

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7313828号

(P7313828)

(45)発行日 令和5年7月25日(2023.7.25)

(24)登録日 令和5年7月14日(2023.7.14)

(51)国際特許分類

F I

G 0 6 N 20/10 (2019.01)

G 0 6 N 20/10

G 0 6 N 20/00 (2019.01)

G 0 6 N 20/00 1 3 0

G 0 6 N 3/02 (2006.01)

G 0 6 N 3/02

G 0 6 T 7/00 (2017.01)

G 0 6 T 7/00 3 0 0 E

請求項の数 12 (全35頁)

(21)出願番号 特願2019-8776(P2019-8776)
(22)出願日 平成31年1月22日(2019.1.22)
(65)公開番号 特開2020-119154(P2020-119154
A)
(43)公開日 令和2年8月6日(2020.8.6)
審査請求日 令和4年1月14日(2022.1.14)

(73)特許権者 000001007
キヤノン株式会社
東京都大田区下丸子3丁目30番2号
(74)代理人 100090273
弁理士 國分 孝悦
(72)発明者 斎藤 侑輝
東京都大田区下丸子3丁目30番2号
キヤノン株式会社内
審査官 佐田 宏史

最終頁に続く

(54)【発明の名称】 情報処理装置、情報処理方法、及びプログラム

(57)【特許請求の範囲】

【請求項1】

認識対象データを取得する取得手段と、
特徴量抽出器を用いて前記認識対象データの対象特徴量を抽出する抽出手段と、
前記対象特徴量に基づいて所定の数 of 代表特徴量を生成する生成手段と、
前記対象特徴量と前記代表特徴量とを比較する比較手段と、
前記比較の結果に基づいた認識処理結果を出力する出力手段と、
を有し、

前記生成手段は、学習済みの代表点生成器を用いて、前記認識対象データに対応する代表点の前記代表特徴量を生成し、

前記認識対象データに対応する代表点は、複数の学習データのうち前記認識対象データに対応する補助データと、前記対象特徴量とに基づいて生成され、

さらに、

前記複数の学習データから前記補助データを選択する選択手段と、

前記対象特徴量と前記選択された補助データの含まれる学習データとに基づいて損失を計算する計算手段と、

前記損失に基づいて、前記特徴量抽出器を含む学習モデルを学習する学習手段と、を有し、

前記比較手段は、複数の代表点と前記補助データとの距離を基に正常か否かを判定し、

前記学習手段は、正常と判定された場合には前記距離を近づけるように学習し、正常でない
と判定された場合には前記距離を遠ざけるように学習することを特徴とする情報処理装

置。

【請求項 2】

前記代表点生成器は、前記認識対象データを特徴空間で表現した前記対象特徴量に基づいて、前記所定の数の前記代表点を生成することを特徴とする請求項 1 に記載の情報処理装置。

【請求項 3】

前記代表点生成器は、前記認識対象データを所定のデータ空間で表現した点に基づいて前記所定の数の前記代表点を生成することを特徴とする請求項 1 に記載の情報処理装置。

【請求項 4】

前記選択手段は、前記認識対象データを示す点について近傍探索することによって前記代表点を選択することを特徴とする請求項 2 に記載の情報処理装置。

10

【請求項 5】

前記選択手段は、前記所定のデータ空間における前記認識対象データを示す点について近傍探索することによって前記代表点を選択することを特徴とする請求項 3 に記載の情報処理装置。

【請求項 6】

前記生成手段は、前記所定の数の特徴量を所定のサイズの特徴量に変換することによって前記代表特徴量を生成することを特徴とする請求項 1 乃至 5 のいずれか 1 項に記載の情報処理装置。

【請求項 7】

前記比較手段は、前記対象特徴量と前記代表特徴量との距離が所定値以上である場合に前記認識対象データが異常であると判断することを特徴とする請求項 1 乃至 6 のいずれか 1 項に記載の情報処理装置。

20

【請求項 8】

前記代表点生成器は、敵対的生成ネットワークのモデルを基に学習されることを特徴とする請求項 1 乃至 7 のいずれか 1 項に記載の情報処理装置。

【請求項 9】

前記代表点生成器は、前記学習モデルの複数階層の各層に基づいて学習されることを特徴とする請求項 1 乃至 7 のいずれか 1 項に記載の情報処理装置。

【請求項 10】

前記代表点生成器は、時系列性のあるデータに基づいて学習されることを特徴とする請求項 1 乃至 7 のいずれか 1 項に記載の情報処理装置。

30

【請求項 11】

情報処理装置が実行する情報処理方法であって、
認識対象データを取得する取得工程と、
特徴量抽出器を用いて前記認識対象データの対象特徴量を抽出する抽出工程と、
前記対象特徴量に基づいて所定の数の代表特徴量を生成する生成工程と、
前記対象特徴量と前記代表特徴量とを比較する比較工程と、
前記比較の結果に基づいた認識処理結果を出力する出力工程と、
を有し、

40

前記生成工程では、学習済みの代表点生成器を用いて、前記認識対象データに対応する代表点の前記代表特徴量を生成し、

前記認識対象データに対応する代表点は、複数の学習データのうち前記認識対象データに対応する補助データと、前記対象特徴量とに基づいて生成され、

さらに、

前記複数の学習データから前記補助データを選択する選択工程と、

前記対象特徴量と前記選択された補助データの含まれる学習データとに基づいて損失を計算する計算工程と、

前記損失に基づいて、前記特徴量抽出器を含む学習モデルを学習する学習工程と、を有し、

前記比較工程では、複数の代表点と前記補助データとの距離を基に正常か否かを判定し、

50

前記学習工程では、正常と判定された場合には前記距離を近づけるように学習し、正常でないと判定された場合には前記距離を遠ざけるように学習することを特徴とする情報処理方法。

【請求項 12】

コンピュータを、請求項 1 乃至 10 のいずれか 1 項に記載の情報処理装置の各手段として機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、映像や画像等のデータから認識対象を認識する技術に関する。

10

【背景技術】

【0002】

映像や画像等のデータから、認識対象としての物体およびその状態を認識するために、その認識対象のデータを用いて特徴量を抽出し、その特徴量から最も近い学習データの特徴量との距離に基づいて、認識対象の状態やクラス等を判別する技術が存在する。例えば異常検知を行う場合の手法として、非特許文献 1 には、認識対象データの近傍に存在する正常データとの距離、ないし正常データのクラスタとの距離に基づいて異常を検知する方法が開示されている。

【先行技術文献】

【非特許文献】

20

【0003】

【文献】Nearest-Neighbor and Clustering based Anomaly Detection Algorithms for RapidMiner. M. Amer and M. Goldstein. 2012

【発明の概要】

【発明が解決しようとする課題】

【0004】

ここで、入力データの認識をより高速に行うためには、認識処理時間を低減した認識システムが望まれる。また、入力データの認識をより高精度に行うためには、認識能力がさらに向上した認識システムが望まれる。

非特許文献 1 のように、認識対象データと近傍データとの距離を算出する場合、近傍データの検索処理の際に学習データの数に依存した処理時間が必要となる。また、非特許文献 1 のように、認識対象データとクラスタ中心との距離を算出する場合、同様にクラスタの数に依存した処理時間が必要となる。一方で、学習データとその量が多様・大規模であると、様々な学習データを表現するために、大量のクラスタの種類やクラス数を用いて学習データを近似する必要があるため、クラスタの数に依存した処理が行われることは望ましくない場合がある。しかしながら、クラスタの数に依存した処理時間を低減するために、クラスタの数を減ずるようにすると、クラスタ中心とクラスタに属するデータとの乖離度合いが大きくなるため、認識対象の検索処理時に誤答の割合が大きくなる場合がある。

30

【0005】

そこで、本発明は、認識処理におけるデータ認識精度の向上と認識処理時間の低減とを可能にすることを目的とする。

40

【課題を解決するための手段】

【0006】

本発明の情報処理装置は、認識対象データを取得する取得手段と、特徴量抽出器を用いて前記認識対象データの対象特徴量を抽出する抽出手段と、前記対象特徴量に基づいて所定の数の代表特徴量を生成する生成手段と、前記対象特徴量と前記代表特徴量とを比較する比較手段と、前記比較の結果に基づいた認識処理結果を出力する出力手段と、を有し、前記生成手段は、学習済みの代表点生成器を用いて、前記認識対象データに対応する代表点の前記代表特徴量を生成し、前記認識対象データに対応する代表点は、複数の学習データのうち前記認識対象データに対応する補助データと、前記対象特徴量とに基づいて生成

50

され、さらに、前記複数の学習データから前記補助データを選択する選択手段と、前記対象特徴量と前記選択された補助データの含まれる学習データとに基づいて損失を計算する計算手段と、前記損失に基づいて、前記特徴量抽出器を含む学習モデルを学習する学習手段と、を有し、前記比較手段は、複数の代表点と前記補助データとの距離を基に正常か否かを判定し、前記学習手段は、正常と判定された場合には前記距離を近づけるように学習し、正常でないとは判定された場合には前記距離を遠ざけるように学習することを特徴とする。

【発明の効果】

【0007】

本発明によれば、認識処理における一タ認識精度の向上と認識処理時間の低減とが可能となる。

【図面の簡単な説明】

【0008】

【図1】異常検知システムの構成例を示す図である。

【図2】異常検知システムの動作（学習時）を示すフローチャートである。

【図3】選択部の動作を示すフローチャートである。

【図4】入力画像と人体領域の例を示す図である。

【図5】学習部の動作を示すフローチャートである。

【図6】CNNの構成例（学習時）を示す図である。

【図7】異常検知システムの動作（検出時）を示すフローチャートである。

【図8】CNNの構成例（検出時）を示す図である。

【図9】Siamese CNNの構成例（学習時）を示す図である。

【図10】CNNの構成例（学習時）を示す図である。

【図11】Autoencoderの構成例（学習時）を示す図である。

【図12】補助情報の利用による代表点生成時の構成例（学習時）を示す図である。

【図13】複数層における代表点生成と階層的クラスタ損失の例を示す図である。

【図14】クラスタごとの代表点生成器の例を示す図である。

【図15】多段階系列長の例を示す図である。

【発明を実施するための形態】

【0009】

以下、添付の図面を参照しながら、本発明の実施形態について説明する。なお、以下の実施形態において示す構成は一例に過ぎず、本発明は図示された構成に限定されるものではない。

< 第1の実施形態 >

以降の説明では、物体およびその状態についての認識対象に関するデータを認識対象データと呼び、認識対象データから抽出した特徴量を対象特徴量と呼ぶことにする。なお認識対象データが特徴空間上に写像された場合でも、一般的に特徴空間上のデータと呼称することが多いと考えられるため、以下の説明で単に認識対象データと記す場合は、例えば元画像空間上のデータのみならず、特徴空間上のデータをも含むものとする。認識対象は人間が指定したものでも、機械が処理対象として指定したものでも、単数であっても、複数であってもよいものとする。以降における個々の事例では、具体例として認識対象がどのような性質のものであるかを述べる場合があるが、本発明において原理的にその場合に限るという意図ではない。本実施形態では、学習データの数に依存しない定数時間で認識対象の認識処理を行う例を第一に示す。また本実施形態では異常検知システムの例を挙げているため、異常検知という目的に合わせた学習のためのアルゴリズム例を挙げている。学習アルゴリズムとしては、例えば正常クラス・異常クラスを前提にしたTwo-Class学習アルゴリズムや、正常データのみを学習に用いるOne-Class学習を挙げることができる。もちろん、本発明はその他の目的にも用いることができ、以降では様々な構成例について順を追って説明する。

【0010】

詳細は後述するが、本実施形態では、認識対象データの近傍（以降、認識対象近傍と呼ぶ）の点を、認識対象データと比較すべき代表点として予測し、その代表点と認識対象データとの間の距離を算出し、その距離を基に異常検知を行う構成及び動作の例を示す。ここで、認識対象データに基づいて予測された代表点は、認識対象近傍の正常データを表現するように予め学習が行われたものとする。ここで述べた学習には、代表点を生成する処理に関する学習と、特徴抽出を行う処理に関する学習と、の二種類があるものとする。なお後述するように、それら二種類の学習処理を一つの学習処理の系として捉えてもよい。また、認識対象データを基に予測された代表点は、認識対象近傍の学習データを表現するものである。このため、代表点を用いることで、学習データの全てと認識対象データとを比較する場合や、学習データを全てクラスタリングした上で全てのクラスタ中心と認識対象とを比較する場合よりも、少ない数の比較により認識対象の認識を行うことが可能となる。本来であれば、認識対象データと、認識対象近傍データ（またはそのクラスタ中心等）と、に基づいて決まる近傍との距離は、認識対象データに基づいて"ここにあるであろう"と予測される"認識対象データと比べるべき点"との距離によって求められる。このため、本実施形態では、代表点を"予測する"という表現を用いる場合があり、これは言い換えると、代表すると予測される点を"生成する"とも言える。また本実施形態では、第一に、代表点は固定のM個として説明を行う。そして本実施形態では、M個の代表点で済むように、特徴抽出器および代表点生成器を学習するものとする。なお、ここでは代表点という用語を用いたが、後述するように構成としては点に限らない。加えて、代表点の学習の仕方によって、"代表点が認識対象近傍の正常データを表現すると考える"というパターン以外の目的にも用いることができることを後述する。

【0011】

以降では、例えば、Convolutional Neural Network (CNN) を特徴抽出および代表点生成に用いて、学習する例を示す。代表点は、CNN の特徴空間上に生成するものとして、以降では代表特徴量と呼称することとする。なお、後述するように、CNN は必須の構成事例ではなく、特徴抽出器の一例とする。

【0012】

本実施形態では、具体例として異常検知システムの構成を挙げており、異常検知システムとしては、監視カメラによって撮像された映像中から異常を検出する機能を備えたシステムを例示する。

本実施形態における異常検知システムでは、監視対象をカメラ等の撮像装置で撮影し、撮影した映像データに基づいて、監視対象に異常があるか否かの判定が行われる。そして、異常があると判定した場合には、警備室等の監視センタに常駐する監視者に対し、異常がある旨の警告、例えば警告表示や警告音の出力による警告を行う。なお、この監視対象には、例えば、一般家庭の屋内及び屋外、又は病院、駅などの公共施設が含まれる。

