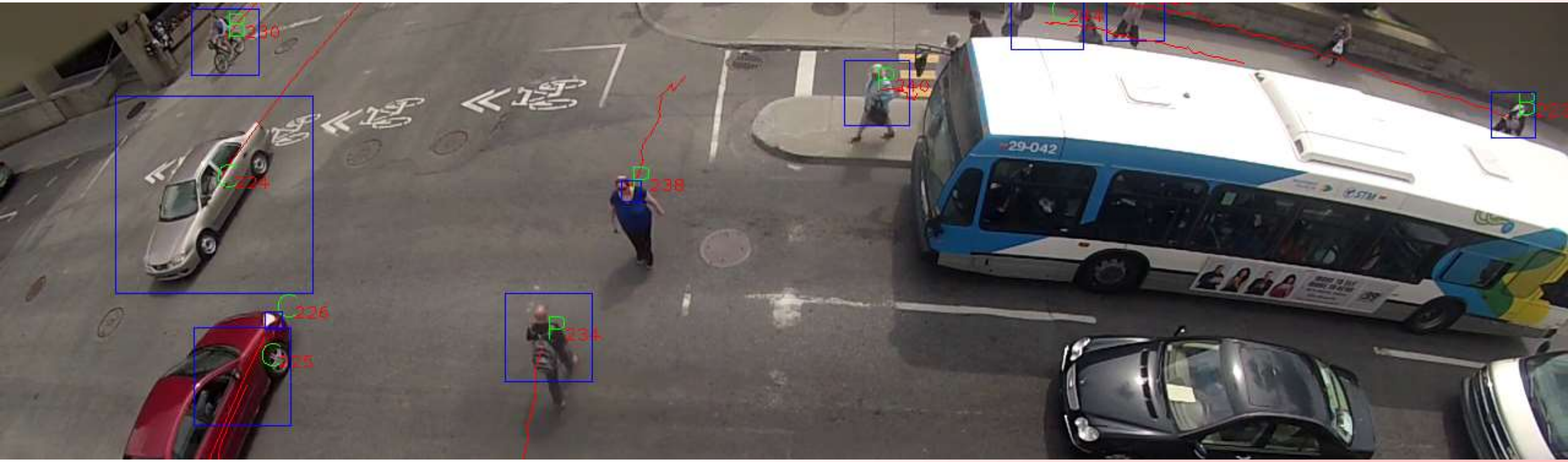
Automated Classification Based on Video Data at Intersections with Heavy Pedestrian and Bicycle Traffic



Sohail Zangenehpour

Luis Miranda-Moreno **Nicolas Saunier**

Department of Civil Engineering and Applied Mechanics
McGill University

ITS Canada, Annual Conference and General Meeting
June 1-4, 2014

Introduction

Introduction

- Two main approaches for studying road safety:
  - Traditional crash and injury data
  - Surrogate measures

- Problems with crash data:
  - Small sample size in short time
  - Lack of detail on the cause of accidents
  - Significant number of crashes need to be recorded before an action can be taken

- Detecting and treating the safety deficiencies before they cause accidents → using Surrogate Measurements

- Examples of surrogate measures:
  - Time To Collision (TTC)
  - Post Encroachment Time (PET)

Introduction

- Shortcoming in availability and quality of data for non-motorized modes
- Few automated methods for collecting microscopic data separately for different road users
- Low accuracy of classification for pedestrians and cyclists
- Problems with classifying pedestrians and cyclists:
  - Non-rigidity
  - Varied appearance
  - Less organized movements
  - Moving in groups close each other
- The main objective of this work: Design an automated method to track and classify objects in video

Object Classification

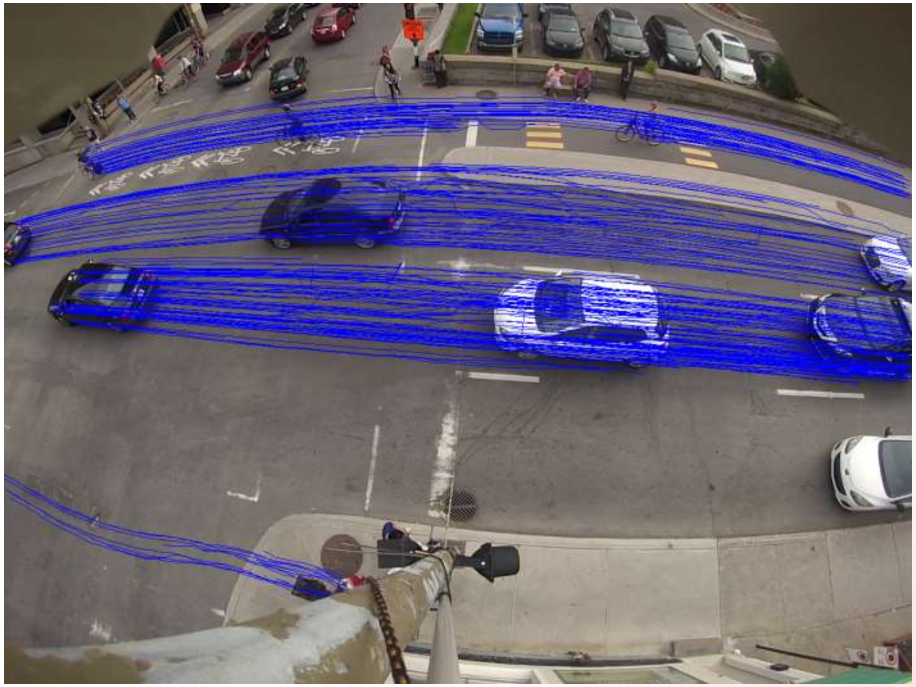
Object Classification Training Step

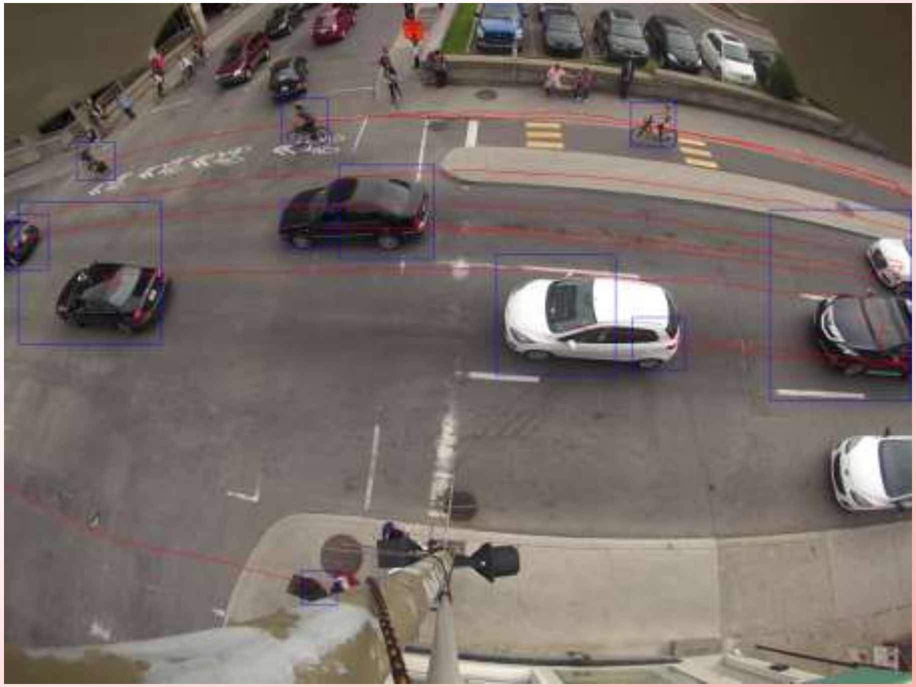
Video → Traffic Intelligence (Tracker) → Trajectory & Image Boxes of all Moving Objects → Manual Classification → Dataset of Training Images → Resizing & HOG → HOG Feature Descriptors → Train the SVM Model

Ordinary video camera

Object Classification Training Step

Video → Traffic Intelligence (Tracker) → Trajectory & Image Boxes of all Moving Objects → Manual Classification → Dataset of Training Images → Resizing & HOG → HOG Feature Descriptors → Train the SVM Model





