北京交通大學

Deep Learning Based Motion Planning For Autonomous Vehicle Using Spatiotemporal LSTM Network

Zhengwei Bai

Mentor：Baigen Cai

2018.12.01

Contents

Self Introduction

Beijing
Jiaotong
University

Zhengwei Bai

2nd Year Master of Beijing Jiaotong University

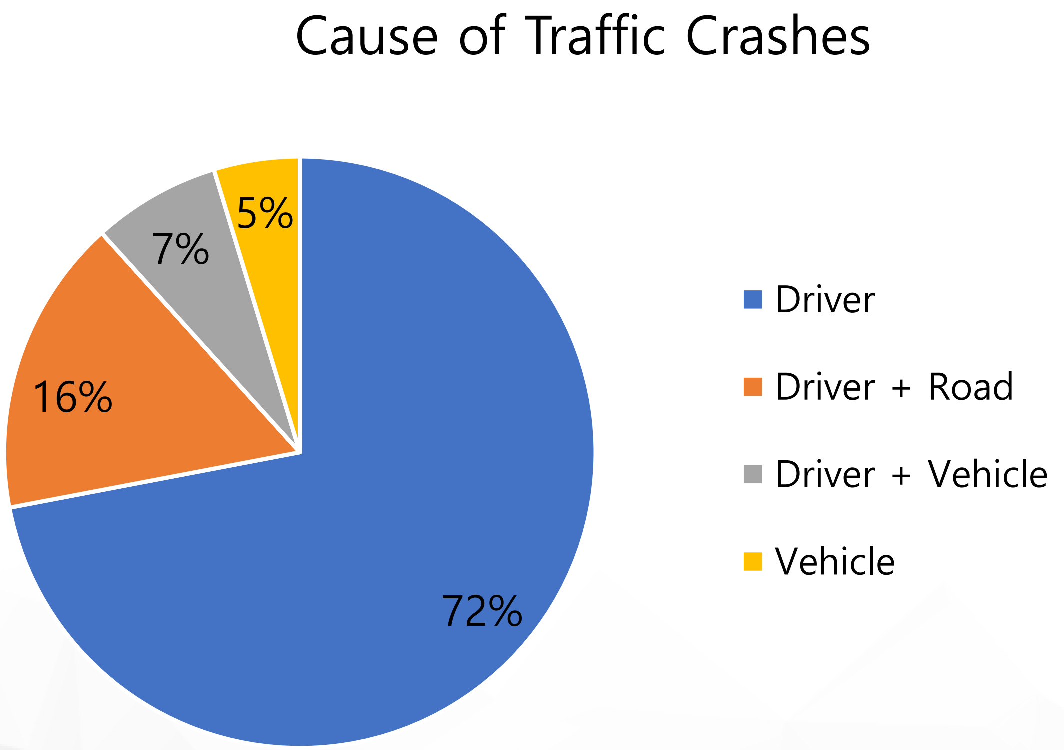
School of Electronic Information Engineering

Autonomous Driving & Vehicle Infrastructure Cooperative Systems(VICS)

Project Intro.

Motivation

- About 1.25 million people die each year as a result of road traffic crashes
- Without sustained action, road traffic crashes are predicted to become the seventh leading cause of death by 2030

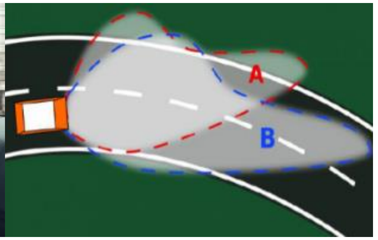


Background

Beijing Jiaotong University

ITS is essentially a multi-variable complex system with human-vehicle coupling

Human2Vehicle



Driving Behavior | Path Planning

Vehicle2Vehicle

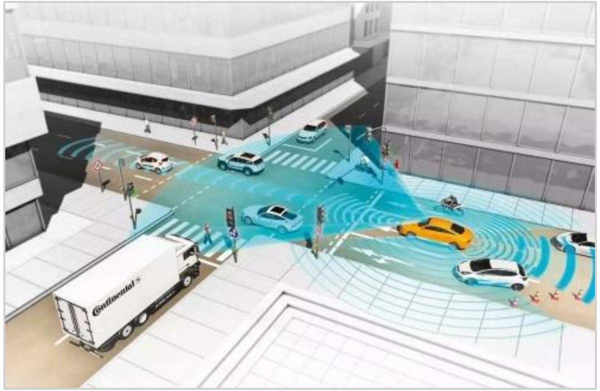
Car-following | Lane change

Vehicle2Road

Interwoven road | Intersection

H-V-R Coupling

Human
↕
Vehicle
↕
Road



Extrinsic Features of Traffic System

- Path planning
- Motion planning
- Motion control
- Road Capacity
- Service Level

