Lane Detection with Spatial Convolutional Neural Networks

1^{st} Soumya Bharatha *
Computer Science of Engineering (AI-ML)
SR UNIVERSITY
Warangal,India
Mail : 2203A52007@sru.edu.in

***Abstract*—Lane detection is among the most important elements of ADAS (Advanced driver assistance systems) and self-driving car technologies where vehicles have to follow the lane and avoid drifting off the lane. This paper proposes a new lane detection algorithm based on Spatial Convolutional Neural Networks (Spatial CNN) on the CULane and TuSimple datasets, which has many images containing roads under various situations. The features of the proposed method that enhance the presence of spatial dependencies and context in the road scene, improve the accuracy of lane detection compared to the methods used. Thus, the presented results expand the existing knowledge of lane detection, and the proposed approach can be used as a solid basis for further development of autonomous driving systems. Therefore, the combination of Spatial CNN with the CULane and TuSimple datasets lays down the best performance for lane detection to improve vehicle safeguarding and driving effectiveness.**

***Index Terms*—Spatial Convolution Neural Network (SCNN), Image Segmentation, Data Augmentation, Occlusion handling, Lane marking detection.**

I. INTRODUCTION

Lane detection is thus a critical feature that helps in improving the ADAS as well as autonomous driving vehicles. It involves the ability to identify and have mean of tracing the lanes on the roads so that vehicles may follow while in use. Especially for some of the features like lane departure warning, lane keeping assist, and adaptive cruise control, Lane detection is very crucial. However, the following factors make the task challenging; the condition of the roads, changes in luminosity, occlusion and the pattern of roads.

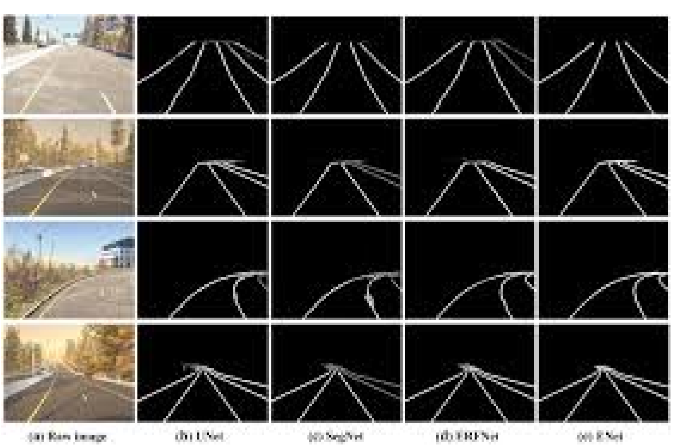


Fig. 1.

Recently, some of the effective techniques provided for lane detection are based on deep learning. From all the above, SCNN and U-net models have become favoured due to their ability to regenerate the spatial dependency and context in road scenes.

On the other hand, for the biomedical image segmentation, a U-net model was developed excelling all the other models for segmentation tasks due to the efficiency of its encoder-decoder architecture with the skip connection. The proposed U-Net can segment lane markings with high accuracy, which is because it incorporates both low-level and high-level features applied within the lane detection process.

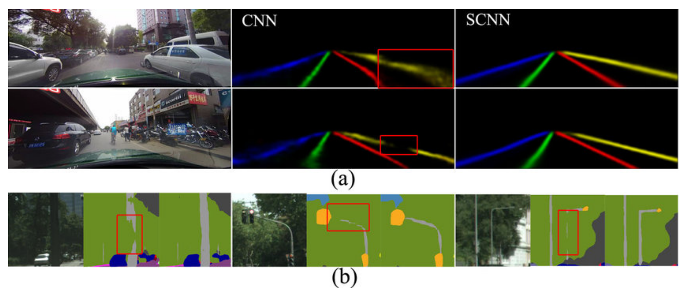


Fig. 2. The overviews of the CNN and SCNN based in (a) lane detection and (b) semantic segmentation are shown in Figure 2 above. For each example, from left to right are the input image, CNN feature map, and SCNN feature map. It could be noticed that SCNN can better fit the long continuous shape prior to lane markings and poles and helped to fix the disconnected fragments in CNN.