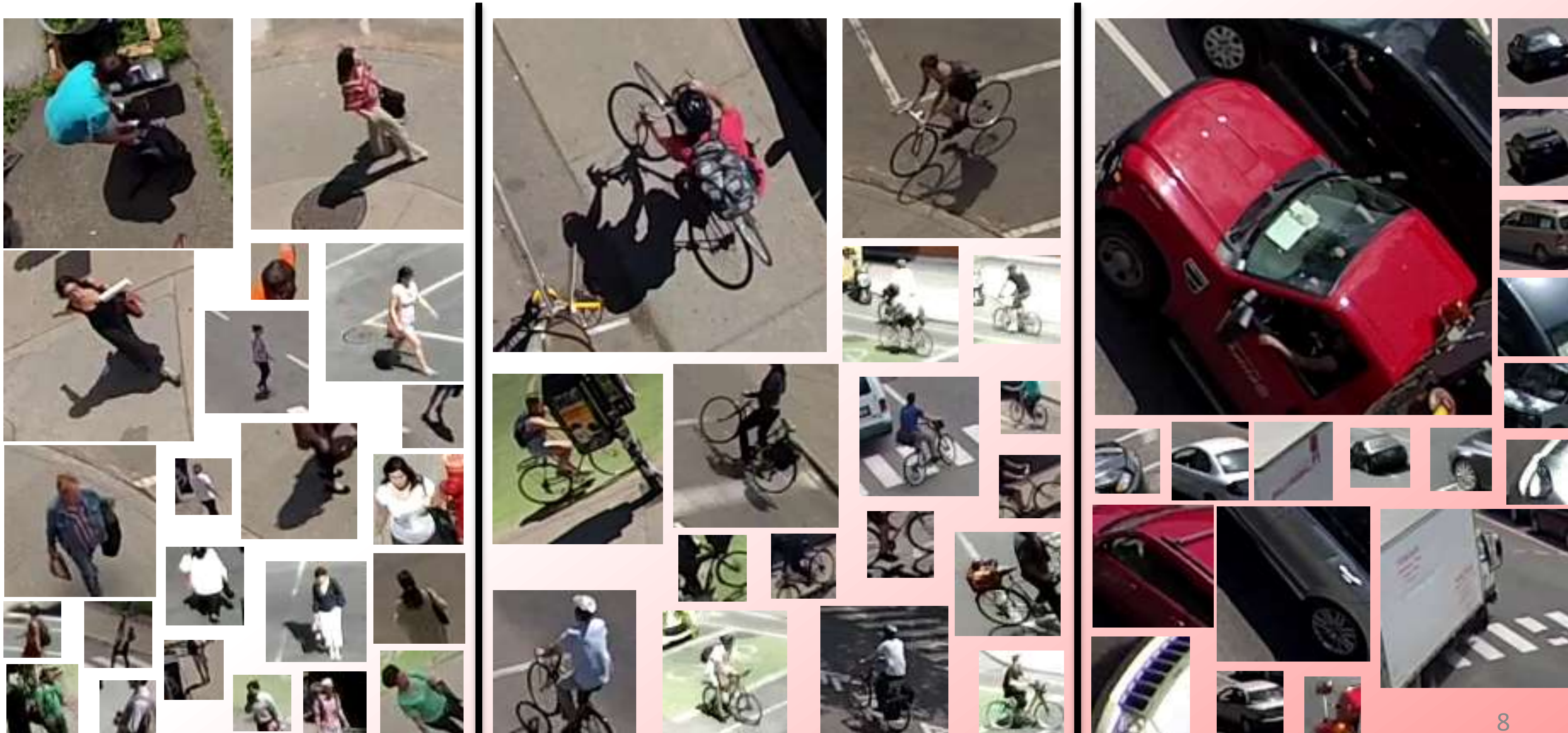
1- Individual pixels (features) are detected and tracked frame to frame

2- Features are grouped based on consistent common motion to make moving objects

Object Classification Training Step

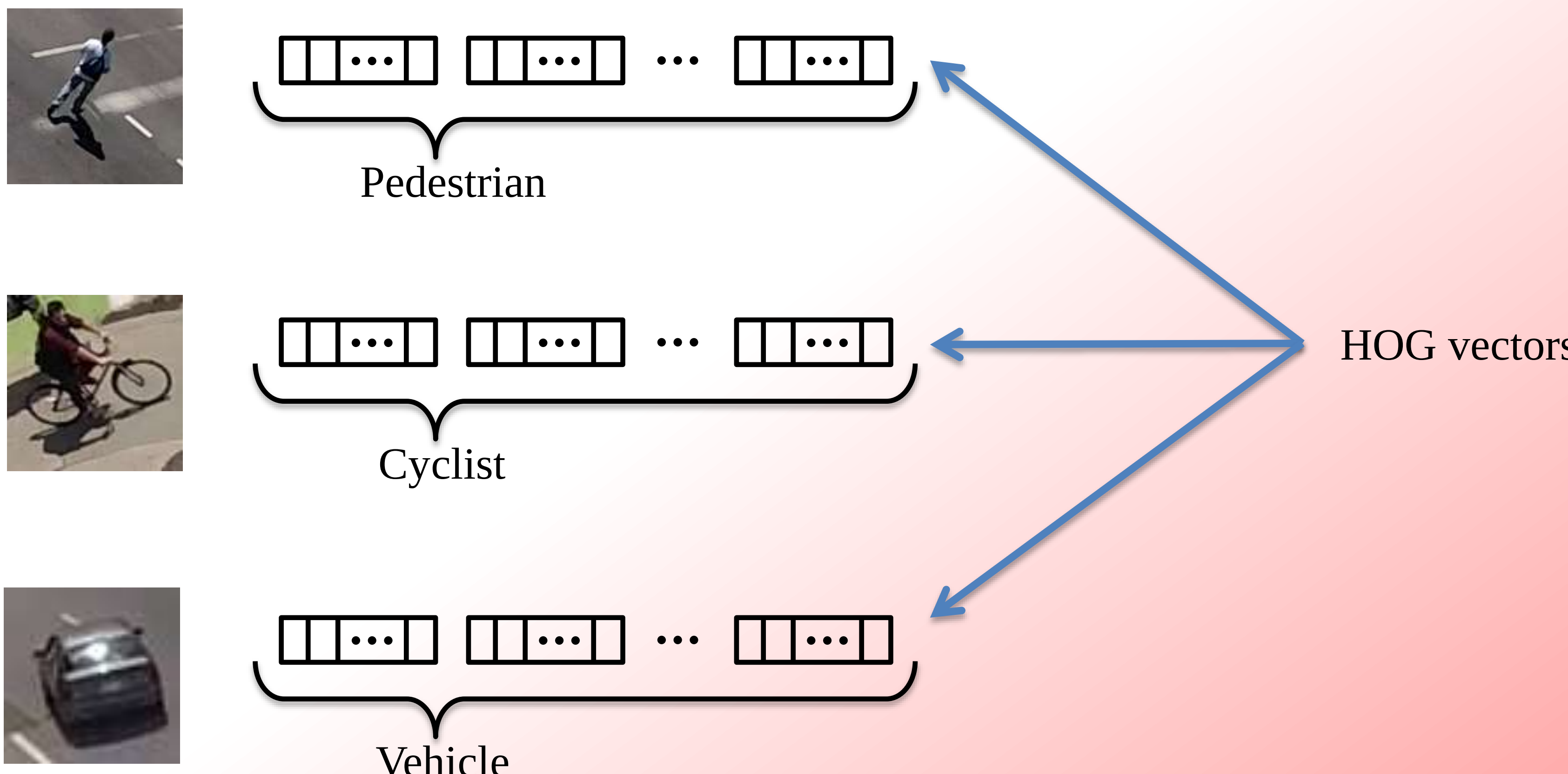
Video → Traffic Intelligence (Tracker) → Trajectory & Image Boxes of all Moving Objects → Manual Classification → Dataset of Training Images → Resizing & HOG → HOG Feature Descriptors → Train the SVM Model

1500 manually classified sample images (training set) for each class:



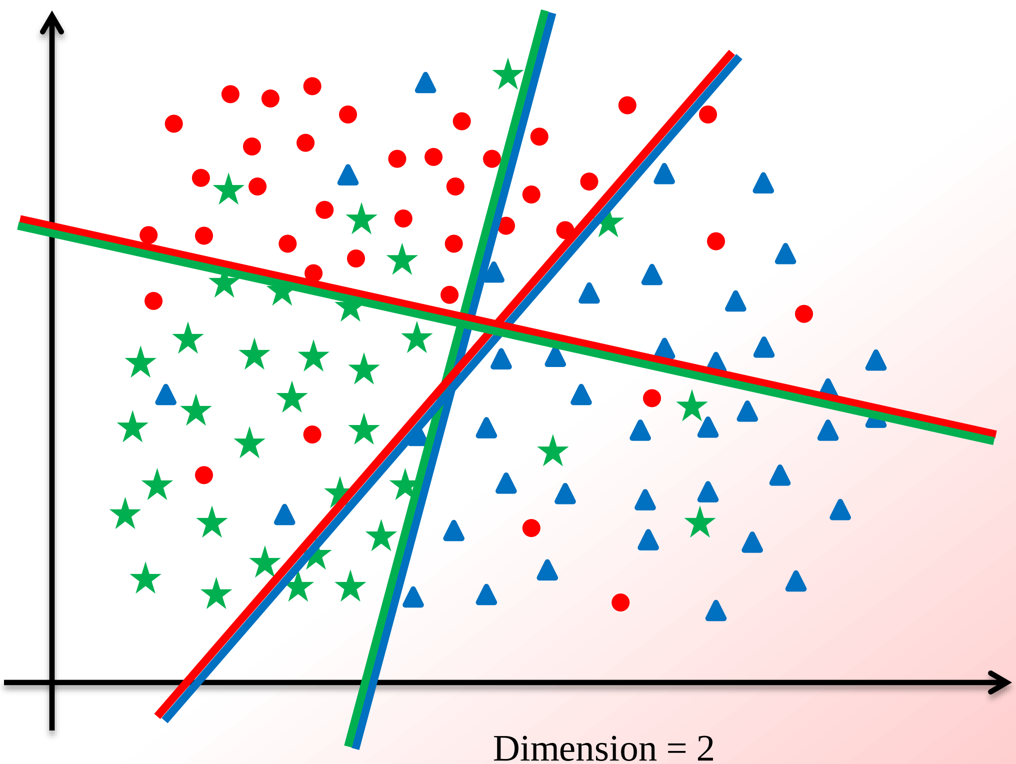
Object Classification Training Step

Video → Traffic Intelligence (Tracker) → Trajectory & Image Boxes of all Moving Objects → Manual Classification → Dataset of Training Images → Resizing & HOG → HOG Feature Descriptors → Train the SVM Model

