Development of the Delfi Project Report 2: Model Development and Testing Results

[Kananat Siangsano]

March 13, 2025

1 Introduction

This is the 2nd report of the Delfi project, this document details the model development process and visual inspection of detection results. Our focus has shifted to model refinement through iterative training cycles and performance evaluation across challenging real-world scenarios.

2 Model Development

The Delfi project progressed through three distinct model iterations, each addressing specific limitations identified in previous versions:

2.1 Model 1: Initial Synthetic Training

The first model implementation utilized YOLOv11m architecture trained exclusively on the synthesized detection dataset described in our previous report.

2.2 Model 2: Active Learning Refinement

Building upon Model 1, the second iteration incorporated active learning principles by training on 100 manually corrected images from the initial model. This approach allowed the model to adapt to real-world image characteristics not fully captured in the synthetic dataset, particularly regarding background variations.

2.3 Model 3: Preprocessing Enhancements

The third and currently the most advanced implementation featured several significant improvements:

- Upgrading to YOLOv12 architecture for enhanced detection capabilities
- Implementation of tiling during the training process to improve small object detection

- Addition of moving window detection during inference for scalability to images with different sizes compared to the training data.
- Expanded training dataset incorporating 150 corrected outputs from previous models.
- Supplementation training data with 6,503 instances of Ceratium (new class).

3 Testing Methodology

To evaluate model performance across diverse real-world conditions, we established four distinct testing scenarios representing common challenges in zooplankton image analysis:

3.1 Test 1: Optimal Conditions

This test represents the best-case scenario with zooplankton specimens against a predominantly black background. These conditions most closely match our synthetic training data and provide a baseline for maximum expected performance.

3.2 Test 2: Light Gradient Background

Occurring in approximately 20-30% of all samples, this test evaluates the models' robustness to irregular background illumination.

3.3 Test 3: Filament Contamination

This test evaluates performance on samples containing numerous filamentous structures mixed with zooplankton.

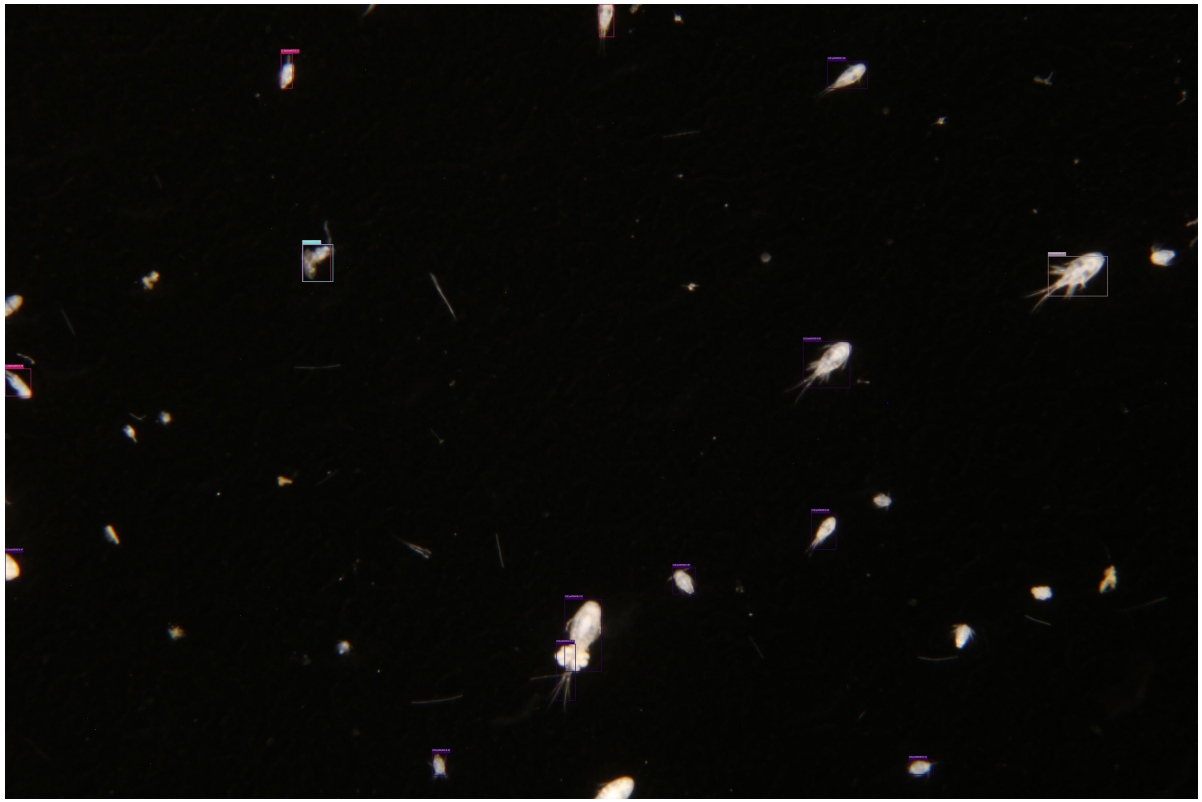
3.4 Test 4: Ceratium Abundance

The final test scenario features high concentrations of Ceratium. This test specifically evaluates the effectiveness of addition of Ceratium to the training set.

