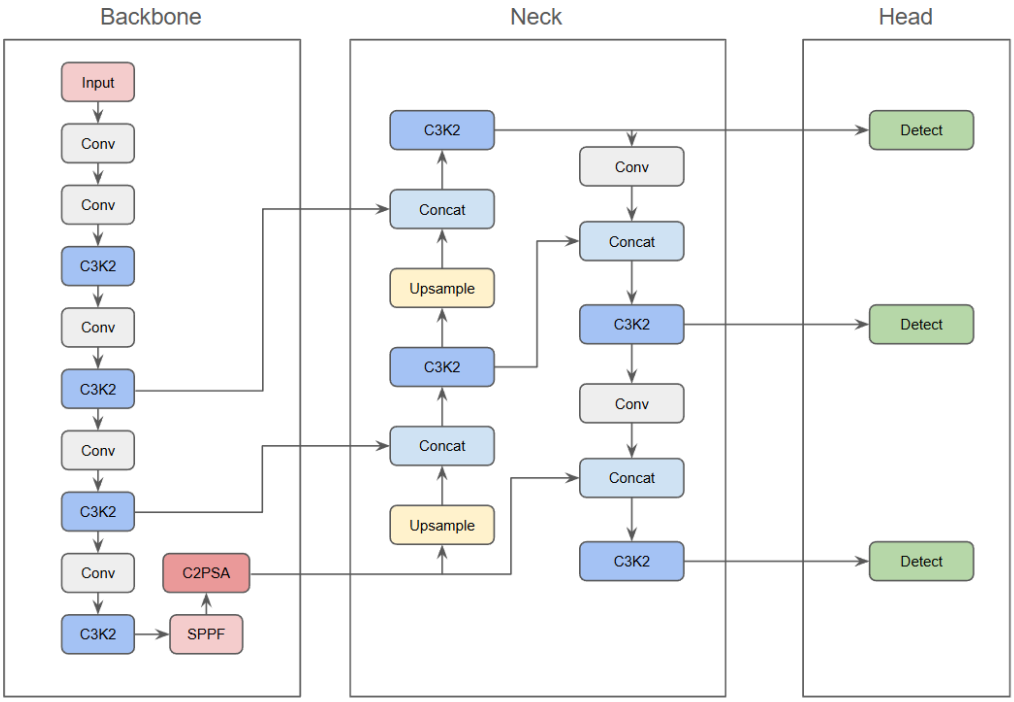
HW1-Object_Detection

1. Model Description

➢ Architecture

- Using the YOLO11x architecture.



2. Implementation Details

➢ Preprocessing

- Remove data where the BoundingBox width or height equals 0.
- Remove groundtruths without corresponding images.
- Convert to YOLO format.

➢ Hyperparameters

- Training

Parameter	Value
epochs	200
patience	30
batch	16
imgsz	640
pretrained	False

- Prediction

Parameter	Value
conf	0.5
iou	0.3
imgsz	640
augment	True

➢ Augmentation

- YOLO11 data augmentation training hyperparameters.

Parameter	Value
hsv_h	0.015
hsv_s	0.7
hsv_v	0.4
translate	0.1
scale	0.5
fliplr	0.5
mosaic	1.0

➢ Loss functions

- box_loss : The part used to optimize the difference between the predicted bounding box and the true bounding box
- cls_loss : Used to optimize the model's prediction accuracy for object categories. Classification loss ensures that the model can correctly identify which category the object in the image belongs to.
- dfl_loss : Solve the class imbalance problem in object detection and improve the performance of the model when dealing with small objects and difficult samples.

