Human & Bike Detection

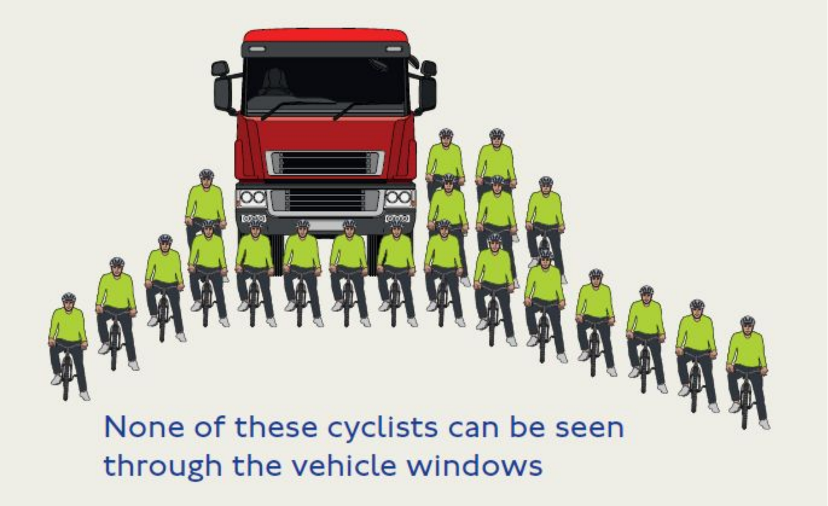
Contents:

- Problem Definition
- Live Demonstration
- Results
- Our Pipeline
- Lessons Learned

Group 6:
Ningxin (Cynthia) Li,
Huanzhang Zhao,
Kaiyi Zhang,
Yuliang Song

1. Problem definition

SAFETY FIRST
PLEASE
WALK YOUR BIKE
FROM HERE
THANK YOU!
SmartSign.com • 800-952-1457 • K-0704


None of these cyclists can be seen through the vehicle windows

HTA 140 (6)/144 (29): Cyclists are not allowed to ride in a crosswalk. The law says you must dismount and walk your bike when utilizing this space where pedestrians may be present.

Contribution

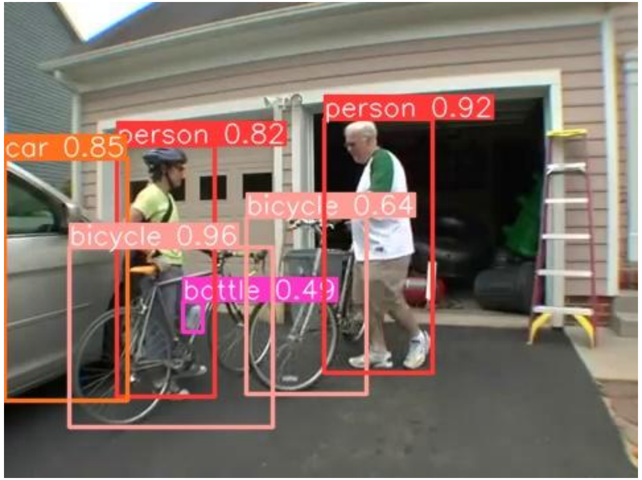
1. Finetune yolo model on identifying a new class “cyclist”
2. Incorporate extracted stem information into our CNN model
3. Compared multiple pretrained model with/out stem information.
4. Introduced our finepicked dataset for cyclist classification
5. Ride: yolo with new class - unride: yolo with optimization

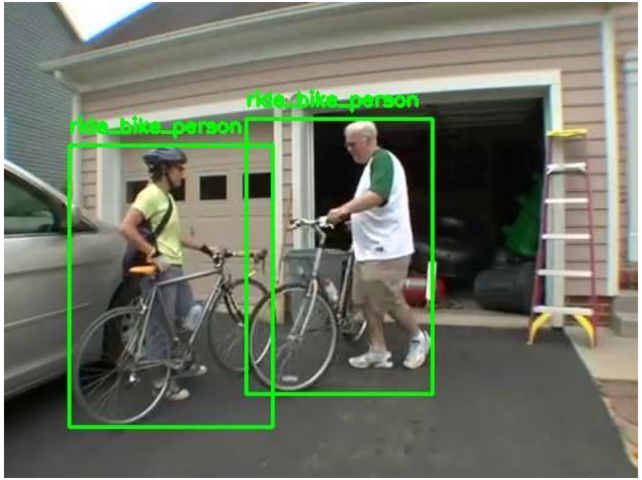
Data description

Task: detection model fine-tuning
We used the CyclistDetectionDataset (CDD)* to finetune our yolo model. This datasets contains 13k train, 1k test, and 500 val images with bounding box information of cyclists.

