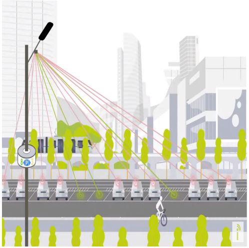
Towards Intelligent Transportation
Urban traffic forecasting and Autonomous parking

Véronique Prinet
Visiting Associate Professor
Chinese Academy of Sciences, Institute of Automation

April 29^{th} 2021
University of Paris (remote)

Table of contents

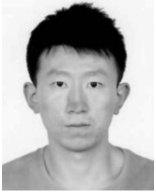
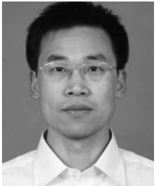
Motivation: connected cities



Urban traffic forecasting

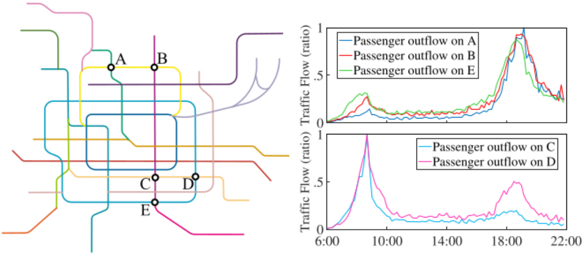
IEEE Trans. on Intelligent Transporation Systems, 2021

Shen Fang Jianlong Chang Michael Werman
ChunHong Pan Shiming Xiang

Objective

Objective: Given past observed data, we want to infer the future traffic state in the forthcoming X minutes/hours/days (typically X=1h).

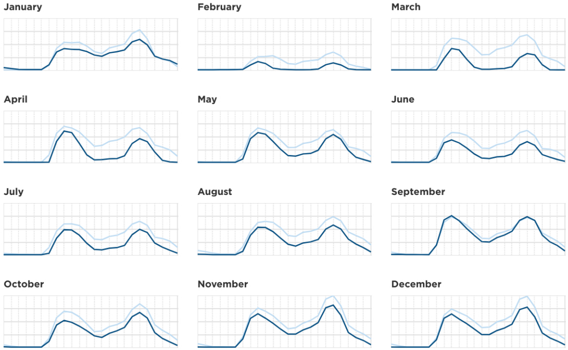


(a) Beijing subway network (b) Traffic flow on five nodes

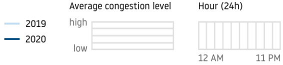
Beijing traffic

WORKING DAY TRAVEL PATTERNS BY MONTH

How did the travel patterns look like during working days in 2020 and 2019?

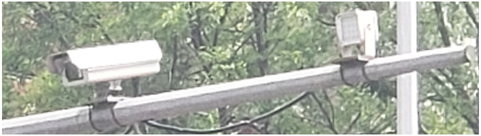


How to read these charts?



https://www.tomtom.com/en_gb/traffic-index/beijing-traffic/

Sensors and data

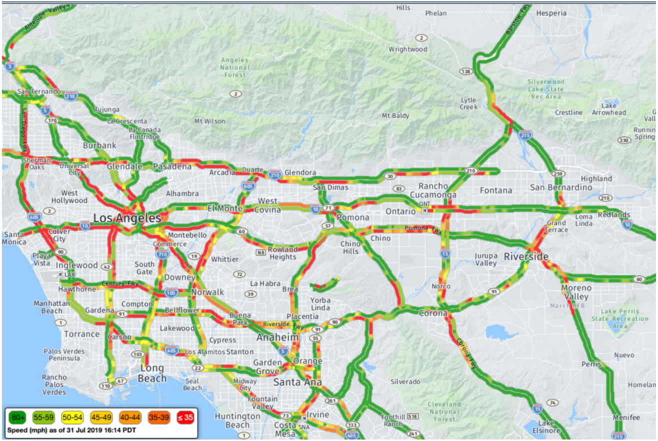


- flow (in / out)
- speed
- occupancy
- ...

Why is it hard?

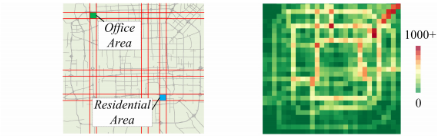
Why is it hard?

CALIFORNIA DEPARTMENT OF
TRANSPORTATION



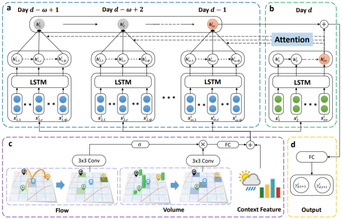
Related work (1/3)

Deep spatio-temporal residual net- works for citywide crowd flows prediction, Zhang & al., AAAI 2017



(a) Grid-based map segmentation (b) Inflow matrix

Modeling Spatial-Temporal Dynamics for Traffic Prediction , Yao & al., arxiv:1803.01254 2018



Related work (2/3)

Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting, Guo & al., AAAI 2019

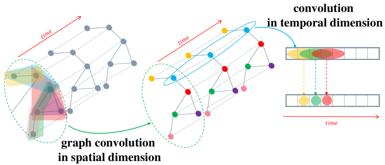


Figure 5: The architecture of spatial-temporal convolutions of ASTGCN.