➢ Training strategies

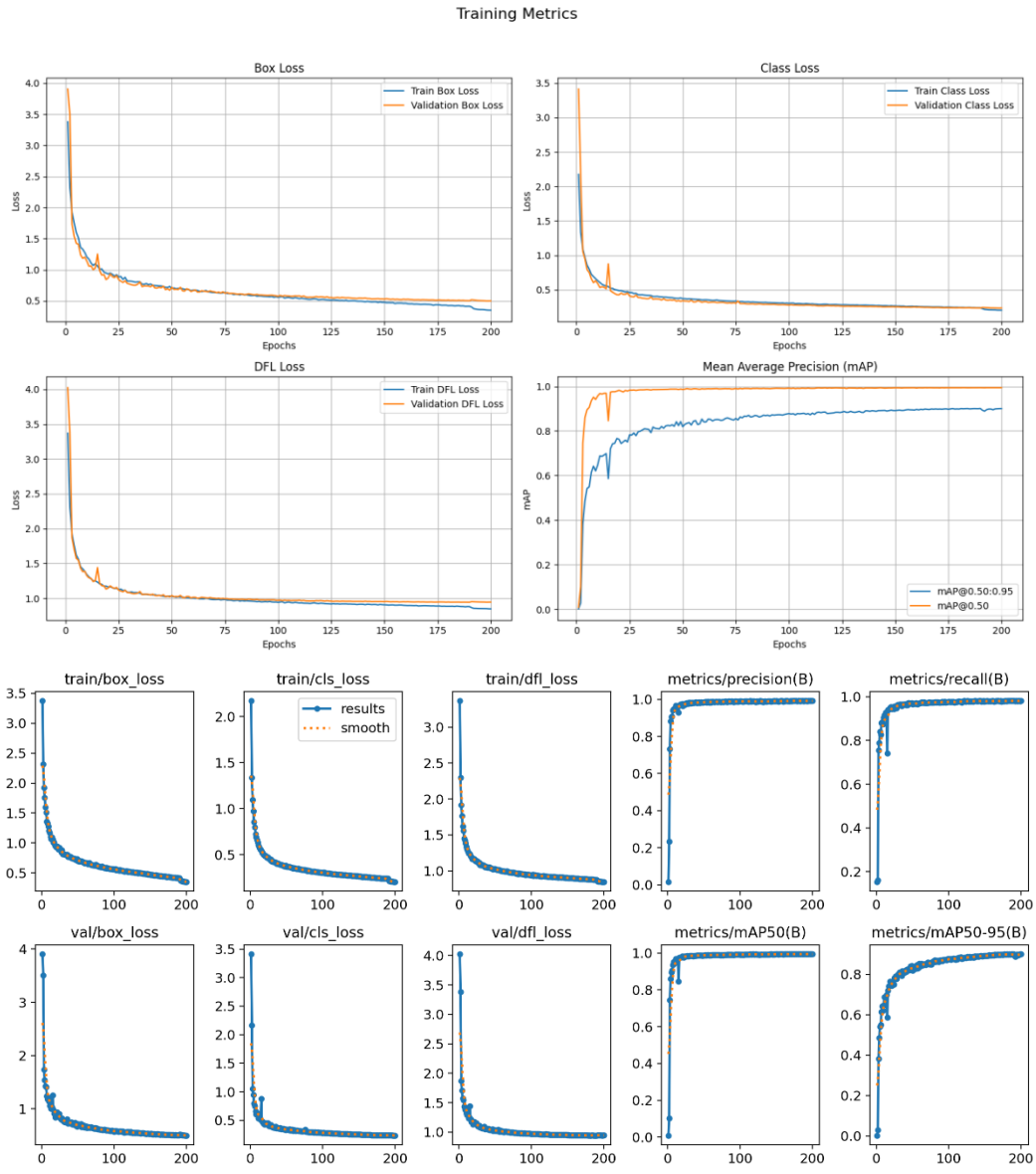
- The YOLO11x architecture serves as the base model. This model is a mid-to-large version of the Ultralytics YOLO series, featuring a deeper backbone and high-capacity feature extraction capabilities, making it suitable for object detection tasks in high-resolution and complex scenes.
- The dataset was split into a 80% training set and a 20% validation set.
- The model is trained for a maximum of 200 epochs. If no validation set performance improvement is seen after 30 consecutive epochs, early stopping is used to prevent overfitting and waste of computing resources. All training is performed on a single GPU.

- To improve the generalization ability of the model, YOLO's built-in enhancement mechanism is used, including color, geometry and combination enhancement.
- In prediction, The confidence threshold is set to 0.4 to balance precision and recall, The IoU threshold is 0.5, Test-time augmentation (TTA) is set to True to improve prediction stability using multi-angle augmentation.

