

自己教師あり学習を導入した局所拡張型異常生成による異常検知の高精度化

Enhancing Anomaly Detection via Locally-Augmented Pseudo-Anomaly Generation with Self-Supervised Learning

本田 晴己¹⁾ 山内 悠嗣²⁾
Haruki Honda Yuji Yamauchi

1 はじめに

深層学習に基づく画像認識技術の発展により、画像からの異常検知が産業検査や医療診断など多様な分野で注目を集めている。特に製造業においては製品に含まれる微細な欠陥を高精度に検出することが求められている。異常検知の代表的なアプローチとしては、正常データのみを用いて学習を行い、異常を外れ値として検出する教師なし学習型 [1, 2] や、正常・異常の両方のラベル付きデータを用いる教師あり学習型が挙げられる。教師あり学習型では、異常と正常を明示的に分類するための 2 値分類モデルを訓練することで高い識別性能が期待できるが、異常クラスの多様性や未知性を十分にカバーするには膨大なラベル付きデータが必要となる。しかし実際の現場では、異常の発生頻度が極めて低く、かつ多様であるため、十分な異常データを事前に収集・ラベル付けすることは困難である。

本研究では、これらの流れを受けて、局所的な変換によって生成した擬似異常画像と正常画像とを SimCLR フレームワークで分離学習することにより、明示的な異常データを必要としない汎用的な異常検知モデルを構築する。このとき、SimCLR[3] に代表される対照学習の枠組みを導入し、正常画像と擬似異常画像が特徴空間において明確に分離されるように表現学習を行う。このような構成により、異常画像を一切使用せずに、異常の検出および局在が可能な汎用的なフレームワークを実現する。

本論文では、まず関連研究の動向について述べた後、提案手法の詳細とその学習手順について説明する。次に、公開データセットを用いた実験結果を示し、既存手法との比較により本手法の有効性を検証する。最後に、本手法の利点・限界について考察し、今後の展望を述べる。

2 関連研究

画像に対する異常検知の研究は、主に使用するデータの性質や学習方法に基づいていくつかの系統に分類される。以下では、正常データのみを用いた異常検知に焦点を当て、代表的なアプローチを紹介する。

2.1 再構成ベースの手法

正常画像のみを用いて学習し、入力画像を再構成した結果との誤差を異常スコアとして用いる手法は、異常検知の初期から広く研究されてきた。畳み込みオートエンコーダ [4] は、正常パターンのみを学習することで、異常なパターンの再構成誤差が大きくなることを利用して異常を検出する。Variational AutoEncoder(VAE)[5] や GAN[6] を用いた手法 [7] も提案されており、生成モデルの多様性により異常の再構成を困難にする試みが行われている。

1) 中部大学 Chubu University

近年では、再構成精度を向上させるために、U-Net 構造 [8] や残差学習 [9] を導入した手法が注目されている。中でも DRAEM[10] は、正常画像と擬似異常画像のペアを用いた再構成と識別を組み合わせ、異常領域の局在性能も向上させた。再構成ベースの手法は、直感的で実装も容易である一方、モデルが異常領域まで正確に再構成してしまう過学習の問題があるため、特徴表現の改善が重要な課題である。

2.2 対照学習に基づく自己教師あり手法

対照学習は、自己教師あり学習の分野で急速に発展しており、異常検知にも応用され始めている。SimCLR[3] は、画像の異なるビューを正例ペアとして扱い、特徴空間上で近づけるとともに、他の画像との表現を引き離すように訓練する手法である。Sehwag らは SimCLR を応用し、正常画像の密度推定に基づく異常スコアの算出を行った [11]。また、PatchCore[12] では、事前学習済みモデルによる中間特徴を取得し、その局所パッチ間の距離をもとに異常スコアを計算する構成をとっている。代りに、特徴空間上で正常・異常の分布をうまく分離できれば、明示的な再構成を行わずとも異常検知は可能となる。

2.3 擬似異常生成を利用した自己教師あり学習

正常画像に対して人工的な変換を施し、擬似的な異常サンプルを生成することで、自己教師ありの識別学習を行う手法も提案されている。CutPaste [13] では、正常画像の一部を取り取り、そのパッチを元画像の別の位置に貼り付けることで不自然な局所変化を導入し、それを異常とみなして識別器を訓練する。この手法は単純ながらも異常の局在性能が高く、視覚的にも異常らしい特徴を持つ擬似データを容易に生成できる利点がある一方、現実では発生し得ない擬似的な異常画像を生成する問題も抱えている。また、Hannah et al. は、正常画像に対して自然かつ局所的な異常を加えることで擬似異常画像を生成し、自己教師あり学習により異常検知および異常領域の局在を同時に実現する手法 [14] を提案している。しかしながら、この手法では異常領域を指定するためにマスク画像を必要とする問題がある。これらの手法は、異常らしさを定義するための設計が性能に大きく影響するため、どのような変換を施すかが重要な設計指針となる。

