

Gaussian-AD を適用した YOLO を用いた建設機材の異常検知 Anomaly Detection of Construction Equipment using YOLO with Gaussian-AD

万代 弦一郎[†] 黒木 啓之[†]
Genichiro MANDAI Takashi KUROKI

1. はじめに

工場製品における建設機材の検品作業は人力によるものが多く、個人による判断の相違により、異常品の出荷や、作業効率の悪化などの問題が発生する可能性がある。近年では、機械学習を用いて製品の異常検知を自動化することで、作業の効率化や人件費の削減をするだけでなく、人間による曖昧な判断ではなく機械的な判断での判別を図る企業が増加している。

これまで筆者らは、現場で使用することを想定した動画の異常検知に焦点を当て、動画のリアルタイム物体検出が可能なモデルである YOLOv7[1]を用いて、自前で用意した建設機材の学習済みモデルを使用し、建設機材の動画のフレームごとの正常、異常のクラスの予測を行った[2]。しかし、学習用データの不足や、動画で正常、異常を判別するための正常、異常フレーム数を用いた異常度の算出方法が不完全であることなどにより、異常検知の精度があまり良くなかった。

そこで本研究では、動画のリアルタイム画像処理に適した YOLOv7 に、正常品の静止画の特徴量を多変量ガウス分布に近似する異常検知手法である Gaussian-AD[3]を適用し、建設機材の動画の異常検知を行う手法を提案する。この手法により、動画のフレーム毎の異常度を計算し、その異常度を総和した値で二値分類を行うことで、異常検知の精度向上と、リアルタイム異常検知の有効性の確認を目的とする。

2. YOLOv7

本研究では、異常検知に用いる CNN に YOLOv7 を使用する。YOLOv7 は、Joseph Redmon が公認している Alexey Bochkovskiy とその他メンバーらが 2022 年に発表した当時の YOLO で最新の物体検出モデルである。5FPS から 160FPS の範囲において、MS COCO データセットで既存のオブジェクト検出器を速度、精度ともに上回り、GPU V100 での 30FPS 以上のリアルタイムオブジェクト検出器の精度で SOTA を達成している。また、YOLOv7 は他のデータセットや事前学習された重みは一切使用せず、MS COCO データセットのみで一から学習している。

YOLOv7 のネットワーク構造は、図 1 のように 3 段階に分割されたピラミッド構造を持つ。A1 から C5 はそれぞれ出力される特徴量マップ、黒色の矢印はそのデータを変換させる複数の畳み込み層を示す。

赤色のピラミッド構造は入力画像を最初に処理する構造である。この時、A1 から A5 と番号が増えるに従い、より入力画像が縮小し、チャンネル数が多くなる特徴量マップとなる。

青色のピラミッド構造は、特徴量マップのアップサンプリングを行う構造であり、B5 から B3 と番号が減るに従い、より画像が拡大された特徴量マップとなる。また、B5 から B4 の途中に A4、B4 から B3 の途中に A3 の特徴量マップを結合する。

緑色のピラミッド構造は、赤色のピラミッド構造のように再度縮小処理を行う構造であり、C4 から C5 と特徴量マップが縮小される。

そして、最終的に B3, C4, C5 の特徴量マップを用いて物体検出処理を行う。

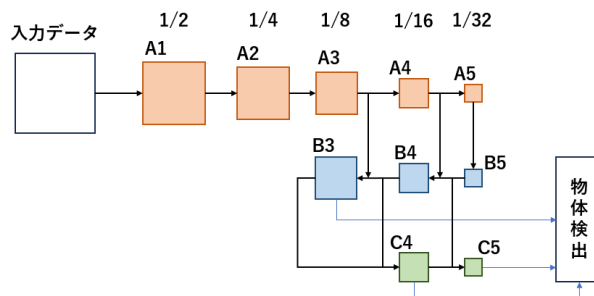


図 1 YOLOv7 のネットワーク構造

3. Gaussian-AD

本研究では YOLOv7 の CNN に、異常検知手法である Gaussian-AD を適用する。Gaussian-AD は、2020 年に Oliver Rippel らが発表した異常検知手法である。

この手法は、大規模自然データセットで学習させた画像分類モデルに入力させた画像の特徴表現が、異常検知における転移学習において正常性を示すという仮定に基づき、ImageNet を学習させた EfficientNet を、特徴表現を抽出するための CNN として用いている。

この手法による異常検知は、学習時と推論時の大きく 2 つの処理に分けられる。

学習時には、まず EfficientNet に正常データのみを学習させ、EfficientNet のネットワークで分割されている 9 つのレベルからそれぞれの特徴ベクトルを抽出し、多変量ガウス分布で近似し、これを正常データの分布とする。

推論時には、同様に EfficientNet で対象画像の特徴ベクトルを算出し、各レベルにおいて、正常データの分布とどれほど離れているかをマハラノビス距離で計算し、異常度

