A Hybrid Spatial-temporal Sequence-to-one Neural Network Model for Lane Detection

Dong, Y.; Patil, Sandeep; Farah, H.; van Arem, B.

Publication date
2022

Document Version
Final published version

Citation (APA)
Dong, Y., Patil, S., Farah, H., & van Arem, B. (2022). *A Hybrid Spatial-temporal Sequence-to-one Neural Network Model for Lane Detection*. Poster session presented at Transportation Research Board 101st Annual Meeting 2022, Washington, District of Columbia, United States.

Delft University of Technology

A Hybrid Spatial-temporal Sequence-to-one Neural Network Model for Lane Detection

Authors: Yongqi Dong| Sandeep Patil | Bart van Arem | Haneen Farah

Event Number: 1415, Event Title: Research on Artificial Intelligence and Advancing Computing Applications, Event Location: Convention Center, Hall A
Event Time: Wednesday, Jan 12, 2022 10:30AM - 12:00PM, Presentation Number: TRBAM-22-03727, Poster-board Location Number: B583

Background & Aim

- ❖ Lane detection is crucial for Automated Vehicles and ADAS
- ❖ Available vision based methods usually use one image to do lane detection
- ❖ Traditional methods usually adopted cumbersome hand-crafted features
- ❖ Deep learning based methods in literature still can not make full use of spatio-temporal information
- ❖ Available methods can not handle challenging driving scenes

The main aim of this study is:

- ➢ **To develop robust detection model handling challenging driving scenes**
- ➢ **To deliver better feature extraction in every single image**
- ➢ **To make the most of spatio-temporal information in continuous frames**

FIGURE 1. Examples of Challenging Driving Scenes.

Proposed Deep Learning Model Architecture

- ➢ **End-to-end Encoder-decoder Structure**
- ➢ **Single Image Feature Extraction Module**
  - ▪ Encoder equipped with the spatial convolutional neural network (SCNN)
- ➢ **Spatial-temporal Feature Integration Module**
  - ▪ Constructed by spatial-temporal recurrent neural network (ST-RNN)
  - ▪ ConvLSTM and ConvGRU are employed and compared
- ➢ **Implementation with widely-used neural network backbones**
  - ▪ SegNet, UNet and its light version UNetLight are adopted
- ➢ **Tested and verified on two commonly used data set**
  - ▪ TuSimple and tvtLANE
- ➢ **12 proposed model variants and 6 baselines are evaluated and compared**