【0013】

本実施形態における異常検知システムの動作は、「学習時」と、「検出時」と、の2つのステージからなる。これら2つのステージのうち、最初の第1ステージは学習を行う際の動作を示しており、また2つ目の第2ステージは学習した結果に基づいて実際の検出処理を行う際の動作を示している。

図1は、第1の実施形態に係る情報処理装置の一例としての異常検知システムの概略構成例を示すブロック図である。本実施形態の異常検知システムは、学習装置10、学習データの記憶部D1、認識装置20、認識対象データの記憶部D2、判定装置30、端末装置40を有して構成されている。

【0014】

まず、本実施形態の異常検知システムにおいて学習時の動作に関わる構成について説明する。

図1の異常検知システムは、学習に関わる構成として、学習装置10を備えている。なお、これらの装置内および装置間は、電子回路を介して接続されていても、外部記憶装置を介して接続されていても、またネットワークを介して接続されていてもよい。このネッ

10

20

30

40

50

トワークには、例えば携帯電話回線網やインターネットが適用できる。以降で説明する装置等に関しても、同様に様々な方法で接続されていてよい。このことは後述する認識装置20等においても同様であるとする。

学習装置10は、選択部11と、学習部12とを備えている。

学習部12は、特徴抽出部121と、代表特徴量生成部122と、損失算出部123とを備えている。

【0015】

次に、異常検知システムにおいて学習時の動作を、図2を用いて説明する。図2は、異常検知システムにおける学習時の動作を示すフローチャートである。なお、以降の説明では、図2のフローチャートにおける各ステップS200～ステップS204をS200～S204と略記する。このことは後述する他のフローチャートにおいても同様とする。

【0016】

まずS200において、学習装置10は、初期化処理を行う。例えば、学習装置10は、学習時に用いるCNNの初期化を行う。具体例は後述する。

次にS201において、選択部11は、学習データ記憶部D1から学習処理の対象となる学習対象データを読み込み、学習部12にmini batch（ミニバッチ）データを送る。特徴抽出部121は、選択部11から学習データを受け取り、その受け取ったデータを基に特徴抽出器により特徴量を抽出し、認識対象特徴量（以下、対象特徴量と呼ぶ）を代表特徴量生成部122に送り、mini batch特徴量を損失算出部123に送る。代表特徴量生成部122は、受け取った対象特徴量から、代表特徴量生成器により代表特徴量を生成して、損失算出部123に送る。損失算出部123は、受け取ったmini batch特徴量と代表特徴量とから、損失関数に基づいて損失値を算出する。以降では、損失値や、損失関数を微分すること等によって得た勾配情報を、誤差に関する情報または簡単に誤差と呼ぶ。これらの詳細については後述する。

【0017】

次にS202に進むと、損失算出部123は、誤差情報を代表特徴量生成部122および特徴抽出部121に送る。代表特徴量生成部122は、受け取った誤差情報に基づいて、さらに代表特徴量生成器の誤差を算出して、特徴抽出部121に送る。特徴抽出部121は、損失算出部123と代表特徴量生成部122とから受け取った（CNNの上段の層の損失である）誤差を用いて、特徴抽出器の学習処理を行う。詳細については後述する。

【0018】

S203に進むと、代表特徴量生成部122は、S202で損失算出部123から受け取った誤差情報を用いて、代表特徴量生成器の学習処理を行う。

その後、S204において、学習装置10は、学習が完了すると、これまで学習した特徴抽出器および代表特徴量生成部を、学習データ記憶部D1に保存する。

【0019】

次に、選択部11の動作に関して、図3に基づいて詳細な説明を行う。図3は、選択部11における学習データの選択動作を示すフローチャートである。

まずS301において、選択部11は、学習データ記憶部D1から学習データを受け取る。ここでの学習データとは、ある一つの監視カメラから得られた監視映像から作成された人物画像であるとする。

【0020】

ここで、ある監視映像から1フレームを抜き出した画像の例を図4に示す。図4に示した画像401は、ある交差点における監視映像の画像例を示しており、この画像401内には撮像された被写体等のオブジェクト402～405が存在しているとする。また画像401内の点線で示したエリア406～409は、選択部11の備える人体領域抽出器によって抽出されたBounding Box（バウンディングボックス）を示している。これらのBounding Boxに囲われた部分画像は、それぞれが人体画像に対応している。なお、ここで示したBounding Boxは、あくまでも人体領域が抽出された際の具体例の一つであり、例えば後述する背景差分法によって撮像されたオブジェク

10

20

30

40

50

トの輪郭に沿った小領域を抽出してもよい。

【 0 0 2 1 】

前述したような人体領域を抽出するための方法は複数存在し、例えば背景差分法、物体検出・追尾法、領域分割法の三つがある。監視対象のオブジェクトが、例えば人体のように予め既知のオブジェクトである場合は、ターゲットのオブジェクトのみを検出・追尾する目的に絞られた物体検出・追尾法が比較的適していると考えられる。物体検出・追尾法には、例えば参考文献 1 に開示された方法があり、この方法を用いてもよい。

【 0 0 2 2 】

参考文献 1 : Real-Time Tracking via On-line Boosting. H. Grabner, M. Grabner and H. Bischof. Proceedings of the British Machine Conference, pages 6. 1-6. 10. BMVA Press, September 2006.

10

【 0 0 2 3 】

さらに選択部 1 1 は、学習データ（監視映像から矩形領域などで人体領域として切り出された人体画像）に予め付与された教師データを利用して、人体画像に対して教師データを付与する。例えば、矩形の人体領域を抽出する場合は、人体領域の B o u n d i n g B o x を定義し、該 B o u n d i n g B o x に教師データを付与することができる。なお、矩形ではなく、例えば人体の存在する輪郭に沿って人体領域を抽出する場合は、画像に対する人体領域のマスクを定義し、該マスクに基づいて人体画像を作成することができる。

【 0 0 2 4 】

20

ここで、教師データは、対象をどのように分類すべきかを示すラベルである。ラベルの種類や、どのような対象にどのようなラベルを付与するか、ということは問題に依存するため、本実施形態では、異常検知システムを使用または導入するユーザが予め決定し、学習データに対して教師データを付与しておくものとする。例えば、正常な映像から得られる人体画像は正常であるとみなし、それらの人体画像に対して正常というラベルを付与してもよい。なお、以降では、監視員が異常であると判断した人体画像には、異常というラベルを付与するものとする。すなわち、ここでの問題設定は、ある特定の監視カメラから得られた映像（に含まれる人体画像）に関する正常 - 異常の 2 クラス問題となる。

【 0 0 2 5 】

学習データに予め付与する教師データは、例えば撮像された被写体の画像上の領域に対して、ユーザが付与することができる。より具体的には、例えば歩行者のオブジェクト 4 0 3 の領域をユーザが手動によって指定し、その領域に対して例えば正常を示すラベルを付与することができる。このとき、抽出した人体領域と、付与された教師データの領域とが重畳している場合、その人体領域に対して最も大きな面積の割合で重畳された教師データを付与してよい。当然ながら、必ず前述のやり方で教師データを付与しなければならないわけではなく、例えば予め人体領域を抽出しておき、それぞれの人体領域に対して、ユーザが教師データを付与してもよい。なお、本実施形態では人体領域を抽出する例を示したが、当然その他の物体領域を抽出してもよいし、また領域を抽出せずに画像全体を認識処理してもよい。どのような領域を抽出し、認識処理の対象とするかは問題依存であるため、問題に応じて設定する必要がある。また、ここでは説明を簡易にするために、学習処理のターゲットとなる全ての監視映像から予め学習データ（人体画像）を作成しておくものとする。

30

【 0 0 2 6 】

次に S 3 0 2 に進むと、選択部 1 1 は、得られた学習データの前処理を行う。本実施形態では、各画像を 2 2 4 × 2 2 4 ピクセルの画像サイズに変形した後、平均画像を引く処理を行うものとする。ここで平均画像は、学習データの平均画像を示す。本実施形態では、学習画像を水平方向に反転して水増しし、水増し前の画像と、水増し後の画像とを、全て学習画像として扱うこととする。

40

【 0 0 2 7 】

次に S 3 0 3 に進むと、選択部 1 1 は、認識対象となる学習データを決定し、それぞれ

50

の認識対象の近傍探索処理を行う。認識対象となる学習データは、全ての学習データをそれぞれ認識対象として扱ってもよいし、重要な学習データを予め選別して、認識対象として選定してもよい。ここで選ぶ認識対象は一枚の学習データとする。すなわち、ある認識対象を選んだとき、その認識対象に類する近傍データが複数選ばれるものとする。なお、複数個のデータによって認識対象を構成することも可能であり、そのような例は後述する。また、選択するのは一つでも可能であるが、認識対象近傍に異なる種類の近傍データの分布が存在しうる場合、複数個がよいと考えられる。

【0028】

近傍探索を行う空間は、ここで述べる第一の例としては画像空間上であるとする。すなわち選択部11は、監視映像から切り出され、前処理された学習データが画像空間上に分布しているとみなして、認識対象データの近傍探索を行う。画像空間以外にも、例えば特徴空間を用いてもよいため、それについては後述する。近傍探索を用いる方法はどのような公知の手段を選んでよいが、ここでは参考文献2に記載されている k -Nearest Neighbor (k -NN)を用いる。なお、 k -NNは、最初に一度実行するだけでもよいし、各エポックの開始時に実行してもよいし、学習用の認識対象を選択する（いわゆるmini batchを作成する）たびに実行してもよいとする。

【0029】

参考文献2：C. Huang et al. Learning Deep Representation for Imbalanced Classification. CVPR2016.

【0030】

次にS304に進むと、選択部11は、学習用データ部分集合として、Mini batchを設定する。本実施形態の第一の例では、Mini batchのサイズは、認識対象1個と、該認識対象近傍データ99個をひとまとめにしたものとする。この認識対象近傍は、前述の近傍探索処理結果によって得られたものとする。このMini batchとしては、学習に用いるために一度に全てのMini batchを作成してもよいし、学習のエポックが進むたびに作成し直してもよいし、その他の学習のどのタイミングで作成してもよいものとする。この例では、最初に一度に全てのMini batchを作成し、Mini batch集合を得るものとする。なお、学習処理の例として、後述する参考文献3の方法を用いてもよい。

【0031】

S305に進むと、選択部11は、Mini batch集合から一つMini batchを選択し、学習部12に送る。このとき選択するMini batchは、例えばランダムな順番でMini batchを選択する。なお、エポックとは、Mini batch集合の全てのMini batchを何回学習したかを表すものとする（ひととおり学習したら1エポック進むというように数える）。本実施形態では、一つの例として、予め決められたエポック数に基づいて学習完了の判定を行うとする。

その後、S306に進むと、選択部11は、既定のエポック数に到達していた場合、学習処理を終了する。

【0032】

次に、学習部12の動作に関して、図5に基づいて詳細な説明を行う。図5は、学習部12における学習動作を示すフローチャートである。

まずS501において、学習部12は、CNNのパラメータ（結合重み・バイアス項・学習パラメータなど）を初期化する。初期化にあたっては、CNNのネットワーク構造を予め決めておく必要がある。ここで用いるネットワーク構造や初期パラメータは、例えば参考文献3と同じものを用いてよいし、独自に定義したネットワーク構造を用いてもよい。

【0033】

参考文献3：A. Krizhevsky et al. ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems 25 (NIPS)、2012.

【0034】

10

20

30

40

50

図6はCNNの模式図の一例を示した図である。この図6を用いて、CNNの構成および動作の例について簡単に説明する。図6の特徴抽出器620と、代表点生成器621とは、本実施形態で用いるCNNのネットワーク構造の一例を示したものである。図6では、特徴抽出器620が、入力層601と、convolution1層602と、pooling1層603と、convolution2層604と、pooling2層1305とから構成されていることを示している。また図6では、2つの階層間の処理方法として、convolution処理610と、pooling処理611と、convolution処理612とが設定されていることを示している。各処理の具体的内容は参考文献3と同様であるためここでは省略する。convolution処理610では畳み込みフィルタを用いたデータ処理が実行され、pooling処理611では例えばmax poolingであれば局所的な最大値を出力する処理が行われる。

10

【0035】

また、図6では、convolution層およびpooling層には複数の特徴マップが存在し、入力層の画像上のピクセルに対応する位置には、複数のニューロンが存在することを示している。例えば学習画像が3原色のRGB形式の画像である場合、RGBチャンネルに対応する3つのニューロンが存在することになる。また学習画像が撮像映像の動き情報を持つOptical Flow画像であれば、画像の横軸方向と縦軸方向とをそれぞれ表現する2種類のニューロンが存在することになる。

【0036】

また、複数の画像を同時に入力として用いる場合は、それら入力画像の数に対応する分だけ入力層のニューロンを増やすことで対応することが可能である。本実施形態では、標準的なRGB画像を対象とする例を示すものとする。図6には、学習時の各層における特徴マップ等のサイズ630~635を例示している。これらサイズ630~635において、(a, b, c, d)という記述のうち、aはデータの数に相当する次元であり、例えばNはminibatchのサイズ、Mは後述する代表点の数である。bは画像のチャンネルに対応する特徴次元の数、cは画像のy軸に対応する特徴次元の数、dは画像のx軸に対応する特徴次元の数と考えることができる。なお、ここでは上述の部分について特徴抽出器であると述べたが、CNNの性質として、CNN全体が特徴抽出器であるとみなすことも可能であり、必要であればそのように解釈してもよい。

20

【0037】

また図6では、代表点生成器621が、認識対象の対象特徴量層605と、代表特徴量層606とから構成されていることを示している。対象特徴量層605は、認識対象の特徴量を保持する層である。認識対象抽出処理613では、minibatchの1番目のデータが認識対象であるとのルールに基づき、Convolution2層の(N, 512, 32, 32)サイズの間の特徴量から、1番目の(1, 512, 32, 32)サイズの特徴量を抜き出す。そしてこの抜き出した特徴量が対象特徴量となされる。当然ながら、どのデータが認識対象であるかを判別可能であれば、minibatchの1番目に認識対象を必ず入れておくという操作は必要がなく、認識対象であることを示す印に基づいて、単に対象特徴量を取り出すという処理に置き換えてもよい。

