GamblePal™

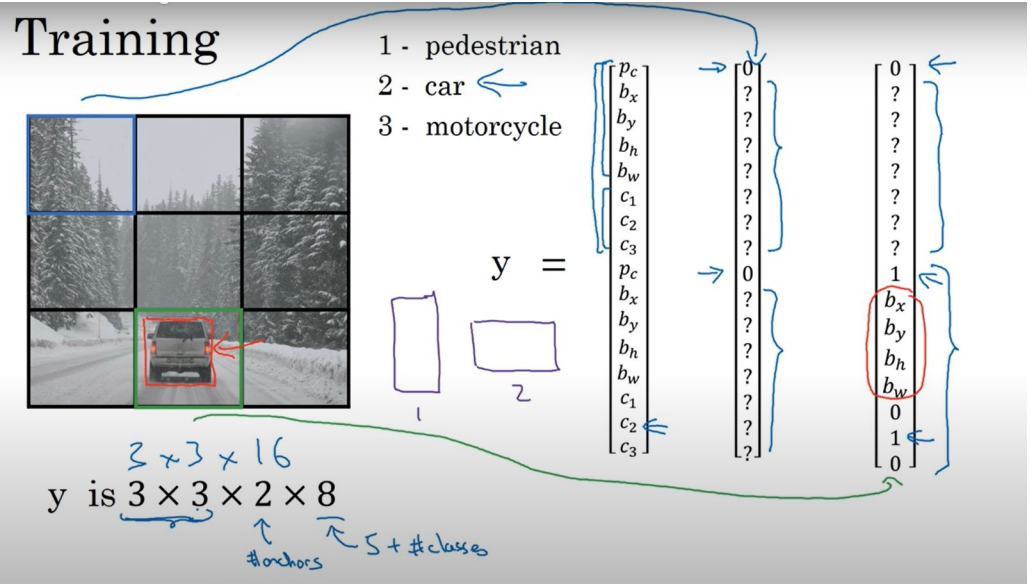
Violet Orange Inc.