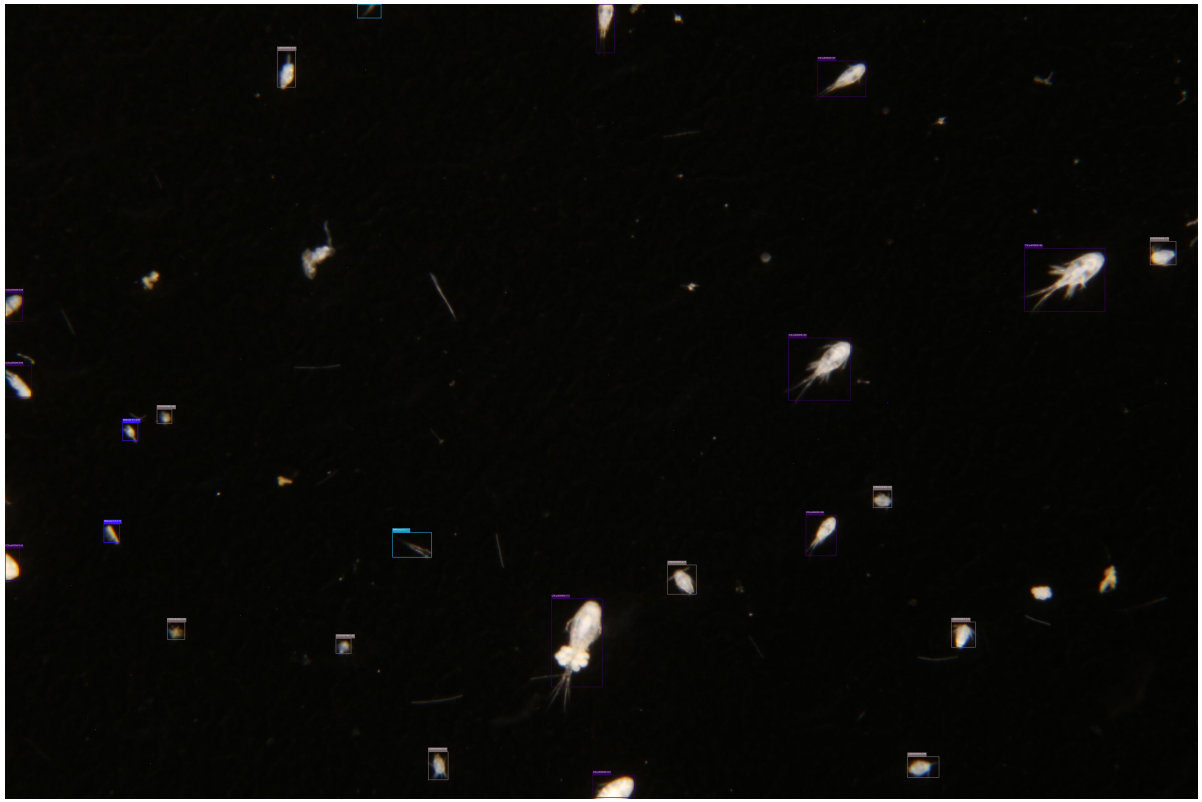
4 Results

Visual inspection of detection results reveals progressive improvement across model iterations, with particularly notable advances in Model 3's performance on Ceratium scenarios. Quantitative performance metrics are currently pending as there is no ground truth available for the test set, since all manually verified data was utilized in the training process.