Dynamic Spatial-Temporal Graph Convolutional Neural Networks for Traffic Forecasting, Diao & al. AAAI 2019

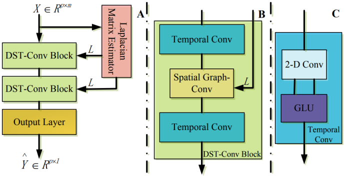
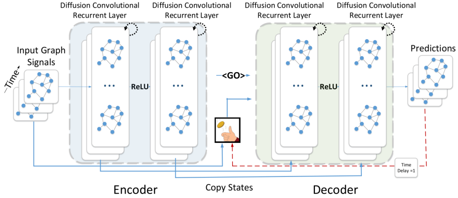


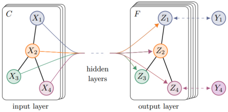
Figure 1: The framework of DGCNN

Diffusion convolution RNN: data driven traffic forecasting, Li and al, ICLR 2018



$$[X^{(t-T'+1)}, \dots, X^{(t-1)}] \rightarrow [X^{(t)}, \dots, X^{(t+T)}]$$

Related work (3/3)



- Spectral graph convolution [Bruna'13, Defferrard'17, Kipf'17]

$$g_\theta \star x = U g_\theta U^T x$$

U: eigenvectors of the Laplacian $L = I_N - D^{-1/2} A D^{-1/2} = U \Lambda U^T$

- Graph convolution on the spatial domain [Hamilton'18]

$$y_i = f_\theta(x_i, \{x_j | \forall v_j \in \mathcal{N}(v_i)\}),$$

AGGREGATION: $\tilde{y}_i = \sum_{j \in \mathcal{N}(i)} x_j W$
UPDATE: $y_i = g(\tilde{y}_i, x_i)$

- Permutation invariance, independant of neigbor size, spatially invariant (weight sharing)

Contributions

Our contributions:

- (Dilated attention) graph convolution operator, accounting for short and long range spatial correlations,

Contributions

Our contributions:

- (Dilated attention) graph convolution operator, accounting for short and long range spatial correlations,
- Temporal integration, accounting for short and long range time dependencies,

Contributions

Our contributions:

- (Dilated attention) graph convolution operator, accounting for short and long range spatial correlations,
- Temporal integration, accounting for short and long range time dependencies,
- Data fusion scheme, accounting for external (exhogenous) factors.

Problem statement

Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$,
$\mathcal{V} = \{v_i\}$, $\mathcal{E} = \{e_{ij} \in [0,1] | e_{ij} = 1 \text{ if } v_j \in \mathcal{N}(v_i)\}$
and nodes attribute $\mathcal{X} = \{x_i\}$:

$$\mathcal{F}_\Theta(\mathcal{G}) \quad : \quad \left(\{\mathcal{X}^{(p)}\}_{p=0}^{P-1}, \mathbf{E}\right) \longrightarrow \hat{\mathbf{X}}_{\mathbf{t}}$$

$$\Theta = \arg\min_{\tilde{\Theta}} \mathcal{L}_2(\hat{\mathbf{X}}_{\mathbf{t}}, \mathbf{X}_{\mathbf{t}})$$

Problem statement

Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$,
$\mathcal{V} = \{v_i\}$, $\mathcal{E} = \{e_{ij} \in [0,1] | e_{ij} = 1 \text{ if } v_j \in \mathcal{N}(v_i)\}$
and nodes attribute $\mathcal{X} = \{x_i\}$:

$$\mathcal{F}_\Theta(\mathcal{G}) \quad : \quad (\{\mathcal{X}^{(p)}\}_{p=0}^{P-1}, \mathbf{E}) \longrightarrow \hat{\mathbf{X}}_\mathbf{t}$$

$$\Theta = \arg\min_{\tilde{\Theta}} \mathcal{L}_2(\hat{\mathbf{X}}_\mathbf{t}, \mathbf{X}_\mathbf{t})$$

Observed historical traffic flow $\mathcal{X}^{(p)}, p = 0, \cdots, P-1$, $\mathcal{X}^{(p)} \in \mathbb{R}^{N \times T \times M}$
External factors $\mathbf{E} = \{\boldsymbol{e}_t^w, \boldsymbol{e}_t^h, \mathbf{E}^p\}$ at future time t,
$\mathbf{E} \in \mathbb{R}^{D^w} \times \{0,1\}^{D^h} \times \mathbb{N}^{N \times D^p}$
Output the traffic flow on all nodes at the next time step t, $\hat{\mathbf{X}}_\mathbf{t} \in \mathbb{R}^{N \times M}$