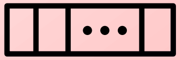


Object Classification Training Step

Video → Traffic Intelligence (Tracker) → Trajectory & Image Boxes of all Moving Objects → Manual Classification → Dataset of Training Images → Resizing & HOG → HOG Feature Descriptors → Train the SVM Model

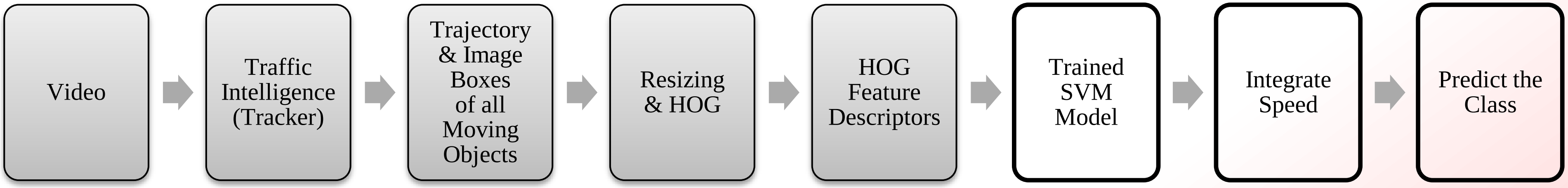


Dimension = 2



Dimension = 1764

Object Classification Prediction

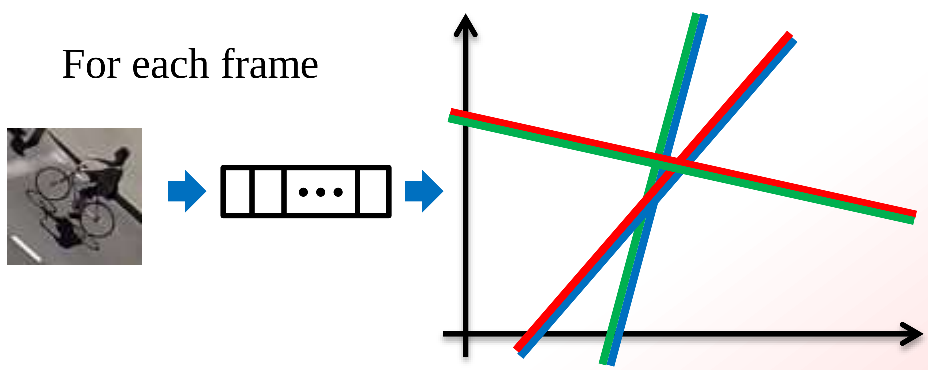


The same as training step

Object Classification Prediction

Video → Traffic Intelligence (Tracker) → Trajectory & Image Boxes of all Moving Objects → Resizing & HOG → HOG Feature Descriptors → Trained SVM Model → Integrate Speed → Predict the Class

For each frame



Most probable class:
1- Pedestrian
2- Cyclist
3- Vehicle

One prediction per frame

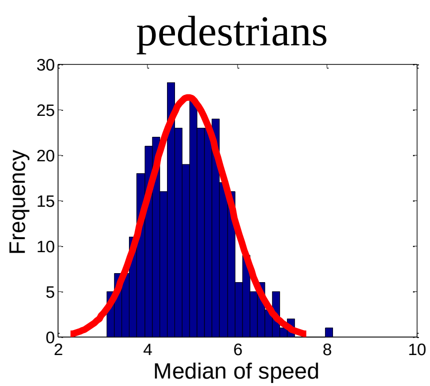
$$\text{P(pedestrian | appearance)} = \frac{\#\ of\ frame\ as\ pedestrains}{\#\ of\ frames}$$

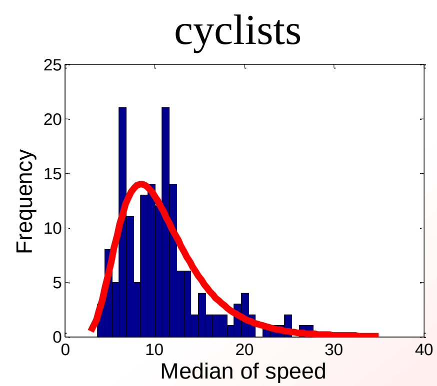
$$\text{P(cyclist | appearance)} = \frac{\#\ of\ frame\ as\ cyclist}{\#\ of\ frames}$$

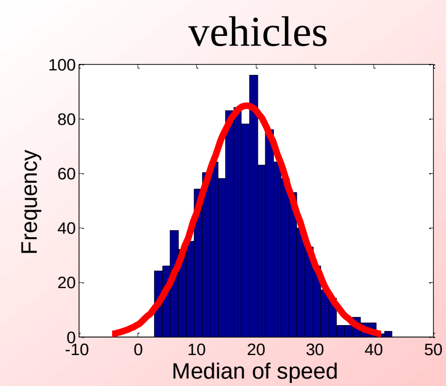
$$\text{P(vehicle | appearance)} = \frac{\#\ of\ frame\ as\ vehicle}{\#\ of\ frames}$$

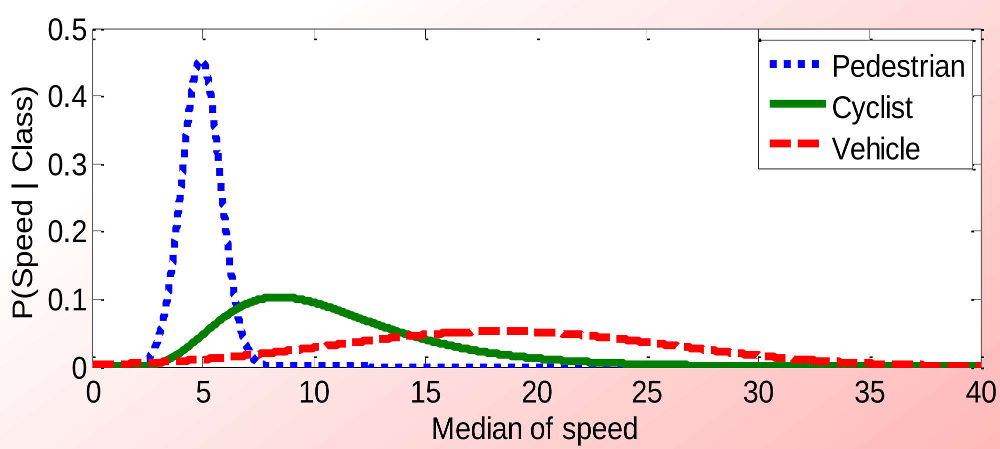
Object Classification Prediction

Video → Traffic Intelligence (Tracker) → Trajectory & Image Boxes of all Moving Objects → Resizing & HOG → HOG Feature Descriptors → Trained SVM Model → Integrate Speed → Predict the Class





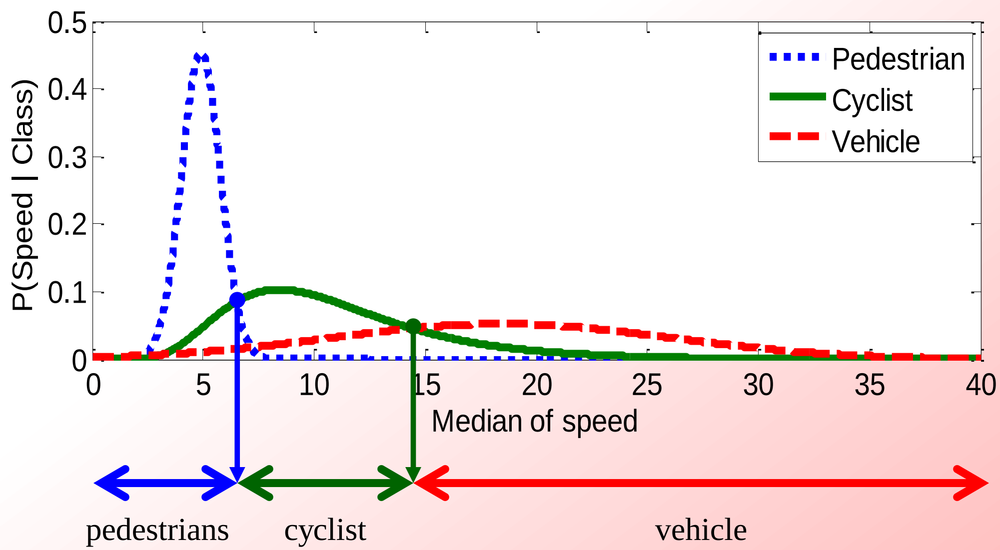




Object Classification Prediction

Video → Traffic Intelligence (Tracker) → Trajectory & Image Boxes of all Moving Objects → Resizing & HOG → HOG Feature Descriptors → Trained SVM Model → Integrate Speed → Predict the Class