Train: classification model training
We gathered around 120 videos of cyclists. For each video, we extracted 3 frames per second with a constant total number of images per video, resulting in around 100k cropped cyclist images . Then, we applied our object detection model to identify potential cyclists to crop. From this cropped set, we manually selected arounds 8000 images across two classes: 'ride' and 'unride' evenly.

Classifier training -Data preparation



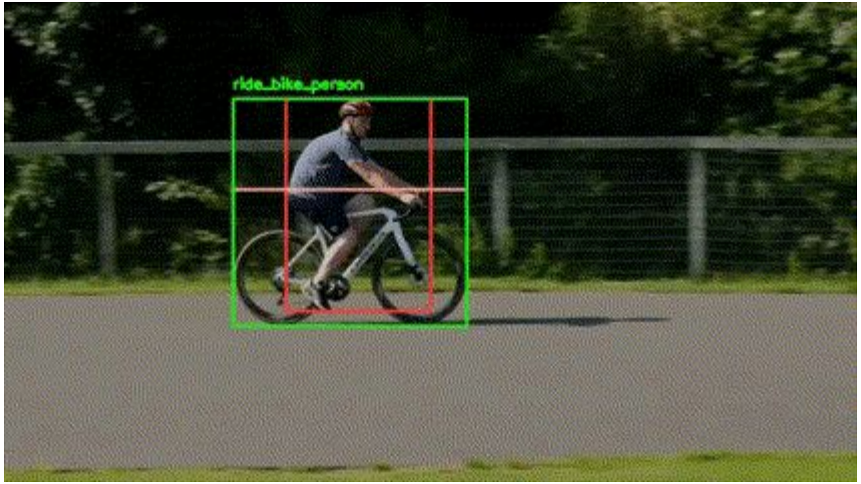


Goal: Min center distance of cyclist and bike while max the overlapping area of them

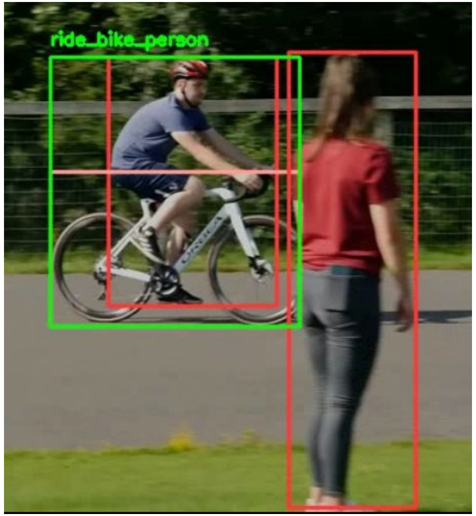
$$\underset{\substack{p\in P, b\in B\\ |b|=|B|}}{\operatorname{argmin}} D(p_i - b_i) - A(p_i, b_i)$$

D is a function measures center distance
A is a function measure overlapping area

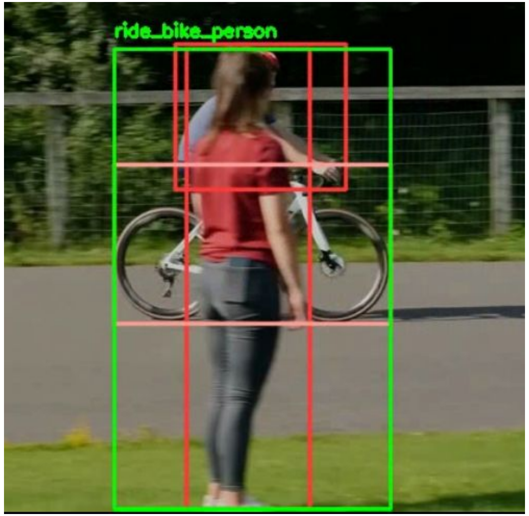
Train classifier -Cyclist detection demo