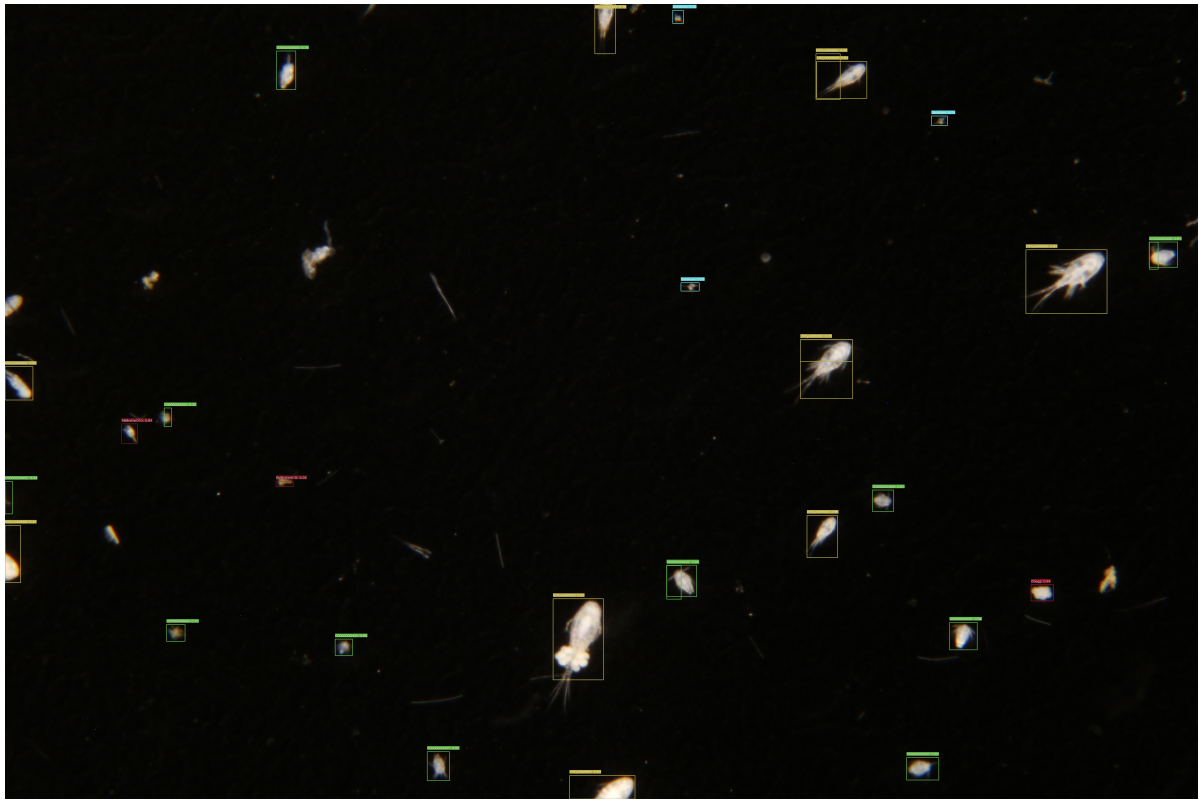
Model 1 result on test 1



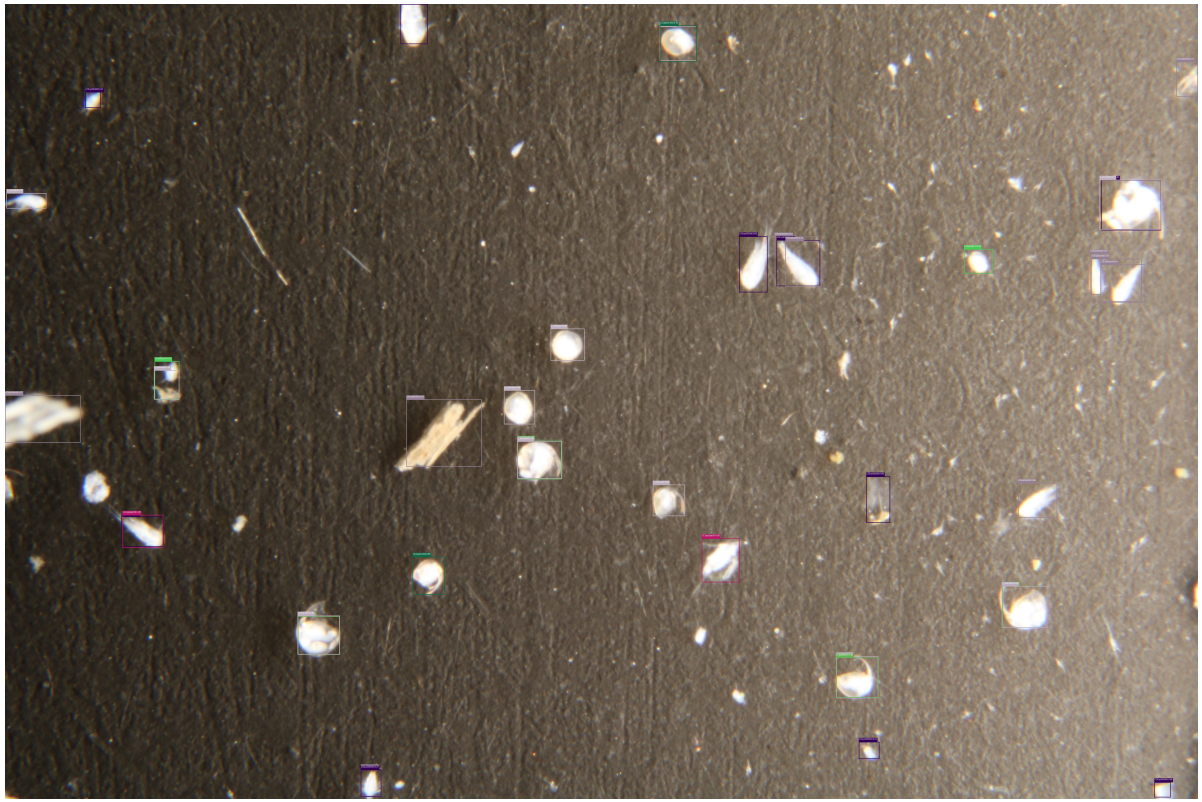
