SOON CHUN HYANG UNIVERSITY

Early Lane Change Prediction for Automated Driving Systems Using Multi-Task Attention-Based Convolutional Neural Networks

Sajjad Mozaffari, Eduardo Arnold, Mehrdad Dianati

SCH Univ.

Dept. of AI and Bigdata

Seokjun Hong

Drone Vision Traffic Prediction

SOON CHUN HYANG UNIVERSITY

Contents

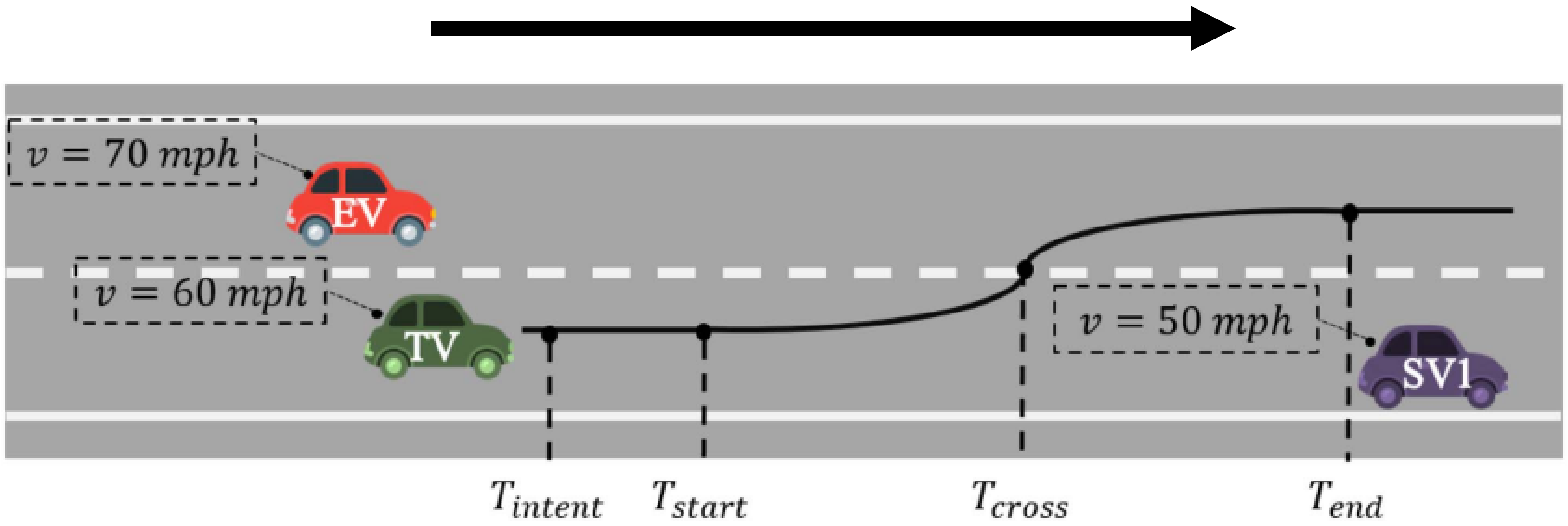
1. Introduction
2. Related Works
3. System Model and Problem Definition
4. Proposed Method
5. Performance Evaluation
6. Conclusion
7. How To Apply

1. Introduction

1. Introduction

- Major Crash → Accidents associated with unsafe LC maneuvers

- Early warning → Alleviate the risk of accidents



Target Vehicle Lane Change Scenario

- EV decelerate as soon as it realises the imminent LC manoeuvre by the TV

- T_{start} ~ T_{end} : 3 ~ 5 seconds

1. Introduction

- Existing studies predicts LC after max 2.5 seconds

- T_{start} ~ T_{end} : 3 ~ 5 seconds

- Predicting LC after already started maneuvering

- Understand traffic context around TV's for long-term predictions

1. Introduction

- Multi-task attention-based prediction model

- Novel CNN using bird's eye view

- Attention model

- Multi-Task Learning (MTL) approach

- Curriculum Learning

2. Related Works

2. Related Works

▸ A. Input Representation

1) TV's States

- TV's lateral position in the lane
- Lateral and longitudinal velocity and Acceleration

2) Environment States

- Relative distance to surrounding vehicles
- Relative velocity
- Distance to the nearest on-or off-ramp
- Existence of the lanes

3) Driver's States

- Head position
- Gaze movement

LSTM

HMMs

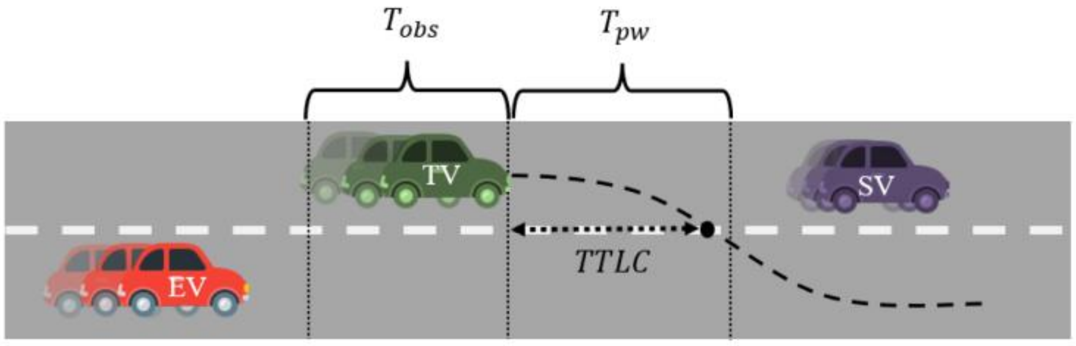
DNNs

Dynamic Bayesian Network

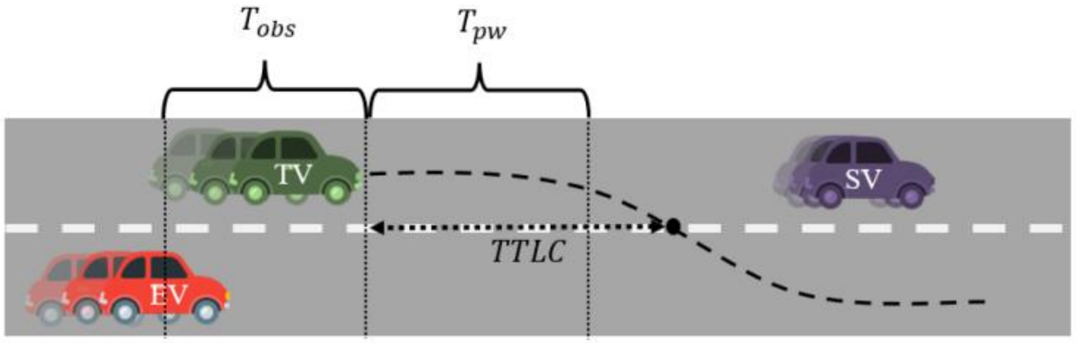
⋮

3. System Model and Problem Definition

3. System Model and Problem Definition



(a) LC Scenario ($0 < TTLC < T_{pw}$)



