AI VIETNAM
Seminar

EDA AND PIPELINE
AICITY CHALLENGE

CVPRW-Track 5

Year

Outline

- EDA Data Track 5
- Top model 2023
- Basic Baseline

Outline

- **EDA Data Track 5**
- **Top model 2023**
- **Baseline Basic**

EDA

Tổng quan về bộ dữ liệu

100 video

Ấn Độ

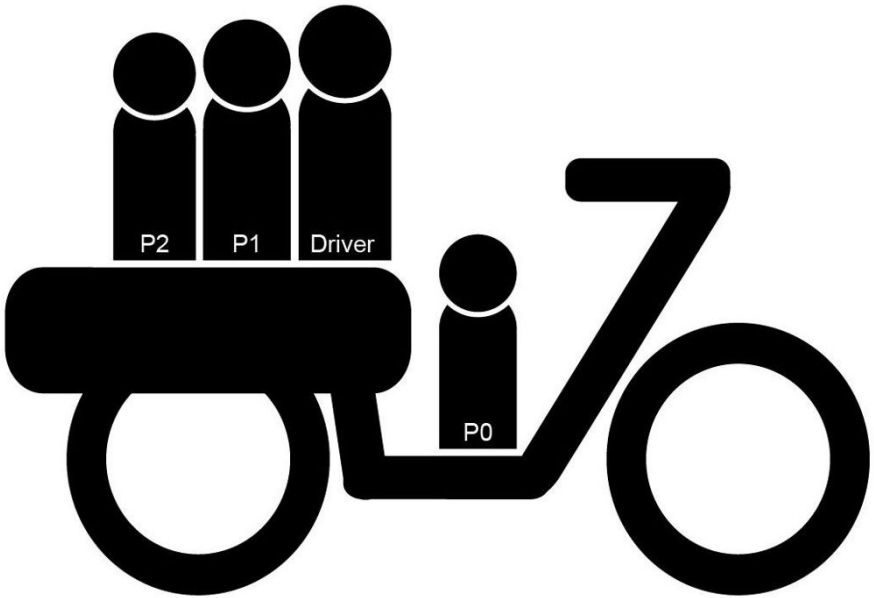
20s/video

10 FPS

1920x1080

EDA

ANNOTATION



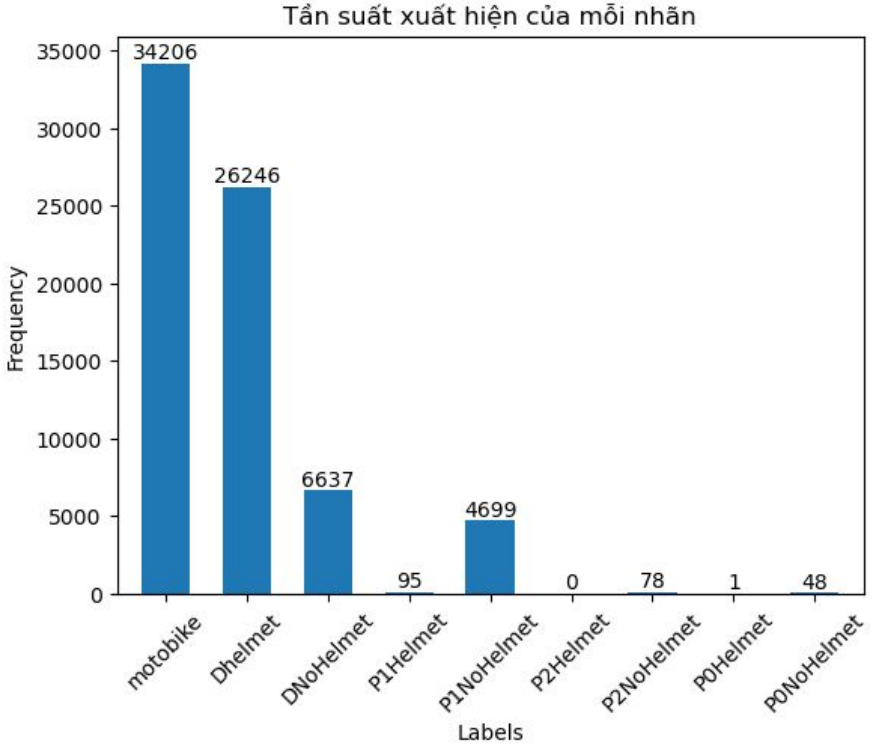
Chú thích label

1. motorbike: bounding box của xe máy
2. DHelmet: bounding box của người lái xe. Nếu người đó đội mũ
3. DNoHelmet: bounding box của người lái xe. Nếu người đó không đội mũ
4. P1Helmet: bounding box của người ngồi sau thứ nhất. Nếu người đó đội mũ
5. P1NoHelmet: bounding box của người ngồi sau thứ nhất. Nếu người đó không đội mũ
6. P2Helmet: bounding box của người ngồi sau thứ 2. Nếu người đó đội mũ
7. P2NoHelmet: bounding box của người ngồi sau thứ hai. Nếu người đó không đội mũ
8. P0Helmet: bounding box của trẻ em ngồi trước người lái. Nếu người đó đội mũ
9. P0NoHelmet: bounding box của trẻ em ngồi trước người lái. Nếu người đó không đội mũ

NOTE: Năm nay bổ sung thêm P0

EDA

Phân phối nhãn



EDA

Ground Truth Format

⟨video_id⟩, ⟨frame⟩, ⟨bb_left⟩, ⟨bb_top⟩, ⟨bb_width⟩, ⟨bb_height⟩, ⟨class⟩

- ⟨video_id⟩ số thứ tự của video bắt đầu từ 1.
- ⟨frame⟩ số thứ tự frame trong video, bắt đầu từ 1.
- ⟨bb_left⟩ là tọa độ x của điểm trên cùng bên trái của bounding box.
- ⟨bb_top⟩ là tọa độ y của điểm trên cùng bên trái của bounding box.
- ⟨bb_width⟩ chiều rộng của bounding box.
- ⟨bb_height⟩chiều cao của bounding box.
- ⟨class⟩ id nhãn dán của object.