[†] 東京都立産業技術高等専門学校
Tokyo Metropolitan College of Industrial Technology

を求める。その各レベルの異常度の総和が、設定した閾値より高ければ異常、低ければ正常と判別する。

4. 提案手法

本研究では、動画像のリアルタイム異常検知を目的とし、Gaussian-AD を適用した YOLOv7 による異常検知手法を提案する。

この手法は、動画像内の複数のフレームの異常度を考慮するため、動画像内の 1 フレームのキャプチャ画像の異常度のみを考慮するよりも、動画像異常検知の精度向上の有効性を示せると考える。

そのため、動画像の高速な画像認識を行う必要があり、YOLOv7 はその処理に適していると判断する。しかし、YOLOv7 による画像認識のみによる異常検知は困難であったため[2]、フレーム画像の異常検知に Gaussian-AD による異常検知手法を用いる。

Gaussian-AD は、ImageNet を学習させた EfficientNet を、特徴表現を抽出するための CNN として用いている。これは、大規模自然データセットで学習させた画像分類モデルの特徴表現が、異常検知における転移学習において、正常性を示すというものである。このことから、同じ大規模自然画像データセットである MS COCO を学習させた YOLOv7 の学習済みモデルによる特徴表現も同様の正常性を示すと仮定し、YOLOv7 を、特徴表現を抽出するための CNN として Gaussian-AD を適用する。

5. 実験

5.1 異常検知の手順

本研究では、以下のような手順で異常検知の実験を行う。

1. 複数の正常の建設機材の画像を YOLOv7 に入力し、いくつかの畳み込み層ごとの特徴ベクトルを抽出
2. 全ての正常画像の特徴ベクトルの平均と共分散行列を算出し、保存
3. 同様に、異常検知対象の入力動画像を YOLOv7 に入力し、フレーム画像の層ごとの特徴ベクトルを抽出
4. 2. で求めた正常画像の学習データと 3. の特徴ベクトルからマハラノビス距離を計算し、層ごとの異常度を算出し、全ての異常度の総和を算出
5. 4. でフレームの異常度が算出されるので、入力動画像中の指定のフレーム数分の異常度の総和を算出
6. フレーム数分の異常度の総和が設定した閾値より高ければ異常、低ければ正常と判別
7. 混同行列による二値分類の性能評価を実施

5.2 データセット

本研究では、YOLOv7 の学習済みモデルに入力する学習用画像として、工場で実際に検品作業の自動化を取り入れる際に模した形式で撮影した建設機材の動画像のフレームを用いる。建設機材が乗せられている台は傾斜があり、この上で建設機材を転がして側面全体を撮影している。図 2

の(a)は正常の建設機材、(b)は異常の建設機材である。学習に使用する画像は正常画像のみであるので、この場合(a)のような正常の建設機材の画像のみを使用する。また、推論に使う入力動画像は、(a)、(b)のような、正常の建設機材と、何種類かの異常の建設機材の動画像を用いる。



図 2 学習・推論に使用する建設機材画像
(a):正常, (b):異常(汚れ)

5.3 畳み込み層の選択

Gaussian-AD では、EfficientNet の 9 つのレベルごとの特徴ベクトルを抽出している。この時、レベル 1 からレベル 9 になるにつれ、特徴量マップのサイズが小さくなり、局所的な特徴から大局的な特徴を表現する特徴ベクトルとなっている。本研究では、YOLOv7 の CNN において、図 1 に示されている A1 から C5 から出力される特徴ベクトルを分析し、同様の性質があるか、また、異常検知においてどの層の選択が最適かを調査する。

5.4 性能評価

建設機材の動画像では、それが異常品である場合、異常であるフレームと正常であるフレームが存在しており、フレーム単位で性能評価を行うのは困難である。そのため、フレーム単位ではなく、動画単位での性能評価を行う。その性能評価が、5.1 の手順 6., 7. である。

評価指標としては、再現率、適合率、F 値、ROCAUC 曲線を使用する。閾値は正常品の異常度の中央値とし、その後逐次変更していく。

6. おわりに

本研究では、YOLOv7 に Gaussian-AD を適用した建設機材の動画像の異常検知手法を提案した。

今後は、MVTec AD データセットにおける実験により有効性を確認し、その後、建設機材における異常検知を行う。また、YOLOv7 の CNN における適切な畳み込み層の選択も調査する。

謝辞

本研究は、ジー・オー・ピー株式会社の助成により行われたことを記し、ここに謝意を表する。

参考文献

- [1] Wang, Chien-Yao and Bochkovskiy, Alexey and Liao, Hong-Yuan Mark Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors", CVPR, pp. 7464-7475 (2023).
- [2] 万代弦一郎, 黒木啓之, "物体検知モデルによる動画像中の建設機材の異常検知", 第 28 回 電子情報通信学会東京支部学生会研究発表会, No131, 2023.
- [3] Oliver Rippel, Patrick Mertens, Dorit Merhof, "Modeling the Distribution of Normal Data in Pre-Trained Deep Features for Anomaly Detection", ICPR, pp. 6726-6733 (2021).