(b) LK Scenario ($TTLC > T_{pw}$)

Classification Problem

- Estimate the probability of LC
- During prediction window, T_{pw}
- LLC, RLC, LK
- T_{pw} : Maximum prediction time
- T_{pw} = 5.2 seconds

3. System Model and Problem Definition

Regression Problem

- Estimate the Time to Lane Change (TTLC)
- TTLC : Shortest time utils the center of the TV

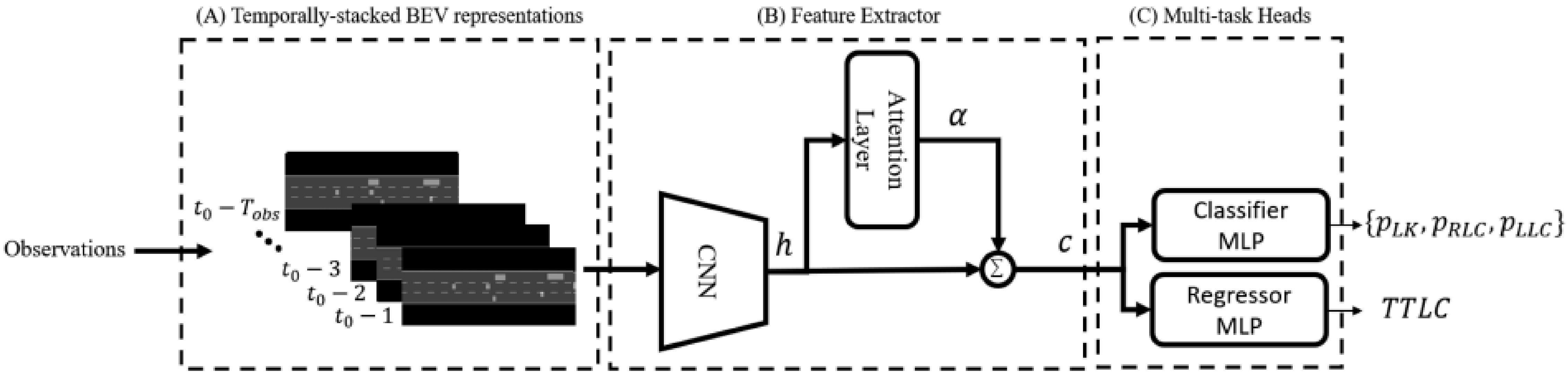
crosses either left of right lane marking

Input Data

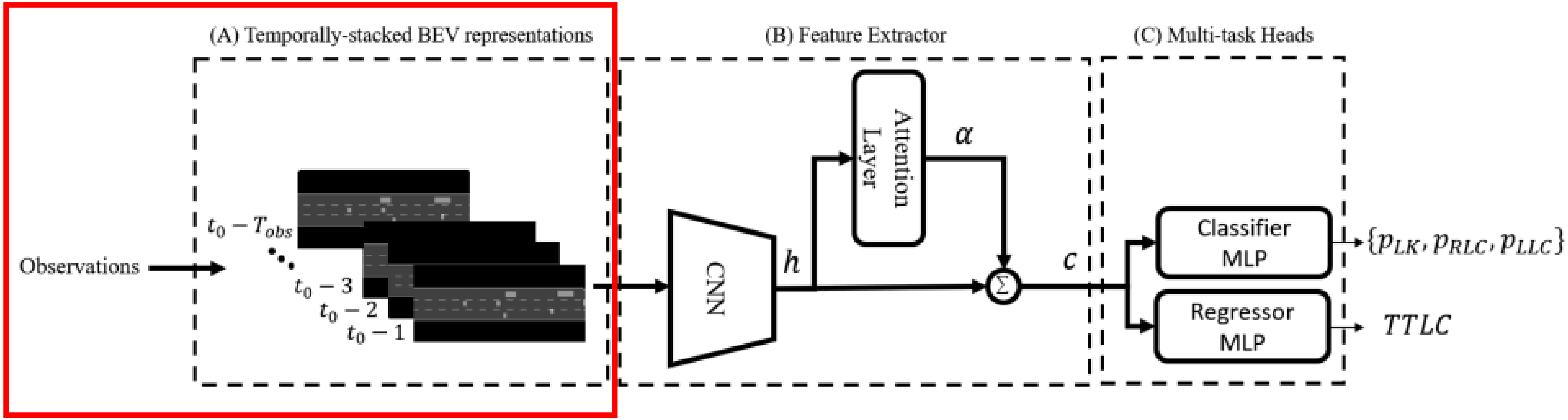
- TV and SV state
- Position of the lane marking
- During observation window of T_{obs}
- T_{obs} = 2 seconds

