

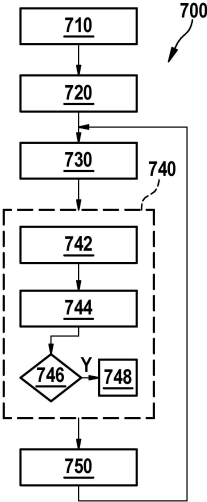
40

50

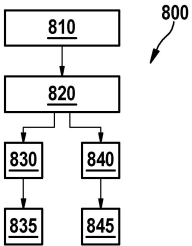
【 図 6 】



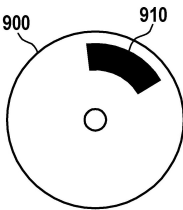
【 図 7 】



【 図 8 】



【 図 9 】



10

20

30

40

50

フロントページの続き

弁理士 永島 秀郎
(74)代理人 100162880
弁理士 上島 類
(72)発明者 カニル パテル
ドイツ連邦共和国 シュトゥットガルト ノイエンシュタイナー シュトラーセ 7
(72)発明者 リアングュ ツォン
ドイツ連邦共和国 シュトゥットガルト アルマンドリング 3 エー ツィマー 1 1
F ターム (参考) 5L096 DA04 GA30 HA11 KA04 KA15 MA07