Model 2 result on test 1



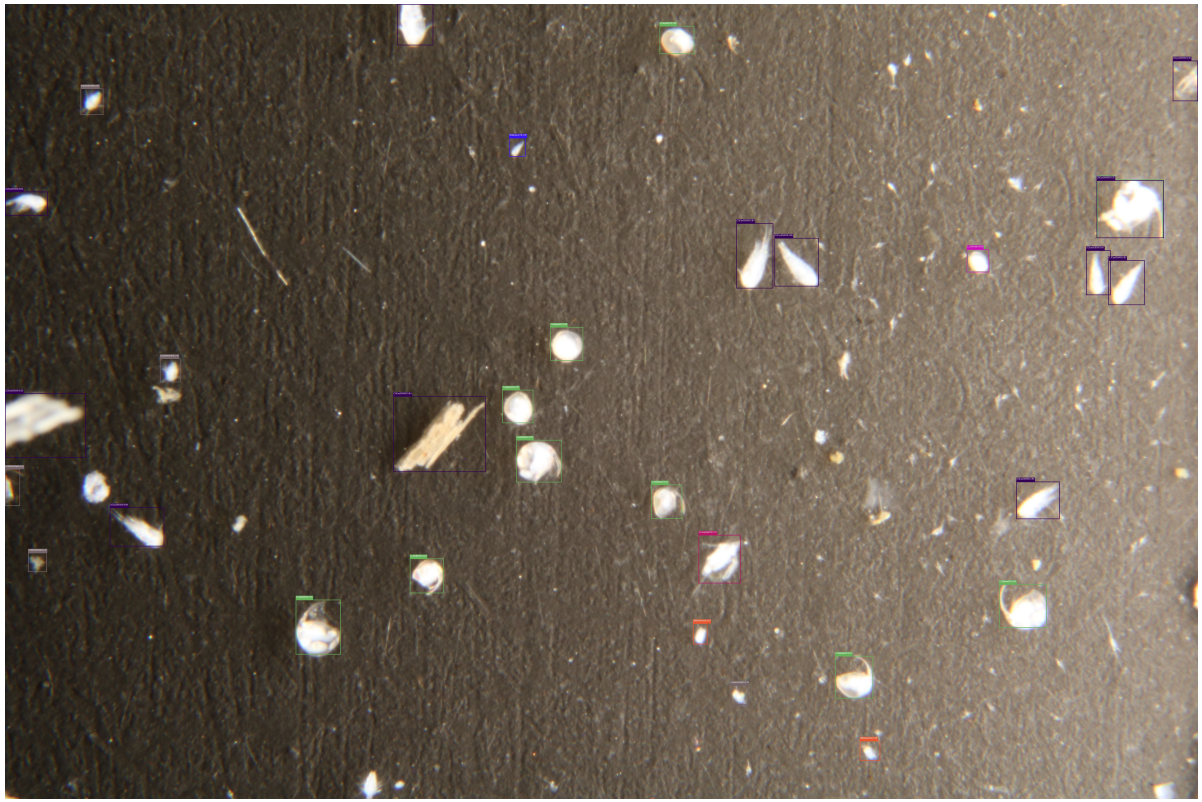
Model 3 result on test 1



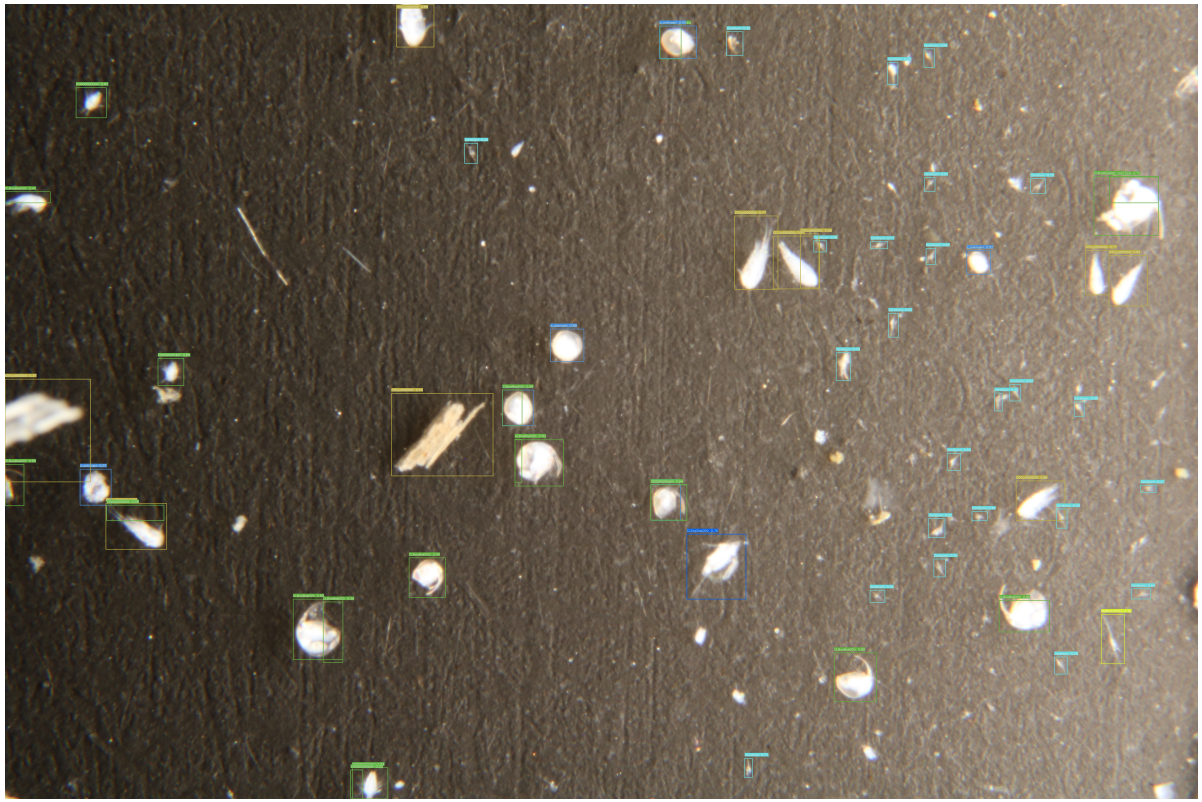
