# Bird Tracking

Zoltán Szarvas, Mohammad Al-Hitawi, Bence Csabai, and Botond A. Nás

June 29, 2022

## 1 Project Description

The goal of this project is to develop and train a state-of-the-art detection and tracking model to track birds in a cage with bounding boxes.

## 2 Bird Dataset

The models are trained with a proprietary dataset. The dataset contains images (train and validation) and annotations. The annotations contain the image ID, the category ID, bounding box values and other meta parameters. See Figures 2 and 2 for examples of images in the dataset.


Figure 1: Dataset sample


Figure 2: Dataset sample

# 3 Tools Used

## 3.1 YOLOX

YOLOX is a high-performing object detector. The YOLO series is a constant exploration of techniques to improve the object detection for optimal speed and accuracy trade-off for real-time applications. Some key features of the YOLOX tracker include: (i) anchor-free detectors to significantly reduce the number of design parameters and (ii) a decoupled head for classification, regression, and localization to improve convergence speed [3].

## 3.2 BYTE

BYTE is a simple and effective association method for multi-object tracking (MOT), which makes full use of detection boxes from high scores to low ones in the matching process. It first matches the high score detection boxes to the tracklets based on motion similarity or appearance similarity. Then, it performs the second matching between the unmatched tracklets, i.e. the tracklet in red box (see Figure 3.2), and the low score detection boxes using the same motion similarity [4].