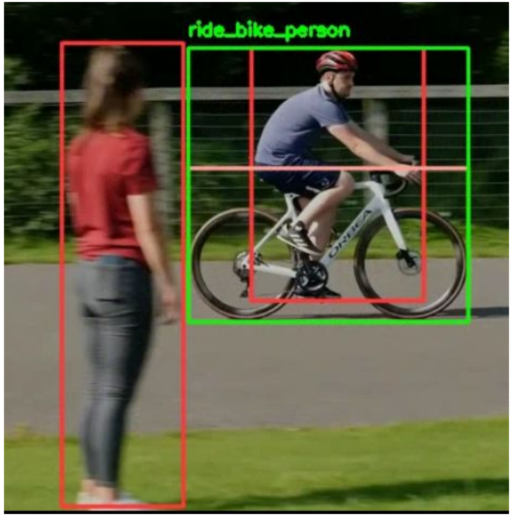
Train classifier -Trade-offs and Triumphs



frame t-1



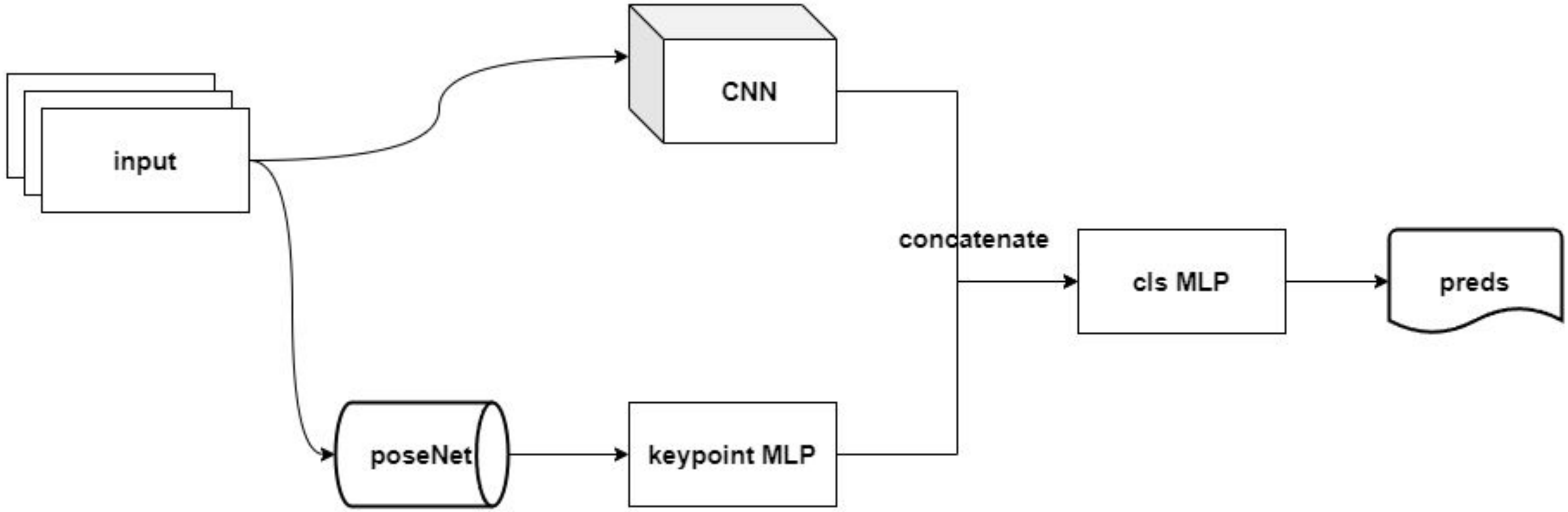
frame t



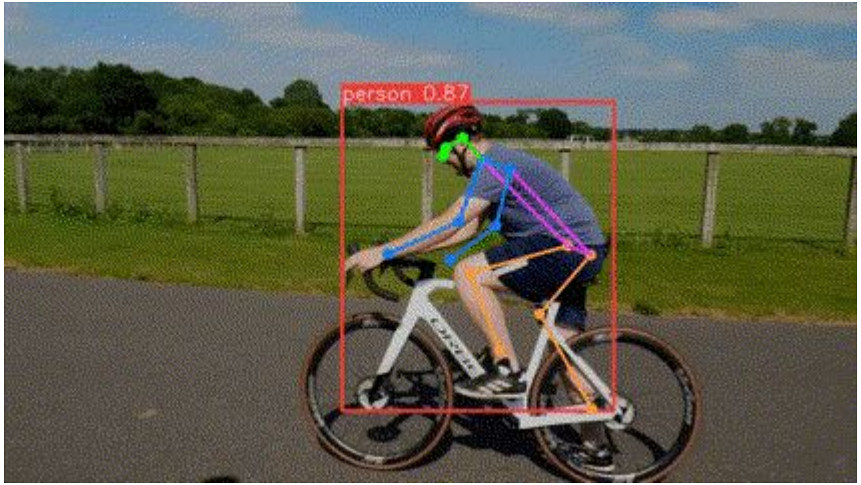
frame t+1

No depth information! But this can be very useful in picking unride class

