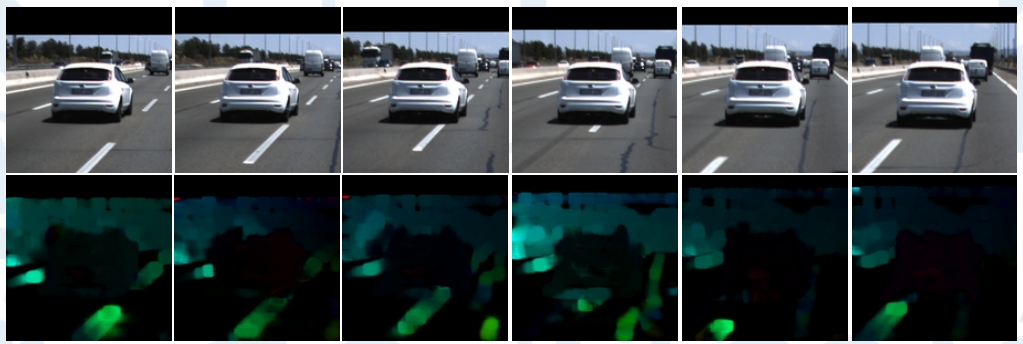
2020 IEEE 23rd International Conference on Intelligent Transportation Systems
ITSC 2020

Two-Stream Networks for Lane-Change Prediction of Surrounding Vehicles



D. F. Llorca[1], M. Biparva[2], R. Izquierdo[1], J. K. Tsotsos[2]

[1]University of Alcalá (Spain)

[2]York University (Canada)

20 September 2020

Index

Index

Application scenario
Motivation

Autonomous navigation on highways in the short term

- Highway Chauffeur (SAE L3): drivers' attention required; limited functionality.
- Highway Autopilot (SAE L4): drivers' attention not required; enhanced functionality.

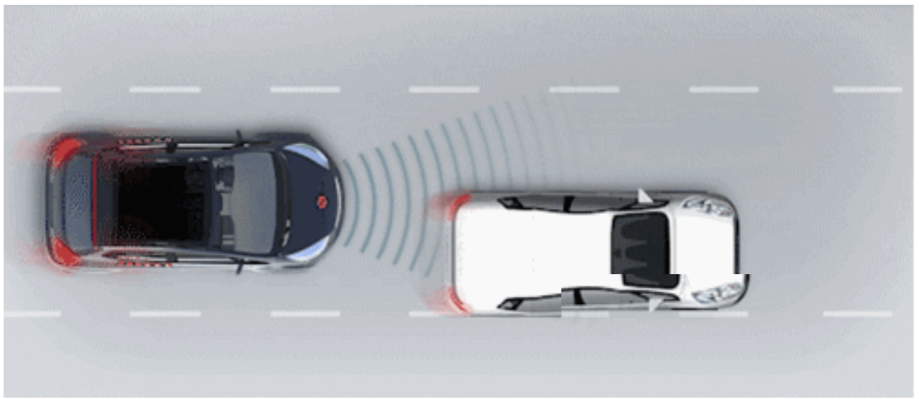
Highway Chauffeur
(SAE L3)

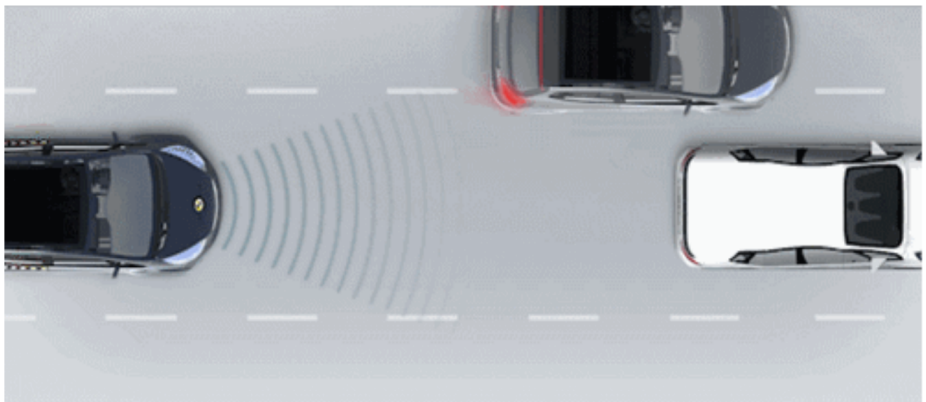
Highway Autopilot
(SAE L4)

Application scenario
Motivation

Most critical and dangerous maneuvers

- Lane-changes: cut-in/cut-out.
- Euro NCAP: since 2018, testing these two scenarios.





Predicting a lane-change

- Lane changes (whether abrupt or not) can be a prelude to a dangerous situation.
- Human drivers are capable of predicting them, and they usually reduce the speed to increase the safety.
- **Goal**: can we design a system to anticipate lane-changes N seconds before they occur?

(Tesla accident April 2018) cut-out maneuver front vehicle not anticipated.
Perception systems did not recognize in path obstacles

Previous works
Motivation

Input variables

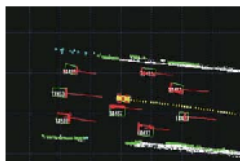
- **Target vehicle dynamics**: lateral and longitudinal distances, velocity, acceleration, time-gap, heading angle and yaw rate (from forward-facing sensors or V2V communications).
- **Context cues**: curvature, speed-limits, number of lanes, distance to the next highway junction, lane markings, distance to the lane end.
- **Visual features**: turn and brake indicators (very few proposals).