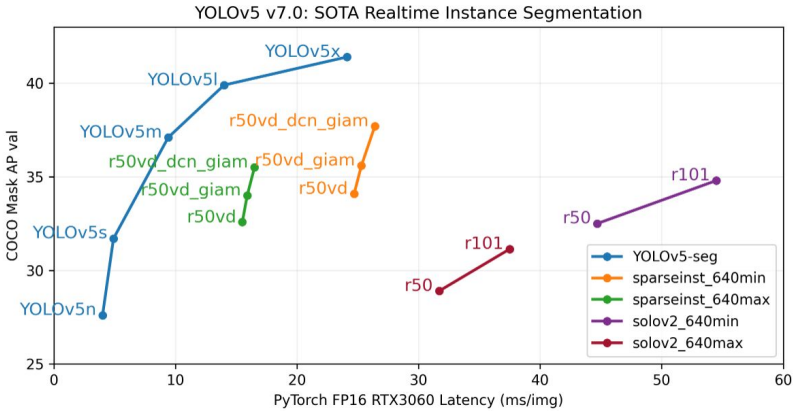
Table of Contents

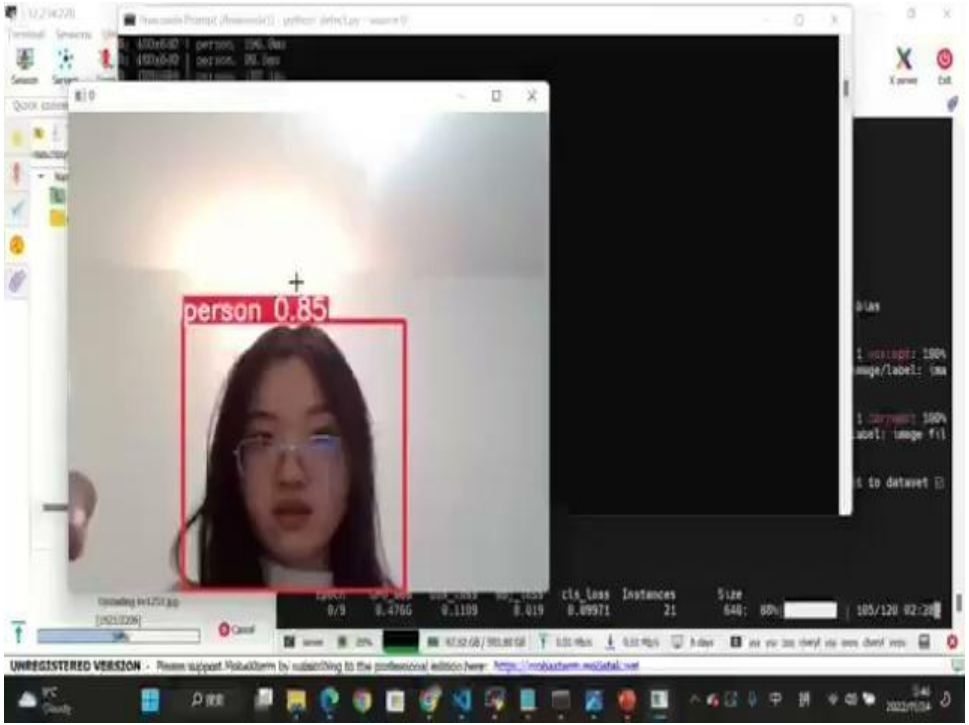
- Hand
- Opening the Pot
- Raise

Hand

Model


Training
1 - pedestrian
2 - car
3 - motorcycle
y =
y is 3 × 3 × 2 × 8
3 × 3 × 16
#anchors
5 + #classes

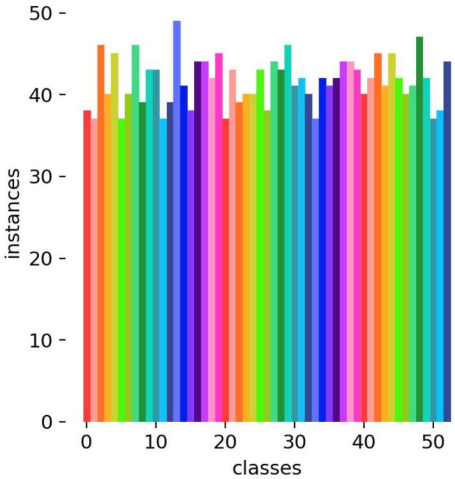

YOLOv5 v7.0: SOTA Realtime Instance Segmentation
COCO Mask AP val
PyTorch FP16 RTX3060 Latency (ms/img)
YOLOv5n
YOLOv5s
YOLOv5m
YOLOv5l
YOLOv5x
r50vd
r50vd_giam
r50vd_dcn_giam
r50
r101
YOLOv5-seg
sparseinst_640min
sparseinst_640max
solov2_640min
solov2_640max



YOLO training set doesn’t contain cards. Need re-train or fine-tuning

Dataset

- *The Complete Playing Card Dataset*
- 53 Classes
- ~40 Images per Class
- Fully annotated (YOLOv5 format)



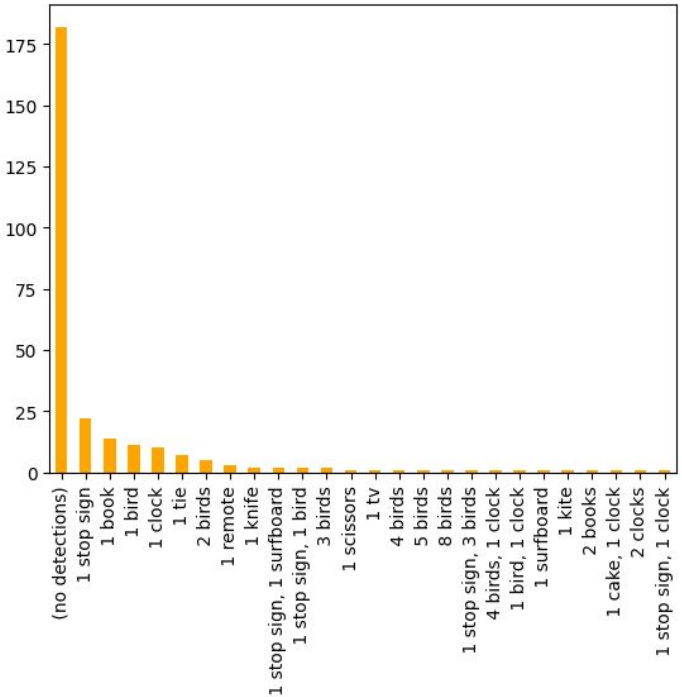
10C0.txt (40 B)

```
37 0.325087 0.665365 0.083767 0.039641
```

Some problems with annotation. How to solve?