4. Proposed Method

4. Proposed Method

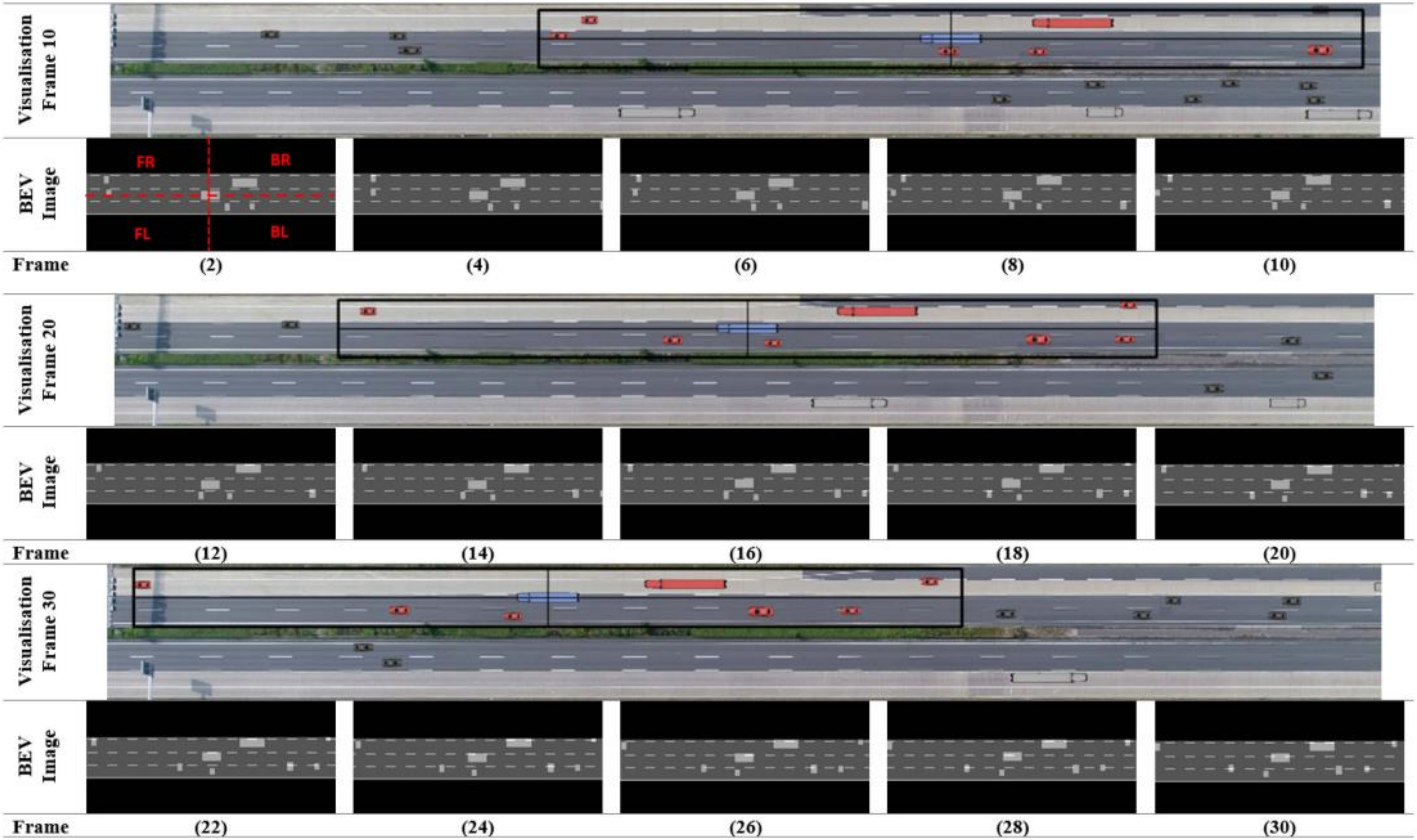


4. Proposed Method



4. Proposed Method

▸ A. BEV Input Data Representation



▸ A. BEV Input Data Representation

1)

- Center the BEV representation on the TV at each time step

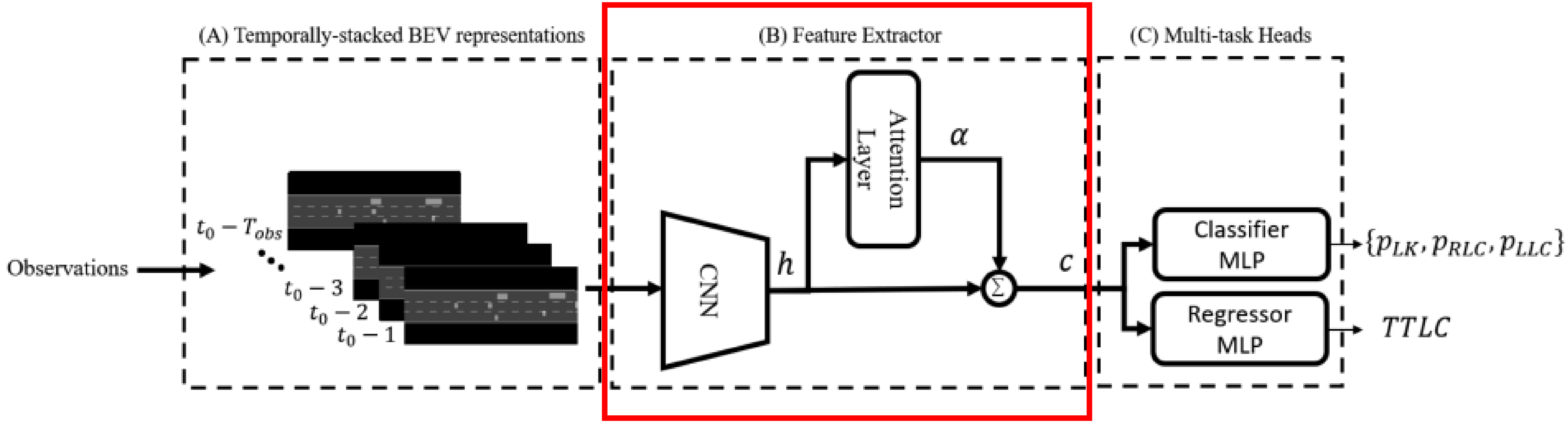
2)

- Lateral dimension resolution four time higher than the longitudinal dimension

- Size of the BEV representation 200 by 80 pixels

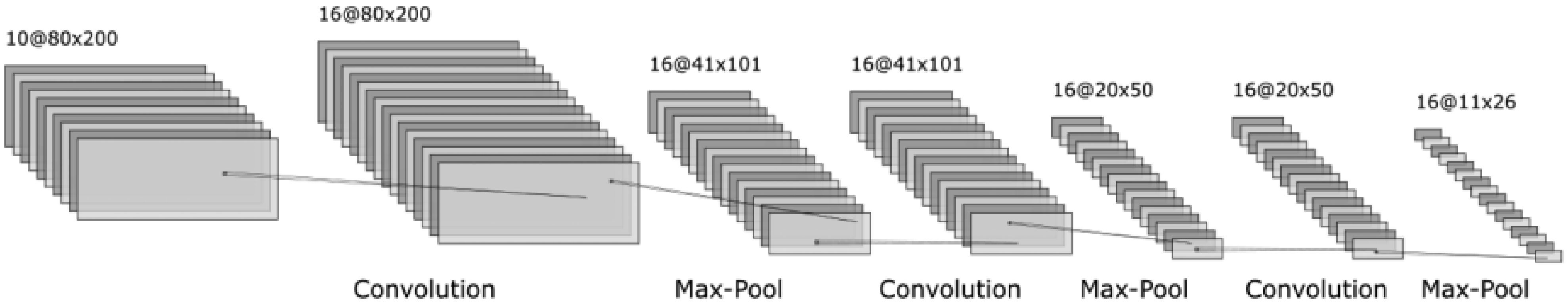
- Input to the CNN is a multi-channel image of size (T_{obs} X FPS) X 200 X 80