Methodologies

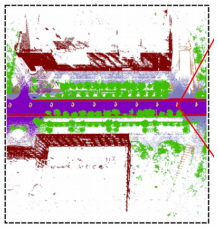
- **Levels**: physical-based, maneuver-based and intention-aware.
- **Approaches**: Bayesian Nets, Structural Recurrent NN, Hidden Markov Models, Random Decision Forest, Support Vector Machines, feedforward CNN, vanilla LSTM, LSTM encoder-decoder, among others.

Datasets

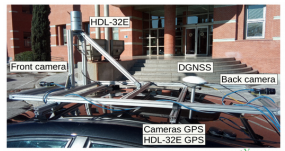
- **Infrastructure- or drone-based**: NGSIM, HighD and INTERACTION.
- **Vehicle-based**: PKU, ApolloScape, and **PREVENTION**.

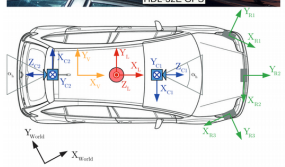
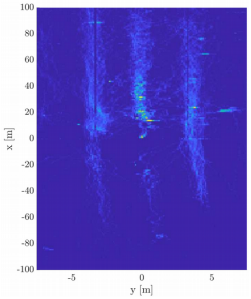



PKU Dataset



Apolloscape Dataset

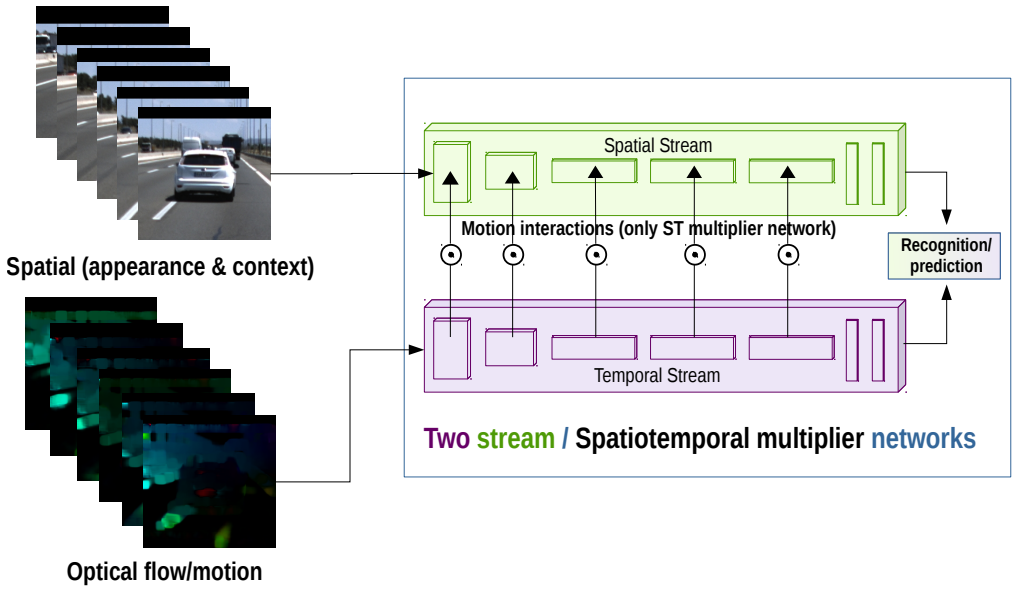


PREVENTION Dataset

Our approach
Motivation

- Rationale: to use the same source of information (visual cues) and the same type of approach (action recognition) that drivers use to anticipate lane changes.
- To apply vision-based (human) action recognition approaches to deal with lane change detection and prediction: Two-Stream Convolutional Networks and Spatiotemporal Multiplier Networks.
- We use spatial and temporal information (sequence of images).



