Deep Bike

Quick introduction

- Great deal of time spent searching for ideas and databases
- Focus on bike sharing
- Aim at tackling common issues from consumer standpoint

Summary

I. **Studying the current bike network**
- Clustering via k-means method
- Social networking method
- Embedding approach

I. **Maintaining the network**
- Predicting availability in bike stations
- Basic models
- A deep learning model

I. **Enforcing security for users**
- Detecting users on bike
  - Extracting faces of these users
  - Testing whether they wear a helmet

I. Studying the current bike network

Two main concerns

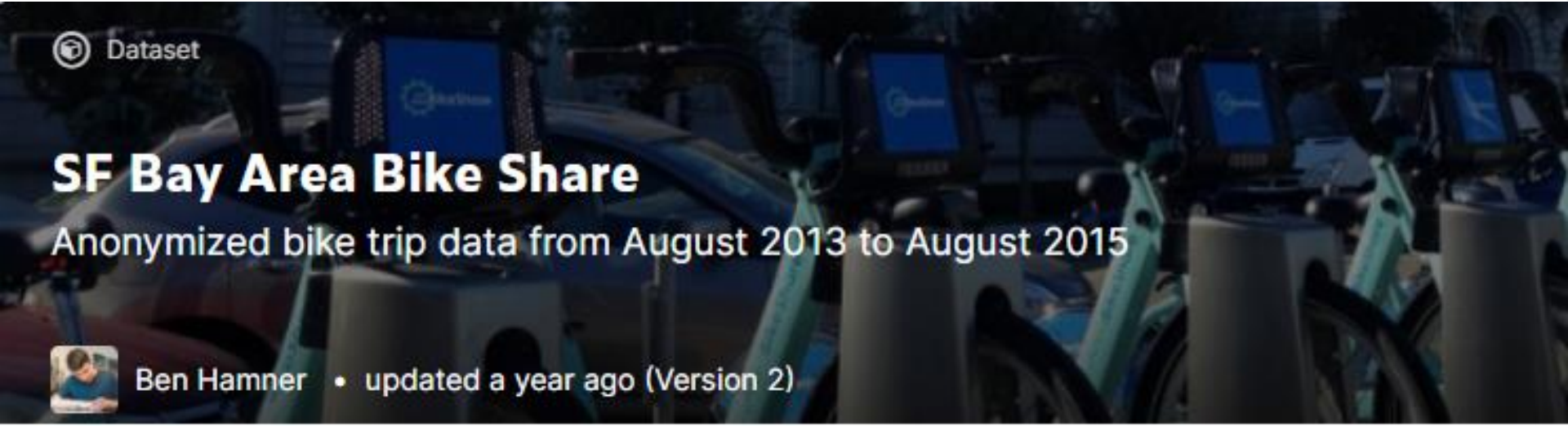
Problem 1 :

Teach a network what is a station from a user-oriented perspective

Problem 2 :