2.4 提案手法の概要

本研究で提案する手法は、擬似異常の生成において CutPaste[13] のような局所的な変換アプローチを踏襲しつつ、SimCLR のような対照学習の枠組みを統合することで、より強力な特徴表現の学習を実現する。本論文の特徴は以下の通りである。

- 現実に発生し得る微小な領域の異常を含む異常画像を自動的に生成する。先行研究では、マスク画像を

必要としていたが、本研究では局所的なデータ拡張により実現する。

- SimCLR[3]に代表される対照学習を導入することで、正常画像と微小な異常を含む異常画像を明確に分離するように学習を行う。通常、SimCLRでは正例と負例の2枚の画像をデータ拡張によりペア画像を作成するが、提案手法では1枚の正常画像から正例と負例のペア画像を作成する。

3 提案手法

本研究では正常画像のみを用いて異常検知モデルを自己教師ありで学習する手法を提案する。提案手法は、図1(a)に示す局所拡張型異常生成、図1(b)に示す対照学習による特徴表現の学習に分けられる。

3.1 局所拡張型異常生成

提案手法は、1枚の正常画像から擬似的な異常画像を生成する。正常画像からランダムに選択した局所領域を切り取りパッチを生成する。生成されたパッチに対して微小な回転及び位置のずれ、色変化等を加えて元の画像に貼り付ける。これにより、現実に発生し得る微小な異常を含む画像を自動的に作成することが可能となる。

しかし、ランダムな領域を切り貼りする過程において、正常画像の背景領域のみが切り取られた場合、異常を含まない異常画像が生成される。このような画像を用いてモデルを学習すると異常検知能力が低下する。そこで、この問題に対処するために、ランダムに領域を切り取った際に異常を含む画像であるかを判定し、異常を含まない画像であれば破棄する仕組みを導入する。

対象物体を含む局所領域をデータ拡張して元の画像に貼り付けると、画像上に変化が発生する。一方、選択された局所領域が背景のみの場合には、画像上の変化は小さい。この特徴を利用して異常の有無を判定する。提案手法では、異常の有無の判定にSSIM(Structural SIMilarity)を使用する。SSIMは、画像の輝度、コントラスト、および構造の3つの要素に基づいて画像間の類似度を評価する指標であり、人間の視覚特性に近い評価が可能であるとされている。元画像 x と局所拡張型

異常生成画像 y の SSIM は、式(1)のように定義される。

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (1)$$

ここで、 μ_x , μ_y はそれぞれの画素の平均値、 σ_x^2 , σ_y^2 は分散、 σ_{xy} は共分散を表す。 C_1 , C_2 は定数で、ゼロ除算を防ぐための安定化項である。

得られた SSIM の値を次式により閾値処理することにより異常の有無を判定する。

$$\text{anomaly image} = \begin{cases} \text{False} & \text{if } \text{SSIM}(x, y) \geq T, \\ \text{True} & \text{else} \end{cases}$$

ここで、 T は閾値である。SSIM の値が閾値を上回る場合には、背景領域のみを切り貼りしたと判断して生成した画像を破棄する。

3.2 自己教師あり学習

提案手法により生成した擬似的な異常画像は、部分的に僅かなデータ拡張を施したものであるため、元画像と非常に似ている。特徴空間において、元画像と擬似的に生成した異常画像が分離されるよう対照学習によりモデルを学習する。本研究では対照学習として SimCLR[3] を採用する。

SimCLRでは、ある入力画像に対して異なる変換を施した2枚の画像を正例ペアとして扱い、別の画像とのペアを負例とみなして学習を行う。本手法では、正常画像同士の拡張ペアを正例として、擬似異常画像の拡張ペアを負例として扱う構成とする。これにより、ネットワークは正常画像の表現を近づけると同時に、擬似異常画像の表現を遠ざけるように学習される。なお、モデルにはResNet18に2層の全結合層を追加したものを採用した。

4 評価実験

提案手法の有効性を検証するために評価実験を実施する。評価実験では、従来手法である CutPaste[13] と提案手法を比較する。CutPasteは、正常画像の一部を切り取り、そのパッチを元画像の別の位置に貼り付けることで不自然な局所変化を導入し、それを異常とみなして

