

令和 5 年 6 月 19 日現在

機関番号：13701

研究種目：基盤研究(C)（一般）

研究期間：2020～2022

課題番号：20K11884

研究課題名（和文）少数不良品サンプルを用いた敵対的自己符号化器による正常モデルの生成と異常判別法

研究課題名（英文）A method for for generating normal model and discriminating anomalies by adversarial autoencoder using a small number of defective samples

研究代表者

加藤 邦人 (Kato, Kunihito)

岐阜大学・工学部・教授

研究者番号：70283281

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：本研究では深層学習の異常検知手法である自己符号化器（AE）の汎化現象に注目し、AEの問題点を克服すべく、AAEにおいて潜在空間の分布を正規分布にし、さらに少量の異常データを学習に用いることで正常と異常の境界を明確に学習する手法を研究する。また、他の異常検知手法との比較で性質を明らかにし、AAEが異常モデルを正規分布に適応させることで識別境界に異常サンプルが集まる特性を、Deep SAD法が正常モデルを超球に配置し境界を明確にする特性を有することを明らかにした。この知見を基に、手法の改良を行い、実応用例としてはハンダ検査、航空外装板の検査で高い認識性能を達成した。

研究成果の学術的意義や社会的意義

本研究の学術的意義は、敵対的学習の一種であるAAEを用いた深層学習による異常検知手法の確立にある。正常サンプルと、少量の異常サンプルによる性能向上を明らかにし、既存のAutoEncoderと比較して異常検知性能が大幅に向上することを明らかにした。

社会的には、異常検知は検査やセキュリティ等への応用が可能であり、本研究はハンダの異常検知への具体的な応用を示した。社会で人手不足問題が深刻化する中、検査の自動化のニーズは高まっており、本手法はその解決に大きく寄与する。さらに、学習サンプルの準備のみで可能な学習手法は、従来の画像処理による検査システムより短時間で導入が可能となる。

研究成果の概要（英文）：This research focuses on the generalization phenomenon of Autoencoders (AE), a deep learning method for anomaly detection. To solve this problem of AE, we studied a method in Adversarial Autoencoders (AAE) where the latent space distribution is fit Gaussian distribution, and a small amount of anomaly samples is used in the training to clearly learn the discriminating boundary between normal and abnormal. Furthermore, We have also characterized our method of using AAE by comparing it to other anomaly detection methods. Based on these results, we improved our method and achieved high recognition performance in practical applications such as solder inspection and aircraft exterior panel inspection.

研究分野：コンピュータビジョン

キーワード：異常検知 敵対的自己符号化器 深層学習 外観検査

## 1. 研究開始当初の背景

異常検知技術は、工業製品の外観検査、機械の監視、セキュリティといった幅広い分野で重要な役割を果たしている。近年ではディープラーニングを用いた方法が多く研究されており、コンピュータビジョンの分野では、ディープラーニングを用いた画像の異常検知技術は外観検査の応用などに広く用いられ始めている。

しかし、ディープラーニングを用いた異常検知技術にはいくつかの問題点がある。まず一つ目の問題として、異常検知の応用が期待される場面では、正常な状態のサンプルは大量に入手可能だが、異常状態のサンプルはその出現頻度が極めて低いため、十分な量を集めることが難しい場合が多い。これは、データ駆動型のディープラーニングにとって大きな障害である。この問題を解決するためには、非常に少ない異常サンプルから有用な情報を抽出し、それを用いて異常を検出する方法が求められている。

二つ目の問題は、正常状態は一定範囲内での変動で収まるが、異常状態のパターンは多種多様で膨大であるという点である。変動が小さい正常状態と、パターンが多岐にわたる異常状態を区別することは、一般的な2クラス分類問題とは異なり、特に困難な課題となっている。これは、ディープラーニングが大量のデータによる学習を必要とする一方、異常サンプルを手に入れることが困難な問題とつながる。

少ない学習サンプルでニューラルネットワークを学習する方法については、既に多くの研究がなされている。しかし、それらの大半はデータ拡張による学習サンプルの水増しや自動生成であり、これらの手法もまた、パターンが多種多様である異常状態のサンプルを大量に取得するという問題の解決には至っていない。また、これらの手法は、訓練時に存在しなかった未知の異常に対する対応能力も保証していない。

