Content

- YOLO-World
- Fish Classification Using YOLO-World

❑ YOLO-World

YOLO-World

- On January 31st, 2024, Tencent's AI Lab released **YOLO-world** , a real-time, open-vocabulary object detection model.

- YOLO-World is a zero-shot model, which means it can run object detection without any training.

Source: **YOLO-World: Real-Time, Zero-Shot Object Detection**
https://blog.roboflow.com/what-is-yolo-world/

Traditional Object Detection:

- Traditional object detection models, such as **Faster R-CNN**, SSD, and YOLO, are designed to identify objects within a predetermined set of categories defined by their training datasets. For instance, models trained on the **COCO dataset** are limited to 80 categories.

- **This limitation restricts their applicability to scenarios that match the training data's scope.** Extending or altering the set of recognizable classes necessitates retraining or fine-tuning the model on a custom dataset tailored to the new categories.

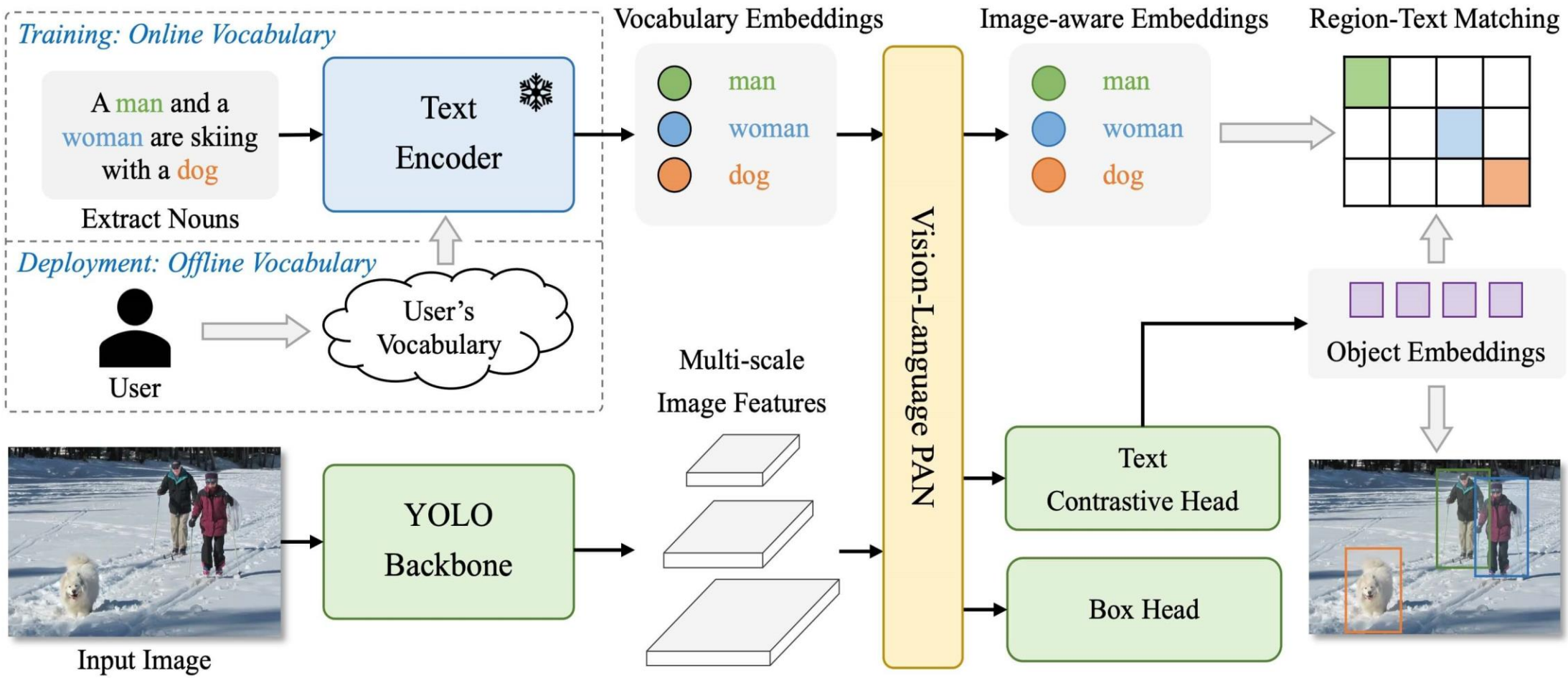
Open-Vocabulary Object Detection:

- As a response to the limitations of fixed-vocabulary detectors, open-vocabulary object detection (OVD) models aim to recognize objects beyond the predefined categories. Early attempts in this direction, such as GLIP and **Grounding DINO**, focused on leveraging large-scale image-text data to expand the training vocabulary, enabling the detection of novel objects.

- **However, they tend to be larger and more computationally intensive, requiring simultaneous encoding of images and texts for prediction**. This approach, while powerful, introduces latency that can delay practical applications.

Source: **YOLO-World: Real-Time, Zero-Shot Object Detection**
https://blog.roboflow.com/what-is-yolo-world/

YOLO-World's Architecture



Source: **YOLO-World: Real-Time, Zero-Shot Object Detection**
https://blog.roboflow.com/what-is-yolo-world/

YOLO-World's architecture consists of three key elements:

•**YOLO detector -** based on Ultralytics YOLOv8; extracts the multi-scale features from the input image.

•**Text Encoder -** Transformer text encoder pre-trained by OpenAI's CLIP; encodes the text into text embeddings.

•**Re-parameterizable Vision-Language Path Aggregation Network (RepVL-PAN) -** performs multi-level cross-modality fusion between image features and text embeddings.

Source: **YOLO-World: Real-Time, Zero-Shot Object Detection**
https://blog.roboflow.com/what-is-yolo-world/

Model Type
YOLOv8s-world
YOLOv8s-worldv2
YOLOv8m-world
YOLOv8m-worldv2
YOLOv8l-world
YOLOv8l-worldv2
YOLOv8x-world
YOLOv8x-worldv2