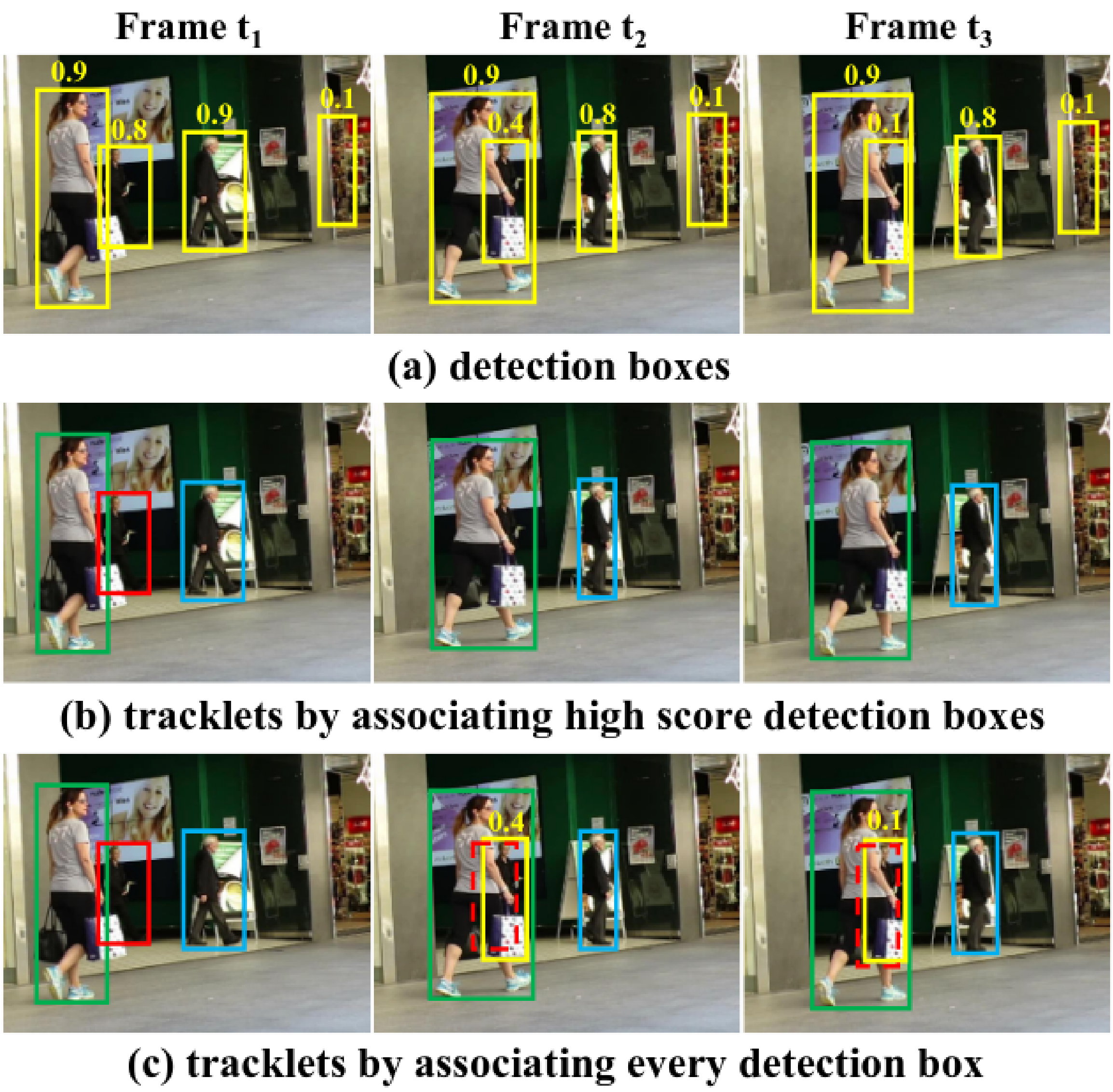


Figure 3: ByteTrack

### 3.3 ByteTrack

ByteTrack performs MOT on a video using the high-performance detector YOLOX and performs association using the BYTE technique. The combination of the two results in a robust and effective MOT algorithm [4].

## 4 Execution and Evaluation

### 4.1 Detection Evaluation

To evaluate detection performance, we used the COCO evaluation metrics [2].

```
Average Precision (AP):
   AP                   % AP at IoU=.50:.05:.95 (primary challenge metric)
   AP^IoU=.50           % AP at IoU=.50 (PASCAL VOC metric)
   AP^IoU=.75           % AP at IoU=.75 (strict metric)
AP Across Scales:
   AP^small             % AP for small objects: area < 32^2
   AP^medium            % AP for medium objects: 32^2 < area < 96^2
   AP^large             % AP for large objects: area > 96^2
Average Recall (AR):
   AR^max=1             % AR given 1 detection per image
   AR^max=10            % AR given 10 detections per image
   AR^max=100           % AR given 100 detections per image
AR Across Scales:
   AR^small             % AR for small objects: area < 32^2
   AR^medium            % AR for medium objects: 32^2 < area < 96^2
   AR^large             % AR for large objects: area > 96^2
```

Figure 4: Detection Evaluation

The values as stated are average precision and average recall on intersection over union (IoU) thresholds between 0.5 and 0.95 where maximum detections over which it is calculated is 100. In other words, precision is calculated over number of positive detections (greater than IoU threshold) by total number of true positives and false positives. Then, the average of precision values over each threshold is calculated to obtain the final value in the last column. The same process is repeated for recall.

## 4.2 Tracking Evaluation

To evaluate tracking performance, we used the MOT Challenge [1] evaluation metrics as described by Figure 4.2.

## 4.3 Baseline Reproduction

We ran the tracker on the dataset using the ByteTrack pretrained model, that is, without any training on the bird dataset.

| $AP$ | $AP^{IoU=.50}$ | $AP^{IoU=0.75}$ | $AR^{max=1}$ | $AR^{max=10}$ | $AR^{max=100}$ |
|---|---|---|---|---|---|
| **0.483** | 0.880 | 0.492 | 0.162 | 0.632 | 0.633 |

Table 1: Baseline detection metrics

| MOTA | MOTP | IDF1 | MT | ML | FP | FN | Rcll | Prcn |
|---|---|---|---|---|---|---|---|---|
| 79.0% | 0.237 | 86.9% | 3 | 0 | 14 | 134 | 81.2% | 97.6% |

Table 2: Baseline tracking metrics

Lower is better. Higher is better.

