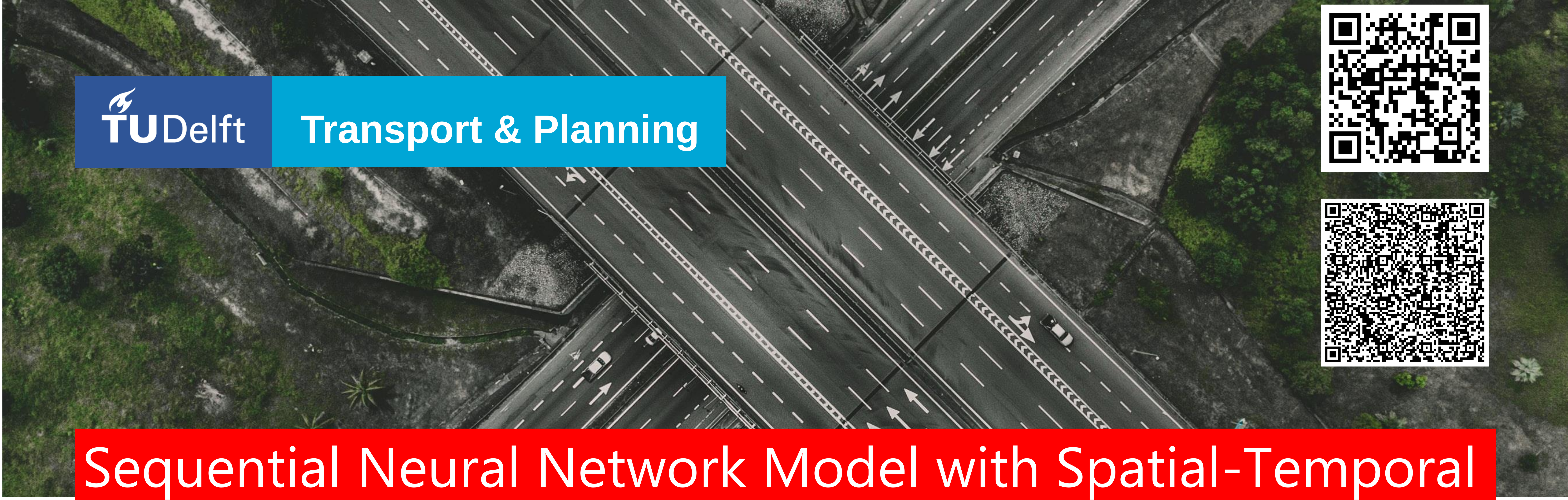
Sequential Neural Network Model with Spatial-Temporal Attention Mechanism for Robust Lane Detection Using Multi Continuous Image Frames

Dong, Y.; Patil, Sandeep; Farah, H.; Hellendoorn, J.

Publication date
2023

Document Version
Final published version

Citation (APA)
Dong, Y., Patil, S., Farah, H., & Hellendoorn, J. (2023). *Sequential Neural Network Model with Spatial-Temporal Attention Mechanism for Robust Lane Detection Using Multi Continuous Image Frames*. 1. Poster session presented at Transportation Research Board 102nd Annual Meeting 2023, Washington, District of Columbia, United States.



Transport & Planning

Sequential Neural Network Model with Spatial-Temporal Attention Mechanism for Robust Lane Detection Using Multi Continuous Image Frames

THE AUTHORS

Yongqi Dong
y.dong-4@tudelft.nl
Sandeep Patil
sandeeppatil538@gmail.com
Haneen Farah
h.farah@tudelft.nl
Hans Hellendoorn
j.hellendoorn@tudelft.nl
Delft University of Technology

Lane detection serves as a fundamental task for automated vehicles and Advanced Driver Assistance Systems. This study develops a novel sequential neural network model with a spatial-temporal attention mechanism that can focus on key features of lane lines and exploit salient spatial-temporal correlations among continuous image frames for the purpose of enhancing the accuracy and robustness of lane detection. Experiments demonstrate the strength and the robustness of the proposed model outperforming available state-of-the-art methods in various testing.

