Left-Right Comparative Recurrent Model for stereo Matching

Zequn Jie, Pengfei Wang, Yonggen Ling, Bo Zhao, Yunchao Wei, Jiashi Feng, Wei Liu

Tencent AI lab, National University of Singapore, University of British Columbia, University of Illinois Urbana-Champaign

Presented by: Xi Fang

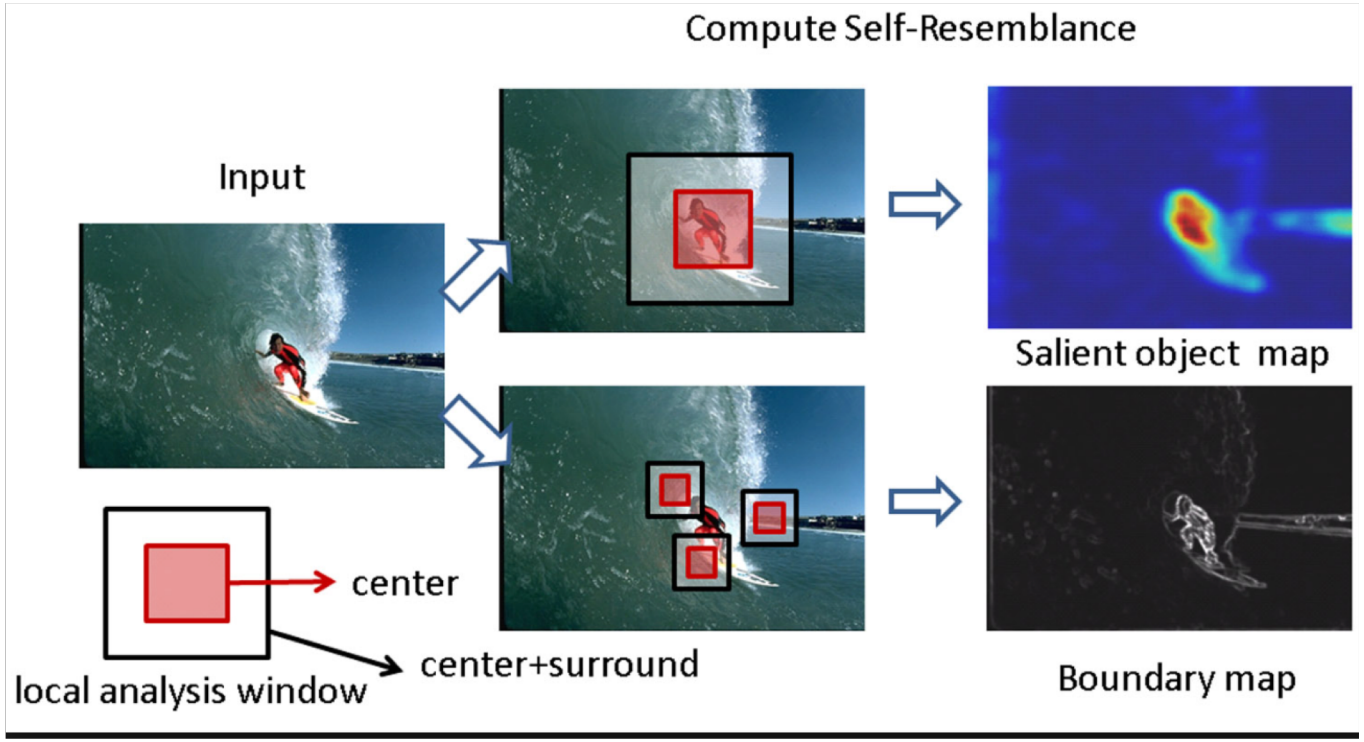
Date: 9/26/2018

Content

- Introduction
- Dataset
- Conventional Methods
- Contribution
- Method
- Result
- Reference