| Measure | Better | Perfect | Description |
|---|---|---|---|
| **MOTA** | higher | 100% | Multi-Object Tracking Accuracy (+/- denotes standard deviation across all sequences) [1]. This measure combines three error sources: false positives, missed targets and identity switches. |
| **IDF1** | higher | 100% | ID F1 Score [2]. The ratio of correctly identified detections over the average number of ground-truth and computed detections. |
| **HOTA** | higher | 100% | Higher Order Tracking Accuracy [3]. Geometric mean of detection accuracy and association accuracy. Averaged across localization thresholds. |
| **MT** | higher | 100% | Mostly tracked targets. The ratio of ground-truth trajectories that are covered by a track hypothesis for at least 80% of their respective life span. |
| **ML** | lower | 0% | Mostly lost targets. The ratio of ground-truth trajectories that are covered by a track hypothesis for at most 20% of their respective life span. |
| **FP** | lower | 0 | The total number of false positives. |
| **FN** | lower | 0 | The total number of false negatives (missed targets). |
| **Rcll** | higher | 100% | Ratio of correct detections to total number of GT boxes. |
| **Prcn** | higher | 100% | Ratio of TP / (TP+FP). |

Figure 5: Tracking Evaluation

## 4.4 Experiments

### 4.4.1 Experiment 1

The first experiment consisted of training the ByteTrack model on the bird dataset for 5 epochs.

| $AP$ | $AP^{IoU=.50}$ | $AP^{IoU=0.75}$ | $AR^{max=1}$ | $AR^{max=10}$ | $AR^{max=100}$ |
|---|---|---|---|---|---|
| **0.724** | 0.934 | 0.807 | 0.230 | 0.770 | 0.771 |

Table 3: Exp1 detection metrics

| MOTA | MOTP | IDF1 | MT | ML | FP | FN | Rcll | Prcn |
|---|---|---|---|---|---|---|---|---|
| 69.7% | 0.176 | 78.2% | 2 | 0 | 33 | 180 | 74.8% | 94.2% |

Table 4: Exp1 tracking metrics

### 4.4.2 Experiment 2

The second experiment consisted of training the ByteTrack model on the bird dataset for 10 epochs. Training for more epochs yielded better results.

| $AP$ | $AP^{IoU=.50}$ | $AP^{IoU=0.75}$ | $AR^{max=1}$ | $AR^{max=10}$ | $AR^{max=100}$ |
|---|---|---|---|---|---|
| **0.720** | 0.926 | 0.851 | 0.223 | 0.761 | 0.761 |

Table 5: Exp2 detection metrics

| MOTA | MOTP | IDF1 | MT | ML | FP | FN | Rcll | Prcn |
|---|---|---|---|---|---|---|---|---|
| 80.4% | 0.161 | 88.1% | 3 | 0 | 18 | 120 | 83.2% | 97.1% |

Table 6: Exp2 tracking metrics

### 4.4.3 Experiment 3

The third experiment consisted of training the ByteTrack model on the bird dataset for 25 epochs. The warmup epoch parameter was also changed to 5 and the learning rate was increased to 0.01/64.0. The weight decay was increased to $5e-3$ and the momentum was changed to 0.7.

| $AP$ | $AP^{IoU=.50}$ | $AP^{IoU=0.75}$ | $AR^{max=1}$ | $AR^{max=10}$ | $AR^{max=100}$ |
|---|---|---|---|---|---|
| **0.622** | 0.937 | 0.675 | 0.207 | 0.714 | 0.716 |

Table 7: Exp3 detection metrics

| MOTA | MOTP | IDF1 | MT | ML | FP | FN | Rcll | Prcn |
|---|---|---|---|---|---|---|---|---|
| 71.6% | 0.189 | 85.5% | 3 | 0 | 86 | 117 | 83.6% | 87.4% |

Table 8: Exp3 tracking metrics

## 5 Results

Using the ByteTrack tracker, we were able to achieve around 80% MOTA score. The experiment that yielded the best results was the base model parameters trained for 10 epochs. Increasing the learning rate helped the model with false negatives, but resulted in more false positives during the tracking. Expanding on this project would consist of tuning the parameters more to find the ideal model for our use case and optionally train the model with more data from different sources.

## References

[1] URL: https://motchallenge.net/results/MOT20/.

[2] *Common objects in context.* URL: https://cocodataset.org/#detection-eval.

[3] Zheng Ge et al. *YOLOX: Exceeding YOLO Series in 2021.* 2021. eprint: arXiv:2107.08430.

[4] Yifu Zhang et al. *ByteTrack: Multi-Object Tracking by Associating Every Detection Box.* 2021. eprint: arXiv:2110.06864.