Lane Detection

ZHANWEI ZHANG and JIANAN XIE, SUSTech, China

Additional Key Words and Phrases: Lanenet, ENet, UNet, DeepLabv3+

1 BACKGROUND

With the development of intelligent transportation, environment perception, as an essential task for autonomous driving, has become a research hotspot. Lane detection is an important part of environmental perception, which is a pivotal technology in autonomous driving field, essential for enabling vehicles to navigate safely on roads.

The existing lane detection methods can be grouped into two categories. One is composed of feature extraction and post-processing. Another one gets the detection and clustering results directly from the input image. To be specific, feature extraction includes heuristic recognition-based algorithm[1][5] and deep learning-based method. And post-processing(like in Vpgnet[12]) mainly contains clustering and fitting.

From technical perspective, there are based on traditional computer vision methods(like Hough transform)[8], segmentation-based methods[9], based on line-by-line classification methods, regression-based methods[11], incorporating attention mechanisms methods[13][6].

2 MOTIVATION

Lane detection is a critical component in autonomous driving and advanced driver assistance systems (ADAS), involving several challenges and difficulties: One is about visual interferences, which includes Worn lane markings, shadows, water stains, and oil. Besides, the variety of environment like different lighting conditions (such as direct sunlight, nighttime driving) and weather conditions (such as rain, fog, snow) can affect the visibility of lane lines. What's more, different kinds of lane markings also cause trouble. The ability to accurately identify and differentiate between various types of lane markings, such as solid lines, dashed lines, double yellow lines, and crosswalks, is necessary. In addition, lane lines maybe missing or occluded by other objects. And undulating terrain can also make the detection task quite hard. Moreover, accurately tracking multiple lane lines and distinguishing the boundaries of each lane in heavy traffic is challenging.

These challenges indicate that lane detection systems need to be highly adaptable and robust to work accurately in various driving environments. Researchers are continuously improving the performance of lane detection systems through technologies like deep learning, computer vision, and sensor fusion.

3 RELATED WORKS

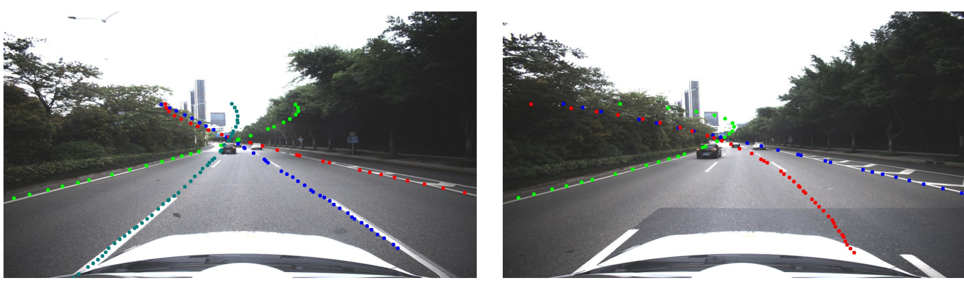
Many efforts have been done during the last decades and many aspects of improvements are achieved. Vpgnet[12] uses vanishing point estimation to improve robustness under various conditions. A reliable CNN based regression approach[11] is used to detect thin and elongated lane boundaries. A newly published work[4] based on transformer can learn BEV representations to eliminate the changes in road height. However, there are some issues and limitations lie in many models. One is that model including post-processing stage can take high computational complexity[12]. In addition, many models only take ego lanes into consideration[7] or can only detect fixed number of predefined lanes[14][11]. Besides, undulating terrain sets up big difficulties for lane detection. Intuitively, learnable perspective transformation will overperform fixed one. Therefore, we want to tackle the problems above and reach better performance based on a present model[9], which

Authors' address: Zhanwei Zhang; Jianan Xie, SUSTech, 1088 Xueyuan Venue, Shenzhen, China.