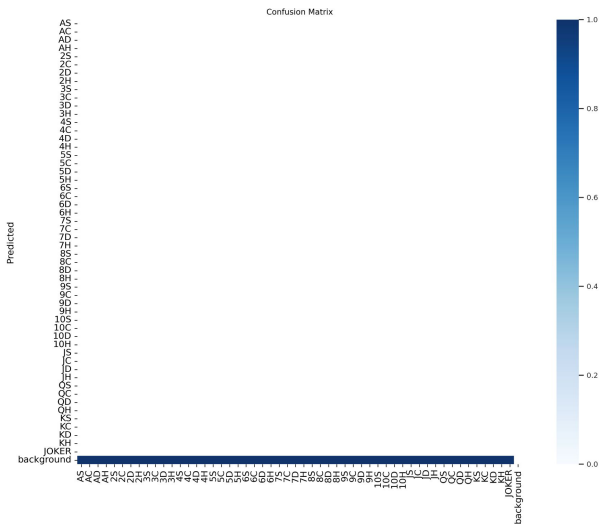
Opening the Pot

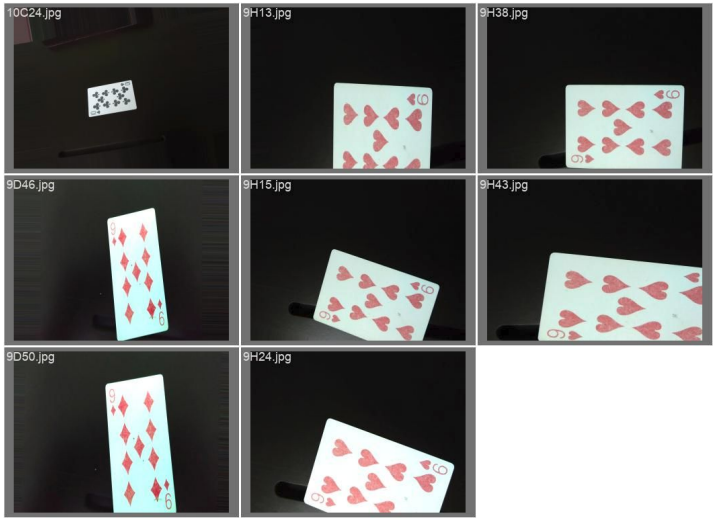
No Training

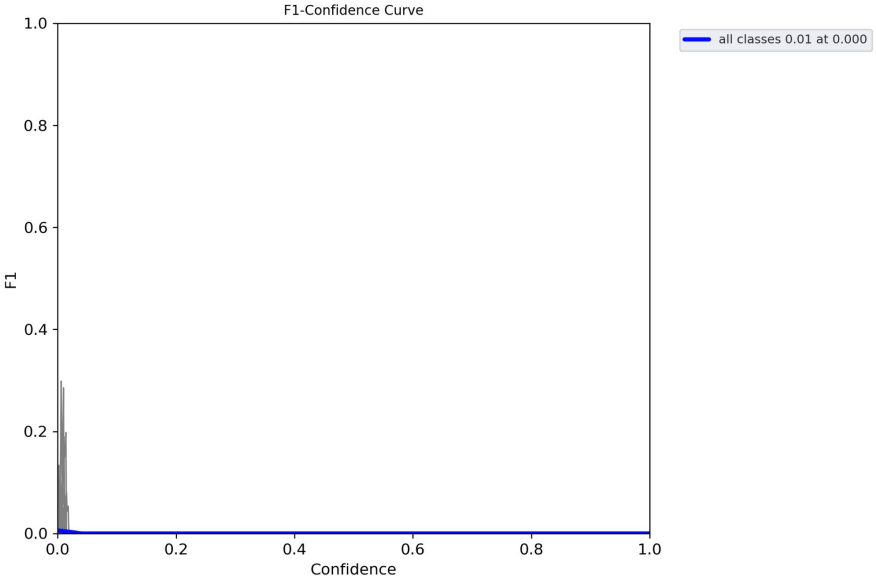

175
150
125
100
75
50
25
0
(no detections)
1 stop sign
1 book
1 bird
1 clock
1 tie
2 birds
1 remote
1 knife
1 stop sign, 1 surfboard
1 stop sign, 1 bird
3 birds
1 scissors
1 tv
4 birds
5 birds
8 birds
1 stop sign, 3 birds
4 birds, 1 clock
1 bird, 1 clock
1 surfboard
1 kite
2 books
1 cake, 1 clock
2 clocks
1 stop sign, 1 clock