4. Proposed Method



4. Proposed Method

▸ B. Attention-Based CNN for Feature Learning



Attention-Based CNN

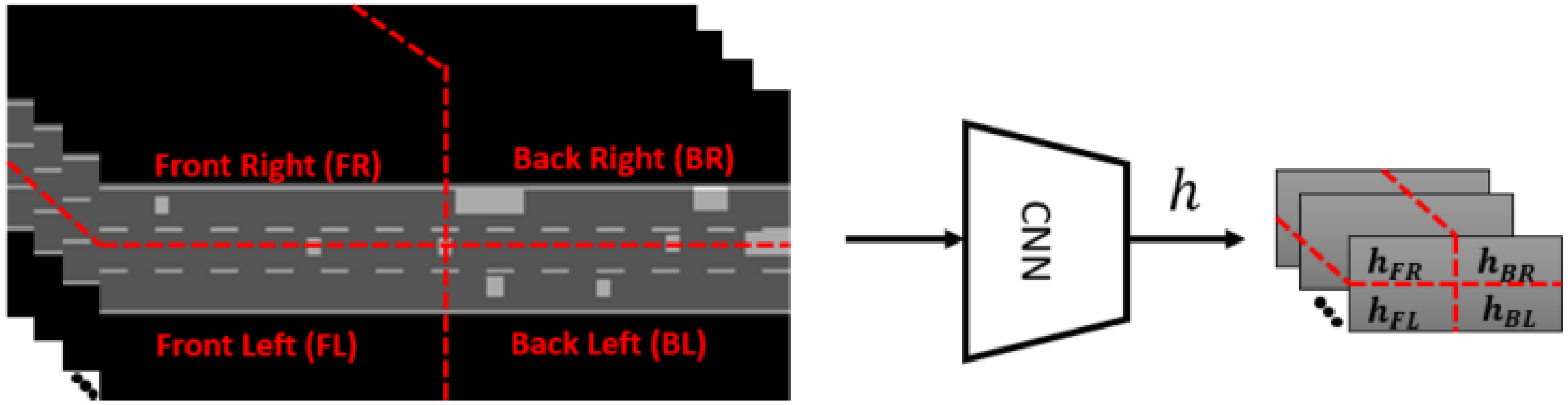
- Extract relevant spatiotemporal features from the temporally stacked BEV representation

Attention model

- Focuses on important parts of the input data
- Processes them based on their importance
- Pays more attention to especially important information among the entire set of data
- Identifying the key elements necessary for solving a problem

4. Proposed Method

▸ B. Attention-Based CNN for Feature Learning



Identify and focus on parts of the environment around TV

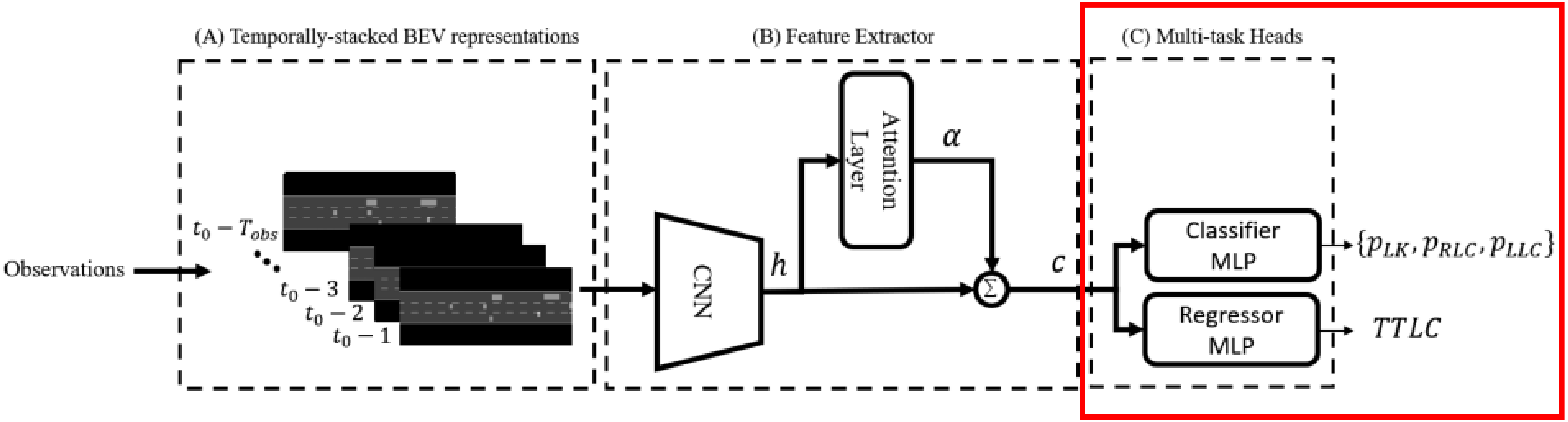
↪ Most impact on the future behavior of the TV

RLC : Slow-moving vehicle in front & Suitable gap in the right lane

Behaviour of SV's driving on the left lane dose not influence RLC decision

Focusing on the relevant areas → Expected to increased the performance of LC prediction

4. Proposed Method



4. Proposed Method

▸ C. Multi-Task Learning (MTL)

- LC manoeuvre classification + TTLC regression tasks

- TTLC regression is more difficult than three-class classification of future LC manoeuvres

- Feature Learnt by the classifier → Enhanced regressor performance

< Classifier>

Activation	ReLU
Dropout	0.5
Hidden	128
Output	3

< Regressor >

Activation	ReLU
Dropout	0.5
Hidden	512
Output	1

$$L_{CE} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{c=1}^{3} y_{i,c}\log \hat{y_{i,c}}$$

$$L_{MSE} = \frac{1}{n}\sum_{i=1}^{n}(x^i - \hat{x^i})^2$$

$$L = L_{CE} + \gamma L_{MSE}$$ γ : ratio between the regressor and classifier losses

4. Proposed Method

▸ D. Curriculum Learning (CL)

$$L = L_{CE} + \boxed{\gamma} L_{MSE}$$

	Initial Training Epochs						Remaining Training Epochs
	0	**1**	**2**	**3**	**4**	**5**	**>5**
Max included TTLC	0.2	1.2	2.2	3.2	4.2	5.2	5.2
Loss Ratio	0	0.2	0.4	0.6	0.8	1	1

Max included TTLC determines the maximum TTLC of included data samples in a epoch.

CL based on two criteria specific to the LC prediction problem

Case1.

- Small TTLC sample → Easier prediction (more explicit predictors found in the TV's motion)
- Start training with sample with near-zero TTLC
- Gradually expose samples with larger TTLC to the prediction model

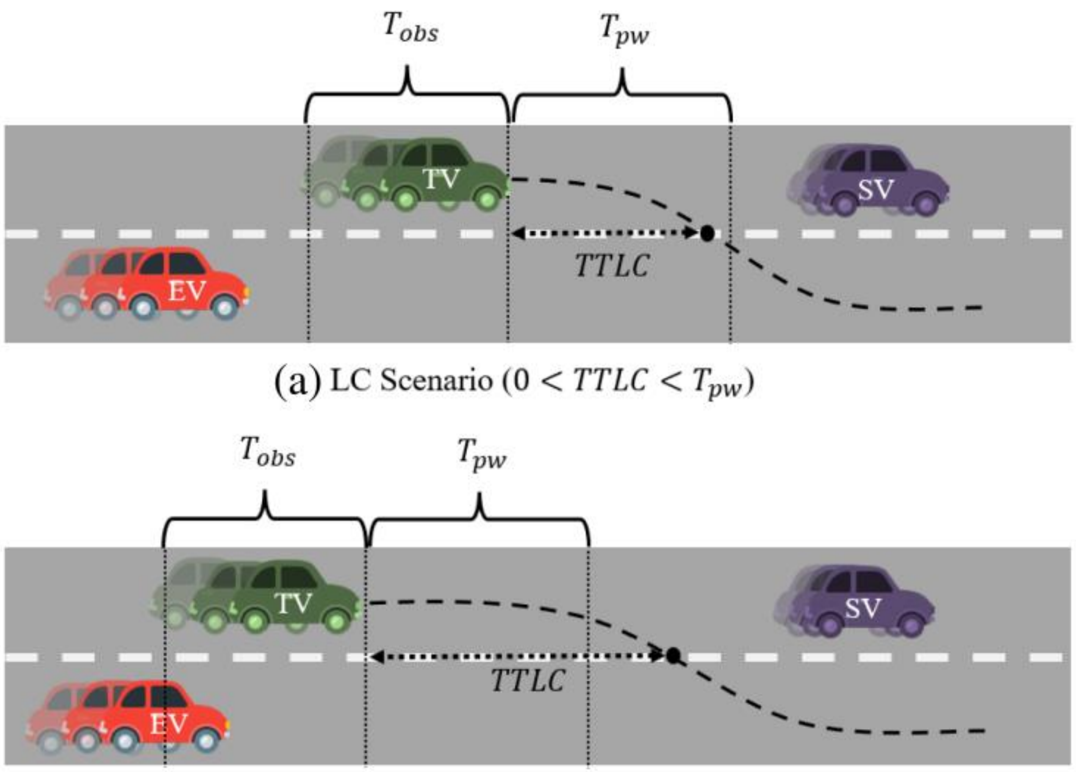
Case2.