3. Result Analysis

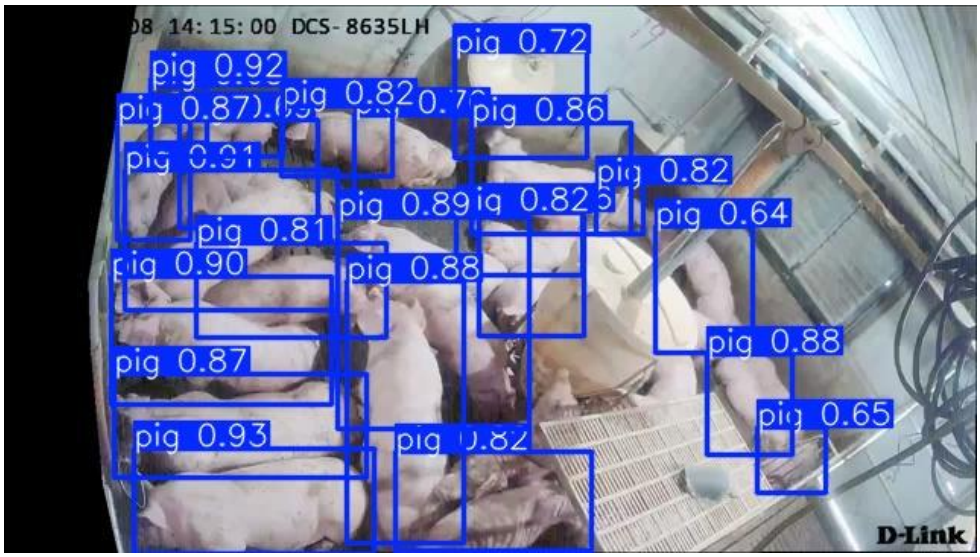
➢ Quantitative improvements

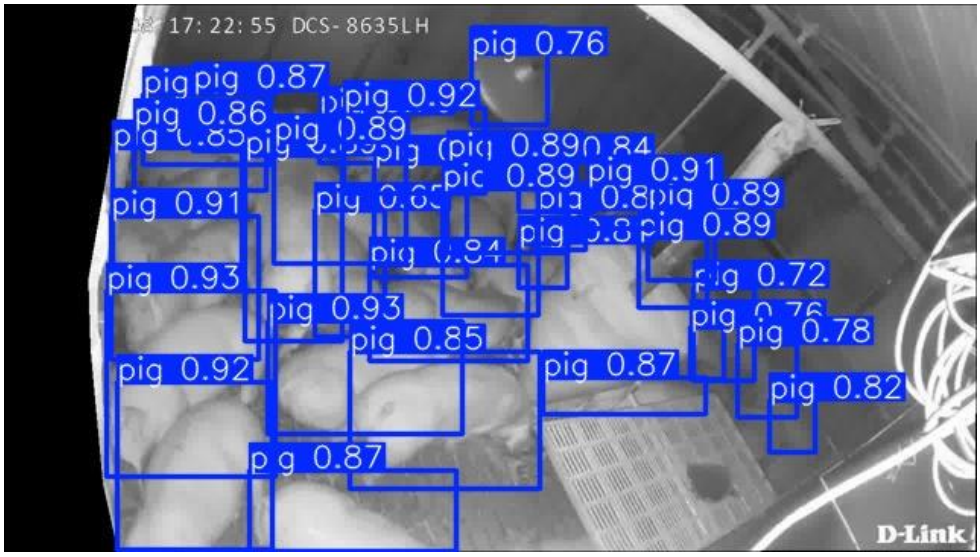
- Loss metrics : Training with YOLO11x and data augmentation converges quickly in the first 10–20 epochs, reaching mAP@50–95 ≈ 0.91 in around 60–80 epochs. The train/val loss is consistent, with no significant overfitting observed.



➢ Visualizations

- Predict results.





4. Conclusion

- For this object detection experiment, I preferred the YOLO series as the detection model due to its one-stage design, low latency, ease of deployment, and low engineering barrier to entry. Furthermore, the Ultralytics implementation provides complete and stable training and inference. Training was completed using YOLO11x with various parameter adjustments and data augmentation. The model achieved stable convergence on the validation set, achieving mAP50-95 of 0.91, and inference improved output stability. Due to the wide variety of adjustable parameters, comprehensive testing to find the optimal

combination was challenging and required a trade-off between performance and efficiency. Overall results show that, without relying on pre-training, appropriate data augmentation and early stopping can effectively improve generalization and demonstrate robustness to variations in lighting, scale, and position.