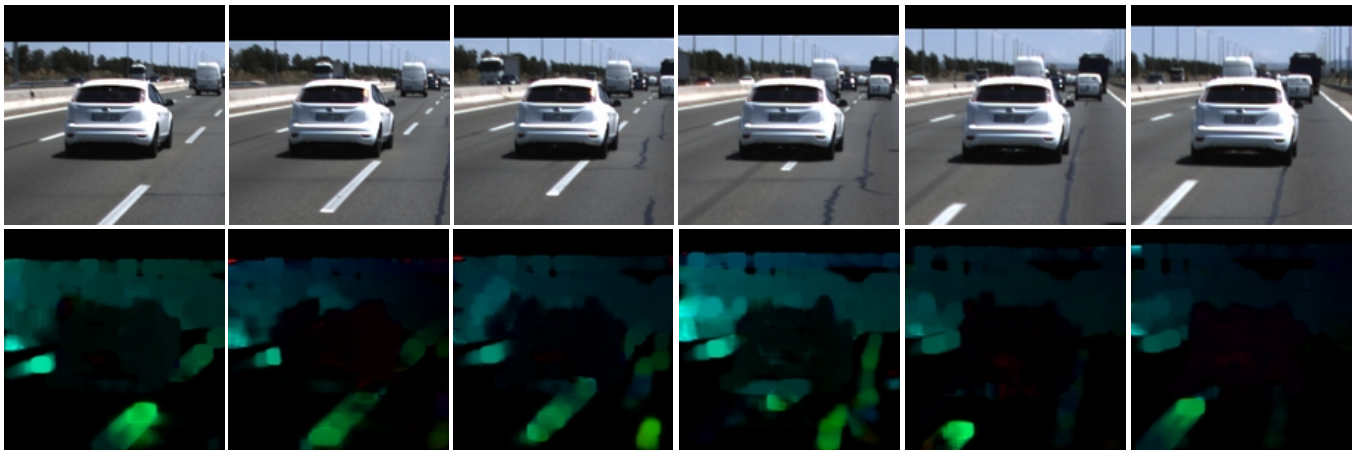
Index

Visual or spatial features

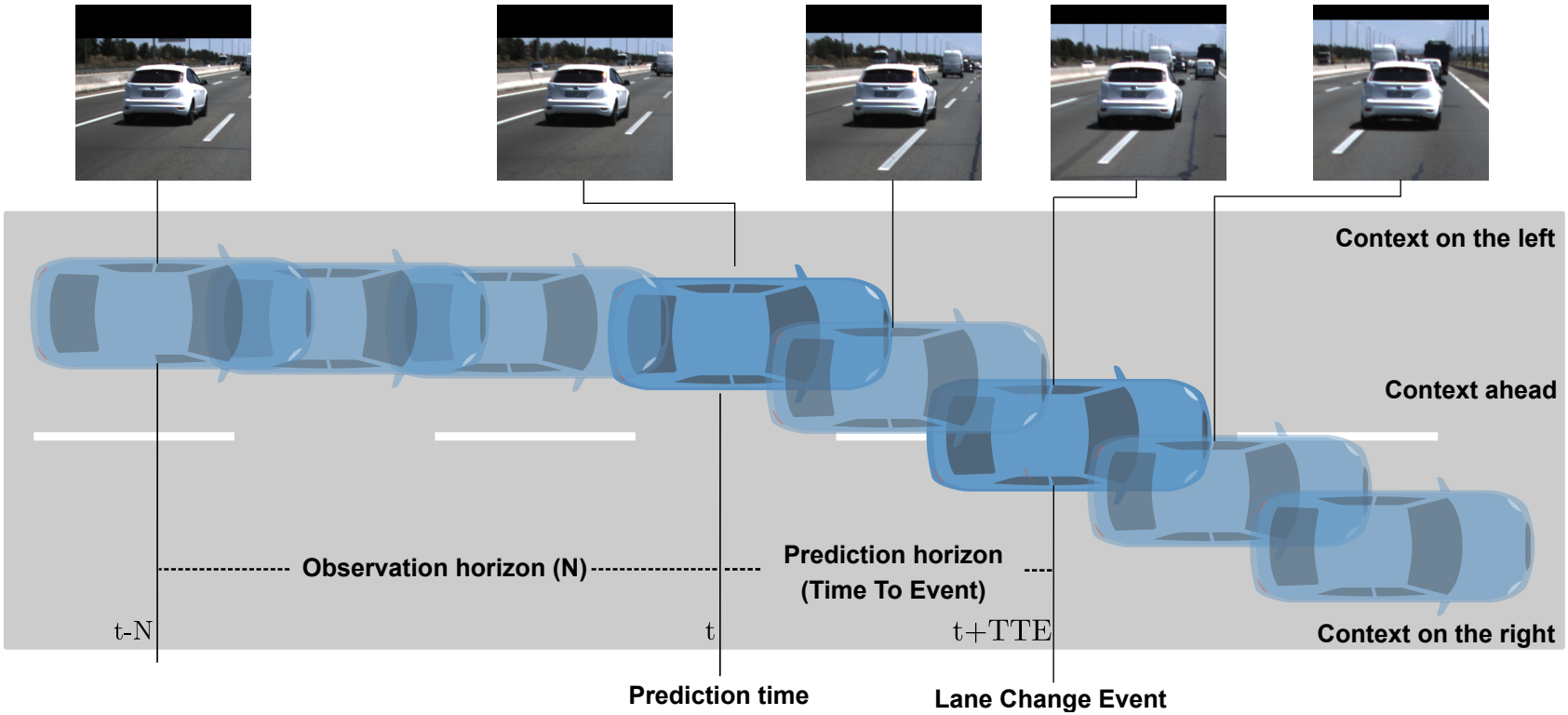
- Regions of interest (ROIs) extracted from the contour labels (from PREVENTION dataset).
- Four different ROI sizes are studied: x1, x2, x3 and x4 the size of the square bounding box around the vehicle contour.
- Zero-padding when ROI exceeds the limits of the image.