Model 1 result on test 2



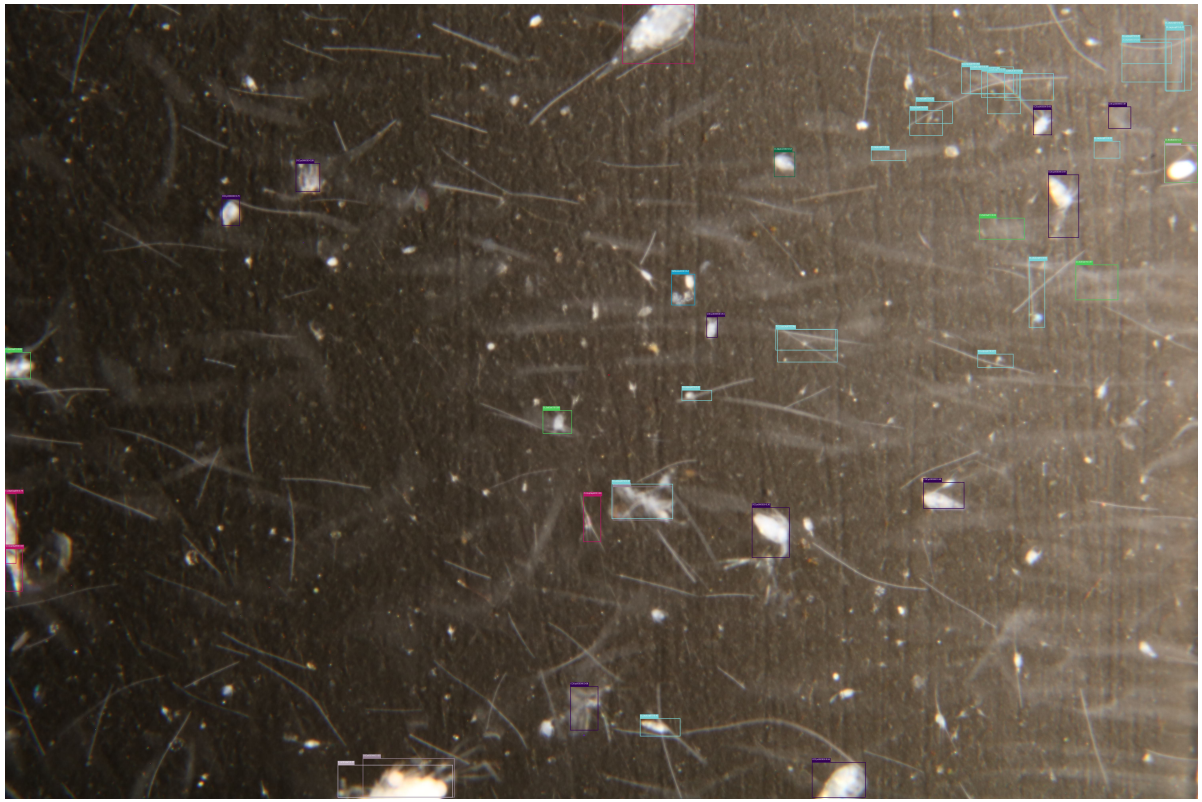
Model 2 result on test 2