EDA

Trường hợp bình thường

Video 2

Video 7

EDA

Abnormal Case (Trường hợp bất thường)

	Sương mù	Bị chói	Bị nhiễu	Đông xe	Nhiều người	Khó nhìn	Không có gì bất thường
	32	13	13	5	8	20	31

Cách phân loại các trường hợp:
- Dễ: Video được đánh giá “Không có gì bất thường”
- Bình thường: Video gặp 1 - 2 tiêu chí bất thường
- Khó: Video gặp 1 trong các tiêu chí: bị nhiễu, giật, có sương mù; hoặc gặp 3 tiêu chí bất thường trở lên

EDA

Abnormal Case (Trường hợp bất thường)

Video 10: Nhiều dây điện

Video 14: Sương mù, chói do đèn xe

EDA

Abnormal Case (Trường hợp bất thường)

Video 23: Sương mù

Video 27: Đông xe

EDA

Abnormal Case (Trường hợp bất thường)

Video 39: Biển hiệu chắn

Video 56: Nhiễu

EDA

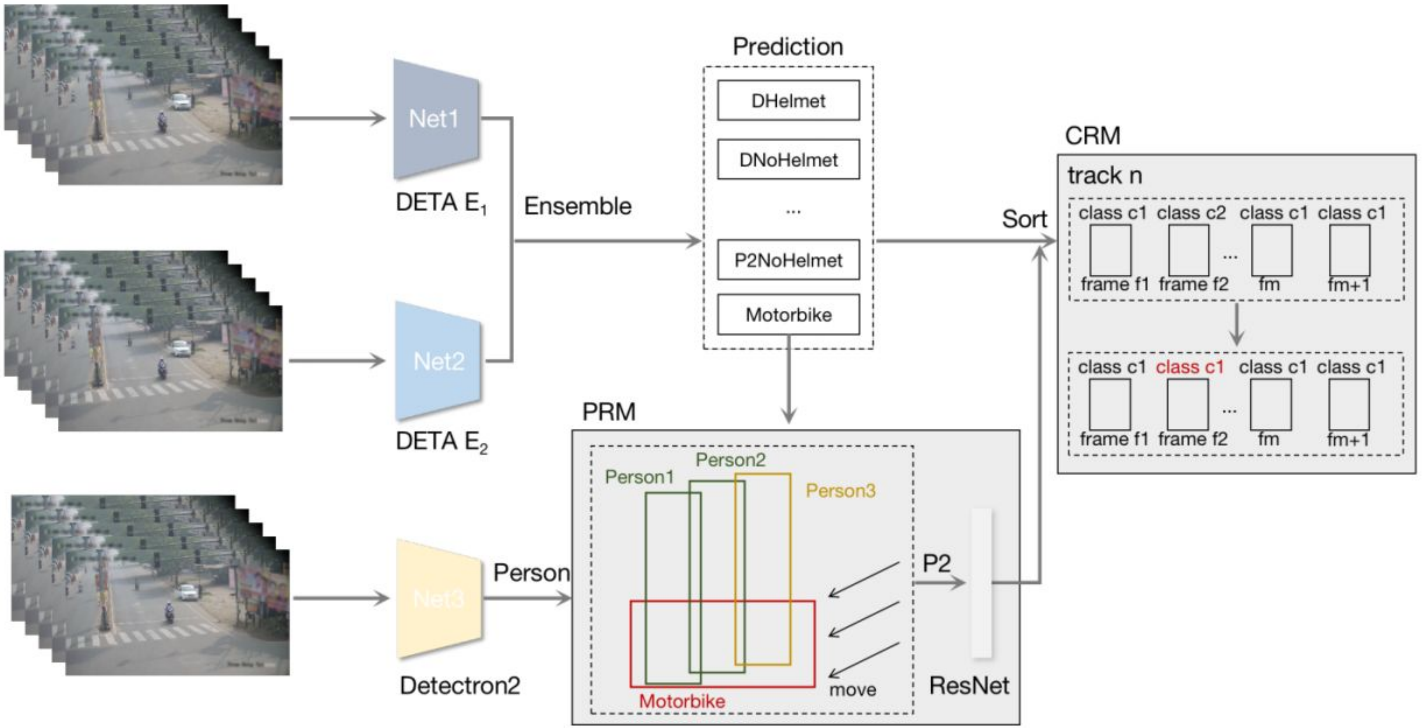
FAQ của AICITY CHALLENGE

- Các đối tượng bé hơn 40px sẽ không được đánh nhãn
- Các vùng bị làm mờ cũng không được đánh nhãn

Outline

- EDA Data Track 5
- **Top model 2023**
- Pipeline Basic

PIPELINE TOP 1 2023



PIPELINE TOP 1 2023

Cách tiếp cận data

- Phân tích data nhận thấy có sự khó khăn trong khi detect trong điều kiện thời tiết xấu

Figure 1. The scenes in the visualization part of the video, from left to right are night, fog, and background chaos, in that order.