Graph convolution operator: DAGC $\mathcal{H}(.)$

Given $\boldsymbol{x}_i \in \mathbb{R}^C$, a feature of dimension C on node v_i, the attention graph convolution, for a given l, is defined as follows,

Graph convolution operator: DAGC $\mathcal{H}(.)$

Given $\boldsymbol{x}_i \in \mathbb{R}^C$, a feature of dimension C on node v_i, the attention graph convolution, for a given l, is defined as follows,

for a one order neighborhood (single hop):

$$\boldsymbol{y}_i = \sum_{v_j \in \mathcal{N}(v_i, l)} \boldsymbol{x}_j \mathbf{W}_a + g(\boldsymbol{x}_i),$$

Graph convolution operator: DAGC $\mathcal{H}(.)$

Given $x_i \in \mathbb{R}^C$, a feature of dimension C on node v_i, the attention graph convolution, for a given l, is defined as follows,

for a one order neighborhood (single hop):

$$y_i = \sum_{v_j \in \mathcal{N}(v_i, l)} f_\psi(x_i, x_j) \cdot x_j \mathbf{W}_a + g(x_i),$$

$$\tilde{f}_\psi(x_i, x_j) = h_\psi(x_i)^{\mathrm{T}} \cdot h_\psi(x_j)$$
$$f_\psi(x_i, x_j) = softmax(\tilde{f}_\psi(x_i, x_j))$$

Graph convolution operator: DAGC $\mathcal{H}(.)$

Given $\mathbf{x}_i \in \mathbb{R}^C$, a feature of dimension C on node v_i, the attention graph convolution, for a given l, is defined as follows,

for a L order neighborhood (L-hop) with dilation d:

$$\mathbf{y}_i' = \sum_{l=1}^{L} \theta_l \sum_{v_j \in \mathcal{N}(v_i, l \cdot d)} f_\psi(\mathbf{x}_i, \mathbf{x}_j) \cdot \mathbf{x}_j \mathbf{W}_a + g(\mathbf{x}_i),$$

$$\mathbf{y}_i = \sigma(\mathbf{y}_i')$$

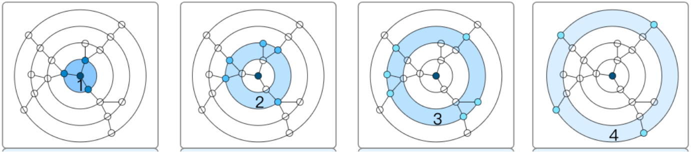
Graph convolution operator: DAGC $\mathcal{H}(.)$

Given $\boldsymbol{x}_i \in \mathbb{R}^C$, a feature of dimension C on node v_i, the attention graph convolution, for a given l, is defined as follows,

for a L order neighborhood (L-hop) with dilation d:

$$\boldsymbol{y}_i' = \sum_{l=1}^{L} \theta_l \sum_{v_j \in \mathcal{N}(v_i, l \cdot d)} f_\psi(\boldsymbol{x}_i, \boldsymbol{x}_j) \cdot \boldsymbol{x}_j \mathbf{W}_a + g(\boldsymbol{x}_i),$$

$$\boldsymbol{y}_i = \sigma(\boldsymbol{y}_i')$$



Space and time integration

Space integrator: Given input traffic record $\mathbf{X}^{(p)} \in \mathbb{R}^{N \times TM}$:

$$\mathbf{Y}^{(p)} = \mathcal{H}^{(p)}(\mathbf{X}^{(p)}, \mathcal{G}), \qquad p = 0, \cdots, P-1$$

Space and time integration

Space integrator: Given input traffic record $\mathbf{X}^{(p)} \in \mathbb{R}^{N \times TM}$:

$$\mathbf{Y}^{(p)} = \mathcal{H}^{(p)}(\mathbf{X}^{(p)}, \mathcal{G}), \qquad p = 0, \cdots, P-1$$

Time integrator: Given $\mathbf{Y}^{c} = [\mathbf{Y}^{(0)}, \cdots, \mathbf{Y}^{(P-1)}] \in \mathbb{R}^{N \times PF}$:

$$\mathbf{Y}^{\tau} = \mathcal{H}^{\tau}(\mathbf{Y}^{c}, \mathcal{G}),$$

External factors embedding

Time varying factors:

Weather (discrete and real-valued variables):

$$\hat{\mathbf{e}}_t^w = FC^w(\mathbf{e}_t^{w,d}, \mathbf{e}_t^{w,r}) \qquad \hat{\mathbf{e}}_t^w = (\hat{\mathbf{e}}_t^{w,d}, \hat{\mathbf{e}}_t^{w,r}),$$

Holiday (categorical variable):

$$\hat{\mathbf{e}}_t^h = FC^h(\mathbf{e}_t^h) \qquad \hat{\mathbf{e}}_t^h \in \mathbb{N}^{F^h},$$

External factors embedding

Time varying factors:

Weather (discrete and real-valued variables):

$$\hat{\mathbf{e}}_t^{w} = FC^{w}(\mathbf{e}_t^{w,d}, \mathbf{e}_t^{w,r}) \qquad \hat{\mathbf{e}}_t^{w} = (\hat{\mathbf{e}}_t^{w,d}, \hat{\mathbf{e}}_t^{w,r}),$$

Holiday (categorical variable):

$$\hat{\mathbf{e}}_t^{h} = FC^{h}(\mathbf{e}_t^{h}) \qquad \hat{\mathbf{e}}_t^{h} \in \mathbb{N}^{F^h},$$

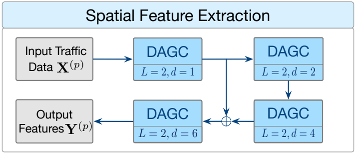
Space varying factors:

Points of interest (categorical variable):

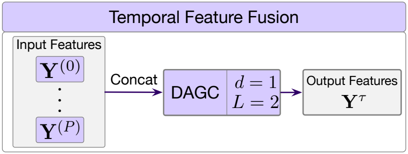
$$\hat{\mathbf{E}}^{p} = FC^{p}(\mathbf{E}^{p}) \qquad \hat{\mathbf{E}}^{p} \in \mathbb{N}^{N \times F^p}.$$

What do we have so far

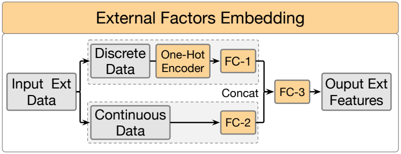
Space integration



Time integration



External (weather) factors embedding



Overall fusion

1- Traffic flow and PoIs features are first fet into a DAGC module:

$$\mathbf{Y} = \mathcal{H}([\mathbf{Y}^{\tau}, \hat{\mathbf{E}}^{p}], \mathcal{G}).$$

2- Weather features and holiday notifications, expanded to space domain, are then fed into an multilayer perceptron:

$$\hat{\mathbf{X}}_{\mathbf{t}} = MLP([\mathbf{Y}, \hat{\mathbf{E}}_{t}^{w}, \hat{\mathbf{E}}_{t}^{h}])$$

Overall fusion

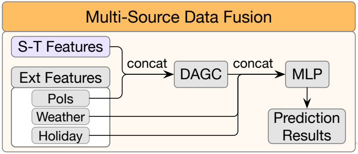
1- Traffic flow and PoIs features are first fet into a DAGC module:

$$\mathbf{Y} = \mathcal{H}([\mathbf{Y}^{\tau}, \hat{\mathbf{E}}^{p}], \mathcal{G}).$$

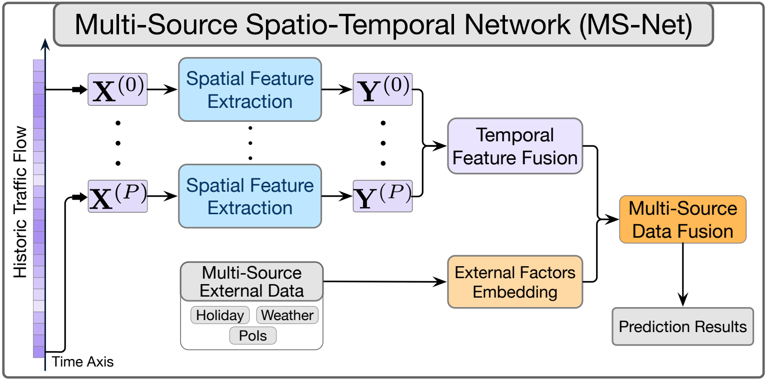
2- Weather features and holiday notifications, expanded to space domain, are then fed into an multilayer perceptron:

$$\hat{\mathbf{X}}_{\mathbf{t}} = MLP([\mathbf{Y}, \hat{\mathbf{E}}_t^{w}, \hat{\mathbf{E}}_t^{h}])$$

Fusion



Overall pipeline



Experiments: Data

Table: Dataset meta information

Properties	Datasets		
	Subway	Bus	Taxi
# traffic nodes	278	4219	300
time interval	10 mins	1 hour	20 mins
time span	2016/6/1 - 2016/6/29		2015/11/28 - 2016/1/26
# train days	15 days		32 days
# valid & test days	7 days		14 days
daily range	6:00-22:00		

Experiments: Data

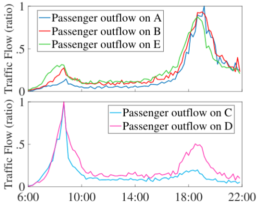
Table: PoIs categories.