METHODS

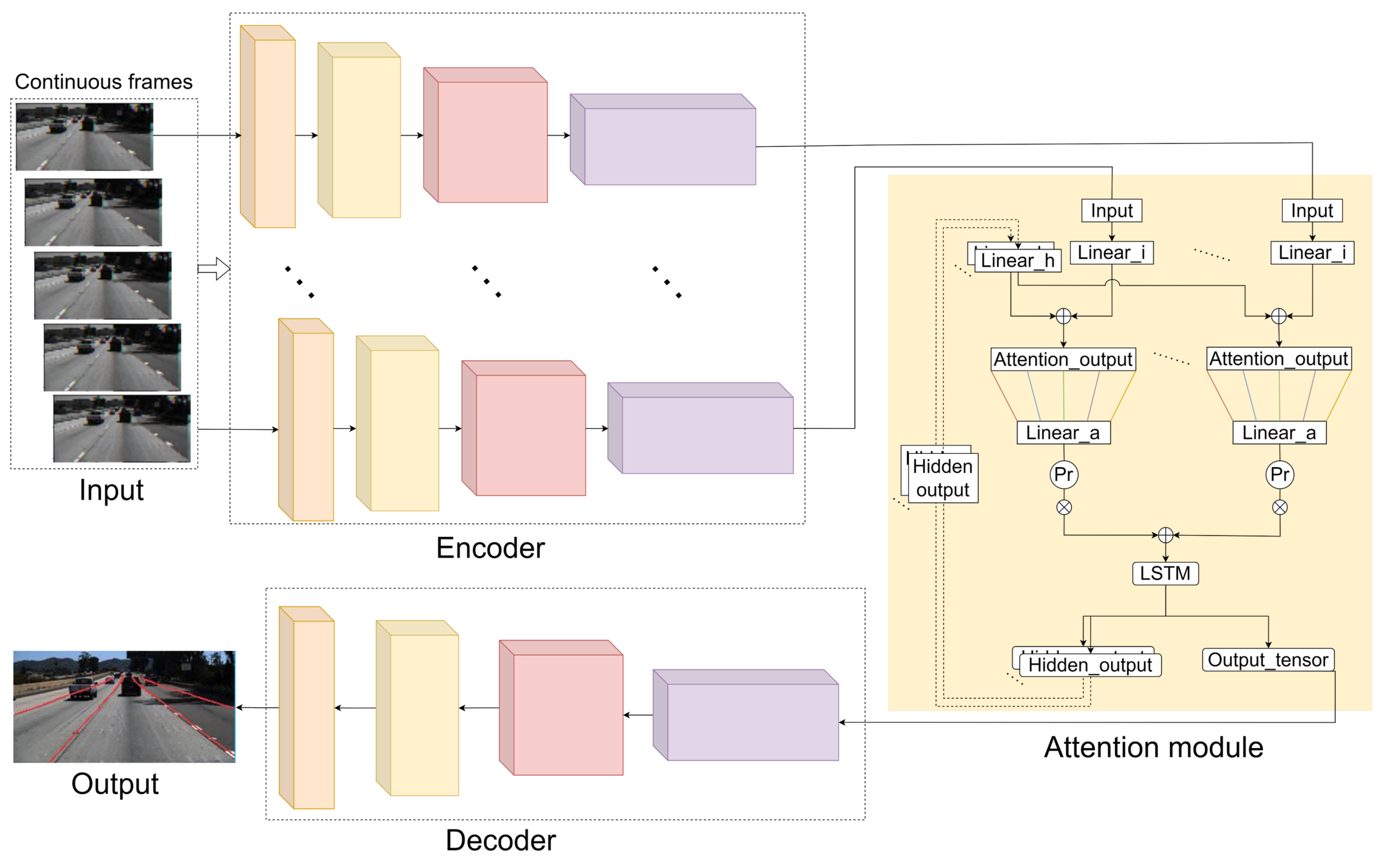


Figure 1. The architecture of the proposed model.