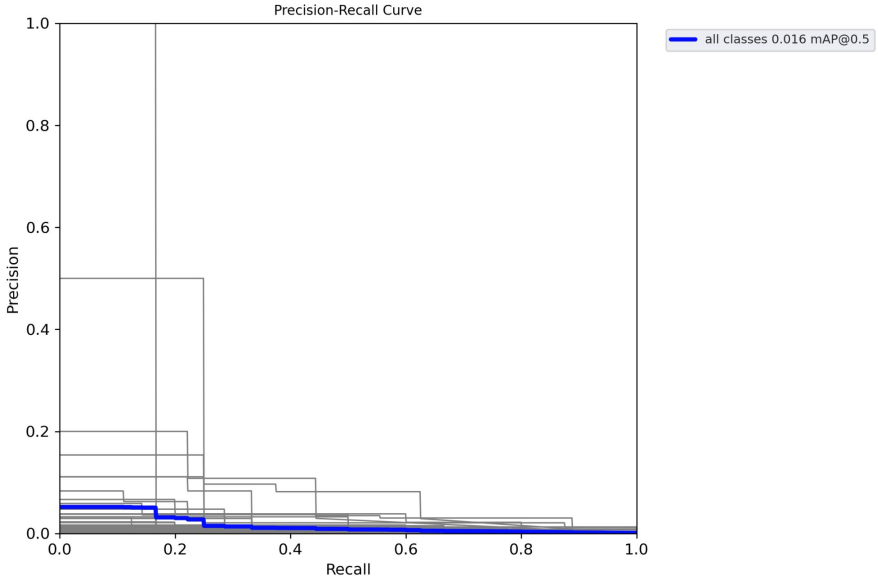
First Training

- 2 Epochs
- 0.8 Training, 0.1 Validation, 0.1 Test



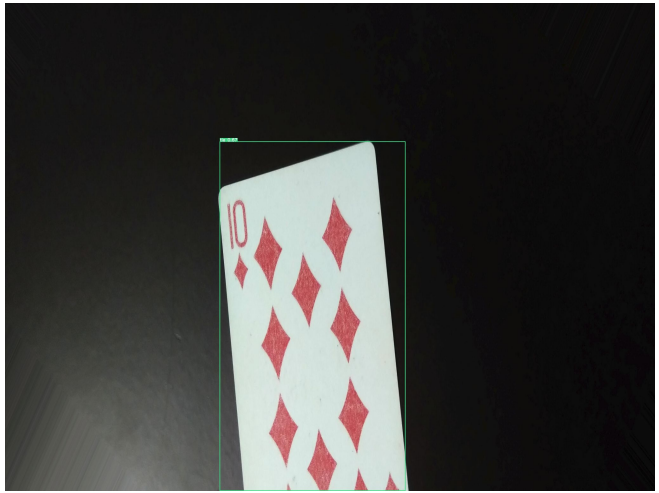



F1-Confidence Curve
all classes 0.01 at 0.000
F1
Confidence
0.0
0.2
0.4
0.6
0.8
1.0


Precision-Recall Curve
all classes 0.016 mAP@0.5
Precision
Recall
0.0
0.2
0.4
0.6
0.8
1.0

But at least:

Before Training:



34.06% wrongly classified

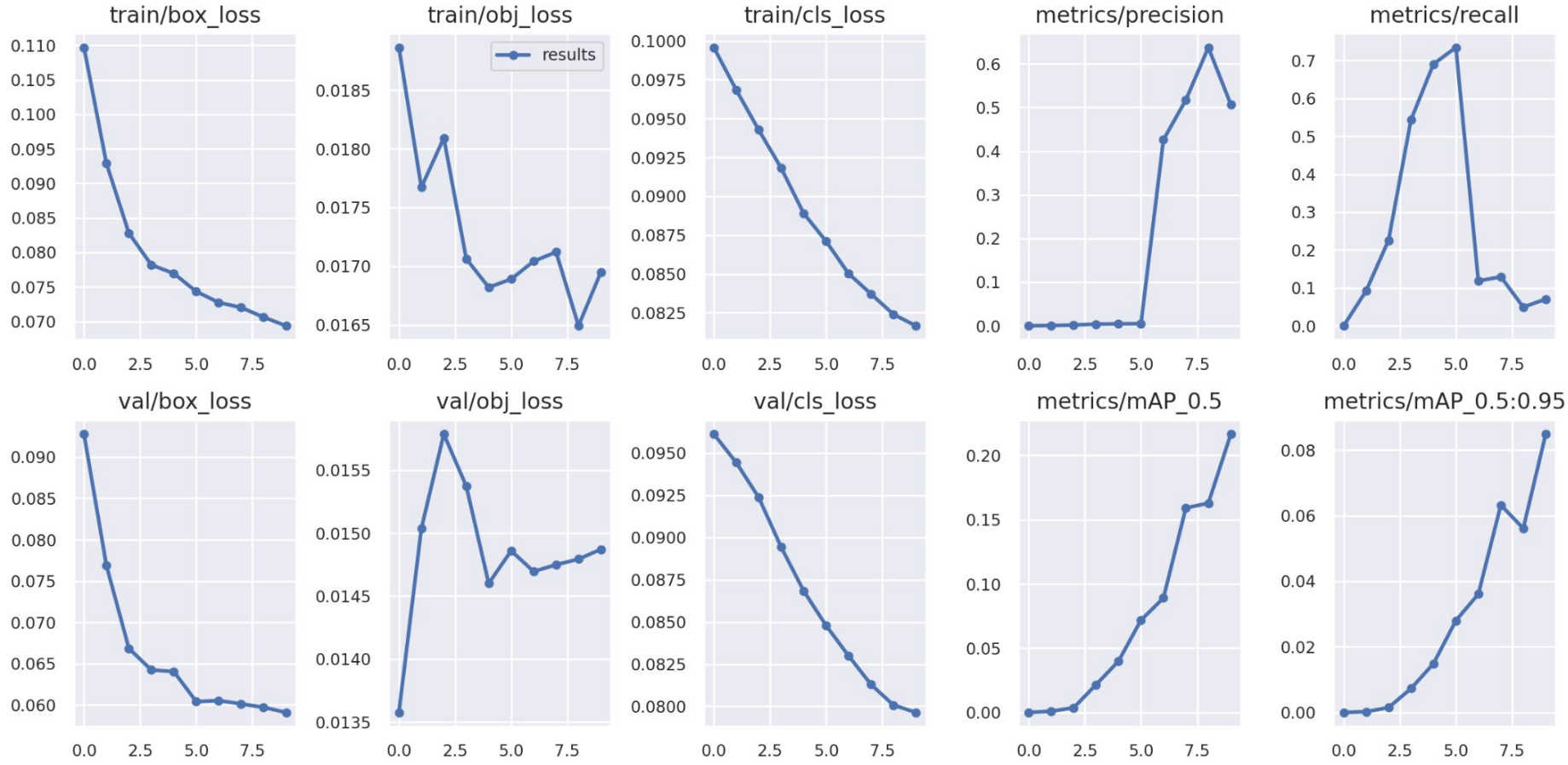
After Training:

