

物体追跡のための物体検出しきい値の深層学習による最適推定

三浦 一真† Sarthak Pathak† 梅田 和昇†

†中央大学

E-mail: miura@sensor.mech.chuo-u.ac.jp

1. 序論

近年、工場内の見えるかが進められている[1]. 工場内で行われている組立作業を数値化して把握することは、作業の効率化において非常に重要な役割を担っている. 従来では、人が目視で作業状況を確認してきたが、人の目に依存せずに作業工程を把握することができれば、現場の負担を減らすことができる. しかし、セル生産方式かつ作業スペースが厳格に規定されておらず、作業が複数日にわたって行われる大型製品の組立工場では、未だに人が紙で進捗管理を行っている場合が存在する. 既存の進捗推定の研究は、いくつか存在する. 作業者に加速度センサーを取り付けて動作認識を行うことで、作業進捗を推定する研究が存在する[2][3]. しかし、ウェアラブルセンサを用いる手法は、ウェアラブルセンサの費用と装着の負担から現場に導入されることは容易ではない. また、カメラを用いて作業者の動作認識を行うことで、進捗推定を行う研究が存在する[4]. しかし、セル生産方式の大型製品の場合、作業している手元を十分な画質で取得することが難しいという問題と、作業時間が複数日にわたるため、動作認識を行うのが難しいという問題が存在する. そこで、作業者以外の情報に着目した研究として、インスタンスセグメンテーションを用いて作業対象を部品ごとに検出することで、進捗推定を行う研究[5]が提案されている. しかし、工場内のようなオクルージョンが多い環境では、セグメンテーションの検出精度が課題となり、十分な検出精度を得るために必要な学習データの枚数が多い問題が存在する. そこで、作業対象を全体で検出し、深層距離学習を用いてクラス分類を行う研究[6]に着目した. 作業者の情報に依存せず、進捗状況ごとに5枚程度の学習データから作業進捗を行える. しかし、事前に物体検出を行い、作業対象を検出する必要があるため、オクルージョンの多い工場内で検出精度の高い物体検出を行う必要がある. 先行研究として、物体検出アルゴリズムである YOLO(You Only Look Once)[7]を用いて、物体検出を行う際に、動画

のフレームごとに、最適な IoU(Intersection over Union)と Confidence Score のしきい値を推定する研究[8]が提案されている. しかし、推定されたしきい値は検出対象のクラスで共通のものを推定しており、しきい値の推定精度に向上の余地があると考えられる.

本研究では、工場内環境において物体検出の精度を向上させることを目的とする. 物体検出アルゴリズムである YOLO を用いて、物体検出を行う. その際、検出対象のクラスごとに最適なパラメータのしきい値を動画のフレームごと推定して使用することで、工場内映像に生じるオクルージョン、ノイズ、輝度変化等に強い物体検出を提案する.

2. 提案手法

本研究では、物体検出に YOLOv5[9]を用いる. YOLO 実行時に使用するしきい値を、CNN を利用したモデルを新規提案することで推定する.

2.1 YOLO 実行時しきい値について

YOLO 実行時に使用するしきい値を、CNN を利用したモデルを新規提案することで推定する. YOLO は BBox(Bounding Box)推定時に、Non-Maximum Suppression を実行する. この際に設ける IoU のしきい値によって、得られる BBox が変化する. また、BBox には検出物体のクラス分類予測確率として、Confidence Score が出力される. 一般的に、Confidence Score にしきい値を設けることで、しきい値以下の Confidence Score を持つ BBox を削除し、FP(False Positive)を減らすことができる. しかし、Confidence Score のしきい値を高くするほど FN(False Negative)も増えるトレードオフの関係があるため、適切なしきい値を設定することが FP と FN の合計を減らすために重要である.

2.2 深層学習によるしきい値推定

最適しきい値で YOLO を実行することで、正確に

物体検出可能なフレームを学習データおよびテストデータに使用する.

勤務時間中の工場内を固定カメラで上から映した動画から、1 分ごとに画像を切り出す。工場内のイメージ図を Fig.1 に示す。切り出した動画に対して、Confidence Score と IoU のしきい値を 0.1 ずつの値で、0.1 から 0.9 までの範囲で変化させ、 $9 \times 9 = 81$ 通りのしきい値の組み合わせで、YOLO を実行する。YOLO は事前に、作業対象、作業者、進捗状況がわからない一部のみに映っている作業対象の 3 クラスを検出できるようにカスタムデータで学習済みのモデルを使用する。その後、YOLO 学習時に使用したアノテーションデータの BBox と比較して、正しく検出できているしきい値の組み合わせと YOLO 検出結果を抽出する。抽出したデータを Fig.2 に示す、提案するモデルで学習することで、特定フレームにおける検出クラスの最適しきい値を推定する。提案したモデルの入力層には、1 フレームの画像と、クラスごとの BBox が入力され、出力層にはクラスごとの最適な Confidence Score と IoU のしきい値が出力される。

推定時には, BBox には YOLO をデフォルトのしきい値で実行した際の結果を入力する. 推定された最適しきい値を使用して, YOLO を再度実行することで, 最適しきい値での YOLO 実行結果を得る.

