FVHNet: Homography Matrix Estimation for Virtual Camera

Zhiming Hou, Yuanzhouhan Cao*

***Abstract*—Robust lane detection in real-time is one of the foundations for advanced autonomous driving, which can provide substantial amounts of useful information for Autonomous Driving Systems, vehicle self-control, localization, and map construction. Since the in/extrinsic parameters of different vehicles are various which has a significant impact on the results of 3D lanes, we introduce the Virtual Camera that unifies the in/extrinsic parameters of cameras mounted on different vehicles to guarantee the consistency of the spatial relationship among cameras. It can effectively promote the learning procedure due to the unified visual space. Different from using fixed intrinsic parameters and extrinsic parameters to transform front-view images to virtual images, we train a network, termed FVHNet, that estimates the parameters of an ideal perspective transformation, conditioned on the input image. In order to train FVHNet for outputting the transformation matrix that is optimal for perspective transformation, we construct a loss function referred FVHLoss. In summary, we propose a method which can unify the in/extrinsic parameters of different vehicles with learnable homography matrix. We verify our method on the Apollo 3D Lane Synthetic Dataset and achieve competitive results.**

***Index Terms*—Homography, Virtual Camera, 3D Lane Detection.**

I. Introduction

As one of the foundation for safety in autonomous driving, many research efforts in lane detection focus on making detection model accurate and robust. Over the past few years, 2D lane detection methods have shown impressive performances [1–5]. However, due to the lack of depth information, transforming 2D images to 3D space still remains highly challenging. Recently, large-scale datasets with 3D lane annotations have been proposed. This has been promoted the research of 3d lane line representation and detection [5–11]. Compared to 2D lane, 3D lane detection can directly obtain the slope information of the lane, to help autonomous vehicles make better decisions.

Among these methods for 3D lane line detection, BEV-based methods [6, 7, 9, 10, 12] have received attention from researchers due to the accuracy, robustness, and speed. These methods need a spatial transformation module to transform front-view features to bev-view. In Bev-LaneDet[12], spatial transformation module is fixed spatial mapping by MLP, which are difficult to be integrated with the camera in/extrinsic parameters, resulting in poor performance. To address this issue, it realizes a preprocessing method of quickly unifying the camera in/extrinsic parameters by establishing a Virtual Camera with standard in/extrinsic parameters, which are fixed parameters derived from the mean value of the in/extrinsic parameters of the training dataset. However, different cameras mounted on the vehicle have different internal/external parameters, which have a significant effect on features transformation and 3D lane results. As illustrated in Fig.1(a), if a fixed transformation matrix is employed to get homography matrix for features transformation, the projection becomes less accurate. Therefore, we need unify the in/extrinsic parameters of front-facing cameras on different vehicles.

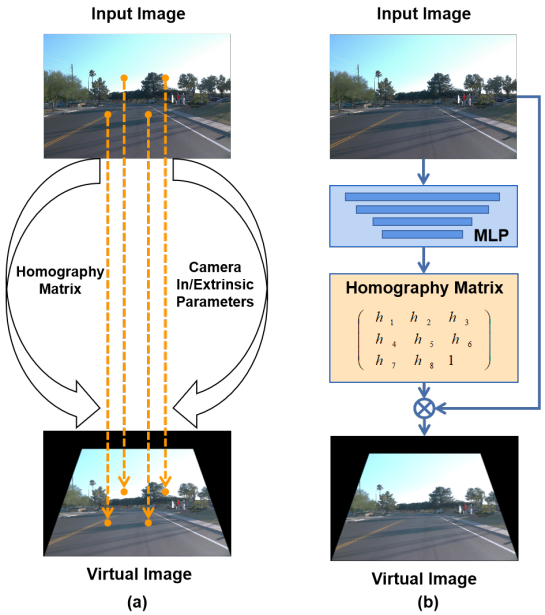

Fig. 1. (a) Fixed parameters transformation matrix, which based on camera in/external parameters. (b) MLP-based transformation matrix, which learns a perspective transformation conditioned on the input image.