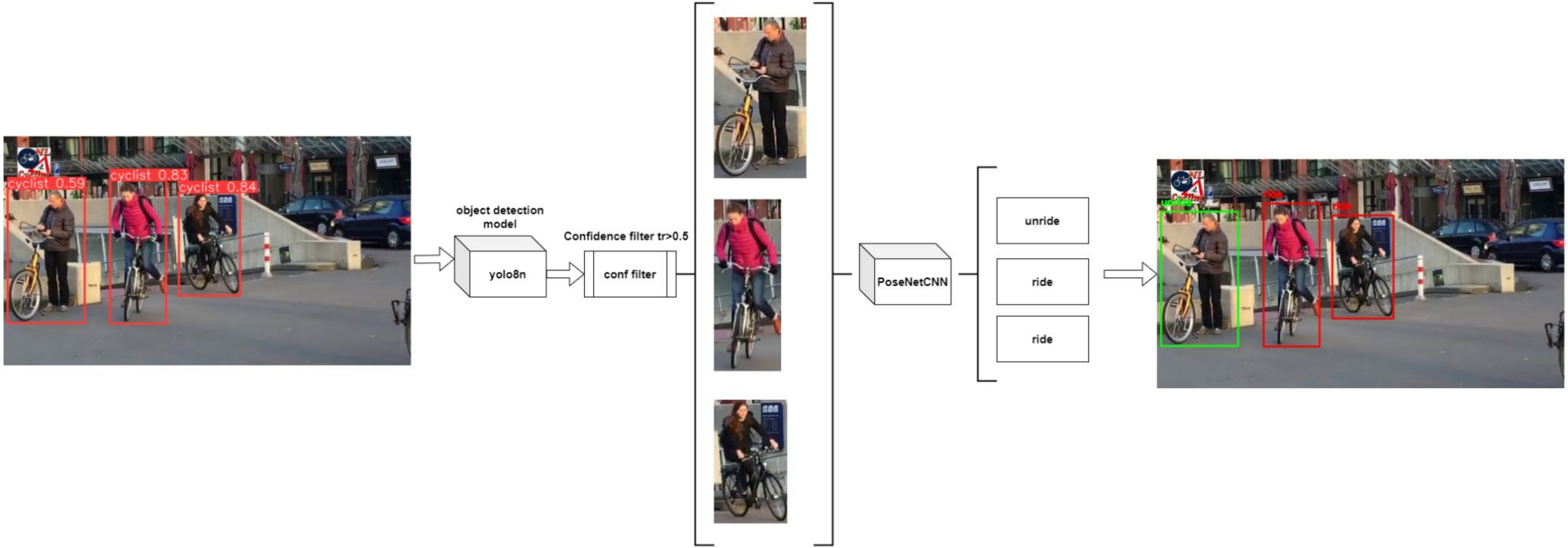
Classifier model structure



Classifier -stem information



System Overview



object detection model

yolo8n

Confidence filter tr>0.5

conf filter

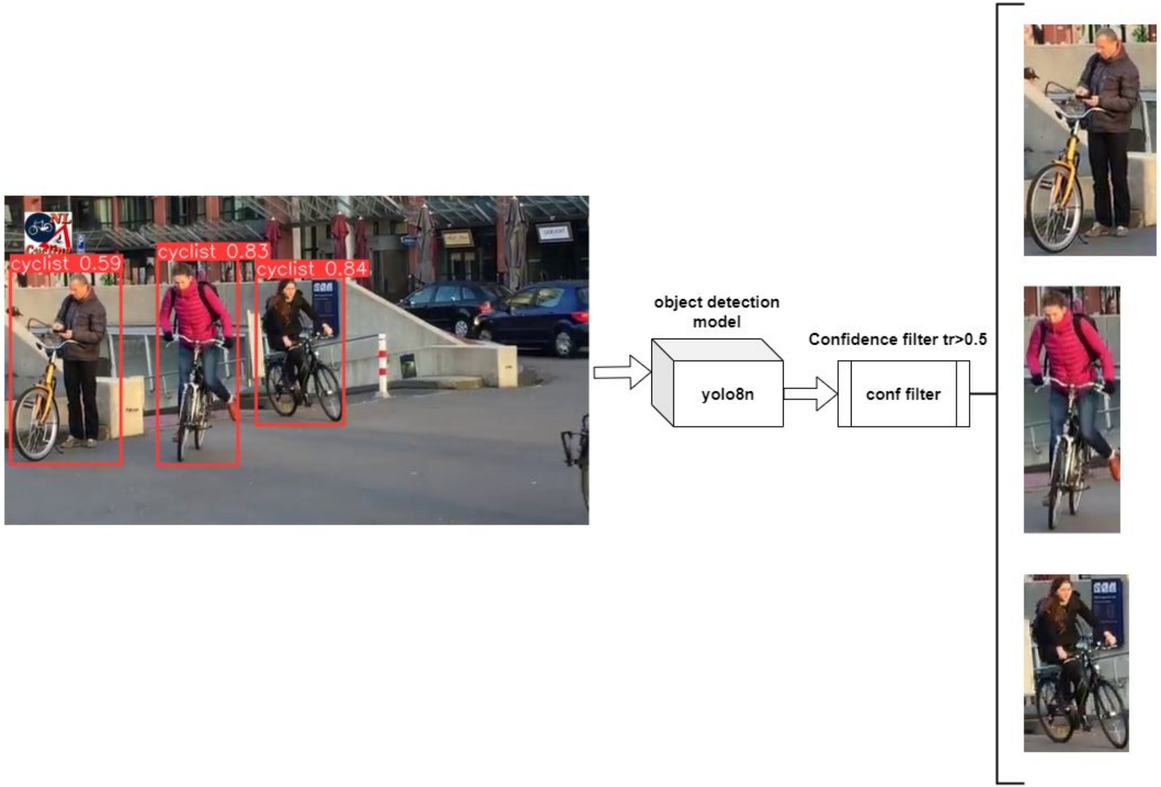
PoseNetCNN

unride

ride

ride