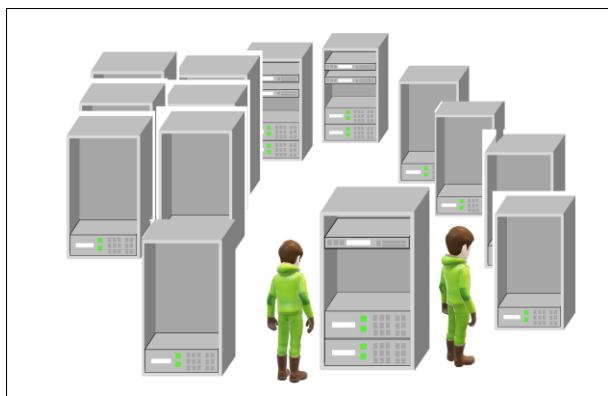


Fig1. 工場内イメージ図

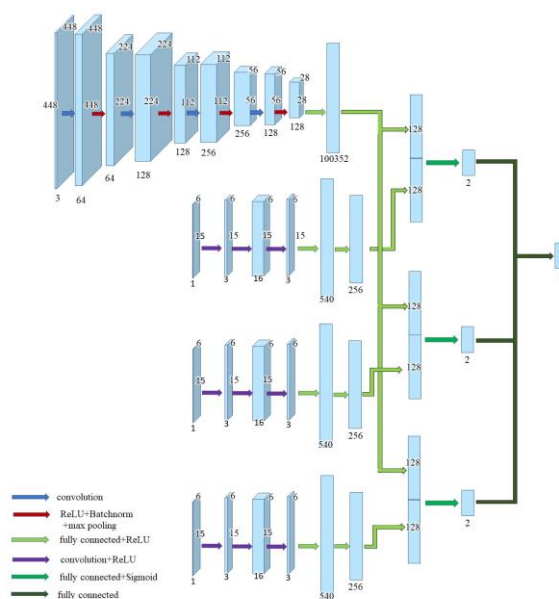


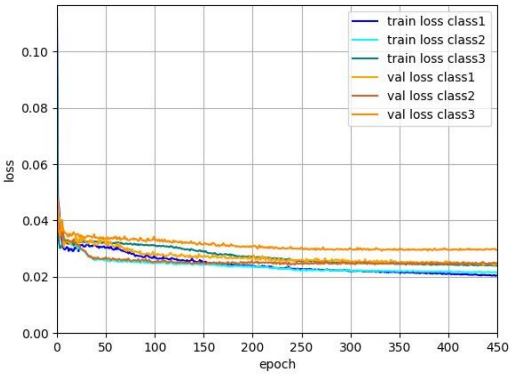
Fig. 2 しきい値学習に用いるモデルの概要

3. 実験

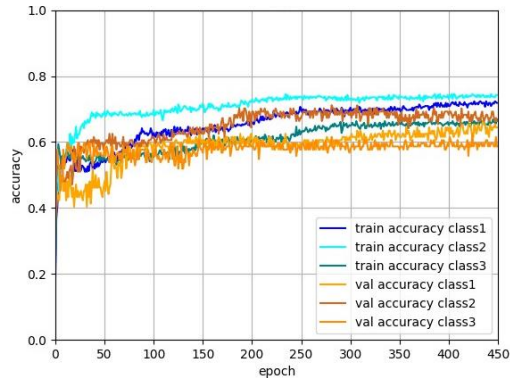
YOLO の最適しきい値の学習及び、最適しきい値を利用した YOLO の実行に関して実験を行った。