Source: YOLO-World Model
https://docs.ultralytics.com/models/yolo-world/

❑ Fish Classification Using YOLO-World

Purpose

Classifying Fishes into Five Different Categories

Class 1- Fish
Class 2-Gibuna
Class 3-tamoroko
Class 4-Aburahaya
Class 5-Tanago

Dataset Information

Number of Images	Number of Classes
1098	5

Number of Labels

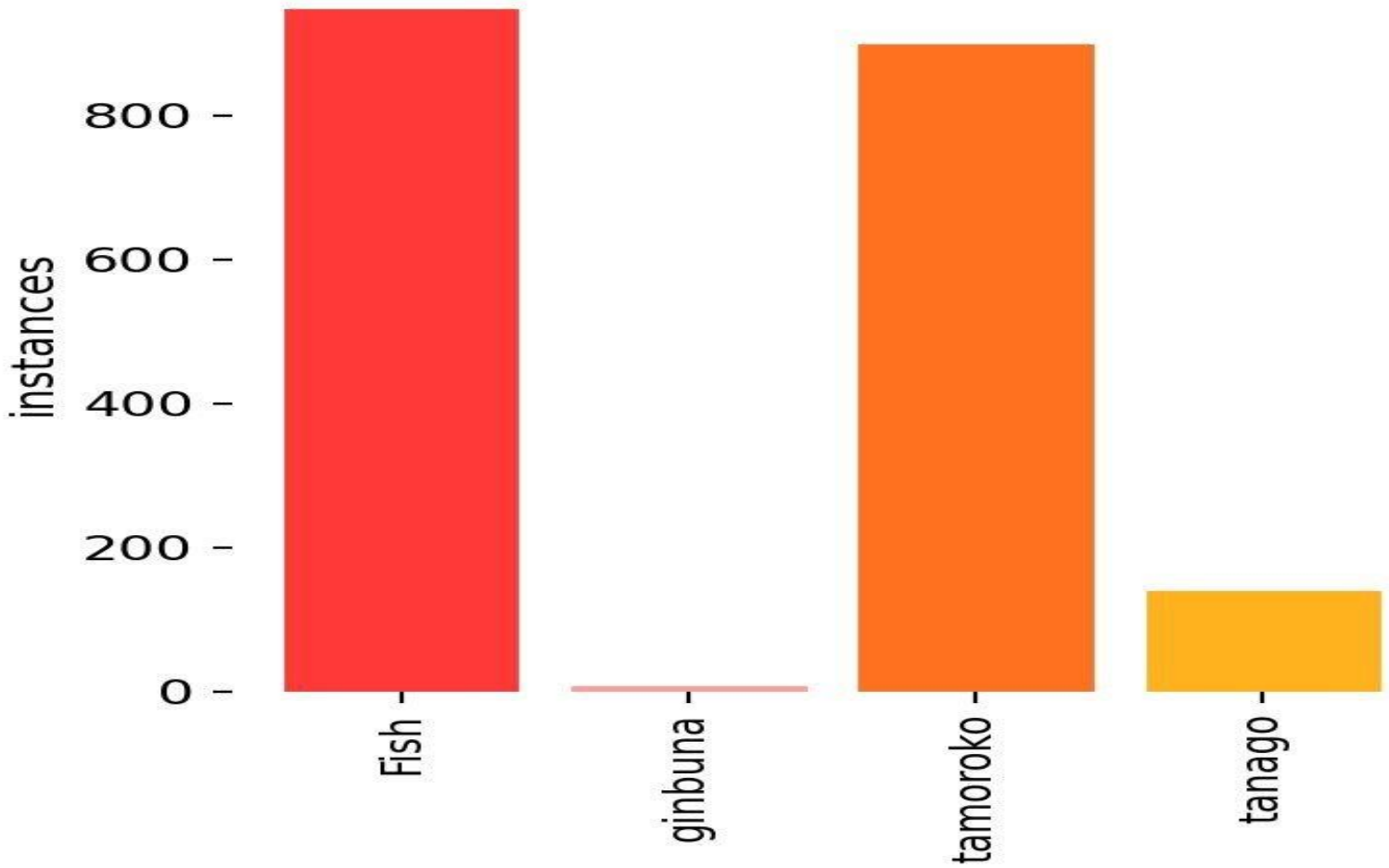
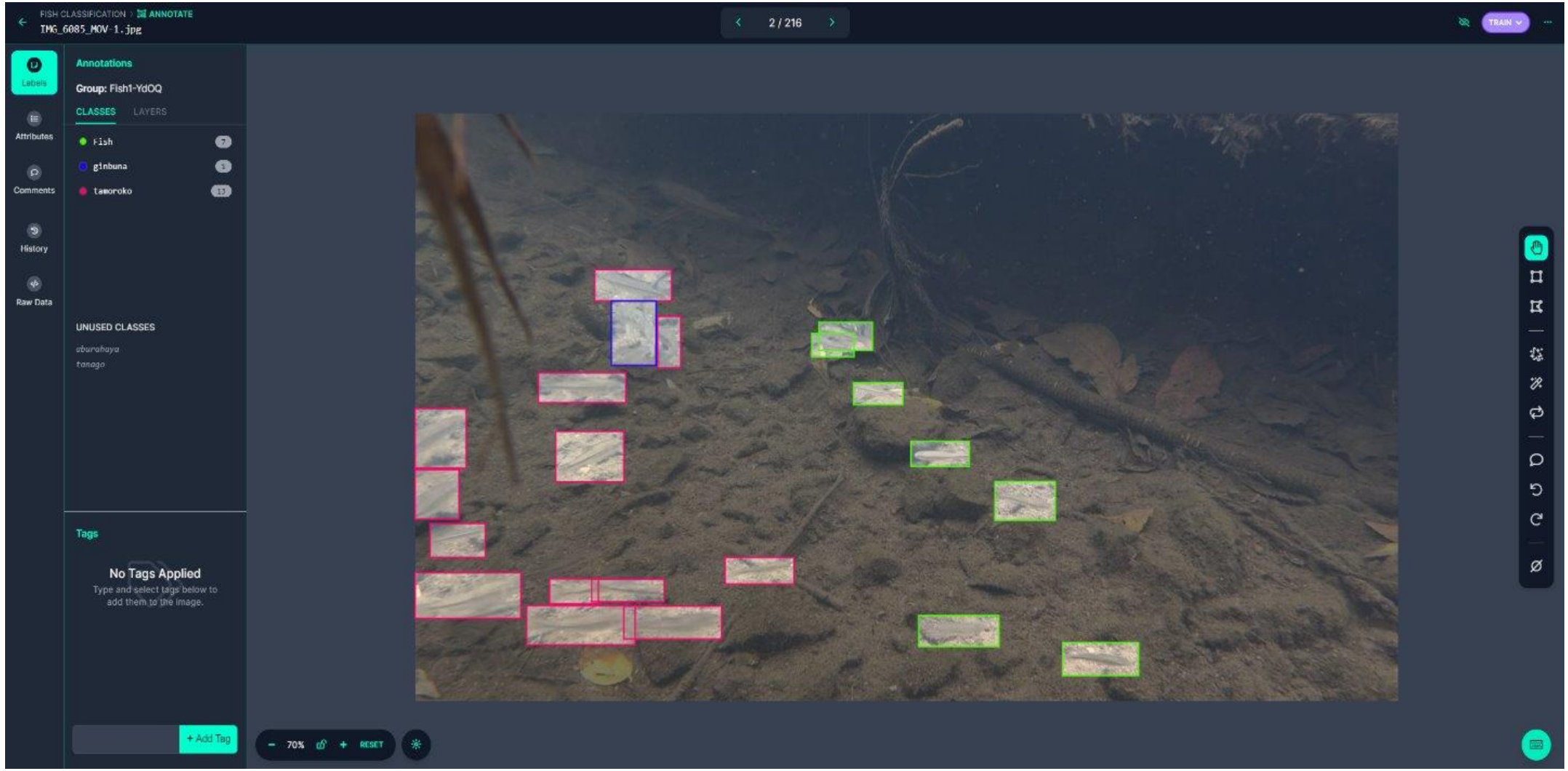


Image Annotation Tool

Roboflow



Training YOLOv8x-Worldv2 Model

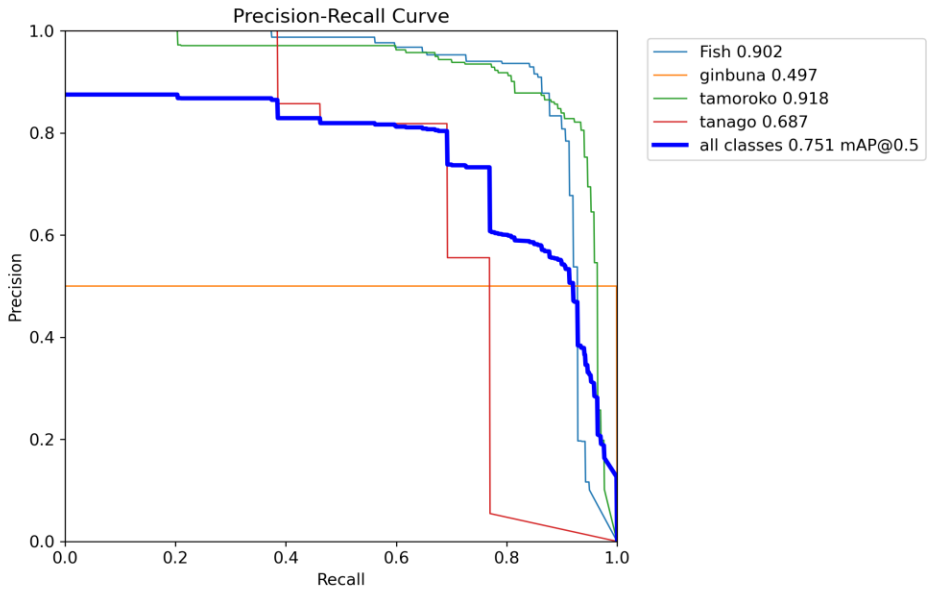
YOLOv8x-Worldv2 Model Package:

YOLO-World Github Package

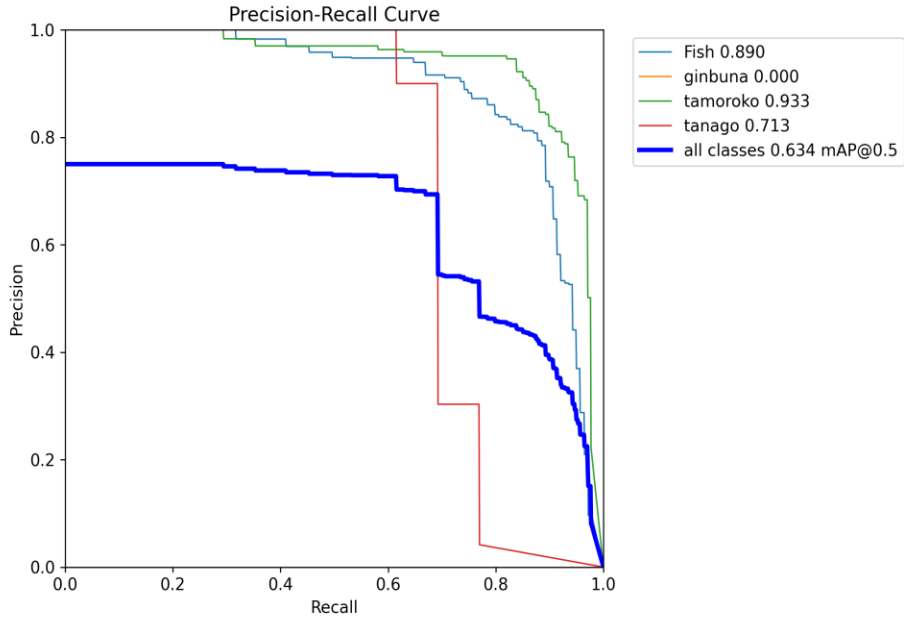
Training Platform:

Anaconda Cloud

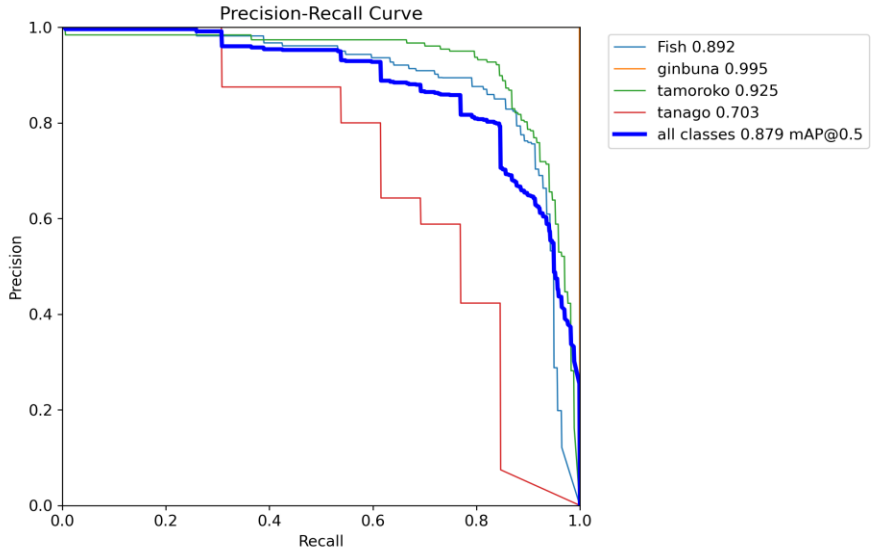
Precision Recall Yolov8s-Worldv2-Model



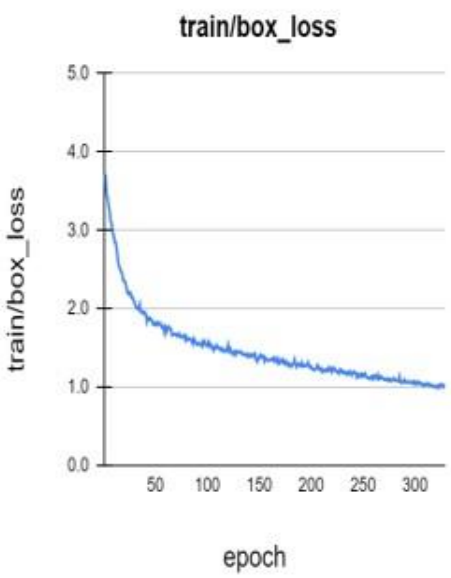
Precision Recall Yolov8m-Worldv2-Model

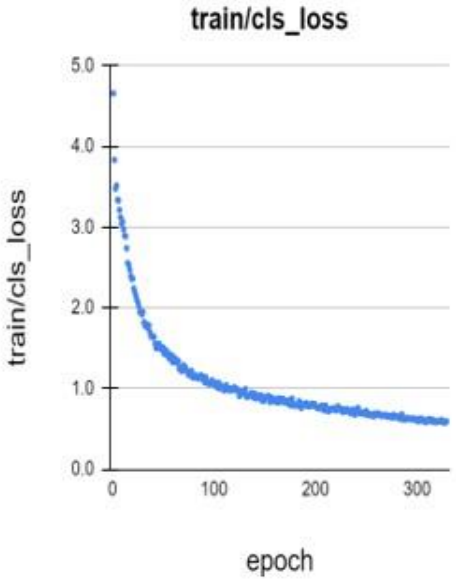


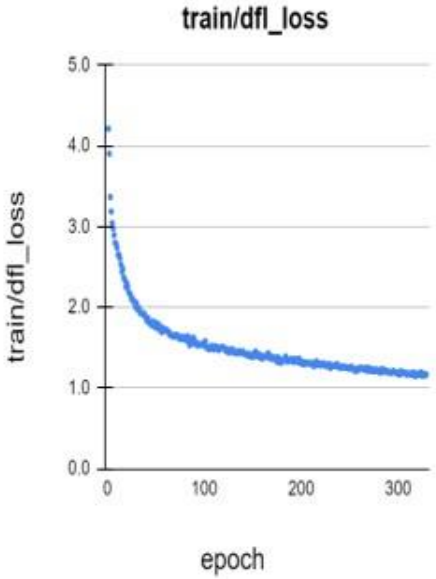
Precision Recall Yolov8x-Worldv2-Model

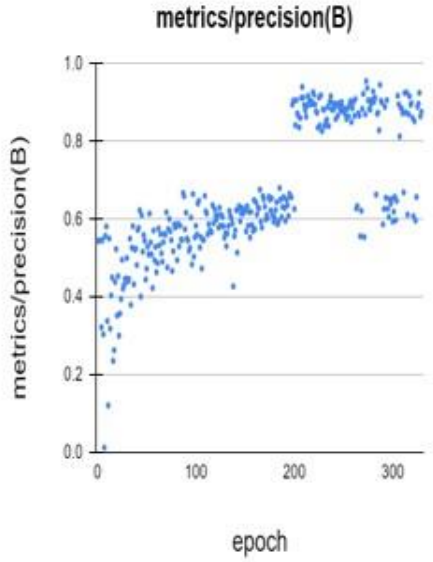


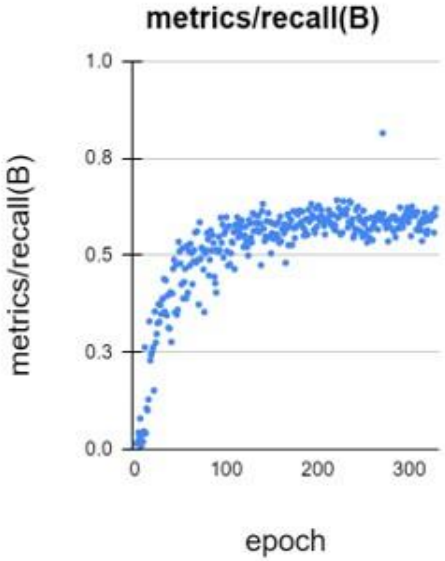
Training Result Yolov8s-Worldv2 Model

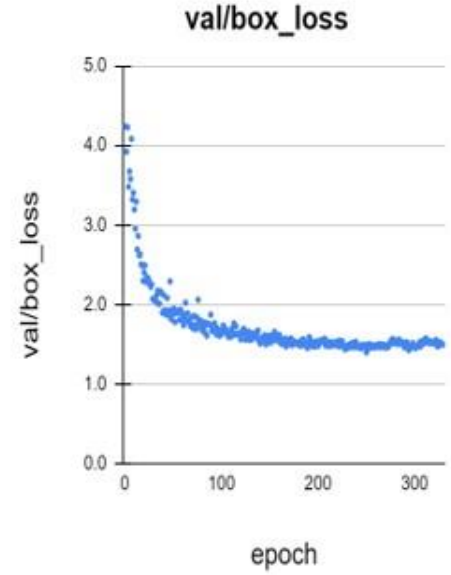