System part1 -cyclist detection



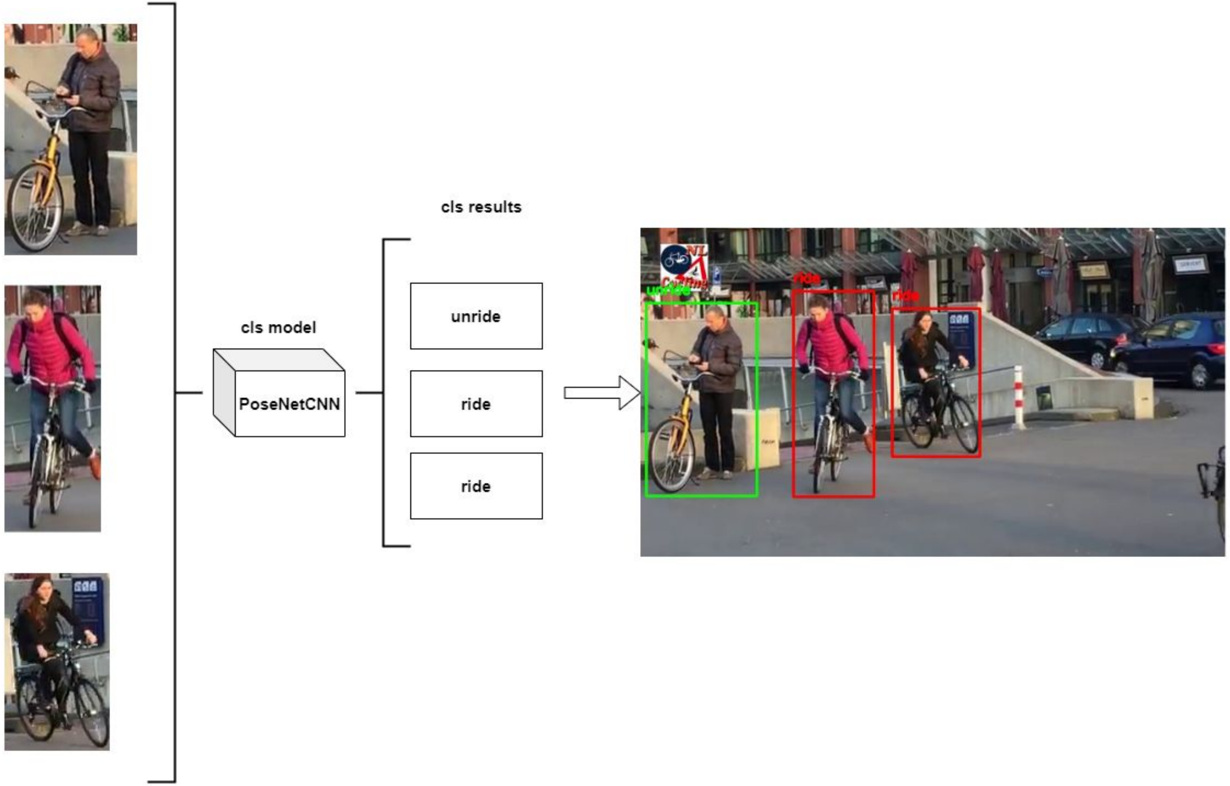
object detection model

yolo8n

Confidence filter tr>0.5

conf filter

System part2 -cyclist classification



cls model

PoseNetCNN

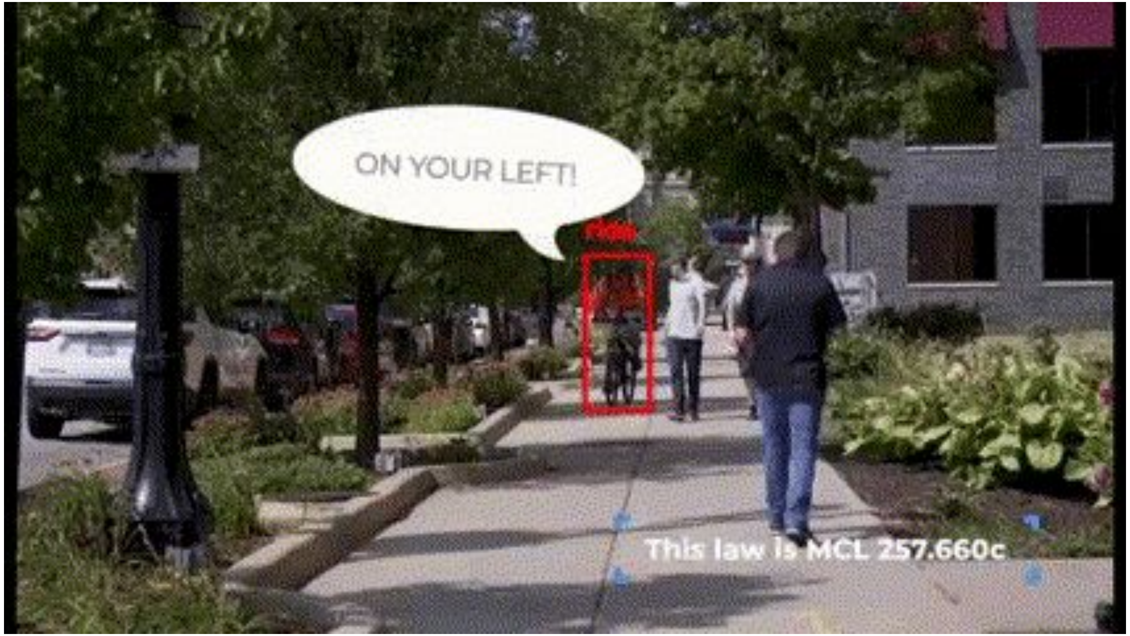
cls results

unride

ride

ride

Demo1



2. Demonstration (>60 seconds)

Find a test video

Identify bike and person: video to image

Model Prediction: ride/unride

……

3. Results - Measurement

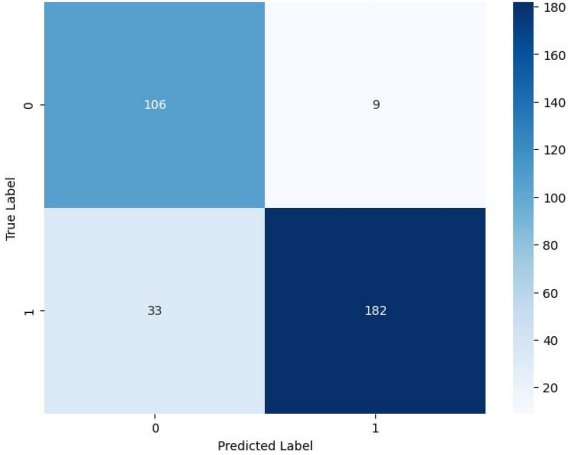
Accuracy = # of correct predictions / # of total images