PIPELINE TOP 1 2023

Cách tiếp cận data

- Thống kế các nhãn và nhận ra sự mất cân bằng về dữ liệu

Class Id	Class Name	Instances
1	motorbike	31121
2	DHelmet	23220
3	DNoHelmet	6856
4	P1Helmet	94
5	P1NoHelmet	4280
6	P2Helmet	0
7	P2NoHelmet	40

Table 1. Category statistics for all targets in the training set.

PIPELINE TOP 1 2023

Cách tiếp cận data

- Thể hiện nhãn dán thông qua ảnh để nhận xét về nhãn dán từ đó tìm ra phương pháp PRM

Figure 2. Image visualization of Passenger 2.The left and right of the image respectively Passenger 2 appeared in the video 005 and 091.

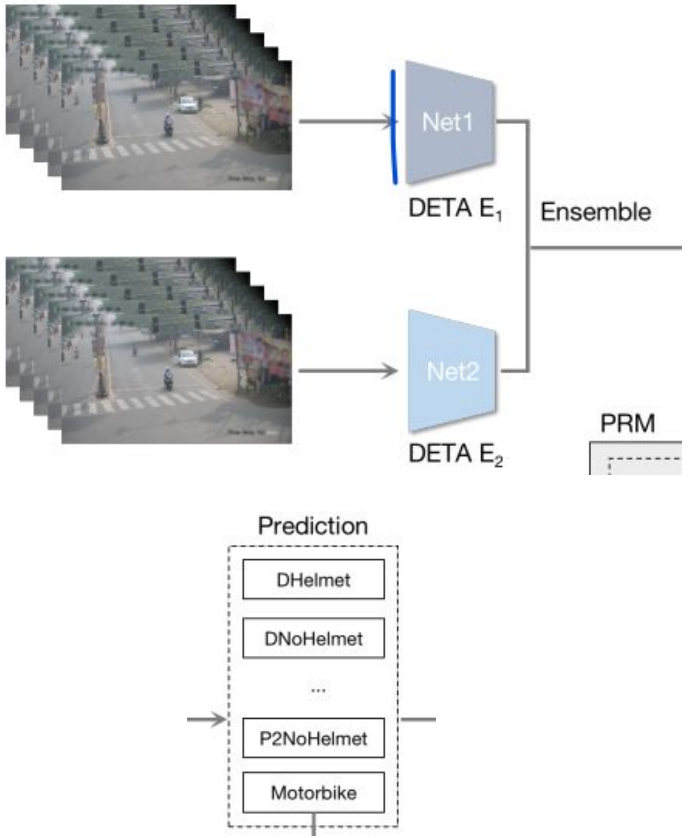
PIPELINE TOP 1 2023

Tổng quan về pipeline

- Đầu tiên dùng 2 mô hình DETA để detect các class và ensemble bằng NMS
- Dùng Detectron 2 để nhận diện người
- Dùng PRM (Passenger Recall Module) dùng để xác định P2
- Sau đó dùng CRM (Category Refine Module) đảm bảo các object khi ra xa khỏi camera vẫn đảm bảo độ chính xác

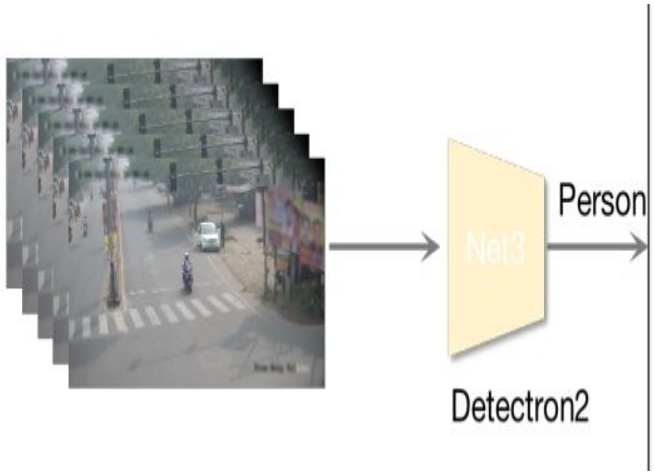
PIPELINE TOP 1 2023

- Dùng 2 mô hình DETA để phát hiện class
- Dùng NMS để ensemble



PIPELINE TOP 1 2023

Dùng Detectron2 để detect person



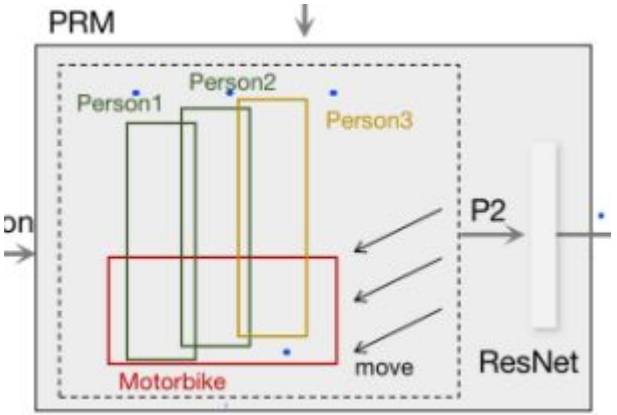
PIPELINE TOP 1 2023

PRM

Xây dựng kỹ thuật để tìm kiếm P2:

- Khi detect người cùng với xe gắn lại với sau
- Dùng SORT để tracking từ đó đoán hướng di chuyển của xe để từ đó để xác định P2
- Sau khi có P2 thì dùng ResNet để detect

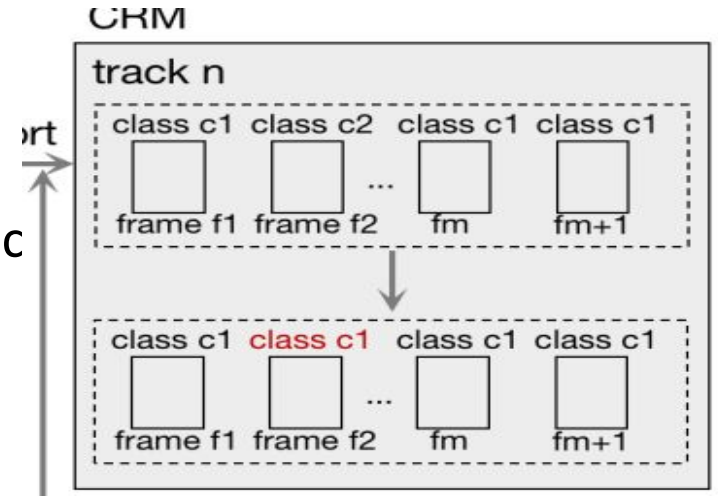
  có mũ bảo hiểm hay không



PIPELINE TOP 1 2023

CRM

Dùng SORT để trackid các class, chọn class c
có số lượng lớn nhất trong trackid
nếu số lượng lớn hơn 50%
thì đổi tất cả class trong trackid thành class c



$$T_{id} = \{(id, b_i, f_i) | i \in v\},$$

PIPELINE TOP 1 2023

KẾT QUẢ THỰC NGHIỆM

Approach	Score
Baseline (DETA)	0.5259
Baseline + Ensemble	0.6973
Baseline + Ensemble + PRM	0.8333
Baseline + Ensemble + PRM + CRM	0.8340

Sau khi dùng PRM độ chính xác của mô hình tăng lên đáng kể.

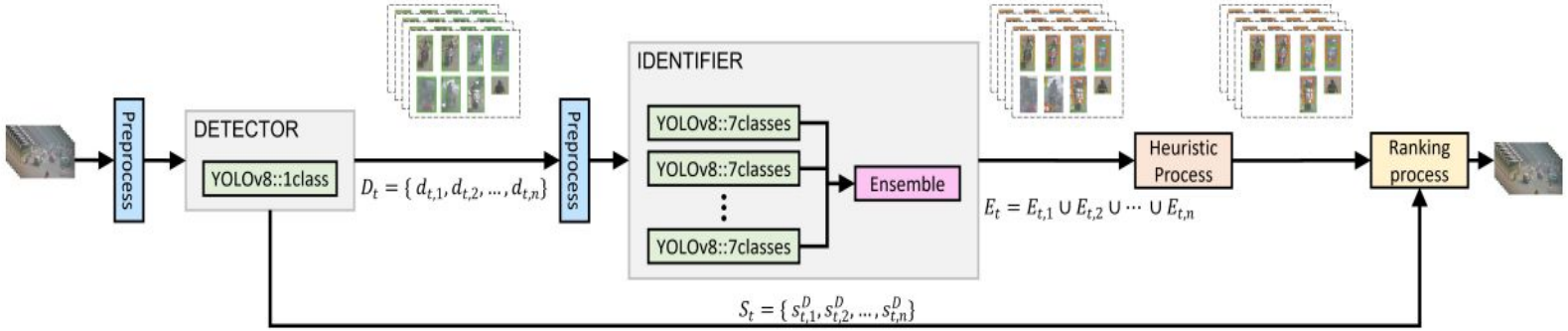
PIPELINE TOP 1 2023

Nhược điểm đối với pipeline top1 đối với data năm nay

Vì sử dụng PRM để tìm kiếm P2 nên sẽ bị fix cứng bởi số người có trên 1 xe

Năm nay bổ sung thêm P0 sẽ khiến cho PRM bị sai lệch.

PIPELINE TOP 2 2023



Cách tiếp cận data

- Thể hiện các điều kiện môi trường gặp của bộ dữ liệu

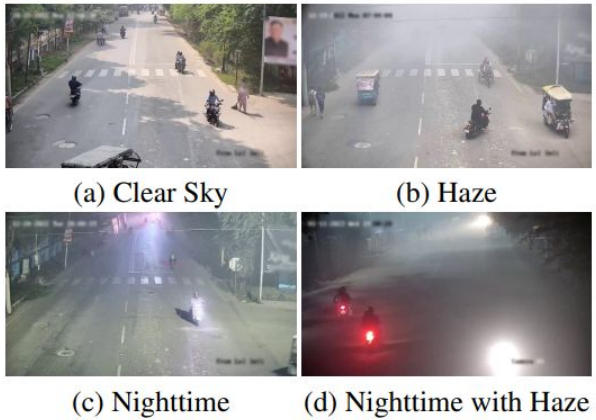


Figure 3. Visualization of various outside environment on one location in dataset.

