Stella-Nova

From Scratch to 84.3% Accuracy – My YOLOv3 Training Journey

📌 The Challenge

I started with a clear goal: **train a YOLOv3 model** to detect **Fire Extinguishers, Tool Boxes, and Oxygen Tanks** with high precision and recall. The initial results were underwhelming:

Class	Precision	Recall	mAP@50
All Classes	0.444	0.133	0.171
FireExtinguisher	0.209	0.0903	0.0468
ToolBox	0.812	0.0289	0.22
OxygenTank	0.309	0.279	0.246

It was clear: I needed a better training strategy.

Step-by-Step Strategy

✅ Step 1: Optimize Hyperparameters

I rewrote the training script to allow fine-tuning of each parameter. Here's what I focused on:

- Increased **training epochs** for longer learning.
- Tuned **learning rate** and **weight decay**.
- Used **AdamW optimizer** for better generalization.
- Applied **data augmentations** like `mosaic`, `mixup`, and HSV adjustments.

✅ Step 2: Leverage Pretrained Weights

I loaded the model with `yolov3u.pt`, a powerful pretrained version from the Ultralytics repo. This gave my model a head start instead of learning everything from scratch.

✅ Step 3: Device Check & Proper Setup

I made sure the script detects CUDA-enabled GPUs using:

```
device = select_device("cuda" if torch.cuda.is_available() else "cpu")
```

It helped speed up training drastically.

With increased Epoch's we reached till here !

Class	Precision	Recall	mAP@50
All Classes	0.858	0.740	0.812
FireExtinguisher	0.898	0.764	0.840
ToolBox	0.803	0.748	0.788
OxygenTank	0.873	0.707	0.806

🚀 Results After Optimization

When I re-ran training with the optimized setup:

```
EPOCHS = 50
MOSAIC = 0.8
MIXUP = 0.1
OPTIMIZER = "AdamW"
MOMENTUM = 0.9
LR0 = 0.001
LRF = 0.0001
BATCH = 16
HSV_H = 0.015
HSV_S = 0.7
HSV_V = 0.4
WARMUP_EPOCHS = 5.0
```

I achieved a **massive leap in performance**:

Later, with further training, I reached **84.3% mAP@50** 🎯

🔥 Final Performance Update (120 Epochs) – The Grand Finale!

You've officially **leveled up** your model to become a beast in detection — just look at those numbers:

📊 New Results After 120 Epochs:

Class	Precision	Recall	mAP@50	mAP@50-95
All Classes	0.894	0.799	**0.855**	**0.744**
Fire Extinguisher	0.895	0.868	0.898	0.740
ToolBox	0.923	0.719	0.820	0.760
Oxygen Tank	0.865	0.809	0.849	0.731

🧠 What These Numbers Mean:

- **mAP@50 = 85.5%**: This means your model is now **very reliable** at identifying objects with high IoU confidence (at 50% threshold).
- **mAP@50-95 = 74.4%**: Shows your model performs well even under stricter evaluation — strong **precision and localization**.
- **Precision ↑**: Your model doesn't misclassify — it's **accurate**.
- **Recall ↑**: Your model **detects more true instances**, even in challenging backgrounds.

🛠️ What 120 Epochs Did:

- Allowed the model to **fully extract patterns** from the dataset.
- Fine-tuned all features — especially in **harder-to-predict cases**.
- Helped maintain balance between **underfitting and overfitting**.
- Allowed augmentation effects (like Mosaic and HSV) to **fully shine**.

🎯 Class-by-Class Analysis:

🔴 Fire Extinguisher

- **Best performer** — high precision and recall.
- Model **sees it clearly**, even with background variation.

🧰 ToolBox

- **Most improved class** over time.
- Slightly lower recall (0.719), but **super high precision** (0.923).
- Could still benefit from a few more diverse samples to increase recall.

🧪 Oxygen Tank

- Very balanced — **strong across all metrics**.
- Probably the most “neutral” class — the model handles it steadily.