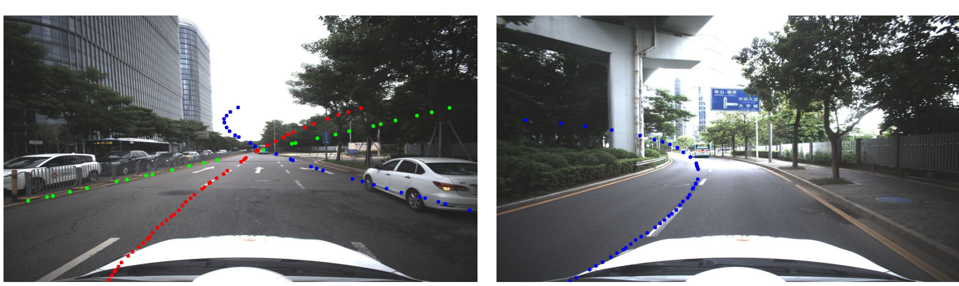
proposed a H-net to estimate IPM transformation matrix and used instance segmentation to detect multiple lanes.

4 INNOVATION

LaneNet's architecture is based on the encoder-decoder network ENet[2], which is consequently modified into a two-branched network. Given that we have tried the original LaneNet model on SUSCape dataset, and find that there are two fatal problems. One is the incomplete lane detection which can lead to great disaster. The other one is that the detected lanes is distorted at the end.



(a) (b)



(c) (d)

Fig. 1. Pre-trained model in SUSCape dataset

We believe that the missed detection of lanes is related to the backbone network's ability to learn representations of images. Because ENet is a simple encoder-decoder architecture used in segmentation tasks, but there are some newly released networks which have the more powerful capacity to learn the representation of images. Therefore, in order to solve these two problems and improve the safety of lane detection, our work is trying to substitute the encoder network of LaneNet, using UNet[10] and DeepLabv3+[3] as backbone to improve the performance of LaneNet.

5 METHODOLOGY

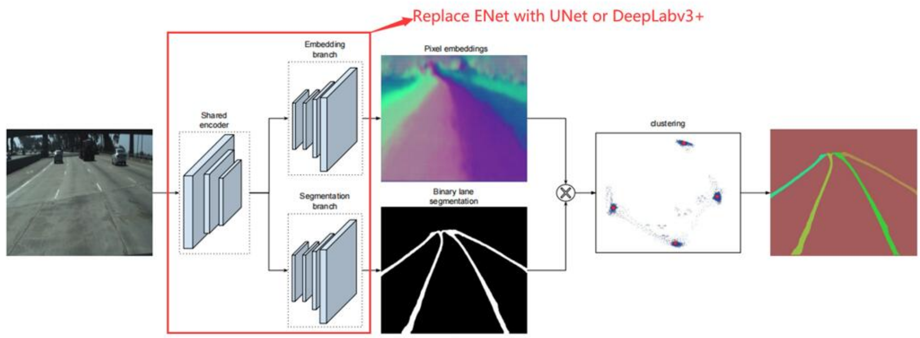


Fig. 2. our model using UNet/DeepLabv3+ as backbone

Our model adopts the framework of Lanenet[9] which is described as below:

- LaneNet: This neural network performs instance segmentation, treating each lane as a separate instance. It has two branches: a segmentation branch that creates a binary lane mask, and an embedding branch that generates an N-dimensional embedding per lane pixel. The network assigns each pixel an ID, clustering pixels from the same lane together.
- Curve Fitting Using H-Net: H-Net is used to estimate parameters for a perspective transformation, tailored to the specific image. This transformation is then applied to lane points for optimal polynomial fitting (usually 2nd or 3rd order). Unlike fixed "bird's-eye view" transformations, H-Net's approach is adaptable and robust against changes in the road plane.

Our model substitutes the backbone of Lanenet[9] from ENet[2] to UNet[10] and DeepLabv3+[3] respectively in expectation of higher performance.