- Four methods for integrating speed:
  1. Without using appearance, classification just based on speed, two speed thresholds



Object Classification Prediction

Video → Traffic Intelligence (Tracker) → Trajectory & Image Boxes of all Moving Objects → Resizing & HOG → HOG Feature Descriptors → Trained SVM Model → Integrate Speed → Predict the Class

- Four methods for integrating speed:
  2. Without using speed, just based on appearance:

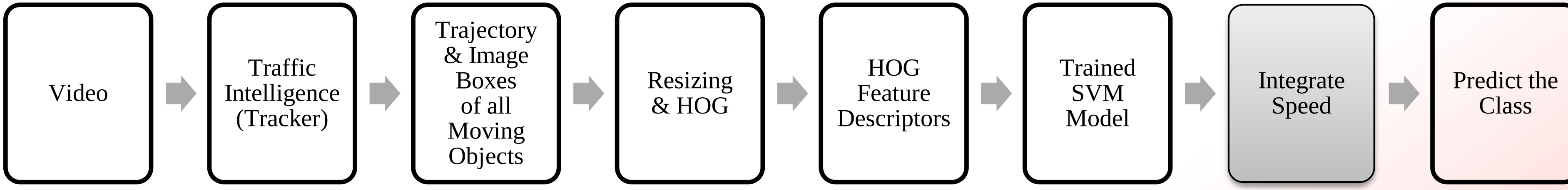
Predicted class is the class with maximum P(class | appearance)

$$\mathrm{P(pedestrian \mid appearance)} = \frac{\#\ of\ frame\ as\ pedestrains}{\#\ of\ frames}$$

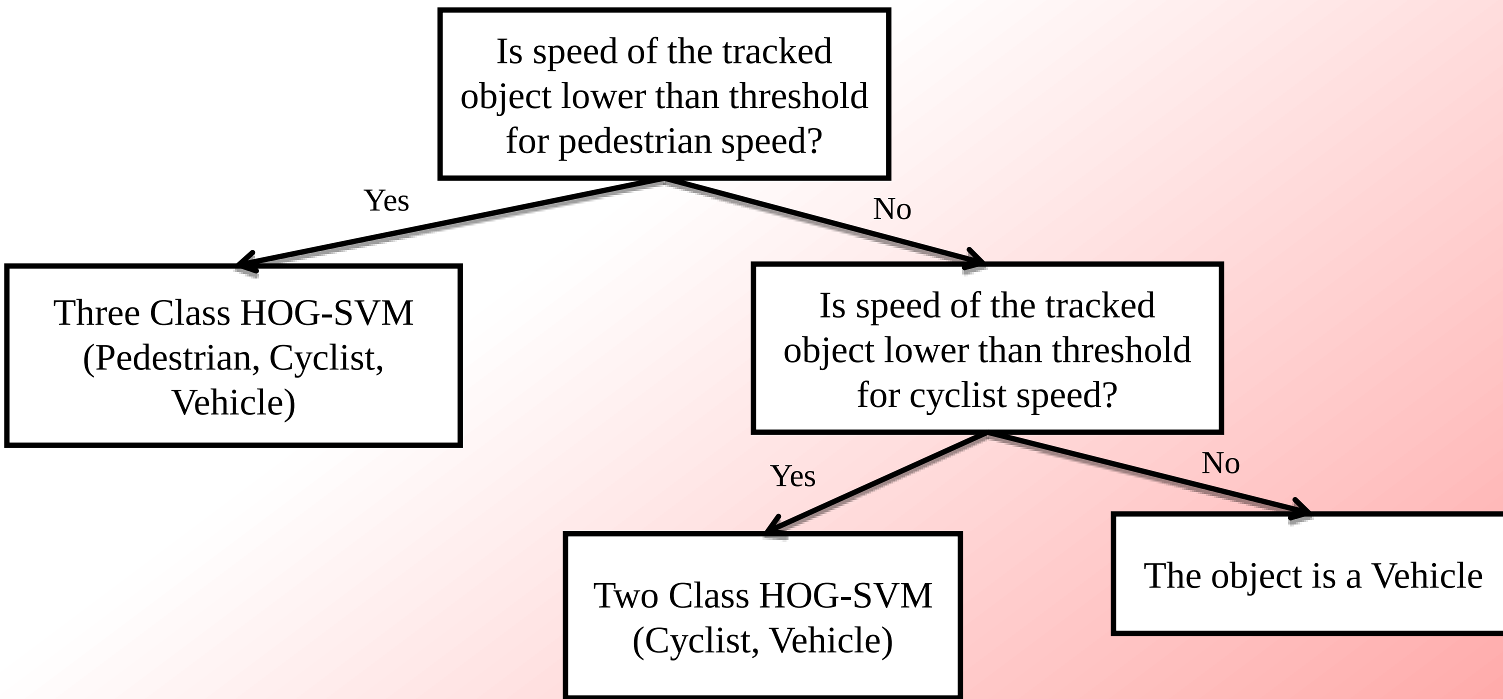
$$\mathrm{P(cyclist \mid appearance)} = \frac{\#\ of\ frame\ as\ cyclist}{\#\ of\ frames}$$

$$\mathrm{P(vehicle \mid appearance)} = \frac{\#\ of\ frame\ as\ vehicle}{\#\ of\ frames}$$

Object Classification Prediction



- Four methods for integrating speed:
  3. Using speed thresholds for switching between different SVM models



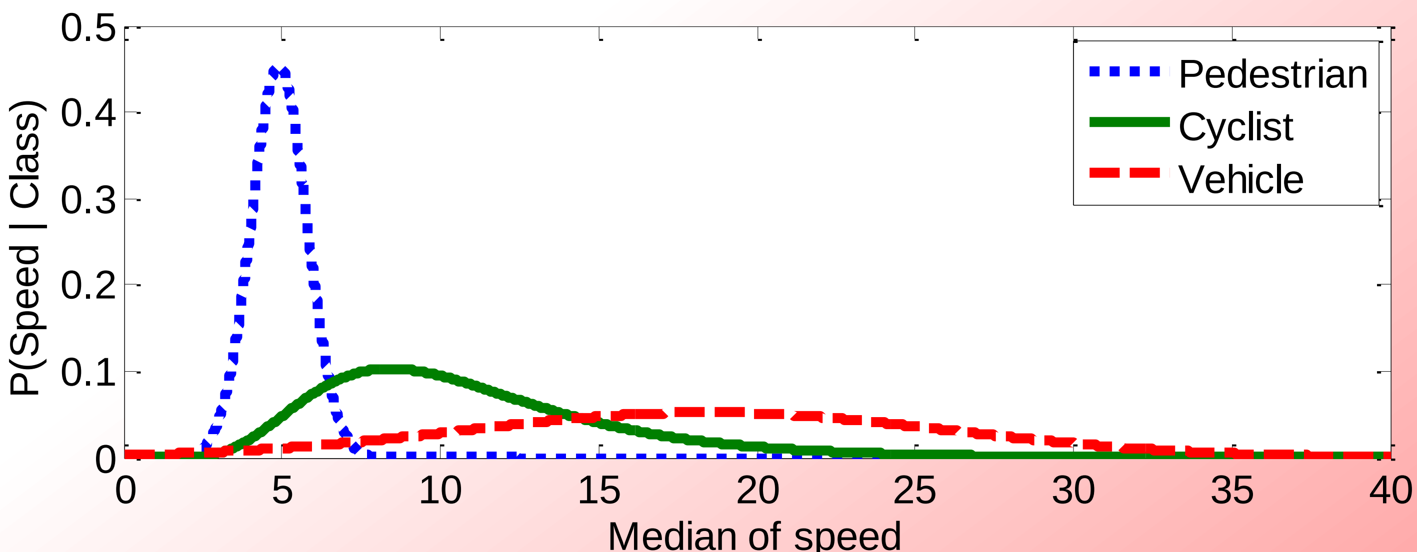
Object Classification Prediction

Video → Traffic Intelligence (Tracker) → Trajectory & Image Boxes of all Moving Objects → Resizing & HOG → HOG Feature Descriptors → Trained SVM Model → Integrate Speed → Predict the Class

- Four methods for integrating speed:
    4. Combining the probability taken from appearance to the probability taken from speed:

$$P(Class \mid Speed, Appearance) \propto P(Class|Appearance)\ P\ (Speed|Class)$$

Predicted class is the class with highest P(Class | Speed, Appearance)