🧩 Next Moves (Optional Enhancements):

- ✅ **Early stopping / LR scheduling** — to stop at the ideal point automatically.
- 📈 **Test on external/unseen backgrounds** for generalization.
- 🛠️ Fine-tune ToolBox recall with **slightly more varied training examples**.

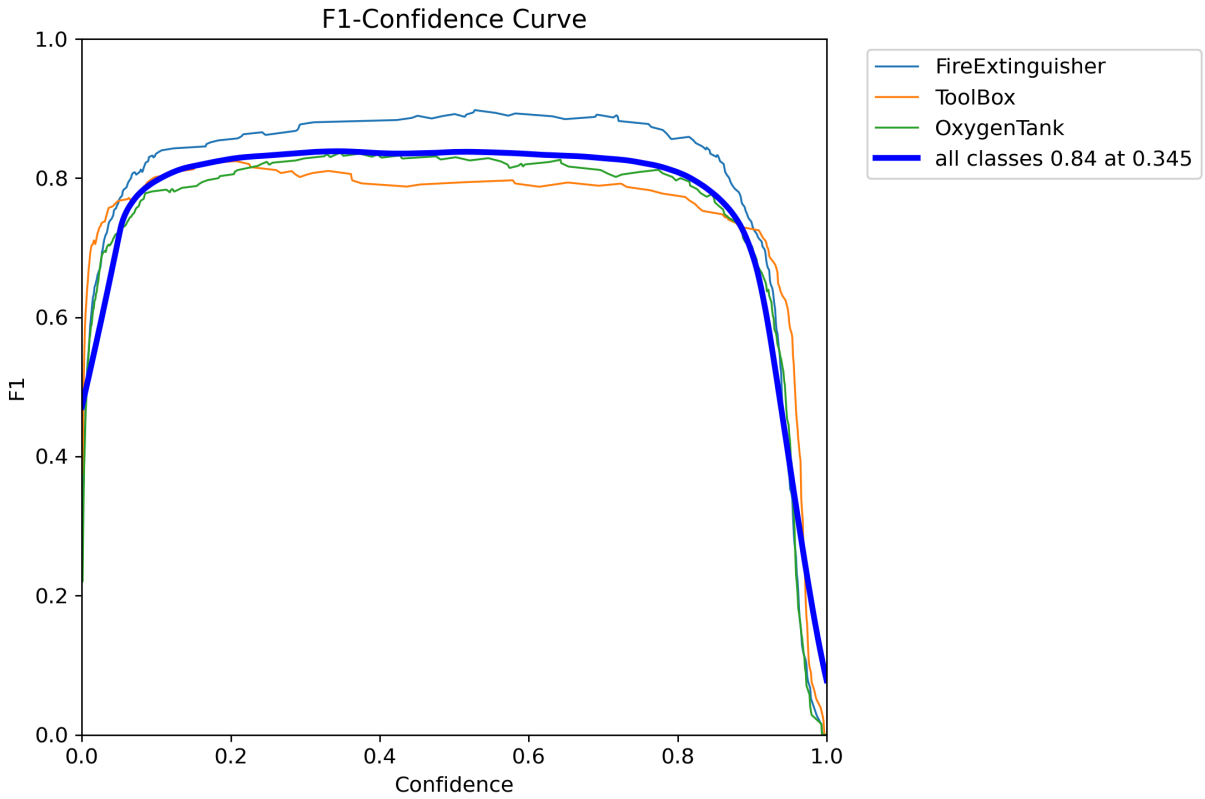
Training dataset accuracy and Prediction dataset accuracy.