30

【0038】

次に、学習部12は、代表点生成処理614によって、対象特徴量層の特徴量を用い、代表特徴量層606のM個の代表特徴量を生成する処理を行う。代表点生成処理614の具体例は後述する。

40

最後に、学習部12は、損失算出処理615によって、損失値607を得る。損失算出処理615の具体例は後述する。

以上が学習時のCNNの構成と動作の例である。なお、初期化の方法としては、公知の手法でCNNのパラメータを初期化することができる。学習率などのパラメータは、必要に応じて任意に決定することができる。

【0039】

次にS502に進むと、学習部12は、変数Iterationを0に初期化する。こ

50

ここで変数 `Iteration` は、`minibatch` を何回読み込んで `CNN` の更新を行ったか、という意味を持つ。必要であれば、変数 `Iteration` やエポック数に基づいて、学習率などの学習に必要なパラメータが変更されてもよい。例えば、安定的な学習のため、エポック数が 10 を越えたときに、学習率を $1/10$ 倍するといった処理を導入することができる。

【0040】

次に S503 に進むと、学習部 12 は、学習用の `Minibatch` を選択部 11 から受信する。

さらに S504 に進むと、学習部 12 は、`CNN` の学習時認識処理を実行する。具体的には、学習部 12 は、S5041 において階層的特徴抽出処理を行い、S5042 において代表点生成処理を行い、S5043 において損失算出処理を行う。これらの処理の概要は、図 4 の内容について説明した際に上述したが、詳細な処理内容については後述する。ここでは基本的に、`Convolution` 2 層の N 個の特徴量と、生成された M 個の代表点（代表特徴量）と、に基づいて損失を算出する。なお、ここでは特徴抽出処理の後に代表点生成処理を行い、その後に損失算出処理を行うという順番を例示したが、これ以外にも、後述するように繰り返し実行する場合など、他の様々な構成にすることも可能である。また例えば、特徴抽出器のみを学習し、それが終わった後に代表点生成器のみを学習するなど、排他的に学習を行ってもよい。特徴抽出器のみを学習する場合の損失算出処理については後述する。

【0041】

次に S505 に進むと、学習部 12 は、S5043 において算出した損失に基づいて、学習誤差を `CNN` に逆伝播する。本実施形態では、`CNN` の学習を行うための方法として、誤差逆伝播法と `SGD` とを組み合わせた方法を用いる。誤差逆伝播法と `SGD` とを組み合わせた方法は参考文献 3 に詳細に説明されているため、ここでは詳細な説明は行わないが、基本的には `Minibatch` を選択し、`CNN` のパラメータを逐次更新するという手順を繰り返すことを特徴とする。なお、誤差情報は認識処理のデータフローとは逆方向に伝播するのが一般的であり、ここでも特徴抽出器および代表点生成器などに対して誤差情報が伝播するものとする。

【0042】

次に S506 において、学習部 12 は、S505 で伝播した誤差情報を用いて、`CNN` のパラメータの更新を行う。例えば、学習部 12 は、変数 `Iteration` に 1 を加算する。

次に S507 において、学習部 12 は、`Minibatch` を全て学習に利用したか否かの判定を行う。そして、学習部 12 は、全ての `Minibatch` を学習に利用したと判定した場合には S508 の処理に進み、そうでなければ S503 の処理に戻る。

【0043】

S508 に進んだ場合、学習部 12 は、エポックに 1 を加算する。

次に S509 に進むと、学習部 12 は、予め設定した上限にエポックが達しているか否かに関する判定を行う。そして、学習部 12 は、上限に達していると判定した場合には `CNN` の学習を終了して S510 に進み、そうでなければ S503 に進む。なお、多くの場合、`NN` の学習停止条件は、エポック数の予め決定した上限値に達成したか否か、学習曲線の勾配などを用いて自動的に決めるか、それらのどちらかを採用する。本実施形態では、エポック数の上限値に達成したか否かを学習停止条件とする。例えば本実施形態では、エポック数の上限値は例として 20000 回とする。なお、エポック数が増えることによって `NN` の学習率を低下させる方法など、エポック数に基づく学習処理の工夫を導入してもよい。

【0044】

S510 に進むと、学習部 12 は、学習済みの `CNN` のモデルを学習データ記憶部 D1 に保存する。ここでは後段の認識処理の動作において認識装置が用いることを考慮し、学習時に使用したコンピュータのメモリ上に保存しておくことを想定しているが、例えばそ

10

20

30

40

50

の他の記憶部を用意して保存してもよいし、その他の保存手法を用いてもよい。

【0045】

なお、上述のCNNによる特徴抽出処理では、特徴学習を通して、損失関数に基づく損失値を低減するように学習が行われることになる。例えば、代表点と認識対象データとを近づけるように学習がなされる。これにより、精度よく認識処理を行うことができる。

【0046】

次に、代表点生成処理の具体的な処理内容について例を挙げて説明する。本実施形態の第一の例では、先述したようにM個の代表点は認識対象データに基づいて生成される。図6に示した代表点生成処理614の例では、 $(1, 512, 32, 32)$ サイズの特徴量を、 $(M, 512, 32, 32)$ サイズの特徴量に変換する処理が、代表点生成処理となる(特徴量のサイズは一例である)。

10

【0047】

代表点生成処理は、認識対象に基づいてM個の代表点を生成する処理であればよい。なお、上述の変換処理を適用することができる。なお必要に応じて、どのような代表点生成処理を採用することも可能である。ここでは、以下の代表点生成処理の例を挙げて説明する。

まず学習部12は、 $(1, 512, 32, 32)$ サイズの特徴量をM個分コピーし、 $(M, 512, 32, 32)$ サイズの特徴量を作り、次に、 $(M, 512, 32, 32)$ サイズの特徴量を $(1, 512 \cdot M, 32, 32)$ サイズに変形する。なお、“ \cdot ”はスカラー値同士の掛け算を表す。変形には、 $(1, 512, 32, 32)$ サイズのM個の特徴量を特徴量の二次元目でconcatenate処理してもよいし、その他の公知の手段を用いてもよい。ここで、concatenate処理とは、二つのデータや特徴量のある指定された次元において結合する処理である。

20

【0048】

次に学習部12は、 $(1, 512, 32, 32)$ サイズの特徴量から、 $(1, 512 \cdot M, 32, 32)$ サイズの特徴量への変換処理を行う。このとき変換前は、同じ特徴量がM個あるだけであるが、変換処理によって、異なるM個の特徴量が現れることを狙う。この処理には、公知の方法を用いることができるため、詳細な説明を省略する。例えば、特徴量に対する大きな非線形変換が必要ないのであれば、変換前の特徴量の周辺に1サイズの余白を作るとともに 3×3 の畳み込み関数を用いて変換処理を行えばよい。また前述の例では対象特徴量をM個コピーした上で変換処理する例を挙げたが、例えばDecconvolution(逆畳み込み関数)を用いた公知の方法を用いてもよい。つまり、Decconvolution処理により、 $(1, 512, 32, 32)$ サイズの対象特徴量を $(1, 512 \cdot M, 32, 32)$ サイズの特徴量に変換してもよい。また、その他の公知の逆畳み込み関数以外の方法を用いてもよい。

30

【0049】

最後に、学習部12は、変換した $(1, 512 \cdot M, 32, 32)$ サイズの特徴量を $(M, 512, 32, 32)$ サイズの特徴量に変形する処理を行う。ここで用いる変形処理は、例えば前述の方法で行ってもよい。なお、これらの例は参考にした代表点生成器の構成例であり、その他の公知の方法を用いて代表点生成器(の一部)として用いてもよい。

【0050】

40

以降で、今回以外の方法で代表点生成器を構成する例を後述する。なお、後述する第2の実施形態では、代表点生成に近年発達してきた生成モデル系の方法を用いる例を説明する。なお、Mは1であってもよいが、認識精度を向上させるために、複数の代表点を用いたほうがよいと考えられる。またNを学習データ全体の数としたとき、クラスタリングのような目的関数を得ることができる。

【0051】

次に、損失算出処理の具体的な処理内容について例を述べる。本実施形態の第一の例では、先述したようにM個の代表点が認識対象データを含むN個のデータからなるMinibatchと比較され、損失が算出される。このとき、どのように損失算出処理を行うかという点でバリエーションが存在するが、先述したように正常-異常の2クラス問題とし

50

て解く場合には、以下の式（１）に示す損失関数を用いることができる。

【 0 0 5 2 】

【 数 1 】

$$2class\ Loss := \begin{cases} \frac{1}{N-1} \sum_i^M \sum_{j=1}^{N-1} \|X_j - A_i\| - \sum_i^M \|X_0 - A_i\| & \text{if 認識対象のラベルが異常} \\ \frac{1}{N-1} \sum_i^M \sum_{j=0}^{N-1} \|X_j - A_i\| & \text{else} \end{cases} \quad \text{式 (1)}$$

【 0 0 5 3 】

10

ここで、式（１）のXはMini batchデータの特徴量、Aは代表特徴量である。特に、X₀は認識対象データとする。また、少なくとも認識対象以外のMini batchデータは正常とする。||・||は二つのデータ（特徴量）間の距離を求める関数であり、例えば二乗フロベニウスノルムなどが用いられる。式（１）は、代表特徴量のAが正常なMini batchデータ（認識対象以外は正常）を精度よく近似するほど小さい値となる。また、認識対象のラベルが異常である場合は、Aが認識対象データから離れているほど損失関数の値が小さくなることから分かる。すなわち、Aは正常な代表特徴量を表していると解釈ができる。

【 0 0 5 4 】

20

式（１）では全てMini batchデータの特徴量と代表特徴量との距離を算出しているが、認識対象近傍のデータが多峰である例を考慮する場合には、各代表特徴量に対して最近傍となるMini batchデータの特徴量との間の距離のみを用いてもよい。これにより、多峰性を考慮したマッチングが可能となる。その場合、最近傍は、例えば特徴空間上でk-NNを用いて見つけることができる。なお、各Mini batchデータの特徴量に対して最近傍となる代表特徴量との間の距離のみを用いてもよい。また、必要であれば、k=1ではなく任意の数としてk個との距離を計算してもよい。例えばk=Mとしたとき、損失関数は式（１）と同じになる。

【 0 0 5 5 】

30

なお、式（１）は、場合分けしたそれぞれで微分可能であり、先述した誤差逆伝播法の目的関数として用いることができる。また、例えば代表点の多様性を維持するために、ある代表点にすでに割りあつたMini batchデータがある場合には、他のMini batch中のデータを同じ代表点に割り当てないようにしてもよい。すなわち、N>Mの場合、M個のデータのみしかMini batch中で学習に寄与しないことになる。一方、N<Mの場合は、N個の代表点のみ学習に用いられることになる。すなわち、動的な代表点数になる。更に後述のマルチタスクラーニングにより、例えばMini batch中の各データがノイズデータ（外乱）であるか否かを判定し、そのノイズ度合いを基にMini batch中のデータを選択してもよい。この場合、Nは変動することになり、これも動的な代表点数になりうる。なお、一個しか割り当てない場合のように制約がきつすぎると考えられる場合は、少数のMini batch中のデータのみ代表点に割り当てることを許すようにしてもよい。

40

以上が、学習装置10の動作の説明である。各部のさらに詳細な説明は後述する。

【 0 0 5 6 】

次に、異常検知システムの検出時の動作に関わる構成に関する説明を行う。

図1の異常検知システムの認識装置20は、入力部21と、認識部22とを備えている。認識部22は、特徴抽出部221と、代表特徴量生成部222と、距離計算部223と、を備えている。

判定装置30は、異常判定部31を備えている。

端末装置40は、表示部41を備えている。なお、端末装置40としては、例えばPC（Personal Computer）のディスプレイやタブレットPC、スマートフォン、フューチャーフォン等が適用できる。

50

【 0 0 5 7 】

以下、異常検知システムの検出時の動作に関する説明を図 7 に基づいて行う。図 7 は、異常検知システムにおける異常検出動作のフローチャートである。

まず S 7 0 1 において、入力部 2 1 は、認識対象データ記憶部 D 2 から認識対象データ（検知対象のデータ）を受け取り、認識部 2 2 に送る。このとき認識対象データには、学習時と同様の前処理を施すとする。前処理では、平均画像を引くなどの処理を行い、画像反転処理などは行わなくてもよい。認識部 2 2 は、受け取った認識対象データを用いて、C N N の認識処理を行う。このとき用いられる C N N のモデルは、学習時に得られたモデルを用いるものとする。なお、もし必要であれば、異常検知システムの学習時に得られたモデルを用いず、その他のモデルを用いてもよい。本実施形態では前述のように学習したモデルを用いるとする。

10

【 0 0 5 8 】

図 6 を例に挙げて説明すると、検出時に認識対象が一つ与えられるのは、 $N = 1$ としたときの図 6 のようになる。具体例を図 8 に示す。なお図 8 において、図 6 と同様の処理内容のものは、図 6 と同じ符号を付与し、重複する説明は省略する。

図 8 において、サイズ 8 3 0 ~ 8 3 3 は、本実施形態において検出時に認識対象が一つだけ入力され、それが認識処理にかけられることを表している。また、図 8 の場合、損失算出処理はなく、対象特徴量と代表特徴量との距離を算出する距離算出処理 8 1 5 を用いて距離 8 0 7 を得る。

【 0 0 5 9 】

20

次に S 7 0 2 に進むと、判定装置 3 0 は、認識対象データに対する C N N の認識処理結果に対して、異常判定処理を行う。このとき現れる C N N の認識処理結果は、本実施形態においては、認識対象と生成された代表点との距離とする。

【 0 0 6 0 】

ここで、本実施形態の学習時において、代表点は認識対象近傍の正常なデータを表現するように学習されるため、認識対象が与えられた際に、認識対象に基づいて生成される代表点は認識対象近傍に存在すると予測される代表点となる。すなわち、認識対象と代表点との間の距離は、認識対象と認識対象近傍とに存在するであろう正常な点との距離であるといふことができる。そして、距離が大きいほど異常であると考えられるため、ここで得られた距離に基づいて異常検知を行うことができる。このとき、異常判定を行う際の閾値は、ここでは予め決定した値を用いるとする。判定装置 3 0 は、距離が閾値以内である場合には正常と判定し、閾値を超える場合には異常として検知する。閾値は、操作者が決定してもよいし、機械的に決めてもよいし、その他の学習方法などによって決めてもよいとする。