ID	PoIs categories	ID	PoIs categories
0	food & beverage service	7	automobile service
1	shopping center	8	education
2	hotel	9	medical treatment
3	public transportation service	10	tourism service
4	entertainment	11	enterprises and institutions
5	residence	12	finance & insurance
6	living service	13	government agency

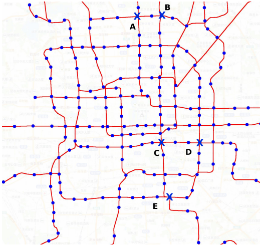
Table: Weather categories.

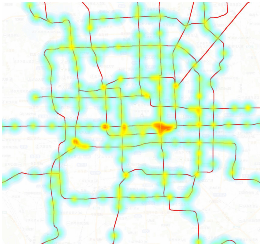
ID	categories	ID	categories
0	cloudy	4	thunder
1	partly cloudy	5	fog
2	mostly cloudy	6	snow / hail
3	light rain	7	sunny

Experiments: Data - Beijing subway



(b) Traffic flow on five nodes





[Inflow weekday] [Inflow weekend] [Outflow weekday] [Outflow weekend]

Experiments: Results and comparison with SOTA

Table: Experimental results of the subway (top) and bus (bottom) evaluation datasets.

	MAE	MAPE (%)	RMSE
HA	45.08	31.02	94.94
GAT	36.68 $\pm$ 2.58	28.97 $\pm$ 2.29	65.35 $\pm$ 6.31
GRU	23.33 $\pm$ 0.20	20.29 $\pm$ 0.67	41.92 $\pm$ 0.33
ChebNet	22.91 $\pm$ 0.59	19.38 $\pm$ 0.39	40.02 $\pm$ 0.98
DCRNN	22.49 $\pm$ 0.22	19.50 $\pm$ 1.08	38.63 $\pm$ 0.47
STGCN	21.69 $\pm$ 0.62	19.13 $\pm$ 1.74	36.49 $\pm$ 0.48
STGCNAction	21.65 $\pm$ 0.27	18.97 $\pm$ 1.32	37.06 $\pm$ 0.39
GSTNet	21.33 $\pm$ 0.13	18.63 $\pm$ 0.72	36.08 $\pm$ 0.22
MS-Net	**19.44** $\pm$ **0.14**	**16.97** $\pm$ 0.30	**32.19** $\pm$ 0.17

	MAE	MAPE (%)	RMSE
HA	35.93	55.47	73.54
GAT	26.40 $\pm$ 0.29	46.88 $\pm$ 3.00	52.73 $\pm$ 0.38
GRU	24.07 $\pm$ 0.22	40.73 $\pm$ 1.59	53.46 $\pm$ 0.25
ChebNet	27.06 $\pm$ 1.10	42.89 $\pm$ 2.67	56.01 $\pm$ 2.73
DCRNN	27.06 $\pm$ 0.12	43.95 $\pm$ 0.47	55.23 $\pm$ 0.09
STGCN	23.42 $\pm$ 0.31	39.01 $\pm$ 1.57	48.80 $\pm$ 2.64
STGCNAction	21.05 $\pm$ 0.72	36.12 $\pm$ 1.01	40.93 $\pm$ 1.90
GSTNet	N / A	N / A	N / A
MS-Net	**19.15** $\pm$ **0.28**	**33.12** $\pm$ **1.19**	**36.42** $\pm$ **0.39**

Metrics: MAE (Mean average error), MAPE (Mean average percentage error), RMSE (root mean square error)

Experiments: Results and comparison with SOTA

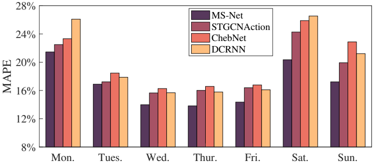
Table: Experimental results of the taxi evaluation datasets.

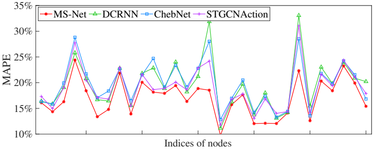
	MAE	MAPE (%)	RMSE
HA	26.18	40.24	55.95
GAT	22.05 ± 1.01	35.27 ± 1.47	45.44 ± 1.66
GRU	20.24 ± 0.19	32.77 ± 1.73	40.04 ± 0.15
ChebNet	19.81 ± 0.07	31.97 ± 0.38	38.39 ± 0.58
DCRNN	20.46 ± 0.34	31.58 ± 1.38	42.03 ± 0.18
STGCN	19.34 ± 0.24	31.34 ± 1.16	37.30 ± 0.26
STGCNAction	19.78 ± 0.13	31.45 ± 1.90	39.41 ± 0.14
GSTNet	19.17 ± 0.32	30.77 ± 1.35	37.01 ± 0.35
MS-Net	**18.60 ± 0.06**	**29.37 ± 0.46**	**35.62 ± 0.12**

GRU: Gated Recurrent Unit
GAT: Graph Attention Network, Velickovic & al. , ICLR 2017
STGCN: Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting, Yu & al. IJCAI 2018