- Three-class classification task is normally considered easier than regression a continuous variable
- Start training process by giving more importance to the classification task
- Gradually shift the focus to the regression task
- During training phase loss ratio γ 0 → 1

5. Performance Evaluation

5. Performance Evaluation

▸ A. Dataset and LC Scenario Extraction



(a) LC Scenario ($0 < TTLC < T_{pw}$)

(b) LK Scenario ($TTLC > T_{pw}$)

Highway Drone Dataset (highD)

- 110,500 Vehicle
- 420m at six different locations
- [t_0 – T_{obs}, t_0 - 1]
- RLC, LLC, LK
- Under sample the LK class
- Train : 7487 / Validation : 932 / Test : 698

< LC / LK scenarios train >

Optimizer	Adam
Epoch	20
Learning rate	0.001
Batch size	64

1) Classification Metrics

- Accuracy
- Precision, F1 score, Recall
- ROC, AUC
- First prediction time : τf

↪ first correct prediction, T_{cross}
How quickly it makes correct predictions

- Robust prediction time : τc

↪ Continuously predict correctly, T_{cross}
How long it maintains consistent accurate predictions

2) Regression Metrics

- MSE
- RMSE
- Box plot

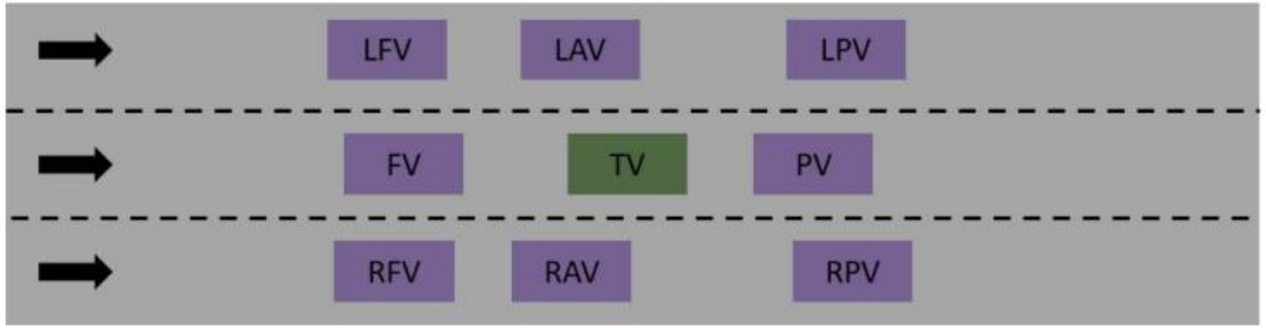
5. Performance Evaluation

▸ D. Quantitative Results

Baseline	List of Features
MLP1 [36]	(1) Existence of left lane, (2) Existence of right lane, (3) Lane width, (4) Longitudinal distance of TV to PV, (5) Longitudinal distance of TV to RPV, (6) Longitudinal distance of TV to FV, (7) Lateral distance of TV to the left lane marking, (8) Lateral distance of TV to RV, (9) Lateral distance of TV to RFV, (10) Relative longitudinal velocity of TV w.r.t. PV, (11) relative longitudinal velocity of TV w.r.t. FV, (12) Relative lateral velocity of TV w.r.t. PV, (13) Relative lateral velocity of TV w.r.t. RPV, (14) Relative lateral velocity of TV w.r.t. RV, (15) Relative lateral velocity of TV w.r.t. LV, (16) Longitudinal acceleration of the TV, (17)Relative longitudinal acceleration of the TV w.r.t RPV, (18) Lateral acceleration of the prediction target
MLP2 [29]	(1) Existence of left lane, (2) Existence of right lane, (3) Longitudinal distance of TV to RPV, (4) Longitudinal distance of TV to PV, (5) Longitudinal distance of TV to LPV, (6) Longitudinal distance of TV to RV, (7) Longitudinal distance of TV to LV, (8) Longitudinal distance of TV to RFV, (9) Longitudinal distance of TV to FV, (10) Longitudinal distance of TV to LFV, (11) Relative velocity of TV w.r.t. RPV, (12) Relative velocity of TV w.r.t. PV, (13) Relative velocity of TV w.r.t. LPV, (14) Relative velocity of TV w.r.t. RV, (15) Relative velocity of TV w.r.t. LV, (16) Relative velocity of TV w.r.t. RFV, (17) Relative velocity of TV w.r.t. FV, (18) Relative velocity of TV w.r.t. LFV
LSTM1 [8]	Same as MLP1 [36]
LSTM2	(1) Lateral velocity, (2) Longitudinal velocity, (3) Lateral acceleration, (4) Longitudinal acceleration, (5) Lateral distance of TV to the left lane marking, (6) Relative longitudinal velocity of the TV w.r.t. PV, (7) Longitudinal distance of TV to PV, (8) Relative longitudinal velocity of the TV w.r.t. FV, (9) Longitudinal distance of TV to FV, (10) Longitudinal distance of TV to RPV, (11) Longitudinal distance of TV to RV, (12) Longitudinal distance of TV to RFV, (13) Longitudinal distance of TV to LPV, (14) Longitudinal distance of TV to LV, (15) Longitudinal distance of TV to LFV, (16) Existence of left lane, (17) Existence of right lane, (18) Lane width

\+ CS – LSTM Proposed

w.r.t : “with respect to”



