Cross-Camera Multi-Target Vehicle Tracking

[1,*] *Yu-Jen Chen*（陳郁仁）, [2] *Tzu-Chia Tung*（董子嘉）, [1] *Jia-Hao Kang*（康家豪）, [2] *Yu-Chin Chu* (朱昱瑾), [1] *Chiou-Shann Fuh* (傅楸善), [2,3] *Ping-Han Chen* (陳稟翰), [2] *Yu-Tang Liao* (廖昱唐), [2] *Chung-Ming Yang* (楊仲民), [1,2] *Li-Jin Huang* (黃麗瑾), [2] *Shu-Ru Huang* (黃淑如)

[1] Department of Computer Science and Information Engineering,
National Taiwan University, Taipei, Taiwan,
[2] Internet of Things Laboratories, Chunghwa Telecom Laboratories, Taoyuan, Taiwan
[3] Department of Engineering Science, National Cheng Kung University, Tainan, Taiwan
[*] E-mail: b11902113@ntu.edu.tw d04944016@ntu.edu.tw d04944016@ntu.edu.tw
chevady@cht.com.tw fuh@csie.ntu.edu.tw rayche522@cht.com.tw yutang@cht.com.tw
cmyang@cht.com.tw srhuang@cht.com.tw

ABSTRACT

This study provides a method for cross-camera multi-target vehicle tracking based on the BoT-SORT model applying the FastReID model and the You Only Look Once version 7 (YOLOv7) model. We determined hyperparameters for both FastReID and YOLOv7 models to achieve the best performance. This research explores the implementation of deep learning for cross-camera multi-target vehicle tracking and the potential advancements in traffic enforcement cameras. The fine-tuned BoT-SORT model shows high performance on the task and ranks 25th out of 52 active participants at the AICUP 2024 private leaderboard.

Keywords: *Cross-camera Tracking, Bot-SORT, FastReID, YOLOv7, Hyperparameters*

1. INTRODUCTION

Traffic enforcement cameras are necessary in modern cities. FastReID is a machine-learning model mainly used in multiple object tracking. YOLOv7 is a deep-learning model for object detection. With their high accuracy, using these models to implement cross-camera multi-target vehicle tracking is possible. This study aims to find the best hyperparameters for models to track vehicles across images from different traffic enforcement cameras.

2. RELATED WORK

2.1 FastReID

Re-IDentification algorithms (ReID) are able to assign unique identification numbers to each object in different frames in time series images. FastReID is a ReID application provided by Lingxiao He et al. [1]. It provides state-of-the-art models including vehicle ReID, which could be applied to cross-camera multi-target vehicle tracking tasks.

2.2 YOLOv7

You Only Look Once version 7 (YOLOv7) is a deep learning model for real-time object detection based on Convolutional Neural Networks (CNN) [2]. YOLOv7 has high accuracy on vehicle detection tasks after pre-training on the Common Object in COntext (COCO) dataset [3].

2.3 Bot-SORT

Simple Online and Real-time Tracking (SORT) is a method for multiple-object tracking based on the Kalman filter and Hungarian algorithm [4]. Bot-SORT is a multiple-object tracker using SORT and FastReID for pedestrian tracking [5]. With the combination of Bot-SORT and YOLOv7 pre-trained on the COCO dataset, this method can be applied to cross-camera multi-target vehicle tracking tasks.

3. METHODOLOGY

3.1 Data Preparation

The dataset used in this study is provided by AI CUP 2024. The dataset consists of images from seven different traffic enforcement cameras recorded at a frame rate of 1 frame per second. Data formats are transformed to meet the requirements of Bot-SORT and YOLOv7.

Each image in the dataset has a resolution of 1280 * 720. The label files recorded the default bounding boxes and ID numbers of each car.

3.2 Fine-tuning FastReID Model

Several hyperparameters of the FastReID Model can be modified to achieve the best performance, including learning rate, epoch number, and batch size. The model is trained under different hyperparameter settings to find out which setting yields the best outcome. The model's performance is examined by the accuracy metric.

3.3 Fine-tuning YOLOv7 Model

Several hyperparameters of the YOLOv7 Model can be modified to achieve the best performance, including learning rate, epoch number, and batch size. The model is trained under different hyperparameter settings to find out which setting yields the best outcome. The model's performance is examined by the mAP@0.5 metric.