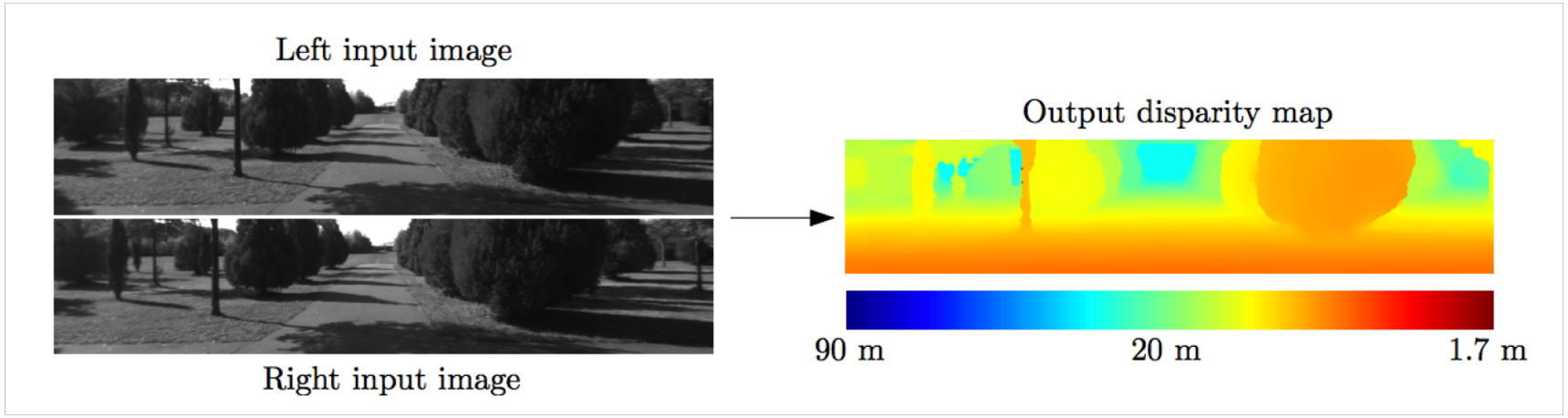
Introduction

- Compute the saliency map in one image



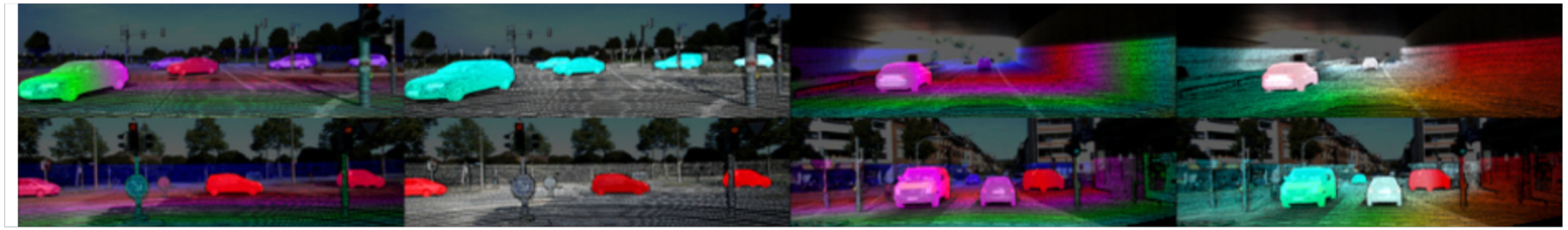
Introduction

- Compute the dense disparity map between a rectifier stereo pair of images
- Core to many computer vision applications, including robotics and antunomous vehicles