5. Performance Evaluation

▸ D. Quantitative Results

Task	Model	Accuracy	Recall	Precision	F1-score	AUC	τ_f	τ_c	RMSE
Classification	MLP1 [36]	0.75	0.65	**0.94**	0.77	0.84	3.97	2.73	-
	MLP2 [29]	0.59	0.52	0.74	0.61	0.61	3.49	2.02	-
	LSTM1 [8]	0.79	**0.90**	0.75	0.82	0.86	4.24	2.98	-
	LSTM2	0.78	0.84	0.81	0.82	0.84	4.43	3.76	-
	CS-LSTM [31]	0.74	0.78	0.81	0.72	0.76	3.92	3.61	-
Regression	LSTM1 [8]	-	-	-	-	-	-	-	0.841
	LSTM2	-	-	-	-	-	-	-	0.976
Dual	Proposed	**0.83**	0.85	0.85	**0.85**	**0.88**	**4.75**	**3.96**	**0.629**

Accuracy 4% ↑

τf 0.32 ↑

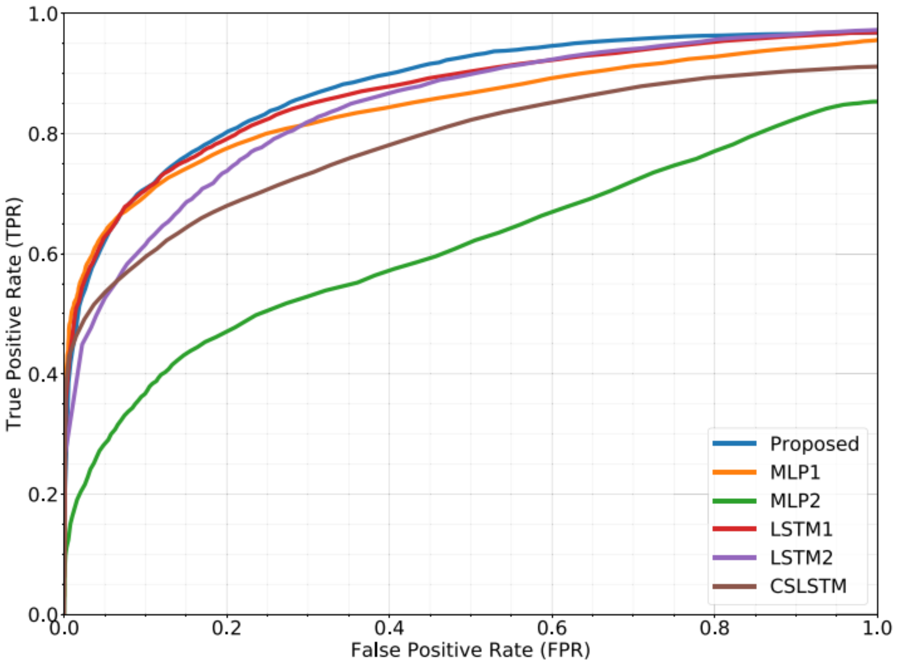