3.4 Evaluation of Bot-SORT Model

The private testing set without labels is provided by AI CUP 2024 [6]. We implemented the Bot-SORT model using fine-tuned Fase-ReID model and YOLOv7 model. After we predicted the bounding boxes and ID for each car in the private testing set using Bot-SORT model, the results were submitted to the website of AI CUP 2024 and evaluated. The evaluation criteria are IDF1 score plus MOTA score.

4. EXPERIMENT

4.1 Experiments of Fine-tuning FastReID Model

During the experiments, only one hyperparameter is modified at a time. The following table shows the default hyperparameter configuration.

Table 1. Default Hyperparameter Configuration.

Variable Name	Values
Learning Rate	0.00035
Epoch Number	10
Batch Size	256

Different learning rates are tested, including 0.0001, 0.00035, and 0.001. Epoch numbers are modified between 5, 10, and 50. Batch sizes are changed from 16 to 64 to 256.

4.2 Experiments of Fine-tuning YOLOv7 Model

During the experiments, only one hyperparameter is modified at a time. The following table shows the default hyperparameter configuration.

Table 2. Default Hyperparameter Configuration.

Variable Name	Values
Learning Rate	0.01
Epoch Number	3
Batch Size	16

Different learning rates are tested, including 0.01, 0.0035, and 0.1. Epoch numbers are modified between 3, 10, and 20. Batch sizes are changed from 3 to 10 to 30.

5. RESULTS

5.1 Results of Fine-tuning Bot-SORT Model

The following table shows the model accuracy after training with different learning rates.

Table 3. Model Accuracy ↑ with Different Learning Rates.

Learning Rate	Accuracy
0.0001	96.29 %
0.00035	**98.83 %**
0.001	97.66 %

As we can see, the model with a learning rate of 0.00035 has the highest accuracy after training for 10 epochs.

The following table shows the model total loss (cross-entropy loss) after training with different learning rates.

Table 4. Model Loss ↓ with Different Learning Rates.

Learning Rate	Total Loss
0.0001	1.571
0.00035	**1.388**
0.001	1.515

The result shows that the model with a learning rate of 0.00035 has the lowest total loss after training for 10 epochs. The results of both the accuracy metric and the total loss imply that the model with a learning rate of 0.00035 has the best performance.

The Learning rate of a model also affects its convergence of accuracy. The following figures illustrate how the accuracy changes with iteration number increases of different model learning rates.

Fig. 1. Learning rate = 0.0001.

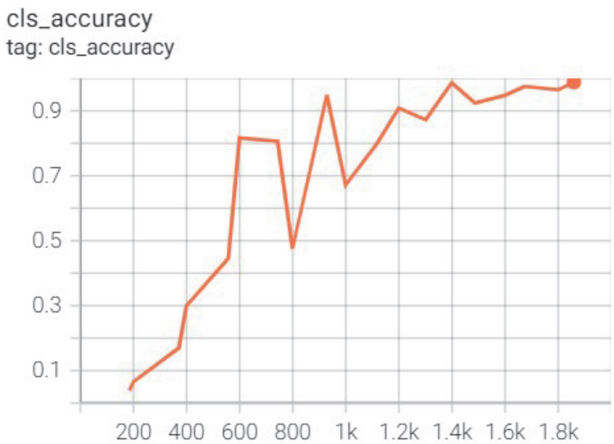

Fig. 2. Learning rate = 0.00035.

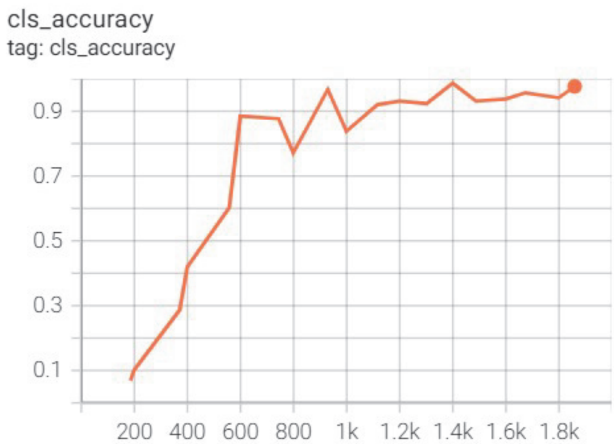

Fig. 3. Learning rate = 0.001.

The results show that the accuracy converges earlier when the learning rate is greater. The same pattern can be found in the relation of total loss and iteration numbers. The following figures illustrate how the total loss (cross-entropy loss) changes with iteration number increases of different model learning rates.

Fig. 4. Learning rate = 0.0001.

