Cross or Not: Pedestrian Prediction

David Chu (dc788), Kelvin Hu (sh2442)
with Eric Jing under the supervision of Harry Chao and Jose Nino

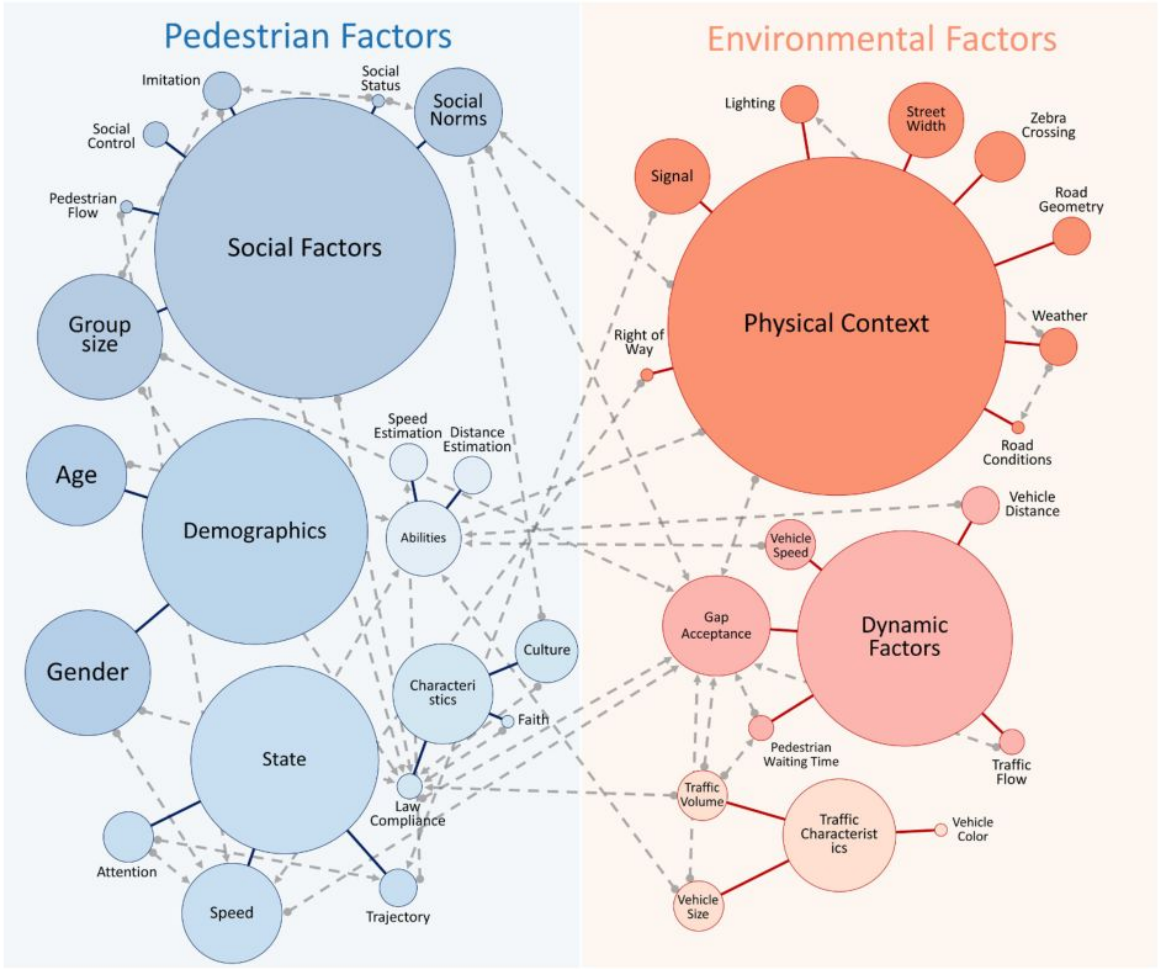
If you're a self-driving car...

