

AI 画像検査におけるモデル劣化検知の検討

内田 美幸[†], 鷺谷 泰佑[†]

株式会社東芝 生産技術センター 光学・検査技術研究部[†]

1. はじめに

近年, AI による画像分類の発展に伴い, 生産現場においても AI を使った外観画像検査が推進されている. しかし, 運用開始当初は高い検査精度であっても, 運用開始後に経時変化に伴って画像が徐々に変化し, 検査精度が低下する課題がある(図 1). このような AI モデルの劣化が発生した場合, 検査性能低下が顕在化するまで異常に気付くことが出来ず, 不良品流出のリスクが高まる. そのため, 運用時には AI モデルをモニタリングし, AI モデル劣化の発生を早期に検知することが必要である. 今後, さらなる AI 検査の普及に伴い, このような保守管理業務の作業負担は年々増大していくことが予想されるため, AI 検査を自動で保守管理する技術を開発している.



図 1 経時変化に伴う AI 検査モデル劣化

2. 先行研究

AI モデルの劣化を検知するためには, 入力データが学習データから変化したかどうかを判定する必要がある. このような学習していない未知の入力データを検知する手法として, 分類外データ検出技術が挙げられる. 従来, 分類外データの検出手法としては, AI モデルの最終層を利用する手法が知られてきた^[1]. 最終層の値から分類の確信度を算出する手法で, 確信度の最大値もしくは最大値と 2 番目の値の比を指標として未知データを推測する.

しかしながら, ReLU を活性化関数とする一般的な画像分類モデルでは, 従来の最終層の値を用いる手法では未知のデータを検知できない可能性が指摘された^[2]. そこで, 新たに注目されているのが, 中間層における特徴量空間を用いる手法である. 特徴量空間の利用により, 最終層

を利用した従来手法よりも高精度な未知データの検知ができることが報告されている^[3].

一方, 既存のドリフト検出手法による AI モデル劣化検知は困難であると報告されており^[4], 経時変化で生じる僅かな画像変化を検知する手法は確立されていない.

3. AI モデル劣化検知

我々はこれまで, 特徴量空間を利用した分類外データ検出技術を活用し, AI 画像検査における未知の異常検出技術を開発してきた^[5]. 特徴量空間におけるデータ分布を制御する距離学習の適用により, 異常検出の高精度化を実現した.

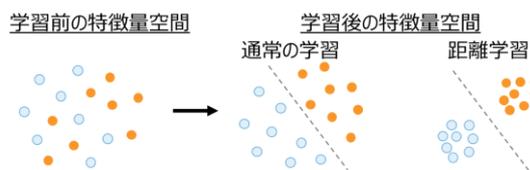


図 2 距離学習の概念図

今回, この技術を応用した AI モデル劣化検出手法を開発した. 開発した手法は, 参照データの特徴量空間におけるデータ分布を元にあらかじめ基準点と基準点からの距離の閾値を定め, 運用時には試験データ群の特徴量空間における分布の代表値と基準点の距離が, あらかじめ定めた閾値を超えたことを検出するアルゴリズムである. さらに, AI モデル劣化の検知性能を向上するため, 距離学習を適用して学習した.

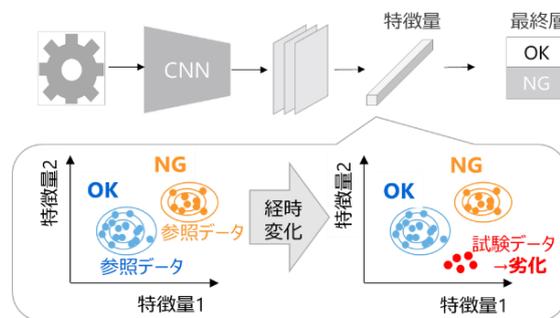


図 3 AI モデル劣化検出技術の概略

The method to detect degradation in accuracy of AI inspection.