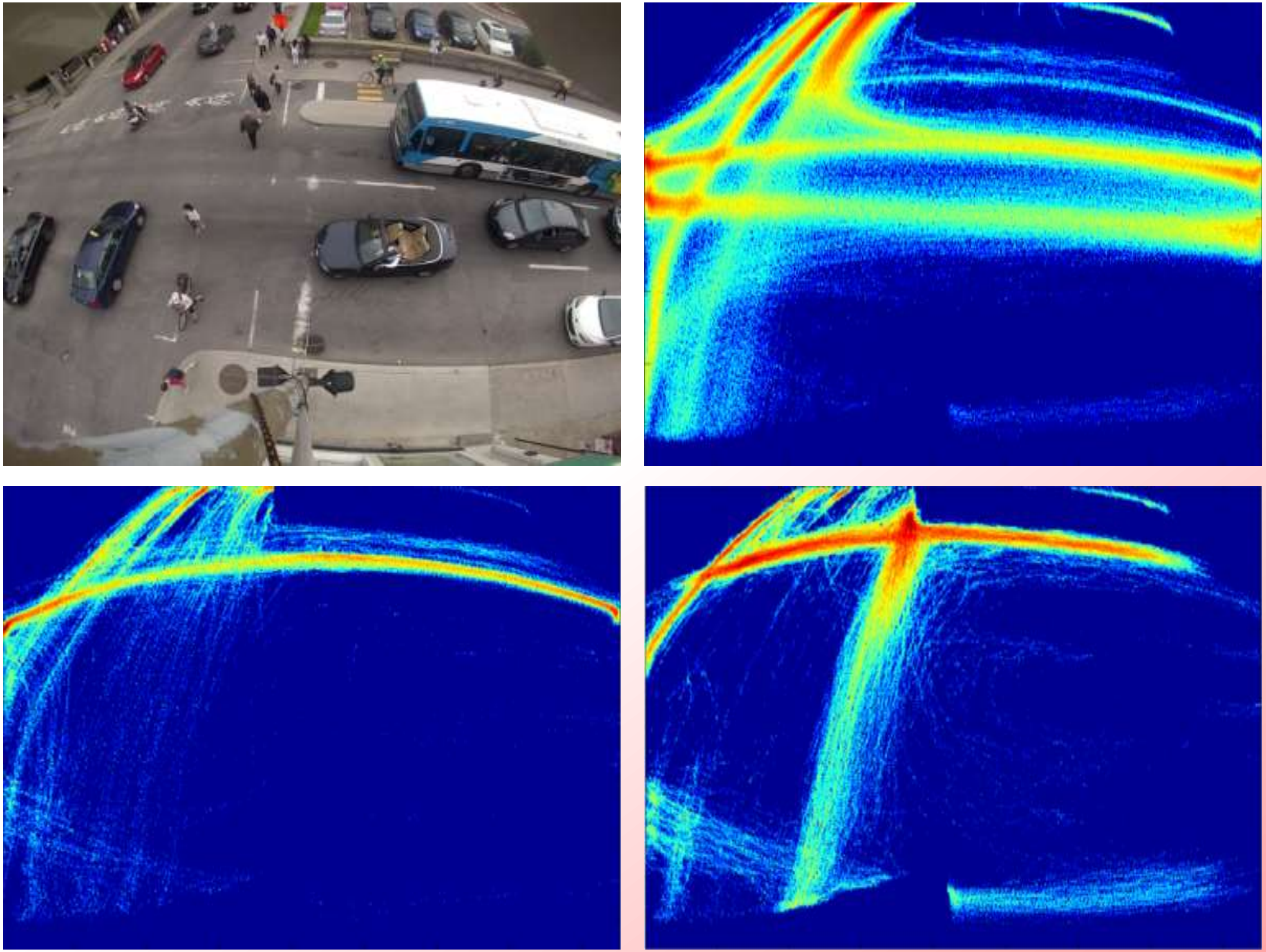
Object Classification Prediction

Video → Traffic Intelligence (Tracker) → Trajectory & Image Boxes of all Moving Objects → Resizing & HOG → HOG Feature Descriptors → Trained SVM Model → Integrate Speed → Predict the Class

Object Classification Accuracy

Confusion Matrix			Ground Truth					Accuracy
			Pedestrian	Bike	Vehicle	Total	**Precision**	
Predicted	**Classifier I**	Pedestrian	946	86	277	1309	**72.3 %**	**72.4 %**
		Bike	77	324	793	1194	**27.1 %**	
		Vehicle	0	78	2175	2253	**96.5 %**	
		Total	1023	488	3245	4756		
		Recall	**92.5 %**	**66.4 %**	**67.0 %**			
	Classifier II	Pedestrian	742	191	584	1517	**48.9 %**	**75.9 %**
		Bike	121	244	37	402	**60.7 %**	
		Vehicle	160	53	2624	2837	**92.5 %**	
		Total	1023	488	3245	4756		
		Recall	**72.5 %**	**50.0 %**	**80.9 %**			
	Classifier III	Pedestrian	726	43	64	833	**87.2 %**	**86.3 %**
		Bike	131	373	177	681	**54.8 %**	
		Vehicle	166	72	3004	3242	**92.7 %**	
		Total	1023	488	3245	4756		
		Recall	**71.0 %**	**76.4 %**	**92.6 %**			
	Classifier IV	Pedestrian	969	53	180	1202	**80.6 %**	**88.5 %**
		Bike	42	371	198	611	**60.7 %**	
		Vehicle	12	64	2867	2943	**97.4 %**	
		Total	1023	488	3245	4756		
		Recall	**94.7 %**	**76.0 %**	**88.4 %**			