F1-score 3% ↑

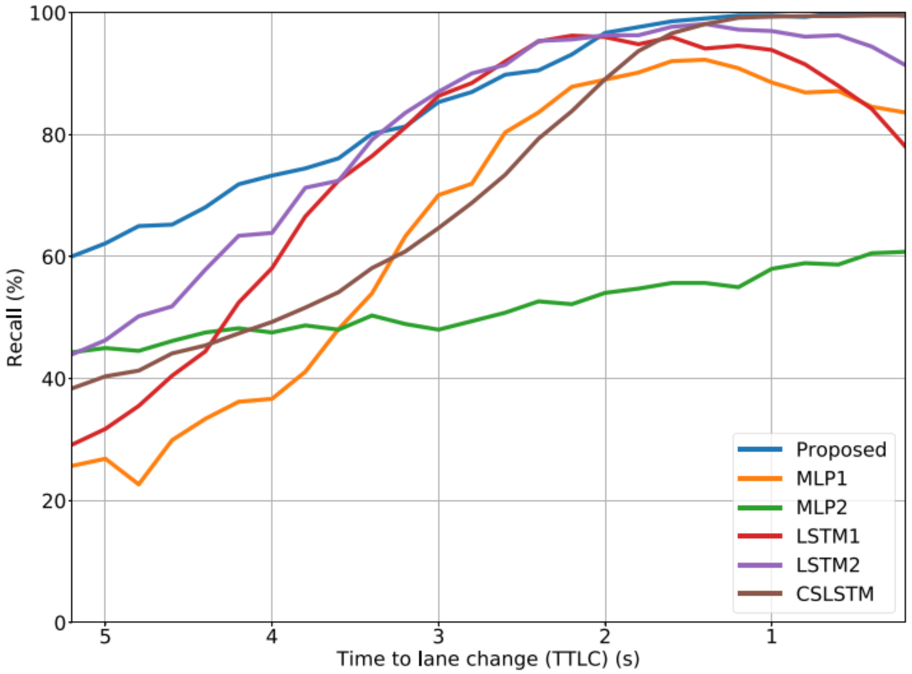
τc 0.2 ↑

RMSE 0.2 ↓

5. Performance Evaluation

▸ D. Quantitative Results





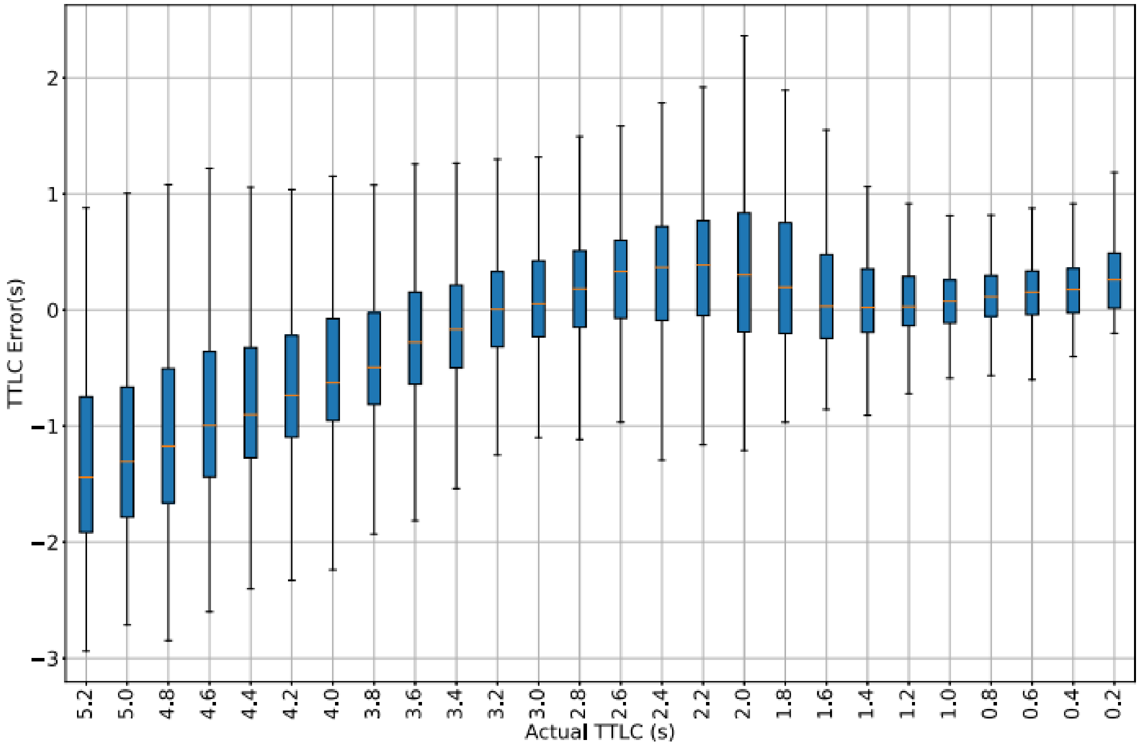
- TTLC 1.5 seconds ↓ close 100% recall

- TTLC 5.2 seconds, 60% recall

- MLP2 only longitudinal features

5. Performance Evaluation

▸ D. Quantitative Results



- Higher TTLC → higher median error and variance

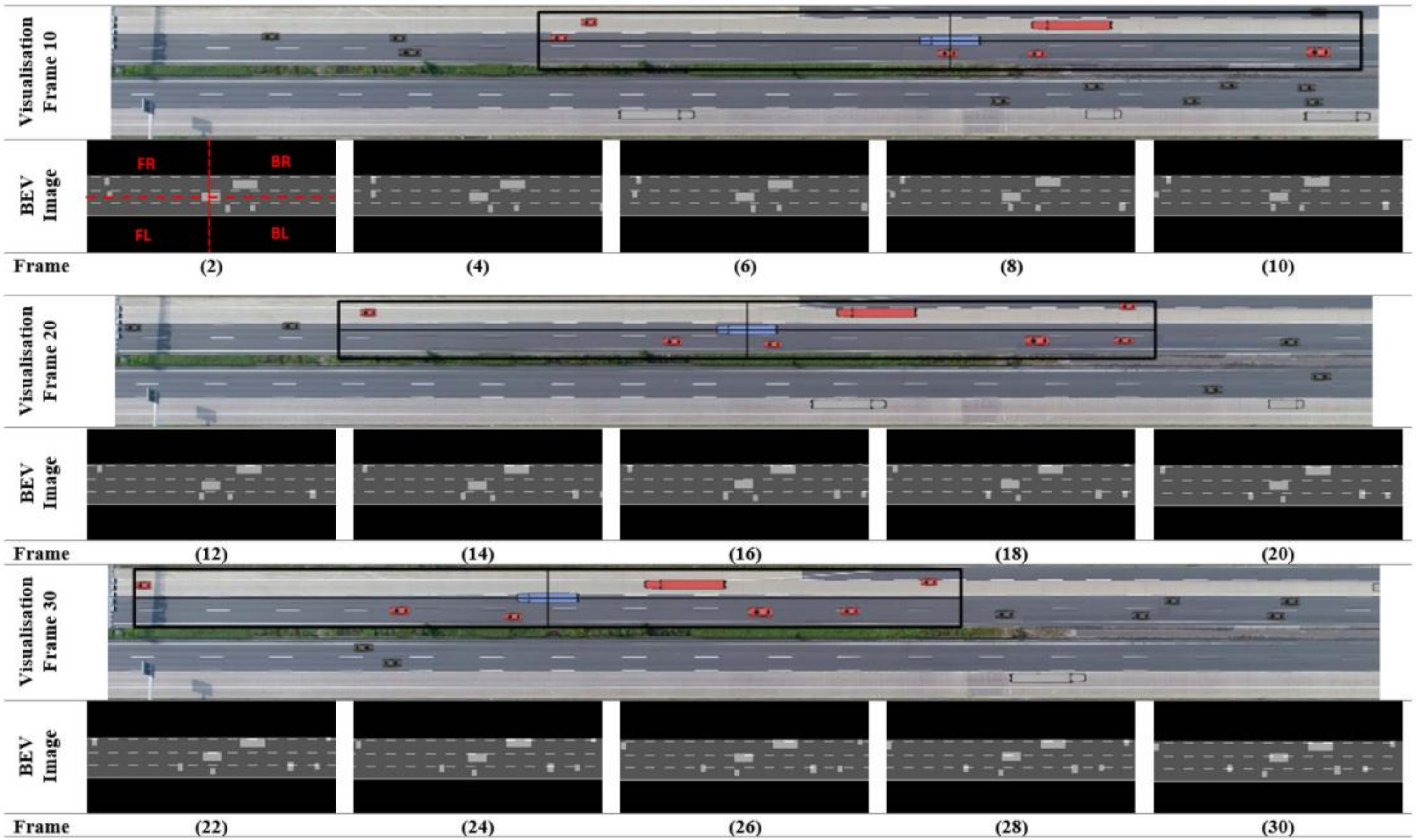
- 3.2 seconds ↑, model tend to predict TTLC less than actual TTLC

- 3 seconds ↑ TTLC sample, do not exhibit any explicit change in lateral movement of the TV