[†]Miyuki Uchida and Taisuke Washitani · TOSHIBA CORPORATION

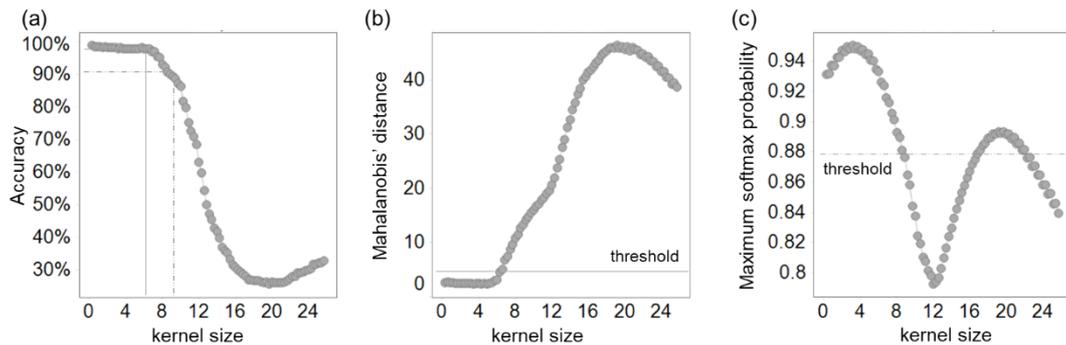


図4 検証結果

4. 検証

本検討では実際の外観検査画像への適用可能性を検証するため、公開画像データセットである MVTEC AD^[6]を用いた。MVTEC ADには15の製品カテゴリーがあるが、本検証では bottle 画像を 256x256 画素にリサイズし、OK/NG の2クラスに分類する AI 検査モデルとした。AI モデル劣化検知の性能を向上するため、距離学習を適用して学習した。

撮像系の経時変化を模擬するため、画像に blur filter を作用させた。画像ボケの度合いを示す kernel size を徐々に増大させ、経時変化による AI モデル劣化を疑似的に再現した。kernel size と検査精度の関係を図 4(a)に示す。

このような経時変化を模擬したデータを用いて、開発した AI モデル劣化検知手法を評価した。blur filter を作用させていないデータを参照データ、blur filter を作用させた各 kernel size 毎のデータを試験データとし、参照データの基準点と試験データの代表値との距離が上限閾値を超えた時点の kernel size を劣化検知点とした(図 4(b))。ここで、閾値は、特徴量空間における参照データの分布広がりをもとに算出した。また、従来の最終層を用いる手法との比較のため、試験データの分類確信度の平均値を指標とし、提案手法と同様の手法で算出した下限閾値を超えた時点の kernel size を、従来手法の劣化検知点とした(図 4(c))。検証の結果、提案手法では劣化検知点が 6.5 画素、検査精度の低下は 1.0%であった。一方、従来手法では劣化検知点が 8.8 画素で、参照データと比較して検査精度は 8.9%と大きく低下していた。また、kernel size が 12~20 画素では、検査精度と分類確信度の増減が矛盾している。そのため、kernel size が突如大きく変化するような急激な劣化が生じた場合には、従来手法では AI モデルの劣化を検知できないことを確認した。

以上の結果より、特徴量空間を利用した提案手法は AI モデル劣化検知に有効であり、従来手

法と比較して AI モデルの劣化を早期に検知可能であることを確認した。

5. おわりに

AI 検査モデルの保守管理自律化に向けて、AI モデル劣化検知手法を開発した。特徴量空間の分布の変化を検出する手法は、従来手法と比較し、正解ラベルなしで高精度な AI モデル劣化検知が可能であることを確認した。

参考文献

- [1] Dan Hendrycks and Kevin Gimpel. A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks. In ICLR, 2017.
- [2] Matthias Hein, Maksym Andriushchenko, and Julian Bitterwolf. Why ReLU networks yield high-confidence predictions far away from the training data and how to mitigate the problem. arXiv:1812.05720v2, 2019.
- [3] Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A Simple Unified Framework for Detecting Out-of-Distribution Samples and Adversarial Attacks. arXiv:1807.03888v2, 2018.
- [4] 石田勉, 金月寛彰, 横田泰斗, 大川佳寛, 小林健一, 中澤克仁, ラベル無し運用ドリフトデータに対するコンセプトドリフト検知手法の評価, 人工知能学会 第 34 回全国大会, 2020.
- [5] 鷲谷泰佑, ディープラーニングを用いた製品の画像検査における異常検知, 情報処理学会 第 84 回全国大会, 2022.
- [6] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. MVTEC AD – A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection. In CVPR, 2019.