そこで、新たなアプローチとして、正常状態の学習サンプルから、正常分布の確率密度を学習するという方法が注目を集めている。このアプローチでは、大量の正常状態のサンプルから正常状態の特性を把握し、その基準から大きく逸脱した状態を異常として検出するという考え方を基礎にしている。これにより、ディープラーニングの異常検知における前述の問題を解決しようという試みが進められている。

## 2. 研究の目的

本研究の目的は、大量の正常状態の学習データとわずかな異常状態の学習データから、ニューラルネットワークを効率的に訓練し、異常検知に最適な確率密度関数を生成することである。この目的を達成するために、図1に示されるような敵対的自己符号化器(Adversarial Auto Encoder, 以下AAE)<sup>[1]</sup>を採用する。AAEを用いると、各特徴を低次元の確率密度関数に圧縮することができ、標準正規分布に従った正常状態のモデルを生成することが可能となる。その際、少量の異常データを学習サンプルに付加することで、正規分布の中心に正常が、周辺に異常が分布する確率密度分布を得る。

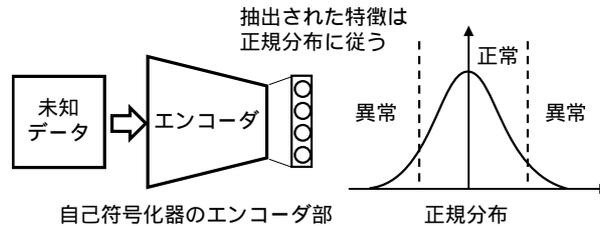


図1 異常検知方法

ディープラーニングにおける異常検知で一般的な方法は、自己符号化器(Auto Encoder: 以下AE)を用いる方法であるが、AEでは学習データに存在しないデータは、他のサンプルから確率密度関数内に内挿するように学習する「汎化」が起こることがあり、汎化が起こると未知の異常データは正しく異常と判定されるか、内挿され正常と判定されるか不確実となる問題がある。この問題は、潜在変数の確率分布を明示的に表現するVAEや、敵対的学習(Generative Adversarial Network: 以下GAN)を用いたAnoGAN<sup>[2]</sup>やEGBAD<sup>[3]</sup>等でも同様である。

そこで、正常データと少量の異常データを用い、敵対的学習により正規分布に従う確率密度関

数を推定することで、潜在空間上での正常と異常の識別境界が明確に学習できることを解明する。

### 3. 研究の方法

本研究では、まず正常サンプルと少量の異常サンプルにより敵対的自己符号化器(AAE)を用いて正規分布に従う正常モデルを学習し、異常を正確に検出できるようなネットワーク構造と学習法の研究を行う。これには、AAEが入力画像群の確率密度関数を学習し、ディスクリミネータがその特徴が正規分布になるようにAEに働くことを確認する。AAEは2つのネットワークが敵対して学習することで、正規分布の特徴抽出を実現するが、学習の安定性と推定される確率密度関数が適切な正常モデルとならない場合があるため、これを回避するネットワーク構造の改良と安定した学習方法の確立を行う。

### 4. 研究成果

#### 4.1 手法の概要

Adversarial AutoEncoder (AAE)は、AutoEncoderと敵対的学習であるGANを組み合わせて構成されている。通常のAutoEncoderではボトルネックにおいて入力データ分布の低次元特徴を表現する潜在空間にはどのような分布が写像されるか未知であるが、AAEではDiscriminatorネットワークを別に用意し用いることで、データ分布 $p_{data}(x)$ に対する潜在分布 $q(z)$ を任意の分布 $q(z|x)$ に近づけることができる。本研究では、Hotelling's T squareを潜在分布に適用するために、 $q(z)$ を標準正規分布に設定し学習を行う。ここで、 $q(z)$ は $q(z|x)$ を用いて式**エラー! 参照元が見つかりません**のように表すことができる。 $q(z|x)$ はEncoderにおける $x$ が入力されたときに $z$ を出力する条件付き確率であるが、AAEでは決定論的に一意に定まる出力のみ行うと仮定していることに留意する必要がある。

$$q(z) = \int_x p_{data}(x) q(z|x) dx \quad (1)$$

AAEにおけるDiscriminatorは、 $q(z)$ を $p_{data}(x)$ へ近づけたいため、 $p_{data}(x)$ を本物の分布、 $q(z)$ をGeneratorの生成分布と捉えて識別を行う。また、Discriminatorを騙すGeneratorの役割はEncoderが担っていると考えられる。