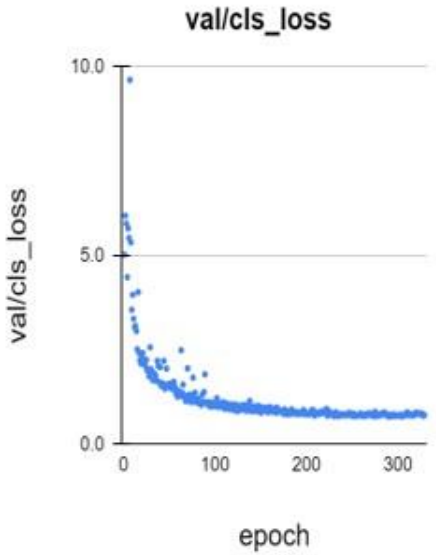


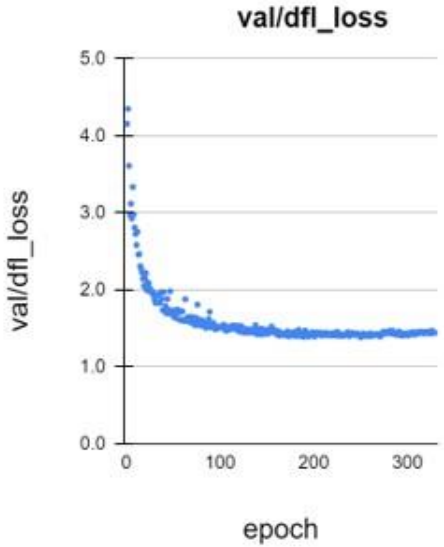


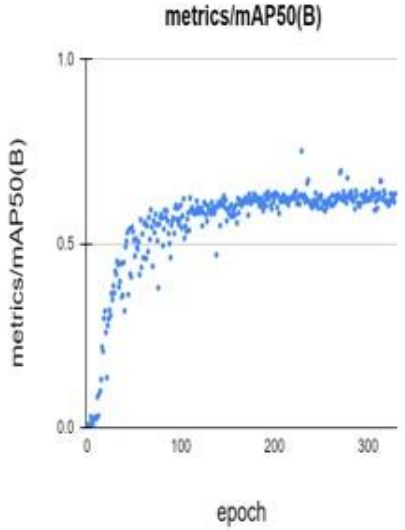


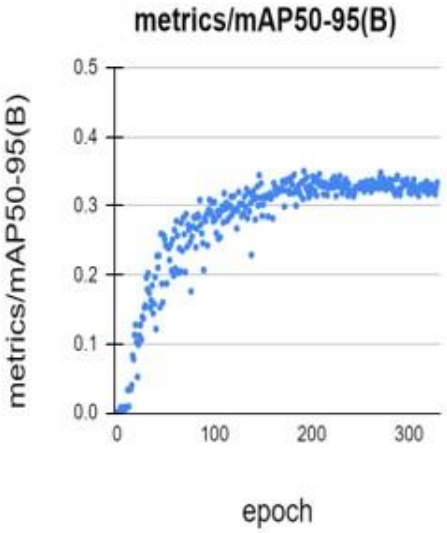




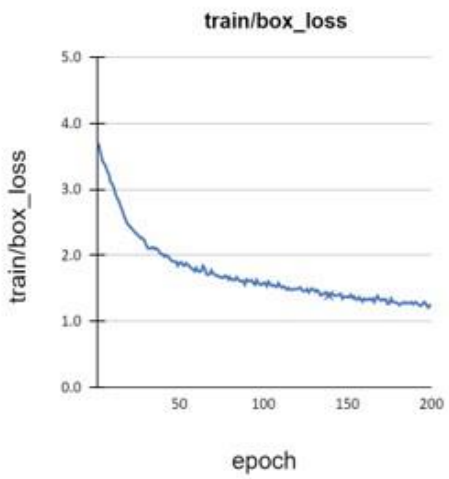


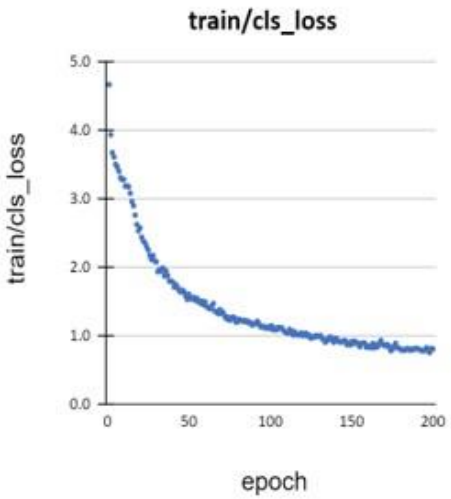


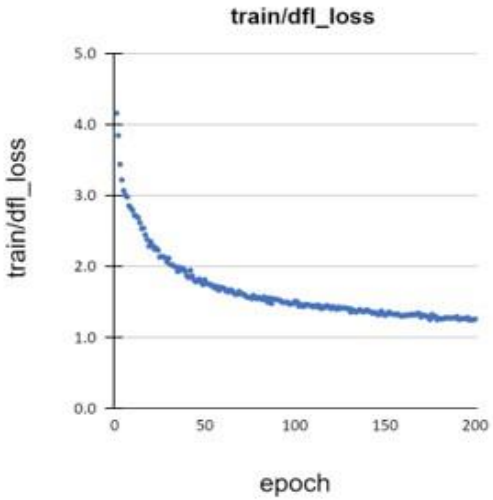


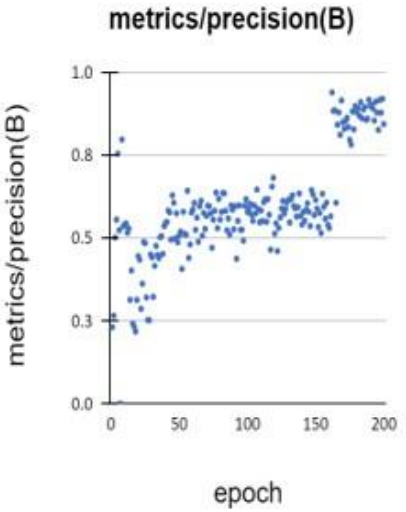


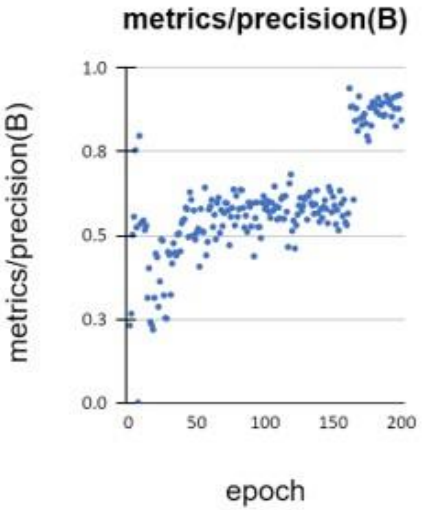
Training Result Yolov8m-Worldv2 Model

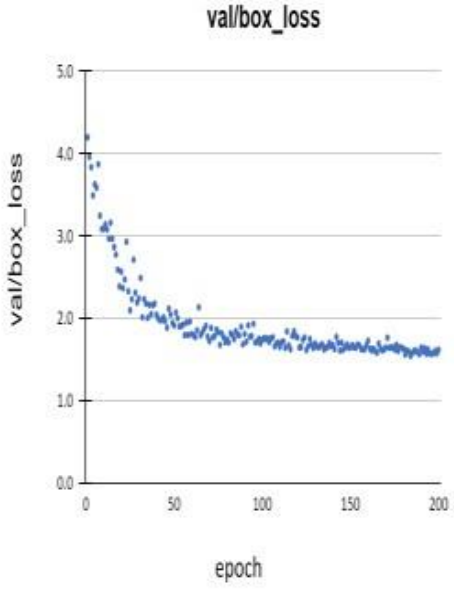