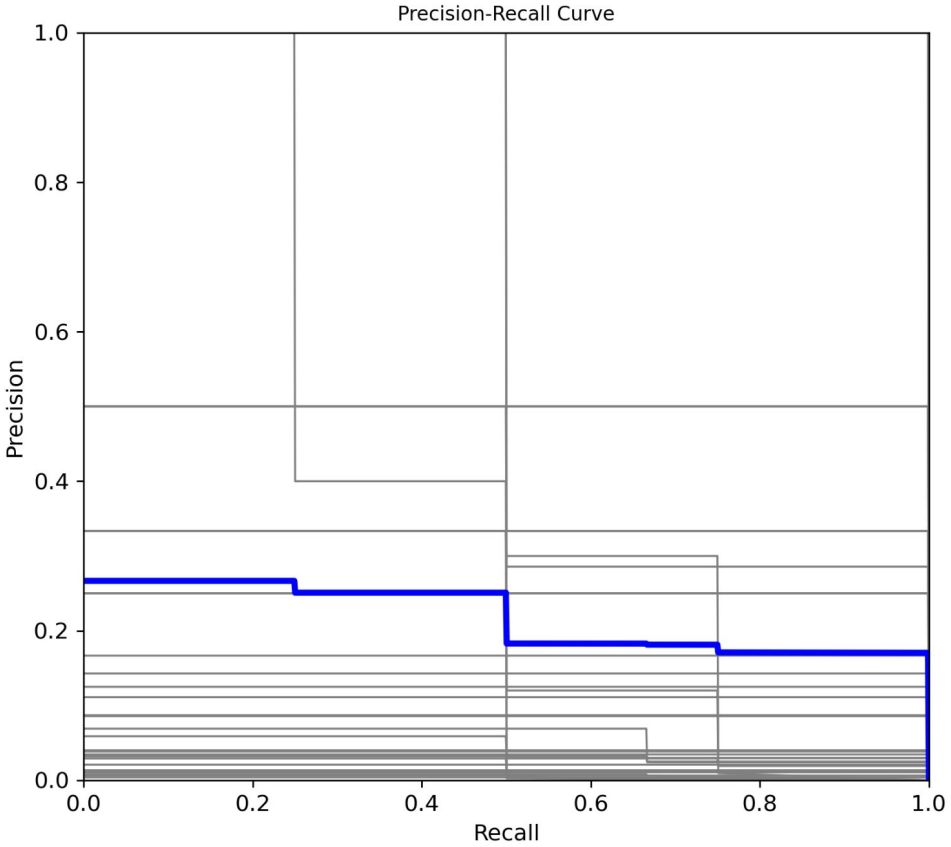
0% wrongly classified

→ no more false positives

Second Training

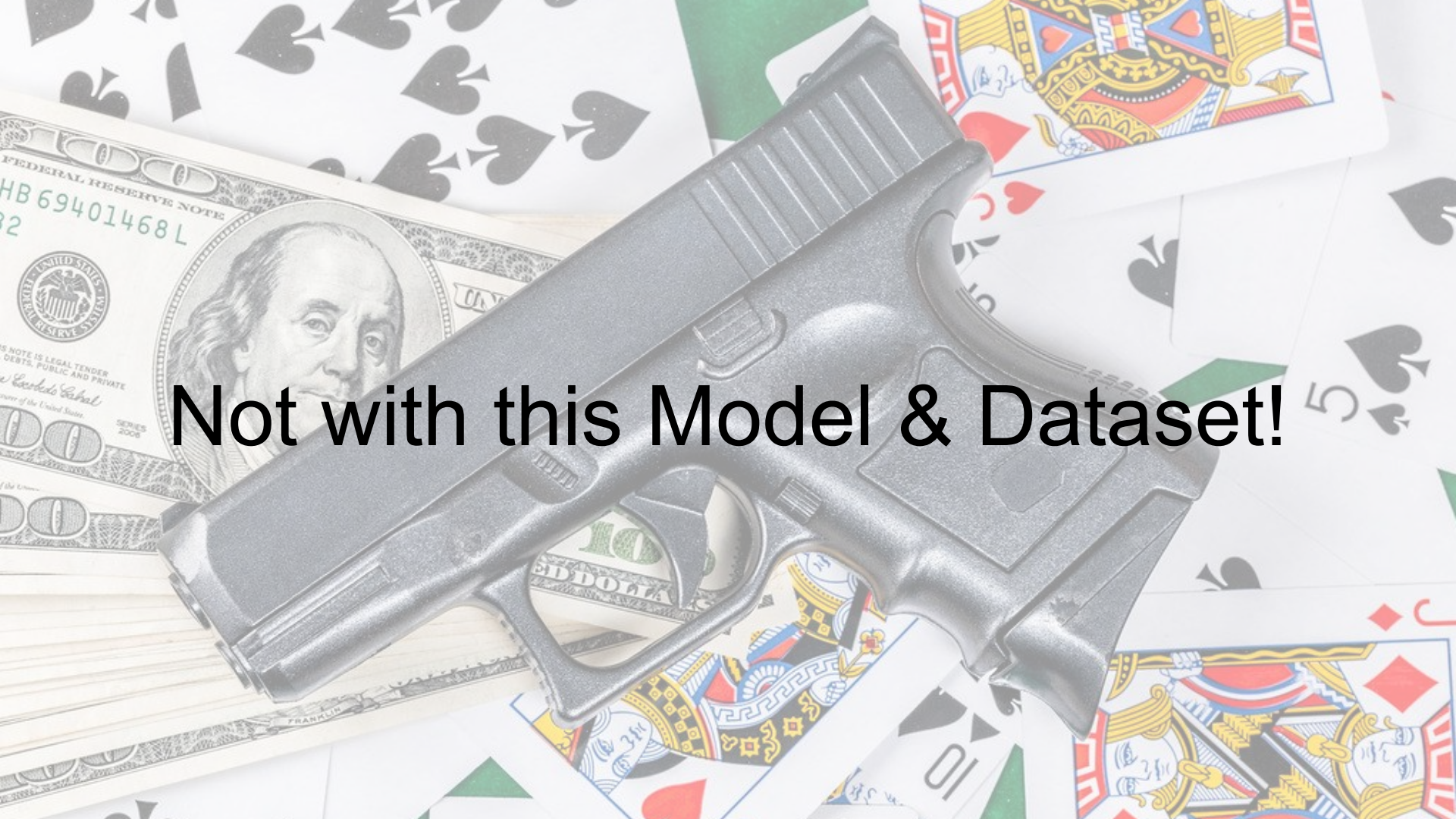
```
Transferred 343/349 items from yolov5s.pt
AMP: checks passed
optimizer: SGD(lr=0.01) with parameter groups 57 weight(decay=0.0), 60 weight(decay=0.0005), 60 bias
train: Scanning D:\Violet\2022Fall\AI\ObjectDetection\dataset\train\labels... 2205 images, 0 backgrounds, 0 corrupt: 10
train: New cache created: D:\Violet\2022Fall\AI\ObjectDetection\dataset\train\labels.cache
train: Caching images (1.9GB ram): 100%| 2205/2205 02:03
Traceback (most recent call last):
  File "<string>", line 1, in <module>
  File "E:\Anaconda3\envs\Object_Detection\lib\multiprocessing\spawn.py", line 116, in spawn_main
    exitcode = _main(fd, parent_sentinel)
  File "E:\Anaconda3\envs\Object_Detection\lib\multiprocessing\spawn.py", line 126, in _main
    self = reduction.pickle.load(from_parent)
MemoryError
```


train/box_loss
train/obj_loss
results
train/cls_loss
metrics/precision
metrics/recall
val/box_loss
val/obj_loss
val/cls_loss
metrics/mAP_0.5
metrics/mAP_0.5:0.95