- UNet: UNet is a popular convolutional neural network for biomedical image segmentation, introduced in 2015, with a U-shaped architecture for capturing context and precise localization. Compared to ENet, a real-time semantic segmentation network introduced in 2017 for mobile devices, UNet has advantages for segmentation tasks due to its larger capacity, skip connections preserving spatial information, and symmetric architecture ensuring balanced representation of local and global features.
- DeepLabv3+: DeepLabv3+ is a powerful convolutional neural network for semantic image segmentation, introduced in 2018, with advanced features such as atrous spatial pyramid pooling and encoder-decoder architecture. Compared to ENet, DeepLabv3+ has advantages for segmentation tasks due to its higher accuracy, robustness to various scales and aspect ratios, and ability to capture detailed structures.

6 EXPERIMENT SETTINGS

6.1 Hardware Configuration

- *CPU:* Intel(R) Xeon(R) Gold 6240 CPU @ 2.60GHz
- *GPU:* NVIDIA GeForce RTX 2080 Ti, equipped with four units to facilitate parallel processing and accelerate computational tasks.

6.2 Software Environment

- *OS:* Rocky Linux release 8.5 (Green Obsidian).
- *Programming Language:* Python 3.7
- *Deep Learning Framework:* TensorFlow 1.15.5, PyTorch 2.1.0

7 EXPERIMENT

We trained our LaneNet using UNet and DeepLabv3+ as backbone on the TuSimple dataset respectively. The model was trained for 25 epochs with the loss function set to Focal Loss.

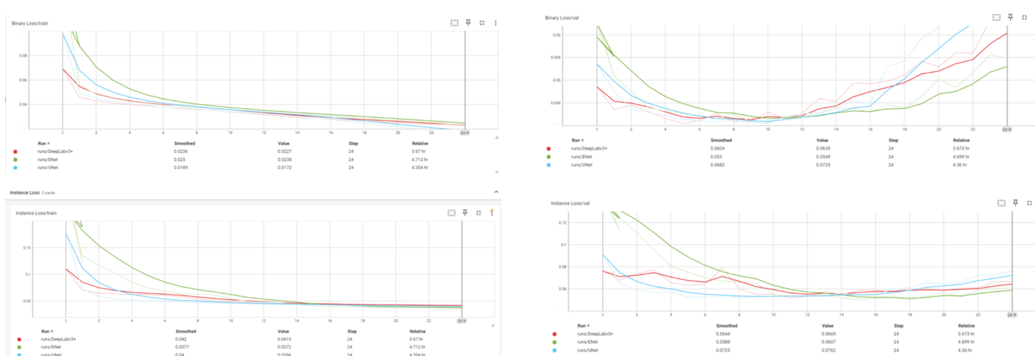


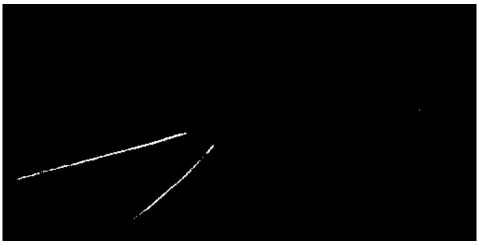
Fig. 3. Loss Curve

In the following part of our experiment, we selected an image with exceptionally high difficulty for lane detection. The challenges presented in this image include significant variations in lighting, with some areas being underexposed while others are overexposed, a highly cluttered scene, and the presence of curves, which are inherently more challenging to recognize than straight roads.

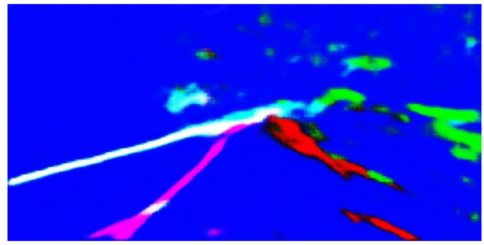
With the ENet model, we observed that the lane markings were faintly identified. Notably, the model failed to detect the rightmost lane. The instance image revealed a chaotic segmentation on the right side of the scene, indicating a potential risk to driving safety due to poor lane recognition.

Upon switching to the UNet architecture, there was a noticeable improvement. The rightmost lane, previously undetected by the ENet, was now correctly identified. This change indicated an enhancement in the network's learning capabilities.