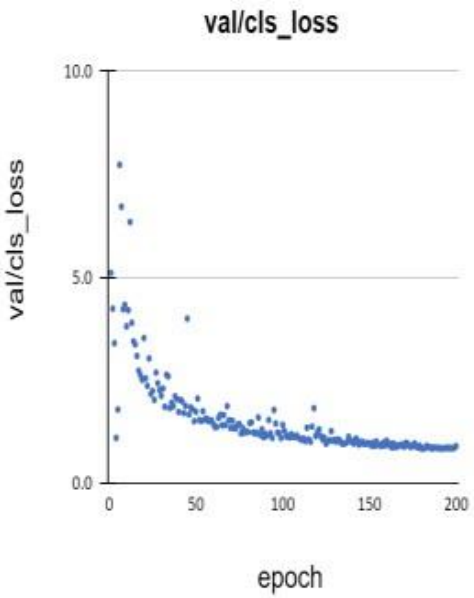


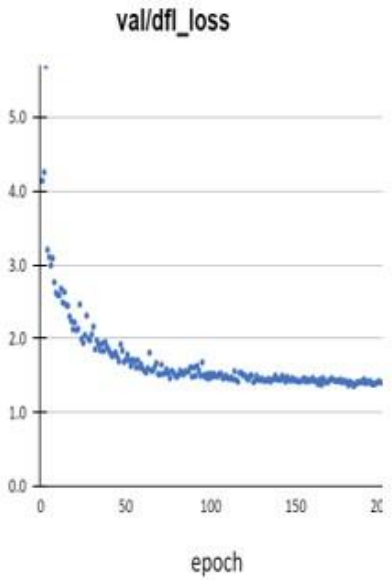


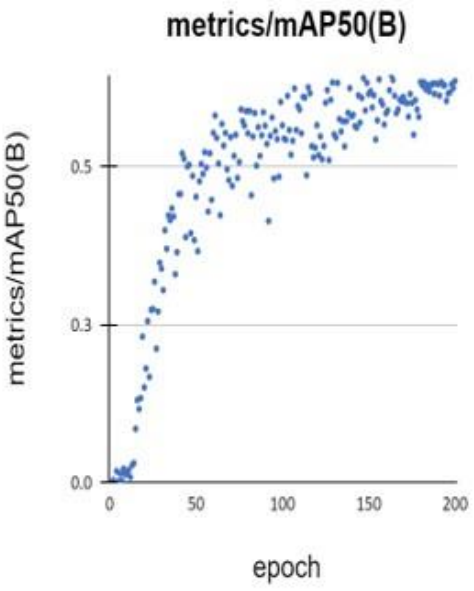


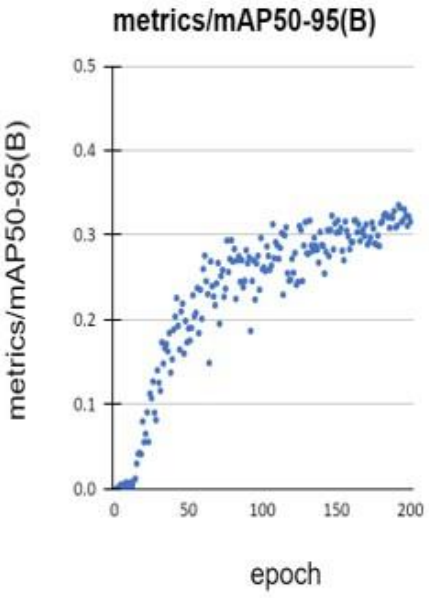




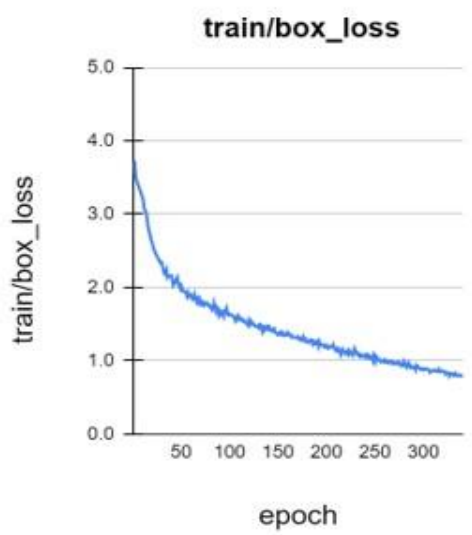


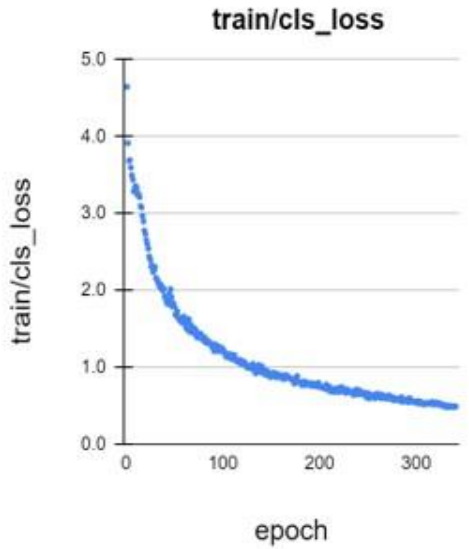


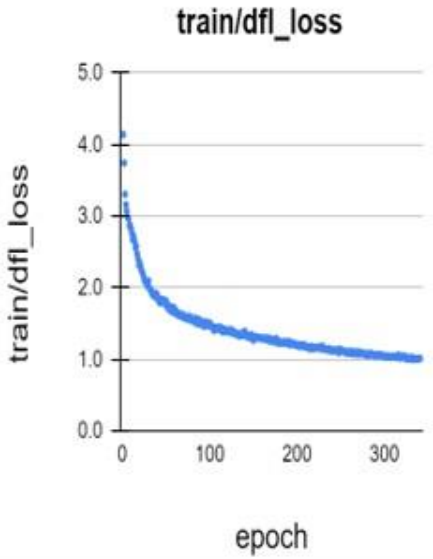


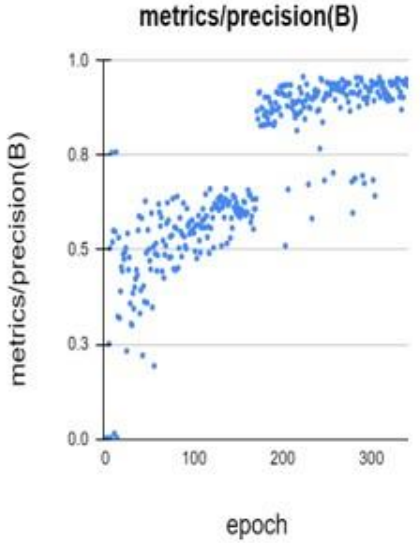


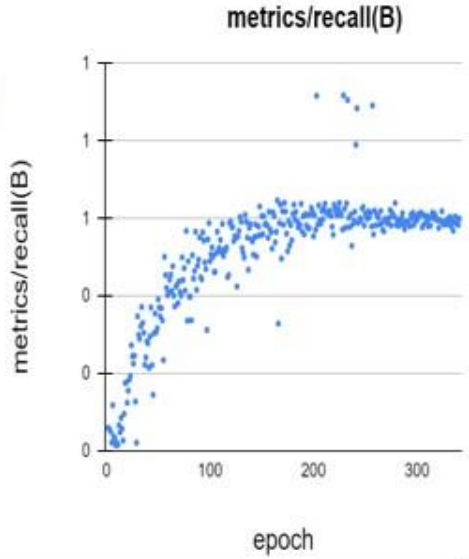
Training Result Yolov8x-Worldv2 Model

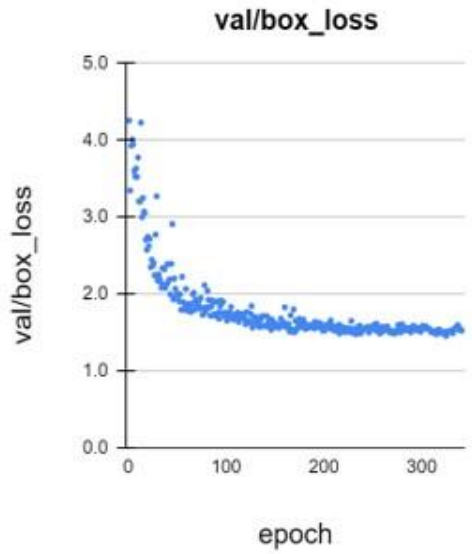


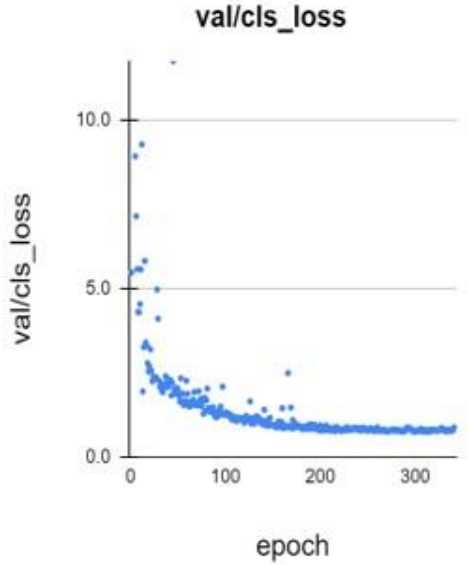