Motion (temporal) features

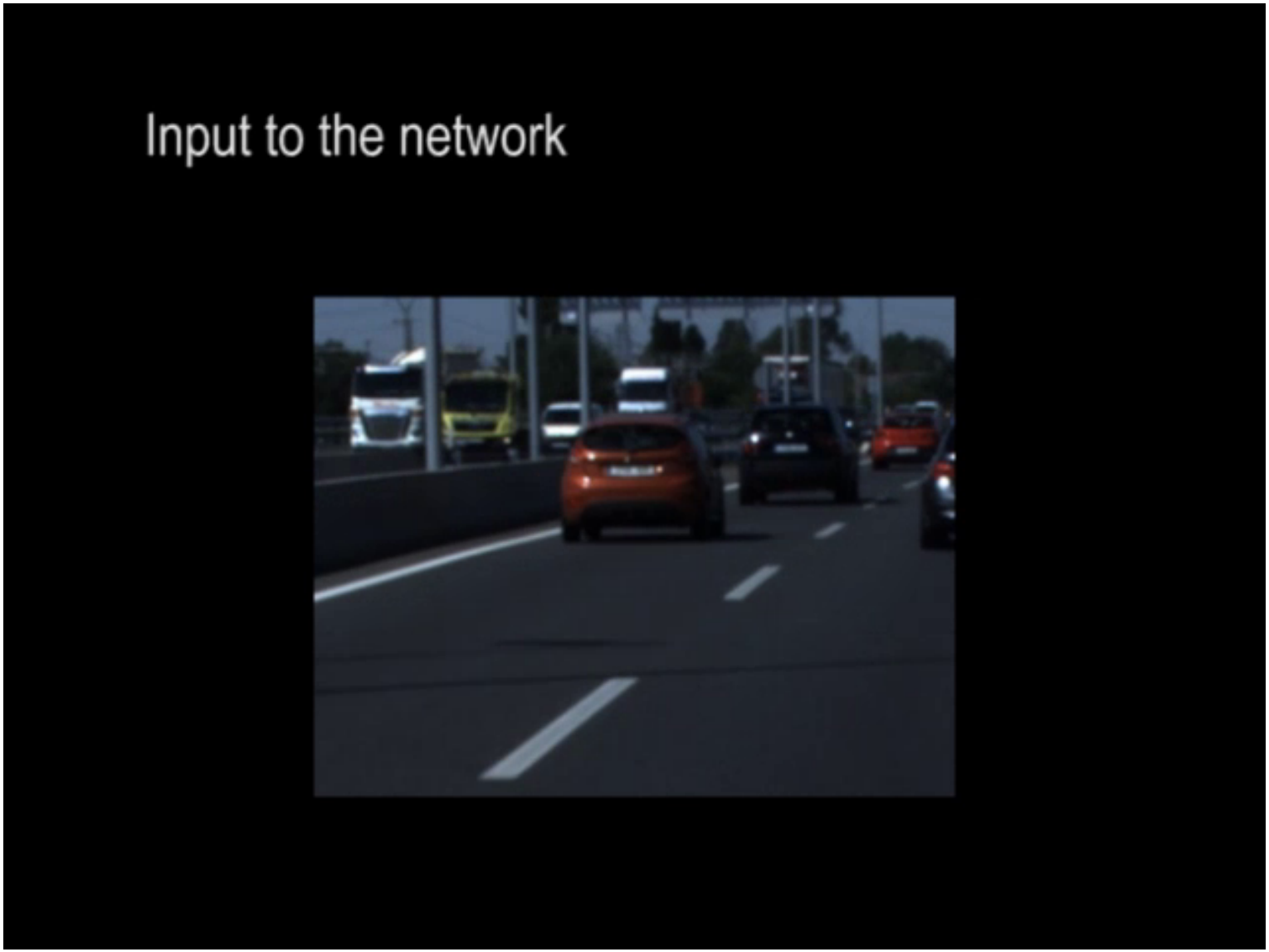
- Dense optical flow (OF) is generated from the ROIs.
- Vehicle is centered on the ROI (canonical view): OF measures the motion of the context around the vehicle.



Problem formulation
System Description



No lane-change
System Description



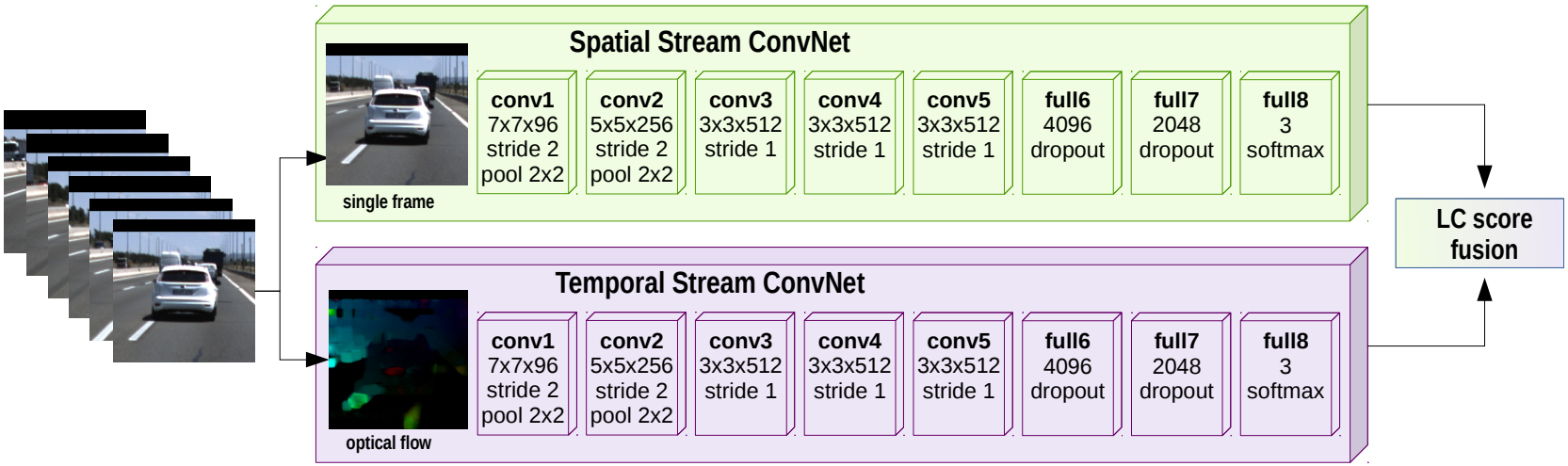
Left lane-change
System Description

Right lane-change
System Description

Disjoint Two-Stream Convolutional Networks

System Description

- Two streams (spatial and temporal), same structure.
- Last FC layer with 3 outputs: left lane-change (LLC), right lane-change (RLC) and no lane-change (NLC).
- Dense OF using polynomial expansion.
- Spatial stream pre-trained using ImageNet.
- Temporal stream pre-trained using UCF-101 and HMDB-51.

