

Noto Sans

Gloved vs Bare Hand Detection

End-to-End Safety Compliance Pipeline Built 100% from Scratch

Rayavarapu Hemanth Kumar

Ai Engineer — Currently leading real-time edge CV systems

November 27, 2025

Final Technical Report – Part 1 Submission

No public datasets — No pre-trained annotation models — Full lifecycle ownership

Executive Summary

A complete, production-grade safety compliance system has been designed, developed, and deployed from absolute scratch — including raw video capture using only my own hands, frame extraction, manual annotation, custom training, inference engineering, logging, CLI design, and public cloud deployment.

- **Model:** YOLOv8n fine-tuned from COCO (only initial weights — no detection model used for data creation)
- **Dataset:** 100% self-captured, self-annotated (1,600 frames → 503 final images)
- **Performance:** mAP@0.5 = 0.980 — mAP@0.5:0.95 = 0.820
- **Deployment:** Live 24/7 web app → <https://ml-biztech-analytics-hemanth.streamlit.app/>
- **Inference:** 80 FPS (GPU) — 30 FPS (CPU) — Full JSON logging

1 100% In-House Data Collection (Truly from Scratch)

- Raw material: Videos recorded using only my own hands performing real factory-like tasks
- Environments: Indoor/outdoor, low light, bright light, cluttered backgrounds, motion
- Glove types: black rubber,
- Poses: Grasping tools, pointing, holding objects, partial occlusion, overlapping hands
- Frame extraction: Custom OpenCV script → extracted 1,600+ clean frames
- No pre-trained model was used at any point for auto-selection or filtering

2 Manual Annotation Process

- Tool: labellmg (open-source)
- Total bounding boxes drawn: 4,000 by hand
- Classes defined: `GLOVE` → maps to `gloved_hand`, `NO_GLOVE` → `bare_hand`
- Quality control: Every annotation personally reviewed for tightness and correctness
- Final curated training set: 503 highest-quality images (402 train — 101 val)

3 Model Training – Full Ownership

Parameter	Value
Architecture	YOLOv8n (nano)
Initial weights	Official COCO <code>yolov8n.pt</code>
Training resolution	1248 × 1248
Epochs	250
Optimizer	AdamW (lr=0.001)
Scheduler	Cosine Annealing
Augmentations	Mosaic, MixUp, HSV, Flip, Rotation, Scale
Hardware	NVIDIA Tesla T4

Final Validation Metrics

Metric	Score
Precision	0.952
Recall	0.941
mAP@0.5	0.980
mAP@0.5:0.95	0.820

4 Production Inference Pipeline

- `detection_img.py` – For batch of Images
- `detection_live.py` – Real-time RTSP/webcam support
- Exact JSON logging format as specified
- Annotated output images saved
- Class mapping: internal → `gloved_hand` / `bare_hand`

5 Live Cloud Deployment

<https://ml-biztech-analytics-hemanth.streamlit.app/>

- Fully functional public demo (available 24/7)
- Headless OpenCV compatibility resolved (`opencv-python-headless`)
- Upload any image → instant compliance result

6 System Requirements

- Python 3.8+
- ultralytics = 8.2.0
- opencv-python or opencv-python-headless (for Streamlit)
- GPU optional (CPU inference fully supported)

7 Conclusion

This submission is not based on public datasets or automated annotation tools — it is a genuine end-to-end computer vision pipeline built entirely from raw video I recorded myself, through manual annotation, training, production engineering, and live deployment.

The same rigor and ownership I apply daily as a Computer Vision Engineer is reflected here.

Ready to contribute at senior capacity from Day 1.

Thank you for the opportunity.