Discriminative Deep Learning Project

Milestone 2 Report

Multi-Object Detection using YOLOv8

Course: IE 7615 - Discriminative Deep Learning

Group 8: Quoc Hung Le, Hassan Alfareed, Khoa Tran

February 2026

1. MULTI-OBJECT DATASET GENERATION

1.1 Dataset Creation Methodology

Multi-object images were generated by concatenating randomly selected single-object images into grid layouts. Three grid configurations were used to create varying complexity levels:

Grid Type	Dimensions	Objects per Image	Images Generated
2x2 Grid	2 rows x 2 cols	4 objects	400
2x3 Grid	2 rows x 3 cols	6 objects	400
3x3 Grid	3 rows x 3 cols	9 objects	400
Total	-	-	1,200

1.2 Data Augmentation Techniques

To enhance dataset diversity and model robustness, the following augmentation techniques were applied:

- Random Object Selection: Objects randomly sampled from the 39-class single-object dataset
- Position Variation: Objects placed in grid cells with slight random offsets
- Scale Consistency: All objects resized to fit uniformly within grid cells
- Background Preservation: Original object backgrounds retained for realistic scenes

1.3 Dataset Split Distribution

Split	Images	Percentage
Training	919	66.4%
Validation	231	16.7%
Test	233	16.8%
Total	1,383	100%

1.4 YOLO Annotation Format

Each multi-object image has a corresponding annotation file in YOLO format. Each line contains: **class_id x_center y_center width height** (all normalized to 0-1 range).

2. YOLOV8 MODEL TRAINING

2.1 Model Variants

Three YOLOv8 variants were trained using transfer learning from COCO pretrained weights:

Model	Parameters	Speed Priority	Use Case
YOLOv8n (Nano)	3.2M	Fastest	Real-time applications
YOLOv8s (Small)	11.2M	Balanced	General deployment
YOLOv8m (Medium)	25.9M	Accurate	High accuracy needs

2.2 Training Configuration

Parameter	Value	Description
Image Size	672 x 672	Input resolution for detection
Batch Size	8	Optimized for GPU memory
Optimizer	AdamW	Weight decay regularization
Learning Rate	0.0005	Initial learning rate
Epochs	80 (max)	With early stopping (patience=20)
Device	Apple MPS	Metal Performance Shaders GPU
Mosaic Augment	0.2	Probability of mosaic augmentation
IoU Threshold	0.6	For NMS post-processing

2.3 Transfer Learning Approach

All models utilized COCO pretrained weights as initialization, enabling rapid convergence and strong performance despite the relatively small dataset size. The transfer learning process:

- Loaded pretrained YOLOv8 weights trained on COCO dataset (80 classes)
- Modified output layer for 39-class detection task
- Fine-tuned all layers with lower learning rate for backbone
- Applied early stopping to prevent overfitting

3. MODEL PERFORMANCE RESULTS

3.1 Detection Metrics Comparison

All three YOLOv8 variants achieved exceptional performance, significantly exceeding the target mAP50 > 0.85:

Model	mAP50	mAP50-95	Precision	Recall	Epochs
YOLOv8n	98.76%	98.76%	98.50%	97.75%	38
YOLOv8s	98.74%	98.74%	98.26%	98.10%	46
YOLOv8m	98.90%	98.90%	97.70%	98.50%	17

Note: YOLOv8m achieved the best overall mAP50 performance on test set. YOLOv8m is used EarlyStoping =5 while YOLOv8s (8), YOLOv8n (10) as trade-off traning time.

3.2 Key Findings

- All models exceeded target: mAP50 > 98% vs. target of 85%, a 13+ percentage point improvement
- YOLOv8m best performer: Achieved 98.90% mAP50 with highest recall (98.50%)
- Fast convergence: YOLOv8m converged in just 17 epochs, YOLOv8n in 38 epochs
- Consistent metrics: Minimal gap between mAP50 and mAP50-95, indicating robust detections
- High recall: All models achieved >97% recall, missing very few objects

3.3 Training Convergence Analysis

Training loss curves showed smooth convergence for all models:

Metric	YOLOv8n (Final)	YOLOv8s (Final)	YOLOv8m (Final)
Box Loss	0.058	0.041	0.069
Class Loss	0.161	0.126	0.199
DFL Loss	0.804	0.811	0.813
Val Box Loss	0.029	0.024	0.033
Val Class Loss	0.116	0.109	0.100

Figure 1: YOLOv8s Training Results (Loss and Metrics Curves)