PIPELINE TOP 2 2023

Cách tiếp cận data

- Lên kế hoạch về data augmentation

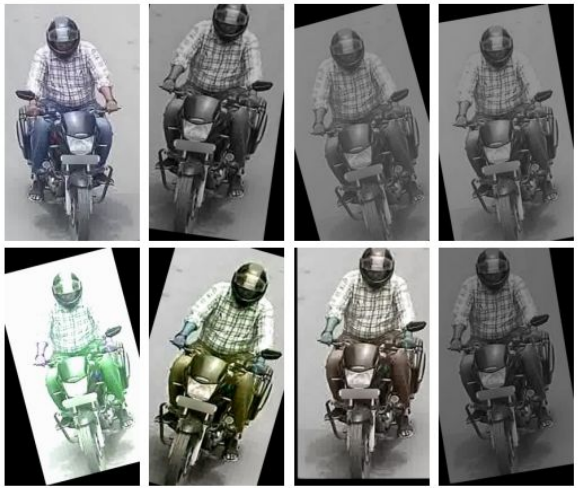


Figure 4. Example of data augmentation in training dataset for Identifier.

PIPELINE TOP 2 2023

Cách tiếp cận data

- Gộp các label để thực hiện cho bước detector

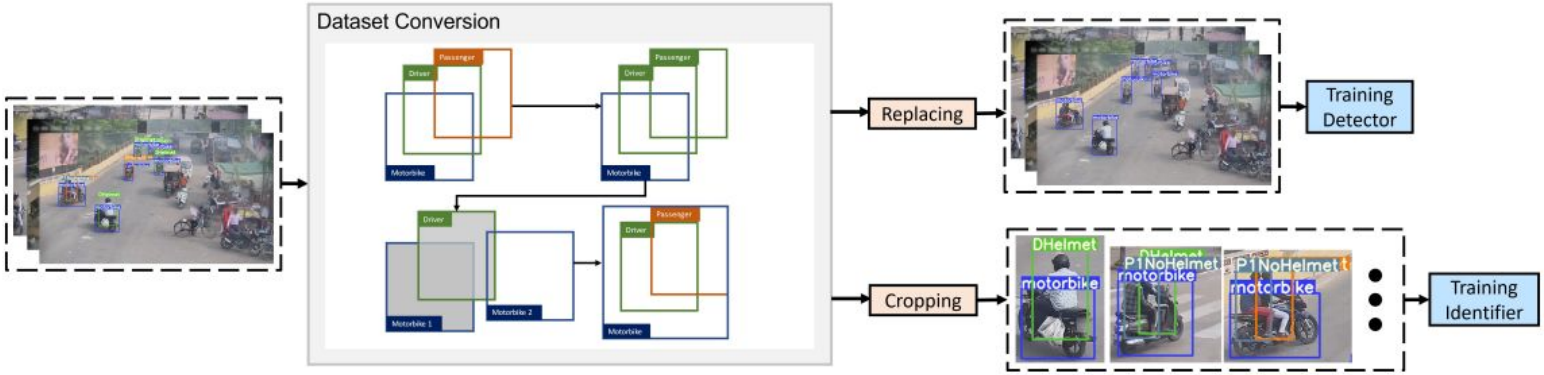


Figure 2. The diagram of data conversion for training both the Detector and Identifier. The input is the groundtruth of the given dataset with 7 classes and 1920x1080 resolution; we convert them into two new datasets, including 1 class with 1920x1080 image size and 7 classes with the cropped image.

PIPELINE TOP 2 2023

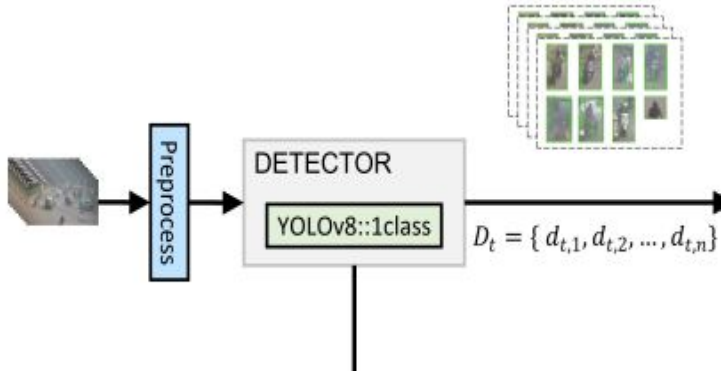
Detector

Họ gộp người lái xe cùng với xe máy thành 1 Class duy nhất motorbike

Sau đó họ cắt bbox trả về chỉ lưu chọn các Bbox có độ lớn >= 40 và scale up bbox lên 1.5% hoặc 50 pixel

Lấy được confidence score thông qua mô hình (the first rank)

$$S_t = \{s_{t,1}^D, s_{t,2}^D, ..., s_{t,n}^D\}$$

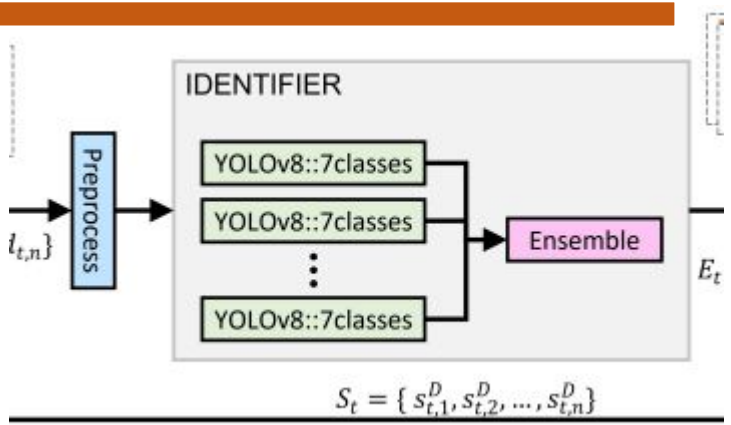


PIPELINE TOP 2 2023

IDENTIFIER

- Dùng 5 model yolov8 với size ảnh khác nhau sau
- Ensemble bằng WBF thì thấy được là 5 model với size 320, 384, 448, 512, 576 có acc cao nhất
- Tính second rank bằng cách merge