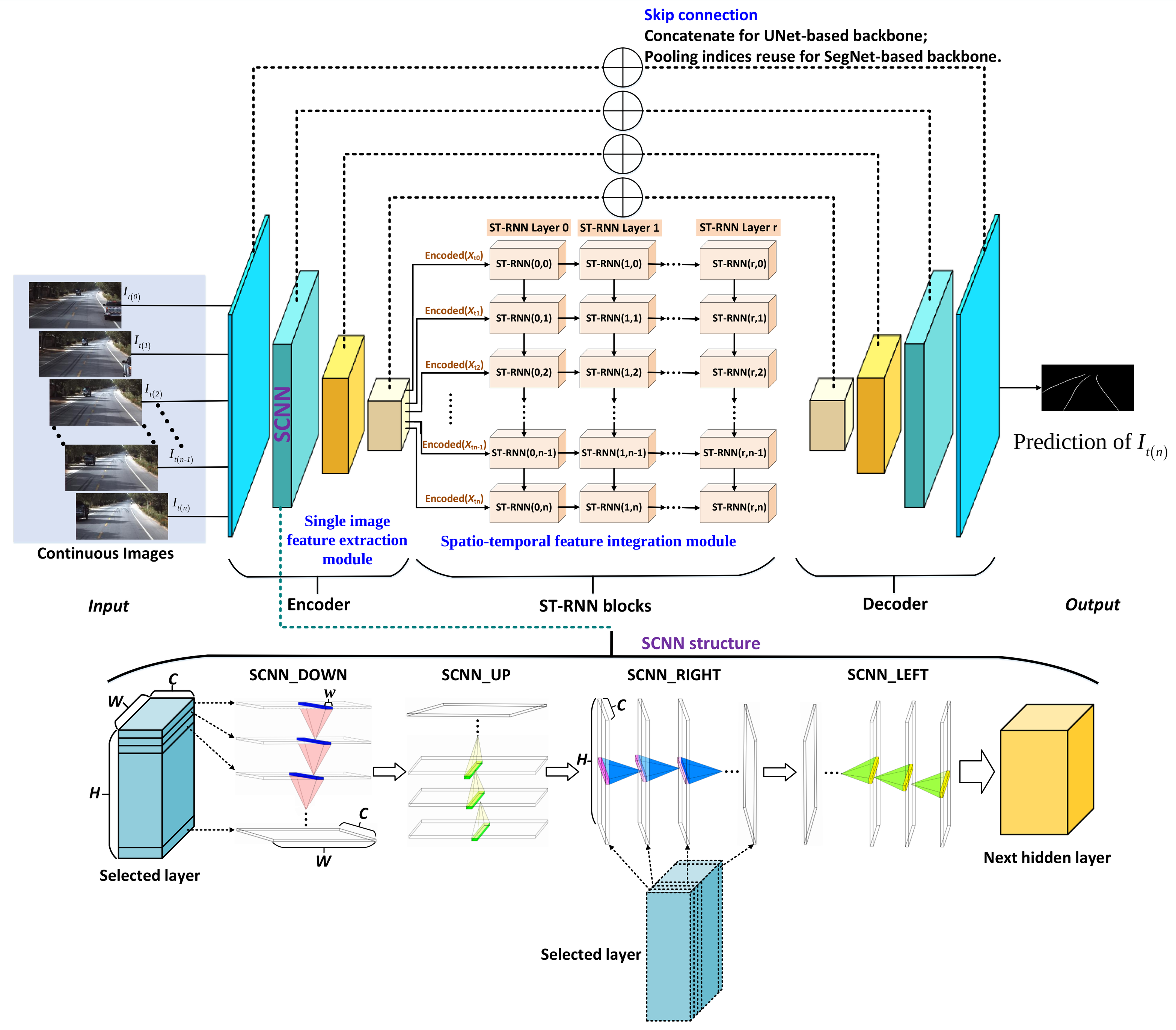

FIGURE 2. The architecture of the proposed hybrid spatial-temporal deep neural network

Evaluation Metrics

- ➢ **Accuracy**
- ➢ **Precision**
- ➢ **Parameter Size**
- ➢ **F1-Meassure**
- ➢ **Recall**
- ➢ **MACs (Multiply-accumulate operations)**

Results

		Test_Acc (%)	Precision	Recall	F1-Measure	MACs (G)	Params (M)
Models using **single image** as input	**Baseline Models**						
	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** as inputs	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 (SegNet-Based)						
	SCNN_SegNet_ConvGRU1	98.00	0.878	0.935	0.905	219.2	43.7
	SCNN_SegNet_ConvGRU2	98.05	0.888	0.918	0.903	221.5	57.9
	SCNN_SegNet_ConvLSTM1	98.01	0.881	0.935	0.907	220.0	48.5
	SCNN_SegNet_ConvLSTM2	98.07	**0.893**	0.928	0.910	223.0	67.3
	Proposed Models (UNet-Based)						
	SCNN_UNet_ConvGRU1	98.13	0.878	0.957	0.916	77.9	27.7
	SCNN_UNet_ConvGRU2	**98.19**	0.887	0.950	0.917	87.0	41.9
	SCNN_UNet_ConvLSTM1	98.18	0.886	0.948	0.916	81.0	32.4
	SCNN_UNet_ConvLSTM2	**98.19**	0.889	0.950	**0.918**	93.0	51.3
	Proposed Models (Light Version UNet-Based)						
	SCNN_UNetLight_ConvGRU1	97.83	0.850	0.960	0.902	19.6	**6.9**
	SCNN_UNetLight_ConvGRU2	98.01	0.863	0.950	0.905	21.9	10.5
	SCNN_UNetLight_ConvLSTM1	97.71	0.830	0.950	0.886	20.4	8.1
	SCNN_UNetLight_ConvLSTM2	97.76	0.840	0.953	0.893	23.4	12.8