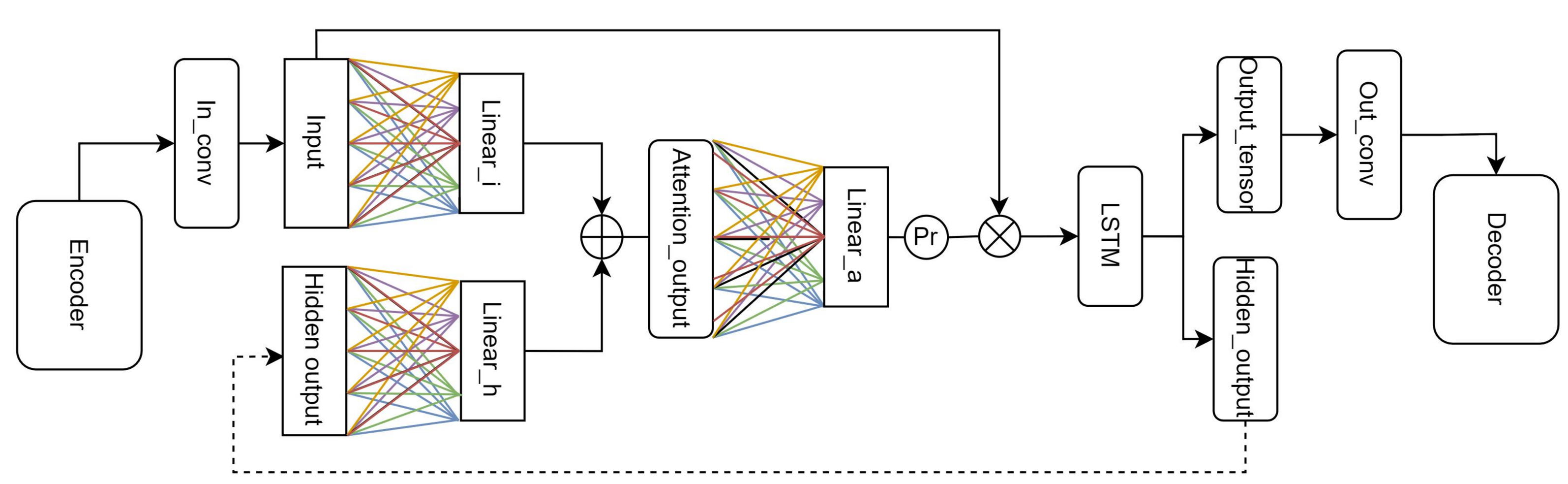


Figure 2. Illustration of spatial-temporal attention with fully connected layers (STFC_Att).

Spatial-temporal attention mechanism

$$x^{(t+n)} = \left(x_{down4}^{(t+n)} * k_{in}\right) \quad (1)$$

$$z^{(t+n)} = \left(Ux^{(t+n)} * Vh\right) \quad (2)$$

$$w^{(t+n)} = softmax\left(Wz^{(t+n)}\right) \quad (3)$$

$$\overline{x^{(t)}} = \sum_{n=1}^{N} w^{(t+n)} x^{(t+n)} \quad (4)$$

After processing the $\boldsymbol{N}$ images, and getting the average $\overline{x^{(t)}}$ for the selected sequence,

$$o^{(t+N)}, h^{(t+1+N)} = F(\overline{x^{(t)}}, h) \quad (5)$$

$$x_{out} = (o^{(t+N)} * k_{out}) \quad (6)$$

$$h = h^{(t+1+N)} \quad (7)$$

RESULTS

Table 1 Quantitative comparison on tvtLane testset #1 (normal)

		Test_Acc (%)	Precision	Recall	F1-Measure	MACs (G)	Para (M
	Baseline Models						
Models using **single image**	U-Net	96.54	0.790	**0.985**	0.877	15.5	13.4
	SegNet	96.93	0.796	0.962	0.871	50.2	29.4
	SCNN*	96.79	0.654	0.808	0.722	77.7	19.2
	LaneNet*	97.94	0.875	0.927	0.901	44.5	19.7
Models using **continuous images sequence**	SegNet_ConvLSTM	97.92	0.874	0.931	0.901	217.0	67.2
	UNet_ConvLSTM	98.00	0.857	0.958	0.904	69.0	51.1
	Proposed Models						
	Tem_Att-UNet_LSTM	98.08	0.877	0.936	0.906	44.7	13.5
	ST_Att-UNet_LSTM	98.09	0.879	0.941	0.909	44.8	13.5
	STFC_Att-UNet_LSTM	**98.14**	**0.887**	0.941	**0.911**	44.9	13.5
	*STFC_Att-SCNN_UNet_LSTM***	***98.20***	***0.906***	*0.936*	***0.921***	*68.9*	*13.7*

Table 2 Quantitative comparison on tvtLane testset #2 (12 challenging scenes)