3.1 YOLO の最適しきい値学習

二日分の工場内の動画から、1 分ごとに画像を切り出し、計 1179 枚の画像を用意した。294 枚を train および validation 用のデータとし、885 枚を test 用のデータに分割した。検出クラスの class 0 を作業対象、class 1 を作業者、class 2 を一部のみ映っている作業対象として、81 通りのしきい値で検出を行った。294 枚のうち、class 0、class 1、class 2 をしきい値の調整により正しく検出できる画像はそれぞれ 283 枚、270 枚、280 枚だった。すべてのクラスを共通して正しく検出できる画像は 260 枚であった。test 用に分割した 885 枚のうち、すべてのクラスを共通して検出できた画像は 475 枚だった。260 枚を train と validation に 8:2 で 208 枚:52 枚に分割し、提案したしきい値推定モデルの学習に使用した。学習において、epochs = 450, batch size = 4, learning rate = 0.00003 とした。Loss は平均絶対値誤差により評価した。Accuracy は、正解として与えた Confidence Score と IoU のしきい値に対し、差が 0.025 以内であれば正解とした。学習結果を Fig.3 に示す。



(a) Loss 評価



(b) Accuracy 評価
Fig. 3 学習評価

3.2 物体検出の評価実験

Test 用に分割した、475 枚の画像を使い、推定したしきい値で YOLO を実行することで、デフォルトの YOLO のしきい値で実行した場合と比較して、FP、FN の数および、検出失敗画像の枚数が減少するかを評価した。デフォルトのしきい値は、Confidence Score=0.25, IoU=0.45とした。Table1 にその結果を示す。FP, FN, 検出失敗画像のすべてがデフォルトのしきい値で YOLO を実行した場合と比較して減少している結果が得られた。FP&FN の数は 537 個から 410 個に約 23.6%減少し、検出失敗画像は 322 枚から 236 枚に約 26.7%減少した。よって、提案したしきい値推定モデルの有効性が認められた。

Table 1 推定しきい値による物体検出

	デフォルトしきい値	推定しきい値
FP	510	387
FN	27	23
FP & FN	537	410
検出失敗画像	322	236

4. 結論と今後の展望

本論文では、物体検出時に最適しい値を設定することで、物体検出の精度を向上させることを提案した。深層学習を用いた最適しい値を推定するモデルを提案し、フレームごとに検出対象に合わせた最適しい値を推定できることを確認できた。具体的には、475 枚のオクルージョンを含む工場内画像に対して物体検出を行った際に、従来手法に比べて、False Positive と False Negative の合計を 537 個から 410 個に減少させ、検出失敗画像を 322 枚から 236 枚に減少させることを確認した。今後は、最適しい値を利用して得られた物体検出結果から、製品の画像を用いた作業進捗推定を行う。

参考文献

[1] Wang, J., Ma, Y., Zhang, L., Gao, R. X., & Wu, D, “Deep learning for smart manufacturing”, Methods and applications. Journal of manufacturing systems, 48, pp.144-156, 2018.

[2] 夏 清 心, “Research on Sensor Data Processing Methods for Recognizing and Understanding Factory Work Activities”, 研究報告ユビキタスコンピューティングシステム (UBI), 31, pp.1-2, 2022.

[3] Q.Xia, J. Korpela, Y. Namioka, and T. Maekawa, ”Robust unsupervised factory activity recognition with body-worn accelerometer using temporal structure of multiple sensor data motifs”, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT), Vol. 4, No. 3, Article 97, pp. 1–30, 2020.

[4] 吉川裕, 金子真也, 浦野雄大, 永吉洋登, 太田俊広, “映像解析技術を核とした作業認識ソリューション”, 日立評論, pp. 739-743, 2020.

[5] 湯本茂樹, 三浦一真, Sarthak Pathak, Alessandro Moro, 梅田 和昇: “アノテーションデータを活用したデータ拡張によるインスタンスセグメンテーションを用いた組立進捗推定”, 動的画像処理実利用化ワークショップ DIA2024, IS1-16, pp.73-77, 2024.3.

[6] 橘 川 拓 実 , 湯 本 茂 樹 , Sarthak Pathak, Alessandro Moro, 梅田 和昇: “オクルージョンを考慮した深層距離学習による組立作業の進捗段階

の推定”, 動的画像処理実利用化ワークショップ
DIA2023, IS1-15, pp.107-113, 2023.3.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi,
“You Only Look Once: Unified, Real-Time Object
Detection”, *IEEE Conference on Computer Vision
and Pattern Recognition (CVPR)*, 2016.

[8] Kazuma Miura, Takumi Kitsukawa, Sarthak Pathak,
Kazunori Umeda, “Estimation of Object Detection
Threshold for Object Tracking by Deep Learning”,
Joint Conference on 14th France-Japan and 12th
Europe-Asia Congress on Mechatronics & The 9th
Asia International Symposium on
Mechatronics(AISM2023), Mecatronics23-
AISM23-000424, 2023.9.

[9] G. Jocher, ultralytics/yolov5,
<http://github.com/ultralytics/yolov5>.