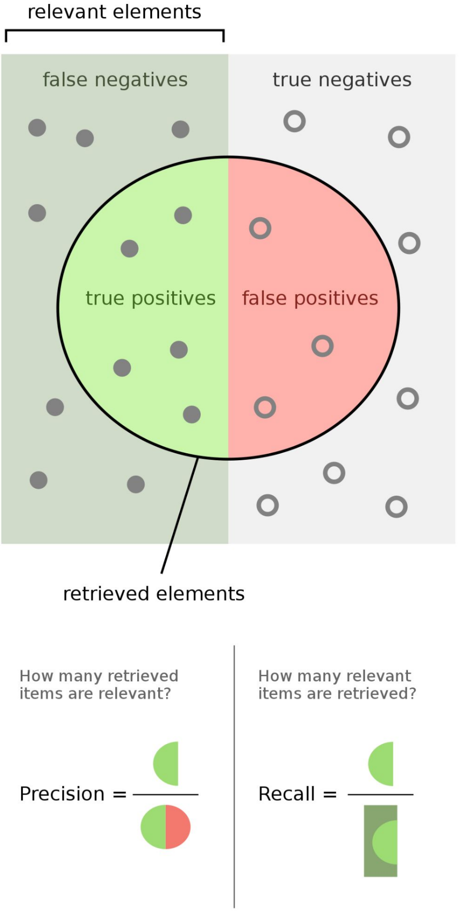
F1 score: $F_1 = \frac{2}{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}} = 2\frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} = \frac{2\mathrm{tp}}{2\mathrm{tp} + \mathrm{fp} + \mathrm{fn}}$

Confusion Matrix:

0 (negative): Ride

1 (positive): Unride


Confusion Matrix for Best Val Acc Epoch: 6



3. Results - Comparing Models

Without YOLO Pose information:

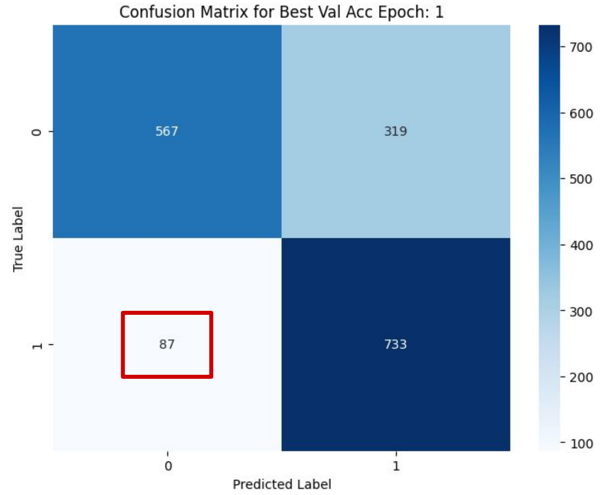
	Simple CNN	SqueezeNet	ResNet18
Best Val Accuracy	0.757	0.687	0.762
Best Val F1 Score	0.779	0.738	0.789

With YOLO Pose information:

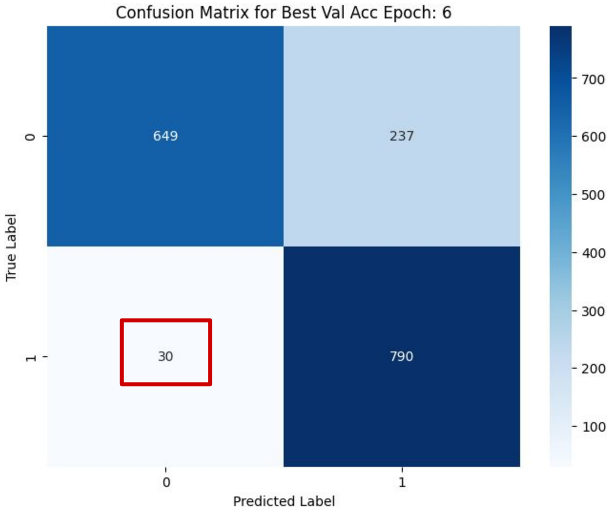
	PoseNet+Simple CNN	PoseNet+SqueezeNet	PoseNet+ResNet18
Best Val Accuracy	0.842	0.548	0.843
Best Val F1 Score	0.842	0.129	0.855

3. Results - Pose Information

ResNet



ResNet + PoseNet



Misclassification

Examples

in ResNet:

Label: 1
Pred: 0

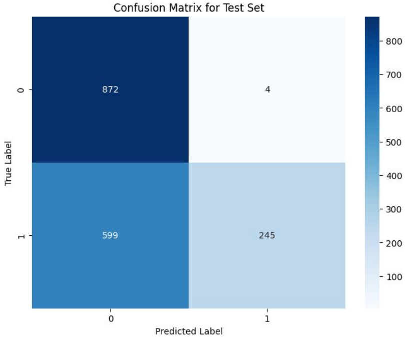
Label: 1
Pred: 0

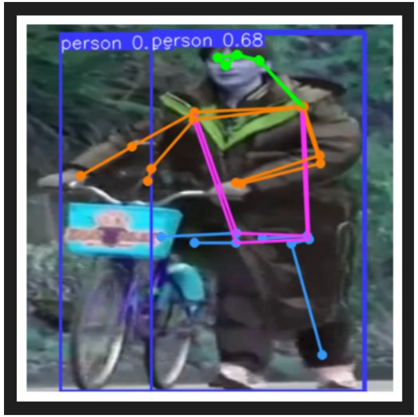
Label: 1
Pred: 0

Label: 0
Pred: 1

Label: 0
Pred: 1

3. Results - Testing & Troubleshooting (Discussion)





Test accuracy: around 65%

Identified Problem (<u>still working to fix</u>):