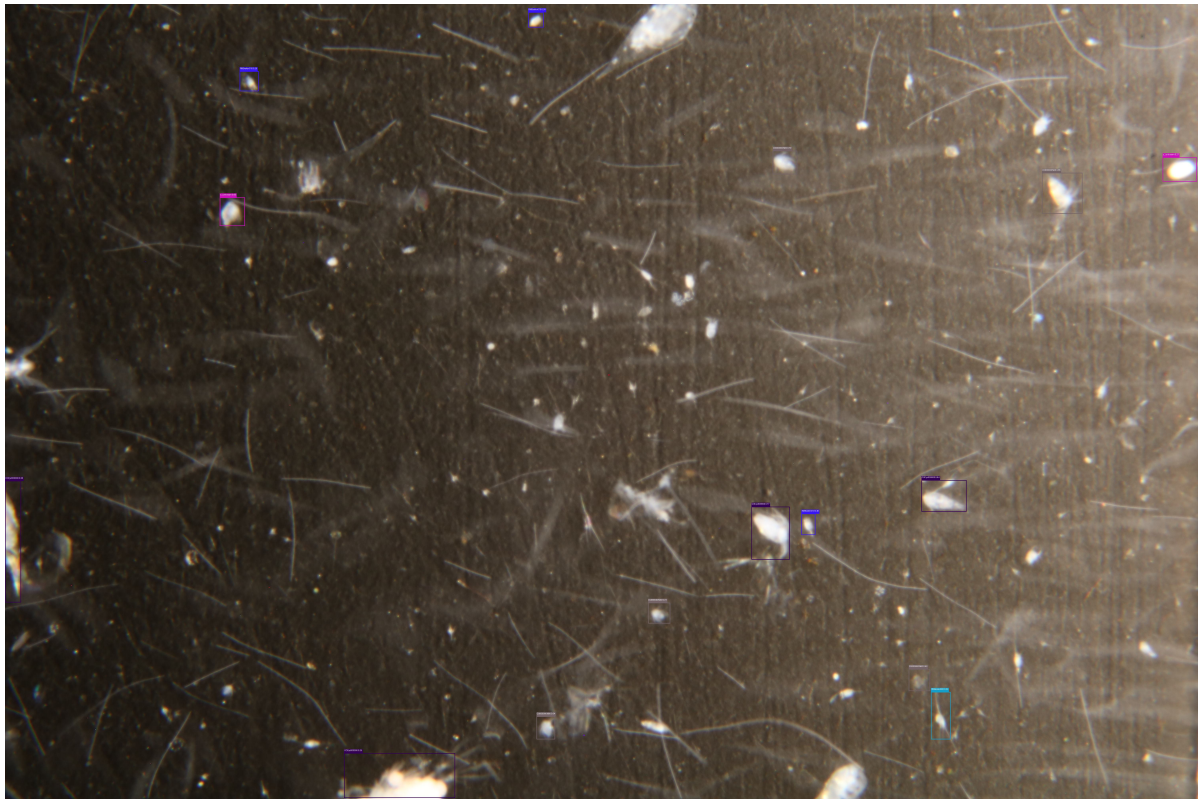
Model 3 result on test 2



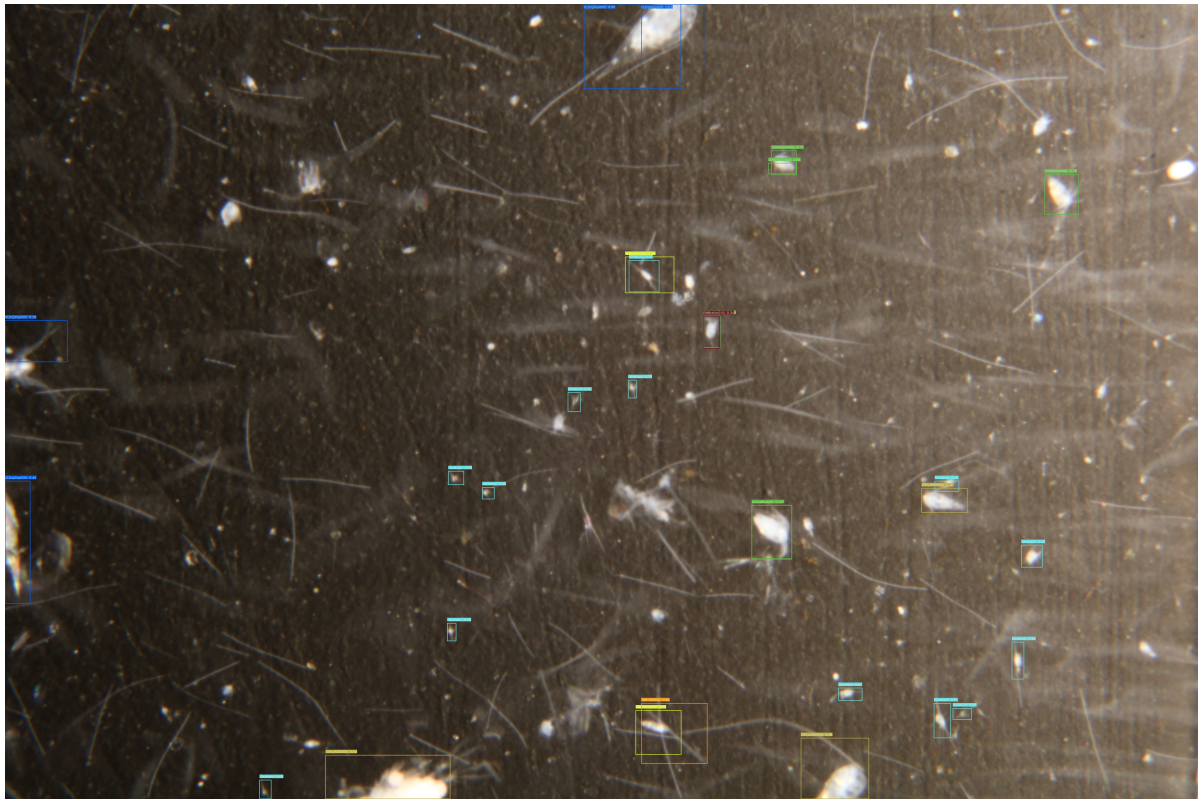