30

【 0 0 6 1 】

次に S 7 0 3 に進むと、端末装置 4 0 は、判定装置 3 0 から異常判定結果を受け取り、その判定結果に基づいた表示処理を行う。例えば、表示部 4 1 は、検出対象が異常であることを示す異常判定結果が送られてきた場合、それを受け取り、警告処理を行ってもよい。このとき、端末装置 4 0 の備える機能に応じて、警告処理を行ってもよい。例えばランプとサイレンが備わっている場合はランプ等の点滅と共に警告音を鳴らしてもよく、映像確認用のディスプレイが備わっている場合は、通常の監視映像に対して異常な場所を強調表示してもよい。ただしこのとき、異常な場所を強調表示するには、映像上の異常な箇所を特定する必要がある。そのためには、例えば入力される検出対象データに画面座標データを付与しておき、異常判定結果に応じてその座標データを利用してよい。

40

【 0 0 6 2 】

次に S 7 0 4 に進むと、認識装置 2 0 は、認識対象データ記憶部 D 2 に検出対象となるデータが残っている場合には S 7 0 1 に戻り、そうでなければ、検出時の動作を終了する。

【 0 0 6 3 】

前述した説明では、2 クラス問題として解く場合の例として式 (1) を示したが、本実施形態は 2 クラス問題以外の問題にも適用可能である。例えば、1 クラス問題（異常検知

50

の文脈では正常学習とも呼ばれる)においては、例えば以下の式(2)に示す損失関数によって損失を計算し、学習に用いることができる。

【0064】

【数2】

$$1Class\ Loss := \frac{1}{N} \sum_i^M \sum_{j=0}^{N-1} \|X_j - A_i\| \quad \text{式(2)}$$

【0065】

式(2)において、 X は正常なデータのみを含むMiniBatchデータ(の特徴量)である。なお、前述した式(1)の場合と同様に、全ての代表点との比較を行わずに、近傍の代表点との距離だけを求めてもよい。正常学習は、学習時には正常なデータしか用いられない場合に特に有効であり、監視カメラに撮像された人物の行動を分析する場合や、工場等における製品の外観検査等にも用いることができる。このような場合は、応用のユースケースによっては、CNNなど多くの処理能力を必要とする特徴抽出器以外の特徴抽出器が望まれる場合がある。そのような場合は、CNNではなく、その他の公知の特徴抽出器等を用いてもよい。特徴学習が困難な場合は、代表点生成器の学習のみを行い、特徴学習を行わないようにしてもよい。

【0066】

また、監視用途の異常行動検知においては、継続して映像を収集することを通して、今まで存在しなかった異常データが得られる場合がある。その際、追加学習を行い、教師データとして用いてもよい。すなわち最初は式(2)に基づいて学習を行い、追加的に式(1)に基づいて学習を行ってもよい。なお、追加学習においては、上述の問題設定以外でも、新たに学習データが取得された都度行ってもよい。その場合、CNN全体を再学習してもよいし、構成の一部(例えば特徴抽出器や代表点生成器)のみを再学習するようにしてもよい。例えば、監視カメラを用いたユースケースにおいて、特徴学習を監視カメラごとに行うのは計算コスト・学習コストが大きい。しかし、もし代表点生成器の学習コストが比較的低い場合は、特徴抽出器を複数のカメラで共通に利用して、代表点生成器のみをカメラごとに学習することができる。ここで用いる特徴抽出器は、他のドメインで予め学習済みのものを用いてよい。

【0067】

なお、1クラス問題または2クラス問題として解く場合の例を上述したが、さらに多いクラス数の問題としても解くことが可能である。例えば、画像認識等の分類問題にも本発明を適用することができる。具体的には、 M 個の代表点に対して、例えば排他的なラベルを最大 M 種類付与することができ、以下の式(3)に示すような損失関数を用いることが可能である。

【0068】

【数3】

$$Multi-Class\ Loss := \frac{1}{|i \in C_j|} \sum_i^M \sum_{j=0}^{N-1} \|X_j - A_i\|_{i \in C_j} - \frac{1}{|i \notin C_j|} \sum_i^M \sum_{j=0}^{N-1} \|X_j - A_i\|_{i \notin C_j} \quad \text{式(3)}$$

【0069】

式(3)において、 $\|\cdot\|$ の右下の項は距離を計算する条件について示しており、 C は認識対象データのクラスを表す。また、 $i \in C_j$ は"MiniBatch中に含まれる j 番目のデータのクラスに i 番目の代表点のラベルが対応する場合"という条件を示し、 $i \notin C_j$ は逆を示す。 $|\cdot|$ は集合の要素を数え上げたときの数を表し、分母として与えることで総和された値を正規化する役割を持つ。例えば、 $|i \in C_j|$ は"MiniBatch中に含まれる j 番目のデータのクラスに i 番目の代表点のラベルが対応する場合の数"であり、

何回条件に当てはまったかを数え上げた数に相当する。上述の対応関係は予め決める必要がある。例えば、 $j = \{0, 1, \dots, 30\}$ のとき C_j は 0 番目のラベルであり、 $j = \{31, 32, \dots, 60\}$ のとき C_j は 1 番目のラベルであるなどである。すなわち、 X_0 に基づいて生成された各代表特徴量には予めラベルが付与されており、当該ラベルに予め紐付けられた認識対象のクラスのデータと近づいて生成されるように学習がなされると考えることができる。そして、検出時には、認識対象データに基づいて生成された代表特徴量と対象特徴量との距離を計算し、最も対象特徴量との距離が近い代表特徴量のラベルに対応する認識対象のクラスが、対象特徴量のクラスと考えることができる。なお、上述の場合は代表特徴量のラベルが排他的である場合について述べたが、排他的ではなく、例えばマルチラベルのような構成にしてもよい。

10

【0070】

なお上述の例では、代表点に基づいて損失を計算して学習処理を行う場合や、代表点に基づいて検出処理を行う場合を示したが、それ以外のタスクをマルチタスクラーニングとして同時に解いてもよい。例えば、認識対象に不審度などの値が教師データとして付与されている場合、代表点の生成と並行して、不審度の推定（回帰）を行うことができる。回帰の学習には、例えば二乗誤差を用いることができる。代表点の生成と回帰とを CNN で行う場合には、例えば、図 6 の Convolution 2 層に基づいて回帰処理を行う CNN を別途用意し、End-to-end で学習処理を行ってもよい。そして、その結果を検出時に用いてもよい。あるいは、生成した代表点に基づいて回帰などの処理を行ってもよい。図 6 の例では、 $(M, 512, 32, 32)$ サイズの代表点と、Convolution 2 層の $(N, 512, 32, 32)$ サイズの Minibatch から選択した $(1, 512, 32, 32)$ サイズの認識対象データとを concatenate 処理する。そして、concatenate 処理による $(M+1, 512, 32, 32)$ サイズの特徴量を、 $(1, 512 \cdot (M+1), 32, 32)$ サイズの特徴量に変形処理し、認識処理に用いる。これによって、Minibatch 中のデータを考慮して認識処理を行うことができる。なお、通常のニューラルネットワークでは、Minibatch 中の他のデータが、あるデータの特徴抽出処理に対して関与することは行われず、一方、この抽出処理等によって Minibatch 中のデータを特徴次元上に移す処理を通すことで、Minibatch 中のデータを考慮した特徴抽出を行うことができるようになる。また、マルチタスクラーニングにおいては、それぞれのタスクの学習をどのような順番で行ってもよいし、同時に学習を行ってもよい。

20

30

【0071】

また上述の特徴量のサイズの表現（例えば $(N, 512, 32, 32)$ ）は、あくまでも特徴量のサイズを例示するために使用された表現であって、計算機内部のデータとしてどのように特徴量を保持していてもよい。例えば $(N, 512, 32, 32)$ というサイズの特徴量は、内部的には $(1, 512 \cdot N, 32, 32)$ という配列のサイズであってもよい。上述の処理においては、特徴量の変形処理を繰り返し行う場合があるため、効率のよいデータ表現が望まれる。また、例えば $(N, 512, 32, 32)$ というサイズの特徴量として内部的に保持している場合であっても、 $(1, 512 \cdot N, 32, 32)$ のように変形処理せず扱うために、例えば畳み込み処理を第一の次元（データ数の次元）にまで拡張してもよい。つまり、通常の畳み込み処理の一例としては、第二の次元（チャンネル数の次元）数だけ畳み込みフィルタを用いる。この場合、 $(N, 512, 32, 32)$ の例では 512 個分の畳み込みフィルタを用いることになるが、第一の次元まで拡張された畳み込み処理は、 $(N, 512, 32, 32)$ の例では $512 \cdot N$ 個分の畳み込みフィルタを用いることになる。これは、 $(1, 512 \cdot N, 32, 32)$ のように変形処理したうえで畳み込み処理を行う場合と等価な表現であるため、変形処理の分だけ計算コストを削減しうる。

40

【0072】

また上述の例では、代表点は主に特徴空間上の点である場合について述べたが、例えば入力データの空間と同じ空間上の点であってもよい。具体的には、入力データが画像であ

50

る場合、画像空間上に代表点を生成して用いてもよい。このような場合、例えば参考文献4のように、Autoencoderを用いて画像空間上に認識対象データの再構成画像を生成し、代表点と比較する形で用いることができる。代表点の生成の仕方はさまざまなやり方があるが、CNNを例にした方法と同様の方法を用いて代表点を生成することができる。なお、ここではAutoencoderを用いて画像空間上で代表点を生成する場合の例について述べたが、Autoencoderの中間層で代表点を生成してもよい。その場合、上述したマルチタスクラーニングの形で、再構成誤差と代表点との距離とに関して最適化を行うようにしてもよい。なお、例えば元となる空間（例えば、画像空間とする）上で代表点を生成した場合、Minibatch中のデータを画像空間上に再構成したデータだけでなく、入力された学習データをも代表点生成器の学習に用いることが可能である。例えば、入力された学習データに近づけるように代表点を生成するための損失関数の項を追加してもよいし、その他の公知な方法等を応用して、入力された学習データと代表点との関係に関する損失関数を設計してもよい。また、もし入力された学習データに異常と正常などの教師情報が付与されているならば、例えば異常の学習データと生成された代表点との距離が大きくなるように学習を行ってもよい。具体的には、例えば、異常の学習データと生成された代表点との二乗誤差が大きくなるほど損失が大きくなる形で損失関数の項を追加することで上記の目的を達成することができる。

【0073】

参考文献4：C. Zhou et al. Anomaly Detection with Robust Deep Autoencoders. KDD, 2017.

【0074】

また、例えば入力データが画像データとテキストデータとである場合など、マルチソースな構成である場合に、一部の入力データの空間上にだけ代表点を生成してもよいし、全部の入力データの空間上に代表点を生成してもよい。もちろん、マルチソース構成でなくとも、一部の入力データの空間上にだけ代表点を生成してもよい。

【0075】

また、本実施形態では主として入力データの空間上で近傍探索を行い、近傍画像等を選択する場合の例について述べたが、任意の特徴空間上で近傍探索を行ってもよい。またさらに、学習ベースの方法を用いて近傍を選択してもよい。例えば、特徴学習とクラスタリングの学習を同時に行う方法を用いて、特徴抽出・クラスタリング・代表点生成の3者を学習することができる。例えば認識対象データを含むランダムに選択した100個のデータをまずMinibatchとして特徴抽出・クラスタリングする。そして、そのうち認識対象と同じクラスに割り当てられた50個のデータを認識対象近傍として特徴抽出・代表点生成に用いてもよい。特徴抽出・クラスタリング・代表点生成は、共通のCNNによって行うことができ、効率的に学習処理を行うことができる。例えば、図6のConvolution2層まで共通にCNNを利用し、その後の処理をクラスタリング・代表点生成とで別々の系統に分けてCNNを学習することができる。なお、クラスタリングを用いる上述の例はあくまでも一例であって、その他の近傍選択・探索が可能な公知の技術を用いてもよい。

【0076】

また本実施形態では、第一の例として、近傍探索を行い、近傍データを近い順から選択してMinibatchを作るための構成や動作の例を挙げたが、他の方法として、例えば認識対象近傍の一定の距離以内に存在するデータを集めてランダム選択をしてもよい。これにより、同じ認識対象に対して異なるMinibatchを構成することが可能となる。またその際に、学習のエポックやイテレーションが進むほど、認識対象近傍の選択する範囲を狭めるようにしてもよい。これによって、徐々にFine-grainedな特徴表現や代表点生成器を獲得しうる。なおこれらの近傍選択については、先述のように特徴空間上で行ってもよい。

【0077】

また、認識対象近傍のデータを選択する際に、近傍性のみならず、その他の基準を用い

10

20

30

40

50

てもよい。この基準の定義によって様々な性質の代表点生成器を学習することが可能である。例えば、認識対象近傍のデータとして、認識対象と同一の属性を持つデータを選択してもよい。具体的には、例えば人物照合等のタスクにおいて、認識対象と同一人物のデータを認識対象近傍のデータとして扱うことで、同一人物らしい代表点の生成を行うことが期待できる。この属性は、例えば物体種に関する属性などでもよい。

【0078】

その他にも、例えば監視映像を用いた行動認識などにおいて、監視映像中に撮像された同一グループと思われる人物らを同一属性として扱ってもよい。また、認識処理が誤りやすいデータをより多く近傍データに選択すれば、学習が早く進むことが期待できる。認識処理が誤りやすいとは、例えば誤分類が多いデータであったり、所望の代表点との距離が大きいデータであったり、必要であればどのような定義を用いてもよい。より多く選択されやすくするために、ある近傍の範囲内にある誤りやすいデータを必ず `MiniBatch` に含めたり、確率的に選択されやすくしたりすることができる。誤りやすいデータは、学習のエポックごとに見直してもよいし、イテレーションごとに見直してもよいし、端末装置40等を用いたユーザからのフィードバックに基づいて選択してもよい。

10

【0079】

また、間違えやすいという意味のみならず、各データについて重要度が分かっているならば、重み付け学習を導入してもよい。ここで重み付け学習とは、例えば損失関数において、代表点とデータとの距離などを求める際に、各データの重みに基づいてその大きさを変えることを意味している。例えば重要度が1から10の範囲である場合は、その範囲で決められた重要度と、例えば代表点とデータとの差を二乗した値とを掛ける処理を行って損失を計算してもよい。また、近傍データを選択する際に、さらにその他の基準を用いる場合の例として、様々な照明条件下で撮影したデータや、異なる視点・カメラ等で撮影されたデータを含むようにしてもよい。

20

【0080】

このようにすることで、変動に対してロバストな代表点生成器等を学習することが可能になる。このとき、例えば人物画像検索において、認識対象と同一の人物であって、様々な環境下で撮影した画像を認識対象近傍として選択してもよく、そのようにすれば、さらに認識処理のロバスト性が向上すると考えられる。

【0081】

なお、もし必要であれば、認識対象近傍データを選択するのではなく、その他の異なる基準を用いて `MiniBatch` を作成してもよい。その場合、近傍性を考慮した代表点生成ではなく、異なる性質を持つ代表点生成器を学習することができる。

30

【0082】

また、人物画像の照合等、1対1のマッチングに関するタスクにも本発明を用いることができる。まずここで例として、照合対象の人物画像が2枚与えられており、これら2枚の画像が同一人物の画像であるか否かを認識処理によって確認することを考える。そのために、例えば、参考文献5に示されるような `Siamese` 構造のCNNを用いるとする。

【0083】

参考文献5：E. Ahmed et al. An Improved Deep Learning Architecture for Person Re-Identification. CVPR, 2015.