#### 4.2 AutoEncoderとの比較による性能評価

入力するデータを低次元な潜在空間に写像するために、AutoEncoderを用いる。異常検知手法でHotelling's T squareを適用するためには、対象とする分布が正規分布である必要があるため、AAEを用いることで入力データの低次元特徴を表現している潜在空間を標準正規分布に近づける。この潜在分布に対してHotelling's T squareを用いて特徴ベクトルの異常度を算出することで、正常・異常の分類を行う。学習データを正常データのみもしくは大量の正常データと極少数の異常データで構成して学習することで、潜在空間における分布は正常特徴によって構成される。異常データの特徴は正常特徴と異なるため、正規分布の確率密度が低い部分に集合し、Hotelling's T squareで異常度が高く算出される。AAEは図2に示すモデルを基礎として用いた。

データセットには、実際に現場で生産されたはんだ製品の検査画像を用いる。画像サイズは135x135であるが、128x128へクロップしてデータセットを構成する。正常データ・異常データの一例を図3に示す。図からわかるように、異常データには製品の形状が変形しているものや、異物が製品上に付着しているものが該当する。

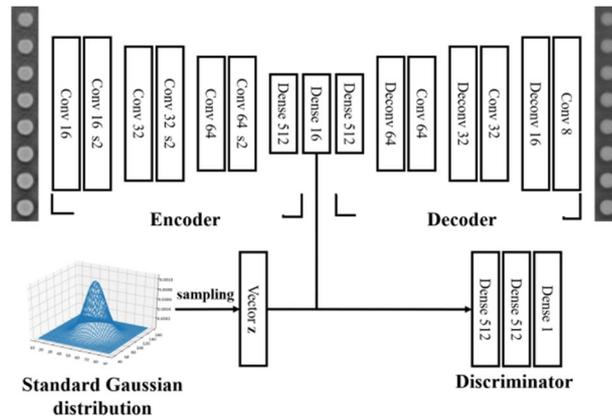


図 2 Adversarial AutoEncoder の概略図

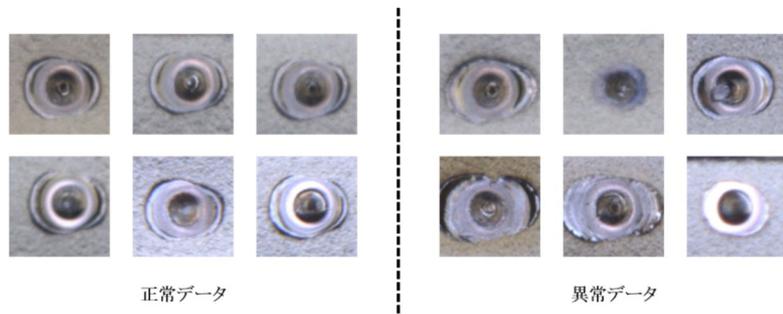


図 3 データセット

AutoEncoder での ROC 曲線と AUC を図 に示す . 図中の凡例における 4l は畳み込みブロックが 4 つであることを示し, 0% は学習サンプルに含まれる異常が 0% であることを示す . 5d や 6d は AutoEncoder のボトルネックの次元数を示している . また, その右の 0.6463 等の数字は AUC を表している . これらから, 特徴次元 5 次元の AutoEncoder では, 十分な異常検知性能が得られていないことがわかった .

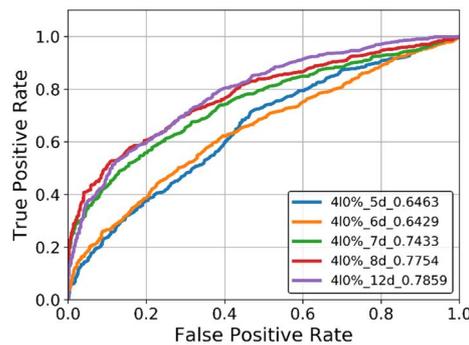


図 4 ROC 曲線

次に, AAE による異常検知について, ボトルネック次元数, 入力画像サイズの比較実験を行い, ボトルネックは 3 層, 入力画像サイズが 32x32 が最適であることが確認された . 精度の改善がみられた畳み込みブロック数・入力画像サイズの変更における共通点として, ネットワークのパラメータ数を減少させていることが挙げられる . パラメータ数を減らすことでより重要な特徴抽出の消失を低