This paper's purpose is to consider the options for employing the SCNN and U-Net modes utilizing the CULane and TuSimple databases, which consist of pictures taken in various road contexts. The challenge arises from street lights, shadows for the same at night time, and occlusion for the CULane dataset while the TuSimple dataset provided clear images for comparison with other lane detection models. The models SCNN and U-Net are trained, tested and compared on these datasets based on the measures of accuracy, stability, and the amount of computations performed.

Currently, the emerging topic of autonomous vehicles has garnered interest across both academic and industrial domains. Another one of them is the traffic scene understanding, which is one of the most difficult tasks of autonomous driving and includes such computer vision tasks as lane detection and semantic segmentation. Lane detection assists in steering the vehicle and can be applied in driving assistance applications (Urmson et al. 2008), whereas semantic segmentation gives enhanced approximate positions of nearby objects like other vehicles and people. These tasks could be very challenging in real applications taking into consideration the multitude of harsh conditions such as bad light, rain or even fog, etc As has been observed in traffic scenes most of the time one is faced with objects that have strong structural priors but few appearance cues such as lane markings and poles which have long continuous contour and might be occluded. As seen in Fig. 1 the above points can be illustrated by the first example, in which the entire right side of the car that appears fully occludes the rightmost lane marking. To solve this problem, we introduce Spatial CNN (SCNN) – the generalization of the DCNN to the high levels of spatial structure. In a layer-by-layer CNN, a convolution layer takes in the output from the previous layer, conveys filter convolution and nonlinear activation and forwards the output to the next layer. All these steps are carried out sequentially. Similarly, SCNN treats rows or columns of feature maps as layers and passes through them convolution, nonlinear activation, and sum which makes SCNN form a deep neural network. In this manner, information could be passed between neurons within the same layer, which meant that profundity could be implemented.

II. LITERATURE REVIEW

Pan et al.(2018) [1] proposed a Spatial Convolutional Neural Network (SCNN), whose purpose is to capture the long-range dependencies by spreading the spatial information regarding the rows and the columns of the image. This architecture is particularly effective for the lane marking detection scenario based on the nature of what it detects. The proposed SCNN model was tested on the newly introduced CULane dataset in particularly complex conditions such as nighttime, shadowing and occlusion. Moreover, our SCNN won 1st place on the TuSimple Benchmark Lane Detection Challenge, with an accuracy of 96.53 Percentage.

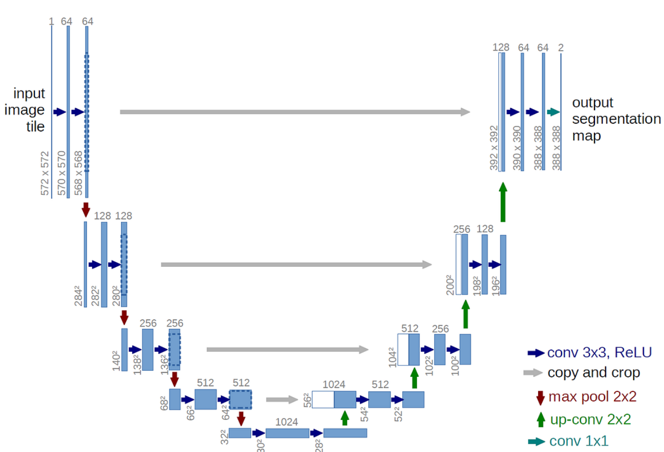


Fig. 3. . U-net architecture (example for 32x32 pixels in the lowest resolution)

