

G-51

敵対的生成ネットワークに基づく外観検査の自動化と判断根拠の解釈可視化による信頼性の向上

Improving reliability through automation of visual inspections based on adversarial generative networks and visualization of judgment criteria

平川 疾† 佐野 睦夫†
Hayate Hirakawa† Mutsuo Sanot†

あらまし 外観検査において敵対的生成ネットワークは不良品が少ない中で効果を発揮するが、判断根拠を明確化する課題があった。XAIに基づき判断根拠の解釈可視化を行い、欠陥の因果関係を明らかにし信頼性向上を目指す。

キーワード Efficient GAN, XAI

1. はじめに

近年、外観検査では目視による製造物の検査に頼っていることが多く、人手不足や高齢化、検査ノウハウの継承といった課題がある [1]。こういった課題を解決するにあたって、インライン検査における人工知能による画像処理技術を用いることで、自動で良品・不良品の検査を行うことができ、省力化及び不良品流出の予防等のメリットを享受することが可能になる。

本研究では、インライン検査における野菜を対象として、敵対的生成ネットワーク GAN(Generative Adversarial Networks) [2]を異常検知に応用した Efficient GAN [3]は不良品が少ない中で効果を発揮するが、判断根拠を明確化する課題がある。

説明可能な AI(Explainable Artificial Intelligence(XAI)) [4]を併用することにより判断根拠の解釈可視化を行い、欠陥の因果関係を明らかにし信頼性向上を目的とする。

2. 関連研究

2. 1. 外観検査に関する研究

松谷等 [5]の研究では、カット野菜の外観検査では、野菜の形状や葉脈のパターンが未知であり、異物が加わると誤検知や見逃しが発生してしまい、画像外観検査だと判断が難しい。

そこで、異常検知フィルターで小さなものを取り除きデータの前処理を施すことで野菜の表面のみを学習させる手法をとることで、葉脈パターンを再現しつつ異物を検知させることができた。

2. 2. 説明可能な AI に関する研究

XAI(Explainable Artificial Intelligence)は、AIモデルの決定理由を人間が理解できるがたちで説明することを目指す技術である。例えば、外観検査における異常検知では、なぜ異常となるか原因が説明されないが、説明可能な AI ではなぜ AI が異常と判断したのか説明される。

山科等 [6]の研究では、画像以上検査モデルと大規模言語モデルを組み合わせることで以上についての説明だけでなく、発生要因や誤検知時に訂正できることが確認でき、よりユーザビリティの高い検知システムに寄与できることを確認できた。

3. 提案手法

本研究では、インライン検査における野菜を対象とした異常検知を想定して、良品学習に基づく高速・敵対的生成ネットワーク(Efficient GAN)による異常検知システムと説明可能な AIを併用するシステムを作成する。

説明可能な AIに基づき半単根拠の解釈可視化を行い、欠陥の因果関係を明らかにし、信頼性向上を目的とする。今回作成するシステム図を図 1 に示す。

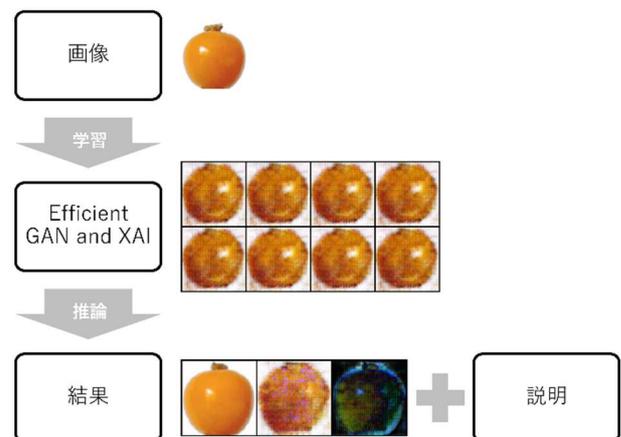


図 1 システム図

3. 1. GAN(Generative Adversarial Network)

GANとは、用意されたデータからデータの特徴を学習し疑似的なデータを生成することができるモデルである。

GANは、生成器と識別機というモデルをニューラルネットワークを用いて誤差逆伝播法という学習下モデルで生成器を使って、疑似的なデータを生成し訓練データと疑似データを比較し本物か偽物かどうかを識別機で識別するシステムである。GANについて図 2 に示す。