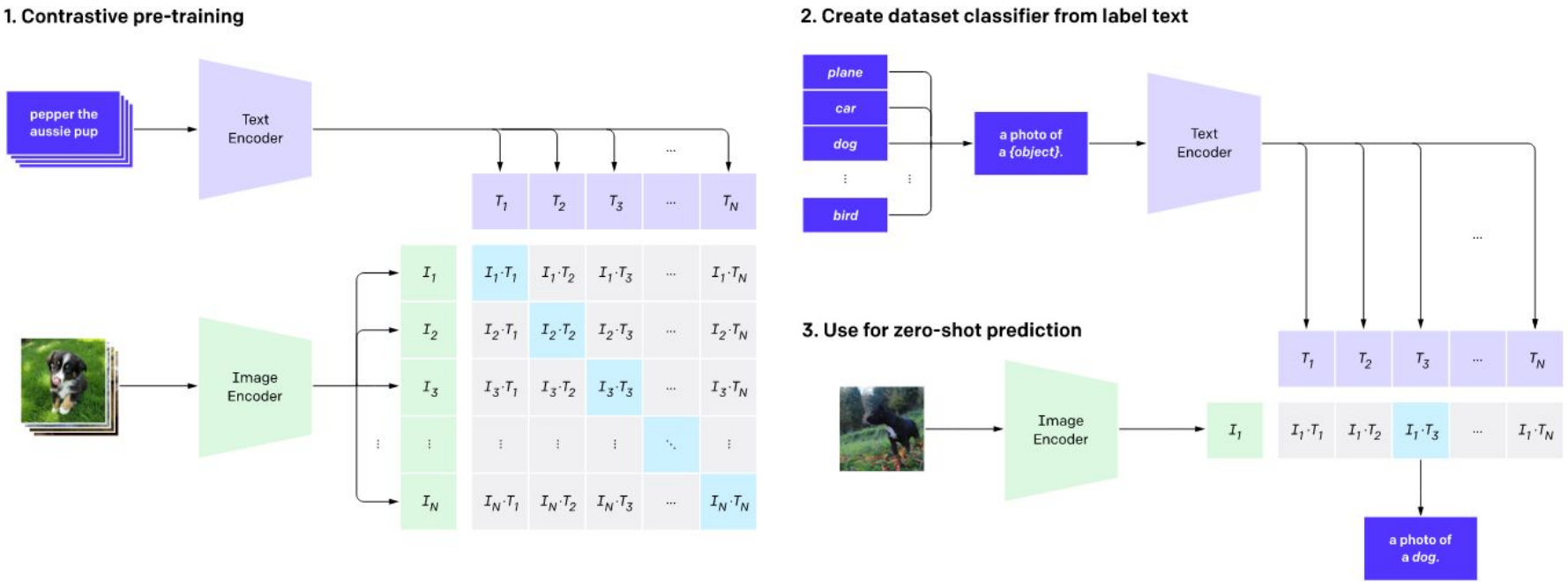
Class	Images	Instances	P	R	mAP50
all	64	64	0.507	0.0706	0.216
AC	64	2	1	0	0.496
AD	64	3	1	0	0.0853
AH	64	1	0	0	0.249
2S	64	2	0.519	0.5	0.662
2D	64	2	1	0	0.00446
2H	64	1	1	0	0.0117
3S	64	1	0	0	0.0343
3H	64	1	0.295	1	0.497
4S	64	2	1	0	0.0865
4C	64	4	1	0	0.376
4D	64	3	0	0	0.0294
4H	64	1	0	0	0.332
5S	64	4	1	0	0.0265
5H	64	2	0	0	0.497
6C	64	1	0	0	0.0398
6D	64	1	1	0	0.0207
6H	64	1	0	0	0.111
7S	64	3	0	0	0.0541
7D	64	1	1	0	0.995
8S	64	2	0	0	0.166
8C	64	1	0	0	0.0106
8H	64	2	0	0	0.0168
9S	64	2	1	0	0.39
9C	64	1	0	0	0.124
9D	64	3	0	0	0.00626

Can we do better?