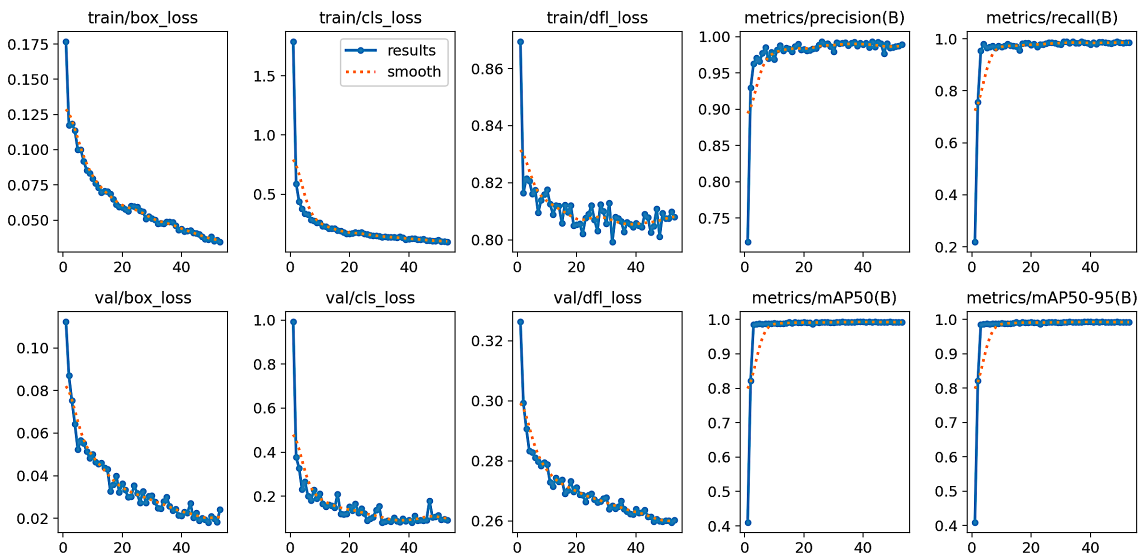


Figure 2: F1-Confidence Curve for YOLOv8s

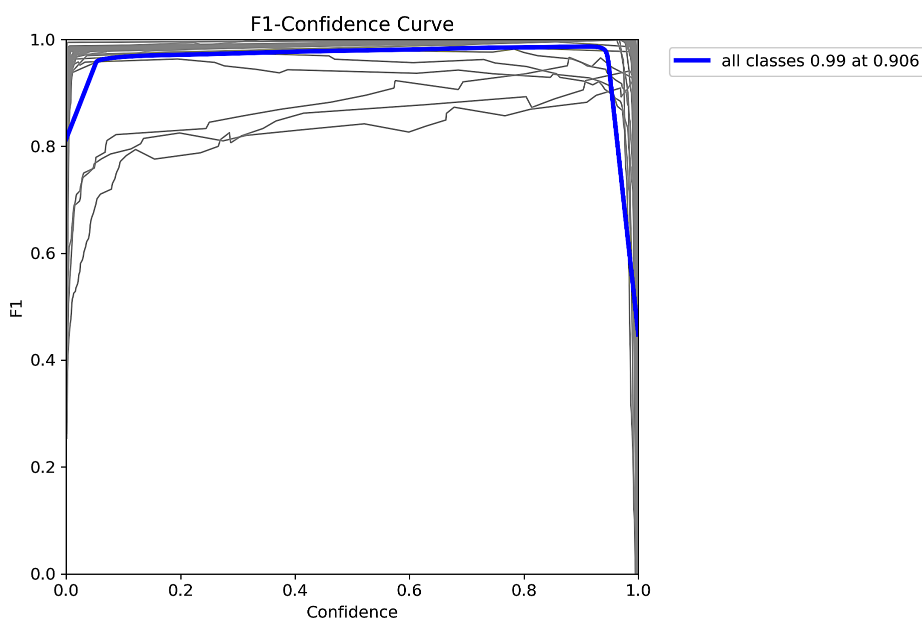
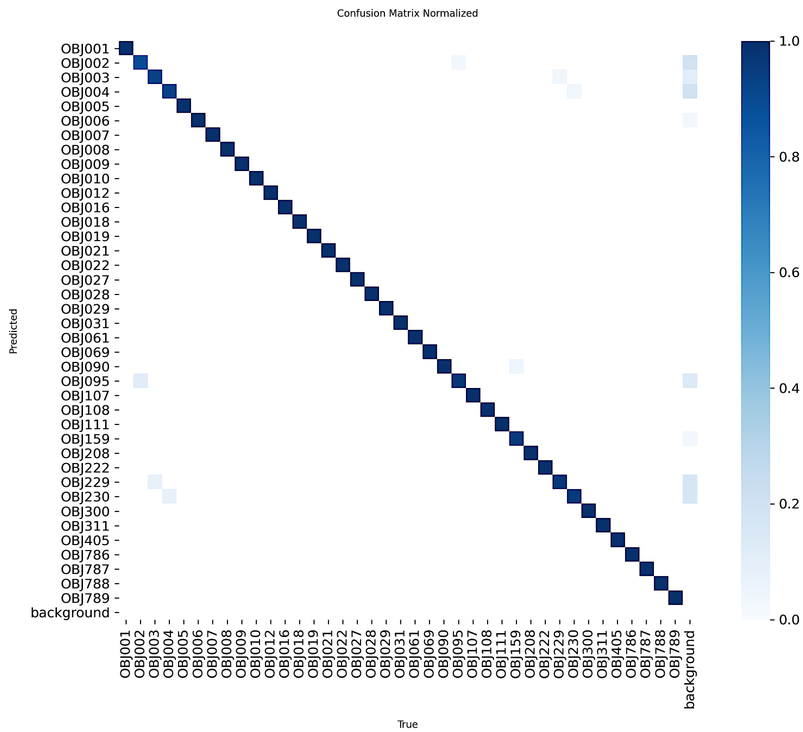