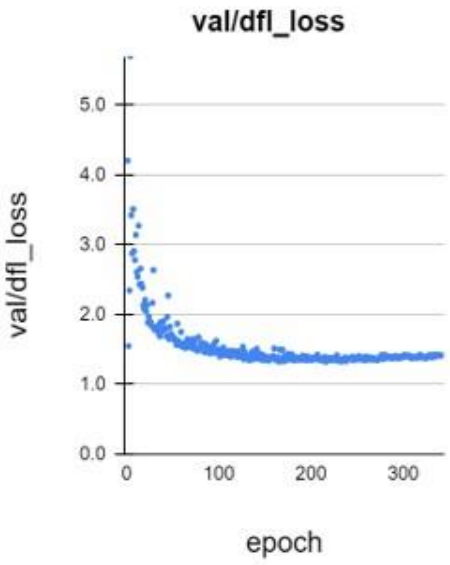


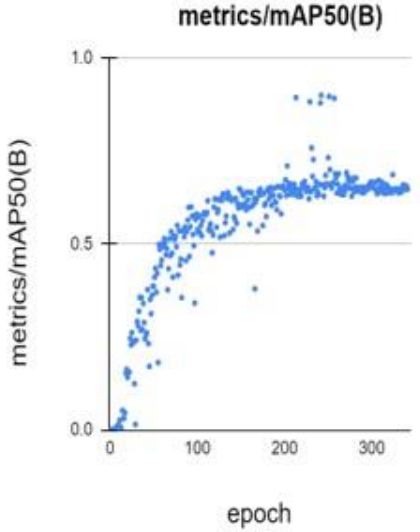


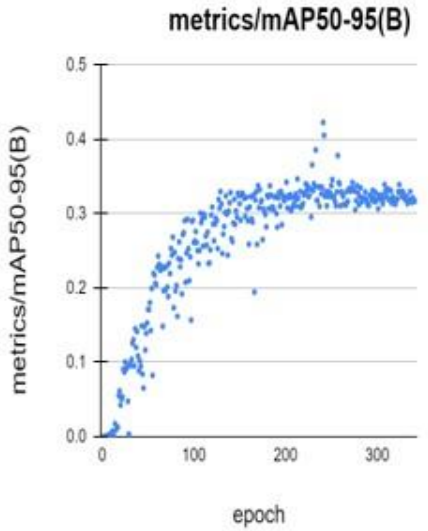




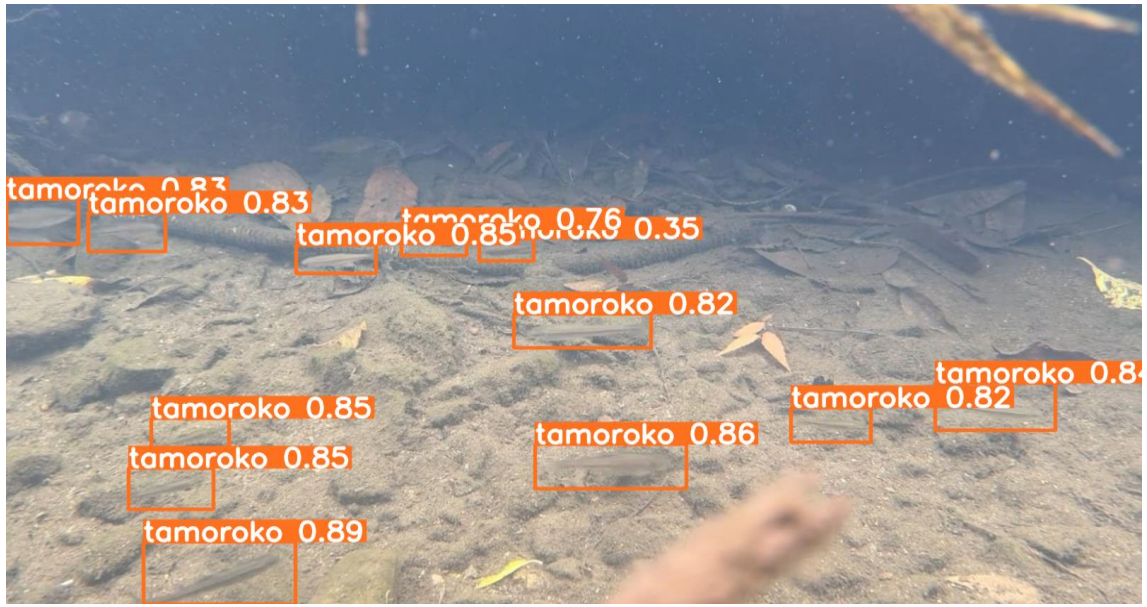


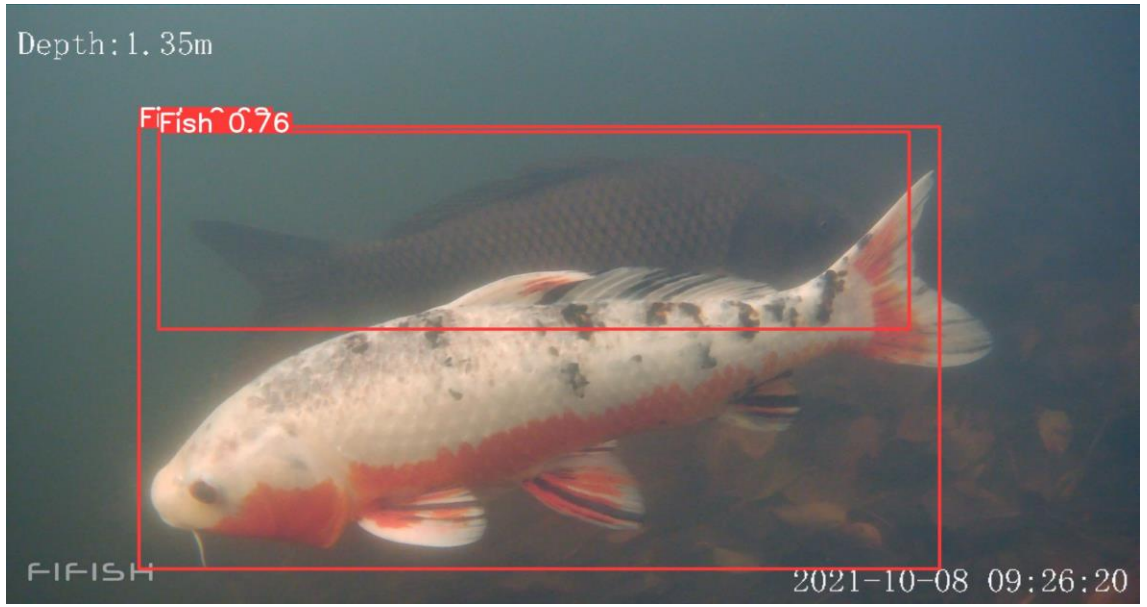


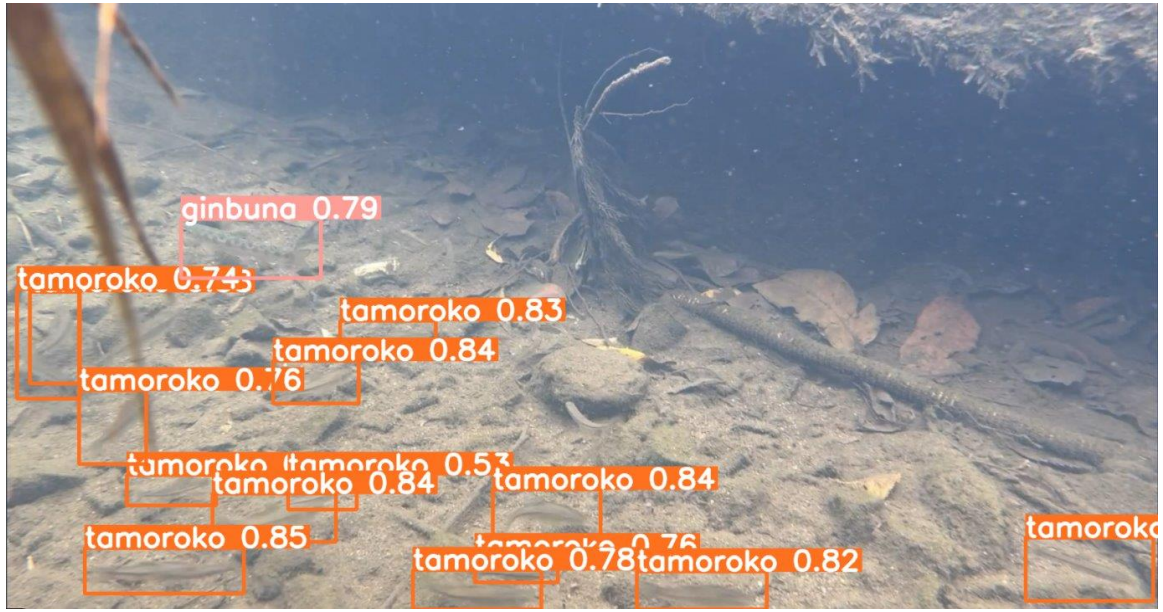




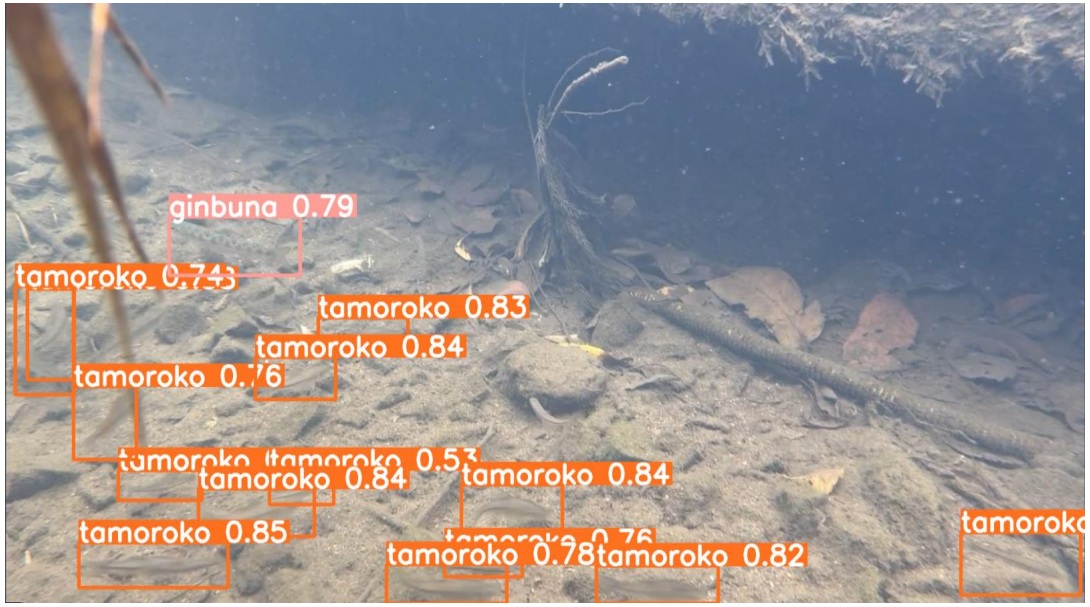
Test Result Yolov8s-Worldv2 Model





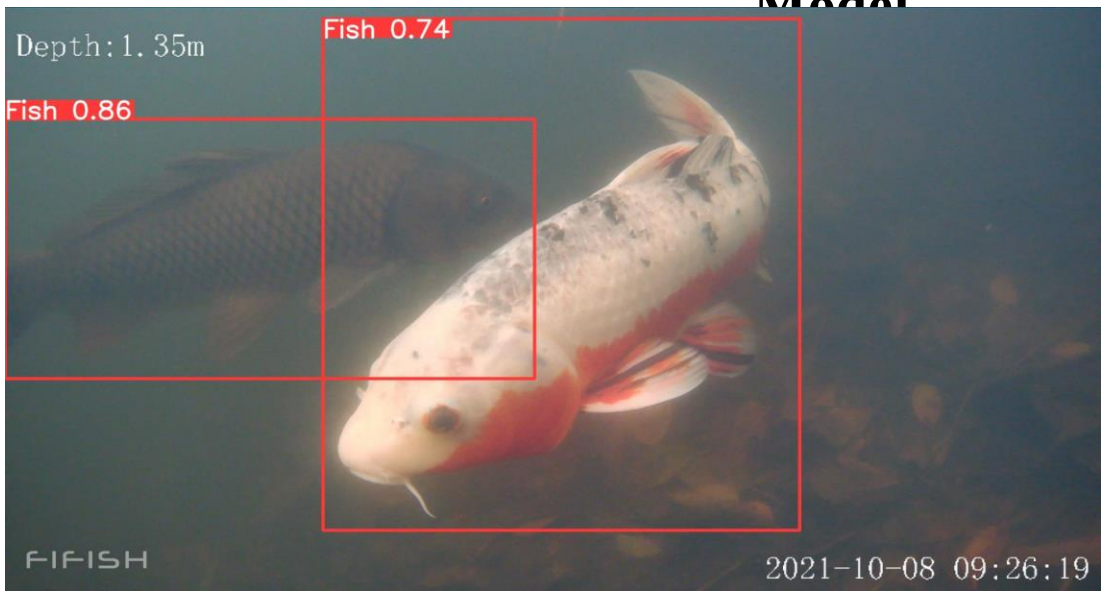


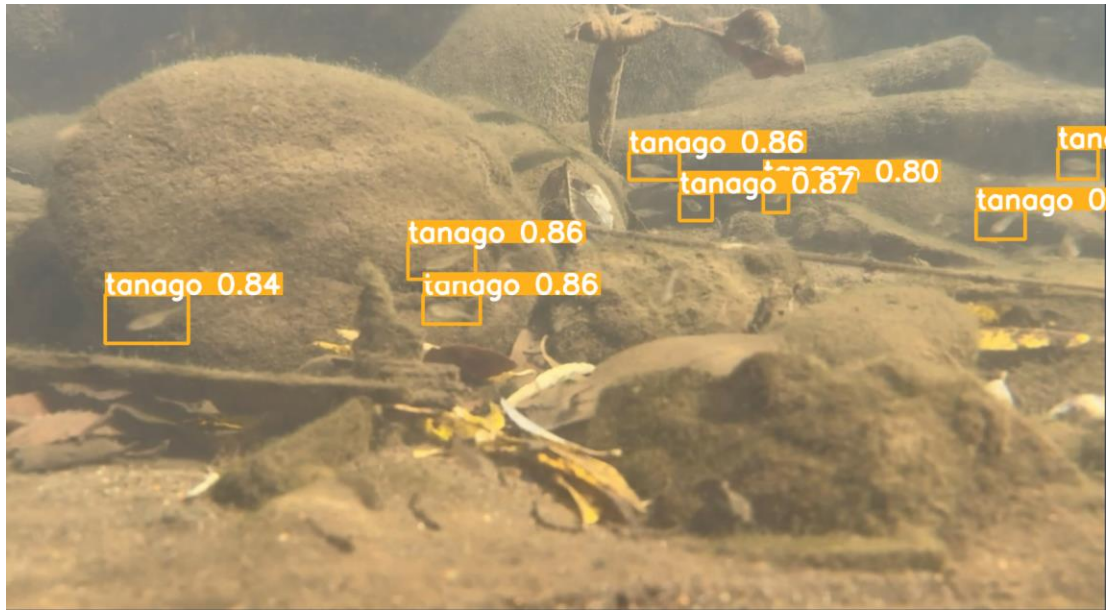
Test Result Yolov8m-Worldv2 Model