$$S_{t,i} = \{s^D_{t,i,1}, s^D_{t,i,2}, \dots, s^D_{t,i,m}\}$$



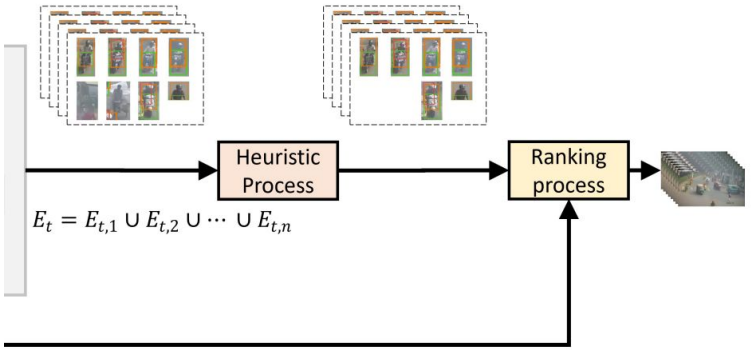
Trainning Size						Inference Size	mAP
256	320	384	448	512	576		
	√	√	√	√		384	0.5861
	√	√	√	√		448	0.5888
	√	√	√	√		512	0.7269
	√	√	√	√	√	512	0.7754
√	√	√	√	√	√	512	0.7718

PIPELINE TOP 2 2023

Heuristic Process và Ranking Process

- Tìm ra một cof max để loại bỏ những label có cof nhỏ hơn
- Lấy cof ở Detector và Identifier để tính cof cho đối tượng

$$r_{t,i,j} = s^{D}_{t,i} \cdot s^{E}_{t,i,j}$$




```
conf_max = 0
for det_crop in dict_temp:
    if det_crop['id'] in [2, 3]:
        conf_max = max(conf_max, det_crop['conf'])
for det_crop in dict_temp:
    if (det_crop['id'] in [2, 3] and det_crop['conf'] == conf_max) or \
            det_crop['id'] not in [2, 3]:
        results.append(det_crop)
```

PIPELINE TOP 2 2023

KẾT QUẢ THỰC NGHIỆM

Model	Image size	mAP
YOLOv8-6e6	1280	0.3814
YOLOv8-6e6	1536	0.3917
YOLOv8-6e6	1920	0.3823

Table 1. Ablation study of image size in the Detector.

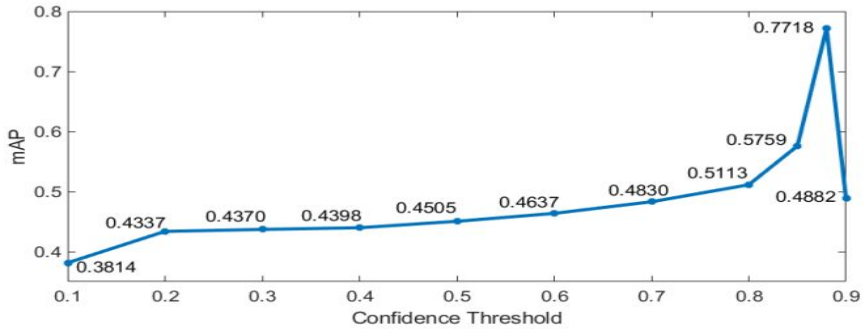


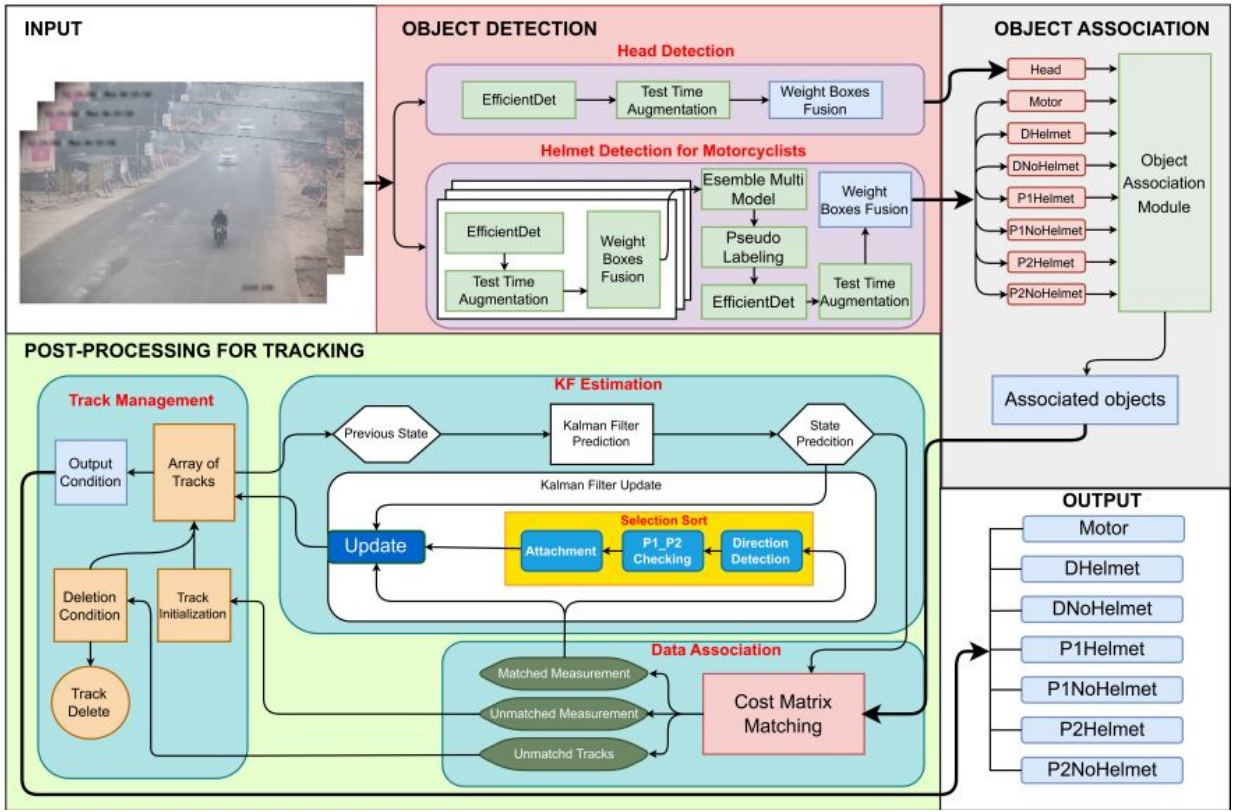
Figure 5. Ablation study of confidence score in Final rank.