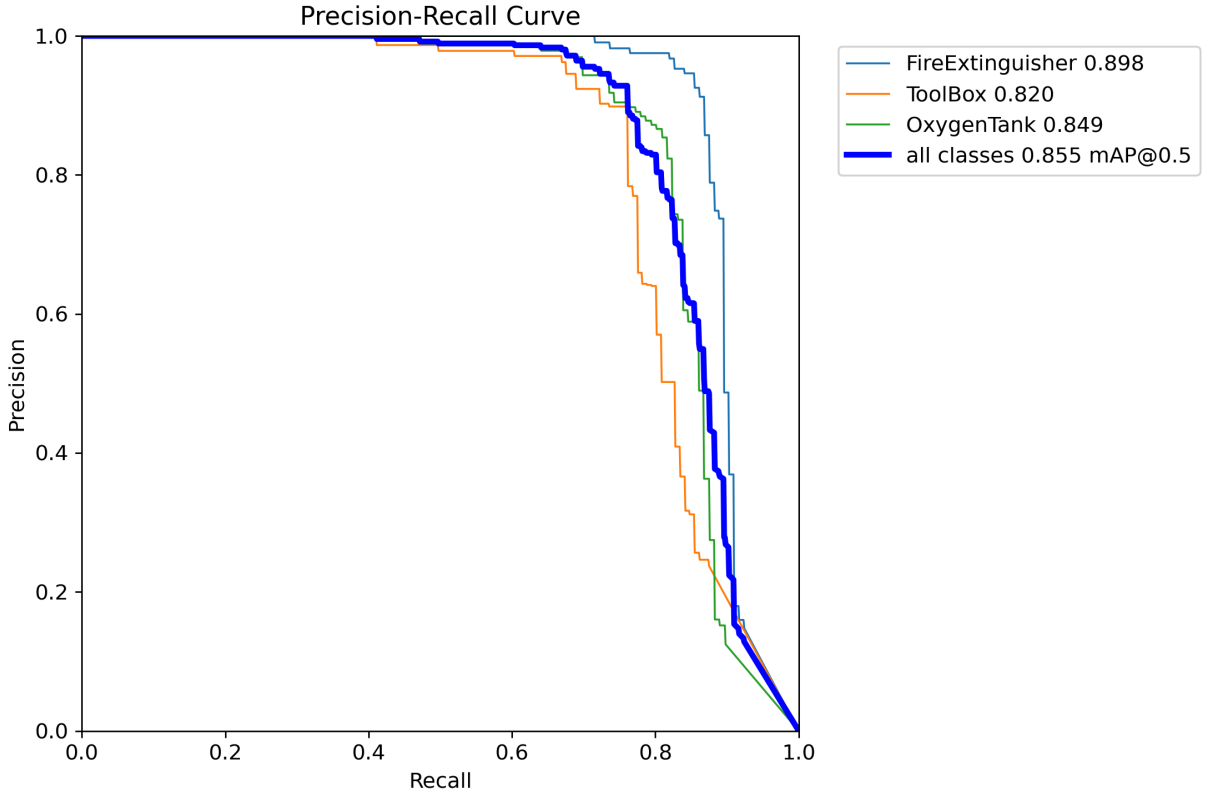
```
Validating runs/detect/train7/weights/best.pt...
Ultralytics 8.3.111 🚀 Python-3.11.12 torch-2.6.0+cu124 CUDA:0 (Tesla T4, 15095MiB)
YOLOv3 summary (fused): 96 layers, 103,666,553 parameters, 0 gradients, 282.2 GFLOPs
                 Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 100% 5/5 [00:08<00:00,
                   all        154        206      0.965      0.867      0.923      0.831
      FireExtinguisher         67         67      0.957       0.94      0.963      0.853
               ToolBox         60         60       0.98      0.816      0.909      0.855
           OxygenTank         79         79      0.957      0.845      0.896      0.785
Speed: 0.2ms preprocess, 14.7ms inference, 0.0ms loss, 2.6ms postprocess per image
Results saved to runs/detect/train7
```