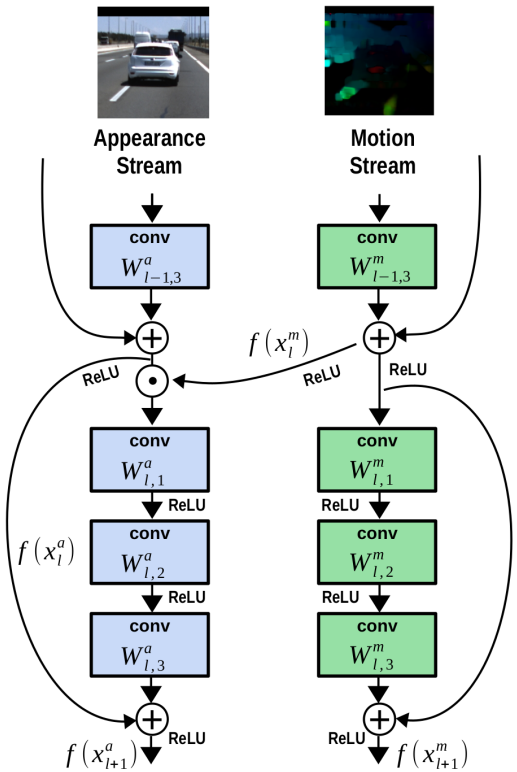


Spatiotemporal Multiplier Networks
System Description

- Appearance and motion streams using ResNet50 with ReLU and batch normalization.
- Multiplicative (element wise) residual connection from the motion path into the appearance stream.

$$\hat{x}^a_{l+1} = f(x^a_l) + \mathcal{F}\left(x^a_l \odot f(x^m_l), W^a_l\right) \quad (1)$$

where x^a_l and x^m_l are the inputs of the l-th layers of the appearance and motion paths respectively, while W^a_l represents the weights of the l-th layer residual unit in the appearance stream and $\odot$ corresponds to elementwise multiplication.



- PREVENTION dataset sampling frequency: 10Hz.
- Observation horizon (N): from t to $t - N$.
- Time-To-Event (TTE):
  - $TTE = 0$ frames: lane-change classification at time t.
  - $TTE = 10$ frames: lane-change **prediction** 1 second ahead $(t + 10)$.
  - $TTE = 20$ frames: lane-change **prediction** 2 seconds ahead $(t + 20)$.

Index

Dataset, evaluation parameters and metrics
Results

- Multi-class problem (3 classes): LLC, RLC and NLC.
- Image size: 112×112.
- Training (85%) and validation (15%).
- Categorical Cross Entropy Loss: $E^p = -log(y_k^p)$.
- Metrics: accuracy (arithmetic mean of precision for all classes), precision and recall (confusion matrices).

	NLC	LLC	RLC
# of sequences	3110	342	438
avg. # of frames	50.9	96.8	80.1

TABLE I

MAIN STATS OF THE DATASET. NLC/LLC/RLC: NO/LEFT/RIGHT LANE-CHANGE.

The following parameters have been evaluated during the experiments:

- ROI sizes: x1, x2, x3 and x4.
- Observation horizon: 20 frames (2 seconds), 30 frames (3 seconds) and 40 frames (4 seconds).
- Time-to-event (prediction horizon): 0 (no prediction), 10 (1 second) and 20 (2 seconds).

Lane change classification results

Results

$TTE = 0$, multiple ROI sizes

		ROI size			
Method	Obs. Horizon	x1	x2	x3	x4
Disjoint	20	83.22	86.18	86.26	87.43
Disjoint	30	83.55	86.69	86.84	86.68
Disjoint	40	84.97	87.69	**89.46**	88.79
ST	20	83.39	85.03	86.51	86.16
ST	30	84.38	84.70	85.36	84.73
ST	40	86.02	87.83	**90.30**	89.64

Table: Disjoint Two-Stream Network and Spatiotemporal Multiplier Network Classification Accuracy (%).

Lane change prediction results

Results

$OH = 20$, $TTE = 10$ and $TTE = 20$, multiple ROI sizes

		ROI size			
Method	TTE	x1	x2	x3	x4
Disjoint	10	84.05	84.54	85.20	85.36
Disjoint	20	85.20	88.82	**91.02**	90.92
ST	10	84.70	85.69	85.20	86.51
ST	20	86.84	90.30	91.45	**91.94**

Table: Disjoint Two-Stream Network and Spatiotemporal Multiplier Network Prediction Accuracy (%). Observation horizon = 20.