……

Traffic Efficiency

Autonomous Vehicle : a Crucial part of ITS

Related Works

Model Based : Pre-defined control strategies

Model Free : DL based Learning Process

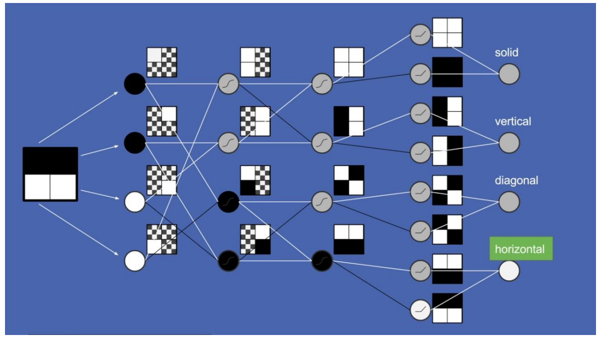
Imagination

Deep Learning: More adaptive to the complexity in real traffic scenarios

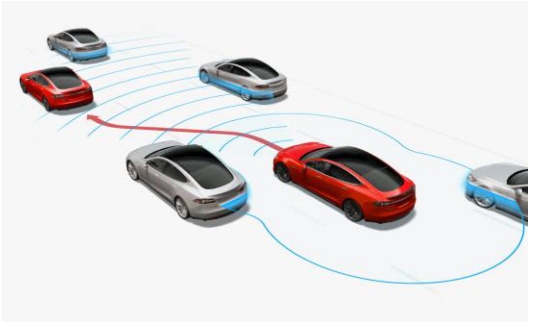
Project Overview

- **Input**: Human-level Control Through Deep Reinforcement Learning (Nature)
- **Algorithm**: Convolutional LSTM + 3D-CNN + FCNN
- **Output**: Motion planning (Steering angle)

