

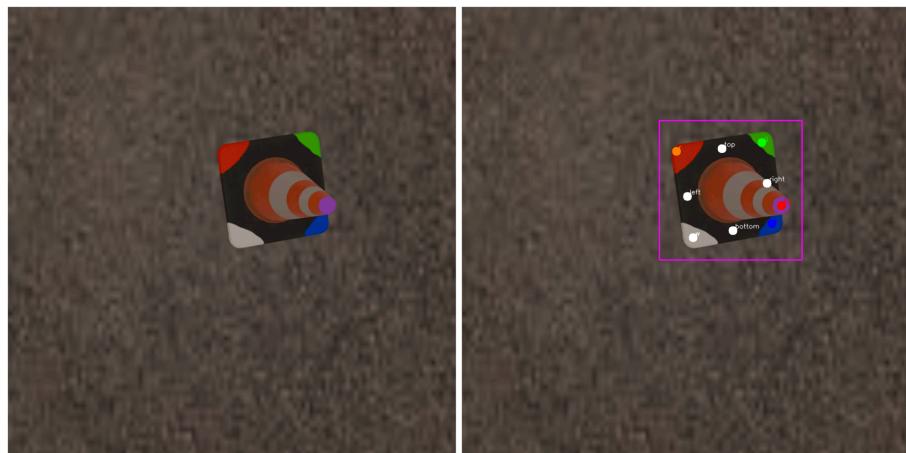
姿勢推定に向けたセグメンテーションモデル構築状況報告

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

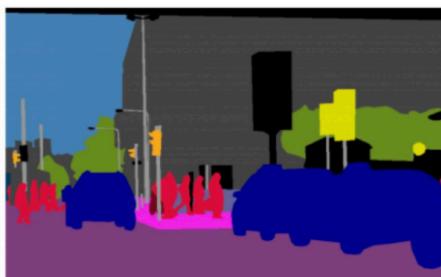
キーポイントと同様に姿勢推定の情報を得られる **セグメンテーションモデル** へ切り替え。

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。



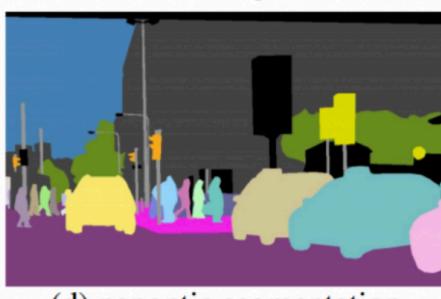
(a) image



(b) semantic segmentation

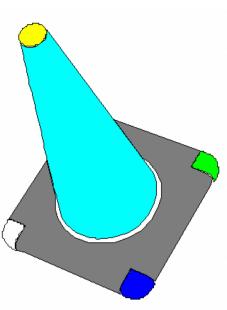
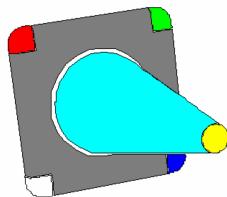
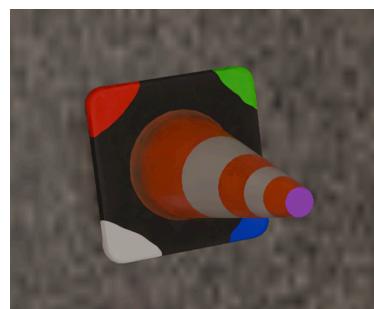