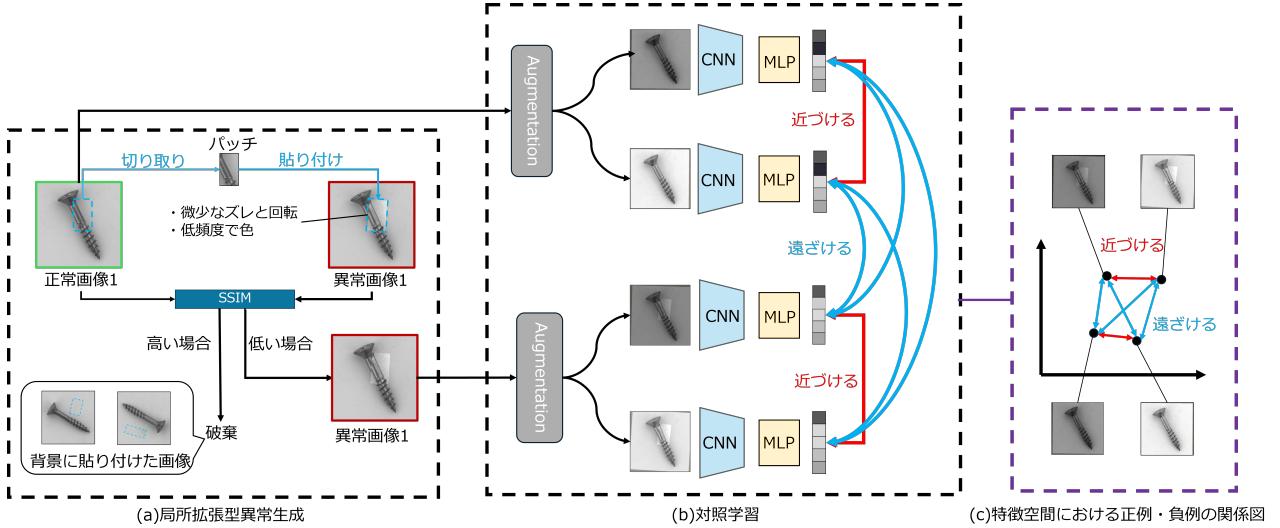


図1 提案手法の流れ

表 1 各手法における分類精度 [%]

手法	bottle	cable	capsule	hazelnut	metal_nut	pill	screw	toothbrush	transistor	zipper	average
CutPaste[13]	100	94.0	93.6	98.7	98.0	95.4	63.5	99.2	95.3	99.7	93.7
Our	100	83.1	97.5	99.6	99.5	83.3	90.2	100	88.4	100	94.2

識別器を訓練する。CutPaste、提案手法とともに局所領域において擬似的な異常付与する手法であるが、CutPasteは現実では発生し得ないような大きな異常を伴う手法である。

評価には、異常検知タスクの標準ベンチマークである MVTec AD データセット [15] を使用した。このデータセットは、10 種類の物体・素材カテゴリから構成されており、各カテゴリについて正常画像と複数種類の異常画像が含まれている。本研究では、訓練には正常画像のみを使用し、異常検知の性能をテスト画像を用いて評価した。評価指標には ROC Area Under the Curve(ROCAUC) を採用した。ROCAUC は、分類の閾値を変化させた際の再現率と適合率をプロットした曲線の下部の面積を表し、分類精度を示す指標として広く用いられている。

4.1 実験結果

表 1 に実験結果を示す。提案手法は従来手法である CutPaste と比較して、ROCAUC の平均が 0.5[%] 向上した。表 1 から半数以上のカテゴリにおいて従来手法を上回る精度を示していることが確認できる。特に screw は精度が 26.7[%] 向上した。