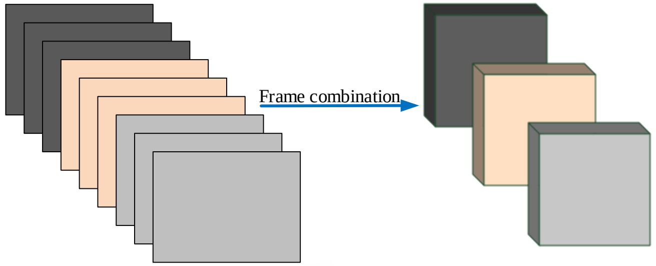
Camera Data



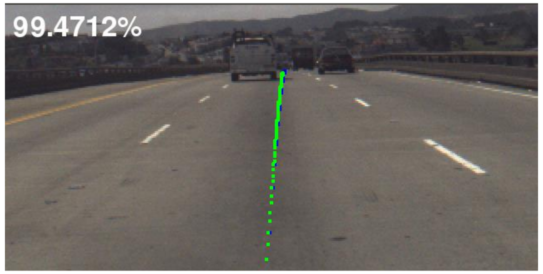
DL Algorithm



Motion Planning



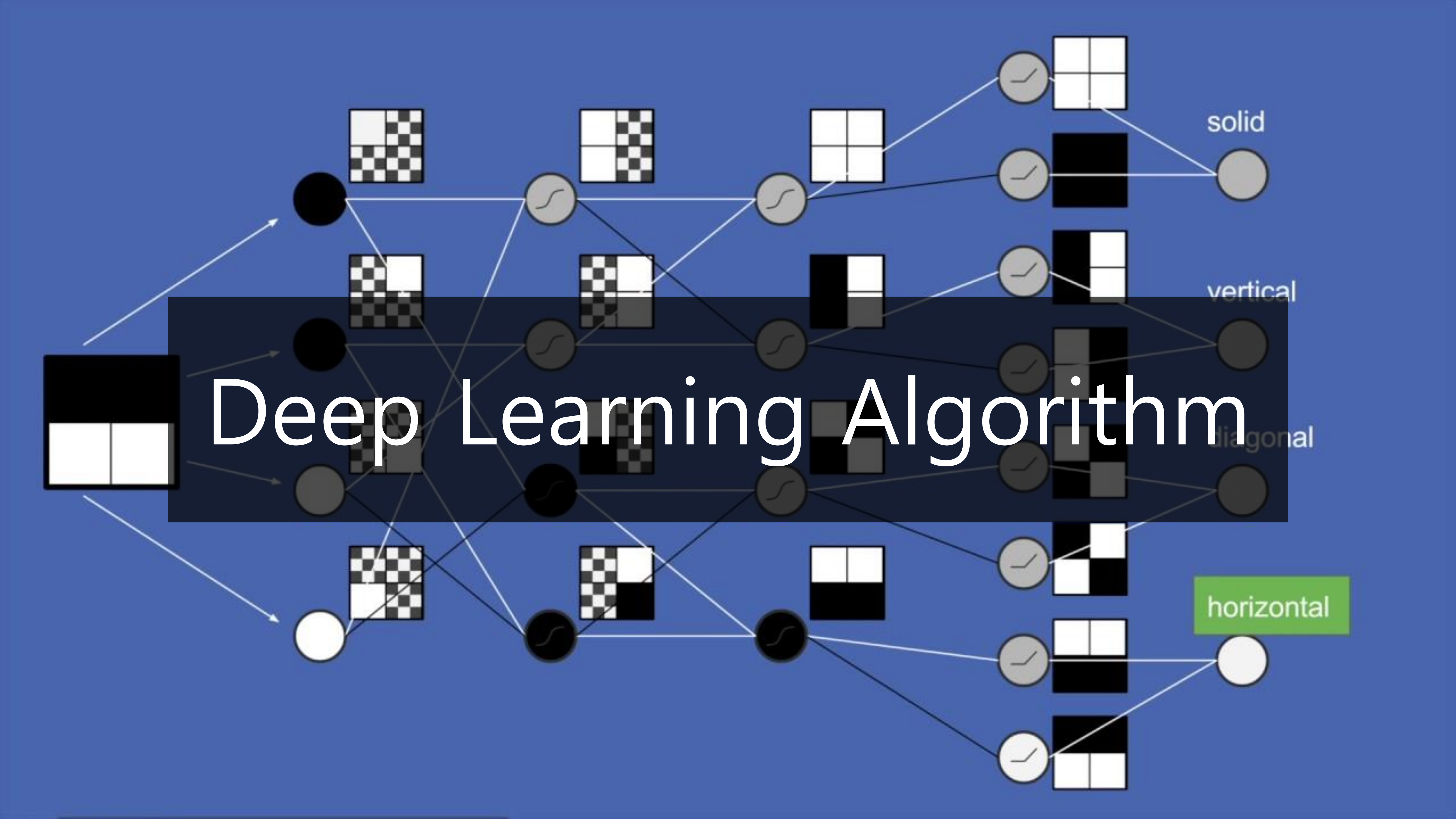
Preprocessing



Visualization

Raw image → Steering angle

Deep Learning Algorithm



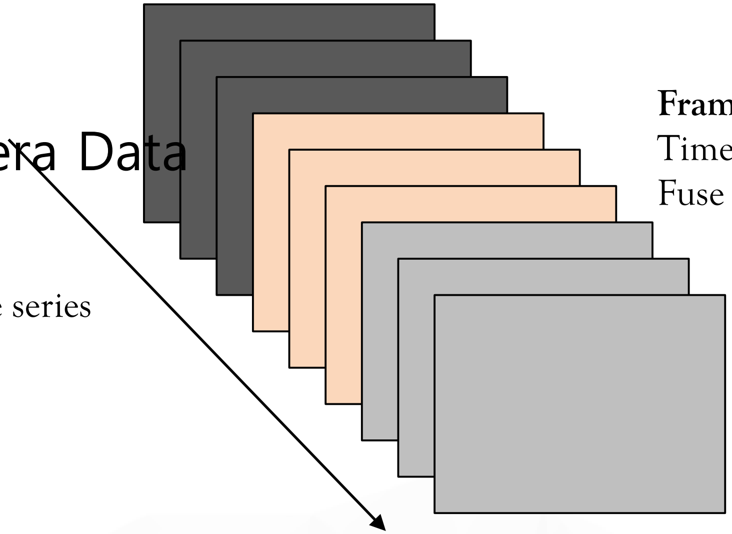
Data Preprocessing

Beijing
Jiaotong
University

Raw image
Dim: 320*160*3

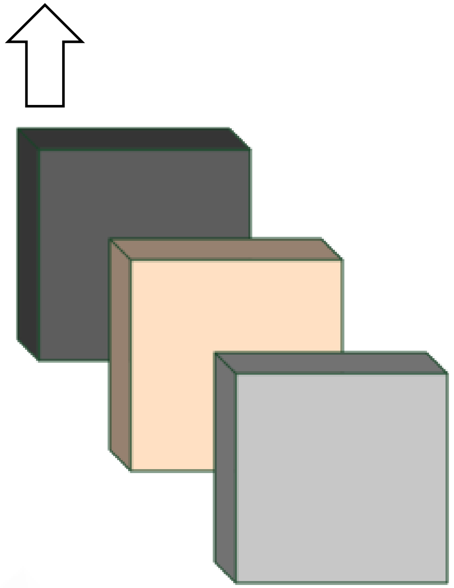
Multi-frame data
Dim: 320*160*3*3

Camera Data



Time series

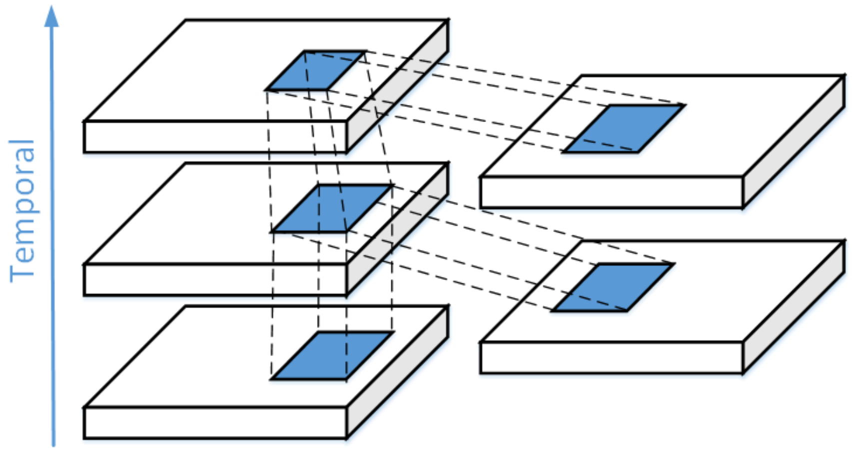
Frame Combination:
Time based 3 continuous frames
Fuse into a higher dimensional data format



Single frame image → Time based visual data

Convolutional LSTM

- Conv-LSTM :
  - Based on multi-frame picture segments
  - Use batch normalization between two layers



$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i)$$
$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f)$$
$$C_t = f_t + i_t \circ tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c)$$
$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o)$$
$$H_t = o_t \circ tanh(C_t)$$