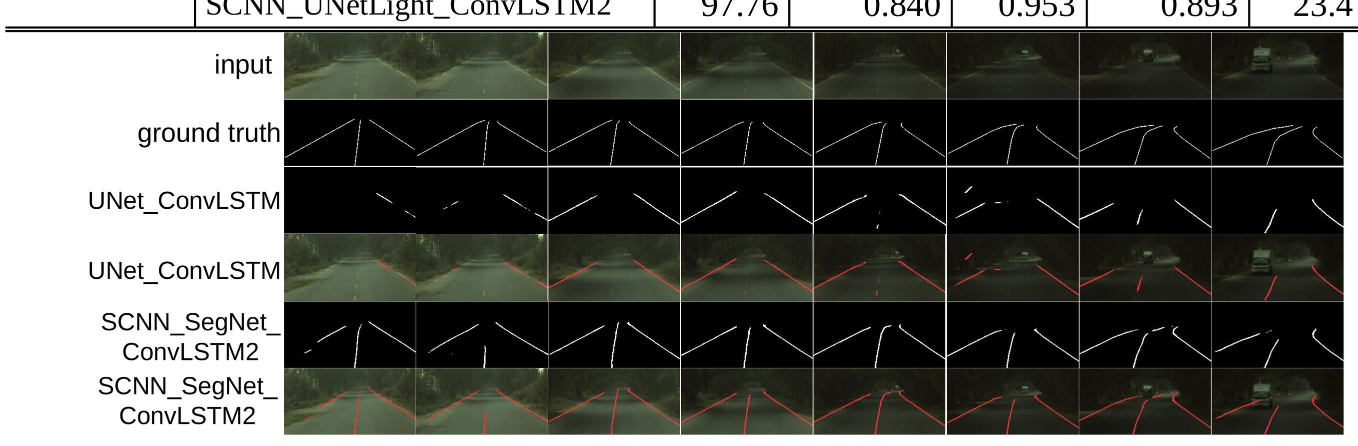

FIGURE 5. Case study of challenging scene 10 shadow-dark

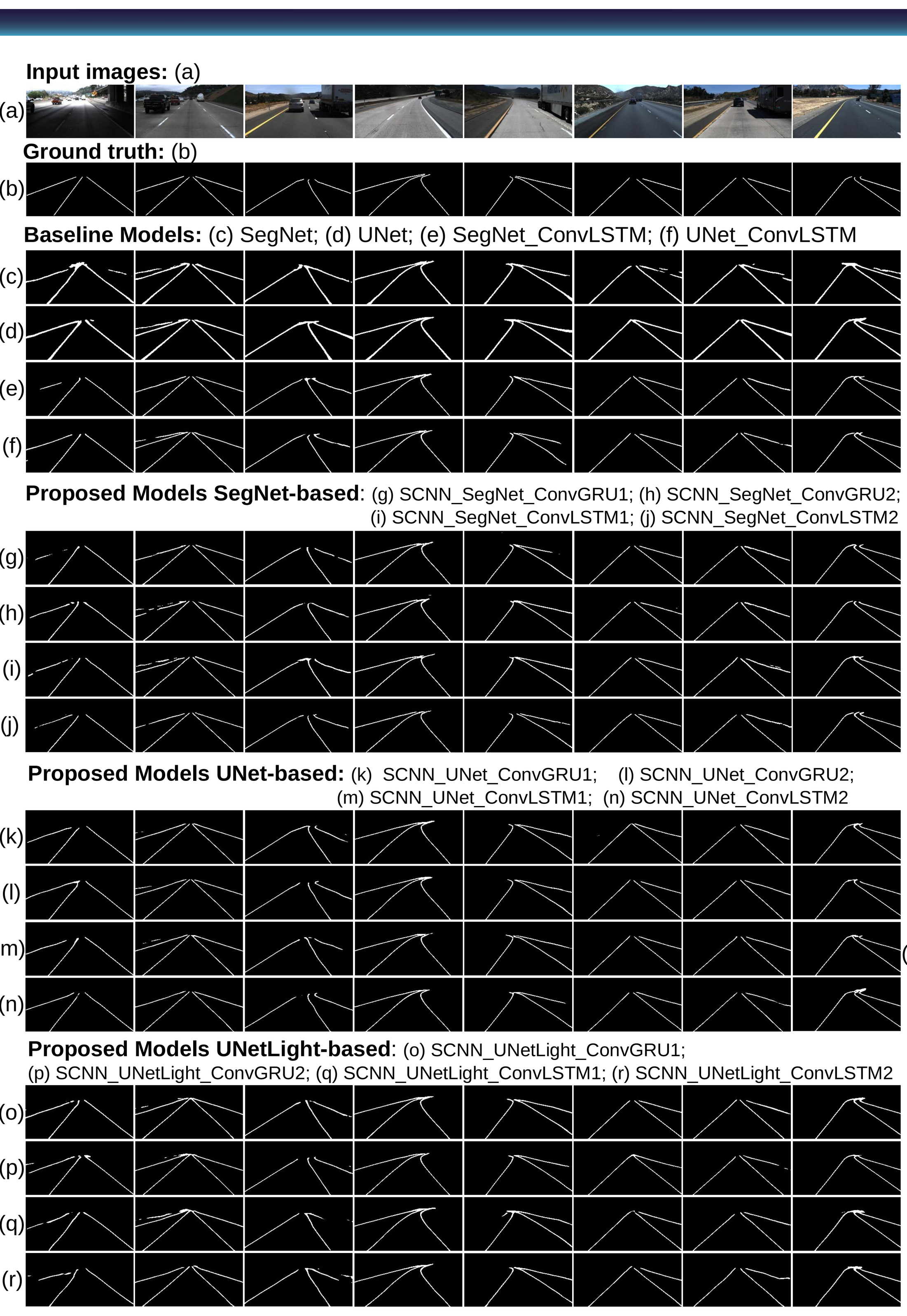

FIGURE 3. Visualization of lane-detection results on normal cases

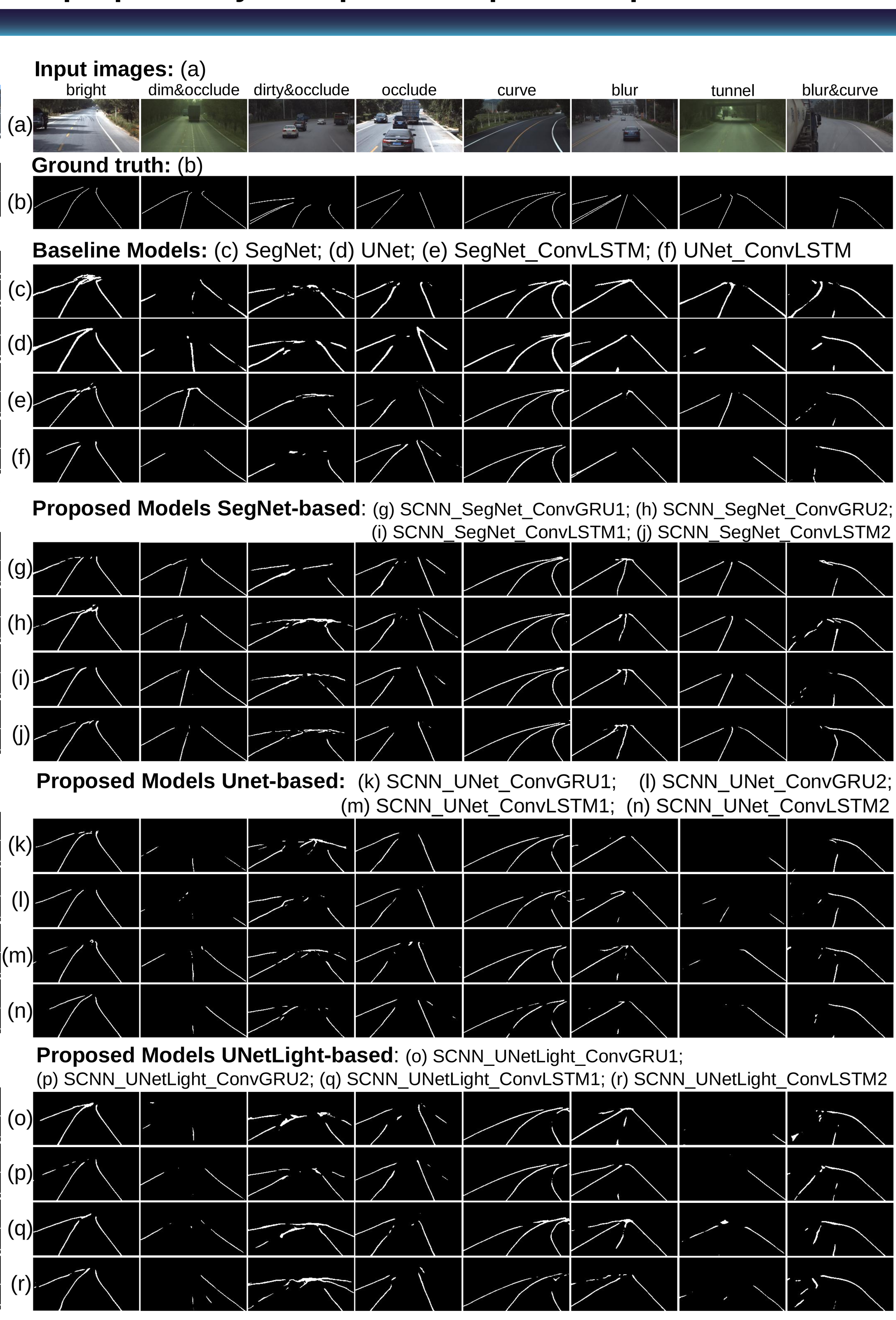

FIGURE 4. Visualization of lane-detection results on 8 challenging driving scenes

Ablation Study

Testing with, without and different locations of SCNN layer