↪ It is not possible to estimate LC time using lateral speed and distance

5. Performance Evaluation

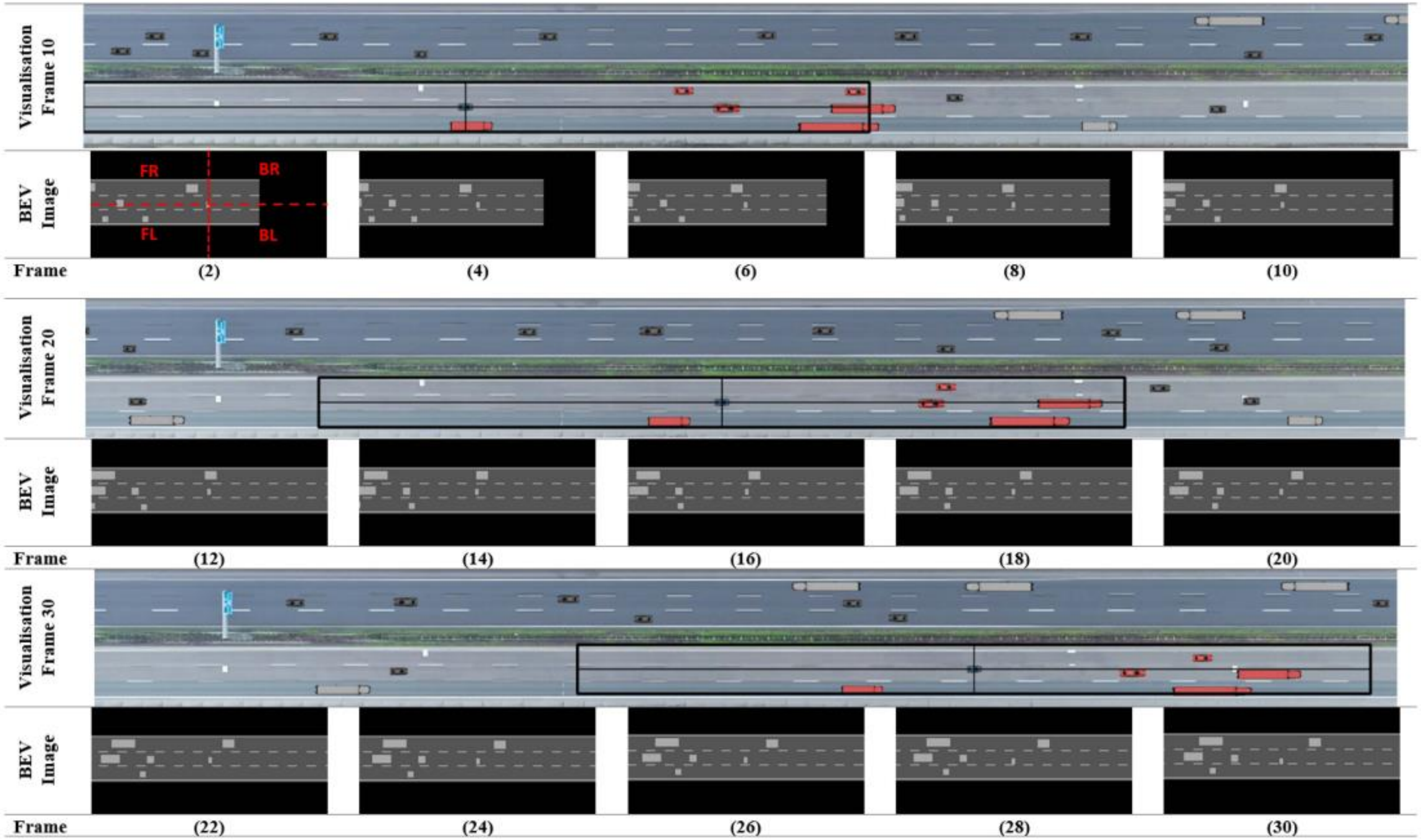
▸ E. Qualitative Results



Frame	Grount Truth TTLC	Predicted TTLC	P(m=LK)	P(m=RLC)	P(m=LLC)	α_{FR}	α_{FL}	α_{BR}	α_{BL}
10	5.2	-	0.5	0.48	0.02	0.32	0.1	0.34	0.23
20	3.2	3.05	0.15	0.85	0	0.2	0.14	0.46	0.2
30	1.2	1.12	0	1	0	0.02	0.07	0.88	0.03

5. Performance Evaluation

▸ E. Qualitative Results



Frame	Grount Truth TTLC	Predicted TTLC	P(m=LK)	P(m=RLC)	P(m=LLC)	α_{FR}	α_{FL}	α_{BR}	α_{BL}
10	5.2	4.87	0.03	0	0.97	0.07	0.73	0.06	0.14
20	3.2	3.3	0	0	1	0.02	0.62	0.04	0.32
30	1.2	0.98	0	0	1	0	0.94	0.05	0.01

5. Performance Evaluation

▸ F. Ablation Study

Task	Attention	CL (Loss)	CL (TTLC)	AUC%	RMSE (s)
C*				88.38	-
R**				-	0.804
MTL				83.11	0.805
MTL	✓			86.89	0.796
MTL	✓		✓	87.79	0.809
MTL	✓	✓		87.97	**0.774**
MTL	✓	✓	✓	**89.43**	**0.774**

[0]* C: Classification, ** R: Regression

6. Conclusion

6. Conclusion

Attention-based CNN

- Extracting interaction-aware features from the surrounding traffic required for long-term prediction

Multi-task approach

- Boosted by two novel curriculum learning

- TTLC and manoeuvre likelihood prediction using shared extracted features

Purposed model

- Outperforms SOTA LC prediction

- 1.5 times better long-term prediction performances

7. How To Apply

▸ A. Difference

CNN		YOLOv8
LC maneuvers, TTLC	⟷	LC O,X
MLP		Transformer

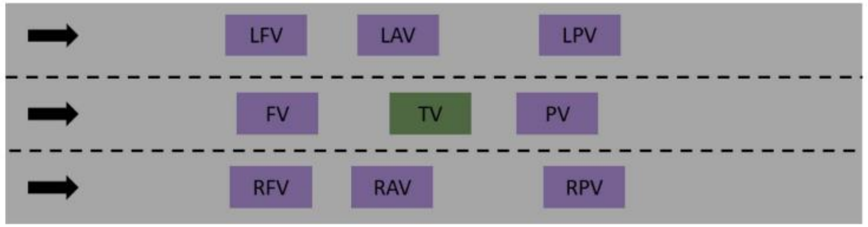
7. How To Apply

▸ B. Feature Select

1) Existence of left lane
2) Existence of right lane
3) Lane width
4) Longitudinal distance of TV to PV
5) Lateral distance of TV to RPV
6) Longitudinal distance of TV to FV
7) Later distance of TV to the left lane marking
8) Lateral distance of TV to RV
9) Lateral distance of TV to RFV
10) Relative longitudinal velocity of TV w.r.t PV
11) Relative longitudinal velocity of TV w.r.t FV
12) Relative lateral velocity of TV w.r.t PV
13) Relative lateral velocity of TV w.r.t RPV
14) Relative lateral velocity of TV w.r.t RV
15) Relative lateral velocity of TV w.r.t LV
16) Longitudinal acceleration of the TV
17) Relative longitudinal acceleration of the TV w.r.t RPV
18) Lateral acceleration of the prediction target

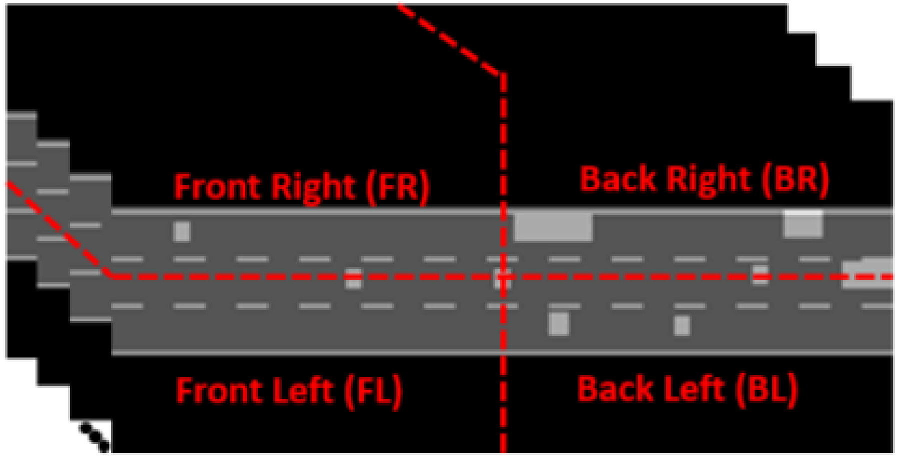
+

Lateral velocity
Longitudinal velocity
Lateral acceleration
Longitudinal acceleration
⋮



7. How To Apply

▸ C. Surround Range



Highway ↔ Local road

200 by 80 pixels → ?

Thank You

☺