Towards this issue, we proposed a MLP-based homography matrix estimation network, called **F**ront-**V**irtual **H**omograph Matrix Net, which makes homography matrix learnable instead of fixed. As show in Fig.1(b). FVHNet takes the RGB image as input and predicts the parameters to generate the homography matrix. We project the input images to the virtual camera by multiplying the homography matrix. After this transformation, we input the virtual-view image features to backbone network which output the front-view features. Then, we can get the 2d prediction by a front-view lane detection head. Finally, the front-view prediction is re-projected back to the original image space by multiplying the inverse homography matrix

This work is supported by Beijing Natural Science Foundation (L221011).
Yuanzhouhan Cao is the corresponding author.

and compute the loss between the prediction and ground truth. Through this way, the model can be trained effectively without providing any extra annotations except for the original lane labels.

In general, our main contributions are three-fold: 1) FVHNet, a MLP-based network that can learn transformation matrix parameters which can transform input images to virtual camera images. **2)** FVHLoss, a loss function to supervise FVHNet training, which only needs the annotations of lane marks. **3)** Our method is validated on the Apollo 3D synthetic dataset, achieving the promising performance.

II. Related Work

A. 2D Lane Detection

2D lane detection [1, 3, 13–16] aims at obtaining the accurate shape and locations of lanes in the images, which is not concerned with the spatial extension of the lane line. Earlier works [17–23] mainly focus on extracting low-level handcrafted features, such as edge and color information. However, these approaches are less robust under changing scenarios and often have complex feature extraction and post-processing designs. Benefiting from deep learning, 2D lane detection has made significant progress. These works are divided into four categories according to segmentation-based methods, keypoint-based methods, curve-based parameters, anchor-based methods.

Segmentation-based methods [2, 14, 24–27] formulate 2D lane detection task as a pixel-wise classification problem, which the computing cost is expensive.

Keypoint-based methods [4, 28–30] focus on identifying and localizing specific key points or landmarks along lane boundaries in images or videos. These methods aim to simplify the lane detection process by directly detecting key features of the lanes. While keypoint-based methods can be efficient, they may face challenges in scenarios where key points are not well-defined or when there is significant noise or clutter in the image. Additionally, accurately identifying key points in varying environmental conditions can be a demanding task.

Curve-based methods [13, 15, 31] focus on identifying and characterizing lane boundaries using mathematical curve representations. These methods aim to model the lanes as curves, such as quadratic or cubic functions, and estimate the parameters of these curves to accurately detect and localize the lanes. These works proposed that the 2D lane detection can be converted into the problem of curve parameter regression by detecting the starting point, ending point, and curve parameters.

anchor-based methods [5, 32–36] design line-like anchors and estimate the offsets between sampled points and predefined anchor points, making them particularly suitable for scenarios with distinct lane patterns. Non-Maximum Suppression (NMS) is then employed to select the lane lines with the highest confidence. LineCNN [32] first defines straight rays emitted from the image boundary to fit the shape of 2D lane lines and applies Non-Maximum Suppression (NMS) to keep only lanes with higher confidence. LaneATT [34] proposes an anchor-based pooling method and an attention mechanism to aggregate more global information. CLRNet [35] learns to refine the initial anchors iteratively through the feature pyramid.

B. 3D Lane Detection

Since projecting 2D lanes into 3D space suffers from inaccuracy as well as less robustness, many researchers have turned their attention to lane detection in 3D space. Unlike traditional 2D methods that operate solely in the image plane, 3D lane detection leverages depth information to provide a more comprehensive understanding of the road environment, enabling vehicles to perceive and navigate lanes in real-world scenarios with varying terrains, elevation changes, and complex road geometries.

Some works restore 3D information using multiple sensor [6, 37, 38]. While 3D lane detection offers significant advantages, it also comes with challenges such as computational complexity, sensor calibration and the collection and annotation cost of multisensor data is expensive. Therefore, monocular camera image based 3D lane detection [5, 7–12] attracts more attention.