PRECISION												
Models / **Challenging Scenes**	**1-curve & occlude**	**2-shadow**	**3-bright**	**4-occlude**	**5-curve**	**6-dirty & occlude**	**7-urban**	**8-blur & curve**	**9-blur**	**10-shadow**	**11-tunnel**	**12-dim & occlude**
U-Net	0.7018	0.7441	0.6717	0.6517	0.7443	0.3994	0.4422	0.7612	0.8523	0.7881	0.7009	0.5968
SegNet	0.6810	0.7067	0.5987	0.5132	0.7738	0.2431	0.3195	0.6642	0.7091	0.7499	0.6225	0.6463
UNet_ConvLSTM	0.7591	0.8292	0.7971	0.6509	0.8845	0.4513	0.5148	**0.8290**	**0.9484**	0.9358	0.7926	**0.8402**
SegNet_ConvLSTM	0.8176	0.8020	0.7200	0.6688	0.8645	**0.5724**	0.4861	0.7988	0.8378	0.8832	0.7733	0.8052
Tem_Att-UNet_LSTM	**0.8430**	**0.8909**	0.7732	0.5740	0.8322	0.4692	0.4567	0.8358	0.8090	0.9244	0.7893	0.8046
ST_Att-UNet_LSTM	0.7938	0.8743	**0.8013**	0.7014	**0.8894**	0.5215	**0.4935**	0.8290	0.8517	0.9286	0.7516	0.8218
STFC_Att-UNet_LSTM	0.8239	0.8782	0.7646	**0.7031**	0.8871	0.5295	0.4848	0.7354	0.9023	**0.9395**	**0.8794**	0.7542
F1-MEASURE												
U-Net	0.8200	0.8408	0.7946	0.7337	0.7827	0.3698	0.5658	0.8147	0.7715	0.6619	0.5740	0.4646
SegNet	0.8042	0.7900	0.7023	0.6127	0.8639	0.2110	0.4267	0.7396	0.7286	0.7675	0.6935	0.5822
UNet_ConvLSTM	0.8465	**0.8891**	**0.8411**	0.7245	0.8662	0.2417	**0.5682**	0.8323	**0.7852**	0.6404	0.4741	0.5718
SegNet_ConvLSTM	0.8852	0.8544	0.7688	0.6878	**0.9069**	**0.4128**	0.5317	0.7873	0.7575	**0.8503**	**0.7865**	**0.7947**
Tem_Att-UNet_LSTM	**0.8933**	0.8657	0.8123	0.6513	0.8306	0.3530	0.5263	0.8290	0.7039	0.5338	0.5225	0.5226
ST_Att-UNet_LSTM	0.8548	0.8977	0.8253	0.7293	0.8254	0.3627	0.5543	**0.8369**	0.7480	0.6197	0.5522	0.5363
STFC_Att-UNet_LSTM	0.8690	0.9059	0.8314	**0.7456**	0.8086	0.3660	0.5277	0.7715	0.7329	0.6543	0.6471	0.5852

Table 3 Quantitative comparison on *TuSimple* testing set

		Test_Acc (%)	Precision	Recall	F1-Measure	MACs (G)	Params (M)
	Baseline Models						
Models using **continuous images sequence**	SegNet_ConvLSTM*	97.96	0.852	0.964	0.901	217.0	67.2
	UNet_ConvLSTM*	**98.22**	0.857	0.958	0.904	69.0	51.1
	UNet_DoubleConvGRU*	98.04	0.875	0.953	0.912	---	13.4
	Proposed Models						
	Tem_Att-UNet_LSTM	98.05	0.876	0.923	0.899	44.7	13.5
	ST_Att-UNet_LSTM	98.14	0.881	0.925	0.902	44.8	13.5
	STFC_Att-UNet_LSTM	***98.20***	**0.886**	0.950	**0.917**	44.9	13.5