U-Net architecture was developed by Ronneberger et al., 2015, [2] for the segmentation of biomedical images which obtained a significant success and can be applied to several image segmentation problems such as lane detection. U-Net design also incorporates encoder-decoder with skip connections that makes it more effective than others in the matters of segmentation of lane markings from the road images. The effectiveness of the suggested model has been confirmed on various datasets: the presented above TuSimple dataset has shown high accuracy of the model in the lane segmentation tasks.
Gaining from the readily available large-scale annotated datasets, the lane detection algorithms have been witnessed to have developed. The CULane dataset which was introduced by Pan et al. (2018) is a large-scale dataset containing end-to-end lane images from different conditions such as urban and rural roads, a highway, and different weather states. Similarly, TuSimple [3], also, is one of the most popular considered datasets to solve the lane detection problem, which contains high-quality images with high variation of shooting environment. It also comes with annotations for the lane markings, thus enabling one to train/evaluate from a sound ground truth point of view. The dataset of TuSimple has been widely employed by the researchers for performance evaluation of their algorithms.

In this step further processing [4] like curve fitting and smoothing are adopted for lane markings and removal of noise. Another recommended approach is adapting the pre-trained models to the specific domain by fine-tuning on relevant data. Image data used in lane detection work can be sourced from the pre-trained models like the ones trained on ImageNet, and one can fine-tune on CULane and/or TuSimple datasets.

III. METHODOLOGY

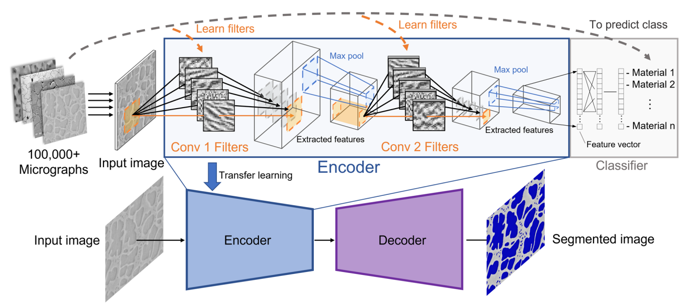


Fig. 4. Microstructure segmentation with deep learning encoders

As a result, in an aim to counteract the difficulties linked to lane detection in intricate and variable location conditions of a vehicle, a method which utilizes the Spatial Convolutional Neural Networks (SCNN) is introduced in this paper. Thus, our strategy should be sensitive enough to capture spatial dependencies as well as contextual information that is useful in identifying the lane markings. The methodology is structured as follows:The methodology is structured as follows:

1. Data Preprocessing: In this case, we start with data preprocessing of CULane and TuSimple datasets in order to get high-quality inputs for SCNN. This includes the following operations – resizing the images, standardizing the pixel intensities and data augmentation for which the model is trained to adapt to different conditions of the roads.

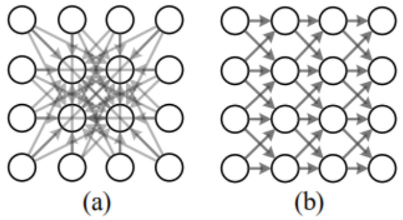


Figure 4: Message passing directions in (a) dense MRF/CRF and (b) Spatial CNN (rightward). For (a), only message passing to the inner 4 pixels are shown for clearance.

For SCNN, we evaluate the results for both when it is applied practically to a scenario and when it is set up like dense CRF. In the practical case, the SCNN is applied on the top hidden layer; therefore, the input contains more channels but fewer heights and widths. In the fair comparison case, the input size of Dense CRF is changed to the same with that of the other methods and the running time of all methods are tested on CPU. The results depict that in the fair comparison case, SCNN is over 4 times faster than the dense CRF despite the latter's efficient implementation made in (Krahenb ¨ uhl ¨ and Koltun 2011). This is caused by the fact that SCNN cuts down the amount of duplicate message passing that occurs as shown in Fig5. Also, SCNN is faster than LSTM especially due to the gate mechanism that functions in LSTM.

