

「正常サンプル」の曖昧さを考慮した異常検知シナリオにおける

従来の外観検査手法と VQ-FastFlow の比較検証

Comparison of VQ-FastFlow and Conventional Methods for Anomaly Detection under Ambiguous Normal Sample Scenarios

齋藤 令次[†] 堀田 一弘[†]
Reiji Saito Kazuhiro Hotta

1. はじめに

近年、製造業では品質管理の高度化と省人化の観点から、外観検査の自動化が求められている。深層学習を用いた外観検査では、異常サンプルの不足や多様性に対応するため、正常サンプルのみを用いた学習による異常検知が有効とされている。しかし、実際の現場では「正常サンプル」の定義は曖昧である。例えば、小さな傷や汚れは許容される一方、異物混入を異常とみなしたい場合がある。また、品質管理上の仕様変更などにより、小さな傷でも許容したくない場合がある。

このような仕様変更に対処するため、文献[1]では 2 つのシナリオが提案されている。1 つ目は、異常サンプルを正常サンプルとみなすシナリオ「Anomaly-to-Normal (A2N)」、2 つ目は、正常サンプルを異常サンプルとみなす「Normal-to-Anomaly (N2A)」である。さらに、仕様変更対象（例えば、小さな傷のある製品）に対して定量的な性能評価が可能となる手法として、「Area Under Anomaly Curve (AUAC) を用いた Area Under the Receiver-Operator Curve (AUROC)」も提案されている。これにより、仕様変更に対して頑健な手法の発見が可能となる。

しかし、文献[1]には 2 つの課題がある。1 つ目は、仕様変更に強いモデルの提案を行っておらず、一般的な外観検査手法との比較のみにとどまっている点である。そのため、仕様変更に対して高い性能を示すモデルを検討する必要がある。2 つ目は、「AUAC を用いた AUROC」が仕様変更に成功している場合であっても評価値が低下する可能性がある点である。この問題点を図 1 に示す。A2N シナリオは異常サンプルを異常と扱う「Standard sub-scenario」と異常サンプルを正常と扱う「Anomaly-to-Normal sub-scenario」の 2 つに分けられる。異常サンプルを正常サンプルとした際にどのくらい仕様を変更できたか評価を行うため、2 つに分けている。図中では前者を「Standard」と示しており、後者を「A2N」と示している。また、図中の数値は AUAC で評価された値であり、黒枠で囲まれているものが 2 種類の仕様変更対象である。この 2 種類の仕様変更は両方とも黒枠内で「Standard」が右に、「A2N」が左に位置しているため、仕様変更に成功している。しかし、文献[1]では、AUROC を用いて閾値の判定を行い、「Standard」全体が右に行くほど、「A2N」全体が左に行くほど良くなる評価指標である。そのため、AUROC では成功例が誤って失敗と評価される可能性がある。

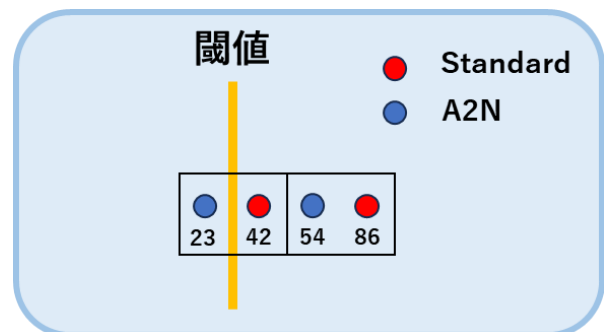


図 1 「AUAC を用いた AUROC」の問題点

これらの 2 つの課題に対して、我々は 2 つの解決策を提案する。1 つ目は、仕様変更に強いモデルを検証するために、VQ-FastFlow[2]を用いる。VQ-FastFlow は VQ-VAE[3] と FastFlow[4]を組み合わせたモデルであり、出現頻度の低い背景のノイズ（歯ブラシの毛など）による過剰な検出を抑制することができる。具体的には、量子ベクトルを用いて背景ノイズをクラスタ化し、それらを正常と判定させることにより、誤検出を防ぐ。我々はこの仕組みを利用し、正常と異常の境界が曖昧なケースの仕様変更を容易にすることができると考えた。2 つ目は、成功例が誤って失敗と評価される問題を解決するために、「Integral of the difference in AUAC」を提案する。この手法は、全体の仕様変更対象に対して評価を行うのではなく、個々の仕様変更対象に対して評価を行う。具体的には、個々の仕様変更対象の AUAC の評価値を用いて「Standard」から「A2N」の評価値を減算する。その後、減算した値を全異常サンプル分行い、数値積分をすることにより評価値を出力する。

評価実験では、文献[1]のシナリオに従い、新たな評価指標を適用して評価を行った。結果として A2N シナリオでは VQ-FastFlow が従来の外観検査手法よりも上回る性能を示し、N2A シナリオでは従来の外観検査手法と同等の性能を示した。

2. 提案手法

本稿では、2 つの課題解決に取り組む。1 つ目は、正常と異常の境界が曖昧なケースに対して、VQ-FastFlow を用いることにより、仕様変更に対する頑健性を向上させることである。2 つ目は、仕様変更対象ごとに個別の評価を可能とする新たな評価指標を提案する。