Model 1 result on test 3



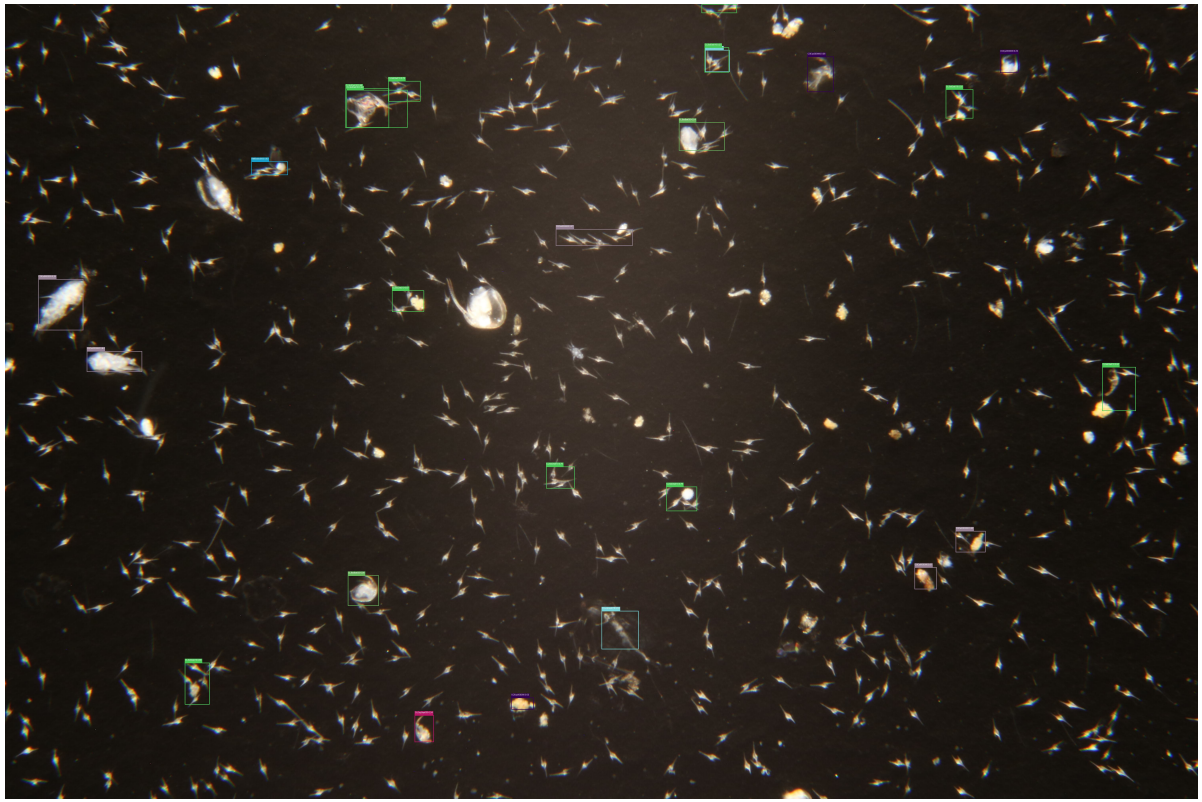
Model 2 result on test 3



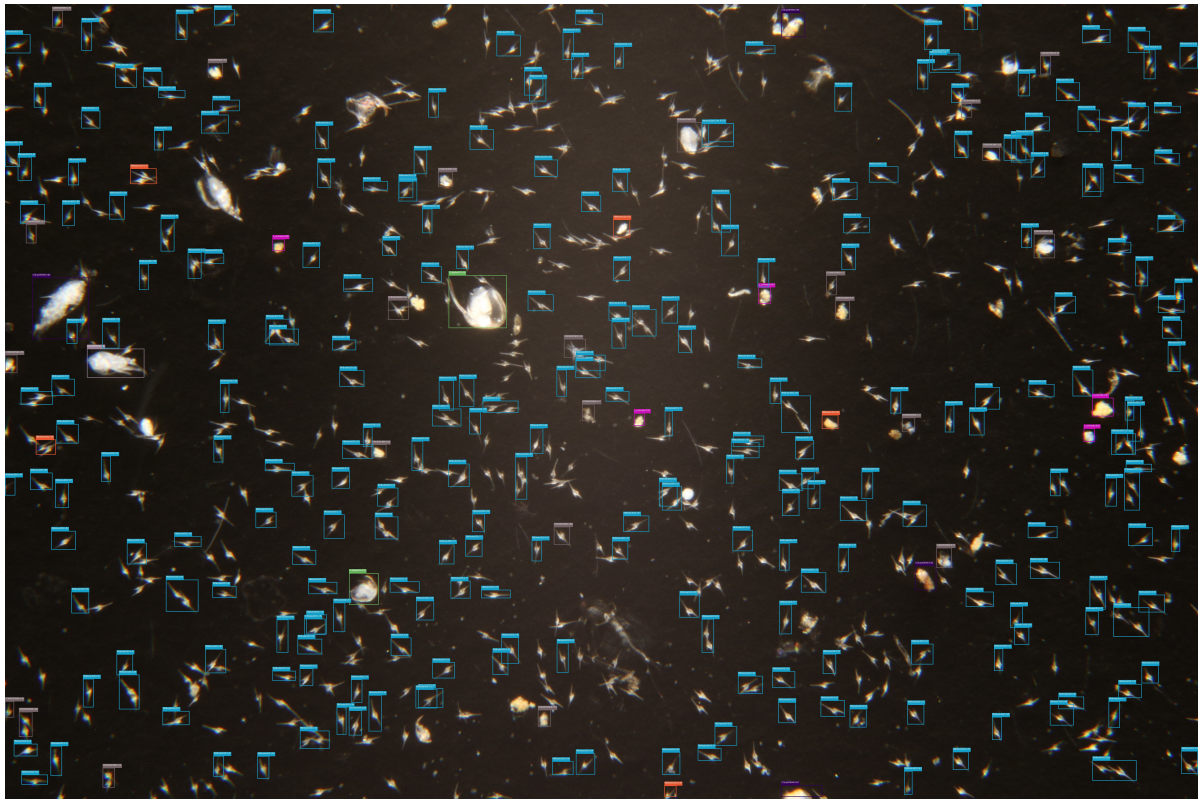
Model 3 result on test 3



Model 1 result on test 4



Model 2 result on test 4



Model 3 result on test 4