(c) instance segmentation



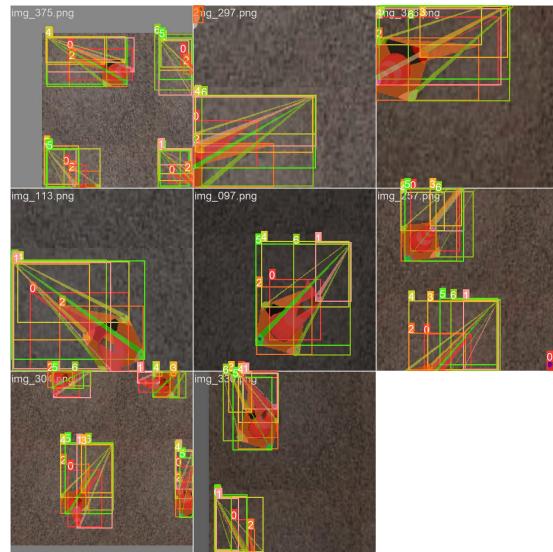
(d) panoptic segmentation

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

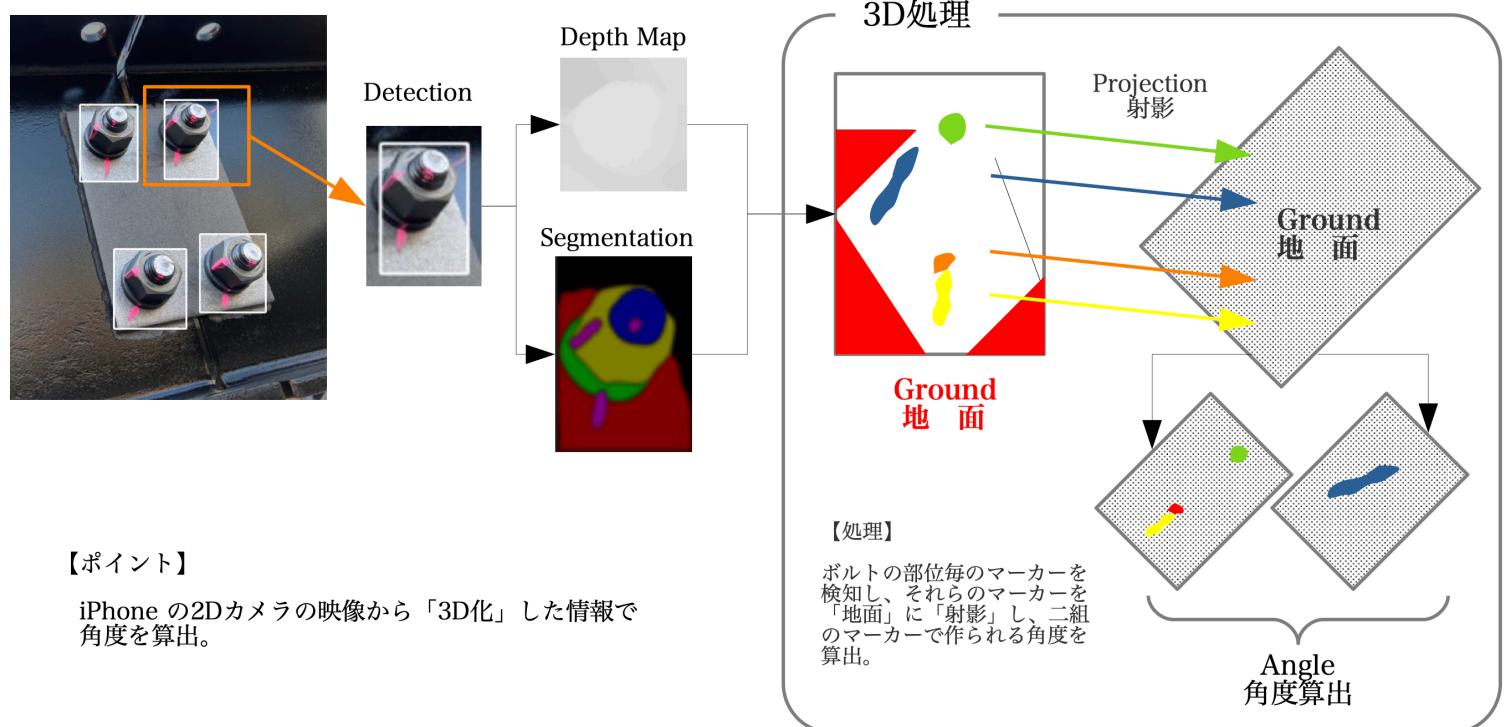
Model	size (pixels)	mAP _{box} 50-95	mAP _{mask} 50-95	Train time 300 epochs A100 (hours)	Speed ONNX CPU (ms)	Speed TRT A100 (ms)	params (M)	FLOPs @640 (B)
YOLOv5n-seg	640	27.6	23.4	80:17	62.7	1.2	2.0	7.1
YOLOv5s-seg	640	37.6	31.7	88:16	173.3	1.4	7.6	26.4
YOLOv5m-seg	640	45.0	37.1	108:36	427.0	2.2	22.0	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。



学習状況：7月22日 16:00

mAP50などの学習評価値が初めて表示、つまり学習している証。

```
...
13/49      0G    0.05222   0.0398    0.0586    0.03851    73     1024: 90%|
██████████| 36/40 [08:20<00:56, 14.08s/it]
13/49      0G    0.05209   0.03974   0.05837    0.03843    86     1024: 92%|
██████████| 37/40 [08:34<00:42, 14.03s/it]
13/49      0G    0.05207   0.04005   0.0586    0.03837    107    1024: 95%|
██████████| 38/40 [08:48<00:27, 13.98s/it]
13/49      0G    0.05197   0.03996   0.05835    0.03844    86     1024: 98%|
██████████| 39/40 [09:01<00:13, 13.75s/it]
13/49      0G    0.05206   0.04019   0.05809    0.03857    83     1024: 100%|
██████████| 40/40 [09:16<00:00, 13.91s/it]
```

Class	Images	Instances	Box(P)	R	mAP50	mAP50-95)	Mask(P)	R
mAP50	mAP50-95):	100% ██████████						
all	80	545	0.632	0.69	0.688	0.295	0.00218	0.00176
0.000103	1.03e-05							

...