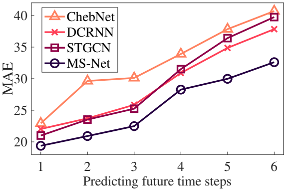
Experiments: Results on subway dataset

Prediction: Next time step





Prediction: Multiple time steps



STGCNAction: Spatial Temporal Graph Conv. Networks for Skeleton-Based Action Recognition, Yan & al. , 2018
ChebNet: Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering, Defferrard & al. , 2016
DCRNN: Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting, Li & al. 2018

Experiments: Ablation study - Beijing subway

Table: Results of temporal feature fusion

Fusion Method	Subway Dataset MAE	MAPE (%)	Time (mins)
Weighted Sum	20.85 ± 0.05	18.08 ± 0.31	0.07
Average	20.49 ± 0.12	17.89 ± 0.49	0.07
Proposed	$\mathbf{19.44} \pm \mathbf{0.14}$	$\mathbf{16.97} \pm \mathbf{0.30}$	0.08

Table: Results of external factors

External Factors	Subway Dataset MAE	MAPE (%)	Time (mins)
w/o factors	20.71 ± 0.04	17.95 ± 0.15	0.06
with factors	$\mathbf{19.44} \pm \mathbf{0.14}$	$\mathbf{16.97} \pm \mathbf{0.30}$	0.08

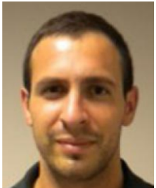
Conclusion

- Dilated attention graph convolution operator
- Integration in space, time and exhogenous data fusion
- On going directions
  - Optimal architecture search (in space domain)
  - Meta learning for exhogenous data fusion

Towards autonomous parking
A vision based system

IEEE Trans. on Intelligent Transporation Systems, sept. 2019

Stanislav Panev   Francisco Vicente   Fernando de la Torre

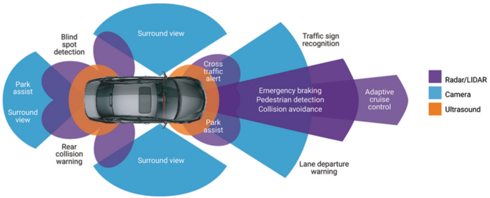


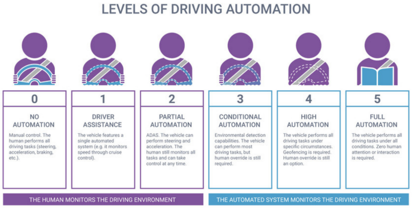
Motivation

- Advanced driver assistance systems (ADAS): to reduce the number of fatalities on the road (Pedestrian detection/avoidance, Lane departure warning/correction, Traffic sign recognition, Automatic emergency braking, Blind spot detection)
- Human Horizons (leading Chinese smart mobility and autonomous driving research company) has already Level 4 Autonomous Valet Parking (AVP) system
- Full vision based

https://www.prnewswire.com/news-releases/world-first-level-4-fully-autonomous-vehicle-parking-system-rolled-out-on-human-horizons-hiphi-x-301132815.html

Motivation





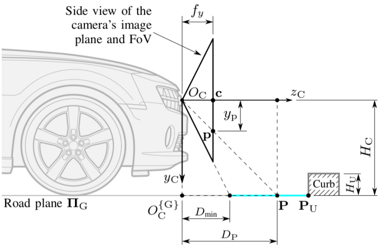
https://www.synopsys.com/automotive/what-is-adas.html,
https://www.synopsys.com/automotive/autonomous-driving-levels.html

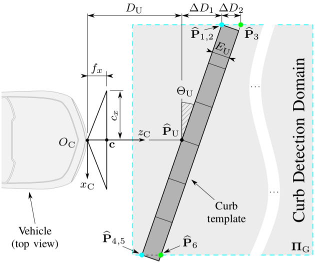
Method

- Objective: detect the curb –vehicle-to-curb distance, curb height, angle
- Fully vision based system (single CDD fisheye camera mouted near the front licence plate)
- Near real time, $>$ 90% accuracy requirement
- Adversarial conditions: rainy/foggy weather, damaged curb
- Perpendicular parking

Youtube video teaser [Video teaser]

Method: Setting and assumptions





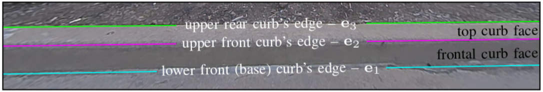
$$D_U = f_y \frac{H_c}{y_U} \qquad \frac{h_U}{H_U} = \frac{f_y}{D_U}$$

(f_x, f_y): cameral focal length
H_c: camera-to-road-plane vertical distance (i.e., along y axis)
h_U: curb's frontal projection face in the image
P_U: point of the curb on the road plane
y_U: vertical's coordinate of P_U's projection in the image