i_t: Input gate at time t

f_t: Forget gate at time t

C_t: LSTM cell at time t

o_t : Out put gate at time t

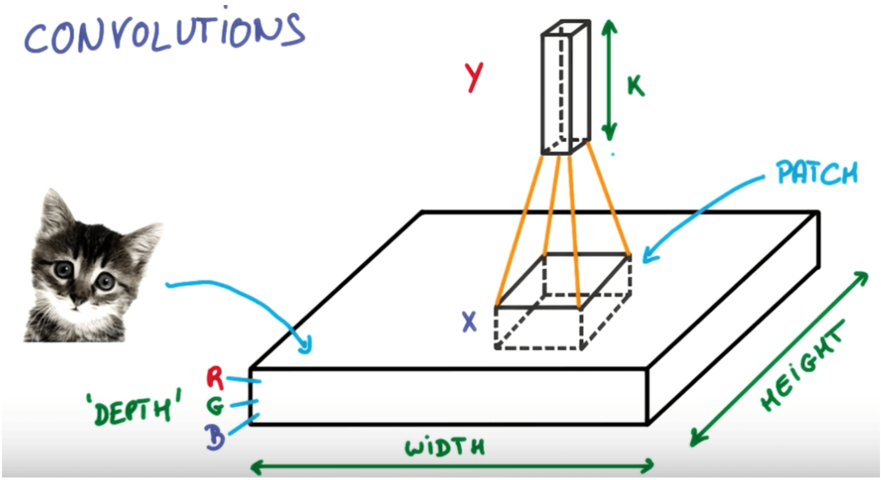
H_t: Hidden gate at time t

Extracting the temporal hidden feature information

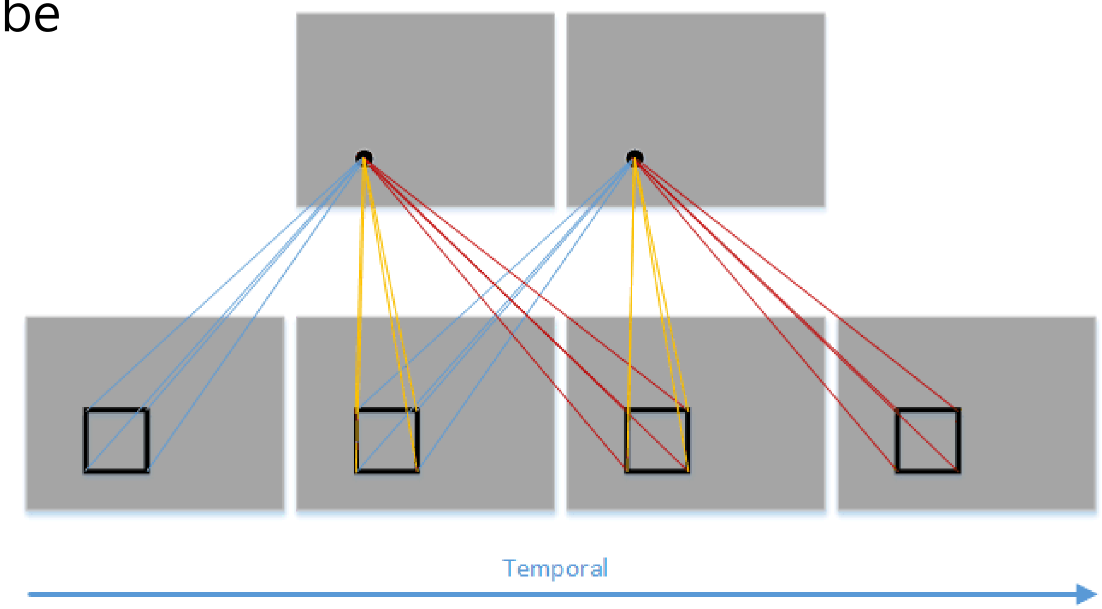
3D-CNN

Beijing Jiaotong University

- 3D-CNN :
  - Build a cube by stacking multiple consecutive frames
  - Use the 3D convolution kernel in the cube



2D-CNN

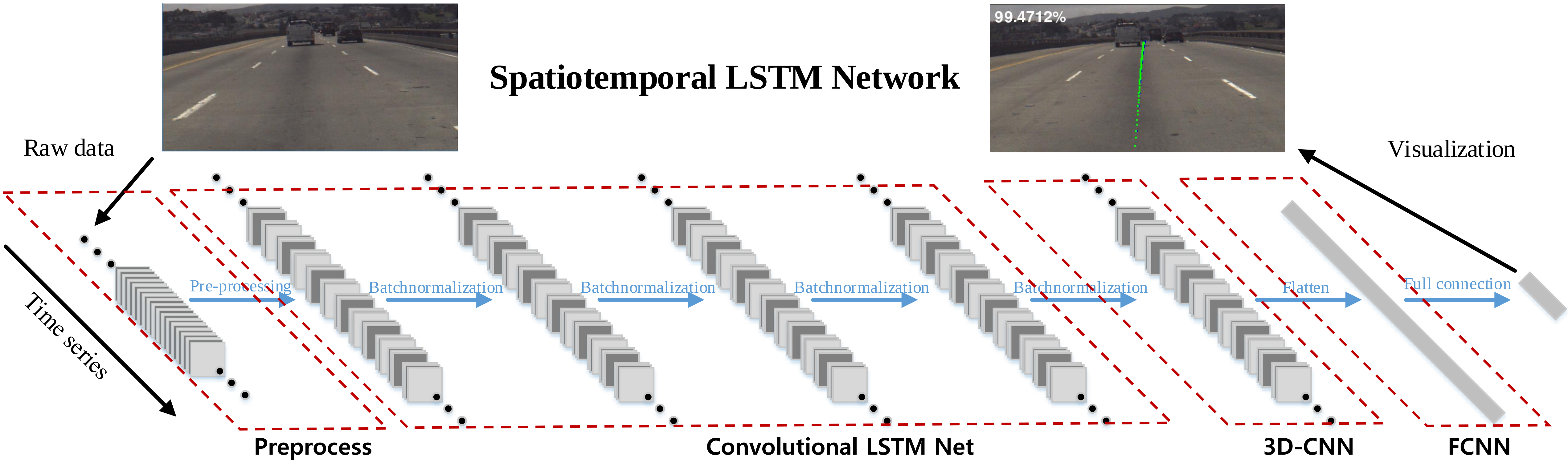


3D-CNN

Capturing the temporal and spatial features of time based image stream

Final DL Model

Beijing Jiaotong University



Activation Function

$$LeakyReLU = \begin{cases} -0.2 * x, x < 0 \\ x, x \le 0 \end{cases}$$

Loss Function

$$MSE(x,y) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} (x_i - y_i)^2$$

Learning the data from both spatial and temporal dimensions

Experiment

- Experimental platform : ubuntu16.04LTS + Tensorflow + Keras

ubuntu + TensorFlow + Keras + Inter i7-6700 CPU @ 3.4GHz
32 GB RAM
NVIDIA GTX 980 GPU