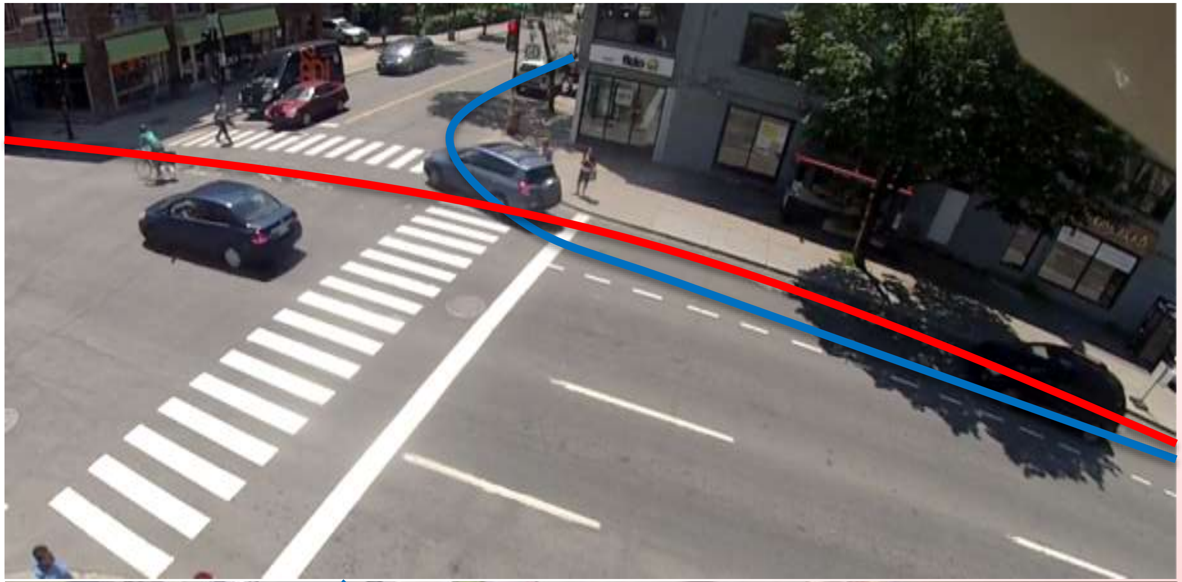
Object Classification Accuracy



Case Studies on Cyclist Safety

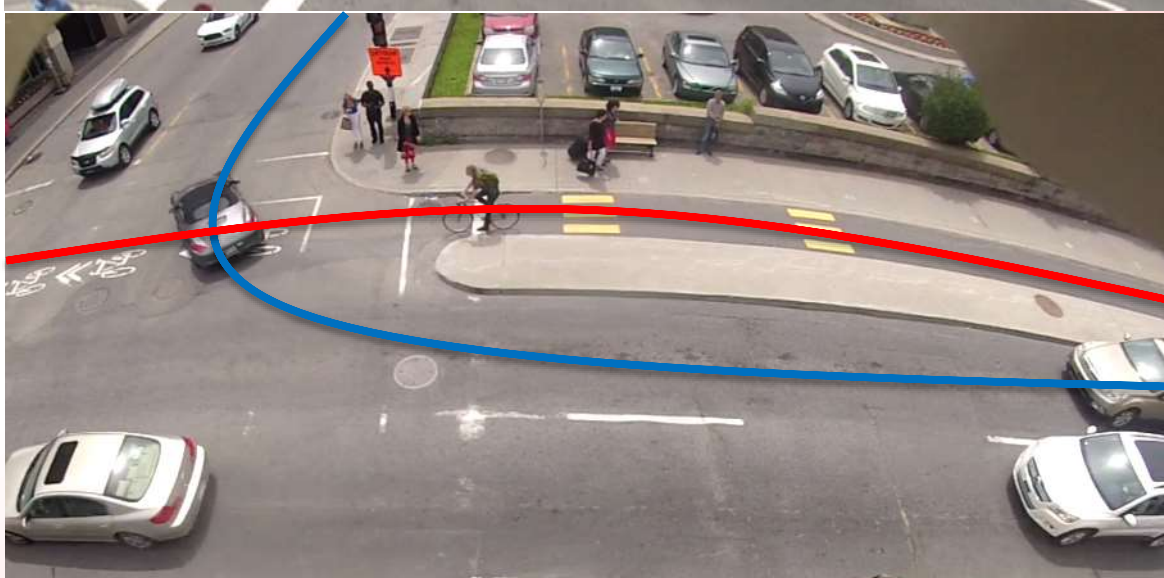
1st Study: Cycle Track

Without cycle track



Saint-Urbain Mont-Royal

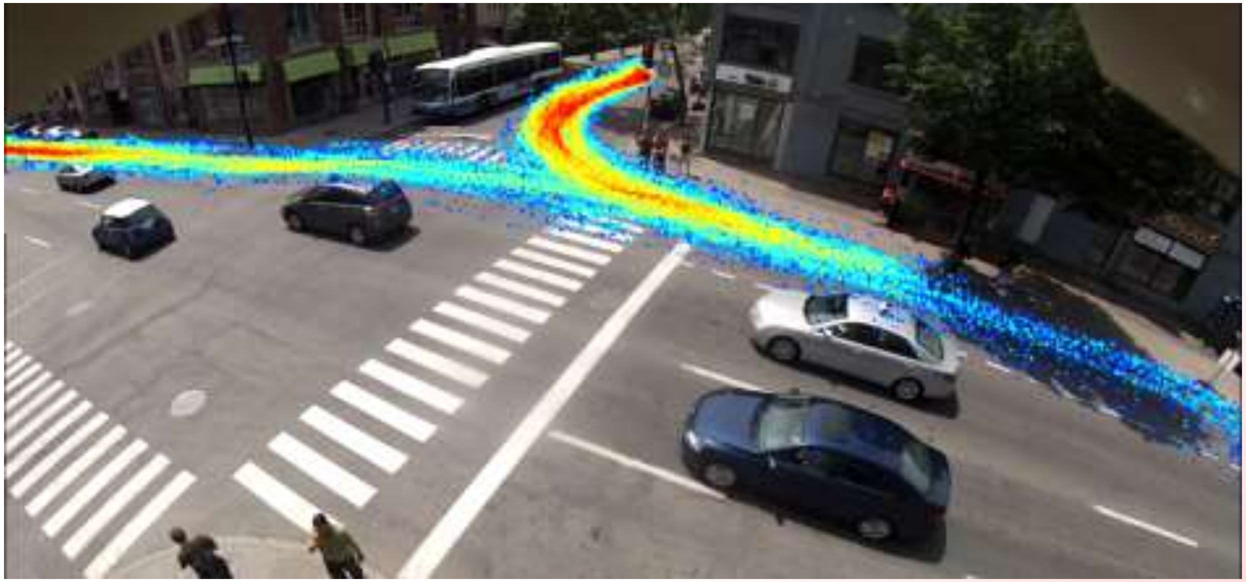
With cycle track



Saint-Urbain Pins

1st Study: Cycle Track

Without cycle track



Saint-Urbain
Mont-Royal

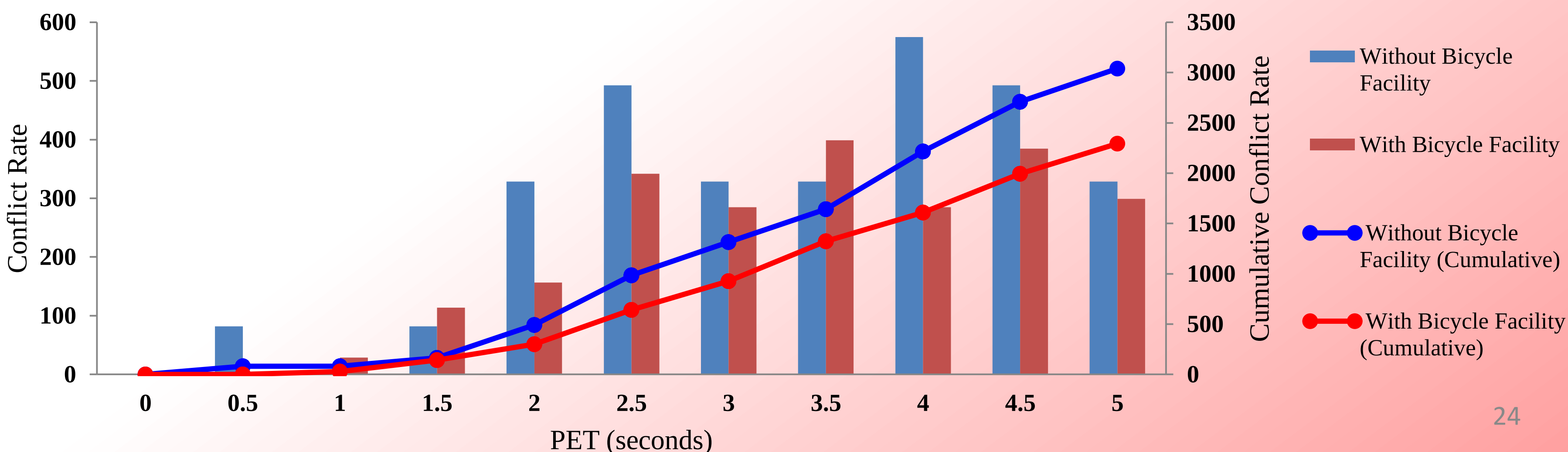
With cycle track

Saint-Urbain
Pins

1st Study: Cycle Track

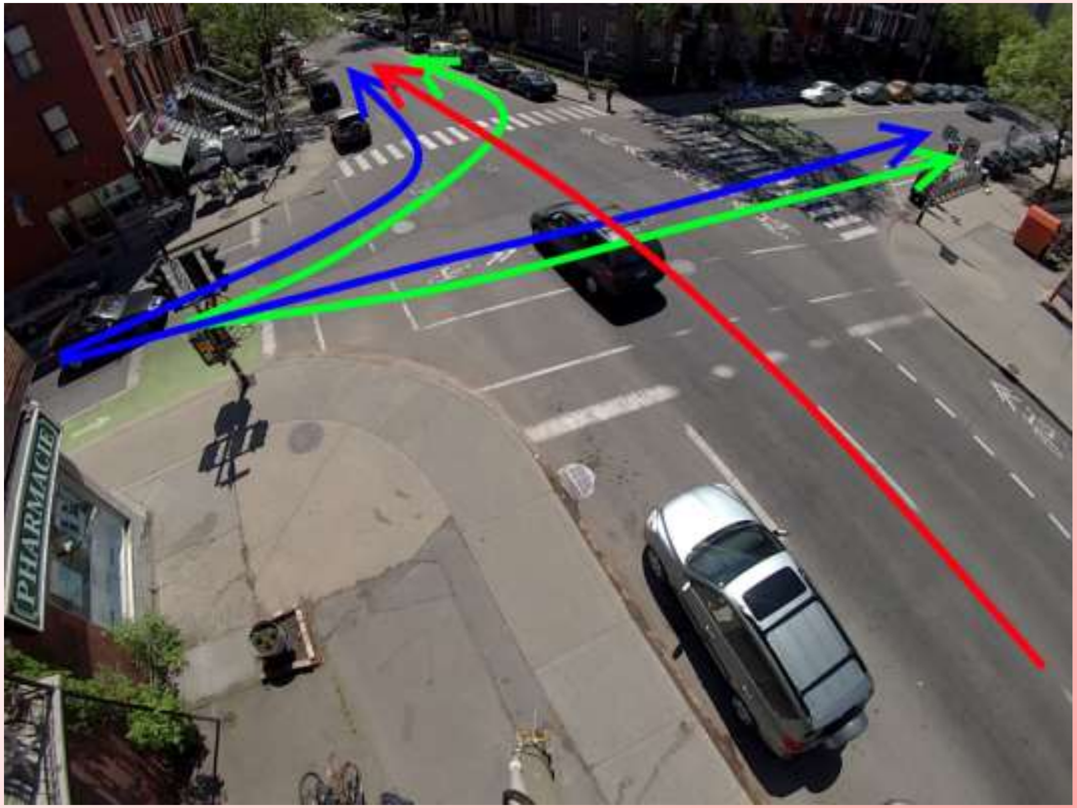
$$Conflict\ Rate = \frac{(Frequency\ of\ Conflicts,\ \ per\ Hour) * 10^6}{(Tracked\ Cyclists, per\ Hour) * (Tracked\ Vehicles, per\ Hour)}$$

	Hours of Video	Cyclists	Right-Turning Vehicles	Average Cyclist Speed	Average Vehicle Speed	TTC^{15} < 5 seconds	TTC^{15} < 1.5 seconds	PET < 5 seconds	PET < 1.5 seconds	TTC Conf. Rate*	TTC Dang. Conf. Rate*	PET Conf. Rate*	PET Dang. Conf. Rate*
Without bicycle facility	2.57	119	263	11.8	12.3	4	2	37	2	**328**	**164**	**3038**	**164**
With bicycle facility	3.88	438	622	15.2	13.7	13	4	161	10	**185**	**57**	**2293**	**142**