Trainning Size						Inference Size	mAP
256	320	384	448	512	576		
	✓	✓	✓	✓		384	0.5861
	✓	✓	✓	✓		448	0.5888
	✓	✓	✓	✓		512	0.7269
	✓	✓	✓	✓	✓	512	0.7754
✓	✓	✓	✓	✓	✓	512	0.7718

Table 2. Ablation study of ensemble in the Identifier.

AI VIETNAM
Seminar

PIPELINE TOP 3 2023



PIPELINE TOP 3 2023

1. Object Detection Module

- Gồm hai module dùng cho head detection và helmets of motorcyclists detection chịu trách nhiệm detec tất cả object trong khung hình
- Dùng one-stage methods : EfficientDet
- Đối với model helmets of motorcyclists detection là model đầu tiên dùng để detec 7 object bao gồm motorbike, driver, and passenger. Nhưng do imbalance giữa các lớp object nên model cho hiệu suất kém ở các lớp nhỏ
- Đối với model head detection là model thứ hai dùng output bổ sung cho model đầu giúp cải thiện hiệu suất tổng thể

PIPELINE TOP 3 2023

1. Object Detection Module

Helmet Detection for Motorcyclists:

- EfficientDet dùng backbone là EfficientNet và feature network là BiFPN
- Baseline chạy thử nghiệm trên 3 biến thể lớn nhất EfficiencyDet (D5, D6 và D7) với các input khác nhau (512 → 1024) để đạt hiệu suất tốt nhất

Head Detection :

- EfficientDet vẫn đạt được hiệu quả
- Module chỉ train head detection model, không kết hợp pseudo-labeling

PIPELINE TOP 3 2023

1. Object Detection Module

Helmet Detection for Motorcyclists:

- EfficientDet dùng backbone là EfficientNet và feature network là BiFPN
- Baseline chạy thử nghiệm trên 3 biến thể lớn nhất EfficiencyDet (D5, D6 và D7) với các input khác nhau (512 → 1024) để đạt hiệu suất tốt nhất

Head Detection :

- EfficientDet vẫn đạt được hiệu quả
- Module chỉ train head detection model, không kết hợp pseudo-labeling

PIPELINE TOP 3 2023

1. Object Detection Module

Data Augmentation:

- Dùng kĩ thuật TTA
- Tổng hợp thông qua Weighted boxes fusion (WBF)

Assembling predicted boxes and pseudolabeling:

- Dùng ensemble và pseudolabeling để cải thiện hiệu suất EfficiencyDet D6(rudeuce variance và 5382 bias)
- Esembling các kết quả từ các model tốt nhất để pseudolabeling, sau đó train model EfficiencyDet trong một vài epoch với pseudolabeling này và chọn model đó làm model cuối cùng. Để tránh loại bỏ rare object sử dụng very low threshold với WBF

2. Object Association Module

- Association output từ object detection để khớp object motor tương ứng với object head và human → ID tracking duy nhất cho mỗi nhóm
- Sau khi gán các object output từ module trước, association sẽ xác định tất các cặp human-motor và human-head và liên kết chúng với nhau. Thực hiện bằng cách tính toán overlap area và vị trí tương ứng của bounding boxes với motor
- Output là danh sách các motor được đính kèm human và head tương ứng

PIPELINE TOP 3 2023

3. Post-processing For Tracking Module

- Thách thức đối với object detection model là phát hiện chính xác số lượng người trên xe (số lượng vượt quá 2) do góc camera khi tiếp xúc xe nên việc phân biệt cá thể trên xe chưa đạt hiệu suất tốt
- Do sự mất cân bằng của tập training dataset (5500 :70) → phân loại sai thành class driver
- Quá trình post-processing sẽ tracking tất cả motor và dùng Selection Sort để gán lại từng human's box trên xe máy trong khi vẫn giữ lại class Helmet hoặc NoHelmet → final output

PIPELINE TOP 3 2023

3. Post-processing For Tracking Module

- Module dựa trên SORT algorithm. Ngoài ra, còn tích hợp Kalman Filter (KF) kết hợp module Selection SORT để cải thiện output detection
- Selection SORT không chỉ cập nhật ID mà còn cập nhật thuộc tính bổ sung cho object trong mỗi frame để xác định lại vị trí của người trên xe

PIPELINE TOP 2 2023

3. Post-processing For Tracking Module

Quy trình cụ thể:

(1) Direction detection:

Figure 2. Illustration of the motorbike detection. The direction of the motorbike is assigned as IN direction (green arrow) or OUT direction (yellow arrow) according to the motorbike's position throughout its appearance in the video.