Due to the good geometric properties of lanes in the perspective of BEV, **3DLaneNet** [8] predict the position of lanes in 3D space. It utilizes an Inverse Perspective Mapping (IPM) technique to transform features from a front-view image into a Bird's Eye View (BEV) representation, where the geometric properties of lanes are more easily discernible. By regressing the anchor offsets in the BEV space, 3DLaneNet can accurately predict the position of lanes without relying on the assumption of a flat ground. **Gen-LaneNet** [9] improves the alignment between the virtual top view generated by an inverse perspective mapping (IPM) and the true top view in 3D space. By distinguishing between these views, Gen-LaneNet enhances the accuracy of lane detection without the need for a bird's-eye view (BEV) transformation. **Persformer** [6] utilizes deformable attention to generate bird's-eye-view (BEV) features more adaptively and robustly, improving the accuracy and reliability of 3D lane detection without relying on the flat ground assumption. **SALAD** [11] tries to get rid of BEV by decomposing 3D lane detection into 2D lane segmentation and dense depth estimation tasks. **Anchor3DLane** [5] predict 3D lanes directly from frontal-viewed (FV), which defines 3D lane anchors in the 3D space and projects them onto the FV features to extract structural and contextual information for accurate predictions, and incorporates a global optimization technique to reduce lateral prediction errors by leveraging the equal-width property between lanes. **BEV-LaneDet** [12] establishes a Virtual Camera with standard in/extrinsic parameters to ensure the consistency in the spatial relationship among cameras, and introduces a Spatial Transformation Pyramid module for transforming front-view features into Bird's Eye View (BEV) features.

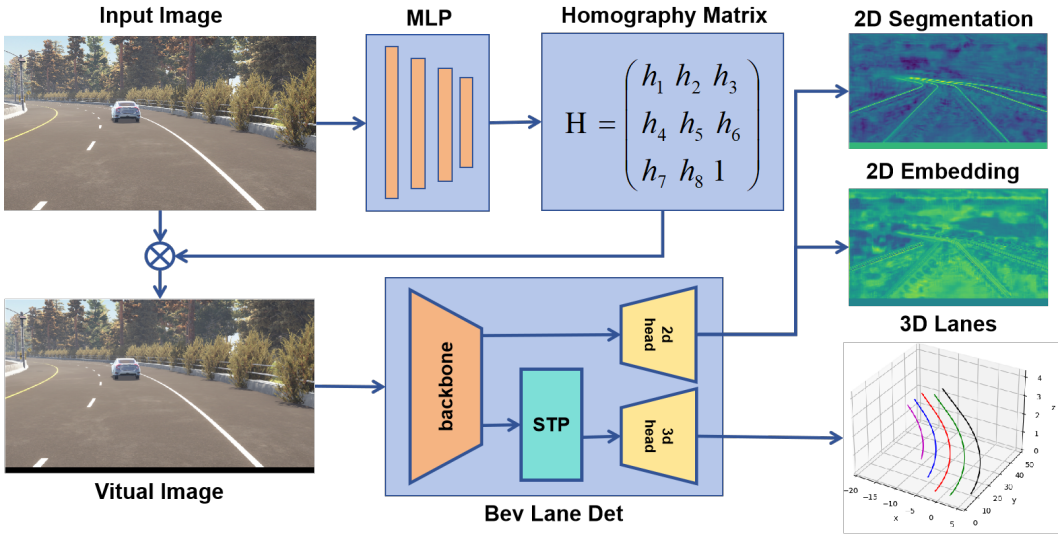

Fig. 2. Overview of the framework

III. Methods

An overview of our entire lane detection framework is illustrated in Fig.2. Our approach takes a single image captured by a front camera mounted on the vehicle as input. The image is projected to the virtual camera view with the homography matrix generated by MLP network [39], which aims to unify the in/extrinsic parameters of front-facing cameras in different vehicles.

We use ResNet18 and ResNet34 [40] as our backbone to extract virtual camera view image features, called front-view features. Spatial Transformation Pyramid [12] transform the front-view features into BEV features. Then, lanes are predicted on the BEV view. We predict the confidence of each cell, the embedding used for clustering, the offset from the center of the cell to the lane in the y-direction, and the height. we added the front view lane detection header as an auxiliary supervision to improve backbone's ability to extract front view features and supervise the homography matrix estimation network's training.

A. Homography Transformation

Homography matrix is the simplest kind of transformation that describes the 2D relationship between two images. It can be mathematically described by a 3D transformation in a homogeneous coordinates space and can be expressed as:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = H \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

where H denotes the homography matrix $H \in R^{3\times3}$ between two two-dimensional planes, which allows us to transform images from one view to another by multiplying the homography matrix. (x, y) is the principle point in the image plane, and (x', y') is the corresponding point in the target image.