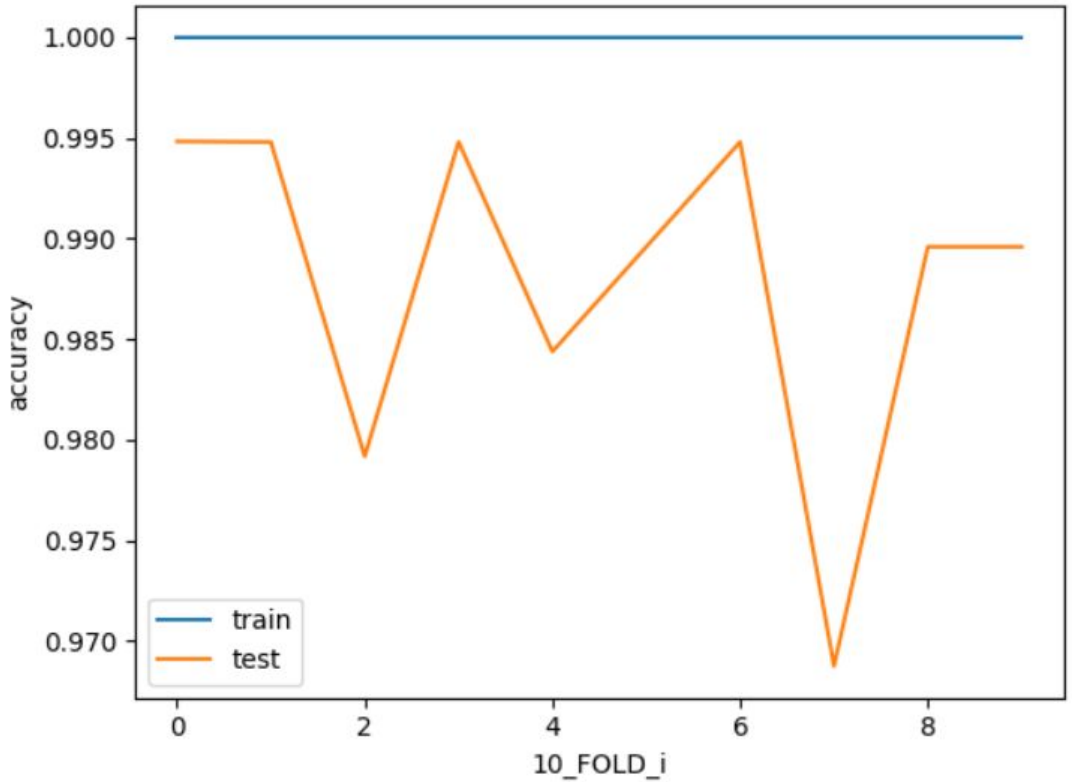

Not with this Model & Dataset!

Raise

CLIP


1. Contrastive pre-training
pepper the aussie pup
Text Encoder
Image Encoder
2. Create dataset classifier from label text
plane
car
dog
bird
a photo of a {object}.
Text Encoder
3. Use for zero-shot prediction
Image Encoder
a photo of a dog.

Classification using CLIP



Final Project Idea:
Representation Learning Algorithms Evaluation

Methods

Method

ALIGN
Scaling Up Visual and Vision-Language Representation Learning With Noisy Text Supervision

CLIP
Learning Transferable Visual Models From Natural Language Supervision

LXMERT
LXMERT: Learning Cross-Modality Encoder Representations from Transformers

ViLBERT
ViLBERT: Pretraining Task-Agnostic Visiolinguistic Representations for Vision-and-Language Tasks

VisualBERT
VisualBERT: A Simple and Performant Baseline for Vision and Language

OSCAR
Oscar: Object-Semantics Aligned Pre-training for Vision-Language Tasks

ViLT
ViLT: Vision-and-Language Transformer Without Convolution or Region Supervision

OFA
OFA: Unifying Architectures, Tasks, and Modalities Through a Simple Sequence-to-Sequence Learning Framework

ALBEF
Align before Fuse: Vision and Language Representation Learning with Momentum Distillation

Browse SoTA > Computer Vision

Computer Vision

3396 benchmarks • 1080 tasks • 2290 datasets • 29384 papers with code

3D

- Semantic Segmentation — 177 benchmarks, 3367 papers with code
- Image Classification — 389 benchmarks, 2759 papers with code
- 3D Reconstruction — 19 benchmarks, 331 papers with code
- Face Verification — 23 benchmarks, 104 papers with code
- Neural Rendering — 75 papers with code

See all 35 tasks

Image Classification

- Image Classification — 389 benchmarks, 2759 papers with code
- Knowledge Distillation — 3 benchmarks, 714 papers with code
- OOD Detection — 164 papers with code
- Few-Shot Image Classification — 95 benchmarks, 155 papers with code
- Fine-Grained Image Classification — 35 benchmarks, 128 papers with code

See all 26 tasks

Semantic Segmentation

- Semantic Segmentation — 177 benchmarks, 3367 papers with code
- Tumor Segmentation — 1 benchmark, 150 papers with code
- Panoptic Segmentation — 15 benchmarks, 114 papers with code
- 3D Semantic Segmentation — 11 benchmarks, 109 papers with code
- Weakly-Supervised Semantic Segmentation — 3 benchmarks, 92 papers with code

See all 21 tasks