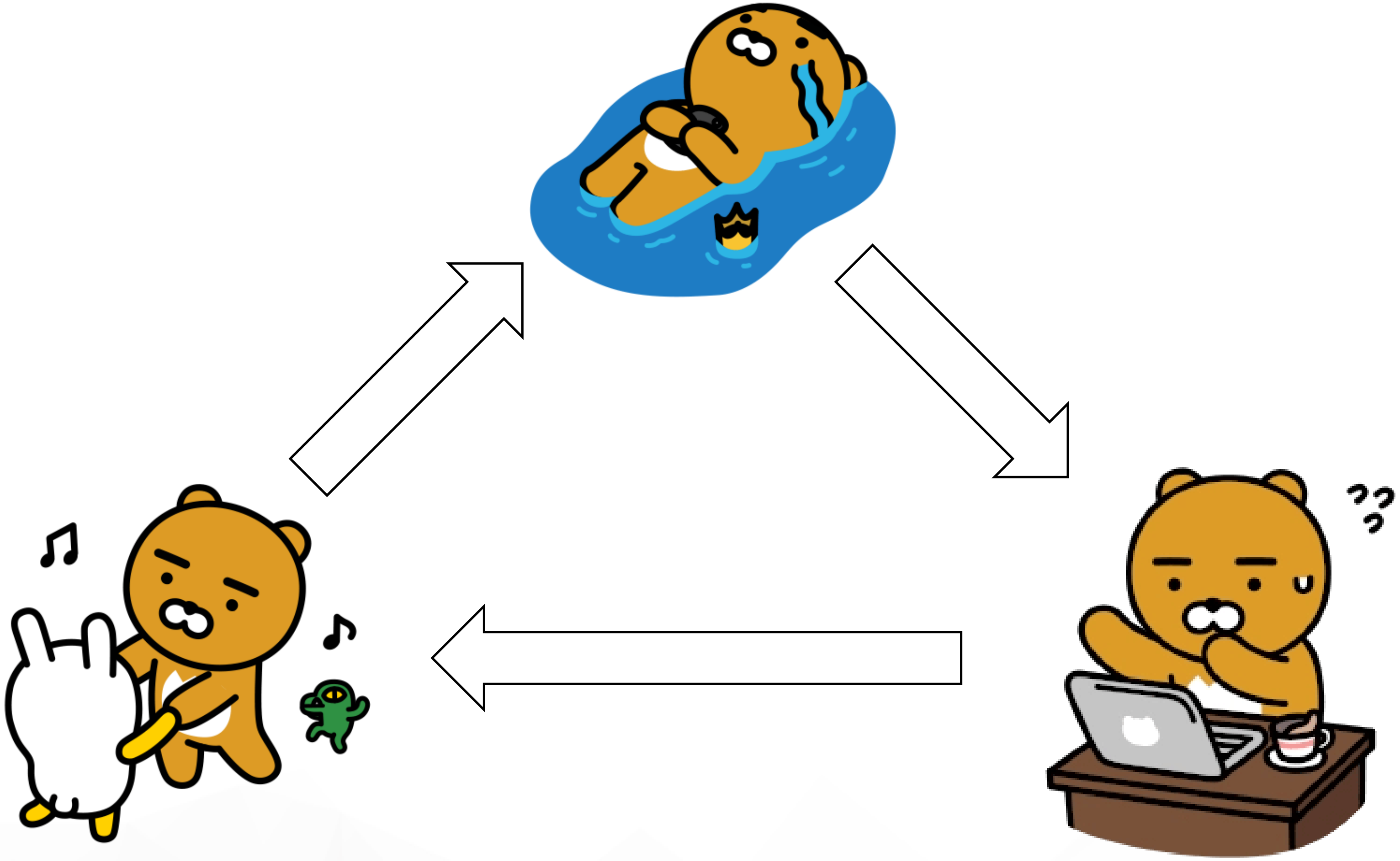
- Database : Comma-ai Dataset

comma.ai

Dataset content:

- **80GB** data
- Raw **image** data and vehicle **actual state** data
- Collected from the **vehicle-based sensors** respectively