Homography matrix is usually parameterized by the elements of a 3×3 matrix, but it has only 8 degrees of freedom, and a simple way to do this is to hardcode $h_9 = 1$. We take the original image as input and use network to estimate the eight parameters.

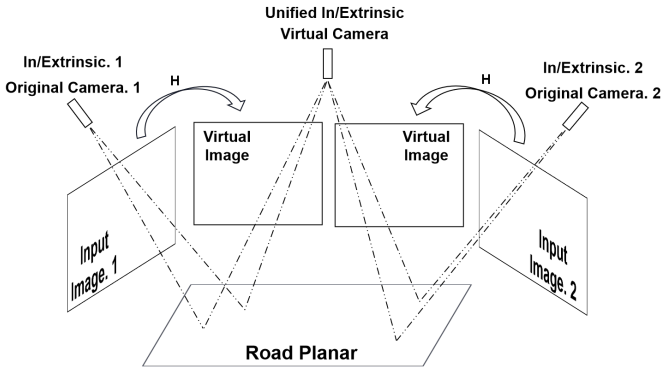

Fig. 3. Virtual Camera

B. FVHNet

Different cameras mounted on the vehicle have different internal/external parameters, which have a significant effect on the 3D lane results. As shown in Fig.3, we can unify the internal/external parameters of different cameras by creating a virtual camera. Therefore, we need transform front-view images to virtual camera images by homography matrix, which can guarantee the consistency of the spatial relationship among cameras. However, if a fixed transformation matrix is employed, the projection becomes less accurate. Towards the issue, we train FVHNet, a simple and efficient network, to output certain crucial parameters in the homography transformation. The network architecture of FVHNet is constructed out of consecutive blocks of 3x3 convolutions, batchnorm and ReLUs. The dimension is decreased using max pooling layers,

TABLE I
FVHNET NETWORK ARCHITECTURE.

Type	Filters	Size/Stride	Output
Conv+BN+ReLU	6	3x3	576x1024
Conv+BN+ReLU	12	3x3	576x1024
Maxpool		2x2/2	288x512
Conv+BN+ReLU	24	3x3	288x512
Conv+BN+ReLU	48	3x3	288x512
Maxpool		2x2/2	144x256
Conv+BN+ReLU	96	3x3	144x256
Conv+BN+ReLU	192	3x3	144x256
Maxpool		2x2/2	72x128
Conv+BN+ReLU	128	3x3	72x128
Conv+BN+ReLU	128	3x3	72x128
Maxpool		2x2/2	36x64
Conv+BN+ReLU	128	3x3	36x64
Conv+BN+ReLU	96	3x3	36x64
Maxpool		2x2/2	18x32
Conv+BN+ReLU	64	3x3	18x32
Conv+BN+ReLU	32	3x3	18x32
Linear+BN+ReLU		1x1	1024
Linear		1x1	8

and in the end 2 fully-connected layers are added. See Tab.I for the complete network structure.

C. FVH Loss

In order to improve the MLP network's ability to extract homography matrix parameters and backbone's ability to extract front view features, a front view lane detection header was added as an auxiliary supervision. Based on this front view lane detection head, we propose the homography loss function. The input image is converted to a virtual camera image by homography matrix H. Then, the front-view features are obtained by backbone, and the front-view lane detection head gets the segmentation and embedding results, which are transformed back to the input camera view by the inverse matrix H^{-1} of the homography matrix. As shown in Fig.4, FVHLoss includes lane segmentation loss and lane embedding loss, referred to LaneNet [24].It is defined as follows,

$$L_h = \lambda^h_{seg} L^h_{seg} + \lambda^h_{emb} L^h_{emb} + \lambda^h_{hm} L^h_{hm}$$

where L^h_{seg} denotes lane segmentation loss, L^h_{emb} denotes lane embedding loss, and L^h_{hm} denotes SmoothL1 loss between predicted and ground truth homography matrix. The ground truth homography matrix is calculated according to the camera intrinsic parameters and extrinsic parameters provided by the current camera and the in/extrinsic parameters of the Virtual Camera which are derived from the mean value of the in/extrinsic parameters of the training dataset.

