

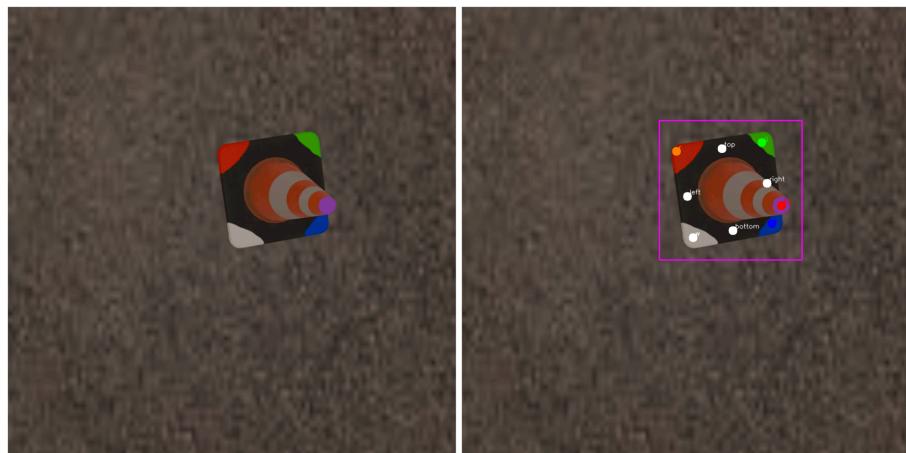
姿勢推定に向けたセグメンテーションモデル構築状況報告

キーポイントモデルからの転換と進捗報告

0. 前回まで

サイドスラスターを「コーン」に見立て、姿勢推進のためのキーポイント（9点）を設定した学習データを作成後、YOLOv5モデル（物体検知モデル）のニューラルネットワークをキーポイント推定モデルに改造して学習。

学習データ



結果：失敗

200件のデータを5時間学習させるも、学習の傾向が見られなかった。ニューラルネットワークの構造に問題はないが、キーポイントに対する「[学習のさせ方](#)」に問題ありと判断。

1. セグメンテーションモデル

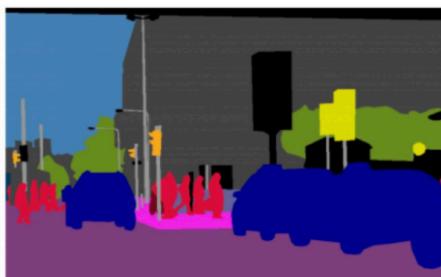
キーポイントと同様に姿勢推定の情報を得られる **セグメンテーションモデル** へ切り替え。

セグメンテーションモデルとは

画像中の各画素を部品カテゴリーに分類するモデル。



(a) image



(b) semantic segmentation

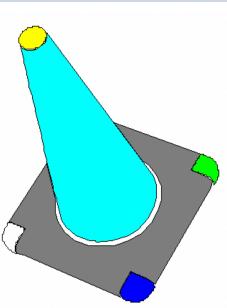
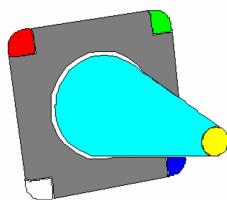
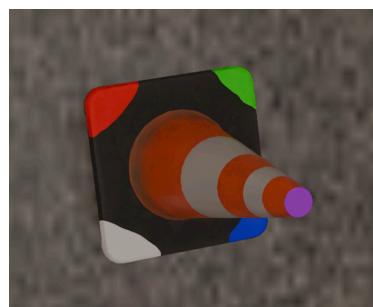


(c) instance segmentation



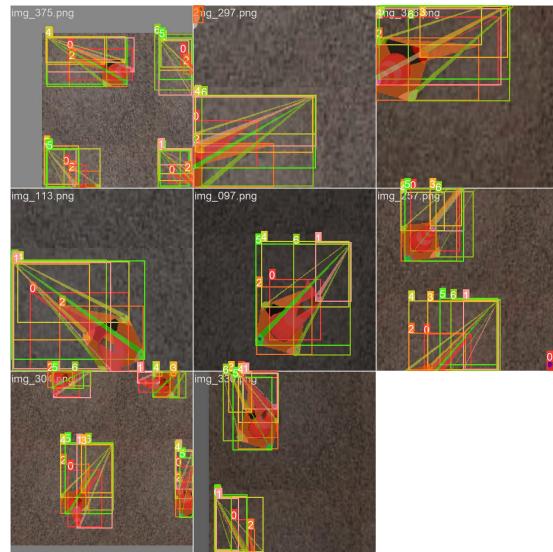
(d) panoptic segmentation

学習データ



2. 現在

400枚のデータで学習中。



なぜセグメンテーションを採用しなかったか

- 学習データの作成が困難

対応方法：3日間かけて、セグメンテーション学習データを自動作成するプログラムを作成

⚠️ ただし 作成に要する時間 1枚あたり 30秒以上 (400枚の場合 3時間半)

- 学習（モデル構築）に時間を要する

3年前の同様の開発 データ1,200枚以上で「CPUの場合は約20時間」「GPUの場合は約4時間」

- ラズパイで動かせるサイズにできる見込み

現在は YOLOv5s-seg、可能であれば YOLOv5n-seg

Model	size (pixels)	mAP _{box} 50-95	mAP _{mask} 50-95	Train time 300 epochs A100 (hours)	Speed ONNX CPU (ms)	Speed TRT A100 (ms)	params (M)	FLOPs @640 (B)
YOLOv5n-seg	640	27.6	23.4	80:17	62.7	1.2	2.0	7.1
YOLOv5s-seg	640	37.6	31.7	88:16	173.3	1.4	7.6	26.4
YOLOv5m-seg	640	45.0	37.1	108:36	427.0	2.2	22.0	70.8
YOLOv5l-seg	640	49.0	39.9	66:43 (2x)	857.4	2.9	47.9	147.7
YOLOv5x-seg	640	50.7	41.4	62:56 (3x)	1579.2	4.5	88.8	265.7

学習の見込み

キーポイントモデルのようにYOLOv5の改造は必要とせず、YOLOの流儀に従った「再学習」(fine tuning) を実施するため、データの量と質の工夫で確実に学習可能。

3. 過去の事例

3年前の「某大手ハウスメーカー向け高力ボルトの締め付け角度推定」で、セグメンテーションモデルを使用。