下させることができたことから、有用な特徴の抽出・精度の向上へつながったと推測する。

次に、正常データとごく少数の異常データで構成された訓練データで学習することで、ボトルネックにおける潜在分布が正常特徴であることを前提としているため、ごく少数の異常データを訓練データに含めた場合の実験を行い、異常率を0.5%としたとき、最も高い精度を達成した。異常サンプルの割合を変えたときのROCカーブを図5に示す。

これら実験により、今回用いたデータセットに対しては少なくともボトルネックを12次元は正常特徴の表現に必要な次元数であったために精度が向上した。同一のボトルネック次元数において、入力画像サイズ32x32、異常率0.5%、畳み込みブロック3層のモデルを用いることで、より正常と異常を分離するような特徴抽出ができた。また、上記条件が同一の場合に、畳み込み回数を増やす・全結合層のユニット数を増やす・チャンネル数を少なくしてから全結合層と結合させるといった変更は特徴抽出性能の改善には繋がらなかったと結論づける。

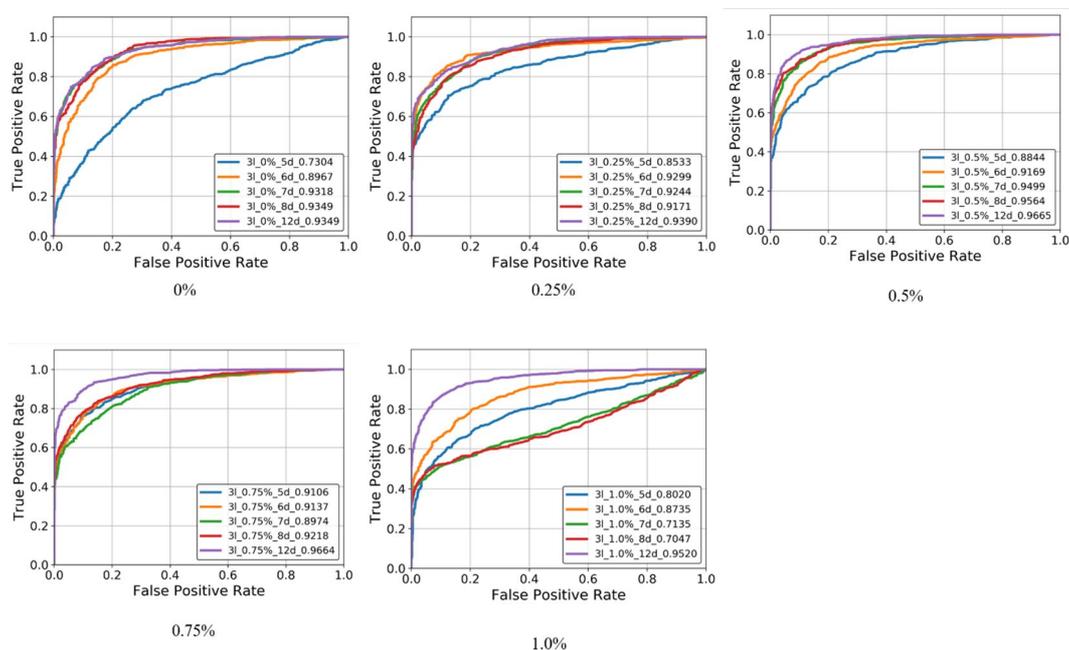


図5 異常率別のROC曲線

## 参考文献

- [1] Makhzani, A., et al.: "Adversarial Autoencoders", arXiv preprint arXiv : 1511.05644, 2015
- [2] Schlegl, T., et al.: "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery". n International Conference on Information Processing in Medical Imaging, pp. 146157, 2017
- [3] Houssam, Z., et al.: "Efficient GAN-based anomaly detection". arXiv preprint arXiv 1802(06222), 2018