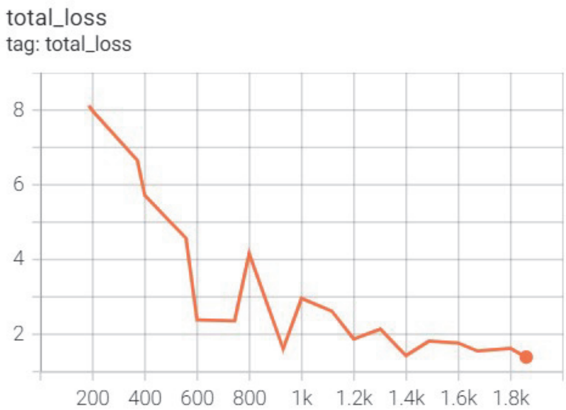

Fig. 5. Learning rate = 0.00035.

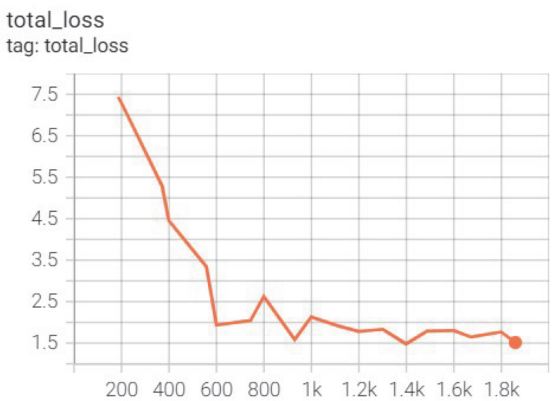

Fig. 6. Learning rate = 0.001.

The results indicate that the total loss converges earlier when the learning rate is greater. Therefore, we decided to use learning rate of 0.00035 in training our final predictor model.

The following table shows the model accuracy after training with different batch sizes.

Table 5. Model Accuracy ↑ with Different Batch Sizes.

Batch Size	Accuracy
16	43.75 %
64	93.75 %
256	**98.83 %**

The result shows that the model with a batch size of 256 has the lowest total loss after training for 10 epochs. However, using larger batch size for training requires more GPU memory spaces. The following figures illustrate how the accuracy changes with iteration number increases of different training batch sizes.

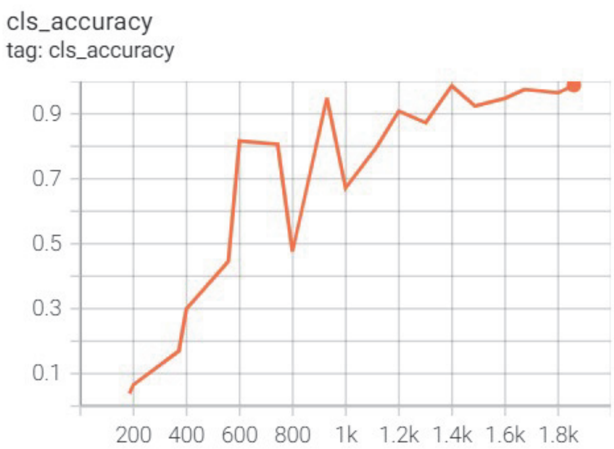

Fig. 7. Batch size = 256.

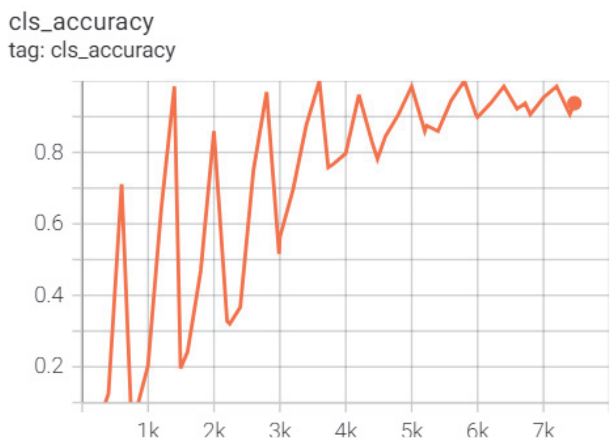

Fig. 8. Batch size = 64

The result shows that the learning curve is more unstable when the training batch size is smaller. Therefore, we decided to use learning rate of 256 in training our final predictor model.

The following table shows the model accuracy after training with different epoch numbers.

Table 6. Model Accuracy ↑ with Different Epoch.

