

# YOLOv3 を用いたロボットアームにおける把持位置と角度の検出

○加藤 博久（山口東京理科大学），永田 寅臣（山口東京理科大学）

## Detection of Grasping Position and Angle in Robot Arm Using YOLOv3

○Hirohisa KATO (Sanyo-Onoda City University), and Fusaomi NAGATA (Sanyo-Onoda City University)

Abstract : You Only Look Once (YOLO) is a fast image recognition algorithm based on deep learning that can simultaneously classify objects and estimate their regions. In this experiment, our proposed fine-tuned YOLOv3 named is applied to the image acquired from a monocular camera attached to the tip of a robot arm. The robot arm is feedback-controlled using the position and angle of the object detected by YOLOv3.

### 1. 緒言

近年，深層学習を中心とした機械学習の進歩により，さまざまな分野において画像認識とそれを応用した自動化や高度な作業の代替が注目を集めている．特に，自動車の自動運転の分野での応用は国内外問わず注力されている．これは車載したカメラや各種センサから得られる情報によってアクセルやブレーキを自動で制御するものである．カメラから得られた画像を一秒間あたり数十から数百フレームの速度で処理し，そこに写っている障害物や人，標識などを認識することを可能としている．このように，自動運転の分野では画像認識技術が応用，研究されているが，自動運転の要求の高さや環境の変化が著しいことから完全な自動運転は実現していない．

一方で，工場の自動化に用いられている産業用ロボットアームの分野では現在でも事前に教示された動作を繰り返す制御が多く用いられている．このような制御では教示されていない動作を行えないため，管理しやすいことが利点としてあげられるが，この教示を用いて人間が行うような高度で柔軟な作業をロボットアームに代替させることは困難である．今後，ロボットアームにより付加価値の高い柔軟な作業を行わせるためには，自動運転に利用されているような，深層学習を用いた物体検出による状況や物体の認識による制御が必要である．

本稿では物体検出アルゴリズムとして YOLO(You Only Look Once) [1] を用いて高速な物体検出を行い，把持物の識別と把持位置，角度を検出する．YOLO は高速に画像認識が可能な R-CNN(Region Based Convolutional Neural Networks) であり，毎秒数十から数百フレームを処理することが可能な深層学習を用いたアルゴリズムである．この YOLO をロボットアームに用いた例はいくつか報告され

Dobot Magician

USB monocular camera

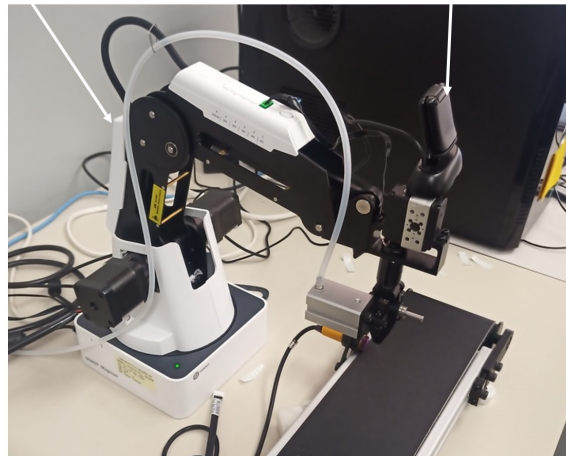


Fig. 1: DOBOT Magician と USB 単眼カメラ

Table 1: 装置の構成

Item	Detail
CPU	Inte <sup>®</sup> Core <sup>™</sup> i7-7700 Processor
GPU	GeForce GTX 1080, 8 GB
Camera	Resolution:960 × 720, 30 fps
Other	Python ver. 3.8.0 Tensorflow ver. 2.3.0

ている [2, 3] が，特に本研究では，対象の一部分を検出することによって姿勢検出への応用を目指す．

### 2. 構成・原理

図 1 に本研究で使用する USB 単眼カメラを取り付けたロボットアーム（Dobot Magician）を示す．Dobot Magician は教育用に製作された Dobot 社製の低価格な汎用小型ロボットアームであり，各プログラミング言語用のサ