Predict where to place a new station

Our first database

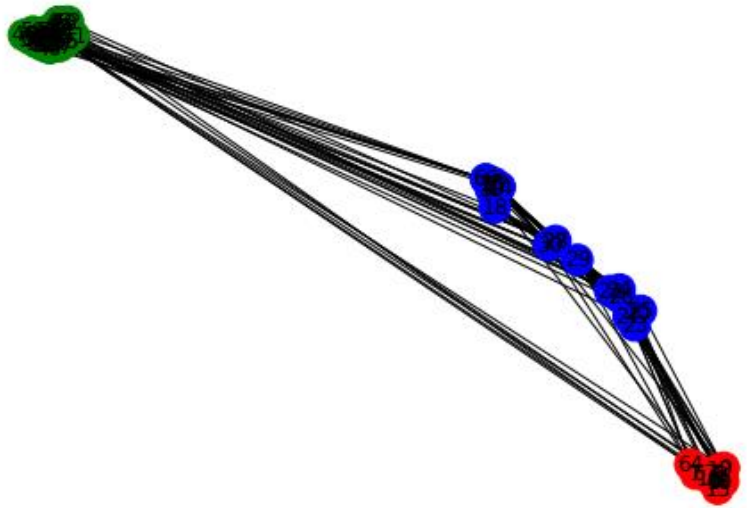


kaggle™

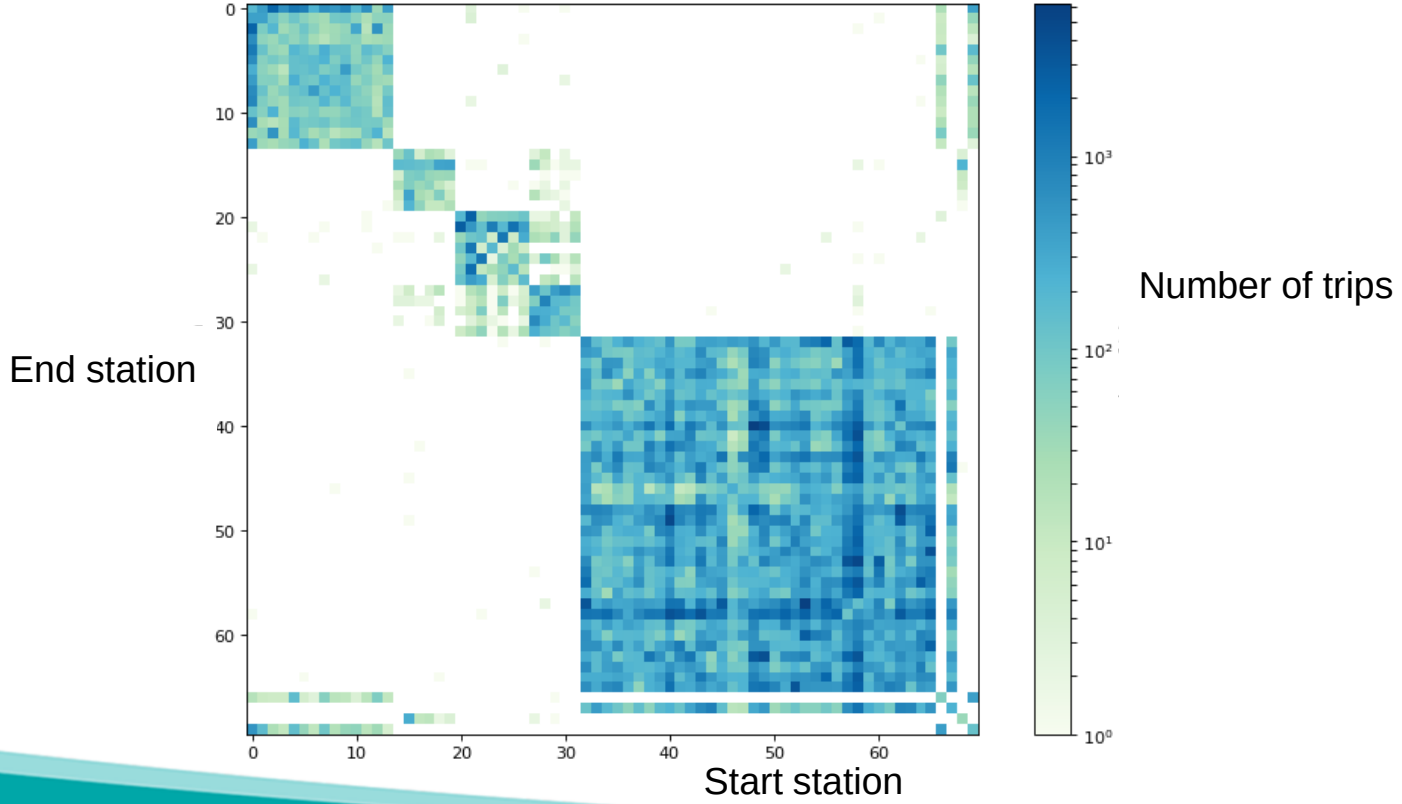
Clustering via k-means method

Objective : Plot data on a map to identify connections

-> Managed to identify 3 distinct geographical clusters : SF, San José, Stanford