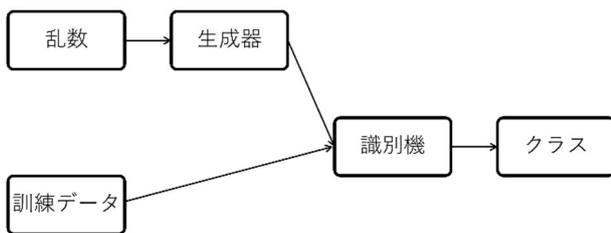


図2 GANの処理フロー

3.2. Efficient GAN

Efficient GANとは、GANでは生成器と識別機の2つのモデルを学習するのと違い、生成器・識別機・エンコーダーの3つのモデルを学習する必要がある。Efficient GANについて図3に示す。



図3 Efficient GANの処理フロー

また、森等 [7]の研究による Efficient GAN による異常検知の時の実データからエンコーダーが求めた潜在変数と実データの差分である再構成誤差(residual loss)と、識別機に入力した判定結果である判別誤差(discrimination loss)の2つを算出し以下の式の通りに計算する。

$$loss = (1 - \lambda) \times residual\ loss + \lambda \times discrimination\ loss$$

λ は森等 [8]の研究により得た結果より 0.1 で計算を行う。この計算結果 loss の値が高いほど異常と判断されたということである。

実装したシステムでは、fruits-360 のデータセット内の柿を利用し 100 エポックで学習した。100 エポック目の学習結果を図4に示す。

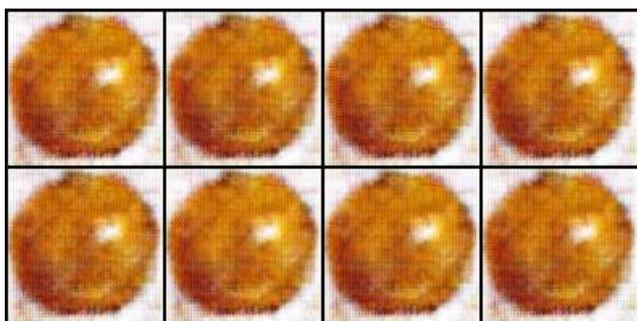


図4 Efficient GANによる学習結果
実際に異常検知を行った結果を図5に示す。

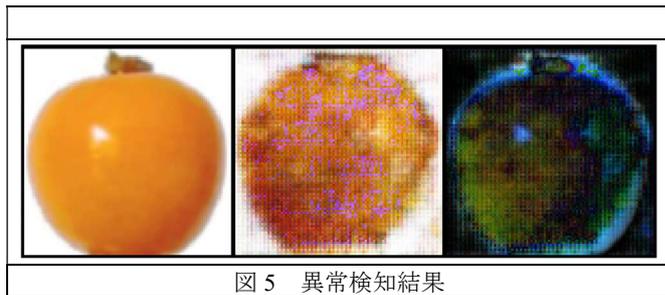


図5 異常検知結果

図5の左の画像から、テスト画像・生成画像・差分画像である。

3.3. XAI (Explainable Artificial Intelligence)

XAIとは、AIモデルの決定理由を人間が理解できるがたちで説明することを目指す技術である。Nagisetty等 [9]の研究では、従来のGANでは識別機から生成器へのフィードバックは損失値だったが、どの特徴に注目したか説明情報を利用して、生成器に送る。

異常の技術を利用することによりなぜ、以上と判断されたかの説明やどの部分が重要だったかの可視化が可能となる。

4. 異常度の評価

今回作成するシステム Efficient GANを利用して、正常画像、異常画像 100 枚異常検知を行った。正常画像の結果をヒストグラムにし図6 最大値・最小値・平均値を表1に示す。

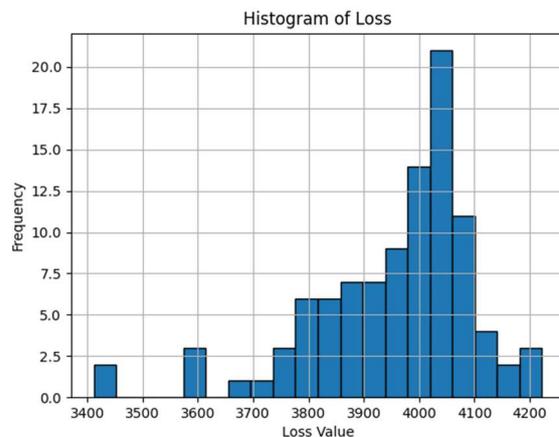


図6 正常画像の異常度の結果

表1 正常画像の異常度の結果

	異常値
max	4222.258
min	3411.803
mean	3955.094

以上の結果から正常画像の異常値の最小値は3411で最大値は4222、平均値は3955であることが分かった。今回は正常な画像を使って異常検知を行った結果なので、異常検