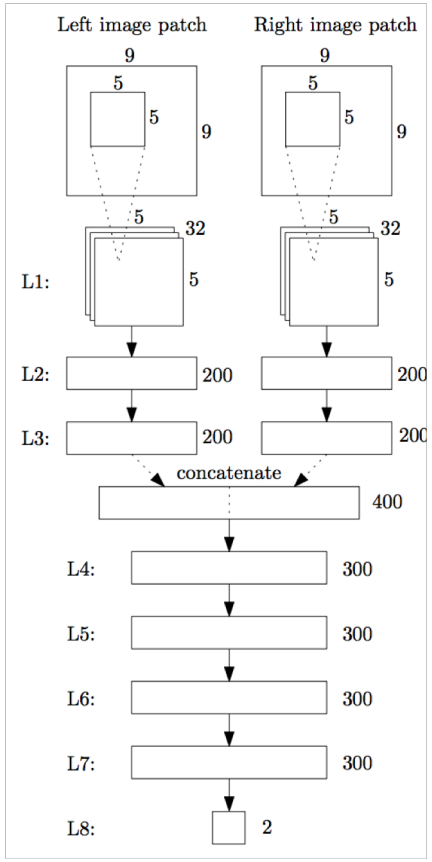


Dataset

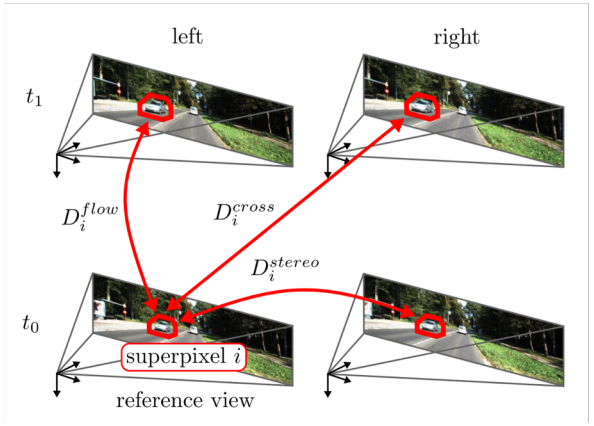
- KITTI 2015 Scene Flow Middlebury



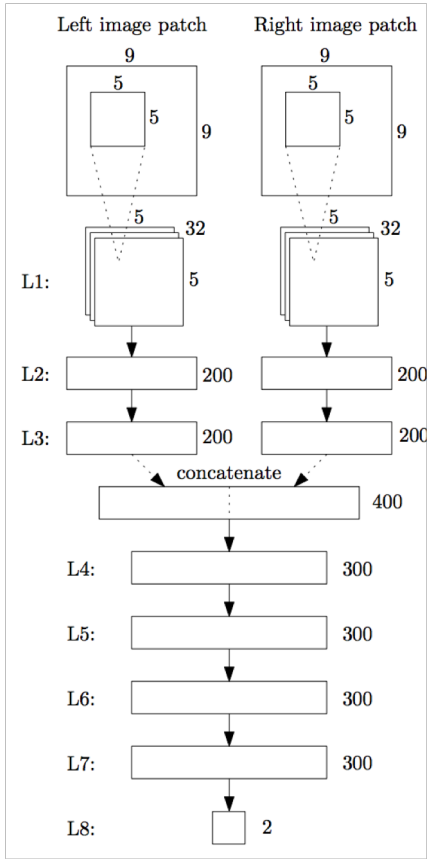
Conventional Methods



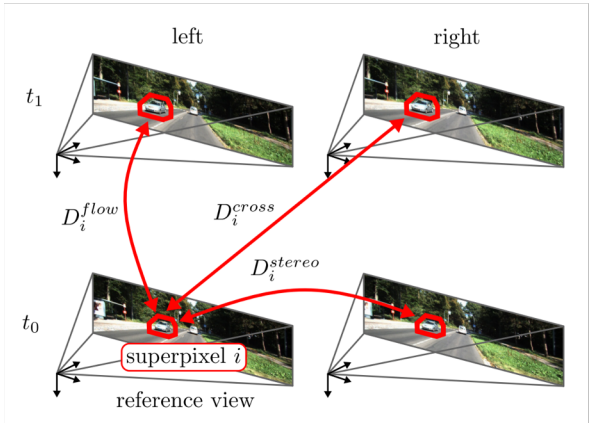
$$C_{\text{CNN}}(\mathbf{p}, d) = f_{\text{neg}}(< \mathcal{P}^{L}_{9\times 9}(\mathbf{p}), \mathcal{P}^{R}_{9\times 9}(\mathbf{pd}) >)$$