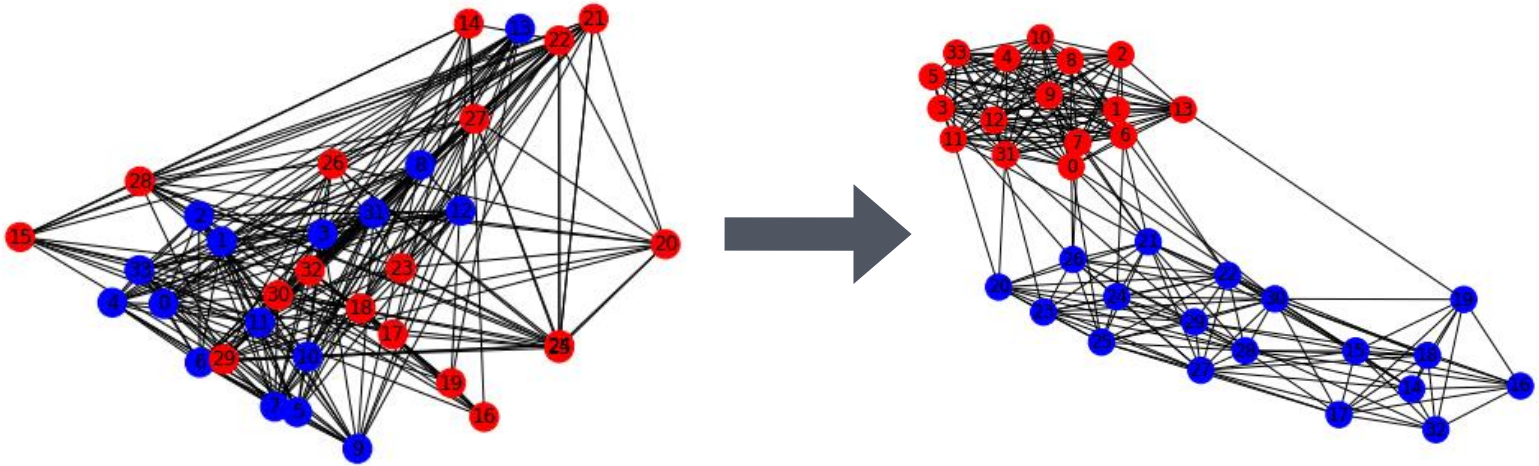
Volume of traffic between stations



Social networking method

Objective : Run a community detection algorithm on the graph of all stations

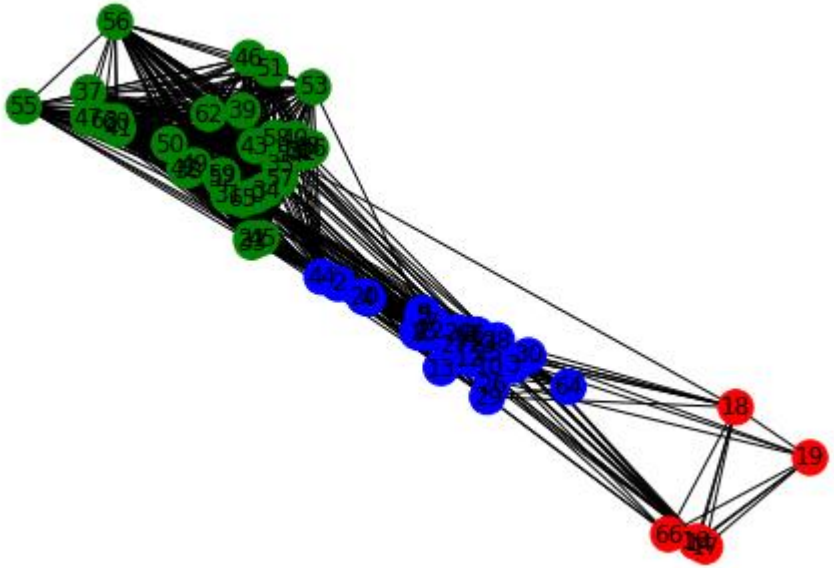
Outcome : Massive improvement -> highlights links between cities previously omitted