Lane change prediction results

Results

Output class	Target class NLC	LLC	RLC	Precision
NLC	473	5	6	97.7%
LLC	10	30	14	55.6%
RLC	12	11	47	67.1%
Recall	95.6%	65.2%	70.1%	90.9%

Table: Disjoint Two-Stream Network Confusion Matrix, OH=20, TTE=20, x4

Lane change prediction results

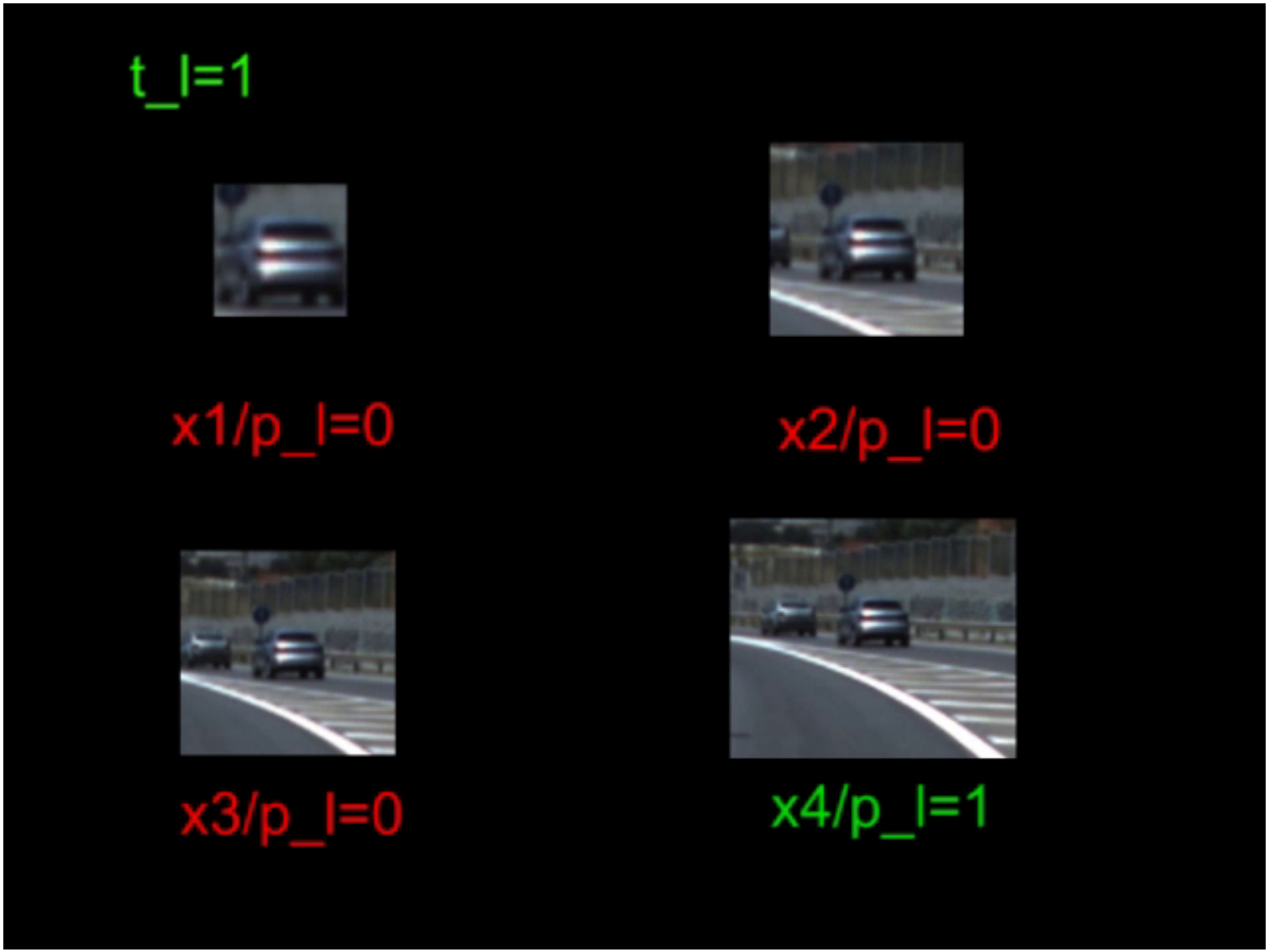
Results

Output class	Target class NLC	LLC	RLC	Precision
NLC	476	5	6	97.7%
LLC	8	33	11	63.5%
RLC	11	8	50	72.5%
Recall	96.2%	71.7%	74.6%	91.9%