40

【0084】

具体例を図9に示す。ここで示すのは学習時の構成とする。説明を簡単にするため、図9のうち、図6に共通する要素は図6と同じ符号を与えてあり、それらの説明は省略する。ここで用いる `MiniBatch` は先述の例と同様に、認識対象データと、認識対象データが与えられたもとでの認識対象近傍のデータとによって構成される。ここでの認識対象データとは、照合対象のうち1枚の人物画像である。近傍データとは、ここでは認識対象の人物が撮像された画像であり、認識対象データと類似するデータであるとする。このようにして構成された `MiniBatch` が各認識器に対して一つ与えられる。なお、学習を効率的に行うために、各 `MiniBatch` に存在するデータはかぶりなく得られる

50

ように排他性を考慮してもよい。図 9 は、認識器 A 9 0 1 と、認識器 B 9 0 2 とを含み、これらは先述の例において与えられた 2 枚の人物画像それぞれを処理する認識器を表している。2 つの認識器の出力を、損失算出処理 9 0 3 によって処理することで、損失値が得られることとなる。各認識器の出力は、各認識器に対する M i n i b a t c h 中のデータの特徴量に抽出したものと、生成された代表点とである。すなわち、S i a m e s e 構成においては、M i n i b a t c h 中のデータの特徴量に抽出したものと、生成された代表点とが、それぞれ 2 つ得られることとなる。1 対 1 のマッチングを行うための学習において、正例の画像のみ用いる場合の損失算出処理は、例えば以下の式 (4) に示す損失関数を用いることができる。

【 0 0 8 5 】

【数 4 】

$$Siamese Loss := \|X_0^A - X_0^B\| + \frac{1}{N-1} \sum_i^M \sum_{j=1}^{N-1} (a_1 \|X_j^A - A_i^A\| + a_2 \|X_j^A - A_i^B\| + a_3 \|X_j^B - A_i^A\| + a_4 \|X_j^B - A_i^B\|) \quad \text{式 (4)}$$

【 0 0 8 6 】

式 (4) において、正例の画像のみ用いるとは、学習に用いる画像のペアが、必ず同一人物であることを示している。ここで、 X^A 、 X^B はそれぞれ認識器 A、認識器 B の M i n i b a t c h データの特徴量 (画像空間上のデータでもよい)、 A^A 、 A^B はそれぞれ認識器 A、認識器 B の代表点である。特に、 X_0^A 、 X_0^B はそれぞれ認識器 A、認識器 B の認識対象データとする。 a_1 、 a_2 、 a_3 、 a_4 は、それぞれの項に対する重みであり、ここでは事前に決めるものとする。式 (4) は、式 (2) と同様に O n e - c l a s s 学習を目的にした損失関数である。ただし、異なる認識器をまたいで、代表点と M i n i b a t c h データの特徴量との距離を考慮している点が異なる。基本的には、それぞれの距離が近いほど損失関数の出力値は小さくなり、そのバランスは式 (4) の a_1 、 a_2 、 a_3 、 a_4 によって決定できる、O n e - c l a s s 学習を目的とした式である。なお、O n e - c l a s s 学習ではなく、その他の目的関数に基づいて学習を行いたい場合は、それぞれの目的に合った関数を用いてもよい。例えば、M u l t i - c l a s s 学習では、本実施形態において先述した方法を用いて、ラベルの異なる代表点と M i n i b a t c h データの特徴量との距離を離し、同じラベル同士は近づけるための損失関数を用いてもよい。具体的には、同一人物か否かを正例・負例について T w o - c l a s s 学習する場合は、以下の式 (5) に示す損失関数を用いて学習を行うことができる。

【 0 0 8 7 】

【数 5 】

$$Siamese Loss := \begin{cases} \|X_0^A - X_0^B\| + \frac{1}{N-1} \sum_i^M \sum_{j=1}^{N-1} (a_1 \|X_j^A - A_i^A\| + a_2 \|X_j^A - A_i^B\| + a_3 \|X_j^B - A_i^A\| + a_4 \|X_j^B - A_i^B\|) & \text{if 正例} \\ -\|X_0^A - X_0^B\| + \frac{1}{N-1} \sum_i^M \sum_{j=1}^{N-1} (a_1 \|X_j^A - A_i^A\| - a_2 \|X_j^A - A_i^B\| - a_3 \|X_j^B - A_i^A\| + a_4 \|X_j^B - A_i^B\|) & \text{else} \end{cases} \quad \text{式 (5)}$$

【 0 0 8 8 】

式 (5) において、右辺上段の式は式 (4) に一致しており、右辺下段は、負例 (画像ペアは異なる人物の場合) の場合の損失関数について示している。負例の場合は、異なる認識器から得られる対象特徴量や代表点等が、より離れるように学習されるように、一部の符号が反転している。また、今回の例では M i n i b a t c h 中のデータのうち、認識対象データとその他のデータは必ず同一人物であるという前提であったが、そうではなく、異なる人物 (それも、非常に似通った人物) にすることもできる。その場合は、異なる人物を表す代表点と、同じ人物を表す代表点とで、異なるラベルの代表点を生成すればよ

10

20

30

40

50

い。以上では、学習時の場合について例示したが、検出時においても距離ベースの認識処理を行うことが可能である。具体的には、例えば、以下の式(6)によって生成された代表点を考慮した距離を計算することができる。

【0089】

【数6】

$$1 \text{ 対 } 1 \text{ 距離測定関数} := \|X_0^A - X_0^B\| + \frac{1}{M} \sum_i^M (a_2 \|X_0^A - A_i^B\| + a_3 \|X_0^B - A_i^A\|) \quad \text{式(6)}$$

10

【0090】

この式(6)による距離と、予め決められた閾値とに基づいて、同一人物が否かを定めることができる。なお、式(6)のように対象特徴量および代表点を総合的に考慮した値ではなく、異なる認識器が出力した対象特徴量および代表点を比べたときに、最も距離の近い組の距離を採用してもよい。例えば、認識器Aが出力した対象特徴量と、認識器Bが出力した代表点との距離などを算出し、最も値の小さい距離値が用いられる。

【0091】

なお、人物画像の照合等に関して、1対1のマッチングを行う場合の例を上述したが、複数の認識対象候補が与えられたもとで、複数の検索候補から最もマッチしたデータを選択する場合も、ユースケースとしてありうる。例えば、スタジアムの観客席等での監視において、他の人物とは異なる行動をしている人物の特定を行いたいとする。具体的な例としては、例えば、参考文献6に「図3スポーツ競技場の観客席における撮影例 競技を観戦している人々の中で、赤丸で示した人物のみが時計やスマートフォンなど別の場所を終始見続けており、本研究では不審行動に分類される。本映像における顔の大きさは、縦方向に約25~30 pixelである。」と記されている。このように、他の人物と異なる行動をとっていることは、不審者を見つけるうえで監視員にとって重要なエビデンスと考えられる。

20

【0092】

参考文献6：黒沢健至 et al. 映像解析を用いた安全安心技術開発のための評価用映像データベースの構築. 第23回画像センシングシンポジウム, 2017.

30

【0093】

そこで、例として、ある人物が画面上の他の人物とは別の行動をしているか否かを判定することを考える。このような場合、複数の認識対象候補が与えられたもとで、複数の検索候補から最もマッチしたデータを選択するという問題設定として捉えることが可能である。以降では、説明を簡単にするために、認識対象候補と検索候補とが重複しており、複数存在する検索候補同士の対応について調べる場合について述べる。重複していない場合・重複が一部しかない場合については後述する。

【0094】

ここで、図10に複数認識対象を用いて複数の検索候補とのマッチングを行う場合の構成の例を示す(学習時)。図10の例において、構成の一部は図6の機能と同様の機能を持ち、その場合は図6と同一の符号を付与して、重複した説明は省略する。図10の変形処理1001は本実施形態で既に述べた変形処理であり、Minibatchのデータを全て第二の次元(特徴量の次元)に持っていく処理を行い、サイズ1002の特徴量を作成する。ここで、Nは可変であってもよい。この特徴量を、サイズ635の特徴量に変換するために、正規化可変代表点生成処理1003を用いる。

40

【0095】

正規化可変代表点生成処理1003の例を図10の左下に示す。ここでは(1, 512, 32, 32)サイズの特徴量がN個あると捉え、(1, 512, 32, 32)サイズの特徴量に対する代表点生成処理をN回繰り返す(N並列の処理を行ってもよい)。代表点生成処理は、本実施形態で説明した代表点生成処理を用いてよい。そして、N回繰り返さ

50

れた代表点生成処理は、足し合わされ、代表特徴量層 6 0 6 に入力される。このとき N 回足し合されるため、代表特徴量層 6 0 6 の入力の前に N で割って正規化処理を行うとする。これによって、任意の数 N は代表特徴量層 6 0 6 で消え、M 個の代表点の特徴量が代表点生成器 1 0 0 4 の出力として得られる。最後に、N 個の対象特徴量 (Convolution 2 層の特徴量) と、M 個の代表点を用いて、損失が計算される。このとき学習データが全て正常であるならば、1 c l a s s 学習を行うことができ、式 (2) を損失関数に用いることができる。正規化可変代表点生成処理 1 0 0 3 と、式 (2) とを用いて学習を行うことによって、N 個のデータの平均的な代表点を学習することが可能である。もし異常ラベルが与えられていれば、異常な認識対象データが代表点から離れるほど損失が減少するような損失関数を与えて学習することが可能である。

10

【0096】

また、検出時には、図 10 の構成において損失算出処理 6 1 5 の代わりに距離計算処理を導入し、検出時に得られた N 個の認識対象データと、それらから生成された M 個の代表点との距離を計算した結果に基づいて、異常判定を行ってよい。本来であれば、N 個の認識対象が存在する場合、それらの距離を計算するには、N の自乗のオーダーでの計算時間増加が見込まれる。図 10 の方法によれば、それを N・M のオーダーの計算時間に圧縮することができる。M が N よりも大きい場合については、例えば N 個の認識対象データをそれぞれ 1 対 1 比較してもよいし、それぞれの認識対象データに基づいて生成した代表点を本実施形態に記載の方法によって比較することで距離を算出してもよい。なお、上述したように、認識対象と検索候補が重複していない場合や、重複が一部しかない場合について

20

【0097】

さらに発展的には、複数個の検索候補を、例えば徐々に絞っていくという場合も考えられる。上述の複数認識対象の例では、認識対象数 N は可変であってよいので、これを用いて徐々に検索候補を絞っていくような使い方を学習時や検出時に行ってもよいし、その他の公知の方法を用いて、検索候補を絞っていくための処理を行ってもよい。複数認識対象の例を用いる場合は、例えば、生成された代表点から近い検索候補を除外していくことで、検索候補を絞っていくための処理を行うことができる。そして、絞られた検索候補に基づいて、再び代表点を生成する処理を行う。

30

なお、より詳しく認識処理を行いたい場合は、式 (2) のように代表点とデータとの間の距離を計算するのみならず、データ同士、代表点同士の距離を計算し、学習時の学習処理に用いてもよいし、検出時に用いてもよい。

【0098】

上述の例では認識対象データに基づいて代表点を生成する例を示したが、これは具体例として CNN を用いる場合の一例であって、その他の公知の方法を用いて代表点を利用してもよい。例えば、すでに生成された代表点をリカレントに用いて、認識処理に再利用してもよい。

【0099】

また上述の例では、画像を用いた認識処理の例を主に記したが、他のデータを対象とした認識処理を行ってもよい。例えば、映像データや、センサデータや、反射光スペクトルデータや、物質組成データや、化学的データなどである。例えば、化合物の新種を発見するために、化合物の正常データのみを学習し、認識対象として新種の化合物が現れたとき、距離ベースの検知によって新種の判定を行うことができる。

40

【0100】

また、上述の例では実データを用いて学習処理を行う場合の例について述べたが、CG (コンピュータグラフィックス) を用いてデータを作成し、学習処理等に用いてもよい。例えば、認識対象データに類似する CG データを作成し、その CG データに光源の変化等で生じるバリエーションを付与した CG データを作成し、それらの CG データや該 CG データから抽出した特徴データを、認識対象近傍データや認識対象データとして用いる。

50

これによって、画像上の様々なバリエーションに対応することができる。

【0101】

なお、上述の例では、光源等の変化に対応するため（例えば異なる光源の画像が与えられても正しく認識処理を行うため）に、学習画像等そのものや近傍の選び方等に工夫を加える例を示した。これに対し、さらに代表点に関しても光源等の変化に対応して認識処理に工夫を加えることで、よりロバストな認識処理を行うことができる。ここでは外観検査における具体例を以下に示す。例えば、本実施形態で示したAutoencoderを用いるとする。なおAutoencoderに畳み込みフィルタを導入したConvolutional Autoencoderでも、その他の再帰的構造を取り入れたモデルでもよい。