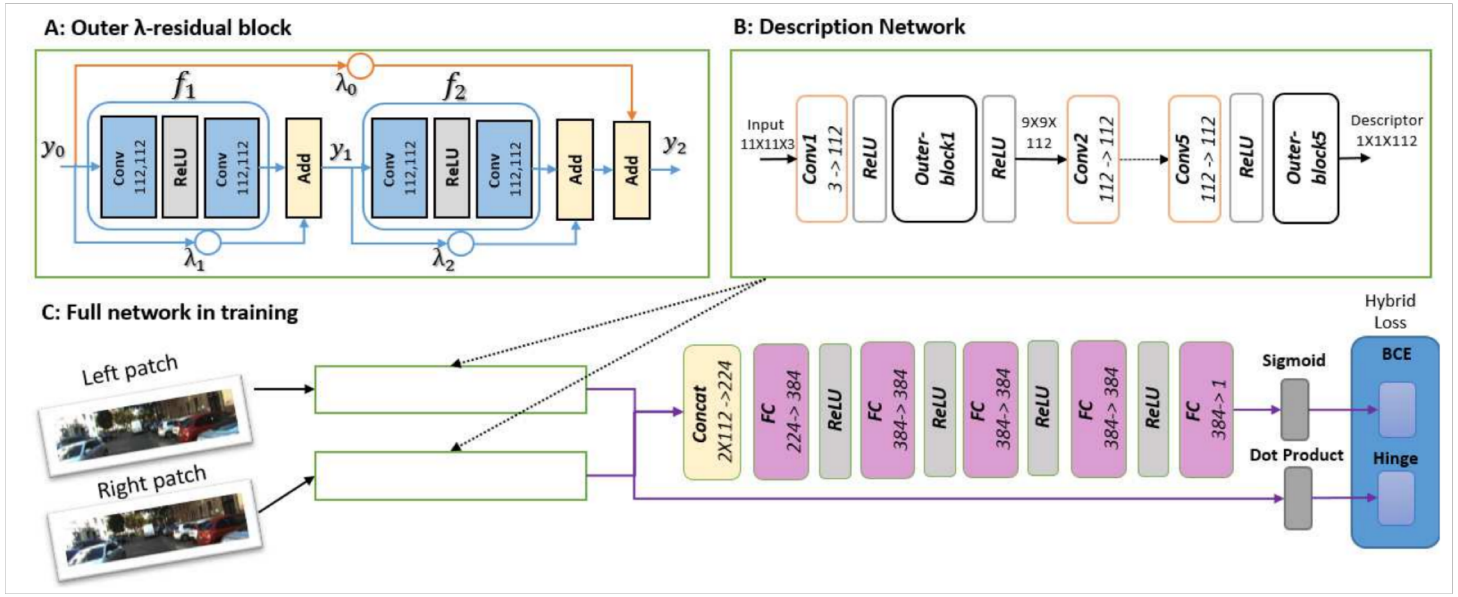


Conventional Methods

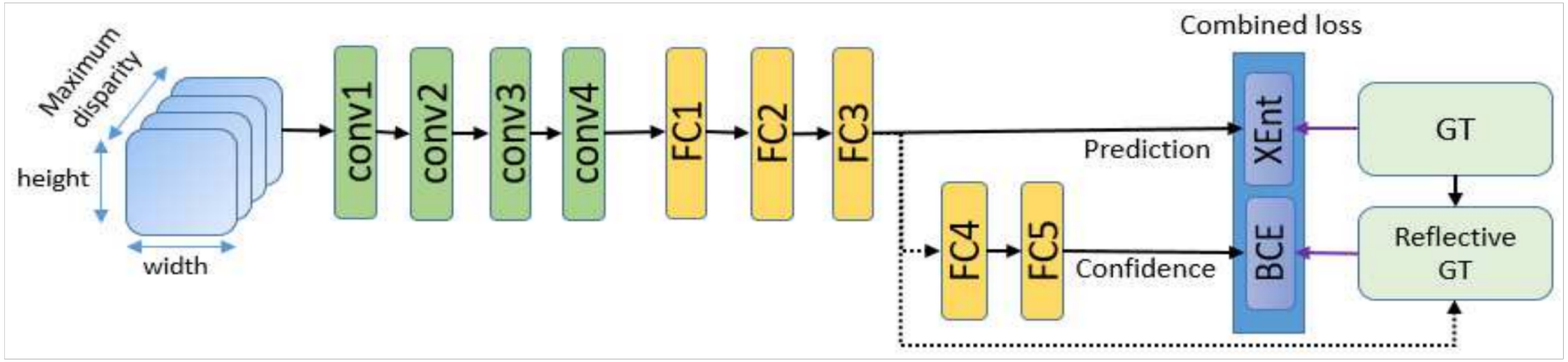


$$C_{\text{CNN}}(\mathbf{p}, d) = f_{\text{neg}}(< \mathcal{P}^{L}_{9\times9}(\mathbf{p}), \mathcal{P}^{R}_{9\times9}(\mathbf{pd}) >)$$