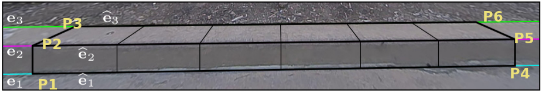
D_U: curb's to camera distance
H_U: curb's height
θ_U: curb's angle
E_U: curb's width

Method: Simple concept

a) curb appearance in image



b) curb's faces and edges



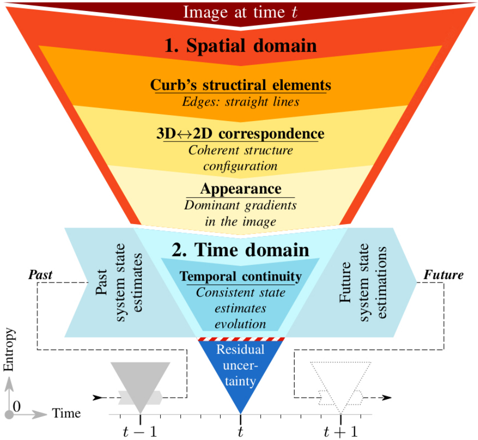
c) 3D curb template's edges and faces

$$\hat{P}_1 = [-W_{max}, H_C, \hat{D}_U + \Delta D_1],$$
$$\hat{P}_2 = [-W_{max}, H_C - \hat{H}_U, \hat{D}_U + \Delta D_1],$$
$$\hat{P}_3 = [-W_{max}, H_C - \hat{H}_U, \hat{D}_U + \Delta D_1 + \Delta D_2],$$

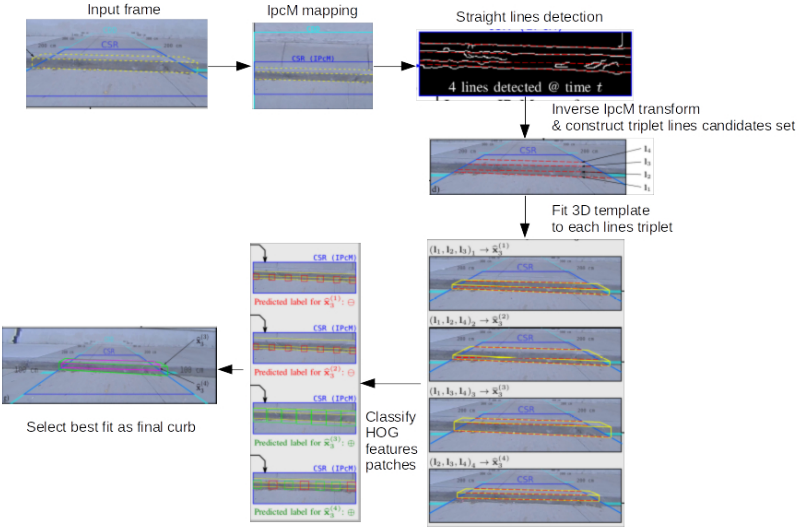
$$\hat{P}_4 = [W_{max}, H_C, \hat{D}_U - \Delta D_1]$$
$$\hat{P}_5 = [W_{max}, H_C - \hat{H}_U, \hat{D}_U - \Delta D_1]$$
$$\hat{P}_6 = [W_{max}, H_C - \hat{H}_U, \hat{D}_U - \Delta D_1 + \Delta D_2]$$

$$\Delta D_1 = W_{max} \tan \hat{\theta}_U, \qquad \Delta D_2 = \frac{\hat{E}_U}{\cos \hat{\theta}_U}$$

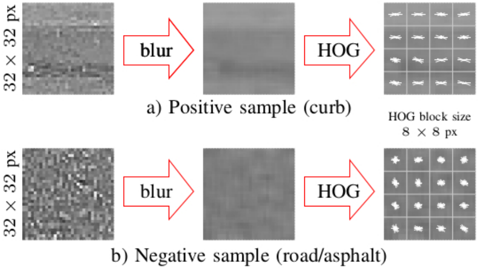
Method: Pipeline

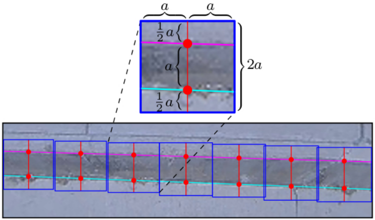


Method: Pipeline



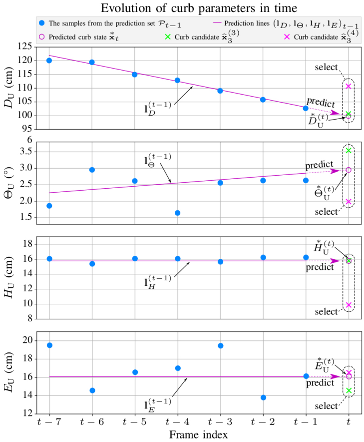
Method: Pipeline



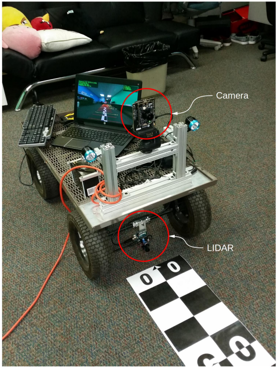


Method: Pipeline

Temporal tracking asssuming small frame to frame displacement:



Experimental results: data



DATASET DETAILS

Vid. seq. #	Curb height (cm)	Curb depth (cm)	Weather conditions	Curb/road physical properties	Frames count‡
1	11.1	20.6	Clear	Co./As.*	702/374
2	13.3	20.6	Clear	Co./As.*	665/344
3	10.6	20.6	Clear	Co./As.*	626/378
4	16.2	15.9	Shadow	Co./As.*	519/332
5	14.6	16.4	Shadow	Co./As.*	497/321
6	10.5	20.6	Clear	Co./As.*	580/345
7	10.8	20.3	Shadow	Co./As.*	545/318
8	9.8	21.6	Shadow	Co./As.*	521/341
9	11.4	20.8	Shadow	Pa./St.†	486/291
10	9.8	20.3	Shadow	Pa./St.†	412/308
11	13.7	20.8	Clear	Co./As.*	555/360

* Concrete/Asphalt
† Painted/Strained
‡ Total number of frames in the sequence/Number of the frames with curb presented in the CCD

Experimental results videos

[Video 2] [Video 3] [Video 5] [Video 8] [Video 9]

https://www.youtube.com/playlist?list=PLeJQFxWyfCj7v1TNN0OOppD2KO8R3Zoib

Experimental results: quantitative evaluation

Classification rate

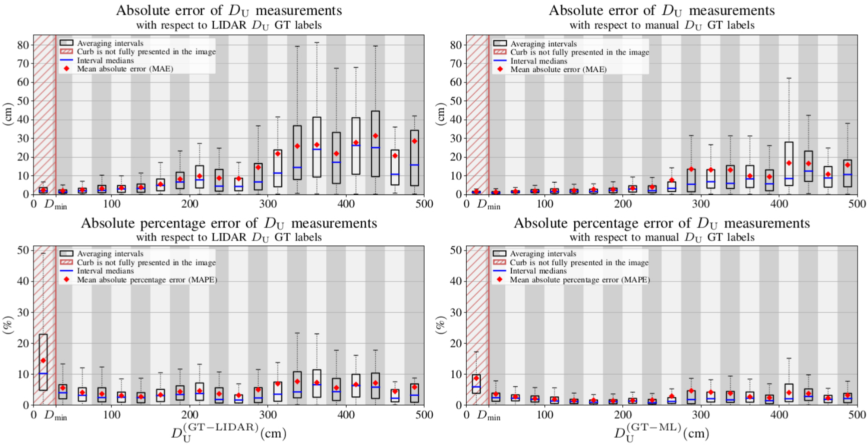
Video sequence #	Accuracy	F_1 score
1	99.7%	0.997
2	99.1%	0.989
3	93.6%	0.926
4	97.4%	0.986
5	97.4%	0.819
6	83.9%	0.901
7	96.2%	0.974
8	81.7%	0.871
9	90.5%	0.940
10	91.6%	0.956
11	81.7%	0.798
Average:	**91.4%**	**0.923**

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

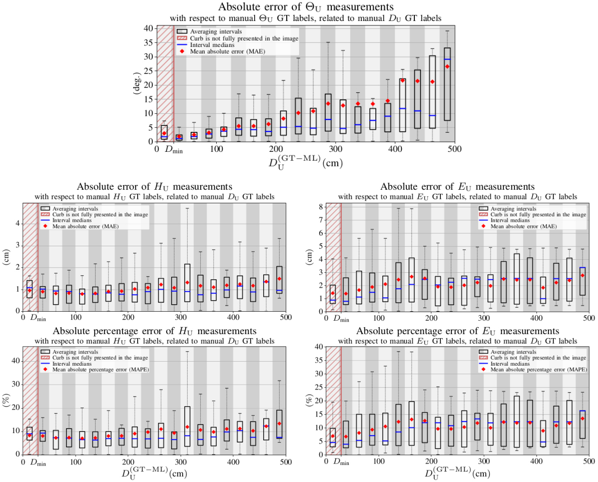
$$F_1 = 2\frac{precision.recall}{precision + recall}$$

$$precision = \frac{TP}{TP + FP}, \quad recall = \frac{TP}{TP + FN}$$

Experimental results: quantitative evaluation D_u



Experimental results: quantitative evaluation Θ_U, E_U



Conclusion

Conclusions and perspectives

- Deep learning, a powerful generic *transversal* toolbox

  agnostic to the application

  enables cross-modality data processing

- For many applications, in particular for large scale natural phenomena/events modeling, numerical (mathematical) models should be coupled with DL

- Many challenging and open problems in the field of dynamic/time-series event analysis (medical, remote sensing)