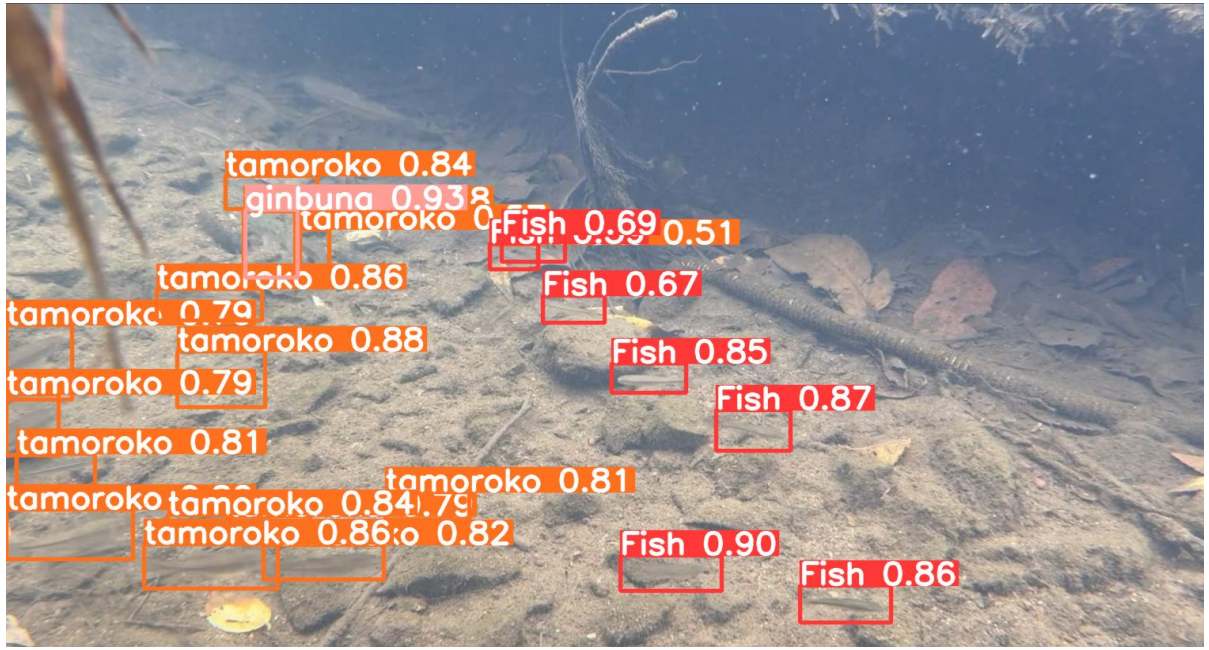


Test Result Yolov8x-Worldv2

Model







Comparision of Test Result

Original Image

Annotated Image

Class 1 - Fish

Comparision of Test Result Yolo-model S,M,X

Model-s

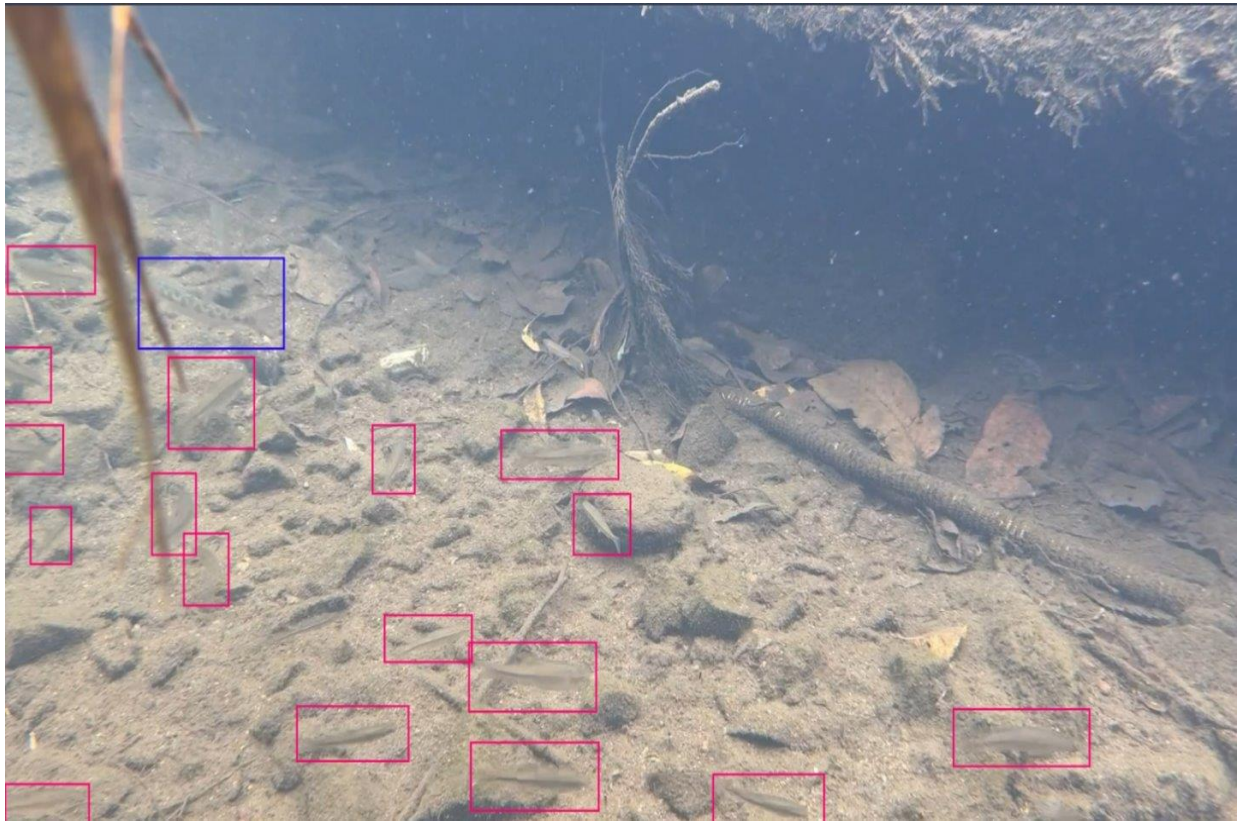
Model-m

Model-x

Comparision of Test Result

Original Image

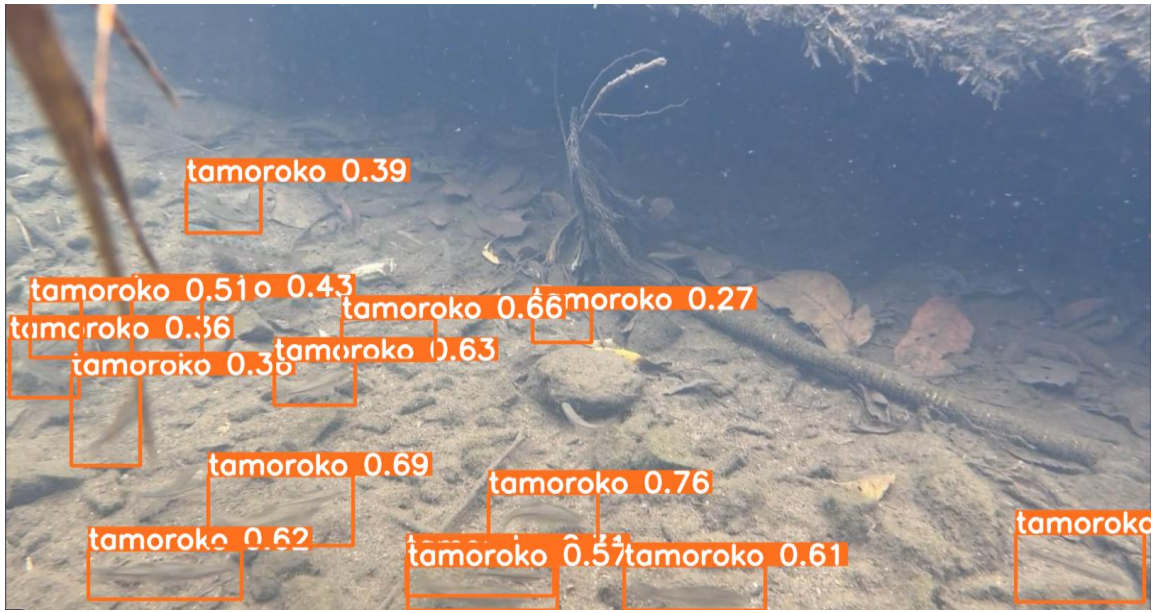
Annotated Image



Class 2 - Gibuna Class 3 - Tomoroko

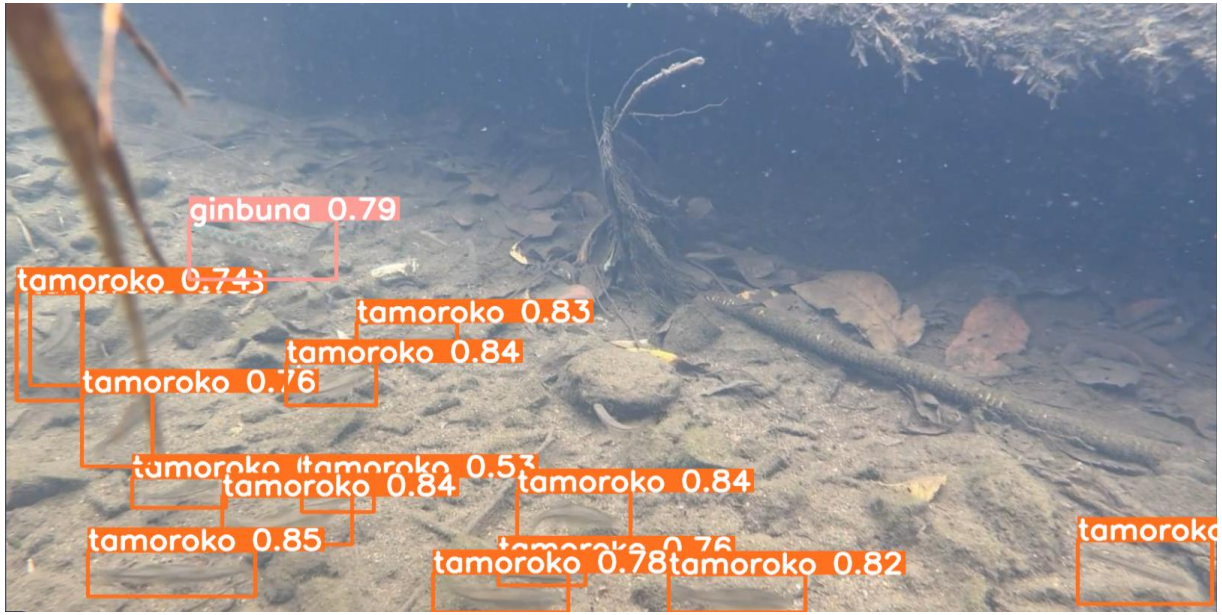
Comparision of Test Result Yolo-model S,M,X