1. Model **cannot detect the stem**

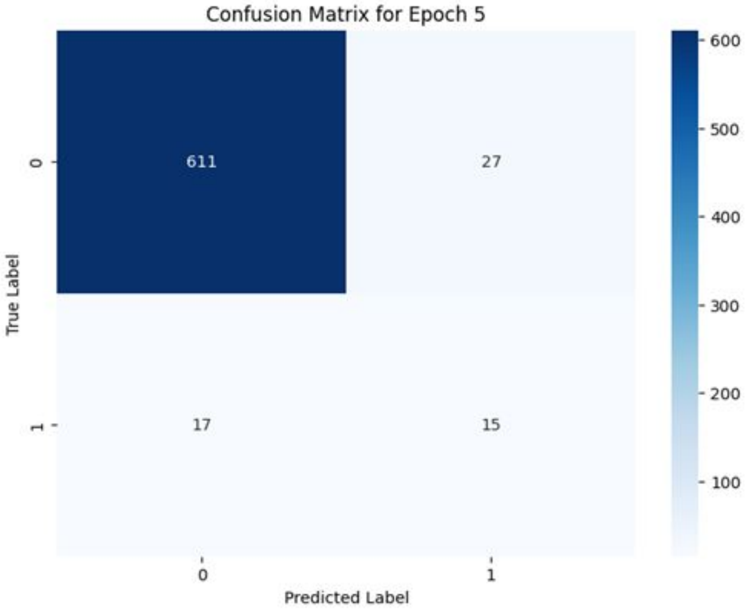
   information due to the **transformer conflict**

   Current data: could detect the pose key-points information on only **16% (277/1720)** of test images

2. If more than one person in image, will generate **multiple stem information**

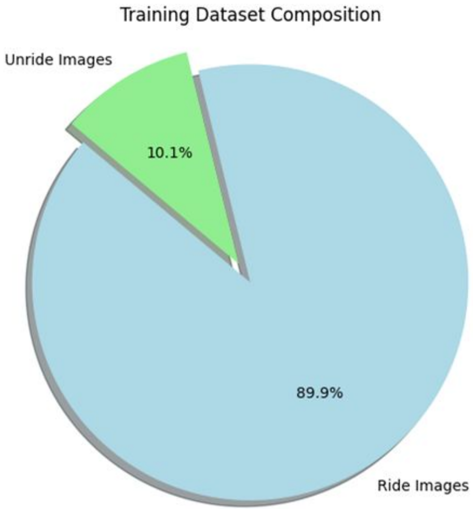
   (sometimes multiple stem will be generated even only one person in image

4. Discussion - Confusion Matrix



Val Accuracy: **94%** Accuracy on the positive Class: **96%**

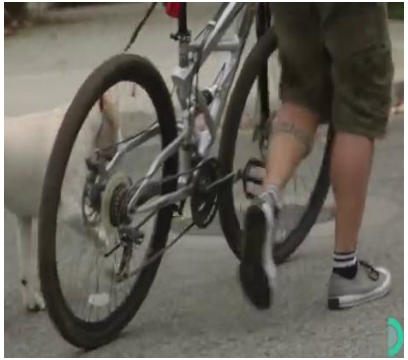
Accuracy on the negative Class: **47%**



Imbalanced Dataset

4. Discussion - Tracking Misclassified Samples

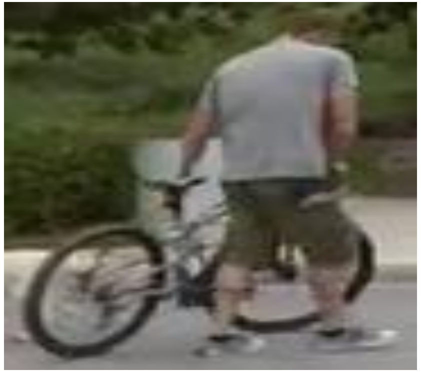
Misclassified Validation Samples



✗

YOLOv8

Improved Image Quality



✓

Thank you!