IV. EXPRIMENTS

A. Dataset

Apollo 3D Lane Synthetic Dataset. Apollo Synthetic dataset [9] consists of over 10k 1080 × 1920 images which are

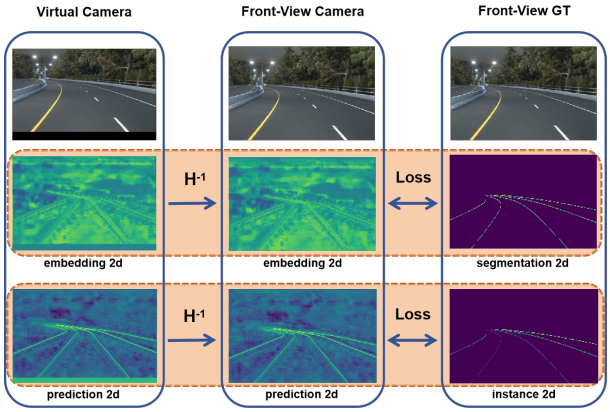

Fig. 4. FVH Loss

built using unity 3D engine, including highway, urban, residential and downtown environments. The dataset is split into three different scenes: balanced scenes, rarely observed scenes and scenes with visual variations for evaluating algorithms from different perspectives. Each scene contains independent training sets and test sets.

B. Experiment Settings

Implementation Details. We use ResNet [40] as backbone. The resolution of our input image is 576 × 1024. The bev range is set to [3, 103] × [-12, 12] along x and y respectively. Each cell represents x × x (x defaults to 0.5m). Our network uses Adam optimizer [41], with a base learning rate of 10^{-3} and weight decay of 10^{-2}. All models are trained from scratch with 200 epochs and the per-GPU batch size is set to 16. All results are tested on four RTX A4000 platform.

C. Evaluation

The evaluation metrics we used are referred from Gen-LaneNet [9], which includes F-Score in different scenes and X/Z error in different regions. We report Average Precision (AP) , Fscore, and errors (near range and far range) to investigate the performance of our model.

D. Comparisons with existing methods

To best show our performance, we also compare with other state-of-the-art methods 3D-LaneNet [8], Gen-LaneNet [9], PersFormer [6]. As shown in II The proposed method was trained using both Apollo 3D Lane Synthetic Dataset training and validation set as previous works did. Compared with other approaches, our method achieves competitive F-Score and lowest X error.

V. CONCLUSIONS

In this paper we have presented a MLP-based network to learn homography matrix parameters instead of fixed, which can accurately unify the in/extrinsic parameters of front-facing cameras on different vehicles. In order to supervise FVHNet, we have designed a loss function, called FVHLoss, to computed the loss between FVHNet outputs and ground

TABLE II
COMPARISON WITH PREVIOUS METHODS ON APOLLO 3D LANE SYNTHETIC DATASET.

Scene	Method	F-Score	X error near	X error far	Z error near	Z error far
Balanced Scence	3D-LaneNet [8]	86.4	0.068	0.477	0.015	**0.202**
	Gen-LaneNet [9]	88.1	0.061	0.496	0.012	0.214
	3D-LaneNet(1/att) [13]	91	0.082	0.439	0.011	0.242
	Gen-LaneNet(1/att) [13]	90.3	0.08	0.473	0.011	0.247
	PersFormer [3]	92.9	0.054	0.356	**0.010**	0.234
	Ours	**97.2**	**0.033**	**0.273**	**0.044**	**0.263**
Rarely Observed	3D-LaneNet [8]	72	0.166	0.855	0.039	**0.521**
	Gen-LaneNet [9]	78	0.139	0.903	0.03	0.539
	3D-LaneNet(1/att) [13]	84.1	0.289	0.925	0.025	0.625
	Gen-LaneNet(1/att) [13]	81.7	0.283	0.915	0.028	0.653
	PersFormer [3]	87.5	0.107	0.782	**0.024**	0.602
	Ours	**95.5**	**0.059**	**0.646**	**0.090**	**0.685**
Vivual Variants	3D-LaneNet [8]	72.5	0.115	0.601	0.032	**0.23**
	Gen-LaneNet [9]	85.3	0.074	0.538	**0.015**	0.232
	3D-laneNet(1/att) [13]	85.4	0.118	0.559	0.018	0.29
	Gen-LaneNet(1/att) [13]	86.8	0.104	0.544	0.016	0.294
	PersFormer [3]	89.6	0.074	0.43	**0.015**	0.266
	Ours	**94.3**	**0.051**	**0.389**	**0.043**	**0.307**

truth. Our method is robust against other methods with fixed homography matrix parameters.

