

**Delft University of Technology** 

#### Robust Lane Detection through Self Pre-training with Masked Sequential Autoencoders and Fine-tuning with Customized PolyLoss

Dong, Y.; Li, Ruohan; Farah, H.

**Publication date** 2023

**Document Version** Final published version

**Citation (APA)** Dong, Y., Li, R., & Farah, H. (2023). *Robust Lane Detection through Self Pre-training with Masked* Sequential Autoencoders and Fine-tuning with Customized PolyLoss. 1. Poster session presented at Transportation Research Board 102nd Annual Meeting 2023, Washington, District of Columbia, United States.

#### Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright** Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology. For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

### **Robust Lane Detection through Self Pre-training with Masked Sequential Autoencoders** and Fine-tuning with Customized PolyLoss **TU**Delft Authors: Yongqi Dong | Ruohan Li | Haneen Farah

**Delft University of Technology** 

### Y.Dong-4@tudelft.nl



**Post-processing** 

Curve

Fitting

# **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
  - spatial-temporal information and correlation
- Available methods can not handle challenging driving scenes well
- The main aim of this study is:
- > To develop robust detection model handling challenging driving scenes
- > To make full use of valuable features and aggregate contextual information
- > To develop pre-training method for sequential vision based lane detection







### Figure 1. Examples of challenging driving scenes.

**Fine-tuning Segmentation** 

# The framework of the proposed pipeline

- End-to-end Encoder-decoder Structure
- Self Pre-training to Reconstruct Images
  - Masked sequential autoencoders
- Fine-tuning Segmentation
  - Transfer pre-trained model weights to the segmentation model
- Customized PolyLoss
- Post-processing with clustering & curve fitting
- Tested and verified on two data sets
  - tvtLANE normal (TuSimple lane)
  - tvtLANE challenging (12 cases)

# **Evaluation Metrics**

- Precision > Accuracy Parameter Size

PolyLoss Fine-tuning loss MSE Pre-training loss DBSCAN Clustering Output Label Output Label **Decoder CNN** Decoder CNN Outconv (64, 3) **Outconv** (64, 2) **Pre-trained** SCNN\_UNet\_ConvLSTM SCNN\_UNet\_ConvLSTM Model ConvLSTM **ConvLSTM** or or UNet\_ConvLSTM UNet\_ConvLSTM **Transfer Weights Encoder CNN Encoder CNN** Input Input Mask(50%) Preprocessin **Continuous Frames Continuous Frames** 

Figure 2. The framework of the proposed pipeline

Self Pre-training

## Results

Models		Test_Acc		Desell	F1-	MACs	Cs Params		<b>Input images:</b> (a)						<b>Input images</b> :(a)						
		(%)	Precision	Recall	Measure	e (G)	(M)	- ( )						and the second							and the second of the second second
	SegNet	Baseline Models					- (a) -														
Using		96.93	0.796	0.962	0.871	50.2	29.4	_	<b>Ground truth:</b> (b)						Ground tr	<b>uth</b> :b)					
single	UNet	96.54	0.790	0.985	0.877	15.5	13.4	– (b)													
image	SCNN*	96.79	0.654	0.808	0.722	77.7	19.2														
	LaneNet <sup>*</sup>	97.94	0.875	0.927	0.901	44.5	19.7	_	<b>Baseline Models:</b> (c)	SegNet, (d) UNe	et, (e) SegNet_	ConvLSTM, (	(f) UNet_Conv	ZLSTM.			egNet, (d) UN	let, (e) SegNet	t_ConvLSTM,	(f) UNet_Con	vLSTM.
	SegNet_ConvLSTM	97.92	0.874	0.931	0.901	217.0	67.2	(c)							$\sim$		$\sim$				
	UNet_ConvLSTM	98.00	0.857	0.958	0.904	69.0	51.1	-													
	UNet_ConvLSTM		<u>Pr</u>	re-trained	<b>J</b> Models			(d)							2		$\sim$	$\langle \rangle$			
	_CE**	98.19	0.882	0.940	0.910	69.0	51.1	- (e)							$\sim$				, '		
Using	UNet_ConvLSTM _PL <sup>**</sup>	98.34	0.921	0.909	0.915	69.0	51.1	(C) (f)													
	SCNN_SegNet		J	Baseline N	Models	<u> </u>		- (1)													
continuous	_ConvLSTM	98.07	0.893	0.928	0.910	223.0	67.3	-	<b>Pre-trained Models</b>	(g) UNet_Con	vLSTM_CE <sup>**</sup>	, (h) UNet_Co	nvLSTM_PL*	* • •	<b>Pre-traine</b>	d Models:	(g) UNet_Co	nvLSTM_CE <sup>*</sup>	<sup>**</sup> , (h) UNet_C	ConvLSTM_PL	** / •
images	SCNN_UNet _ConvLSTM	98.19	0.889	0.950	0.918		51.3	(g)													
	SCNN_UNet	Pre-trained Models					— (h)														
	_ConvLSTM_CE <sup>**</sup>	98.20	0.891	0.952		93.0	51.3	-	<b>Baseline Models:</b> (i)	SCNN_SegNet_	_ConvLSTM, (	(j) SCNN_UN	et_ConvLSTN	1.	Baseline N	<b>Iodels:</b> (i)	SCNN SegNe	et ConvLSTM	1. (i) SCNN_U	JNet_ConvLST	ΓM.
	SCNN_UNet _ConvLSTM_PL**	98.38	0.929	0.915			51.3	(i)													
	Inputs							- (j)	Pre-trained Models		ot ConvI STM		N LINet ConvI	STM DI **	Dro traino						
Marial															Pre-traine		(k) SCININ_UIN	et_ConvLSTW		N-UNet_Convi	LSIM_PL .

Masked images

# **Reconstruct images**

Figure 3. Visualization of the reconstructing results in the pre-training phase.

<b>Ablation Study</b>	Testing	Testse	et #1 (Nor	mal Situa	ations)	Testset #2 (Challenging Situations)				
	Datasets Models	Loss Function	Test_Acc (%)	Precision	F1- Measure	Loss Function	Test_Acc (%)	Precision	F1- Measure	
Testing different loss	UNAt Const STM	CE	98.19	0.882	0.910	CE	98.13	0.7932	0.6537	
functions and models	UNet_ConvLSTM	PL	98.34	0.921	0.915	PL	98.38	0.8331	0.6284	
	SCNN_UNet	CE	98.20	0.891	0.921	CE	98.03	0.8001	0.7327	
	_ConvLSTM	PL	98.38	0.929	0.922	PL	98.36	0.8444	0.6711	

#### Figure 4. Visualization of lane-detection results on

normal cases.

Figure 5. Visualization of lane-detection results on 7 challenging driving scenes.

# Conclusions

- The proposed masked sequential autoencoder based pre-training and customized PolyLoss are useful
- > The proposed pipeline is effective and robust which can improve the performances of SOTA models in **both normal and challenging cases**