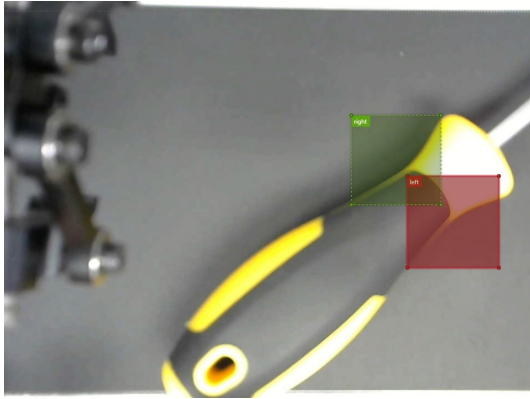


Fig. 2: 把持位置をアノテーションした画像

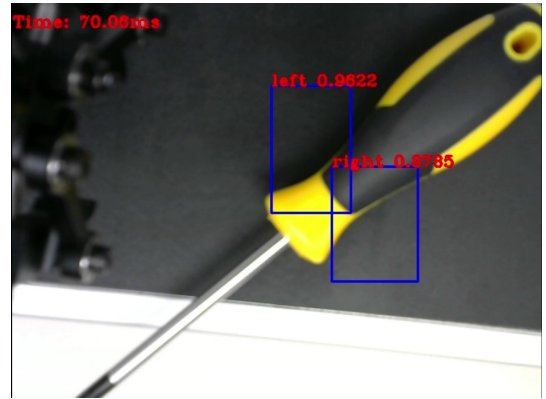


Fig. 3: YOLO を用いた把持位置検出

ンプルプログラムや API が公開されている。この Dobot Magician に解像度  $960 \times 720$ 、フレームレート 30 fps の USB 単眼カメラを取り付けて画像フィードバック制御を行う。表 1 に使用したパソコンと環境の構成を示す。

把持位置と角度検出を行うために、学習済み YOLOv3 [4] に対して把持位置をアノテーションした画像を用いてファインチューニングを行い、把持物であるドライバのくびれ部分を left と right の 2 つの部分として学習させる。角度は 2 つのバウンディングボックスの中心点から算出する。図 2 に学習に使用されたアノテーションされた画像データの一例を示す。把持物をドライバとし、ドライバの把持位置を左右に分けてアノテーションされている。このように、把持物の部分を 2 点以上学習させ検出することにより把持位置と角度を検出する。このような画像データを 100 枚用意し、学習を行った。

本稿では、特定の工業製品の把持を前提としているため、任意の物体の位置・角度検出は考慮していない。従来、このような条件下ではテンプレートマッチングなどの画像処理によって位置と角度を検出する取り組みがなされてきたが、複数の物体や部分を検出対象とした場合、テンプレート画像の増加に比例して処理時間も増加してしまうという問題があった。提案手法のように深層学習である YOLO を適用することで処理時間をほぼ一定にできることが利点として挙げられる。

### 3. 実験結果

実験では学習させたドライバの把持位置と角度の検出を行った。アーム先端に取り付けたカメラから得られる画像に YOLOv3 を適用し、その時の実験結果を以下に示す。

図 3 に YOLO を適用した把持位置検出の様子を示す。青線で示されたバウンディングボックスの 2 つの中心点から直性を求め、把持位置と角度を算出する。YOLOv3 における 1 フレームの処理に 60 ms から 70 ms ほど要して