REFERENCES

[1] D. Jin, W. Park, S.-G. Jeong, H. Kwon, and C.-S. Kim, "Eigenlanes: Data-driven lane descriptors for structurally diverse lanes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 163–17 171.

[2] Y. Hou, Z. Ma, C. Liu, and C. C. Loy, "Learning lightweight lane detection cnns by self attention distillation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1013–1021.

[3] R. Liu, Z. Yuan, T. Liu, and Z. Xiong, "End-to-end lane shape prediction with transformers," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2021, pp. 3694–3702.

[4] J. Wang, Y. Ma, S. Huang, T. Hui, F. Wang, C. Qian, and T. Zhang, "A keypoint-based global association network for lane detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1392–1401.

[5] S. Huang, Z. Shen, Z. Huang, Z.-h. Ding, J. Dai, J. Han, N. Wang, and S. Liu, "Anchor3dlane: Learning to regress 3d anchors for monocular 3d lane detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 17 451–17 460.

[6] L. Chen, C. Sima, Y. Li, Z. Zheng, J. Xu, X. Geng, H. Li, C. He, J. Shi, Y. Qiao *et al.*, "Persformer: 3d lane detection via perspective transformer and the openlane benchmark," *arXiv preprint arXiv:2203.11089*, 2022.

[7] N. Efrat, M. Bluvstein, S. Oron, D. Levi, N. Garnett, and B. E. Shlomo, "3d-lanenet+: Anchor free lane detection using a semi-local representation," *arXiv preprint arXiv:2011.01535*, 2020.

[8] N. Garnett, R. Cohen, T. Pe'er, R. Lahav, and D. Levi, "3d-lanenet: end-to-end 3d multiple lane detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 2921–2930.

[9] Y. Guo, G. Chen, P. Zhao, W. Zhang, J. Miao, J. Wang, and T. E. Choe, "Gen-lanenet: A generalized and scalable approach for 3d lane detection," in *ECCV*, 2018.

[10] R. Liu, D. Chen, T. Liu, Z. Xiong, and Z. Yuan, "Learning to predict 3d lane shape and camera pose from a single image via geometry constraints," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 2, 2022, pp. 1765–1772.

[11] F. Yan, M. Nie, X. Cai, J. Han, H. Xu, Z. Yang, C. Ye, Y. Fu, M. B. Mi, and L. Zhang, "Once-3dlanes: Building monocular 3d lane detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 143–17 152.

[12] R. Wang, J. Qin, K. Li, Y. Li, D. Cao, and J. Xu, "Bev-lanedet: An efficient 3d lane detection based on virtual camera via key-points," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 1002–1011.

[13] Y. Chen, Z. Xiang, and W. Du, "Improving lane detection with adaptive homography prediction," *The Visual Computer*, vol. 39, no. 2, pp. 581–595, 2023.

[14] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, "Spatial as deep: Spatial cnn for traffic scene understanding," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2018.

[15] L. Tabelini, R. Berriel, T. M. Paixao, C. Badue, A. F. De Souza, and T. Oliveira-Santos, "Polylanenet: Lane estimation via deep polynomial regression," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 6150–6156.

[16] J. Yang, L. Zhang, and H. Lu, "Lane detection with ver-

satile atrousformer and local semantic guidance," *Pattern Recognition*, vol. 133, p. 109053, 2023.

[17] M. Aly, "Real time detection of lane markers in urban streets," in *2008 IEEE intelligent vehicles symposium*. IEEE, 2008, pp. 7–12.

[18] Y. He, H. Wang, and B. Zhang, "Color-based road detection in urban traffic scenes," *IEEE Transactions on intelligent transportation systems*, vol. 5, no. 4, pp. 309–318, 2004.