2nd Study: Bicycle Box

- 11.7h of video for intersections without bicycle box (3 intersections)
- 10.1h of video for intersections with bicycle box (2 intersections)
- Two types of conflicts:
  - Conflict Type 1: Cyclist (green) with Vehicle (red)
  - Conflict Type 2: Cyclist (green) with Vehicle (blue)
- Modelling conflicts by logit model
  - Number of lanes
  - Red and green times
  - Land use
  - Presence of bicycle box
  - Any other bicycle facility
  - Traffic flow of cyclists (30s before)
  - Traffic flow of vehicles (30s before)



2^{nd} Study: Bicycle Box

Variables	Conflicts Type 1 (Green - Red)				Conflicts Type 2 (Green - Blue)			
	Conflict (PET < 5s)		Dangerous Conflict (PET < 1.5s)		Conflict (PET < 5s)		Dangerous Conflict (PET < 1.5s)	
	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Constant	-2.99	0.00	-4.35	0.00	-0.56	0.00	-1.95	0.00
Cyclist Flow (green) passing 30s before	-	-	-	-	0.4230	0.00	0.4340	0.00
Vehicle Flow 1 (red) passing 30s before	0.1170	0.00	0.0970	0.00	-0.0857	0.00	-0.0823	0.01
Vehicle Flow 2 (blue) passing 30s before	0.0628	0.00	-	-	0.0908	0.00	0.0399	0.04
Presence of Bicycle Box	**-0.726**	**0.00**	**-2.050**	**0.00**	**-0.739**	**0.00**	**-1.230**	**0.00**
Number of total observations	1074		1074		1074		1074	
Number of positive observations	103		14		291		79	
Final log-likelihood	-299.85		-66.44		-544.00		-251.48	
Constant log-likelihood	-339.37		-74.67		-627.43		-282.19	
Adjusted Rho²	0.592		0.907		0.263		0.655	

Thank You!

Bayes' Rule

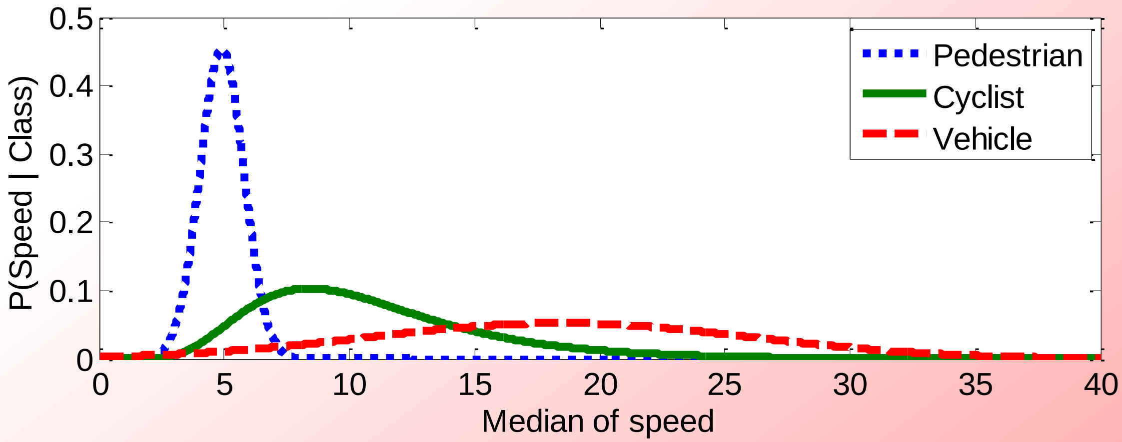
$$P(\text{Class} \mid \text{Speed}, \text{Appearance}) = \frac{P(\text{Class})}{P(\text{Speed}, \text{Appearance})} P(\text{Speed}, \text{Appearance} \mid \text{Class})$$

$$P(\text{Class} \mid \text{Speed}, \text{Appearance}) = \frac{P(\text{Class})}{P(\text{Speed})P(\text{Appearance})} P(\text{Speed}|\text{Class})P(\text{Appearance}|\text{Class})$$

$$P(\text{Appearance}|\text{Class})P(\text{Class}) = P(\text{Class}|\text{Appearance})P(\text{Appearance})$$

$$P(\text{Class} \mid \text{Speed}, \text{Appearance}) = \frac{P(\text{Class}|\text{Appearance})}{P(\text{Speed})} P\,(\text{Speed}|\text{Class})$$

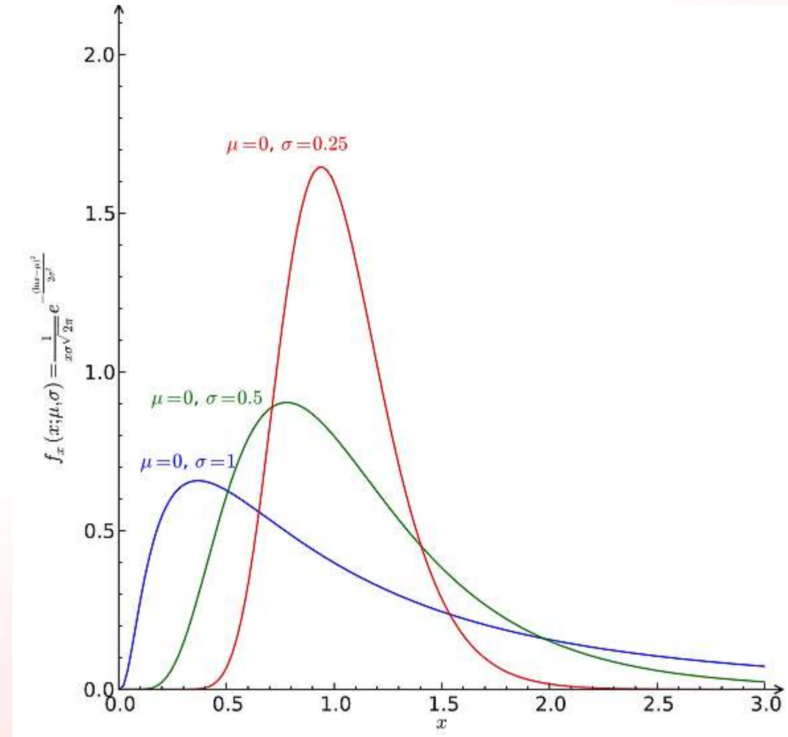
$$P(\text{Class} \mid \text{Speed}, \text{Appearance}) \propto P(\text{Class}|\text{Appearance})\ P\,(\text{Speed}|\text{Class})$$



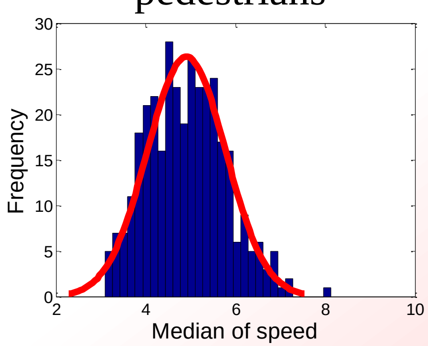
Normal, Log-Normal Distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

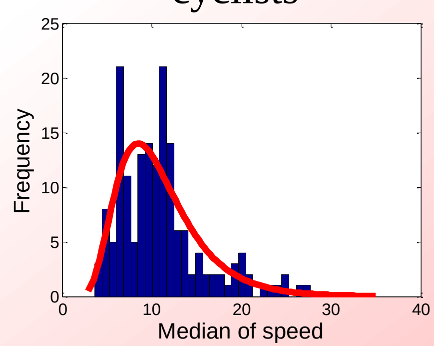
$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{[\ln(x)-\mu]^2}{2\sigma^2}}$$