2.1 VQ-FastFlow の検証

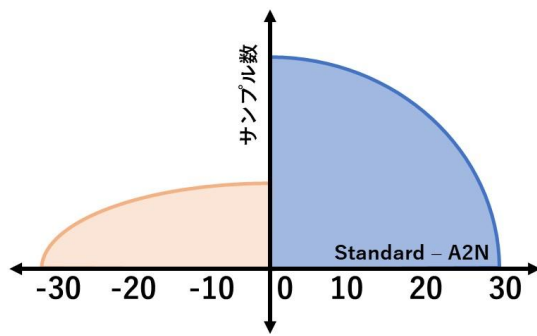


図 2 「Standard sub-scenario」と「Anomaly-to-Normal sub-scenario」の AUAC の差を用いた数値積分

仕様変更に対して高い性能を示すモデルを検討するため、本稿では VQ-FastFlow を用いる。仕様変更対象は、小さな傷やひび割れなど、微小な異常が中心である。本手法では、VQ-VAE に含まれる量子化ベクトルを利用し、微小な異常をクラスタ化し、仕様に応じて異常と正常の定義を切り替えることができる。そのため、仕様変更に対して適応性の高い手法であると考えた。

2.2 Integral of the difference in AUAC

本稿では、仕様変更対象ごとに個別の評価を可能とする新たな評価指標「Integral of the difference in AUAC」を提案する。まず、入力画像に対してモデルから出力される異常度マップを用い、AUAC により各シナリオの評価を行う。A2N シナリオでは、異常サンプルを異常として扱いたい「Standard sub-scenario」と、異常サンプルを正常として扱いたい「Anomaly-to-Normal sub-scenario」の 2 つのシナリオが存在するため、それぞれに対して、AUAC により評価を行う。AUAC は異常度が高いほど大きな値を取り、異常度が低いほど小さな値となる。

仕様変更の達成度を定量的に評価するために、2 つのシナリオにおける AUAC の差分を指標として用いる。これは、異常と正常の定義を変更した際に、その変更がモデルの出力にどの程度適切に反映されたかを示すものである。すなわち、AUAC の差が大きいほど、仕様変更が効果的に反映されていると解釈できる。この差分は 1 種類の異常に対して算出し、これをすべての異常サンプルに対して繰り返すことにより、仕様変更全体への適応度を評価する。

その後、得られた差分を数値積分することにより、最終的な評価値を導出する。図 2 に AUAC の差を用いた数値積分の様子を示す。横軸は AUAC の差 (Standard - A2N)、縦軸は異常サンプルの頻度を示す。数値積分の結果は、図 2 の青色とオレンジ色で示され、青色は仕様変更に適応した領域、オレンジ色は仕様変更に失敗した領域を示す。最終的な評価値は、青色領域を積分した結果からオレンジ色領域を積分した結果を減算することで得られる。これを N2A のシナリオでも行う。これにより、各シナリオに対して適切な評価を行うことができる。

3. 評価実験

評価実験には、MVTecAD[5]データセットを使用し、文献[1]に基づいたシナリオにより、評価を行った。ただし、評価指標には 2.2 節で示した「Integral of the difference in

表 1 従来手法との比較

| | PatchCore | SimpleNet | FastFlow | VQ-FastFlow |
|-----|-----------|-----------|----------|-------------|
| A2N | 8.34 | 17.64 | 18.92 | 24.84 |
| N2A | 11.34 | 38.50 | 34.83 | 38.18 |

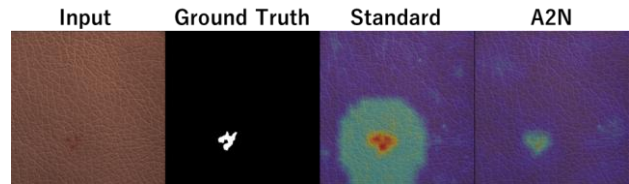


図 3 VQ-FastFlow の A2N シナリオの可視化結果

AUAC」を用いる。この評価指標は値が高いほうが良い指標である。

Table.1 に従来手法と VQ-FastFlow の比較結果を示す。VQ-FastFlow は A2N シナリオにおいて、従来手法よりも高い性能を示した。これは、量子ベクトルが微小な異常をクラスタ化し、仕様に応じて異常と正常の定義を入れ替えることができたためであると考えられる。また、N2A シナリオでは SimpleNet とほぼ同等の性能を示した。この要因として、仕様変更対象に疑似異常を使用しており、両手法において異常と正常の定義を比較的容易に切り替えることが可能であったためであると考えられる。

図 3 に VQ-FastFlow を用いた A2N シナリオにおける可視化結果を示す。左から入力画像、教師ラベル、「Standard sub-scenario」、「Anomaly-to-Normal sub-scenario」である。可視化結果は異常度が高い領域を赤色、低い領域を青色で示している。「Standard」では、異常サンプルを異常として判定したいケースであり、異常箇所が赤く強調されていることから、正しく検出できている。「A2N」では、異常サンプルを正常として判定したいケースであり、異常箇所が赤色ではなく、黄色または青色であることから正しく仕様変更されていることがわかる。

4. おわりに

本稿では、正常と異常が曖昧なケースに対して、VQ-FastFlow を用いて比較を行い、従来手法よりも高い性能を示した。また、仕様変更ごとに個別の評価を可能とする新たな評価指標を提案することにより、精密な評価が可能となった。今後は VQ-FastFlow が高い性能を示すことを確実なものにするために、最新の手法との比較を行いたい。

参考文献

- [1] 齋藤 令次, 神谷 聡, 堀田 一弘, “「正常サンプル」の定義の曖昧さに対処するための新たなシナリオと評価指標”, 第 31 回 画像センシングシンポジウム (SSII), 2025.
- [2] 神谷 聡, 堀田 一弘, “VQVAE を用いた FastFlow の過検出抑制法”, 動的画像処理実用化ワークショップ 2024 (DIA), 2024.
- [3] A. van den Oord et al., “Neural Discrete Representation Learning”, Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.
- [4] Jiawei Yu et al., “Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows”, arXiv preprint arXiv:2111.07677, 2021.
- [5] Paul Bergmann et al., “MVTec AD -- A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019