If you're a self-driving car...
you'd want to know if she will cross

Goal:

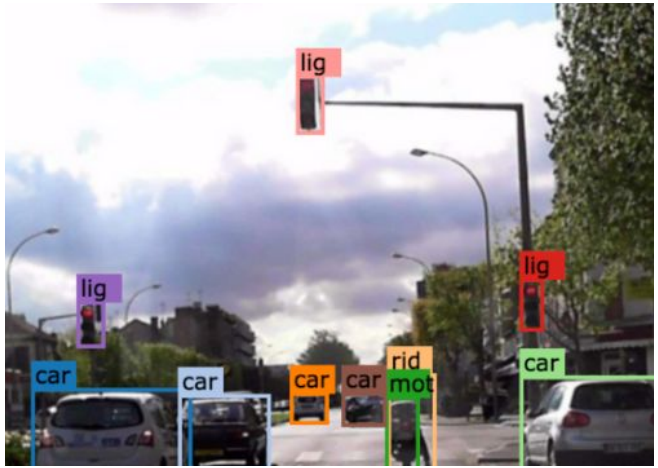
Predict “Will the pedestrian cross?”

Factors


Pedestrian Factors
Environmental Factors
Imitation
Social Status
Social Norms
Social Control
Pedestrian Flow
Social Factors
Group size
Age
Demographics
Gender
State
Attention
Speed
Trajectory
Law Compliance
Speed Estimation
Distance Estimation
Abilities
Characteri stics
Culture
Faith
Lighting
Street Width
Zebra Crossing
Signal
Road Geometry
Physical Context
Weather
Right of Way
Road Conditions
Vehicle Distance
Vehicle Speed
Gap Acceptance
Dynamic Factors
Pedestrian Waiting Time
Traffic Flow
Traffic Volume
Traffic Characterist ics
Vehicle Color
Vehicle Size

Context: environmental cues

- Traffic signs
- Traffic lights
- Vehicles
- Road
- Sidewalk



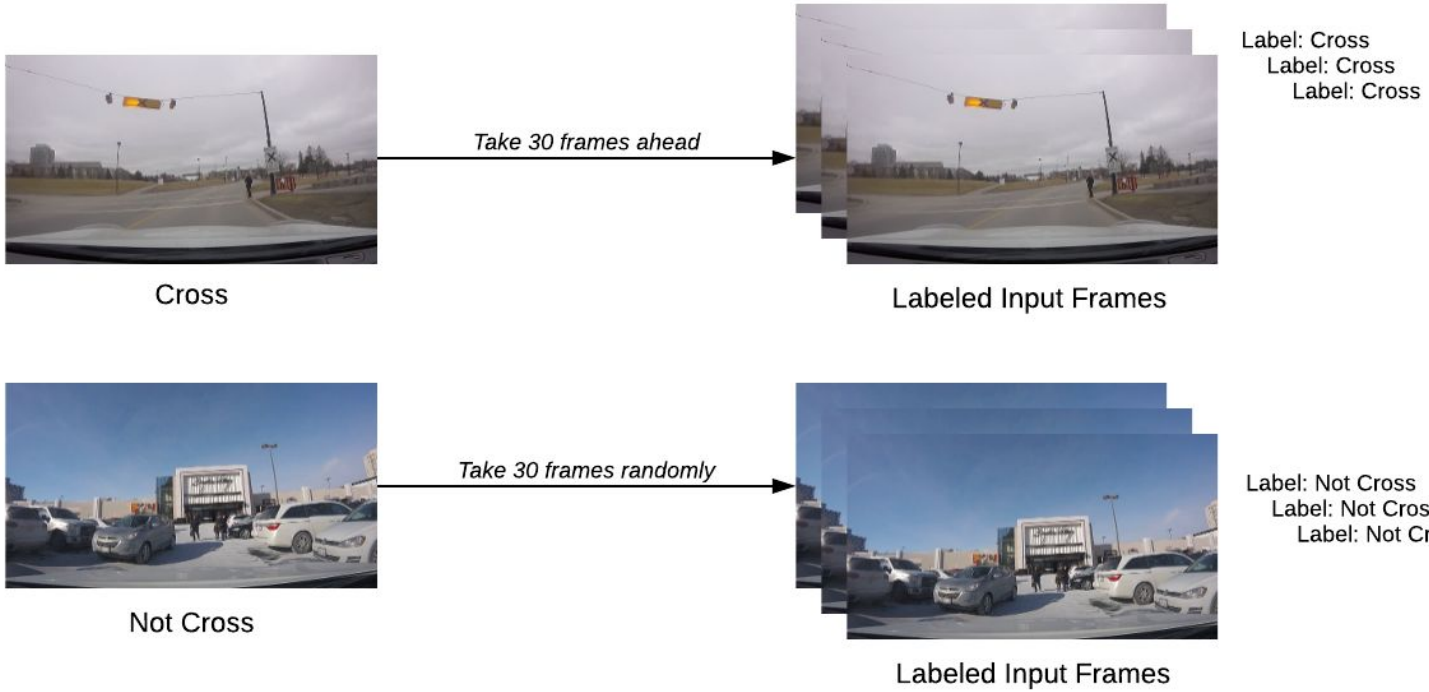
Previous work - Are They Going to Cross?