10

【0102】

まず、検査する対象の製品の認識対象データを選択する。説明を簡単にするために、ここで用いる認識対象データは一つであるとするが、複数あってもよい。

次に、ここで選んだ認識対象データと同一の製品のデータであって、認識対象データとは異なる光源等のバリエーションが存在するデータを選択する。単純にAutoencoderを適用するならば、Autoencoderは独立にそれぞれのデータを再構成するための学習に用いられるが、先述のようにAutoencoderの中間層特徴量に基づいて、代表点を生成することを考える。

【0103】

学習時の具体例を図11に示す。図11の例において、構成の一部は図6の機能と同様の機能を持ち、その場合は図6と同一の符号を付与してそれらの説明は省略する。

20

補助データ選択器1141は、認識対象データ1101に基づいて補助データを選ぶとする。ここで、補助データとは、本実施形態で認識対象近傍データと述べてきたデータに相当する。今回は"選んだ認識対象データと同一の製品のデータであって、認識対象データとは異なる光源等のバリエーションが存在するデータを選択する"ようにしている。これらは特徴空間上等での認識対象近傍とは限らないため、ここでは補助データと呼称することにする。

【0104】

MinibatchサイズをNとしたとき、N-1個の補助データ（補助データ1102～110N）を選択する。補助データは、様々な基準で選ぶことは可能であるが、ここでは"選んだ認識対象データと同一の製品のラベルがついているデータであって、認識対象データとは異なる露光で撮像したデータをランダムに選択する"とする。なお、異なる露光で撮像したデータのみを補助データとしたが、認識対象データと同一の露光で撮像したデータを補助データとしてまじえることで、学習処理を安定化させる試みを行ってもよい。また、補助データは人間が選択してもよい。

30

【0105】

再構成関数1120は、Autoencoderのことであり、再構成関数1120によって再構成認識対象データ1111および再構成補助データ1112～111Nが得られる。ここで、通常のAutoencoderと同様に再構成誤差を最小化するとともに、代表点生成の学習を進めることを考える。認識対象データ1101から再構成認識対象データ1111を得るまでのAutoencoderの中間層において、Autoencoderの中間層特徴量を用いて代表点生成処理614が行われ、代表点1131を得るとする。

40

【0106】

ここで、代表点生成処理614で用いるAutoencoderの中間層は、どの層であってもよく、複数の層の特徴量を用いてもよい。代表点生成処理614で用いるAutoencoderの中間層と同じ層の特徴量を用いて、特徴量抽出処理1121が行われ、特徴量1132～113Nが得られる。ここで特徴量抽出処理1121は、代表点との比較等に用いる補助データの特徴量を抽出するための処理であり、Autoencoderの中間層特徴量をそのまま用いてもよいし、その他の変換処理を行ってもよい。ここで

50

得た特徴量と代表点とを用いて、損失算出処理 6 1 5 が行われて、損失値 6 0 7 が得られる。このとき用いる損失関数は本実施形態で例示したどのようなものを用いてもよく、その他の公知の損失関数を応用したものを用いてもよい。例えば式 (2) によって `Unsupervised` な学習を行うことができる。なお学習データは全て正常であるという仮定をおく意味では、教師が存在する。

【 0 1 0 7 】

このとき、補助データは露光が異なるデータであるから、認識対象データから生成された代表点と、補助データの特徴量との位置関係は、露光の影響によって離れた場所に存在しうる。しかしながら、式 (2) に基づいて学習を進めることで、特徴抽出器および代表点生成器の学習が進み、生成された代表点と、補助データの特徴量との位置関係は、より近づいたものになりうる。これは、露光の影響を減じる効果であると捉えることができる。

10

【 0 1 0 8 】

さらに、式 (2) など代表点に関する損失関数以外に、`Autoencoder` ならではの再現誤差関数 (入力データと再現データとの二乗誤差など) を導入し、前述したマルチタスクラーニングとして解くことで、学習を効率的に進めることができる。また、`Autoencoder` のある中間層の特徴空間を、代表点を生成する特徴空間としてそのまま用いる場合、生成された代表点を用いて、`Autoencoder` の一部の関数によって画像を構成することが可能である。すなわち、生成された代表点が意味する画像を生成することができる。これは、`Autoencoder` の中間層で代表点を生成し、さらにその代表点に基づいて、画像空間上での代表点を生成したとして解釈することができる。つまり、段階的な代表点生成、あるいは階層的な代表点生成処理として解釈できる。

20

【 0 1 0 9 】

なお、例えば複数の層においてラベル付き代表点を生成した際に、それぞれの層で認識対象データが所属すると思われるクラスが異なるという結果であった場合は、例えば複数の層の結果を加味して所属するクラスを決定してもよい。例えば、それぞれの層で得られる平均の距離で正規化した距離を足し合せて、最も距離値の小さいクラスを所属クラスとしてもよい。また、ラベル付きでない場合でも、例えば最大または最小の距離を用いて異常検知等の検出処理を行ってもよい。また、各層において異なる閾値を用いて、それぞれの層で得られた代表点生成結果に基づいて独立に異常検知等の検出処理を行ってもよい。

【 0 1 1 0 】

30

また、上述の `Autoencoder` の例では、画像空間上に生成した代表点を端末装置 4 0 で可視化することで、監視員等に対して、検出結果の理解を促進してもよい。また、これらは代表点であるから、式 (2) 等によって、さらに損失関数を適用して学習を促進することや、画像空間上の代表点と入力データとの距離に基づいて、異常検知に用いることもできる。なお、ここでは学習データが全て正常であるという前提に基づいて説明を行ったが、異常データなどのラベルが存在する場合は、式 (1) などの損失関数を用いて学習し、得られた学習結果に基づいて、検出処理を行ってもよい。なお、`Autoencoder` はデータを再構成する場合の本実施形態における構成の一例であり、その他の公知の方法を用いてもよい。

【 0 1 1 1 】

40

上述の例では露光について述べたが、その他のカメラパラメータや、製品に関するデータ等を補助データとして用いてもよい。例えばカメラパラメータ A の場合、カメラパラメータ B の場合、カメラパラメータ C の場合、・・・、製品 A の場合、製品 B の場合、製品 C の場合、・・・、といったように、外部情報を補助情報として扱い、どのように入力処理を行ってもよい。

【 0 1 1 2 】

本実施形態における上述の例では、検出時に、同一の認識対象が入力された際に、同一の代表点が生成される仕組みであった。例えば監視映像の監視を目的とした際に、状況に応じて正常の定義が変化する場合がありうるため、状況に応じて生成される代表点が変わることで、正常の定義の変化を表現できることは望ましい。例えば、端末装置 4 0 に補

50

助情報入力器 1201 が備えられ、天気や備え付けられた店舗の状態（例えば営業中か否か）等の外部の情報が自動的に入力されるか、またはユーザによって与えられるとし、補助情報は状況の状態を与える情報であるとする。これに基づいて代表点生成器を切り替えてもよいし、また、補助情報を用いた代表点生成器を用いてもよい。

【0113】

以降では、補助情報を用いた代表点生成器の学習時の例について図12を参照しながら示す。なお、学習時の Mini batch は、補助情報に基づいて正常の定義が変化することを表現するために、補助情報ごとに構成する。例えば補助情報が0のときの正常データを集めた Mini batch と、補助情報が1のときの正常データを集めた Mini batch とを用意するとする。なお、これらの Mini batch 中のデータの選定には、本実施形態で述べた方法を用いてよい。図12の例において、構成の一部は図6の機能と同様の機能を持ち、その場合は図6と同一の符号を付与し、それらの説明は省略する。

10

【0114】

図12の補助情報入力器1201は、先述の補助情報を入力するものである。補助情報1202は入力された補助情報を示し、離散値・連続値、次元数を問わないが、ここでは説明を簡単にするために、一次元の{0, 1}の値をとりうるものとする。与えられた補助情報1202と、対象特徴量層605の特徴量とを、補助情報付き代表点生成処理1203によって変換し、代表特徴量層606の代表特徴量を得る。このとき代表特徴量は、補助情報に依存して決まる。これらが代表点生成器1204の役割である。

【0115】

20

補助情報付き代表点生成処理1203の動作に関する具体例を、補助情報の入力に基づく状況変化対応の例1210として、図12の左下枠に示す。図12中の枠の上部は補助情報が0であるときの動作、下部は補助情報が1であるときの動作について説明を行うための図である。Convolution 2層と代表特徴量層606は同一サイズの特徴空間であり、この特徴空間における代表点等の模式図について1220に示す。図中の黒色の点は対象特徴量1213であり、補助情報の値に関わらず、同じ位置に存在する。代表点群1211と代表点群1212は、補助情報の値に依存して、生成される点変動している。この動作が、状況の変化に応じて正常が変化することを表現している。このとき、各代表点群と対象特徴量との距離が、前述の二つの場合で変化しうするため、異常検知結果が変わりうる。このときの損失を、本実施形態で例示した損失関数や、その他の公知の手法を用いて算出して学習に用いてよい。

30

【0116】

補助情報付き代表点生成処理1203は、例えば以下のように構成してもよい。具体的には、対象特徴量層605の特徴量に対して、補助情報を concatenate し、その特徴量を図6に記載の代表点生成処理614に入力し、代表点を作成することができる。ここで、concatenateする補助情報は、one-hotなベクトルを、concatenateする特徴マップと同一サイズのマップに拡大してから、concatenateするものとする。具体的には、補助情報が1のスカラー値であったとしたら、1次元目が0で埋められ、2次元目が1で埋められたマップを concatenate する。このようにして、補助情報を考慮した代表点の生成を行うことができる。なお、その他の公知の方法を用いて、補助情報を考慮した代表点の生成を行ってもよい。

40

【0117】

これらの学習を行ったのち、検出時に、補助情報を考慮した代表点の生成を行うことで、状況の変化に応じた異常検知等に用いることができる。これらの方法を本実施形態で述べたマルチクラス学習に拡張することで、状況の変化に対応した多クラス分類問題等に適用することもできる。

【0118】

本実施形態で述べた学習処理において、特段最適化技法の制約について述べなかったが、必要に応じて制約や、正則化を追加してもよい。例えば、代表点や特徴量を、超球上に生成するように制約してもよい。そのようにすることにより、学習処理が安定化しうる。

50

また、その他の公知の方法を用いて制約や正則化を加えてもよい。また例えば、下記の式（７）に示す代表点類似正則化項を損失関数に追加することで、同じ代表点生成器から生成される代表点同士が似ないように学習を行うことができる。

【 0 1 1 9 】

【 数 7 】

$$\text{代表点類似正則化項} := -\frac{b}{(M-1)^2} \sum_i^M \sum_j^M \|A_i - A_j\|_{i \neq j} \quad \text{式 (7)}$$

10

なお、代表点同士が似すぎると、幅広い認識対象近傍を表現しづらくなる虞があるため、このような正則化項を追加してもよい。なお、式（７）のbは任意のパラメータである。

また、上述の例では、代表点それぞれを点として扱って、認識対象データ等との距離を測ったが、距離ではなく、例えば代表点群とMinibatch中のデータとの分布としての差異を測ってもよい。同様に、距離以外の公知の尺度を用いて、代表点等の差異を測ってもよい。

【 0 1 2 0 】

これまで代表点を生成することについて述べてきたが、点ではなく、その他のものを生成してもよい。例えば、質量をもった超球を生成してもよい。質量をもった超球は、これまで述べた代表点を生成するように生成し、ある半径の範囲の超球内に特徴量が充填されているとみなすことができる。ここで半径は、予め決定されるものとし、例えば1を与えてもよい。質量を持った超球（代表点）と認識対象データ等との距離は、認識対象データ等と、質量を持った超球とが交わる最も短い直線によって定義してもよい。

20

【 0 1 2 1 】

また、可視化のために、生成した代表点と認識対象データを低次元に写像し、それをマップとして表示してもよい。低次元に写像する方法は、どのような公知の技術を用いてもよく、例えば参考文献6に記載の方法を用いてもよい。

【 0 1 2 2 】

参考文献6：S. Wold et al. Principal component analysis. 1987.

【 0 1 2 3 】

また例えばマップ上に配置した代表点や認識対象データ等のうちの一つを、グラフィカルインタフェース等を用いて指定すると、近傍に存在する代表点や認識対象データ等が更に表示され、グラフィカルインタフェースを通した芋づる式の近傍探索が可能である。端末装置40は、そのようなインタフェースを備えてもよい。

30

その他にも、クロスバリデーションによって、最適なMを選択してもよい。また、代表点生成器は、CNNを用いず、例えば線形回帰等によって構成してもよい。CNNを用いる例は、高い非線形性を求められる場合に用いられる一例であり、本発明はCNNを用いる場合に限らない。

【 0 1 2 4 】

以上説明したように第1の実施形態では、認識対象データを用いて、データ中の認識対象人物等やその状態等の認識対象の代表点を特徴空間ないし元のデータ空間上に予測（生成）する代表点生成器を有し、認識対象データや代表点の比較を基に検出処理を行う。これにより、第1の実施形態によれば、高速な認識処理や、高い認識精度の認識処理が実現しうる。

40

【 0 1 2 5 】

< 第2の実施形態 >

第1の実施形態では、認識対象データを用いて、認識対象人物等やその状態等の認識対象の代表点を特徴空間ないし元のデータ空間上に予測（生成）し、認識対象データと代表点との比較に基づいて検出処理を行う例を示した。その際の構成の第一の例では、データの特徴量を抽出する特徴抽出器と、代表点を生成する代表点生成器とがあり、それらを畳

50

み込みニューラルネットワーク（CNN）で構成する例を述べた（CNNを使わない場合の例も述べた）。第2の実施形態では、代表点生成器を識別的に学習することに関して、第一の実施形態とは異なる構成および動作について説明する。

【0126】

第2の実施形態で示す構成は第1の実施形態で例示した構成と大部分が同一であり、一部の構成と動作が異なる。本実施形態において第1の実施形態と構成や動作が異なるのは、先述のように代表点生成器を学習するための構成および動作である。なお、本実施形態における式の表記は第1の実施形態と共通であるとする。

【0127】

本実施形態で、一つの構成例として、第1の実施形態の構成に加えて、代表点生成器を Generative Adversarial Network（GAN）のような敵対的生成ネットワークの一部を応用することによって学習する例を示す。GANは、参考文献7に開示されている方法であり、元のデータ空間上に偽データを生成する生成モデルと、生成された偽データの真贋を識別する識別モデルとがあり、敵対的に学習処理が行われることで、高度な生成器が学習されうる。本実施形態では、この識別モデルを用い、生成された代表点が認識対象データであるか、あるいは生成された代表点であるか（以降、真贋を識別するとする）、という識別を行うための学習を行い、代表点生成器は識別器をだますように学習を行う例を示す。なお、これらは識別器を用いて代表点生成器の精度向上を行う場合の一例であって、上述のようなGANの構成・動作に本発明は制限されない。なお、本実施形態で用いる代表点は、通常のGANとは異なり、元のデータ空間のみならず、特徴空間上での比較を想定していることに注意が必要である。

【0128】

参考文献7：I. J. Goodfellow et al. Generative Adversarial Networks. Arxiv, 2014.