Conventional Methods

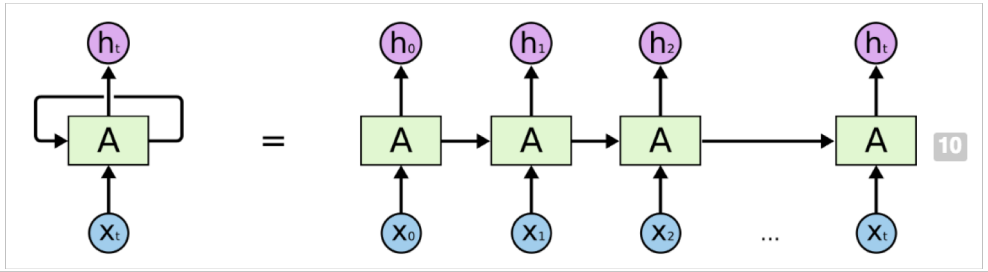


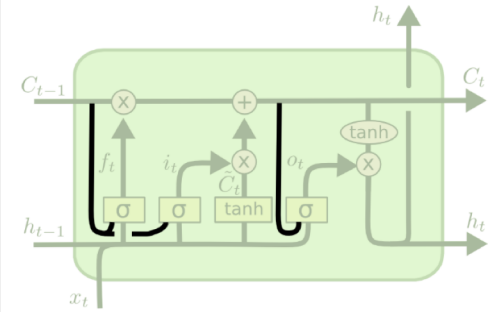
Contribution

- A novel recurrent model for better handling stereo disparity estimation tasks
- A soft attention mechanism is introduced to guide the network to automatically focus more on the unreliable regions
- Perform extensive experiments on KITTI 2015, SceneFlow and Middlebury, and achieve the state-of-the-art results.