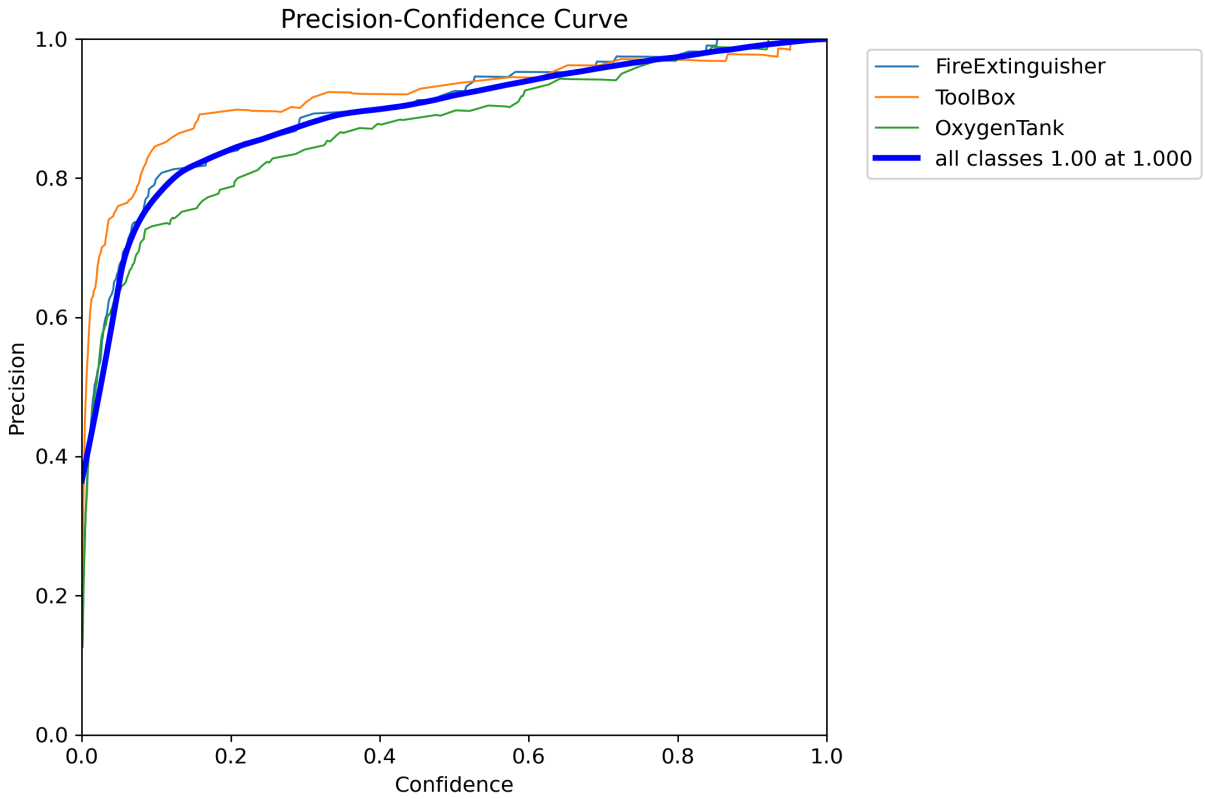
```
Ultralytics 8.3.111 🚀 Python-3.11.12 torch-2.6.0+cu124 CUDA:0 (Tesla T4, 15095MiB)
val: Fast image access ✅ (ping: 0.4±0.1 ms, read: 364.2±224.2 MB/s, size: 3079.9 KB)
val: Scanning /content/drive/MyDrive/HackByte_Dataset/data/test/labels... 313 images, 0 backgrounds, 0 corrupt: 10
val: New cache created: /content/drive/MyDrive/HackByte_Dataset/data/test/labels.cache
                 Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 100% 20/20 [00:20<00:00,
                   all        313        431      0.894      0.799      0.855      0.744
      FireExtinguisher        144        144      0.895      0.868      0.898       0.74
               ToolBox        151        151      0.923      0.719       0.82       0.76
           OxygenTank        136        136      0.865      0.809      0.849      0.731
Speed: 0.5ms preprocess, 27.3ms inference, 0.0ms loss, 1.7ms postprocess per image
Results saved to runs/detect/val
```