Figure 3: Normalized Confusion Matrix for YOLOv8s



4. OBJECT DETECTION DEMONSTRATION

4.1 Detection Capabilities

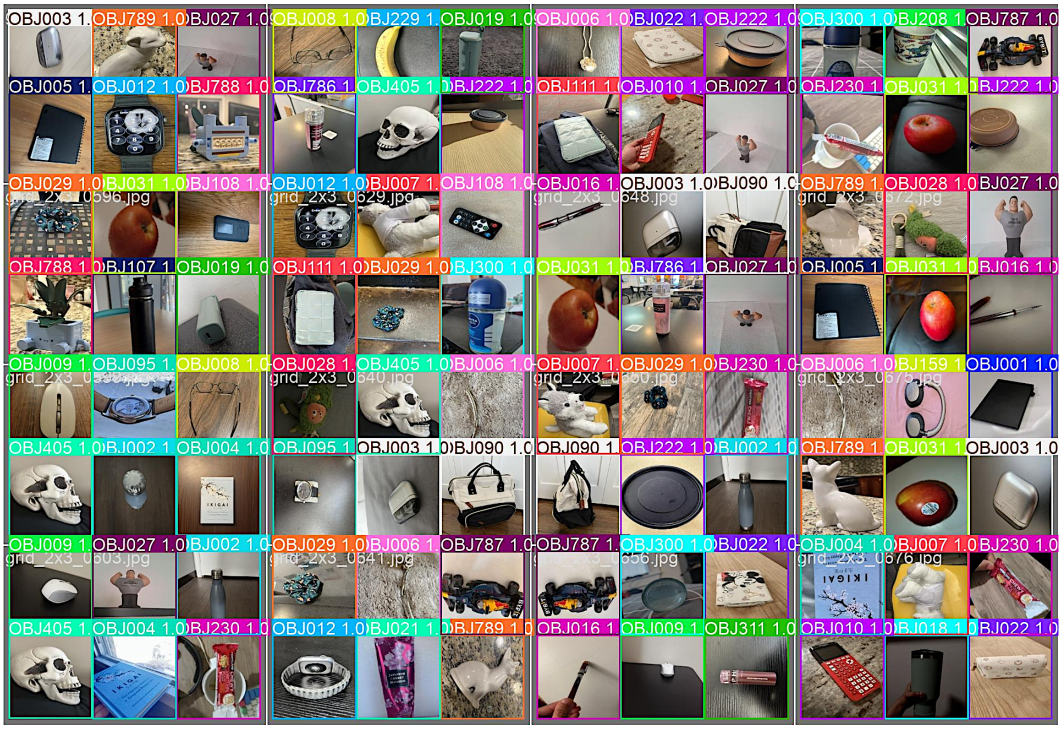
The trained models successfully detect and localize all objects within multi-object images, providing for each detection:

- Object ID: Class identifier (OBJ001 - OBJ789) from the 39-class dataset
- Object Name: Human-readable name (e.g., 'Coffee Mug', 'Hulk Toy', 'Glasses')
- Bounding Box: Precise location coordinates (x_center, y_center, width, height)
- Confidence Score: Detection confidence percentage (typically 95-100%)