【0129】

例えば、GANの目的関数を代表点の生成データに応用することで、以下の式（8）を得ることができる。

【0130】

【数8】

$$1class\ GAN\ Loss := \frac{c_1}{N \cdot M} \sum_i^M \sum_{j=0}^{N-1} \|X_j - A_i\| + \frac{c_2}{N} \sum_{j=0}^{N-1} \log D(X_j) + \frac{c_3}{M} \sum_i^M \log(1 - D(A_i)) \quad \text{式(8)}$$

【0131】

式（8）において、Dは真贋を識別する識別関数（識別器）である。第一項は式（2）と一致するMinibatch中のデータ（の特徴量）と生成された代表点との誤差を表す。第二項は識別器がMinibatch中のデータ（の特徴量）を誤って偽のデータであると識別してしまった誤差を表し、第三項は識別器が代表点を偽のデータであると正しく識別できたか否かの損失を表す。 $c_1 \sim c_3$ はそれぞれの項の重みであり、任意の数を設定してよい。このとき、代表点Aは代表点生成器である生成モデルGによって生成された点であり、認識対象データ X_0 を用いて生成されることから、 $A = G(X_0)$ と書くことができる。ここで、 $A = \{A_1, A_2, \dots, A_M\}$ である。

【0132】

なお、生成モデルGは第1の実施形態で述べた方法を用いて構成してもよいし、その他の公知の方法を用いて構成してもよい。このとき、敵対的学習を行うことを考えると、式（8）を用いて目的関数を例えば $\min_D \max_G 1class\ GAN\ Loss$ と書くことができる。式（8）は微分可能であるから、CNNを用いる場合、通常のCNNの学習と同様に、例えば誤差逆伝播法を用いて学習を行うことができる。また、式（8）は1class学習する場合として真贋を識別する識別器の誤差と式（2）とを組み合わせた例であって、式（2）ではなく、第1の実施形態に示したようななどのような損失関数を用いてもよい。

10

20

30

40

50

【0133】

このように学習された代表点生成器は、検出時において、より実際のデータに近くなった代表点を生成できる代表点生成器によって生成された代表点と、認識対象データと、を比較することで検出処理を行うこととなる。これにより、第2の実施形態では、認識精度が向上しうる。

【0134】

<第3の実施形態>

第1の実施形態では、Autoencoder等の入力データを再構成する構成を用いて、中間層と出力層（再構成後の層）とで、それぞれの層で代表点の生成を行う場合の例を示した。これは、CNN等を用いて複数階層の特徴抽出を行う場合に、複数の層で代表点を生成できることを示す例である。第3の実施形態では、代表点生成器を階層的な損失のもとで学習することに関して、第1の実施形態とは異なる構成および動作について説明する。特に、複数の認識対象が存在する場合として、入力されたデータは全て認識対象である場合について記述する。

【0135】

以降、第3の実施形態について、図面を参照して説明する。第3の実施形態の構成は、第1の実施形態で例示した構成と大部分が同一であり、一部の構成と動作が異なる。第3の実施形態において第1の実施形態と構成や動作が異なるのは、先述のように、代表点生成器を階層的に学習するための構成および動作である。なお、第3の実施形態における式の表記は第1の実施形態と共通であるとする。

【0136】

図13は、複数層における代表点生成と、階層的クラスタ損失に関する模式図（学習時）である。図13において、図6と同様の処理内容のものは、図6と同じ符号を付与し、重複する説明は省略する。図13では、左側に特徴抽出器および代表点生成器の例を示し、右側に代表点生成を行った空間上で階層的な損失計算を行う動作に関して例示している。代表点生成器1321～1323は、特徴抽出器620の特定の層から認識対象抽出処理613に基づいて認識対象データを取得し、代表点を生成する処理を行う。そして、代表点生成器1321～1323は、生成された代表点と、Minibatch中のデータの特徴量とを比較する損失算出処理1301～1303を行う。

【0137】

このとき、損失算出処理1301において、代表点生成器1321で生成された代表点と、Minibatch中のデータの特徴量との近傍探索を行い、例えば最も近い代表点に各データを割り当てるとする。この処理はクラスタリングと解釈することができる。近傍探索処理については、第1の実施形態で記載した方法を用いることができる。そして、ここで得られた割り当ての情報を、クラスタ1351として代表点生成器1322および損失算出処理1302に送る。また、第1の実施形態で記載した方法によって、損失値1311を得る。

【0138】

次に、損失算出処理1302において、代表点生成器1322で生成された代表点と、Minibatch中のデータの特徴量との近傍探索を行ってクラスタリングを行う。このとき、代表点生成器1322は、損失算出処理1301で行ったクラスタリングで得られたクラスタの情報をを用いて、クラスタごとに代表点を生成する。この処理に関しては後述する。また、ここで行うクラスタリングでは、クラスタ1351のクラスタをさらに分割する階層的なクラスタリングが行われる。すなわち、クラスタ1351の各クラスタに割り当てられたデータに関して、代表点を生成し、それぞれをクラスタリングすることで、二段目のクラスタの分割が行われる。このようにすることで、クラスタが段階的に微細になり、よりきめ細かい表現を獲得しうる。また、多段に何度も代表点を生成することで、一度に多数の代表点を生成する必要性が薄れるため、より効率的に学習を行える場合がある。そして、損失算出処理1302により損失値1312が得らる。このような操作を、代表点生成器1323、損失算出処理1303、と行っていく、多段に損失算出処理を

行えるようになる。

【 0 1 3 9 】

なお、前述の例では、特徴抽出器の全ての層で代表点を生成する例を示したが、任意の層を選択して代表点を生成する構成を用いてもよい。図 1 3 の右側には、上述した動作に関する特徴空間上の模式図を示している。図 1 3 の右上上段の図は、損失算出処理 1 3 0 1 で用いられるクラスタの模式図であり、Convolution 1 層 6 0 2 と同じサイズの特徴空間上に、対象特徴量 1 3 3 1、代表点 1 3 3 2、クラスタ 1 3 4 0 がある。ここで、図中の白丸は代表点、黒丸は対象特徴量、破線はクラスタの範囲の模式図を表現している。なお、第 1 の実施形態で述べたように、代表点を生成する先の空間は、元の特徴量の空間と同じでなくてもよい点に注意が必要である。つまり、Minibatch 中のデータの特征量も、代表点を生成する先の空間に写像すればよいためである。

10

【 0 1 4 0 】

ここで例として代表点は 2 点生成されており、複数の対象特徴量が近傍探索によって代表点に割り当てられている。このクラスタが、階層的クラスタリングにおける一段目のクラスタである。そして第 3 の実施形態においても第 1 の実施形態と同様に、このクラスタの割り当てを用いて、損失を計算してよい。例えば、以下の式 (9) を用いて損失を計算することができる。

【 0 1 4 1 】

【数 9】

$$Cluster\ Loss := \sum_i^M \frac{1}{|C'_i|} \sum_{j=0}^{N-1} \|X_j - A_i\|_{j \in C'_i} \quad \text{式 (9)}$$

20

【 0 1 4 2 】

式 (9) において、 C'_i は i 番目のクラスタを表し、 i は代表点に対応する。 $j \in C'_i$ は j 番目のデータが i 番目のクラスタに割り当てられているということを表現し、 $|C'_i|$ は i 番目のクラスタに割り当てられたデータの数を表す。また、各層での損失に関して重み付けを行ってもよい。

【 0 1 4 3 】

図 1 3 の右下の図は、損失算出処理 1 3 0 2 で用いられるクラスタの模式図であり、下層でのクラスタをさらに分割していることがわかる。

なお、以上のように得た複数の損失値は、第 1 の実施形態で述べたように全て学習に用いてもよいし、一部のみ学習に用いてもよい。

【 0 1 4 4 】

次に、クラスタごとに代表点を生成する方法について図 1 4 を参照しながら例示する。図 1 4 において、図 6、図 1 0 と同様の処理内容のものは、同じ符号を付与し、重複した説明は省略する。図 1 4 のクラスタ 1 4 4 0 は、前段でクラスタリングを行ったクラスタリング結果を示す。クラスタ 1 4 4 0 は、クラスタ情報入力 1 4 4 1 によって分割処理 1 4 4 2 の処理に用いられ、対象特徴量層 6 0 5 のデータをクラスタごとに分割する。仮にクラスタ数を L としたとき、対象特徴量層 1 4 0 5 には L 個に分割されたデータが入力される。

30

40

【 0 1 4 5 】

サイズ 1 4 0 1 ~ 1 4 0 L は分割されたデータ数を考慮した特徴量のサイズを示しており、データ数 $N_1 \sim N_L$ はそれぞれのクラスタのデータ数を表す。クラスタのデータ数は可変であるため、第 1 の実施形態で例示した正規化可変代表点生成処理 1 0 0 3 を用いて、各クラスタのデータについて、それぞれ M 個の代表点を生成することができる。 M 個は任意の数を設定でき、クラスタをさらに分割するためには、2 個以上の数を設定するとよい。上述した代表点生成器 1 4 1 0 によって、 $M \cdot L$ 個の代表点が生成される。最後に、これらの代表点を $(M \cdot L, 5 1 2, 3 2, 3 2)$ のサイズのように concatenate

50

e 処理し、損失の計算処理に用いることができる。

【 0 1 4 6 】

なお、前述した例は学習時の構成および動作について挙げているが、同様の代表点生成処理を用いて検出時に用いることができる。その際は、第 1 の実施形態で例示したように、認識対象データと代表点とを比較し、距離等を計算して、検出処理に用いてよい。その際、複数層での代表点生成結果を用いてよい。

【 0 1 4 7 】

また、上述の例は、いわば分割統治型のクラスタリングであるが、凝集型のクラスタリングを行うこともできる。その場合、クラスタを分割したのち、階層的にクラスタを統合していきながら、代表点の生成や損失の計算を行えばよく、上述の分割統治型と類似の処理を行えば実現可能であるため、詳細は省略する。

10

【 0 1 4 8 】

また、上述の例は、複数の認識対象データが存在する場合の例であった。認識対象データが単数であり、認識対象近傍のデータ等を学習に用いる場合にも、上述の方法を用いることができる。代表点の生成には単数の認識対象データを用い、分割統治型のクラスタリングを行う場合には、認識対象データの所属するクラスタのみを階層的にクラスタリングしていくこととなる。その際、認識対象データの所属するクラスタ以外に割り当てられた認識対象近傍のデータは、以降の認識処理を省略することができる。

【 0 1 4 9 】

以上説明したように、第 3 の実施形態によれば、階層的に多段階に学習された代表点生成器は、より精度よく学習処理が行われうるため、検出時の認識精度が向上しうる。

20

【 0 1 5 0 】

< 第 4 の実施形態 >

第 4 の実施形態では、時間などの系列を考慮した場合に、代表点生成や特徴抽出等に関してどのような工夫を導入することが可能であるかを述べる。第 4 の実施形態で示す構成は第 1 の実施形態で例示した構成と大部分が同一であり、一部の構成と動作が異なる。第 4 の実施形態において第 1 の実施形態と構成や動作が異なるのは、先述のように、時系列性を考慮した点である。

【 0 1 5 1 】

なお予め述べるが、第 1 ～ 第 3 の実施形態において説明を行ったいずれの例も、時系列性のあるデータに対して構成・動作しうる。例えば、第 2 の実施形態において、一枚の画像ではなく、映像や動きベクトル等を生成することも可能である。また、ある人物の画像を認識対象とするとき、別の時刻の同一人物の画像を認識対象近傍の画像として用いることも可能である。第 4 の実施形態で述べるのは、さらに時系列性を利用した場合の例である。

30

【 0 1 5 2 】

第 1 の実施形態では、例えば行動認識や不審行動検知において、人物画像を用いる場合の例を述べた。その他の例として、例えば参考文献 8 に開示された方法を用いて、R G B 画像のみならずオプティカルフロー画像を抽出し、認識処理に用いてもよい。

【 0 1 5 3 】

参考文献 8 : K. Simonyan et al. Two-Stream Convolutional Networks for Action Recognition in Videos. NIPS, 2014.