The final transition to a more powerful model, DeepLabv3+, brought significant improvements. The instance image no longer exhibited the previous chaotic segmentation. Furthermore, the ability to recognize curved lanes, especially those in the distance, was markedly better than with both ENet and UNet models.

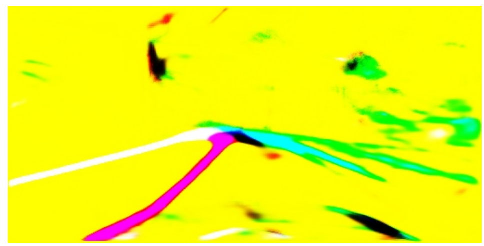


(a) ENet_binary



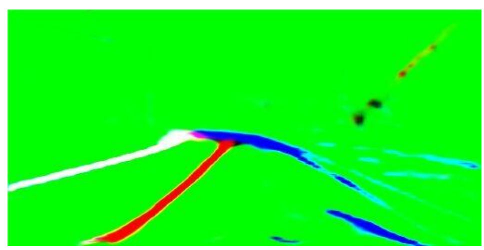
(b) ENet_instance

(c) UNet_binary



(d) UNet_instance

(e) DeepLabv3+_binary



(f) DeepLabv3+_instance

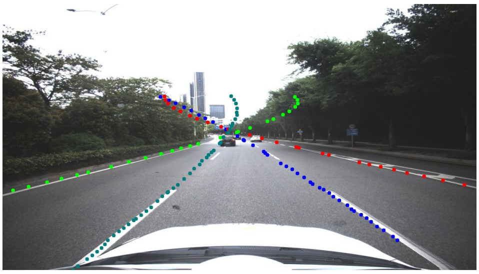
Fig. 4. Binary and Instance segmentation Comparison for nkd_3

8 CONCLUSION

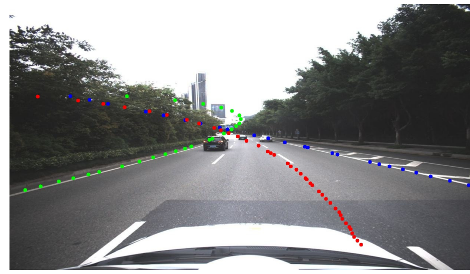
Our models, specifically those using UNet and DeepLabv3+ as backbones, have demonstrated superior performance in real-world scenarios compared to the original ENet-based LaneNet. The DeepLabv3+ based model, in particular, exhibits a significantly more powerful ability to understand and learn the structure of lanes. It demonstrates efficient and accurate lane recognition, typically within a timeframe of 2 to 3 seconds.

This enhanced performance can be attributed to the advanced architectural features of DeepLabv3+, which include an atrous spatial pyramid pooling module that effectively captures multi-scale contextual information, crucial for complex lane detection tasks. Furthermore, the improved feature extraction and segmentation capabilities of this model enable it to handle challenging scenarios, such as varying lighting conditions, complex road geometries, and diverse traffic environments.

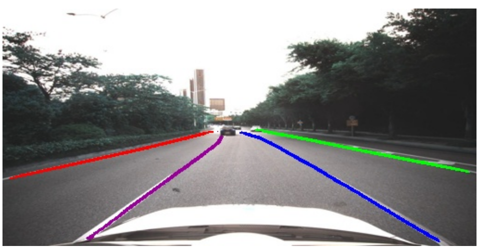
The successful deployment of these models in real-world conditions underscores the potential of advanced deep learning techniques in the realm of autonomous driving and advanced driver-assistance systems. The ability to accurately detect lanes under diverse and challenging conditions is critical for the safety and reliability of these systems. Our findings suggest that the adoption of