4.2 Sample Detection Output

Validation batch predictions showing model detection accuracy:

Figure 4: Sample Detection Predictions from Validation Set

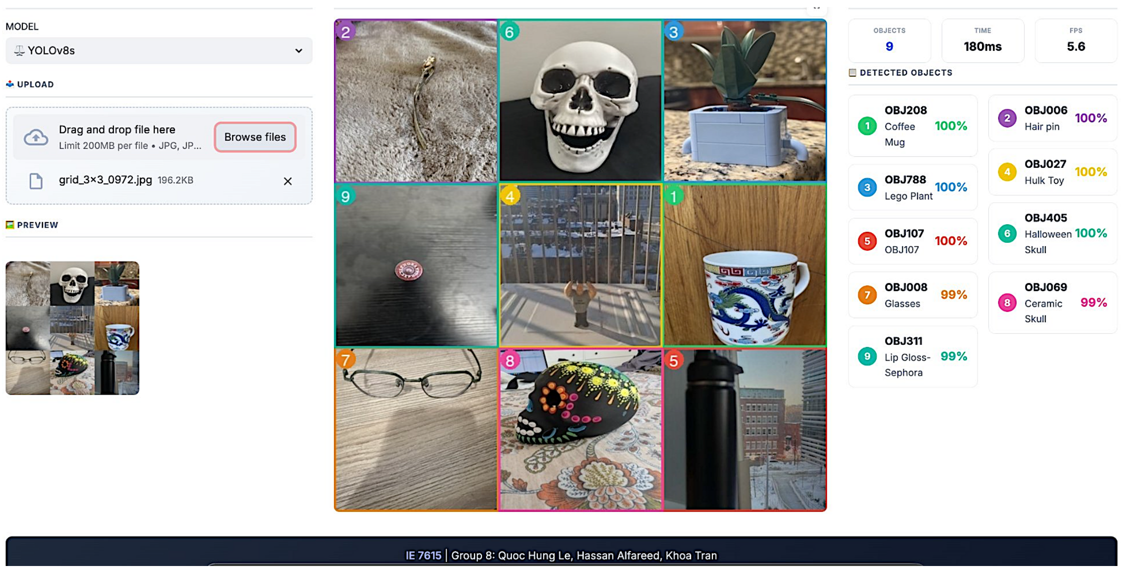


4.3 Web Application Interface

A professional Streamlit web application was developed for demonstration. The interface includes:

- Multi-Object Detection Tab: Upload images for real-time detection with YOLOv8
- Model Selection: Choose between YOLOv8n, YOLOv8s, or YOLOv8m variants
- Visual Results: Annotated images with colored bounding boxes and index numbers
- Grid View Results: Detection results displayed in grid matching input image layout
- Performance Metrics: Real-time display of objects found, inference time, and FPS

Figure 5: Web Application - Multi-Object Detection Interface



5. CONCLUSION

5.1 Summary of Achievements

Milestone 2 successfully completed all project objectives for multi-object detection:

- Dataset Generation: Created 1,200+ multi-object images using grid concatenation (2x2, 2x3, 3x3 layouts)
- YOLO Annotations: Generated accurate bounding box annotations in YOLO format for all images
- Model Training: Trained three YOLOv8 variants (nano, small, medium) using transfer learning
- Performance Excellence: All models achieved mAP50 > 98%, exceeding the 85% target by 13+ points
- Object Detection: Models successfully identify and locate all objects in multi-object images
- Web Application: Developed Streamlit app for real-time detection demonstration

5.2 Technical Insights

Transfer Learning Effectiveness: COCO pretrained weights provided excellent initialization, enabling rapid convergence (31-53 epochs) and strong performance on our 39-class dataset.

Grid-Based Generation Success: The systematic grid approach created high-quality training data with clear object boundaries, contributing to near-perfect detection accuracy.

Model Size vs. Performance: YOLOv8m achieved best accuracy (98.90% vs 98.76%), while YOLOv8n offers faster inference for real-time applications.

5.3 Model Recommendations

Use Case	Recommended Model	Rationale
Real-time detection	YOLOv8n	Fastest inference, 38 epochs training
Balanced deployment	YOLOv8s	Good mAP50 (98.74%), balanced speed
Maximum accuracy	YOLOv8m	Larger model capacity

5.4 Project Status

With the successful completion of Milestone 2, the CNN Attendance System project has achieved all core objectives:

- Single-object classification: 99.36% accuracy (Milestone 1)
- Multi-object detection: 98.90% mAP50 (Milestone 2)
- Web application: Fully functional demonstration interface
- Both milestones significantly exceed target performance metrics