Table: Spatiotemporal Multiplier Network Confusion Matrix, OH=20, TTE=20, x4

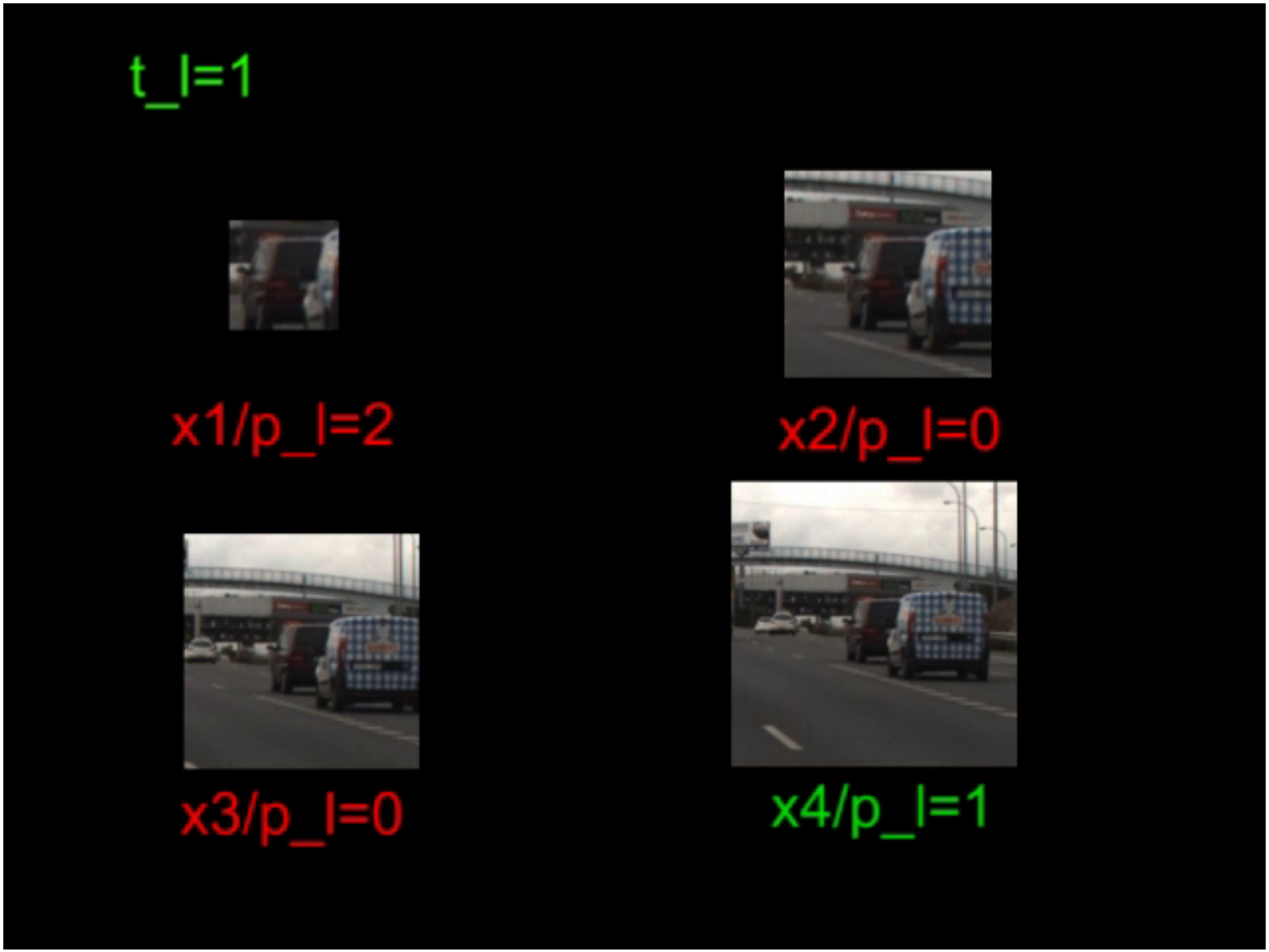
Example 1: left lane-change, x4

Results



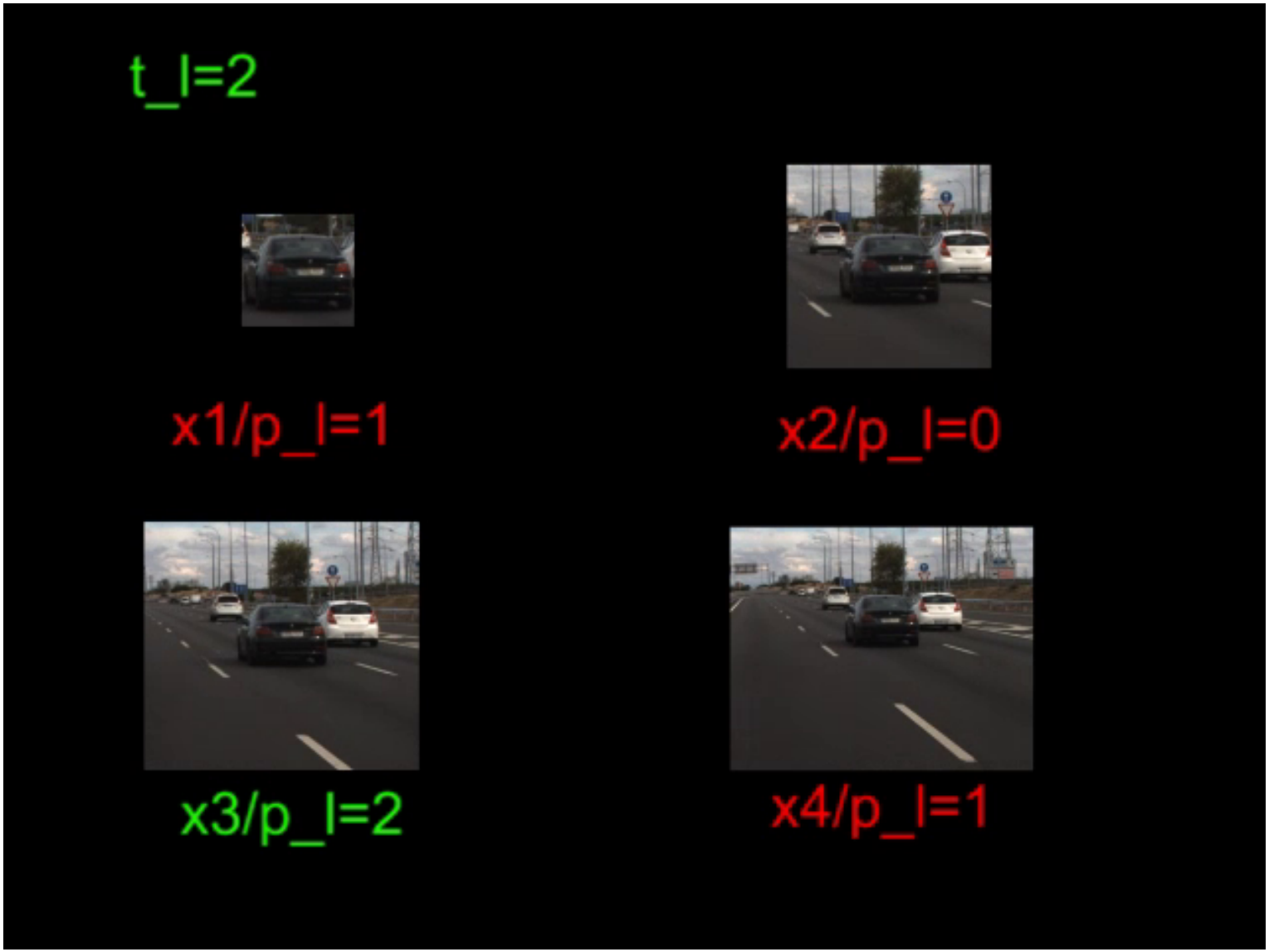
Example 2: left lane-change, x4

Results



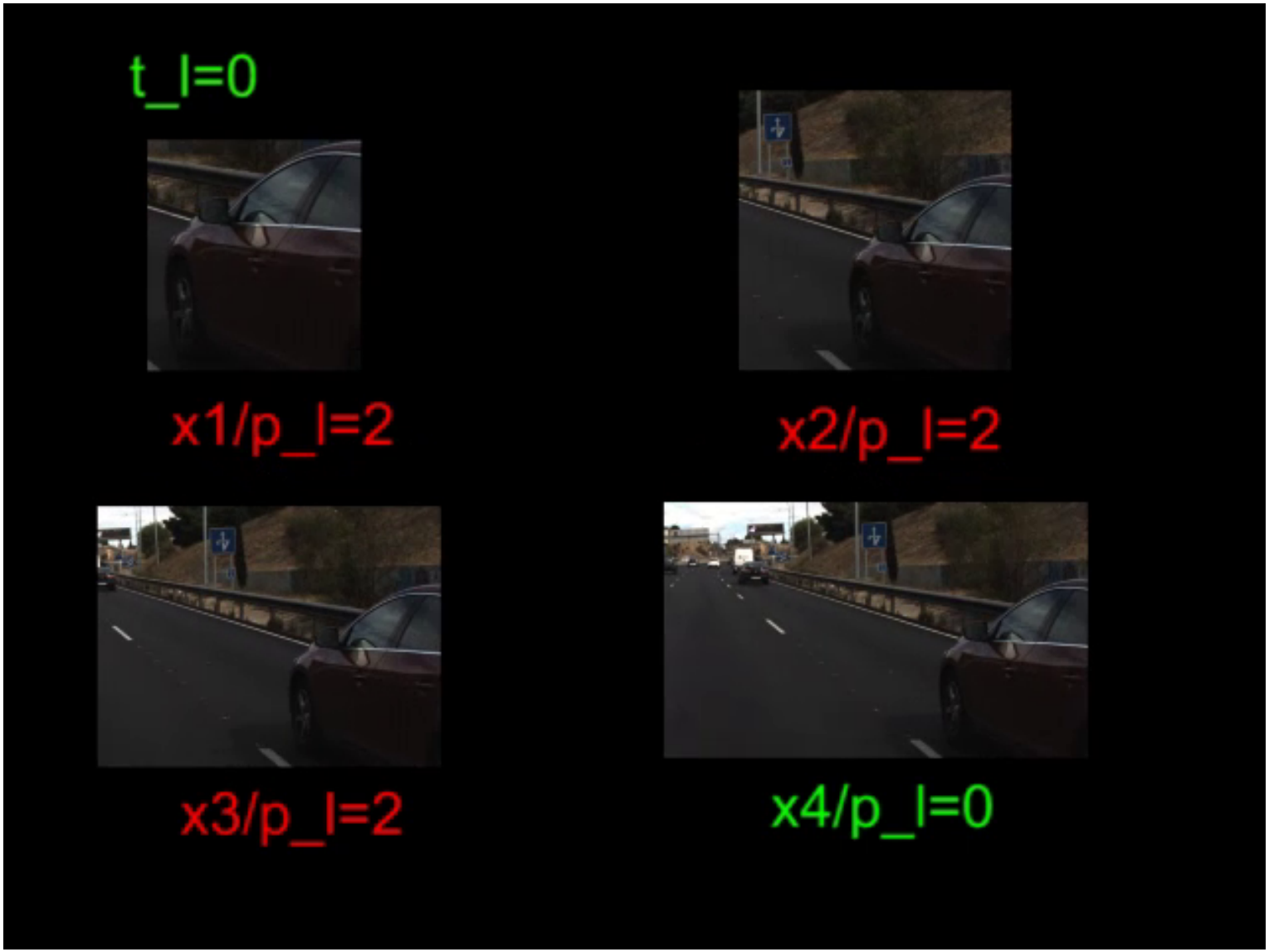
Example 3: right lane-change, x3

Results



Example 4: no lane-change, x4

Results



Index

Conclusions and future works

Conclusions

- Video action classification adapted to perform lane-change classification and prediction of target/surrounding vehicles.
- Highway scenarios using the PREVENTION dataset.
- Visual cues (same as humans) and multi-class classification.
- Spatial (appearance) and temporal (motion, OF) streams.
- Larger ROIs (x3 and x4) provide better performance (context and interaction aware).
- Best configuration: x4, OH=20, 92% 2 seconds ahead.

Future works

- Evaluate other action recognition approaches: I3D, SlowFast, etc.
- Experimental validation with other datasets.

2020 IEEE 23rd International Conference on Intelligent Transportation Systems
ITSC 2020

Two-Stream Networks for Lane-Change Prediction of Surrounding Vehicles

QUESTIONS?

20 September 2020