- Input
  - JAAD dataset (Joint Attention in Autonomous Driving)
  - **15 frames** before crossing event
  - **Global** environment cues:
    Crossing, **traffic sign**, **traffic signal, etc**
  - Pedestrian attributes: Looking, walking
- Output
  - SVM: Cross or not
- Results:
  - 234 crossing events, 81 non crossing events, 315 pedestrians
  - **62.73 ± 13.16%** accuracy

Our goals

- Input
  - JAAD dataset (Joint Attention in Autonomous Driving)
  - **30 frames** before crossing event
  - **Local** environment cues:
    Road, sidewalk, traffic sign, traffic sign, vehicle
  - Pedestrian attributes: Looking, walking
- Output
  - SVM: Cross or not

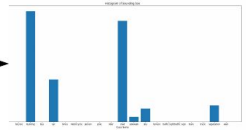
Input data


Take 30 frames ahead
Label: Cross
Label: Cross
Label: Cross
Cross
Labeled Input Frames
Take 30 frames randomly
Label: Not Cross
Label: Not Cross
Label: Not Cross
Not Cross
Labeled Input Frames

Pipeline

Frame

Mask RCNN

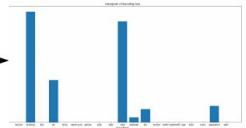


Histogram Filter

Traffic Sign: True
Traffic Light: True
Vehicle: False

Frame

DRN



Histogram Filter

Road: True
Sidewalk True

Pedestrian BBox

AlexNet

PCA

100 Features

Annotated Metadata

Daytime: True
Road Has Snow: False
Road Has Rain: True
Road Is Dry: False
Cloudy: True
Clear: False
Rainy: False
Snowy: False

SVM

Cross or Not Cross

Pipeline recap

1. Instance Segmentation (Mask R-CNN)
    a. Models based on ResNet 101
    b. Pretrained COCO model for **vehicles, traffic lights**
    c. Fine-tuned ImageNet model on Cityscapes for **traffic signs (international)**
2. Semantic Segmentation (Dilated Residual Network)
    a. Model based on Models based on ResNet 101
    b. Pretrained model on Cityscapes dataset
    c. Fine-tuned model on BDD dataset recognizes **roads, sidewalks**
3. AlexNet
    a. Select pedestrians for 30 frames
    b. Take the last layer output: a 2048-dimensional vector
    c. Apply PCA to generate **100 pedestrian attributes**
4. SVM
    a. Combine all features and predict crossing or not

Traffic sign instance segmentation

Vehicle/traffic light instance segmentation

Road/Sidewalk semantic segmentation

Results

- Baseline accuracy:
  62.73 ± 13.16%
- With only pedestrian attributes:
  69.85 ± 5.92%
- With traffic signs, traffic lights, and vehicles:
  69.56 ± 6.28%
- With traffic signs, traffic lights, vehicles, roads, and sidewalks:
  70.06 ± 6.72%

Tally of local objects detected:

- Num roads: 4921
- Num sidewalks: 1998
- Num signs: 210
- Num cars: 2281
- Num lights: 8

References

- Joint Attention in Driver-Pedestrian Interaction from Theory to Practice
  - Amir Rasouli, John K. Tsotsos.
- Are They Going to Cross?
  - A Rasouli, I Kotseruba, JK Tsotsos. Computer Vision Workshop (ICCVW), 2017 IEEE International Conference.