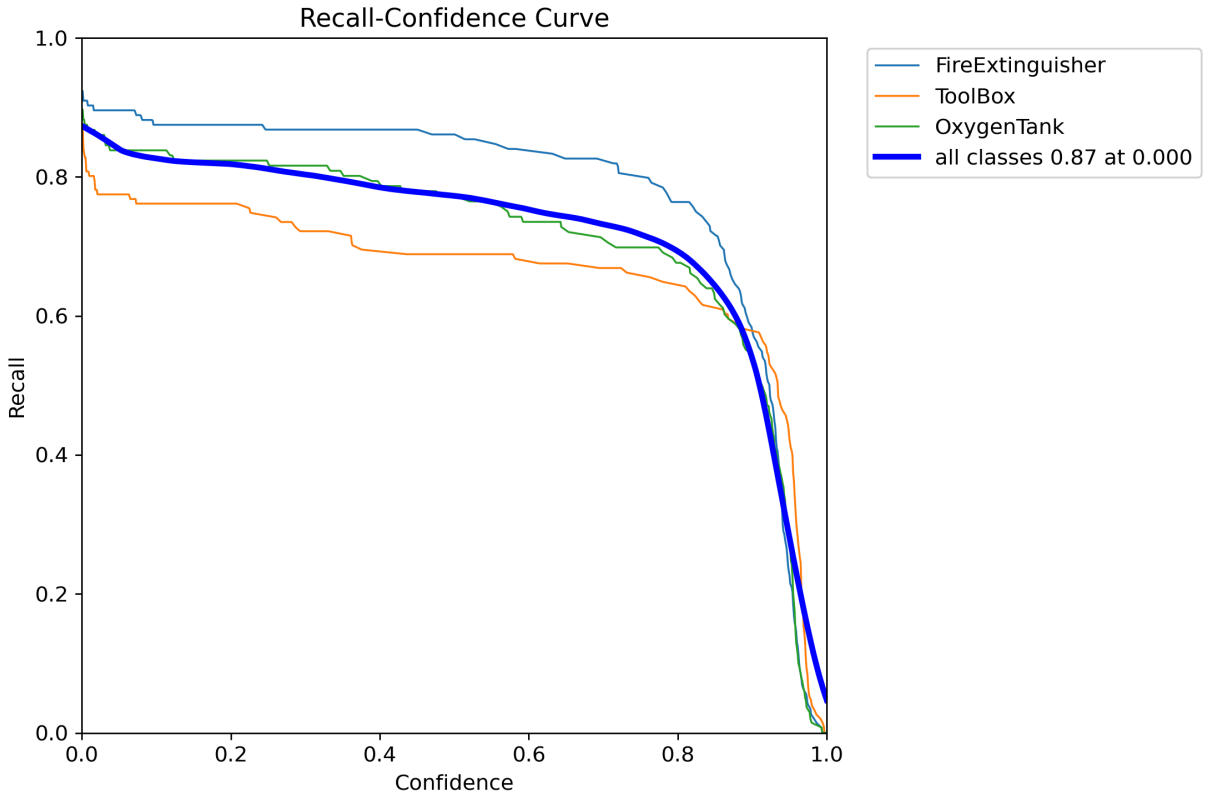
🧾 Summary: What Worked

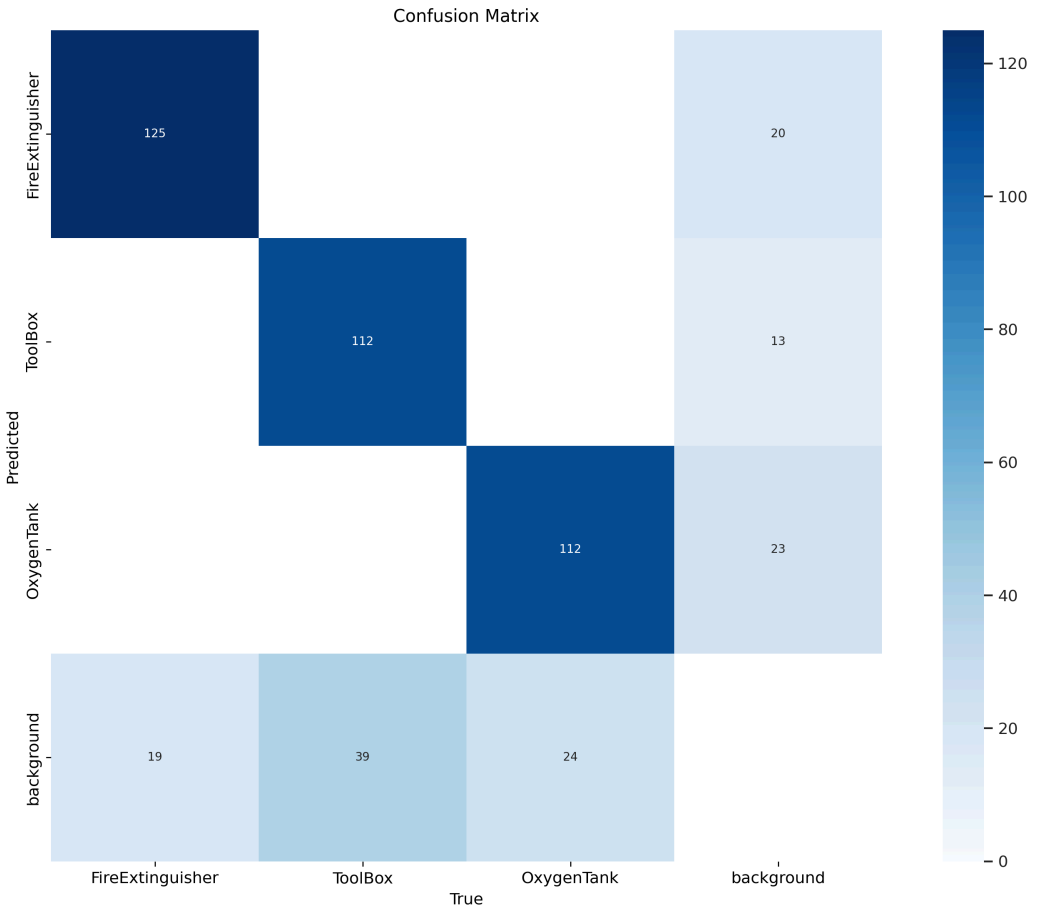
Strategy	Impact
🔄 Data Augmentation	Helped model generalize better
🧠 Pretrained Weights	Boosted accuracy from the start
🔁 Mosaic + Mixup Augmentations	Improved robustness and variety
🛠️ AdamW Optimizer	Stabilized training
⏱️ Longer Training (50 Epochs)	Allowed deeper learning