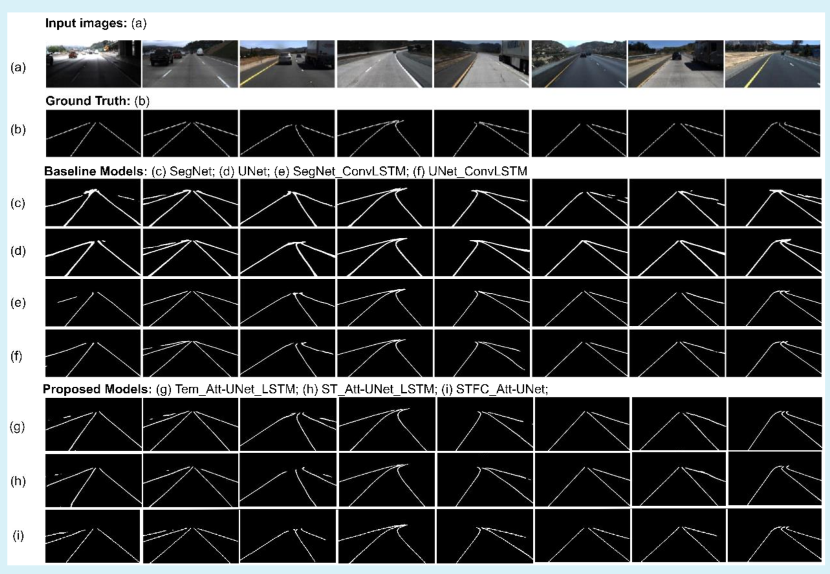


Figure 3. Qualitative comparison on tvtLANE testset #1 (normal).

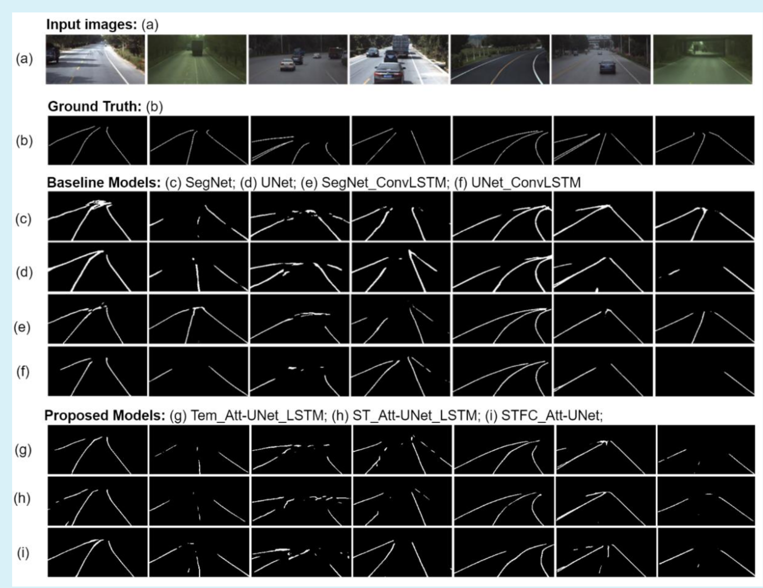


Figure 4. Qualitative comparison on tvtLANE testset #2 (challenging).

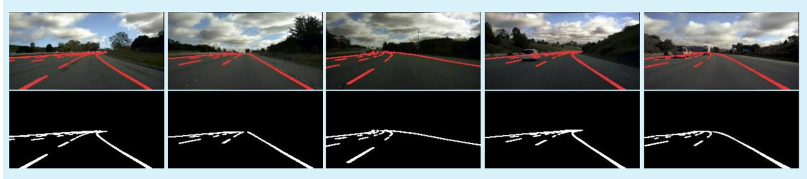


Figure 5. Qualitative evaluation results on the LLAMAS dataset.

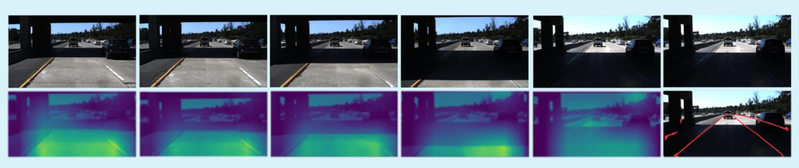


Figure 6. Post-explanation: visualization of lane detection under a bridge with shadow and occlusion.

CONCLUSION

- The proposed spatial-temporal attention mechanism can focus on key features of lane lines and exploit salient spatial-temporal relevance among continuous frames.
- The proposed model can cooperate with other mechanisms or model structures, and outperforms available state-of-the-art methods in various testing.
- The proposed model possess fewer parameters and smaller multiply-accumulate (MAC) operations.

Transport & Planning

SAMEN Safe and Efficient OperAtion of AutoMated and Human DrivEN Vehicles in Mixed Traffic

Lab Traffic and Transportation Safety