Progress of Experiment Making

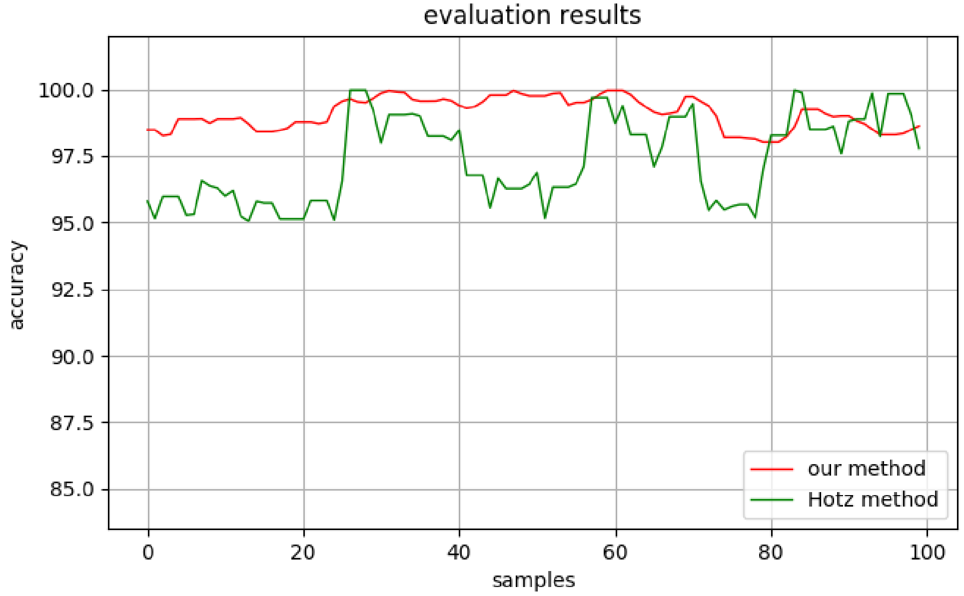


Experiment & Result

Result & Visualization

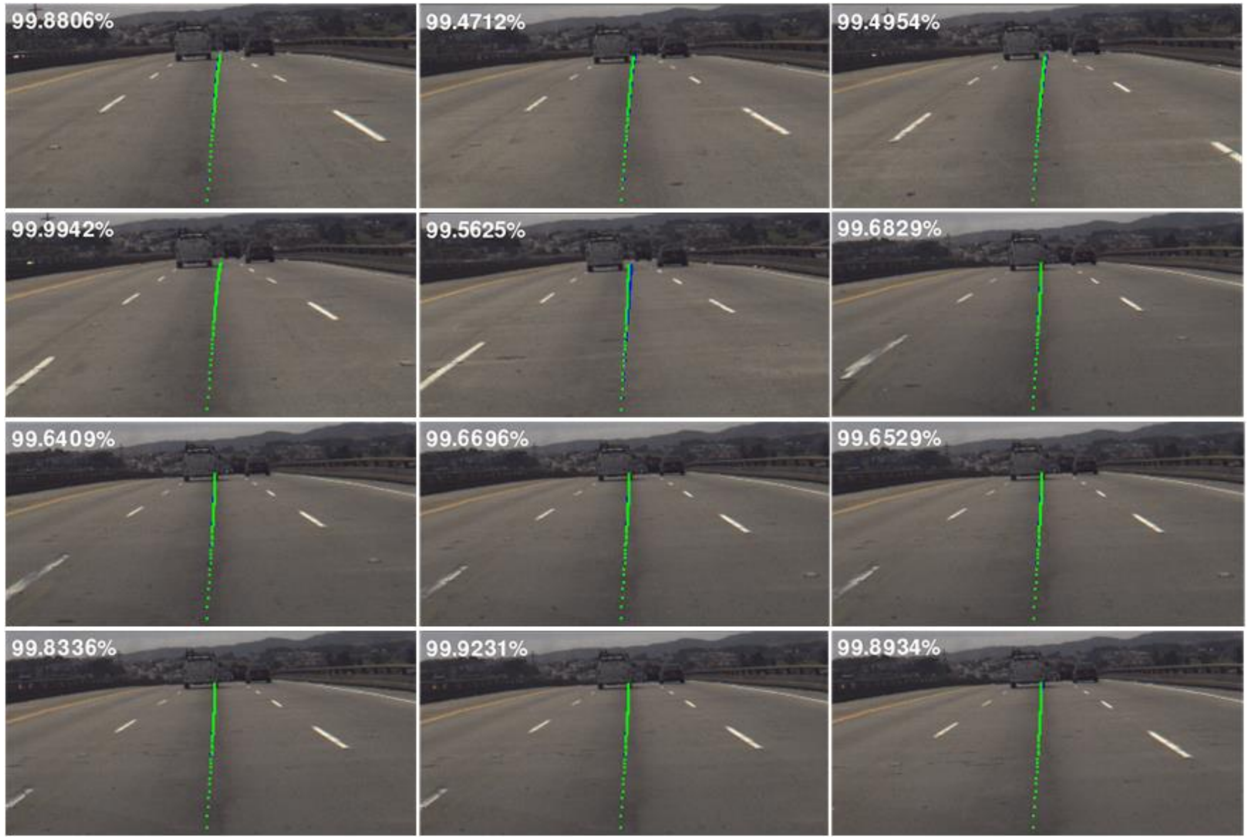
Beijing Jiaotong University

Contrast experimental results



- **Our method**: high accurate and stable
- **Hotz's method**: Can achieve the high accuracy sometime but with higher fluctuation