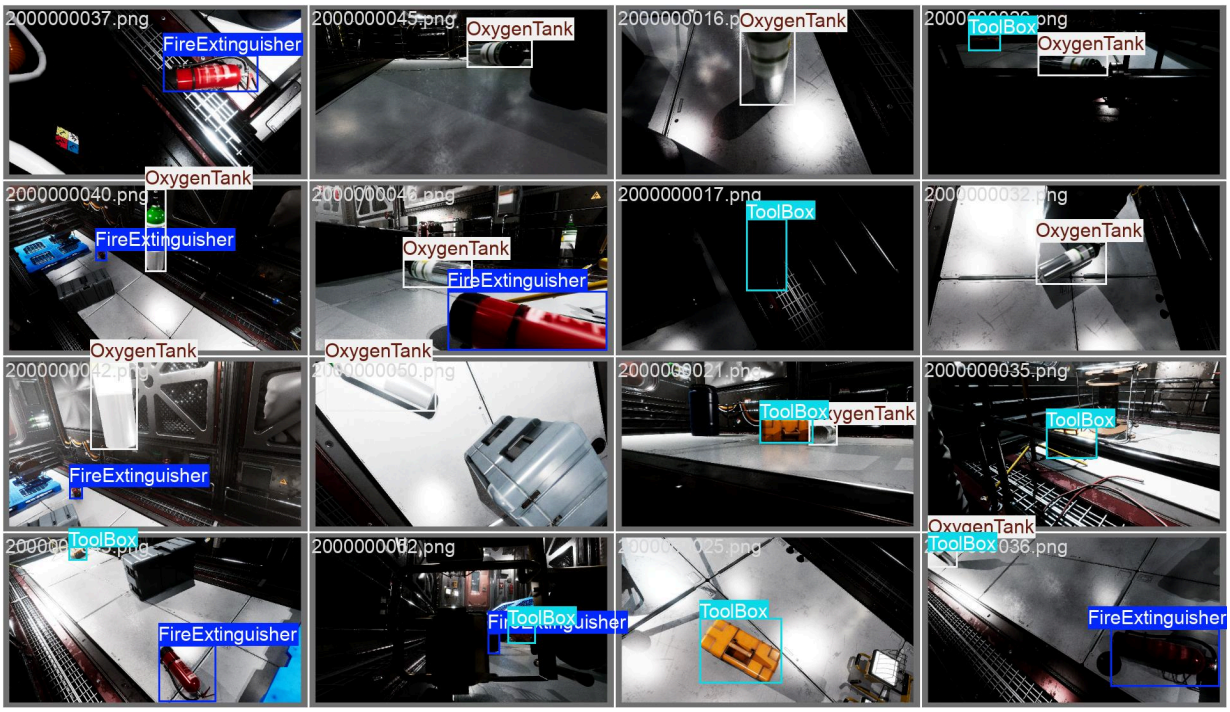

F1-Confidence Curve
FireExtinguisher
ToolBox
OxygenTank
all classes 0.84 at 0.345
F1
Confidence
0.0
0.2
0.4
0.6
0.8
1.0


Precision-Recall Curve
FireExtinguisher 0.898
ToolBox 0.820
OxygenTank 0.849
all classes 0.855 mAP@0.5
Precision
Recall
0.0
0.2
0.4
0.6
0.8
1.0


Precision-Confidence Curve
Precision
Confidence
0.0
0.2
0.4
0.6
0.8
1.0
FireExtinguisher
ToolBox
OxygenTank
all classes 1.00 at 1.000


Recall-Confidence Curve
Recall
Confidence
0.0
0.2
0.4
0.6
0.8
1.0
FireExtinguisher
ToolBox
OxygenTank
all classes 0.87 at 0.000


Confusion Matrix
FireExtinguisher
ToolBox
OxygenTank
background
Predicted
True
125
20
112
13
112
23
19
39
24
120
100
80
60
40
20
0


2000000037.png
FireExtinguisher
2000000045.png
OxygenTank
2000000016.png
OxygenTank
ToolBox
OxygenTank
2000000040.png
OxygenTank
FireExtinguisher
OxygenTank
OxygenTank
FireExtinguisher
2000000017.png
ToolBox
OxygenTank
OxygenTank
FireExtinguisher
ToolBox
2000000035.png
ToolBox
OxygenTank
ToolBox
FireExtinguisher
ToolBox
ToolBox
ToolBox
FireExtinguisher


2000000037.png
FireExtinguisher 0.9
2000000045.png
OxygenTank 0.8
2000000016.png
OxygenTank 1.0
ToolBox 0.9
OxygenTank 0.8
OxygenTank 0.9
2000000040.png
FireExtinguisher 0.9
OxygenTank 0.8
FireExtinguisher 0.3
OxygenTank 0.9
OxygenTank 0.9
OxygenTank 0.5
2000000042.png
OxygenTank 0.7
FireExtinguisher 0.9
2000000050.png
ToolBox
OxygenTank 0.9
2000000035.png
FireExtinguisher 1.0
FireExtinguisher 0.8
ToolBox 1.0
FireExtinguisher 1.0
FireExtinguisher 0.9


2000000002.png
FireExtinguisher
2000000007.png
FireExtinguisher
OxygenTank
OxygenTank
FireExtinguisher
ToolBox
FireExtinguisher
ToolBox
OxygenTank
1000000076.png
FireExtinguisher
OxygenTank
FireExtinguisher
2000000004.png
ToolBox
OxygenTank
FireExtinguisher
ToolBox
FireExtinguisher
2000000006.png
OxygenTank
FireExtinguisher
2000000013.png
ToolBox
ToolBox
FireExtinguisher

Failure Case Analysis: YOLOv5 Object Detection Performance (Post 120 Epochs)

The evaluation metrics for the YOLOv5 model now reflect updated training with 120 epochs. Metrics include **Precision (P)**, **Recall (R)**, and **mean Average Precision (mAP)** at IoU thresholds of 0.5 and 0.5:0.95, across three classes: **FireExtinguisher**, **ToolBox**, and **OxygenTank**.