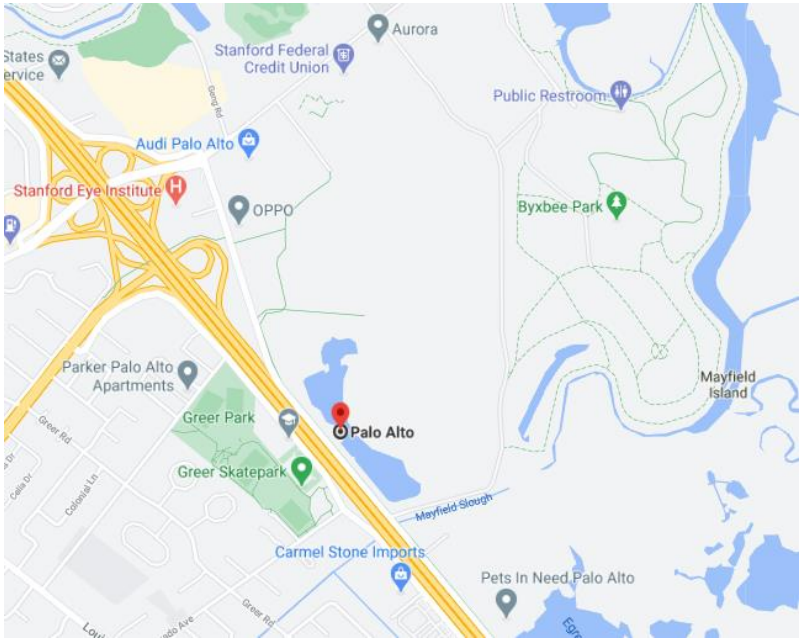
Embedding approach

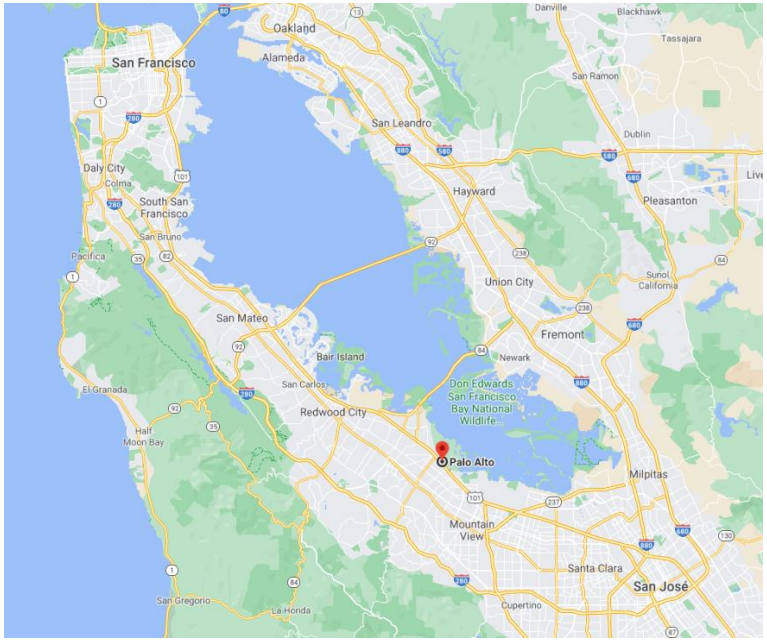
Objective : Train a neural network on maps -> predict where to better place a new station

Use of NLP methods

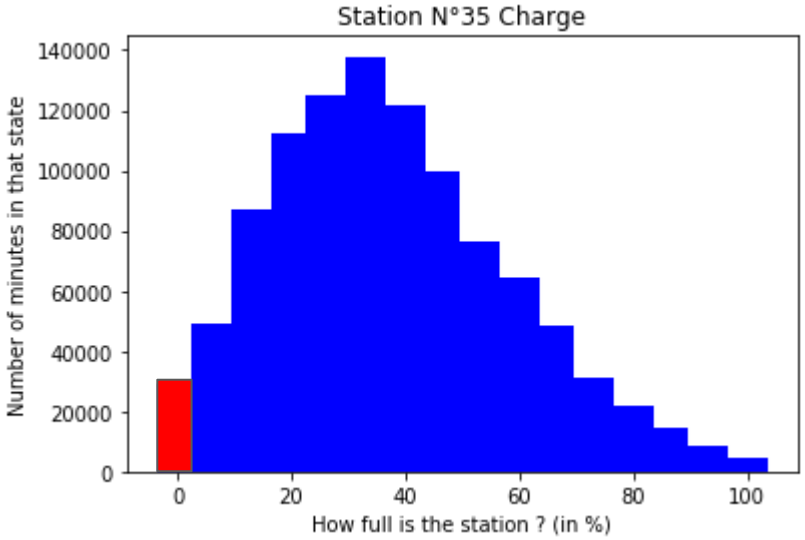


Where to put a new station ?





II. Maximizing bike availability



Some stations are empty for more than 45 minutes a day !