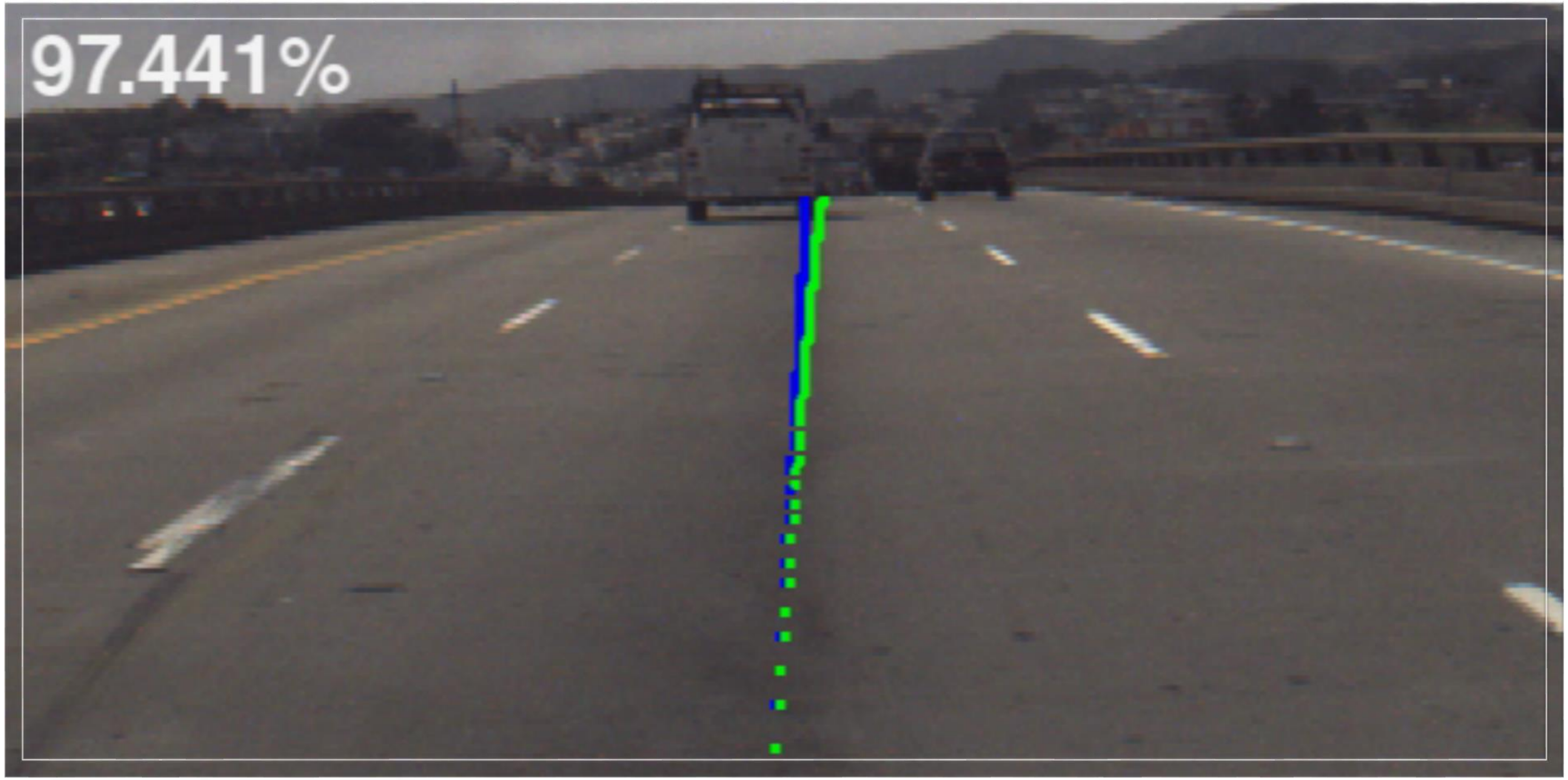
Visualization Results



Spatiotemporal LSTM Network : Get more accurate and stable results

Result Video

Beijing Jiaotong University



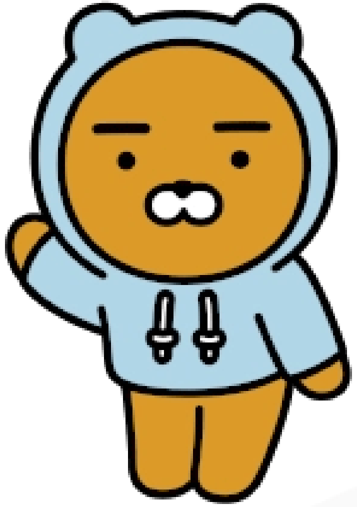
Conclusion

- It can drive vehicle!! But not smart at all scenario
- The Spatiotemporal LSTM Network has the ability to learn the data from both spatial and temporal dimensions.
  - Spatial: more precise results in motion output
  - Temporal: more stable motion in complex scenario
- Changing layers and parameters can cause much better result
  - Need more 3D-CNN layer, more FCNN layer!
  - More training time!!

After This Project...

Beijing
Jiaotong
University

- Training the algorithm with a better computer in lab!
- Read and implement more papers to reinforce my DL algorithm
- Training and testing algorithms with more dataset
- Compare the results in different sensor data
  - Lidar + Image
  - Only Image
  - Only Lidar
- Implement in a real vehicle?



北京交通大學

THANK YOU !

Zhengwei Bai

Mentor: Baigen Cai