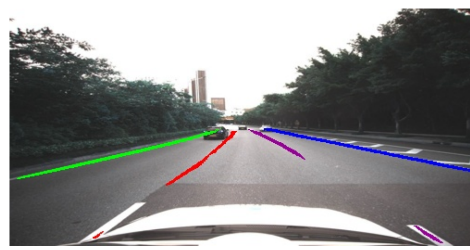
(a) ENet on nkd_0



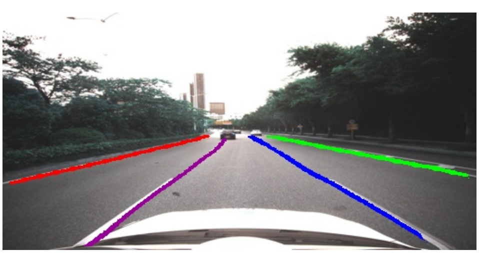
(b) ENet on nkd_1



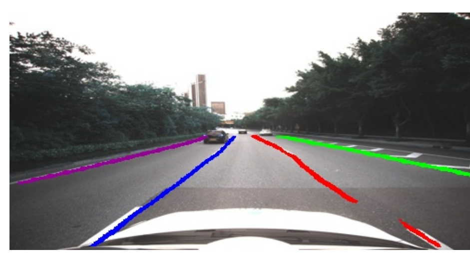
(c) UNet on nkd_0



(d) UNet on nkd_1



(e) DeepLabv3+ on nkd_0



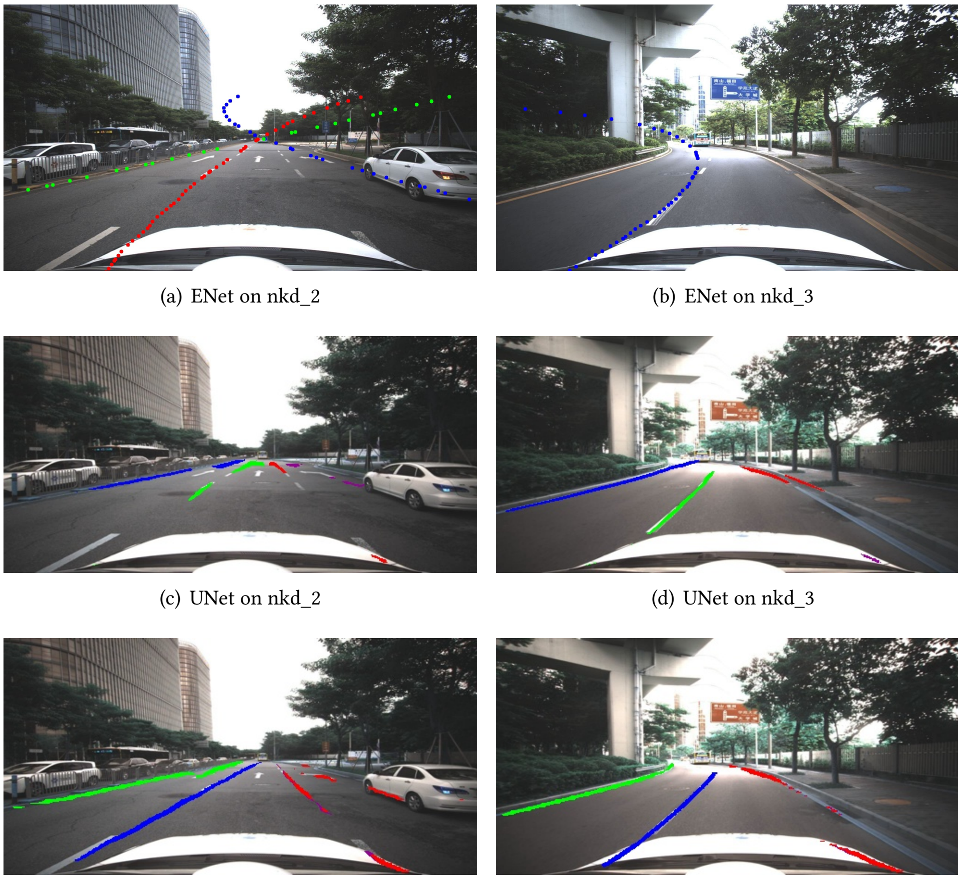
(f) DeepLabv3+ on nkd_1

Fig. 5. Mask image comparison(1)

more sophisticated models like DeepLabv3+ can significantly enhance the perception capabilities of autonomous vehicles.