Conv-LSTM

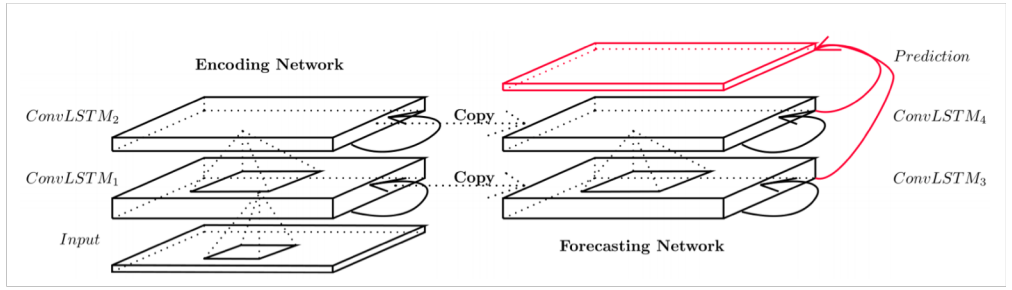
LSTM





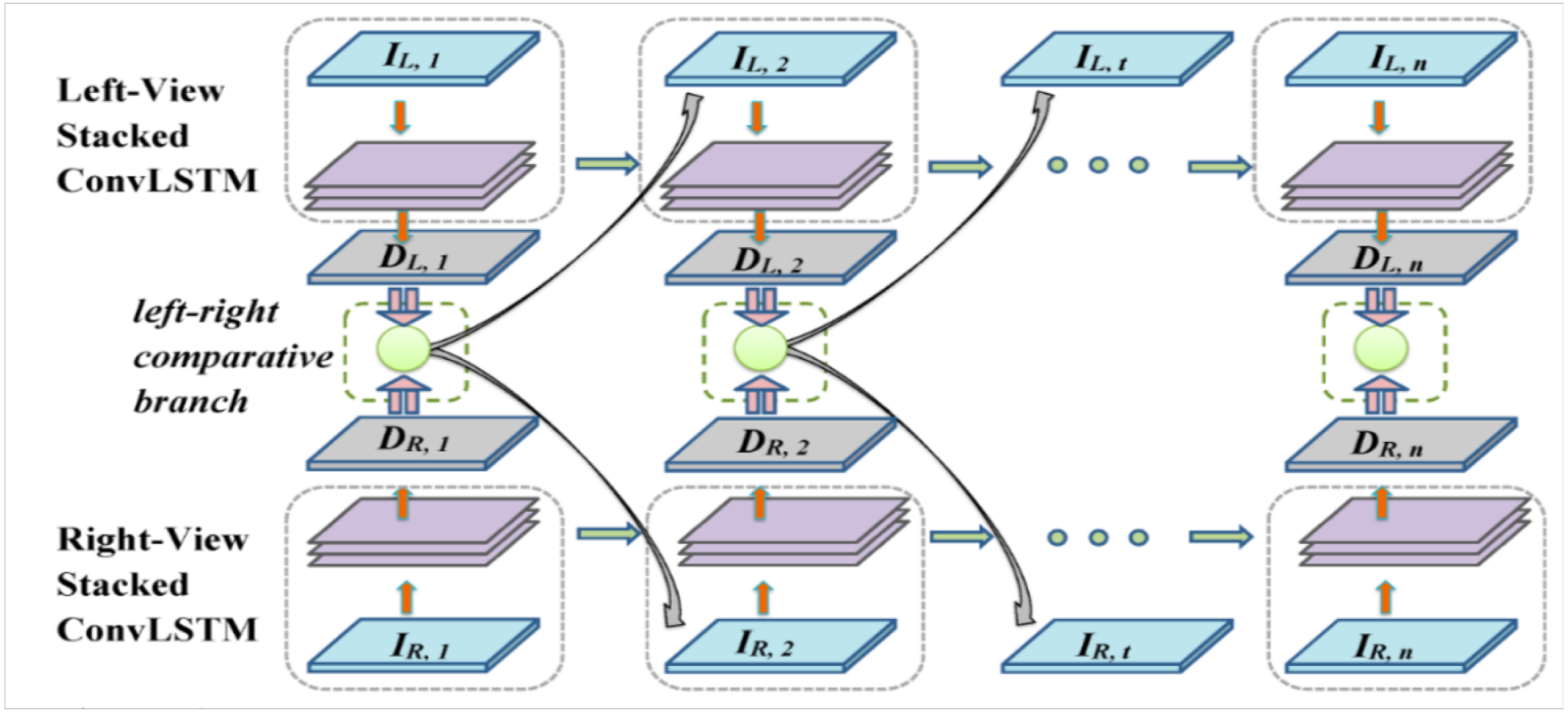
$$f_t = \sigma\left(W_f \cdot [\boldsymbol{C_{t-1}}, h_{t-1}, x_t] \; + \; b_f\right)$$
$$i_t = \sigma\left(W_i \cdot [\boldsymbol{C_{t-1}}, h_{t-1}, x_t] \; + \; b_i\right)$$
$$o_t = \sigma\left(W_o \cdot [\boldsymbol{C_t}, h_{t-1}, x_t] \; + \; b_o\right)$$

