

YOLO と CNN を用いた物体検出法の提案と評価

永田研究室 F118057 村上 結斗

1. 目的

現在の日本では、少子高齢化に伴う労働人口の減少が懸念されている。近年、AI 技術が急速に発展し、機械による業務の自動化や効率化が期待されている。最近では、AI を物体検出に応用しようとする試みも活発になされている。本研究では YOLO (You Only Look Once) と CNN (Convolutional Neural Network) を用いて、単眼カメラの画像から物体の位置・角度・距離の算出を試みる。CNN で算出した物体までの距離と実測値とを比較し、性能を評価する。

2. 研究内容

まず、Windows のウェブカメラを用いてペンの画像を 50 枚撮影し、Microsoft から提供されている Vott というソフトウェアを用いて画像にペンの両端の位置の情報を加えるアノテーションという作業を行う。次に、位置の情報を加えた画像を用いて 8 : 2 の割合でトレーニング用の画像とバリデーション用の画像に分け、学習済みの YOLO の重みに対してファインチューニングを行う。ファインチューニングされた YOLO を用いてペン両端の位置と分類確度を取得し、2 つの検出箇所からペンの中心位置と角度の算出を行う。この角度は、ロボットが把持する場合の姿勢情報として参照できる。最後に、図 1 の矩形枠で示すように YOLO が検出した部分画像を用いて CNN を訓練し、その部分までの距離を求める。

3. 実験結果

図 1 に示した状態からカメラのペンの高さを 25 mm ずつ上昇させながら単眼カメラから得られた画像を訓練データとし、学習後の YOLO, CNN を用いてペンまでの距離を測定し、性能を評価した。図 2 には、CNN が検出した距離と実際の距離の関係を示す。図 1 に示す例では、YOLO によりペンの両端が top と bottom として検出され、また CNN によりそれぞれまでの距離が推定出来ていることが確認できる。その他の測定結果をまとめたものを図 2 に示す。各測定点で 50 回の距離の測定を行い、その最大値、平均値および最小値をエラーバーとして示している。どちらの結果においても実際の距離に比例して良好に検出できていることが確認できる。これらの実験結果より、検出された距離は校正することで利用できると思われる。top のグラフで実距離 70 mm 付近で近似曲線から低く外れている検出値に関しては、これまでの実験から光の反射が強くなっている部分に認識する部分が被ると検出値が小さくなってしまいう現象が確認されていたため、それが原因になったものと考えられる。以上の結果から、自動化ラインなどできるだけ光の反射が一様になっている空間で利用することがより効果的であると思われる。今後の学習においては、CNN の訓練データに対して、光の反射を考慮した画像をより多く追加するなどの対策を行いながら性能の向上を目指していきたい。

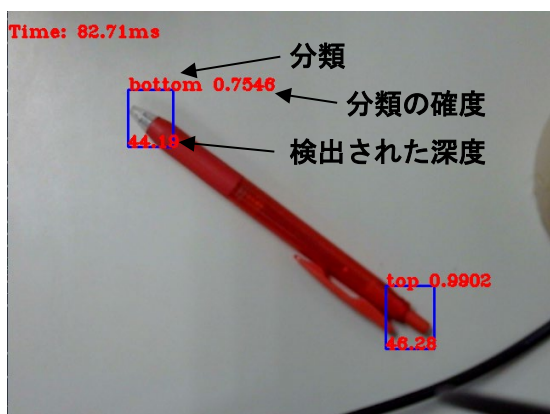


Fig. 1 Partial depth estimation using CNN.

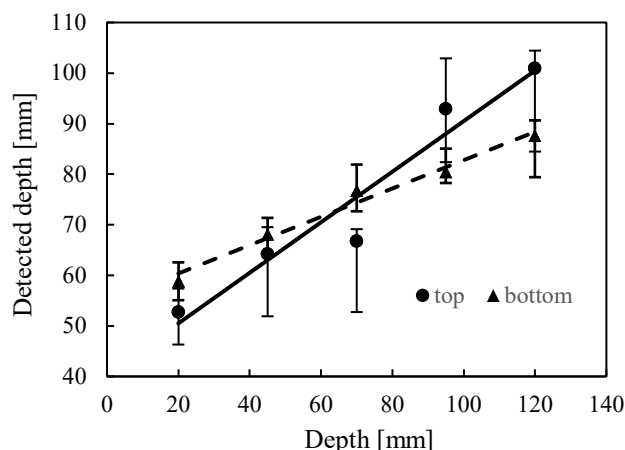


Fig. 2 Relationship between detected depth and actual depth.