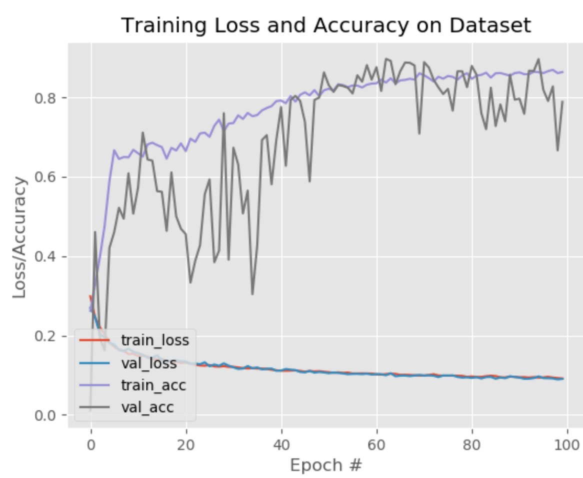
2. Data Visualization: Data Visualization: In the entire methodology, real-time visualization of data is used to analyze, comprehend the performance of the model, and further optimize it. Such representations are useful for assessing problems, evaluating the model's performance, and presenting the findings to others. Key visualizations include:

- Distribution of lane markings in the CULane and TuSimple datasets identified when labelling the lane markings.
- Sample of preprocessed as well as augmented images.
- Intermediary output from the SCNN layers.
- Curves of training and validation loss are utilized.
- P-R curves and IoU Distribution.
- Some of the examples of tests carried out on the images and the lane detection is shown below.

3. Network Architecture: Thus the SCNN is the center of each of our approaches. In the case of SCNN, the authors have presented a new spatial convolution layer that spreads information not only in the horizontal direction but in the vertical direction as well as in the conventional CNNs. This unique layer is good in dealing with elongated and continuous structures such as the lane markings as can be seen in the figure below.

The network architecture is illustrated in Fig3. It is made up of a constricting path (left) and a dilating path (right). The contracting path adheres to the standard architectural layout of having a convolutional network. It is composed of two sequences of 3x3 convolution, for which the results of the previous layer are passed through ReLU rectification and 2x2 max-pooling with a rate of 2 for downsampling. Thus, as in the initial part of the down-sampling process, at each down-sampling step, we increase, on average, the number of channels by a factor of two. In each of the steps in the wide path, an upsampling of the feature map is done and a 2*2 up-convolution that downs the channel size is performed and then the feature map is concatenated with the cropped feature map from the contracting path Two 3x3 convolutions with ReLU activation is also performed. The cropping is required because border pixels are missing in every convolution that takes place. At the final layer, the 1x1 convolution is also applied to convert each 64-component feature vector to the number of classes needed.

4. Model Training: The SCNN model is learnt using supervision and thus the incorporation of being a supervised neural network model. To do so, we use lane markings of CULane and TuSimple datasets as annotated guidelines for the learning process. The training strategy involves techniques like crop, flip, and adjustment of brightness to maximize the model's generalizability.



Comparison of the three largest lane detection datasets and our new CULanes are shown in Tab1. Our new CULanes benchmark has substantially more images, bigger resolution, more average number of lanes and more curved lanes.

TABLE I
DATASET COMPARISON

Datasets	Total Images	Resolution	Road Type	Curves
TuSimple [3]	13.2K	1280x720	Highway	~30%
CurveLane	133.2K	1640x590	Urban & Highway	~2%
BDD100K	100K	1280x720	Urban & Highway	~10%
Our CULane	150K	2650x1440	Urban & Highway	~90%

5. Loss Function: To increase the efficiency of the SCNN model, a specific loss function [11] that is used for lane detection is incorporated. This loss function focuses on the distances of the lane boundary predictions and also takes into consideration incorrect positive and negative predictions. We vary them to find out the best loss function for our immediate use in this application.

We guide the model in learning the real surface and prevent it from learning an arbitrary mapping by introducing a simple grid-based regression loss as

$$\mathcal{L}_{\text{surf}} = \frac{1}{X \cdot Y} \sum_{(u,v) \in X \times Y} 1_{uv} \cdot \| z_{uv} - \hat{z}_{uv} \|_1 , \qquad (1)$$

with 1_{uv} indicating whether surface ground truth $\hat{z}_{uv}$ is available for cell (u, v). The height ground truth is obtained by interpolation of the 3D lane annotations at cell locations.

