

# Pedestrian Intent Prediction

Stefano Zanatta, Pasquale Coscia, Lamberto Ballan

December 19, 2020

## 1 Introduction

We implemented the work of Haziq Razali, Alexandre Alah: Pedestrian-Intention-Prediction<sup>1</sup>. The original work consists in a Convolutional Neural Network - LSTM, with the goal to predict the intention of pedestrians on crossing the road.

The model was originally trained on Lausanne Dataset (not public) and on JAAD Dataset<sup>2</sup>. The former was recorded by static cameras, the latter on moving cameras (placed in the cars).

The goal of this project is to study the pipeline, reproduce the experiments on the JAAD dataset, then adapt and improve the model for the PIE dataset<sup>3</sup>.

## 2 Lausanne Dataset

## 3 JAAD Dataset

The JAAD datasets consists of mp4 videos, each one associated to a rich description, frame to frame, of what is happening on a given moment:

- id of the pedestrian
- coordinates of the pedestrian on the relative frame
- standing, walking
- looking at the traffic

---

<sup>1</sup><https://github.com/lucyze/Pedestrian-Intention-Prediction>

<sup>2</sup>[http://data.nvision2.eecs.yorku.ca/JAAD\\_dataset/](http://data.nvision2.eecs.yorku.ca/JAAD_dataset/)

<sup>3</sup>[https://data.nvision2.eecs.yorku.ca/PIE\\_dataset/](http://data.nvision2.eecs.yorku.ca/PIE_dataset/)

- hand gesture
- speeding up/slowing down
- crossing the road
- occlusion

For the target we only use the information about crossing the road.



The dataset contains some false positive annotations, for example the line `(1.0,0,463.0,730.0,69.0,118.0,0.0,1.0,0.0,0.0,0.0,568.0,0,00000001.0.png)` refers to the frame 0 of the video 02:



in particular there is a non-existent pedestrian with id 0:



This problem occurs for multiple videos. In most cases, every pedestrian is recognized.

## 4 Pipeline

### Adapting the ground truth

The first step is to create the ground truth for the model by transforming the input into a csv. Originally, the Lausanne dataset required a step to locate the pedestrians (with an external object detector) and the (static) coordinates of the road. This is not longer required, because JAAD and PIE include the coordinates of the pedestrians and whether or not they are crossing the road.

frame	id	tlx	tly	width	height	walking	standing	looking	incrossing
1,0	463	730	69	118	0	1	0	0	0
1,0	463	730	69	118	0	1	0	0	1

## Trajectories

The second Step consists in assigning, at each frame, the “lifetime” of the pedestrian in the video, and a “cross” column, indicating if they ever crossed the road in their lifetime.

incrossing	lifetime	cross
0	568	0
0	568	0

## Hungarian Tracker

Remove pedestrians with a lifetime below a threshold, and change pedestrians IDs to a numeric value.

## Crop Pedestrians

Crops each pedestrian into images, organized in sub-folders:

```
dataset
  └── all
      └── crops
          └── name_of_the_video
              └── id_of_the_pedestrian
                  └── id_pedestrian.png
```



Then appends the path of the crop to each frame's annotation.

filename	folderpath
00000001.png	crops_0001000000000
00000001.png	crops_0001000000000

## Save Scenes

Add the whole scenes to the dataset and adds their paths to the annotations.

```
dataset
  └── all
    └── scenes
      └── name_of_the_video
        └── id_video.png
```

The new annotations columns are:

```
scene_filename  scene_folderpath
00000001.png    scenes_0001
00000001.png    scenes_0001
```

## Split train test

In the last step the dataset is split into train and test:

```
dataset
└── train
dataset
└── test
```

## 5 Model

The model is composed of a VGG-16 pre trained CNN, a fully connecter layer (feature embedder), a LSTM and a linear classifier to predict if the pedestrian crossed the road.



## **VGG-16**

This is a pre-trained CNN that works as a feature extractor for each image. This operation is executed for each image in the batch.

## **Feature Embedder**

Embed the CNN output into a 255 units vector. Then perform dropout.

## **LSTM**

Process the sequence through the LSTM with ReLU activation function and dropout.

## **Linear Classifier**

Classify whether the pedestrian crossed the road or not.

## **Changes to the original model**

To improve the results, we added the entire scene for the prediction.

## **6 Results on JAAD**

#todo

## **7 PIE Dataset**

#todo

## **8 PIE Pipeline**

#todo

## **9 Results on PIE**

#todo