Conv-LSTM

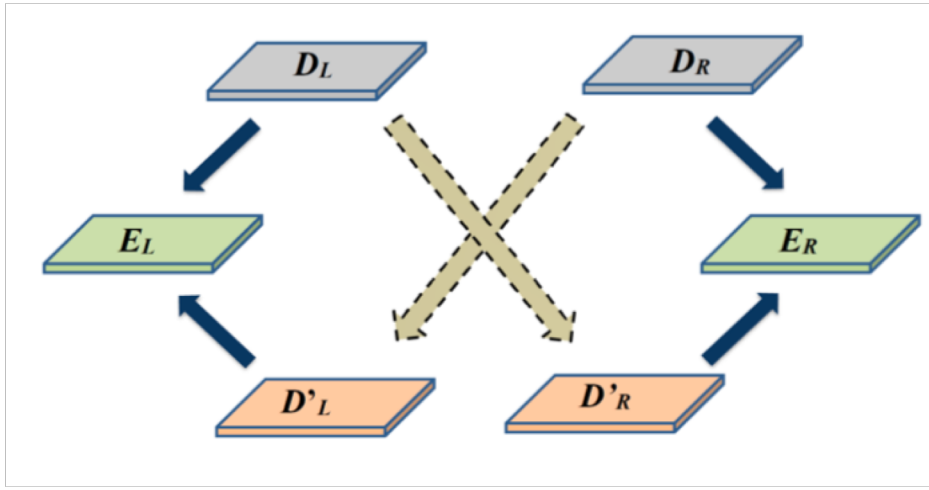


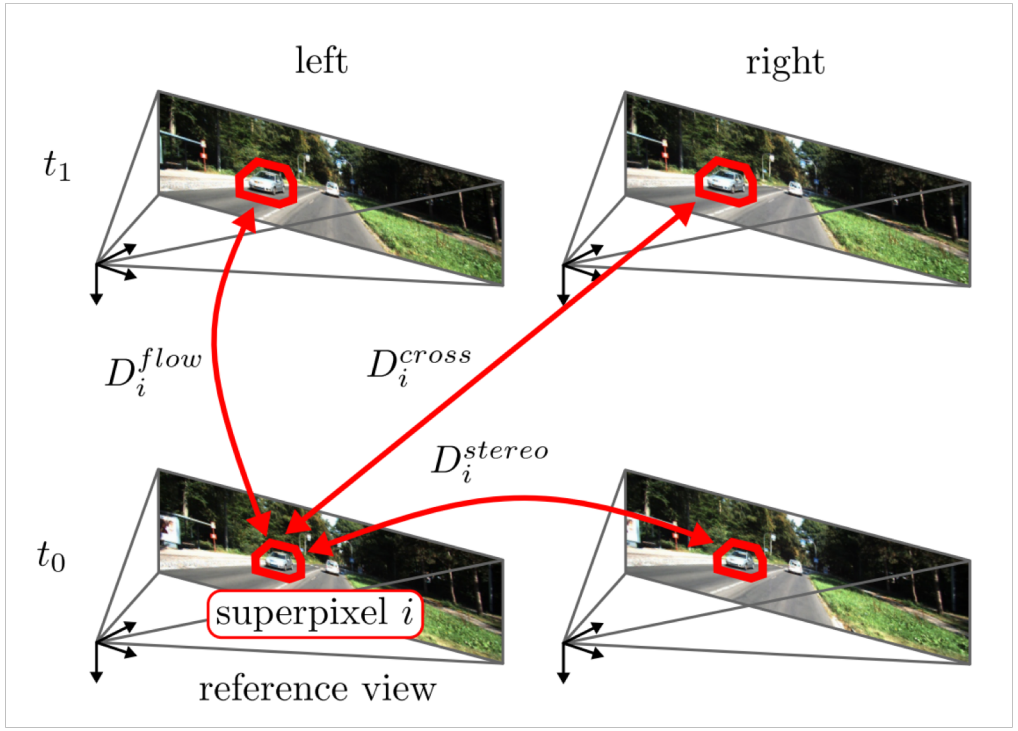
$$i_t = \sigma(W_{xi} * \mathcal{X}_t + W_{hi} * \mathcal{H}_{t-1} + W_{ci} \circ \mathcal{C}_{t-1} + b_i)$$
$$f_t = \sigma(W_{xf} * \mathcal{X}_t + W_{hf} * \mathcal{H}_{t-1} + W_{cf} \circ \mathcal{C}_{t-1} + b_f)$$
$$\mathcal{C}_t = f_t \circ \mathcal{C}_{t-1} + i_t \circ \tanh(W_{xc} * \mathcal{X}_t + W_{hc} * \mathcal{H}_{t-1} + b_c)$$
$$o_t = \sigma(W_{xo} * \mathcal{X}_t + W_{ho} * \mathcal{H}_{t-1} + W_{co} \circ \mathcal{C}_t + b_o)$$
$$\mathcal{H}_t = o_t \circ \tanh(\mathcal{C}_t)$$

Method


Left-View Stacked ConvLSTM
left-right comparative branch
Right-View Stacked ConvLSTM
$I_{L,1}$
$I_{L,2}$
$I_{L,t}$
$I_{L,n}$
$D_{L,1}$
$D_{L,2}$
$D_{L,n}$
$D_{R,1}$
$D_{R,2}$
$D_{R,n}$
$I_{R,1}$
$I_{R,2}$
$I_{R,t}$
$I_{R,n}$

Method


D_L
D_R
E_L
E_R
D'_L
D'_R


left
right
t_1
t_0
D_i^{flow}
D_i^{cross}
D_i^{stereo}
superpixel i
reference view