学習状況：7月22日 8:00

「不正」なデータを直して再度学習。前回は 13 epoch (= 13段階) で学習評価値が表示、今回は 5 epochで表示。つまり良い学習データになった証。

```
...
5/99      0G    0.06147   0.04896   0.07332    0.0478    77     1024: 90%|██████████
```

| 36/40 [06:27<00:44, 11.20s/it]

```
5/99      0G    0.06141   0.04911   0.07311    0.04778    87     1024: 92%|
```

██████████| 37/40 [06:39<00:33, 11.29s/it]

```
5/99      0G    0.06132   0.049    0.0731    0.04775    105    1024: 95%|
```

██████████| 38/40 [06:53<00:24, 12.22s/it]

```
5/99      0G    0.06123   0.04886   0.07292    0.04774    85     1024: 98%|
```

██████████| 39/40 [07:05<00:12, 12.25s/it]

```
5/99      0G    0.06121   0.0487    0.07259    0.04773    85     1024: 100%|
```

██████████| 40/40 [07:18<00:00, 10.95s/it]

Class	Images	Instances	Box(P)	R	mAP50	mAP50-95)	Mask(P)	R
mAP50	mAP50-95):	100% ██████████						
all	80	545	0.305	0.308	0.301	0.0906	0.000173	0.00882
0.000101	1.2e-05							

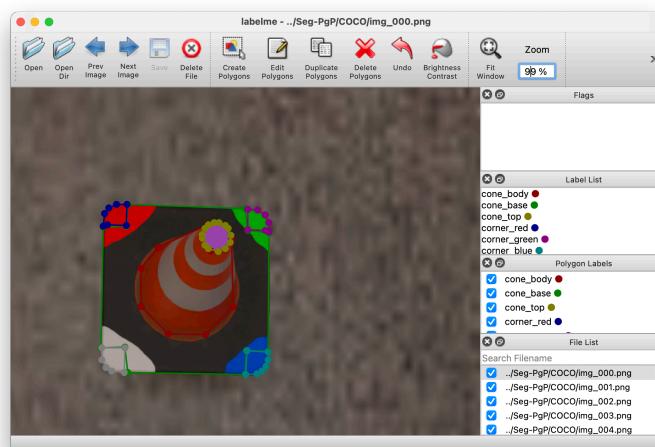
...

追記：学習データ見直し

フォーマット不正改修後



目視確認&修正



学習状況：7月23日 8:00 mAP50などの学習評価値が 1 epoch 目から表示、つまりさらにデータは良くなつた証。

```
...
0/99      0G      0.1095      0.1175      0.09609      0.05791      87      1024: 90%|██████████
| 36/40 [06:23<00:44, 11.24s/it]
0/99      0G      0.1092      0.1161      0.09623      0.05771      102      1024: 92%|██████████
| 37/40 [06:36<00:34, 11.56s/it]
0/99      0G      0.1087      0.115       0.09622      0.05748      97      1024: 95%|██████████
| 38/40 [06:47<00:22, 11.47s/it]
0/99      0G      0.1082      0.1141      0.09603      0.05728      83      1024: 98%|██████████
| 39/40 [06:59<00:11, 11.49s/it]
Class     Images Instances      Box(P      R      mAP50      mAP50-95)      Mask(P      R
mAP50    mAP50-95): 100%|██████████
all        80      537       0.749      0.0462      0.0155      0.00347      0.742      0.0304
0.0197    0.00428
...
```

最終結果

...
30/99 0G 0.03691 0.0203 0.03183 0.007395 76 1024: 100% |
██████████| 40/40 [07:42<00:00, 11.56s/it]
Class Images Instances Box(P R mAP50 mAP50-95) Mask(P R
mAP50 mAP50-95): 100% |██████████|
all 80 537 0.901 0.958 0.975 0.645 0.901 0.958
0.972 0.766

項目	値
Box mAP50	0.975
Box mAP50-95	0.645
Mask mAP50	0.972
Mask mAP50-95	0.766
Precision	0.901
Recall	0.958
学習データ枚数	80 images
インスタンス数	537
エポック	30 / 100

推論結果