Predicting availability in bike stations

- The number of bikes available in each station
- The number of empty docks in each station

- The time of the day
- Informations about the weather

Machine Learning Model

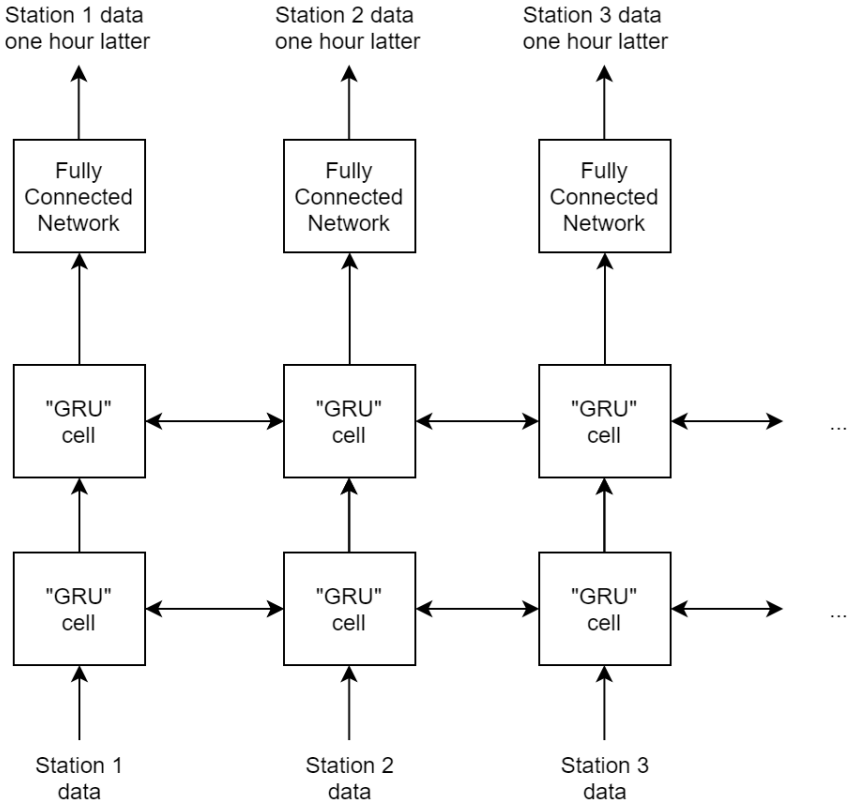
The number of bikes available and docks available for each station **one hour later**

The basic models

The model	Mean Squared Error (after rounding)
Bikes Available at t+1h = Bikes Available at t0	4.1
Linear Regression	3.9
Fully Connected Neural Network	3.8

Note : models were trained on the San Francisco part of the network to accelerate learning