Result

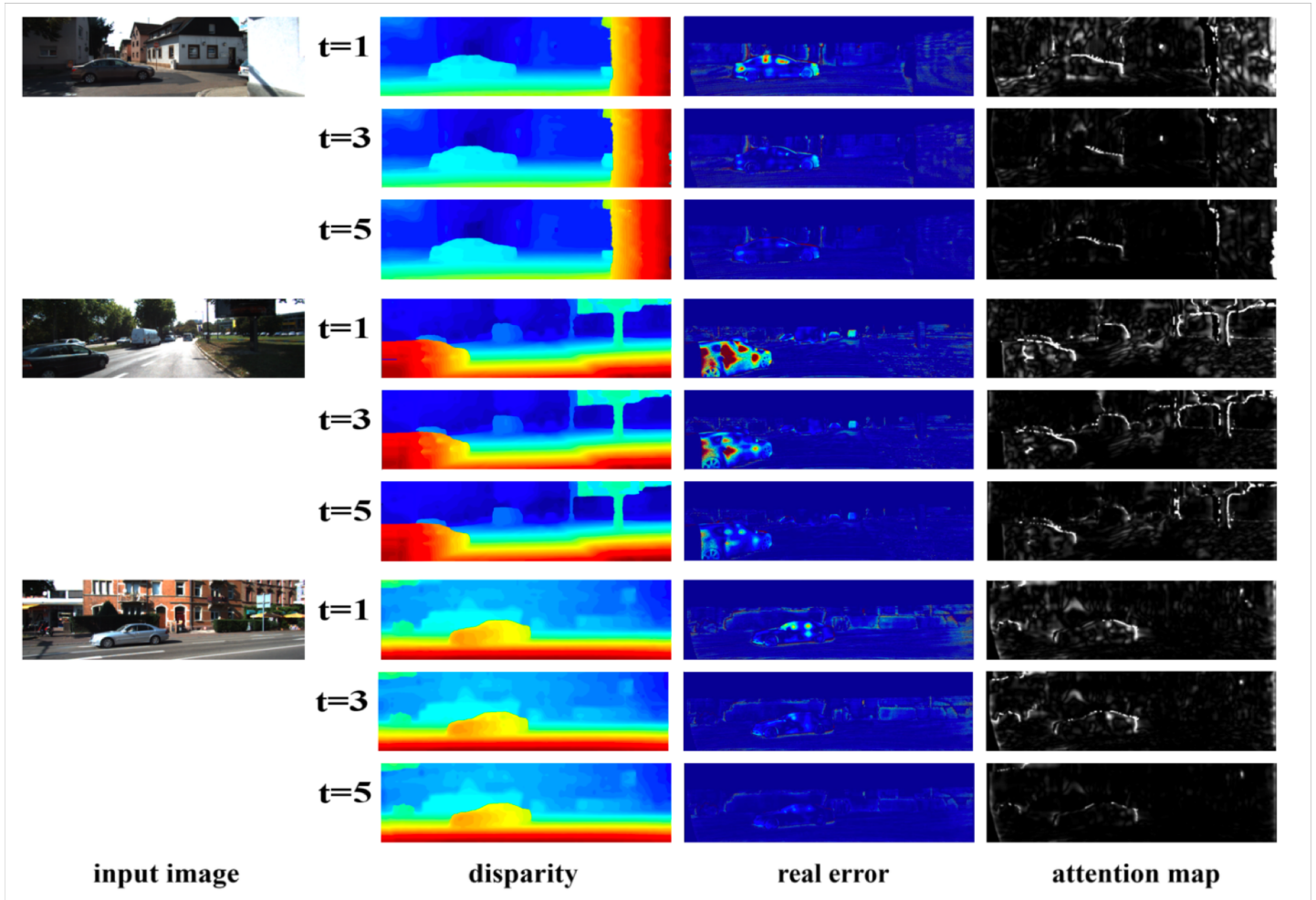
end point error and pixel per- centages with errors larger than 2, 3 and 5 pixels

Model	Recurrent Type	>2 px	>3 px	>5 px	EPE
WTA	–	7.94	4.78	3.11	1.144
Non-recurrent	–	5.70	3.69	2.44	0.939
t=1	w/o comp	6.24	3.92	2.54	0.972
	w/o atten	6.15	3.87	2.53	0.969
	LRCR	6.20	3.90	2.54	0.971
t=2	w/o comp	5.88	3.74	2.46	0.950
	w/o atten	5.79	3.72	2.45	0.945
	LRCR	5.61	3.66	2.42	0.935
t=3	w/o comp	5.64	3.67	2.43	0.936
	w/o atten	5.05	3.44	2.29	0.891
	LRCR	4.85	3.35	2.15	0.877
t=4	w/o comp	5.43	3.61	2.38	0.922
	w/o atten	4.54	3.22	2.08	0.855
	LRCR	4.33	3.12	2.02	0.839
t=5	w/o comp	5.32	3.58	2.36	0.913
	w/o atten	4.14	3.05	1.99	0.825
	LRCR	3.92	2.96	1.93	0.806

error rates of the non-occluded pixels and all the pixels

Type	Method	NOC	ALL	Runtime
Others	MC-CNN-acrt [35]	3.33	3.89	67s
	Content-CNN [15]	4.00	4.54	1s
	Displets v2 [7]	3.09	3.43	265s
	DRR [6]	2.76	3.16	1.4s
End-to-end CNN	GC-NET [13]	2.61	2.87	0.9s
	CRL [21]	2.45	2.67	0.47s
LRCR and the baseline	λ-ResMatch (fast) [29]	3.29	3.78	2.8s
	λ-ResMatch (hybrid) [29]	2.91	3.42	48s
	Ours (fast)	2.79	3.31	4s
	Ours (hybrid)	2.55	3.03	49.2s

Result


t=1
t=3
t=5
t=1
t=3
t=5
t=1
t=3
t=5
input image
disparity
real error
attention map

Reference

- Computing the stereo Matching Cost with a convolutional Neural Network, CVPR 2015
- Object Scene Flow for Autonomous Vehicles, CVPR 2015
- Convolutional LSTM Network:A Machine Learning Approach for Precipitation Newcasting, NIPS 2015
- Improved Stereo Matching with Constant Highway Networks and Reflective Confidence Learning, CVPR 2017