[19] Z. Kim, "Robust lane detection and tracking in challenging scenarios," *IEEE Transactions on intelligent transportation systems*, vol. 9, no. 1, pp. 16–26, 2008.

[20] Y. Wang, E. K. Teoh, and D. Shen, "Lane detection and tracking using b-snake," *Image and Vision computing*, vol. 22, no. 4, pp. 269–280, 2004.

[21] S. Zhou, Y. Jiang, J. Xi, J. Gong, G. Xiong, and H. Chen, "A novel lane detection based on geometrical model and gabor filter," in *2010 IEEE Intelligent Vehicles Symposium*. IEEE, 2010, pp. 59–64.

[22] S. Srivastava, M. Lumb, and R. Singal, "Improved lane detection using hybrid median filter and modified hough transform," *Int. J. Adv. Res. Comput. Sci. Softw. Eng*, vol. 4, no. 1, pp. 30–37, 2014.

[23] P. Chanawangsa and C. W. Chen, "A new color-based lane detection via gaussian radial basis function networks," in *2012 International Conference on Connected Vehicles and Expo (ICCVE)*. IEEE, 2012, pp. 166–171.

[24] D. Neven, B. De Brabandere, S. Georgoulis, M. Proesmans, and L. Van Gool, "Towards end-to-end lane detection: an instance segmentation approach," in *2018 IEEE intelligent vehicles symposium (IV)*. IEEE, 2018, pp. 286–291.

[25] Z. Qin, H. Wang, and X. Li, "Ultra fast structure-aware deep lane detection," in *European Conference on Computer Vision*. Springer, 2020, pp. 276–291.

[26] Y. Dong, S. Patil, B. van Arem, and H. Farah, "A hybrid spatial–temporal deep learning architecture for lane detection," *Computer-Aided Civil and Infrastructure Engineering*, vol. 38, no. 1, pp. 67–86, 2023.

[27] D.-H. Lee and J.-L. Liu, "End-to-end deep learning of lane detection and path prediction for real-time autonomous driving," *Signal, Image and Video Processing*, vol. 17, no. 1, pp. 199–205, 2023.

[28] Y. Ko, Y. Lee, S. Azam, F. Munir, M. Jeon, and W. Pedrycz, "Key points estimation and point instance segmentation approach for lane detection," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[29] Z. Qu, H. Jin, Y. Zhou, Z. Yang, and W. Zhang, "Focus on local: Detecting lane marker from bottom up via key point," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14 122–14 130.

[30] S. Xu, X. Cai, B. Zhao, L. Zhang, H. Xu, Y. Fu, and X. Xue, "Rclane: Relay chain prediction for lane detection," in *European Conference on Computer Vision*. Springer, 2022, pp. 461–477.

[31] Z. Feng, S. Guo, X. Tan, K. Xu, M. Wang, and L. Ma, "Rethinking efficient lane detection via curve modeling," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 062–17 070.

[32] X. Li, J. Li, X. Hu, and J. Yang, "Line-cnn: End-to-end traffic line detection with line proposal unit," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 1, pp. 248–258, 2019.

[33] L. Liu, X. Chen, S. Zhu, and P. Tan, "Condlanenet: a top-to-down lane detection framework based on conditional convolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3773–3782.

[34] L. Tabelini, R. Berriel, T. M. Paixao, C. Badue, A. F. De Souza, and T. Oliveira-Santos, "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, pp. 294–302.

[35] T. Zheng, Y. Huang, Y. Liu, W. Tang, Z. Yang, D. Cai, and X. He, "Clrnet: Cross layer refinement network for lane detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 898–907.

[36] H. Ran, Y. Yin, F. Huang, and X. Bao, "Flamnet: A flexible line anchor mechanism network for lane detection," *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[37] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223.

[38] Y. Luo, X. Yan, C. Zheng, C. Zheng, S. Mei, T. Kun, S. Cui, and Z. Li, "Mˆ 2-3dlanenet: Multi-modal 3d lane detection," *arXiv preprint arXiv:2209.05996*, 2022.

[39] Y. Tang, K. Han, J. Guo, C. Xu, Y. Li, C. Xu, and Y. Wang, "An image patch is a wave: Phase-aware vision mlp," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10 935–10 944.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[41] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.