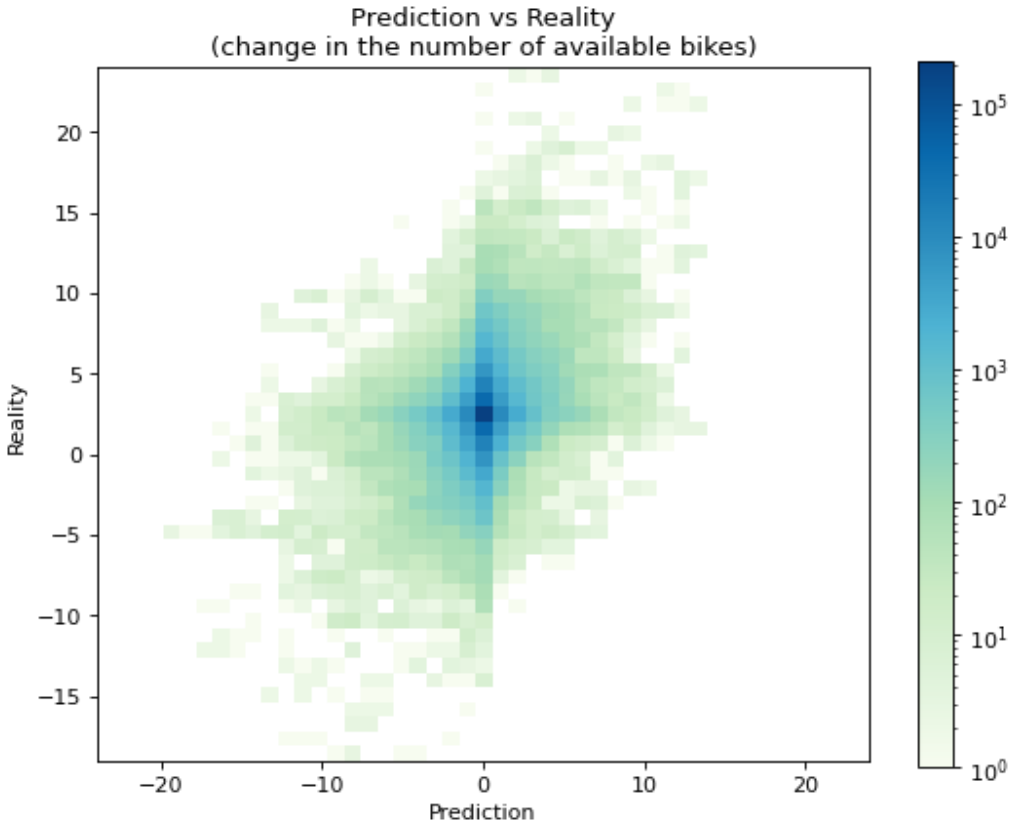
A Better Deep Learning Model



Results

Mean squared error: 3.4

17% improvement over the naïve approach!



Next steps

- Improve predictions using more data
- Improve predictions using better models like Graphnet
- Create a policy to move bikes and reduce empty station time

Make more people use bike sharing systems...

… and improve users safety !

III. Enforcing security for users

Problem :

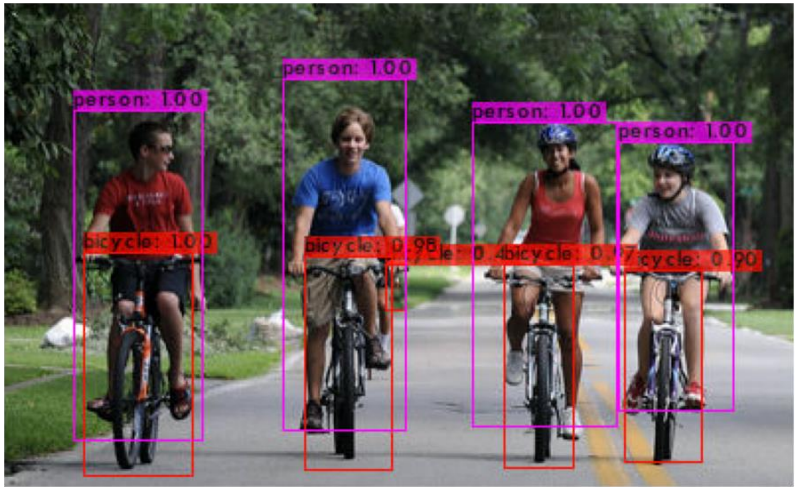
Implement a clever radar, detecting helmets from video feed

=> Discriminate efficiently bike users and pedestrians
=> Detect whether user wears a helmet or not

Finding bike users on a given image

A given image

Use of pre-trained YOLOv3 algorithm to identify boxes containing people and bicycles.



Calculate intersection between people and bike boxes.