Model-s



Model-m

Model-x



Comparision of Test Result

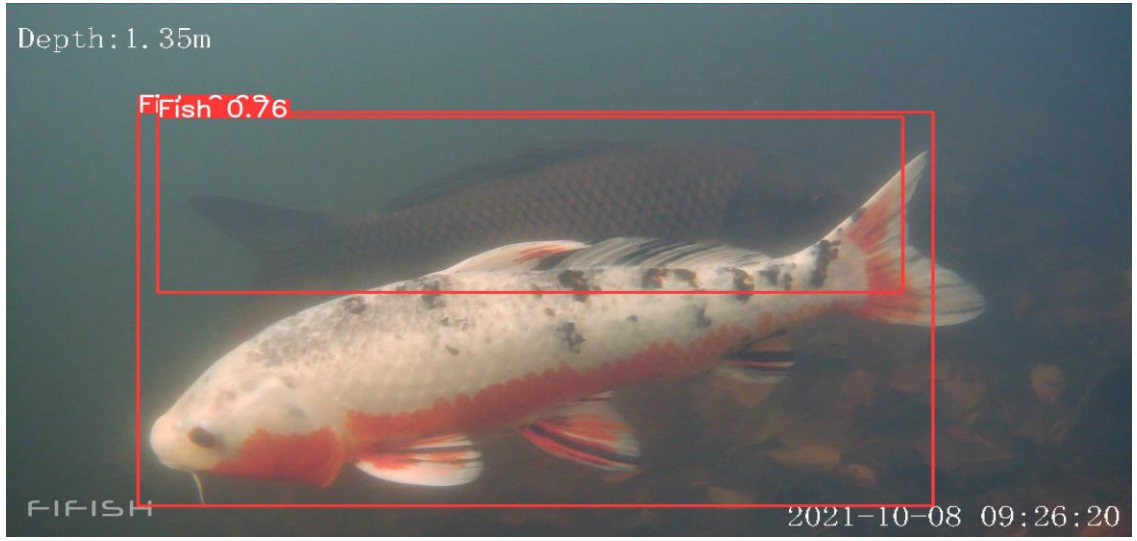
Original Image

Annotated Image

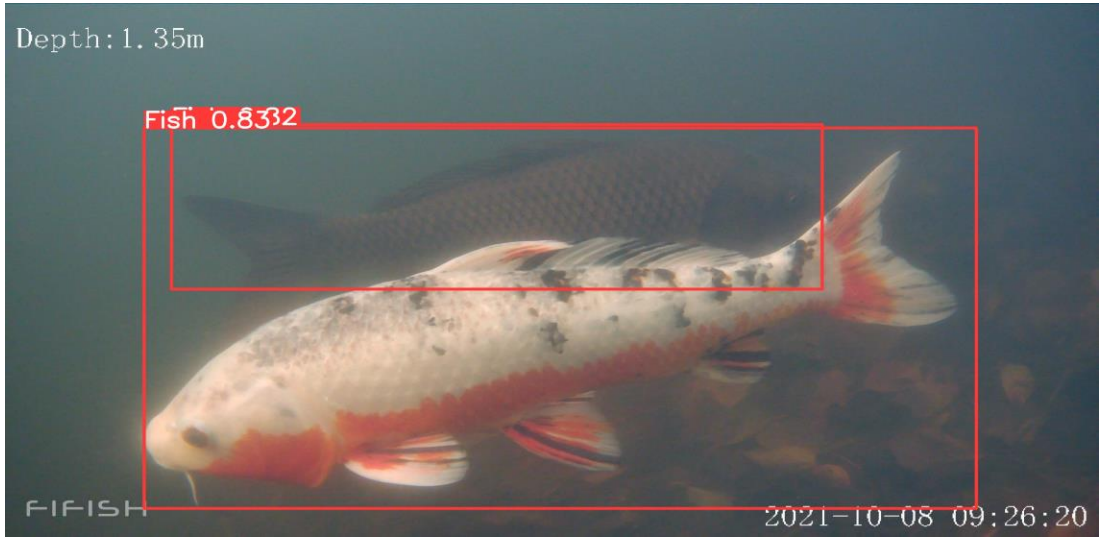
Class 1 - Fish

Comparision of Test Result Yolo-model S,M,X

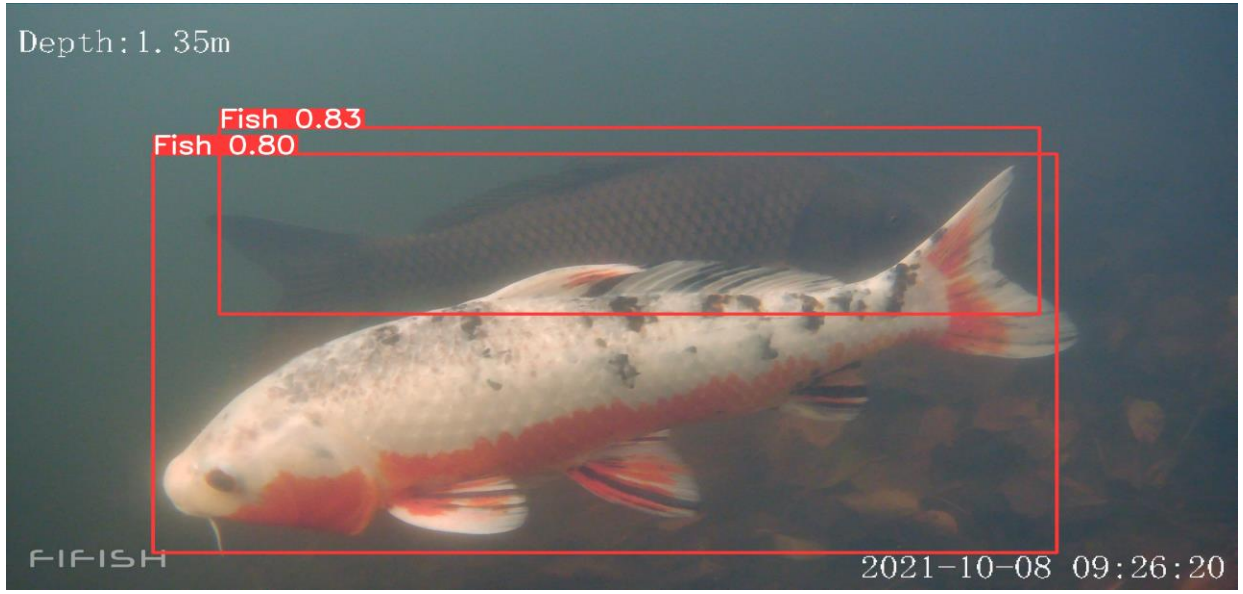
Model-s



Model-m



Model-x



Comparision of Test Result

Original Image

Annotated Image

Class 5 - Tanago

Comparision of Test Result Yolo-model S,M,X

Model-s

Model-m

Model-x