Models (Testing Datasets)	Location of SCNN	Testset #1 (Normal Situations) Test_Acc (%)	Precision	Recall	F1-Measure	Testset #2 (Challenging Scenes) Test_Acc (%)	Precision	Recall	F1-Measure
SegNet_ConvLSTM	Without	97.92	0.874	0.931	0.901	97.83	0.756	0.765	0.761
SCNN_SegNet_ConvLSTM2	Conv1_1	98.00	0.884	0.921	0.902	97.92	0.757	0.757	0.757
	Conv2_1	98.07	0.893	0.928	0.910	97.90	0.767	0.766	0.767
UNet_Conv LSTM	Without	98.00	0.857	0.957	0.904	97.93	0.778	0.660	0.714
SCNN_UNet_ConvLSTM2	In_Conv_1	98.28	0.896	0.939	0.917	98.08	0.776	0.593	0.672
	Conv1_1	98.19	0.889	0.950	0.918	97.95	0.778	0.640	0.702

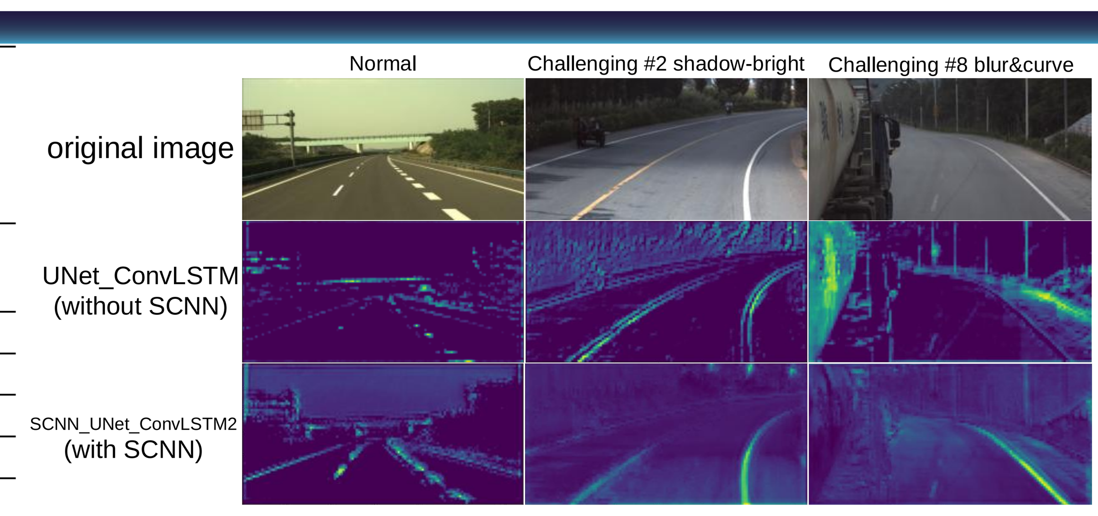

FIGURE 6. Visualization of extracted low-level features

Conclusions

- ➢ **The proposed model architecture is effective and robust beating SOTA baseline models with large margins in both normal and challenging cases**