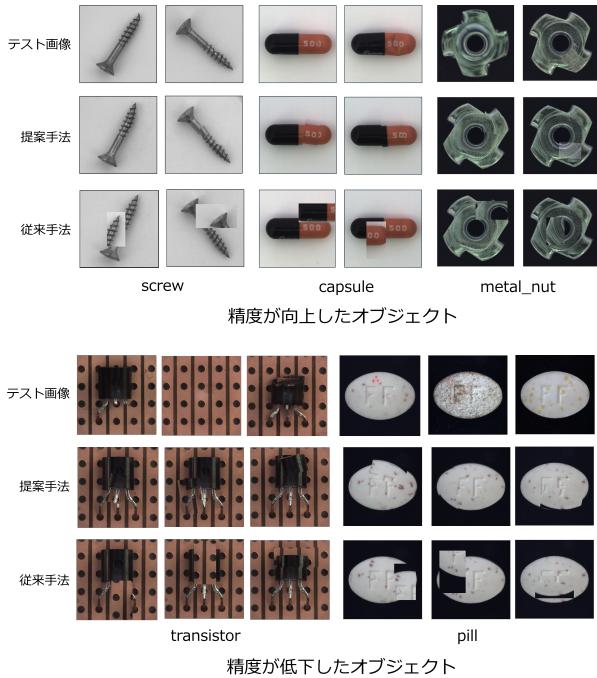


図 2 テスト画像および各手法により生成された擬似的な異常画像

図 2 にテスト画像、従来手法と提案手法により生成した擬似的な異常画像の例を示す。

従来手法では貼り付け処理による不自然さが顕著に表れやすいオブジェクトにおいて、提案手法により精度が向上していることが確認できる。特に、大幅な精度向上

がみられた screw は、背景領域が広く、データセット内で画像の姿勢に一貫性がないことから、従来手法では不自然な画像が生成されやすい傾向にある。一方で、提案手法ではより自然な擬似異常画像が生成されており、テスト画像に近い外観が再現されている。同様に、capsule においても、文字の消失や潰れといった異常が擬似的に再現され、精度の向上に寄与している。また、metal_nut においては、微小な異常に加えて、ひっかき傷のような金属の剥がれを含むテスト画像に対し、局所的な色変化を加えたパッチにより、高い視覚的類似性を持つ擬似異常画像が生成されており、これが精度向上の一因であると考えられる。

一方で、transistor や pill に関しては提案手法の有効性が確認できなかった。transistor においては、提案手法が微小な異常の再現を目的としているのに対し、実際の異常は大きな欠損や構造的な変化が多く、生成された擬似異常との乖離が精度低下の要因と考えられる。また、pill においては、テスト画像の異常が非常に細かい色の変化で構成されているのに対し、提案手法ではそれに相当する擬似異常を十分に再現できず、結果として性能向上には至らなかったと考察される。

5 おわりに

本稿では、自己教師あり学習を導入した局所拡張型異常生成による異常検知手法を提案した。提案手法では、正常画像に対して局所的なデータ拡張を適用することで擬似的な異常画像を自動生成し、SimCLR に基づくコントラスト学習を通じて、正常画像と擬似異常画像の特徴表現を明確に分離するよう訓練を行った。マスク画像や異常ラベルを一切使用しないシンプルな構成ながら、異常の検出および局在の両方において既存手法と同等以上の性能を達成したことを、MVTec AD データセットを用いた評価実験により示した。

今後の課題としては、擬似異常の生成方法のさらなる多様化や、特徴空間における分布モデリングを導入した異常スコア算出の高精度化などが挙げられる。また、実環境におけるリアルタイム異常検知や、マルチスケール異常への対応も将来的な応用に向けた重要な検討事項である。

参考文献

- [1] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, Vol. 13, No. 7, pp. 1443–1471, 2001.
- [2] Lukas Ruff, Robert A Vandermeulen, Nico Görnitz, Lucas Deecke, Alexander Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. *International Conference on Machine Learning*, pp. 4393–4402, 2018.
- [3] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning*, pp. 1597–1607. PMLR, 2020.
- [4] Jonathan Masci, Ueli Meier, Dan Ciresan, and Jürgen

- Schmidhuber. Stacked convolutional auto-encoders for hierarchical feature extraction. In *International Conference on Artificial Neural Networks*, pp. 52–59, 2011.
- [5] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [6] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pp. 2672–2680, 2014.
- [7] Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, Hugo Larochelle, and Ole Winther. Autoencoding beyond pixels using a learned similarity metric. *arXiv preprint arXiv:1512.09300*, 2015.
- [8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.
- [9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [10] Philipp Spoettel, Tobias Rippel, Paul Bergmann, Manuel Fauser, David Sattlegger, and Carsten Steger. Draem - a discriminatively trained reconstruction embedding for surface anomaly detection. In *International Conference on Computer Vision*, pp. 16000–16009, 2021.
- [11] Vikash Sehwag, Bo Wang, Mung Chiang, and Prateek Mittal. Ssd: A unified framework for self-supervised outlier detection. In *Advances in Neural Information Processing Systems*, Vol. 34, pp. 21288–21300, 2021.
- [12] Karsten Roth, Robin Geyer, Edgar Schönfeld, Jonas Müller, Bernt Schiele, and Zeynep Akata. Towards total recall in industrial anomaly detection. In *Conference on Computer Vision and Pattern Recognition*, pp. 14318–14327, 2021.
- [13] Cheng Li, Kihyuk Sohn, Jiyang Hu, Krishna Kumar Singh, Chun-Liang Zhang, Hyun Oh Lee, and Tomas Pfister. Cut-paste: Self-supervised learning for anomaly detection and localization. In *Conference on Computer Vision and Pattern Recognition*, pp. 9664–9674, 2021.
- [14] Hannah M. Schlüter, Jeremy Tan, Benjamin Hou, and Bernhard Kainz. Natural synthetic anomalies for self-supervised anomaly detection and localization. In *European Computer Vision Association*, 2022.
- [15] Paul Bergmann, Manuel Fauser, David Sattlegger, and Carsten Steger. Mvtac ad—a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9592–9600, 2019.