知を行う際には最大値の4222より閾値が大きい必要があることが分かった。

異常画像の結果をヒストグラムに図7最大値・最小値・平均値を表2に示す。

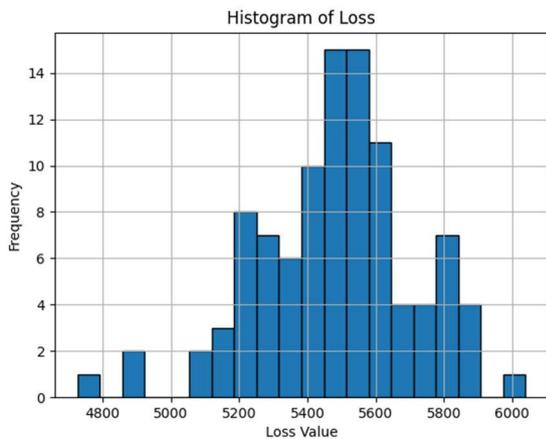


図6 異常画像の異常度の結果

表2 異常画像の異常度の結果

	異常値
max	6039.812
min	4725.235
mean	5481.869

以上の結果から異常画像の異常度の最小値は4725で最大値は6039、平均値は5481であることが分かった。今回は異常画像を使って異常検知を行った結果なので最小値の4725より閾値が小さい必要があることが分かった。

これらの結果から、今回の異常検知での閾値は、4222以上4725未満である必要があることがわかった。

5. おわりに

インライン検査における野菜を対象として、Efficient GANを利用した際の判断根拠の明確化といった課題を解決するため、Efficient GANとXAI技術を併用することにより判断根拠の解釈可視化を行い、欠陥の因果関係を明らかにし信頼性向上を目的としたシステムの提案を行った。

現状、説明可能なAIを実装できていないので今後は、説明可能なAI機能の実装を行う。

参考文献

- [1] 独立行政法人 労働政策研究・研修機構, “ものづくり産業における技能継承の現状と課題に関する調査,” 6 6 2019. [オンライン]. Available: <https://www.jil.go.jp/press/documents/20190606.pdf>. [アクセス日: 25 7 2025].
- [2] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, “Generative Adversarial Nets,” arXiv:1406.2661v1, 10

Jun 2014.

- [3] H. Zenati, C.-S. Foo, B. Lecouat, G. Manek, V. Chandrasekhar, “EFFICIENT GAN-BASED,” arXiv:1802.06222v2, 1 May 2019.
- [4] L. Longo, M. Brcic, F. Cabitza, J. Choi, R. Confalonieri, J. Del Ser, R. Guidotti, Y. Hayashi, F. Herrera, A. Holzinger, R. Jiang, H. Khosravi, F. Lecue, G. Malgieri, A. Páez, W. Samek, J. Schneider, T. Speith, S. Stumpf, “Explainable Artificial Intelligence (XAI) 2.0: A Manifesto of Open Challenges and Interdisciplinary Research Directions,” 30 Oct 2023.
- [5] 松谷直樹, 佐野睦夫, 松田展, 大井翔, “敵対的生成ネットワークに基づくカット野菜の異物混入画像検査方式の検討,” 画像電子学会第303回研究会講演予稿, 2023/02/21.
- [6] 山科勇輔, 須賀圭一, 白井祐典, 市川佳彦, “自然言語での異常解釈: LLMを用いたAI説明モデルの提案,” 言語処理学会第31回年次大会 発表論文集, 2025年3月.
- [7] 森仁美, 仲野美由紀, 丸千尋, 小口正人, “異常度算出手法変更によるGANを用いた異常検知性能の評価,” 2024.
- [8] 森仁美, 仲野美由紀, 丸千尋, 小口正人, “GANを用いた異常検知における異常度算出方法についての考察,” 第16回データ工学と情報マネジメントに関するフォーラム, 2024年.
- [9] V. Nagisetty, L. Graves, J. Scott, V. Ganesh, “xAI-GAN: Enhancing Generative Adversarial Networks via Explainable AI Systems,” arXiv, 29 Mar 2022.