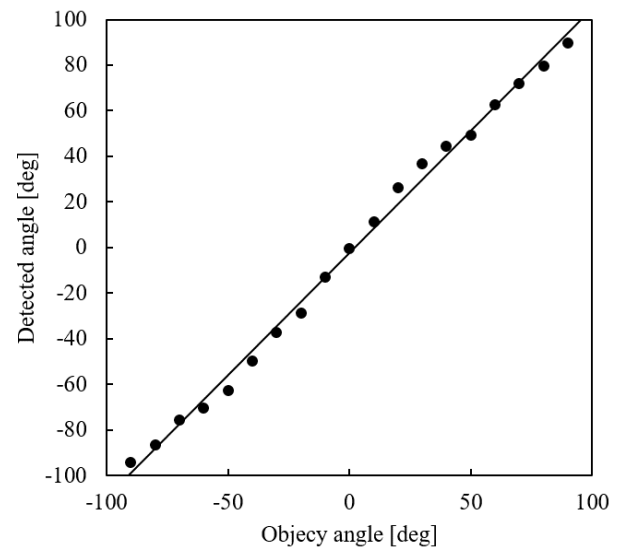


Fig. 4: 検出された角度と実際の角度との関係

おり、15 fps 程度の速度で位置検出が可能であると確認された。

図 4 に検出された角度と実際の角度との関係を示す。実験は USB 単眼カメラから取得された画像の横方向を  $x$  軸、縦方向を  $y$  軸としたとき、角度 0 度を基準としたものである。検出された角度は 2 点の把持位置から算出されるため、YOLO により検出する把持位置間の距離を長くすればそれだけ精度の高い角度が求められる。この実験では実際の把持位置をそのまま学習させたため、2 点間の距離が短く、角度は大きく異なっていることが確認された。最大の残差は 50 度の時の 6.6 度であった。

図 5 に角度 50 度時に検出された 70 フレーム分の把持位置を示す。70 フレーム中の把持位置は USB 単眼カメラと把持物であるドライバを固定した状態で連続して検出された。YOLO から求められる把持位置を画像のピクセルに変換した場合、 $\pm 1$  px の範囲内で把持位置が検出できている。また、この時の処理時間は最大 70 ms、平均 62 ms、

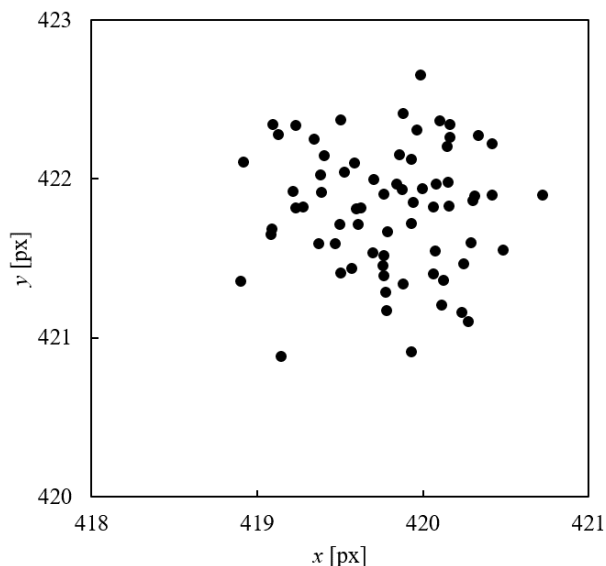


Fig. 5: 角度 50 度時に検出された把持位置

最小 60 ms であった。

#### 4. 結言

本稿ではロボットアームへの応用を目指し、YOLO を用いた把持部分の検出により、把持位置と角度を検出した。実験では学習済み YOLOv3 に対してファインチューニングを行ったものを利用し、15 fps の速度でロボットアームに取り付けた USB 単眼カメラからの画像を処理した。検出された角度は最大 6.6 度の残差があるものの、把持位置は  $\pm 1$  px の範囲内で検出できることが確認された。このような結果から YOLO による把持物の角度と把持位置の検出は柔軟なピッキング動作などロボットアームの高機能化に向けた応用に利用できるものと考えられる。

#### 参考文献

- [1] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi: “You only look once: Unified, real-time object detection”, *Conference on Computer Vision and Pattern Recognition*, pp. 779–788 (2016).
- [2] Jin-Yi Yang, Ui-Kai Chen, Kai-Chu Chang, Ying-Jen Chen: “A Novel Robotic Grasp Detection Technique by Integrating YOLO and Grasp Detection Deep Neural Networks”, *International Conference on Advanced Robotics and Intelligent Systems*, DOI:10.1109/ARIS50834.2020.9205791 (2020).
- [3] Li Tian, Nadia Magnenat Thalmann, Daniel Thalmann, Zhiwen Fang, Jianmin Zheng: “Object Grasping of Humanoid Robot Based on YOLO”, *36th Computer Graphics International Conference*, pp. 476–482 (2019).
- [4] Joseph Redmon: “YOLO: Real-Time Object Detection”, <https://pjreddie.com/darknet/yolo/> (accessed in 2021.10.18)