The overall loss used during training is given as the weighted sum of loss components

$$\begin{aligned} \mathcal{L} = \lambda_{pr}\mathcal{L}_{pr} + \lambda_{cat}\mathcal{L}_{cat} + \lambda_{reg}\mathcal{L}_{reg} + \\ \lambda_{vis}\mathcal{L}_{vis} + \lambda_{prior}\mathcal{L}_{prior} + \lambda_{surf}\mathcal{L}_{surf} . \end{aligned} \qquad (2)$$

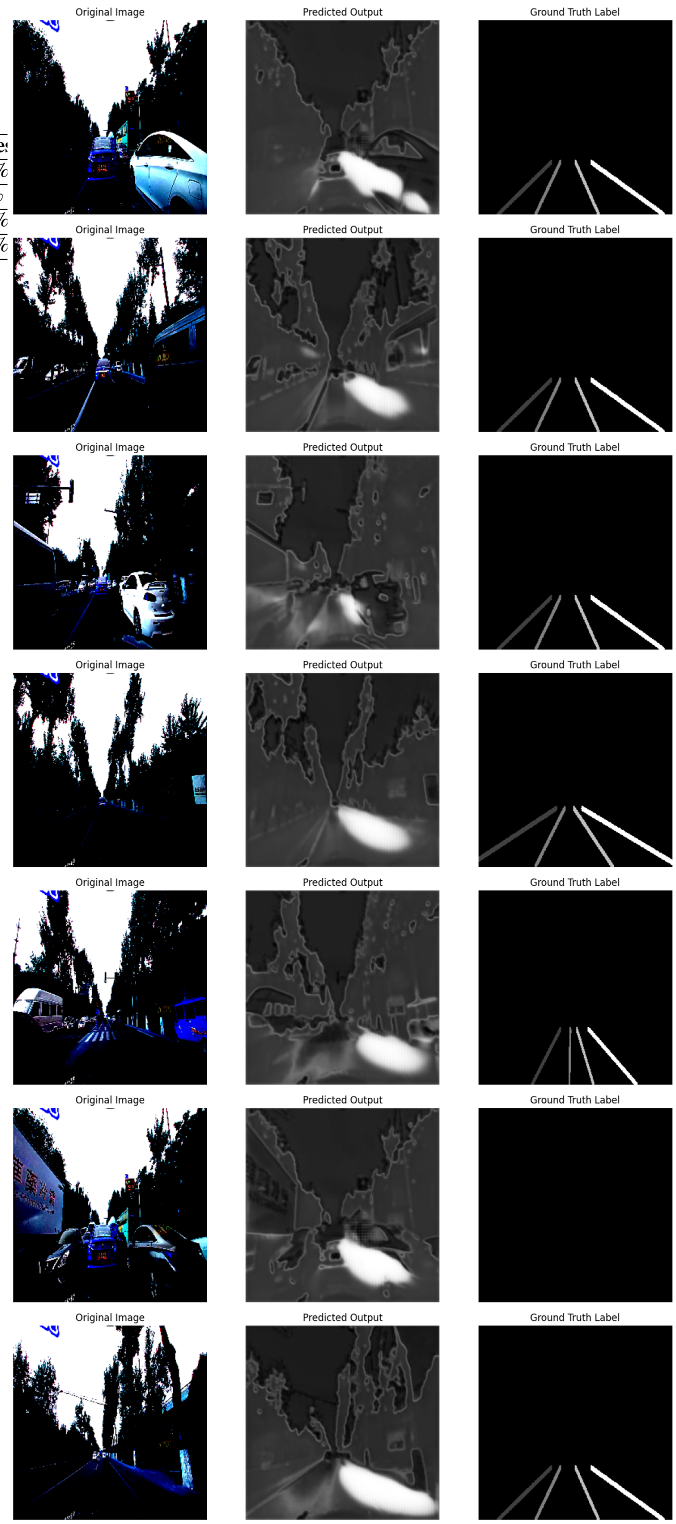
We use focal loss [11] for lane presence $\mathcal{L}_{pr}$ and category classification $\mathcal{L}_{cat}$. For the regression loss $\mathcal{L}_{reg}$, we adapt the formulation of [12] to three instead of two dimensions.