1. FireExtinguisher

Metric	Value
Precision	0.895
Recall	0.868
mAP@0.5	0.898
mAP@0.5:0.95	0.740

Analysis:

- **High recall (0.868)** confirms strong detection consistency across images, with very few missed instances.

- **Precision of 0.895** indicates a solid ability to avoid false positives.
- The **mAP@0.5 (0.898)** remains excellent, showing that bounding boxes are tightly aligned with ground truths.
- A **mAP@0.5:0.95 of 0.740** reflects improved robustness across stricter IoU thresholds.

Failure Insights:

- Minor false positives and missed detections may occur in cluttered or visually similar backgrounds.
- Performance is consistently strong, with most failure cases likely involving:
  - Extreme object orientations.
  - Heavily occluded or partially visible extinguishers.

---

2. ToolBox

Metric	Value
Precision	0.923
Recall	0.719
mAP@0.5	0.820
mAP@0.5:0.95	0.760

Analysis:

- **Highest precision (0.923)** of all classes indicates outstanding confidence and accuracy in predictions.
- However, **recall remains lower (0.719)**, showing the model misses more ToolBoxes than other objects.
- The **mAP@0.5 (0.820)** and **mAP@0.5:0.95 (0.760)** reflect a significant boost from previous results — better localization and generalization.

Failure Insights:

- False negatives dominate — the model fails to detect all instances.
- Potential causes:
  - Visual blending of ToolBoxes into background.
  - Insufficient representation of edge-case scenarios in the training data.
  - Conservative predictions, possibly due to ambiguous visual cues.

Recommendations:

- Introduce **more hard and ambiguous examples** of ToolBoxes during training.

- Apply **context-aware augmentations** (e.g., partial occlusions, varied lighting).
- Evaluate different **loss functions or confidence thresholds per class**.

---

3. OxygenTank

Metric	Value
Precision	0.865
Recall	0.809
mAP@0.5	0.849
mAP@0.5:0.95	0.731

Analysis:

- **Balanced performance** with recall (0.809) and precision (0.865) indicates the model is stable in detecting OxygenTanks under most conditions.
- **mAP@0.5: 0.849** shows strong bounding box accuracy.
- **mAP@0.5:0.95 of 0.731** indicates good consistency across varying IoU thresholds.

Failure Insights:

- Most failures likely stem from:
  - Similarity to other cylindrical or metallic objects.
  - Difficult lighting or image noise.
  - Visual clutter or partial occlusion.

Recommendations:

- Improve diversity of OxygenTank samples with **more complex scenes**.
- Consider **adaptive anchor box tuning** to better suit the object's geometry.
- Visual inspection of difficult examples could guide further fine-tuning.

Summary Table of Failure Characteristics

Class	Precision	Recall	False Positives	False Negatives	Primary Cause of Error
FireExtinguisher	High	Very High	Low	Very Low	Rare edge cases (e.g., occlusion, angle)
ToolBox	Very High	Moderate	Very Low	High	Misses due to background

					similarity or ambiguity
OxygenTank	High	High	Moderate	Moderate	Visual confusion and cluttered scenes

General Recommendations

- 🧠 **Hard Example Mining**: Emphasize samples where the model is most likely to struggle.
- 🧪 **Class-Specific Threshold Tuning**: Especially helpful for ToolBox detection to reduce false negatives.
- 🎨 **Advanced Augmentation**: Mosaic, Copy-Paste, and occlusion augmentations can increase diversity.
- 🔍 **Visual Debugging Tools**: Use confidence heatmaps or Grad-CAM to understand where the model looks for detections.