## 5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件/うち国際共著 1件/うちオープンアクセス 1件）

1. 著者名 GOTO Keisuke, AIZAWA Hiroaki, KATO Kunihiro, HARADA Yoshihiro, NOGUCHI Minori, NISHIKAWA Hideo, KASHIMURA Kazunori, HAMAYA Akifumi, OHIRA Tsukasa, IHARA Michi	4. 巻 87
2. 論文標題 Anomaly Detection for Wide-Field Images Using Gaze Prediction Module	5. 発行年 2021年
3. 雑誌名 Journal of the Japan Society for Precision Engineering	6. 最初と最後の頁 1008 ~ 1012
掲載論文のDOI (デジタルオブジェクト識別子) 10.2493/jjspe.87.1008	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Goto Keisuke, Kato Kunihiro, Saito Takaho, Aizawa Hiroaki	4. 巻 29
2. 論文標題 Adversarial autoencoder for detecting anomalies in soldered joints on printed circuit boards	5. 発行年 2020年
3. 雑誌名 Journal of Electronic Imaging	6. 最初と最後の頁 1 ~ 1
掲載論文のDOI (デジタルオブジェクト識別子) 10.1117/1.JEI.29.4.041013	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する
1. 著者名 AIZAWA Hiroaki, KOMOTO Kyosuke, KATO Kunihiro	4. 巻 87
2. 論文標題 Mutual Consistency-Ensured Bi-Directional GAN for Anomaly Detection	5. 発行年 2021年
3. 雑誌名 Journal of the Japan Society for Precision Engineering	6. 最初と最後の頁 61 ~ 70
掲載論文のDOI (デジタルオブジェクト識別子) 10.2493/jjspe.87.61	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 NAKATSUKA Shunsuke, KATO Kunihiro	4. 巻 87
2. 論文標題 Visual Inspection Method by Learning Only Normal Data using Complementary GAN with Multi-scale Map	5. 発行年 2021年
3. 雑誌名 Journal of the Japan Society for Precision Engineering	6. 最初と最後の頁 120 ~ 126
掲載論文のDOI (デジタルオブジェクト識別子) 10.2493/jjspe.87.120	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計5件（うち招待講演 1件 / うち国際学会 1件）

1. 発表者名 富田悠斗, 相澤宏旭, 加藤邦人
2. 発表標題 自己蒸留を用いたドメイン普遍的表現学習によるFew-Shot Learning
3. 学会等名 動的画像処理実利用化ワークショップ DIA2022
4. 発表年 2022年

1. 発表者名 Yuto Tomita, Hiroaki Aizawa, Kunihito Kato
2. 発表標題 Universal Representation Learning from Multiple Domains for Few-Shot Classification with Self Distillation
3. 学会等名 Proceeding of 28th International Workshop on Frontiers of Computer Vision(IW-FCV2022) (国際学会)
4. 発表年 2022年

1. 発表者名 相澤 宏旭, 神本 恭佑, 加藤 邦人
2. 発表標題 Mutual Consistency-Ensured Bi-directional GANによる異常検知
3. 学会等名 第26回画像センシングシンポジウム SSII2020
4. 発表年 2020年

1. 発表者名 後藤 圭亮, 相澤 宏旭, 加藤 邦人, 原田 佳周, 野口 稔, 西川 英雄, 鹿志村 和典, 濱谷 章史, 大平 司, 井原 巳知
2. 発表標題 視線領域推定モジュールによる広視野画像からの異常検知
3. 学会等名 ビジョン技術の実利用ワークショップ ViEW2020
4. 発表年 2020年

1. 発表者名 中塚俊介, 相澤宏旭, 加藤邦人, 後藤圭亮, 稲垣達也
2. 発表標題 深層学習を用いた外観検査・異常検知
3. 学会等名 精密工学会画像応用技術専門委員会第5回研究会(招待講演)
4. 発表年 2021年

〔図書〕 計2件

1. 著者名 加藤邦人他	4. 発行年 2020年
2. 出版社 技術情報協会	5. 総ページ数 605
3. 書名 工場・製造プロセスへのIoT・AI導入と活用の仕方	

1. 著者名 加藤邦人他	4. 発行年 2020年
2. 出版社 情報機構	5. 総ページ数 390
3. 書名 データ分析の進め方 及びAI・機械学習導入指南	

〔出願〕 計1件

産業財産権の名称 異常検知システムおよび方法	発明者 後藤圭亮, 相澤宏旭, 加藤邦人他	権利者 同左
産業財産権の種類、番号 特許、特願2020-197113	出願年 2020年	国内・外国の別 国内

〔取得〕 計0件

〔その他〕

-

6. 研究組織	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------