6. Evaluation Metrics: For the purpose of evaluating our proposed SCNN-based system on lane detection the following parameters were used Precision, Recall, F1 score, and IoU. These measures give a good insight about the model showing the level of correctness and stability of the model in real life.

Models	ExtraConv	SCNN_D	SCNN_DU	SCNN_DURL
F1 (0.3)	77.6	79.5	79.9	80.2
F1 (0.5)	64.0	68.6	69.4	70.4

TABLE II
EXPERIMENTAL RESULTS ON SCNN WITH DIFFERENT DIRECTIONAL SETTINGS. F1 DENOTES THE F1 MEASURE, AND THE VALUE IN THE BRACKET DENOTES THE IOU THRESHOLD. THE SUFFIXES'D', 'U', 'R', AND'L' DENOTE DOWNWARD, UPWARD, RIGHTWARD, AND LEFTWARD RESPECTIVELY.

7. Results and Discussion: Since PyTorch is an open-source library with good GPU support, our proposed SCNN model is implemented using PyTorch. Describing the exact details of the implementation that can be discussed are the hyperparameters used, the number of epochs for training, and the computational resources employed.



IV. CONCLUSION

In this paper, the CNN-like architecture with effective spatial information propagation is present with the help of

'Spatial CNN'. Thus, it was easily integrated into deep neural networks and could be trained in an end-to-end manner. It is evaluated at two tasks in traffic scene understanding: ed lane detection and thus semantic segmentation. Such figures also proved that the effects from the diffusion of SCNN-enabled end-to-end learning for thin shapes retain the continuity of the long thin structure without discontinuity breaks in the middle. In detail, incorporating SCNN to the LargeFOV model, our 20-layer network shows a slightly better result in the lane detection compared to MRF and U-net. Finally, we supposed that the dataset introduced for large challenging lane detection would promote the research on autonomous driving.

REFERENCES

1 X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, "Spatial As Deep: Spatial CNN for Traffic Scene Understanding," *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018.

2 O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015.

3 TuSimple, "TuSimple Benchmark for Lane Detection," 2019. [Online]. Available: https://github.com/TuSimple/tusimple-benchmark

4 C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019.

5 J. Bergstra and Y. Bengio, "Random Search for Hyper-Parameter Optimization," *Journal of Machine Learning Research*, vol. 13, no. Feb, pp. 281-305, 2012.

6 Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 2018, pp. 3-11.

7 J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How Transferable Are Features in Deep Neural Networks?," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

8 Yan Tiana, Judith Gelernter b, Xun Wanga, Weigang Chena, Junxiang Gaoc, Yujie Zhanga, Xiaolan Li a *Lane marking detection via deep convolutional neural network.*, 2017.http://www.elsevier.com/locate/patcog

9 Jigang Tanga,b, Songbin Li a,b, Peng Liua,b *A review of lane detection methods based on deep learning.*, 2020.

10 Li Chen1, Chonghao Sima1, Yang Li1, Zehan Zheng1, Jiajie Xu1,Xiangwei Geng1, Hongyang Li1,2, Conghui He1, Jianping Shi3, Yu Qiao1, and Junchi Yan *PersFormer: 3D Lane Detection via Perspective Transformer and the OpenLane Benchmark*, 2020.

11 Tsung-Yi Lin, Priya Goyal, Ross B. Girshick, Kaiming He, and Piotr Dollar.*Focal loss for dense object detection. In Proc. of the IEEE International Conf. on Computer Vision (ICCV)* 2017.

12 Maximilian Pittner, Alexandru Condurache, and Joel Janai.*3d-splinenet: 3d traffic line detection using parametric spline representations. In Proc. of the IEEE Winter Conference on Applications of Computer Vision (WACV)* 2023.