Epoch Number	Accuracy
5	94.92 %
10	98.83 %
20	**99.21 %**

The result shows that the model accuracy is higher when the training epoch number is higher. Therefore, we decided to use epoch number of 20 in training our final predictor model.

5.2 Results of Fine-tuning YOLOv7 Model

Figure 9 shows the training result of the model using default hyperparameter configuration.

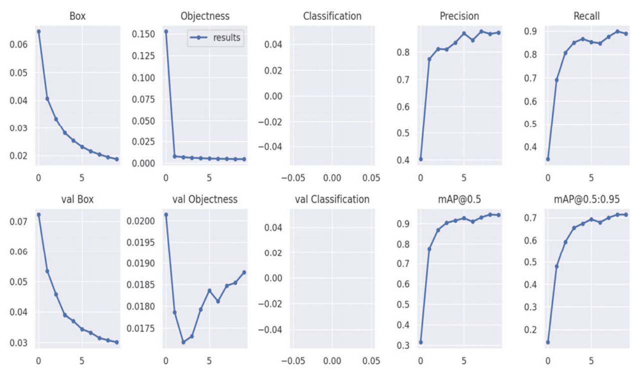

Fig. 9. Result of YOLOv7 model.

The result shows that the mAP@0.5 value increases as epoch number increases. The value of mAP@0.5 converges to 0.95. The box loss also decreases smoothly as epoch number increases.

The following images are demonstrations of YOLOv7 model predictions. The bounding boxes of each cars can be correctly identified and labeled.

Fig. 10. YOLOv7 model predictions

Due to long training time and memory limitations, the experiments of fine-tuneing YOLOv7 model are not complete yet.

5.3 Results of the BoT-SORT Model

The final public score (IDF1 + MOTA) of the fine-tuned BoT-SORT model is 0.766359 and ranked 29th out of 52 active participants. The final private score (IDF1 + MOTA) of the fine-tuned BoT-SORT model is 0.868704 and ranked 25th out of 52 active participants.

6. CONCLUSION

Applying FastReID and YOLOv7 models to resolve cross-camera multiple-vehicle tracking is possible. The fine-tuning experiments show that using a learning rate of 0.0035 on the FastReID model yields the best performance. In addition, we set the batch size to 256 and the epoch number to 20 in order to achieve the highest accuracy. The result of YOLOv7 model experiments also shows a high mAP@0.5 value after training of 10 epochs.

After these hard works, we are ranked 25th out of 52 active participants at the AICUP 2024 private leaderboard.

7. FUTURE WORK

7.1 Fine-tuning FastReID Model

More hyperparameter configuration experiments should be conducted, including but not limited to input size, drop ratio, and optimizer.

Data augmentation could be applied to increase the size of the training set, and to improve the model's performance.

7.2 Fine-tuning YOLOv7 Model

Experiments of fine-tuning the YOLOv7 Model should be completed. In the future, we will use GPU to speed up experiments.

More hyperparameter configuration experiments should be conducted, including but not limited to input size, drop ratio, and optimizer.

7.3 Implementing BoT-SORT Model

BoT-SORT Model is implemented based on both the FastReID model and the YOLOv7 model. Some experiments should be conducted to find out which combination of the FastReID model and the YOLOv7 model performs the best based on IDF1 + MOTA score.

8. REFERENCES

[1] Lingxiao He, Xingyu Liao, Wu Liu, Xinchen Liu, Peng Cheng, and Tao Mei. 2023. FastReID: A Pytorch Toolbox for General Instance Re-identification. *Proceedings of the 31st ACM International Conference on Multimedia (MM '23)*. Association for Computing Machinery, New York, NY, 9664–9667. https://doi.org/10.1145/3581783.3613460

[2] C. Wang, A. Bochkovskiy, H. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 7464-7475.

[3] N. S. Munir, N. Hossain, R. R. Zame and G. Sarowar, "Vehicle Detection of Bangladesh using YOLOv7 with Hyper-parameter Tuning," 2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP), VIJAYAWADA, India, 2023, pp. 1-5, doi: 10.1109/AISP57993.2023.10135057.

[4] N. Wojke, A. Bewley and D. Paulus, "Simple online and realtime tracking with a deep association metric," *2017 IEEE International Conference on Image Processing (ICIP)*, Beijing, China, 2017, pp. 3645-3649, doi: 10.1109/ICIP.2017.8296962.

[5] N. Aharon, R. Orfaig, B. Bobrovsky. "BoT-SORT: Robust Associations Multi-Pedestrian Tracking," in arXiv preprint arXiv:2206.14651, 2022.