Helmet detection : the second database

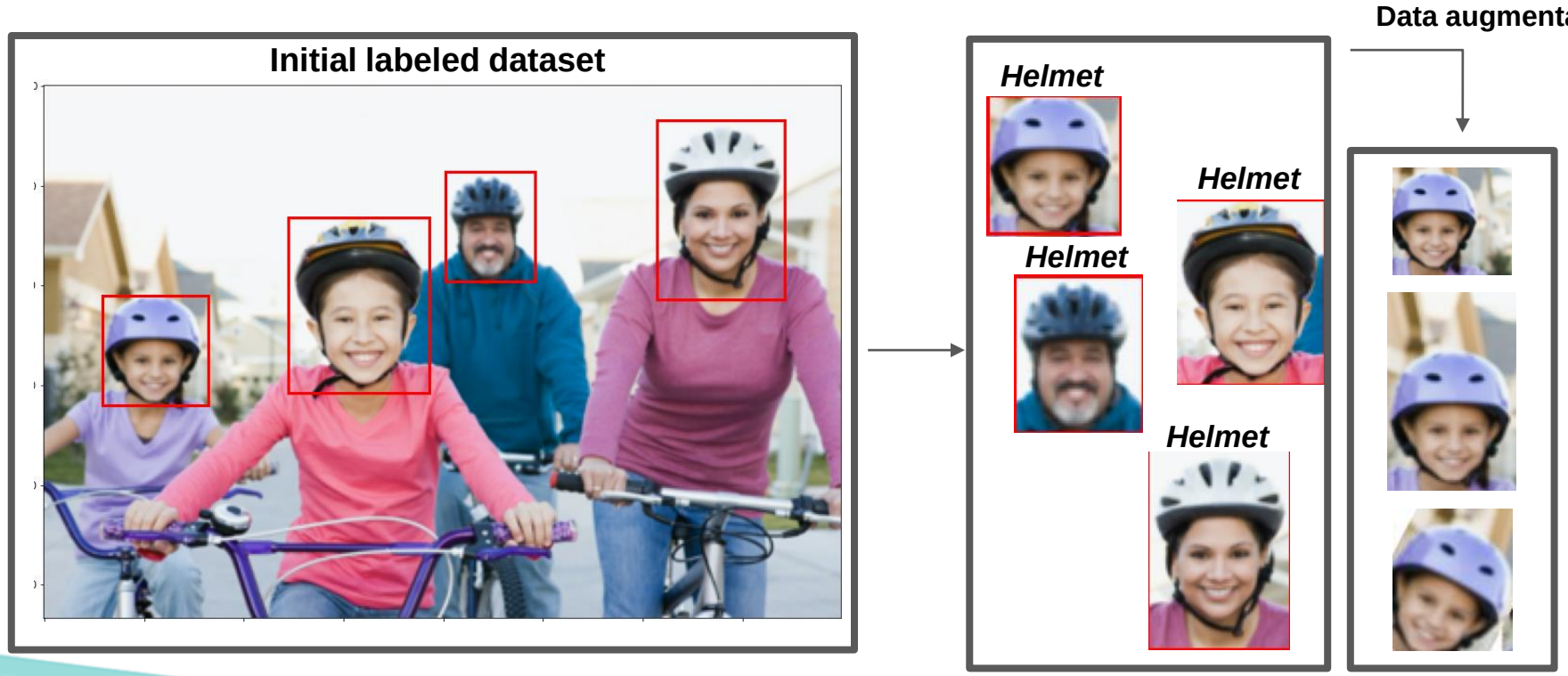
kaggle™

Helmet detection : faces of bike users

Face detection algorithm **ON** boxes of bike users previously detected -> yields **face boxes**

Said face boxes -> expanded with margin to ensure helmet inside **box**

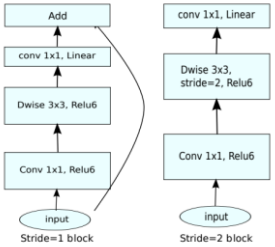
Helmet detection : Training the network



Creation of a new dataset made up of labeled faces with or without helmet

Helmet detection : Training the network

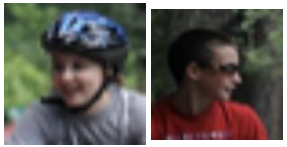
Pre-trained MobileNetV2, with two additional FC layers + Dropout + L2 regularization



(d) Mobilenet V2

Train	Test
99%	93%

Helmet *No helmet*



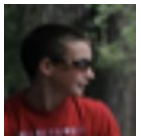
Pre-trained EfficientNetB7, with two additional FC layers + Dropout + L2

EfficientNetV2: Smaller Models and Faster Training

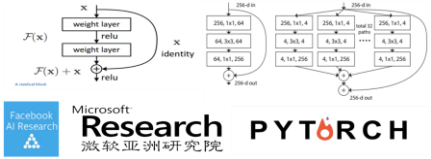
Mingxing Tan [1] Quoc V. Le [1]

Train	Test
89%	82%

Helmet *No helmet*



Pre-trained ResNet18, with some additional FC layers + learning rate schedule



Facebook AI Research

Microsoft Research 微软亚洲研究院

PYTORCH

Train	Test
96%	95%

Three different models giving test scores up to 95 %

Proof of concept

A picture taken today without helmet

Possible improvements

- Speed
- Improved accuracy of the helmet detection
- A bigger database (merely 700 pictures)
- Strive for speed detection, and other features ...

Conclusion

Thanks for listening !

Links

- The SF dataset : https://www.kaggle.com/benhamner/sf-bay-area-bike-share
- The helmet dataset : https://www.kaggle.com/brendan45774/bike-helmets-detection
- Empty station picture : https://www.google.com/url?sa=i&url=https%3A%2F%2Fmlmpages.wordpress.com%2Ftag%2Fvelib%2F&psig=AOvVaw2022C29kqr6Fg9-cDg0E2M&ust=1618241982643000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCOjlqajD9u8CFQAAAAAdAAAAABAD