40

【 0 1 5 4 】

またその場合、各ストリームにおいて独立に代表点を生成して認識処理に用いてもよい。また、異常検知等に用いる場合、例えば以下の式 (1 0) を損失関数として用いてもよい。

【 0 1 5 5 】

【 数 1 0 】

50

$$\text{TwoStream Loss} := \sum_i^M \sum_{j=0}^{N-1} (d_1 \|X_j^T - A_i^S\| + d_2 \|X_j^S - A_i^T\|) \quad \text{式 (10)}$$

【0156】

式(10)において、 X^T 、 X^S はそれぞれ認識対象データに対応する時間的ストリーム(Temporal Stream)の特徴量、空間的ストリーム(Spatial Stream)の特徴量を表している。 A^T 、 A^S はそれぞれ認識対象データに対応する時間的ストリーム上の代表点、空間的ストリーム上の代表点を表している。 d^1 、 d^2 は重み付けのパラメータであり任意に設定してよい。式(10)は、空間的ストリームの対象特徴量から生成された代表点が、時間的ストリームの対象特徴量とどれくらい近いかを測り、時間的ストリームの対象特徴量から生成された代表点が、空間的ストリームの対象特徴量とどれくらい近いかを測る式になっている。そして、式(10)に基づいて正常学習されることを通して、一方のストリームの正常な特徴量から、一方の正常な特徴量を予測するための特徴抽出器・代表点生成器が学習されうる。

10

【0157】

検出時には、式(10)と同様にこれらの特徴量および代表点との距離を測り、もし距離が大きくなっていけば、正常学習を行った際の正常な関係が崩れたとみなし、異常であると判定することができる。なお、式(10)は特徴量と代表点との複数の組み合わせについて考慮するが、他の実施形態で示した例と同様に、最近傍の点のみを考慮するように変更してもよい。また、式(10)は時間的ストリームと空間的ストリームの入力データをそれぞれN個用いている例であるが、これらのデータは、映像の連続する区間において抽出したデータであってもよい。また、ある映像からランダムないし何らかの基準によって取り出したデータであってもよい。

20

【0158】

なお、第1の実施形態における1対1のマッチングのように、ある系列とある系列とのマッチングを行いたい場合は、系列長を階層的・多段階に複数パターン考慮して代表点を生成し、マッチングに用いてもよい。例えば第3の実施形態において、階層的に代表点を生成し、階層的に損失関数を計算する場合の例を述べたが、この際、多段階に系列長を考慮することが可能である。

30

【0159】

図15にその例を述べる。時間方向1501は、右にいくほど時間が進むことを表している。CNN特徴量1502のように、複数のタイミングでCNN特徴量が抽出されるとする。範囲1503は複数のCNN特徴量をまとめあげる範囲を示しており、動作1504は代表点を生成する動作を表している。すなわち、複数のCNN特徴量を認識対象データとして、代表点を生成する例を示す。また、さらにその上段でCNN特徴量をまとめあげて、認識対象データとして、代表点が生成されている。このようにして、多段階に代表点を生成することができる。図15では、上段と下段とで系列1、系列2の1対1マッチングの例を示しており、例えば上段の系列1における動作1504と下段の系列2における動作1505とは対応している。第4の実施形態では、このように対応する場所のそれぞれで、代表点および認識対象データを比較することを考えることができる。比較する方法は、前述の他の実施形態に記載のどのような方法を用いてもよい。また、図15に示した系列は時間的に持続しうるので、CNN特徴量をまとめあげる範囲を、例えば右に1つ移動することで、また新たに1対1のマッチングを行うことが可能である。このように学習した結果と、前述の他の実施形態で示した方法とを用いて、例えば行動認識や、類似行動検索、異常行動などの検出処理に用いることができるようになる。

40

【0160】

以上述べたように、第4の実施形態によれば、階層的・多段階に系列を考慮して学習された代表点生成器では、系列データに対する認識性能を向上することができる。

【0161】

50

本発明は、前述の実施形態の１以上の機能を実現するプログラムを、ネットワーク又は記憶媒体を介してシステム又は装置に供給し、そのシステム又は装置のコンピュータにおける１つ以上のプロセッサがプログラムを読み出し実行する処理でも実現可能である。また、１以上の機能を実現する回路（例えば、ＡＳＩＣ）によっても実現可能である。

【０１６２】

前述の実施形態は、何れも本発明を実施するにあたっての具体化の例を示したものに過ぎず、これらによって本発明の技術的範囲が限定的に解釈されてはならないものである。即ち、本発明は、その技術思想、又はその主要な特徴から逸脱することなく、様々な形で実施することができる。

【符号の説明】

10

【０１６３】

１０：学習装置、１１：選択部、１２：学習部、２０：認識装置、２１：入力部、２２：認識部、３０：判定装置、３１：異常判定部、４０：端末装置、４１：表示部

20

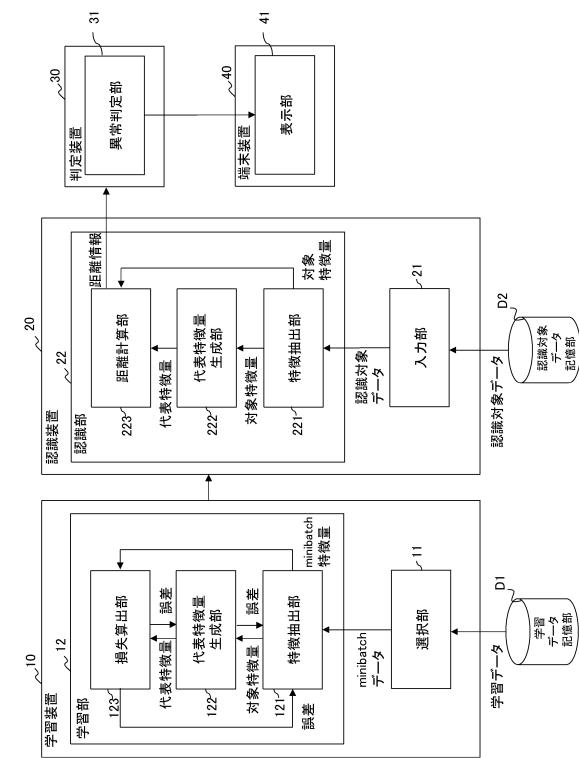
30

40

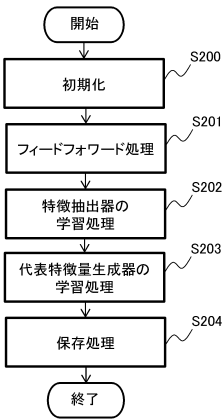
50

【図面】

【図 1】



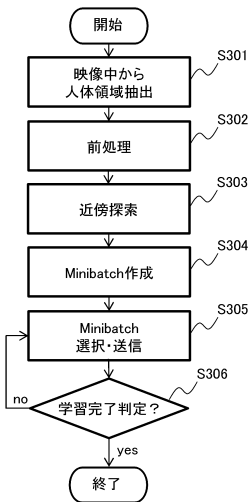
【図 2】



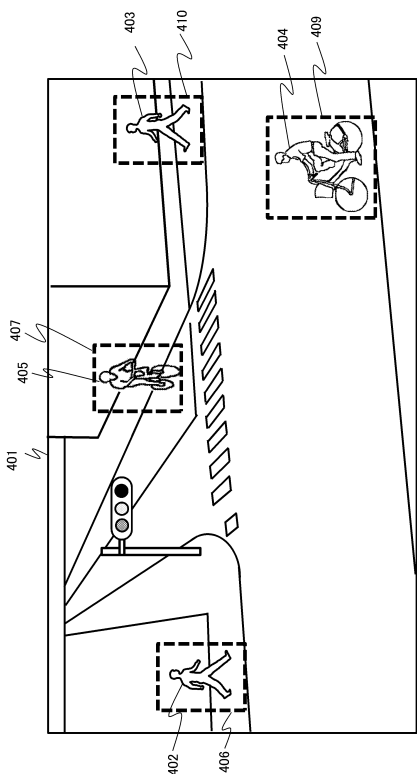
10

20

【図 3】



【図 4】

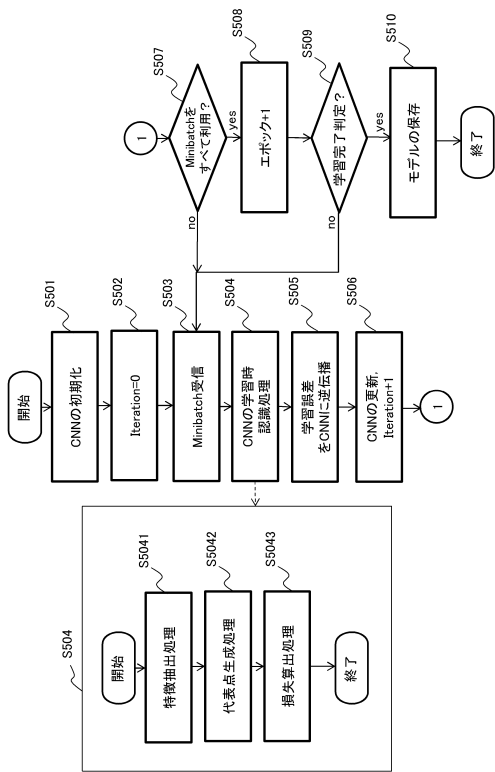


30

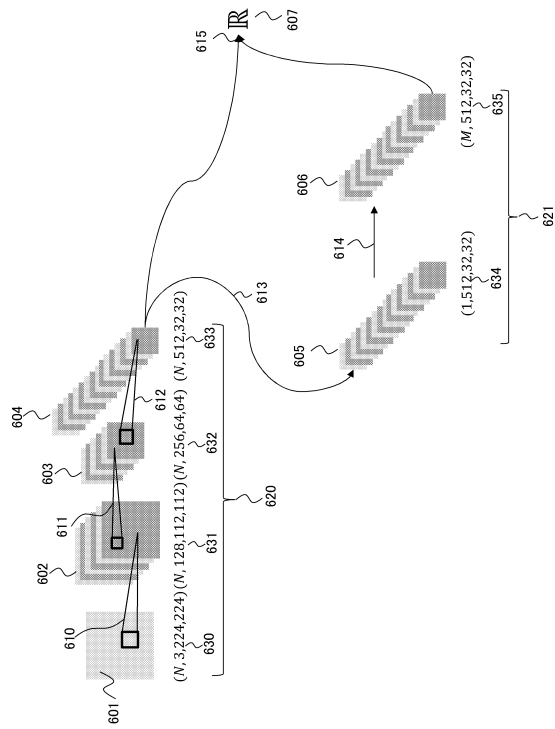
40

50

【図 5】



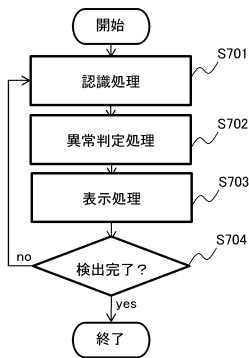
【図 6】



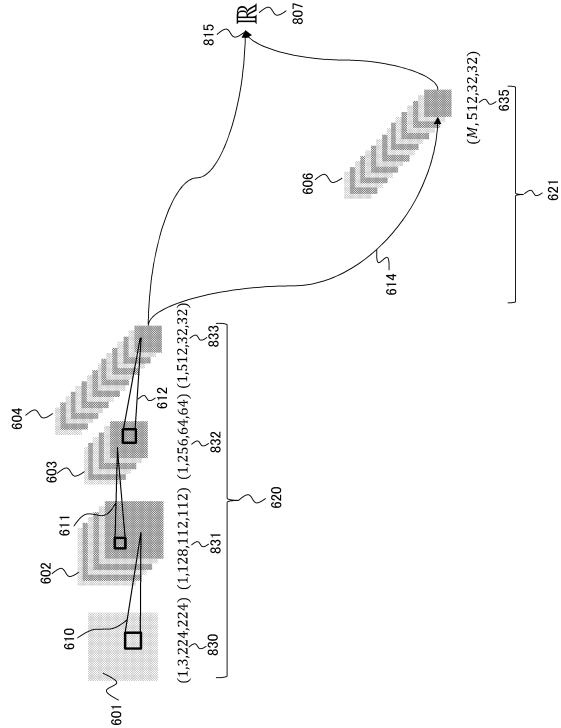
10

20

【図 7】



【図 8】



30

40

50

フロントページの続き

- (56)参考文献
- 特開 2 0 1 6 - 1 3 3 9 4 4 (J P , A)
- 特開 2 0 1 6 - 0 1 5 1 1 6 (J P , A)
- 特開 2 0 0 4 - 2 2 0 5 5 5 (J P , A)
- 国際公開第 2 0 1 6 / 1 1 7 3 5 8 (WO , A 1)
- 国際公開第 2 0 1 8 / 1 8 0 1 9 7 (WO , A 1)
- 倉西 佐知、外 2 名, “ 時系列画像に基づく異常検出のための学習画像の検討 ”, 映像情報メディア学会技術報告, 日本, (一社) 映像情報メディア学会, 2016年02月15日, Vol.40, No.6, pp.269-274
- 鍛冶 一祐、外 2 名, “ パケットのヘッダ情報に基づく不審な通信挙動の検知 ”, 情報処理学会研究報告, 日本, 情報処理学会, 2018年02月26日, Vol.2018-IOT-40, No.33, pp.1-7
- Dong-Won Shin et al., "Local Patch Descriptor Using Deep Convolutional Generative Adversarial Network for Loop Closure Detection in SLAM", 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 米国, IEEE, 2017年12月15日, pp.546-549
- Alexander Sage et al., "Logo Synthesis and Manipulation with Clustered Generative Adversarial Networks", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 米国, IEEE, 2018年06月18日, pp.5879-5888
- 関口 浩、外 2 名, “ 複数音声の分離再構成に向けた聴覚数理モデル導出型ラダーネットワーク ”, 電子情報通信学会技術研究報告, 日本, 一般社団法人電子情報通信学会, 2018年07月19日, Vol.118, No.160, pp.9-13
- 小向 航平、外 8 名, “ オートエンコーダによるデータ圧縮と分類器のラベル追加を両立したセンサネットワークモデル ”, マルチメディア, 分散, 協調とモバイル (D I C O M O 2 0 1 8) シンポジウム論文集, 日本, 一般社団法人情報処理学会, 2018年06月27日, Vol.2018, No.1, pp.1507-1516
- Masoud Abdi et al., "Discriminative Clustering of High-Dimensional Data Using Generative Modeling", 2018 IEEE 61st International Midwest Symposium on Circuits and Systems (MWSCAS), 米国, IEEE, 2018年08月05日, pp.799-802
- Shagan Sah et al., "Multimodal Reconstruction Using Vector Representation", 2018 25th IEEE International Conference on Image Processing (ICIP), 米国, IEEE, 2018年10月07日, pp.3763-3767
- Mennatallah Amer et al., "Nearest-Neighbor and Clustering based Anomaly Detection Algorithms for RapidMiner", Proceedings of the 3rd RapidMiner Community Meeting and Conference (RCOMM 2012), ドイツ, Shaker Verlag GmbH, 2012年08月28日, pp.1-12, <https://www.researchgate.net/publication/230856452>
- (58)調査した分野 (Int.Cl., D B 名)
- G 0 6 N 3 / 0 2 , 2 0 / 0 0 , 2 0 / 1 0
- G 0 6 T 1 / 0 0 , 7 / 0 0 - 7 / 9 0
- G 0 6 V 1 0 / 0 0 - 1 0 / 9 8