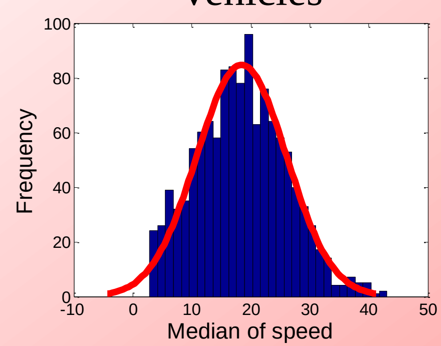
pedestrians



cyclists

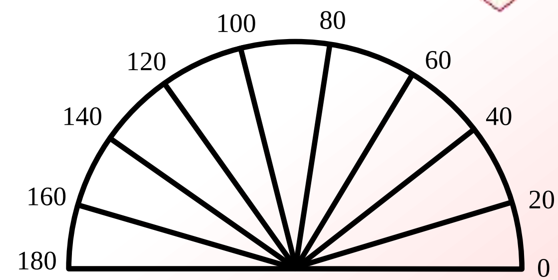


vehicles



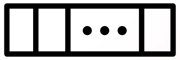
HOG

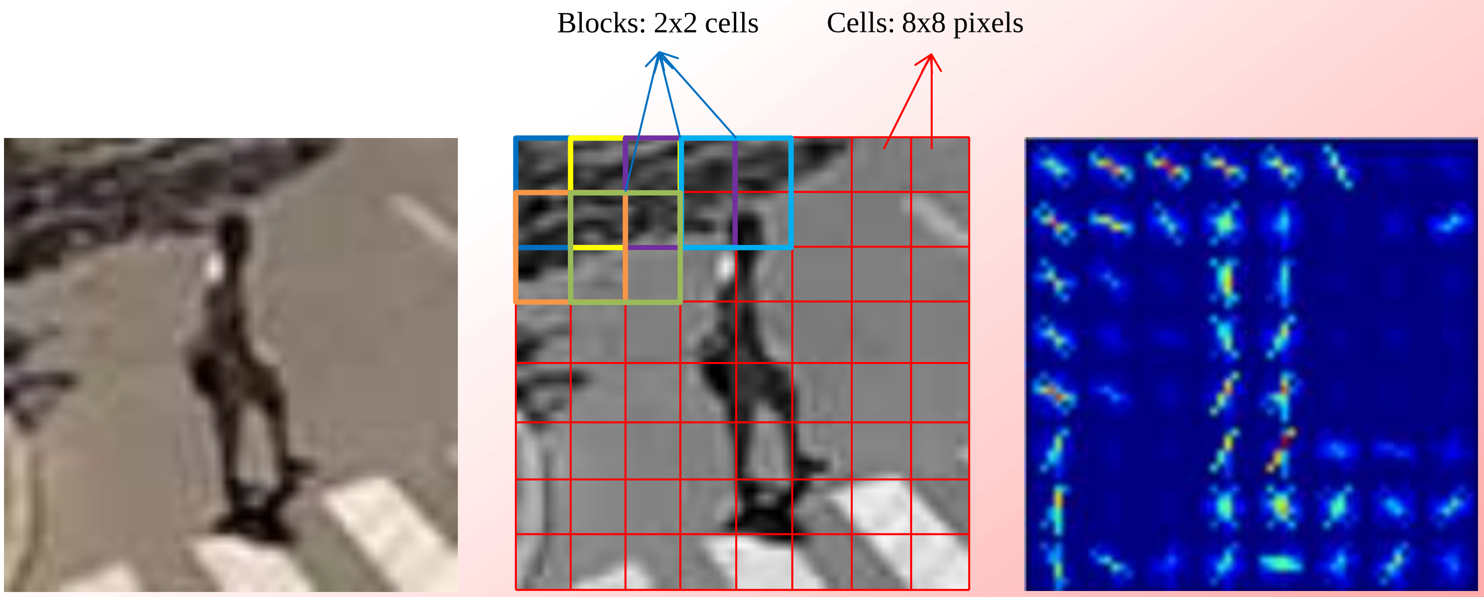
- Normalized image size: 64x64 pixels
- Number of pixels per cell: 8x8
- Number of cells per block: 2x2
- Number of orientations: 9



Number of orientations: 9 bins

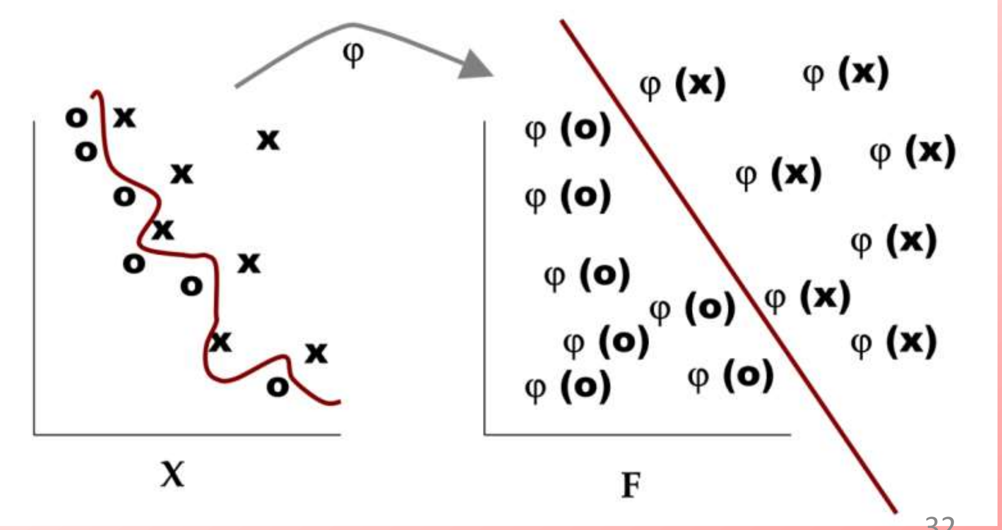
- Normalization over the blocks for each cell: $v \leftarrow \frac{v}{\sqrt{\|v\|^2+\varepsilon}}$
- Vector dimension: 49 x 4 x 9 = 1764





SVM

- Non Linear SVM
- Here we used RBF kernel (Radial Basis Function)



Recall – Precision - Accuracy

Confusion Matrix			Ground Truth					Accuracy
			Pedestrian	Bike	Vehicle	Total	**Precision**	
Predicted	**Classifier IV**	Pedestrian	969	53	180	1202	**80.6 %**	**88.5 %**
		Bike	42	371	198	611	**60.7 %**	
		Vehicle	12	64	2867	2943	**97.4 %**	
		Total	1023	488	3245	4756		
		Recall	**94.7 %**	**76.0 %**	**88.4 %**			

$$Recall_k = \frac{c_{kk}}{\sum_j c_{kj}}$$

$$Precision_k = \frac{c_{kk}}{\sum_i c_{ik}}$$

$$Accuracy = \frac{\sum_k c_{kk}}{\sum_i \sum_j c_{ij}}$$

ROC

Confusion Matrix			Ground Truth					Accuracy
			Pedestrian	Bike	Vehicle	Total	**Precision**	
Predicted	**Classifier IV**	Pedestrian	969	53	180	1202	**80.6 %**	**88.5 %**
		Bike	42	371	198	611	**60.7 %**	
		Vehicle	12	64	2867	2943	**97.4 %**	
		Total	1023	488	3245	4756		
		Recall	**94.7 %**	**76.0 %**	**88.4 %**			

- True positive rate: true positive out of all the positives
- False positive rate: false positive out of all the negatives

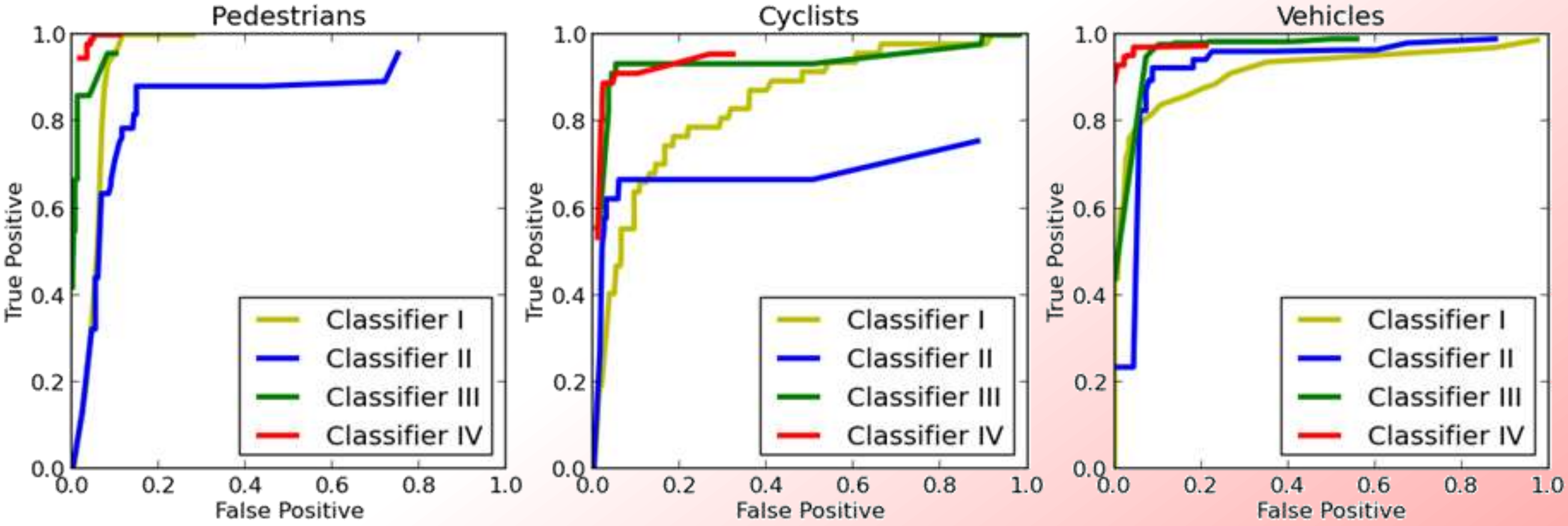
- For example for pedestrian:

True Positive Rate (pedestrian) = Recall =

False Positive Rate (pedestrian) =

Object Classification Accuracy

- Receiver Operating Characteristic (ROC)
- To reduce the effect of poor choice of parameters



TTC

- Time To Collision
  - Is a measure of remaining time (at any time t) before two objects collide, in case of no reaction from them