9 FUTURE WORK

Although DeepLabv3+ has achieved excellent detection results, there are still some problems. On the one hand, we noticed that the outer contour of the car is sometimes detected as a lane line, and the disconnection of the lane line mark cannot be recognized in the UNet and ENet networks, nor can it be perfectly recognized in the DeepLabv3+ network, which makes us have to consider realizing that it is not enough to simply replace the more powerful backbone network, some methods or models with learning context should be used to represent lane lines with topological structures. For example, introducing some special method allows us to characterize an image. Small and medium-sized lane line pixels are processed, and the corresponding pixel instances are classified according to their straight line or curve structure, or the encoder structure of the transformer is used to learn the context structure of the lane line pixels. On the other hand, after improving the model learning ability, we found that the original distortion at the end of the lane line no



(a) ENet on nkd_2 (b) ENet on nkd_3

(c) UNet on nkd_2 (d) UNet on nkd_3

(e) DeepLabv3+ on nkd_2 (f) DeepLabv3+ on nkd_3

Fig. 6. Mask image comparison(2)

longer existed, but we did not consider operating H-Net in our work, which can be taken into consideration in subsequent work.

REFERENCES

[1] M. T. Smith A. Borkar, M. Hayes. 2012. A Novel Lane Detection System With Efficient Ground Truth Generation. *IEEE Trans. Intelligent Transportation Systems* (2012).

[2] S. Kim E. Culurciello A. Paszke, A. Chaurasia. 2016. ENet: A deep neural network architecture for real-time semantic segmentation. *CoRRabs/1606.02147* (2016).

[3] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. arXiv:1802.02611 [cs.CV]

[4] Ziye Chen, Kate Smith-Miles, Bo Du, Guoqi Qian, and Mingming Gong. 2023. An Efficient Transformer for Simultaneous Learning of BEV and Lane Representations in 3D Lane Detection. arXiv:2306.04927 [cs.CV]

[5] S. Reuter M. Szczot M. Konrad K. Dietmayer H. Deusch, J. Wiest. 2012. A random finite set approach to multiple lane detection. *ITSC* (2012).

[6] Ma Z. Liu C. Loy C. C Hou, Y. 2019. Learning lightweight lane detection cnns by self attention distillation. *In Proceedings of the IEEE/CVF international conference on computer vision* (2019).

[7] C. Park J. Kim. 2017. End-to-end ego lane estimation based on sequential transfer learning for self-driving cars. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2017).

[8] M. Xu P. Lv X. Zhao J. Niu, J. Lu. 2016. Robust lane detection using two-stage feature extraction with curve fitting. *Pattern Recognit* (2016).

[9] De Brabandere B. Georgoulis S. Proesmans M. Van Gool Neven, D. 2018. Towards end-to-end lane detection: an instance segmentation approach. *IEEE intelligent vehicles symposium (IV)* (2018).

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597 [cs.CV]

[11] A. Ismail G. Adam V. Narayan M. Schulze S. Chougule, N. Koznek. 2018. Reliable multilane detection and classification by utilizing CNN as a regression network. *Proceedings of the European Conference on Computer Vision(ECCV)* (2018).

[12] J. Shin Yoon S. Shin O. Bailo N. Kim T.-.H. Lee H. Seok Hong S.-.H. Han I. So Kweon S. Lee, J. Kim. 2017. Vpgnet: vanishing point guided network for lane and road marking detection and recognition. *Proceedings of the IEEE International Conference on Computer Vision* (2017).

[13] Berriel R. Paixao T. M. Badue C. De Souza A. F. Oliveira-Santos T Tabelini, L. 2021. Keep your eyes on the lane: Real-time attention-guided lane detection. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021).

[14] D. Neven M. Proesmans L. Van Gool W. Van Gansbeke, B. De Brabandere. 2019. End-to-end lane detection through differentiable least-squares fitting. *Proceedings of the IEEE International Conference on Computer Vision Workshops* (2019).