Algorithm 1: Direction Detection

Input : $centers$: list of center points' coordinator for 1 motorbike
Output: 1 (IN) / 0 (OUT) / $None$: direction of motorbike

```
checks ← [];
for i in [1: len(centers)] do
    checks.append(centers_i - centers_{i-1});
    if len(checks) ≥ 3 then
        break;
end
if len(checks) < 3 then
    return None
else
    num_T ← count_true_in_checks;
    num_F ← count_false_in_checks;
    return num_T > num_F
```

PIPELINE TOP 3 2023

3. Post-processing For Tracking Module

(2) P1_P2_Checking: xác định số lượng người trên xe

Algorithm 2: P1_P2_Checking

Input : *bbox_head*: list of heads attached to 1 motorbike
P_type: *P*1 or *P*2
Output: P_type

```
if P_type = P1 then
    num_heads ← 2;
else
    num_heads ← 3;
for heads in bbox_head do
    if len(heads) = num_heads then
        counter ← counter + 1;
    if counter = 3 then
        return P_type;
end
return None;
```

PIPELINE TOP 3 2023

3. Post-processing For Tracking Module

(3) Đặt lại lớp đúng cho các đối tượng con người dựa trên kết quả từ thuật toán Kiểm tra P1 P2 và thuật toán phát hiện hướng.:

(a) Detection output (b) Final output

Figure 3. Illustration for reassignment process. Figures 3a and 3b depict examples of the OUT direction case.

PIPELINE TOP 2 2023

3. Post-processing For Tracking Module

- Dựa vào output của giai đoạn trước thuật toán gán lại chính xác class human
- Các human boxes được sắp xếp theo tọa độ để có vị trí tương đối so với xe
- Dựa vào thuật toán 3 để gán lại class thích hợp trong suốt thời gian ID xe đó trong video

PIPELINE TOP 3 2023

Algorithm 3: Attachment 1 motorbike on 1 frame

Input : *center*: center point's coordinator
bbox_head: list of head's coordinator
bbox_human: list of human body coordinator and class
Output: Update class in *bbox_human*

```
if d is None then
    update d with algorithms 1
if is_P1 is None then
    update is_P1 with algorithms 2
if is_P2 is None then
    update is_P2 with algorithms 2
if is_P2 ≠ None then
    bbox_human ← sorted(bbox_human);
    if len(bbox_human) = 1 then
        update class according to table 1
    if len(bbox_human) = 2 then
        update class according to table 1
    if len(bbox_human) = 3 then
        update class according to table 1
else if is_P1 ≠ None then
    bbox_human ← sorted(bbox_human);
    if len(bbox_human) = 1 then
        update class according to table 1
    if len(bbox_human) = 2 then
        update class according to table 1
```

Table 1. Class transform on different cases. d is the direction of motorbike, h_i human with original class from the detection module, h'_i human after being assigned class. The class values are defined in table 2

	P1				P2					
	1 human	2 humans			1 human	2 humans		3 humans		
d	$h_1 \to h'_1$	$h_1 \to h'_1$	$h_2 \to h'_2$		$h_1 \to h'_1$	$h_1 \to h'_1$	$h_2 \to h'_2$	$h_1 \to h'_1$	$h_2 \to h'_2$	$h_3 \to h'_3$
0	2,6 4 3,7 5	4,6 2 5,7 3	2,6 4 3,7 5		2,4 6 3,5 7	- - - -	2,4 6 3,5 7	2,4 6 3,5 7	2,6 4 3,7 5	4,6 2 5,7 3
1	4,6 2 5,7 3	2,6 4 3,7 5	4,6 2 5,7 3		- - - -	2,4 6 3,5 7	- - - -	4,6 2 5,7 3	2,6 4 3,7 5	2,4 6 3,5 7

PIPELINE TOP 3 2023

KẾT QUẢ THỰC NGHIỆM

Table 3. Comparison of the performance of different ensemble methods on the training set of fold 1, following the experimental setup described in Section 4.2.

Method	WBF	NMS
mAP	44.46	40.72

Table 4. Ablation study on impact of applied methods: Ensemble(Ens), Pseudo Labeling (Ps), and our Post-processing for Tracking (PPT) respectively. The first row is the baseline results from EfficientDet-D6 with image size of 768.

Ens	Ps	PPT	mAP
			44.09(baseline)
✓			47.53 (+3.44)
✓		✓	67.85 (+23.76)
✓	✓		53.59 (+9.5)
✓	✓	✓	**69.97 (+25.88)**

Table 5. Leaderboard of Track 5 in the AI City Challenge 2023.

Team ID	mAP
58	83.4
33	77.54
37 (Ours)	**69.97**
18	64.22
16	63.89

Outline

- EDA Data Track 5
- Top model 2023
- Pipeline Basic

PIPELINE cơ bản