

Fine-Tuning Object Detection Models for Syringe Detection in Pediatric Medication Preparation Video

Viktor Krastev · CSE3000, TU Delft · Supervisors: Nergis Tömen, Xucong Zhang · Data: Erasmus MC

1 MOTIVATION

Medication preparation errors are a leading cause of pediatric patient harm. Manual preparation steps remain visually unverified — automated detection could close this gap.

RQ: How accurately can pretrained models be fine-tuned to detect syringes in medication preparation video?

2 DATASET

56 videos	5,760 frames	70/15/15 split
---------------------	------------------------	--------------------------

Video-level split prevents data leakage. Negative examples reduce background bias.

3 MODELS COMPARED

YOLO11n	YOLO26n
C3k2 + C2PSA attention DFL · TAL · 2.6M params NMS required	NMS-free · ProgLoss STAL · MuSGD · 2.4M 43% faster on CPU

Nano variants — edge deployment + avoids overfitting on small datasets

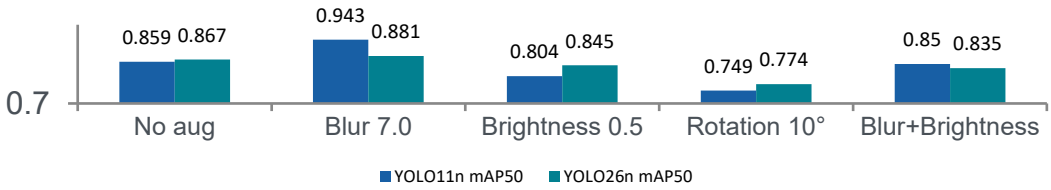
4 HYPERPARAMETER SEARCH

50 random trials on 20% data fraction (Bergstra & Bengio 2012) Shleifer & Prokop: 50% of data explains 95% of variance. YOLO11n dominated — 12 of top 15 configs.

lr: 0.001–0.1	wd: 4 values	imgsz: 320/640
blur: 0–10	bright: 0–0.6	rot: 0–20°

Best on 20% data ≠ best on full dataset — proxy search limitation (Shleifer & Prokop)

5 AUGMENTATION EXPERIMENTS



6 Hyperparameter Search TOP RESULTS (20% DATA FRACTION)

Model	Rank	P	R	mAP50	mAP50-95
YOLO11n	1	0.929	0.828	0.930	0.577
YOLO11n	2	0.866	0.628	0.779	0.415
YOLO11n	4	0.735	0.736	0.763	0.308
YOLO11n	5	0.855	0.715	0.760	0.300
YOLO26n	9	0.668	0.702	0.746	0.379

7 Axis-aligned BOUNDING BOX RESULTS (TEST SET)

Model	Precision	Recall	mAP50	mAP50-95
YOLO11n	0.832	0.857	0.859	0.551
YOLO11n+blur	0.927	0.911	0.943	0.542
YOLO26n	0.846	0.854	0.867	0.473
YOLO26n tuned	0.853	0.903	0.878	0.475

YOLO11n wins on both mAP50 and recall for standard detection.

9 Optimizer Experiment Results (AdamW vs Default)

Model, Optimizer	Ep	P	R	mAP50	mAP50-95
YOLO11n-obb SGD	50	0.998	0.997	0.995	0.820
YOLO11n-obb AdamW	50	0.989	0.977	0.993	0.729
YOLO26n-obb MuSGD	80	0.934	0.974	0.987	0.805
YOLO26n-obb AdamW	80	0.979	0.973	0.994	0.781

8 ORIENTED BOUNDING BOX RESULTS (TEST SET)

Model	Epoch	Aug	P	R	mAP50	mAP50-95
YOLO11n-obb	50	None	0.998	0.997	0.995	0.820
YOLO11n-obb	100	None	0.998	0.999	0.994	0.785
YOLO11n-obb	100	Blur	0.999	0.999	0.995	0.715
YOLO26n-obb	100	None	0.960	0.959	0.989	0.766
YOLO26n-obb	100	Blur	0.947	0.975	0.991	0.833
YOLO26n-obb	150	None	0.959	0.869	0.969	0.800

10 Image Resolution Results

Model	Ep	Imgsz	P	R	mAP50	mAP50-95
YOLO11n-obb	50	640	0.998	0.997	0.995	0.820
YOLO11n-obb	50	1024	0.998	0.998	0.995	0.828
YOLO26n-obb	80	640	0.934	0.974	0.987	0.805
YOLO26n-obb	80	1024	0.976	0.977	0.993	0.833

11 FAILURE CASES

- Keyboard/floor as syringe — background bias (fixed by adding negatives)
- Medicine vials — shape similarity (cylindrical + text markings)
- Duplicate detections — NMS limitation (fixed in YOLO26n-obb)

12 LIMITATIONS

- Single clinical environment and little data variety
- Single annotator
- No external test set

13 Conclusion

YOLO11n-obb	YOLO26n-obb
Best mAP50 0.995	Best mAP50-95 0.833
OBB dramatically outperforms standard boxes. YOLO26n has the best mAP50-95, but YOLO11n-obb is the best overall, despite having mAP50-95 0.828.	
KEY RESULT	
0.011 pretrained mAP50	→ ~90x
	0.995 best OBB mAP50