

TopoMask: Instance-Mask-Based Formulation for the Road Topology Problem via Transformer-Based Architecture

M. Esat Kalfaoglu

Halil Ibrahim Ozturk

Ozsel Kilinc

Alptekin Temizel

Graduate School of Informatics, Middle East Technical University, Ankara, Turkey

esat.kalfaoglu@metu.edu.tr, atemizel@metu.edu.tr

Abstract

Driving scene understanding task involves detecting static elements such as lanes, traffic signs, and traffic lights, and their relationships with each other. To facilitate the development of comprehensive scene understanding solutions using multiple camera views, a new dataset called Road Genome (OpenLane-V2) has been released. This dataset allows for the exploration of complex road connections and situations where lane markings may be absent. Instead of using traditional lane markings, the lanes in this dataset are represented by centerlines, which offer a more suitable representation of lanes and their connections. In this study, we have introduced a new approach called TopoMask for predicting centerlines in road topology. Unlike existing approaches in the literature that rely on keypoints or parametric methods, TopoMask utilizes an instance-mask based formulation with a transformer-based architecture and, in order to enrich the mask instances with flow information, a direction label representation is proposed. TopoMask have ranked 4th in the OpenLane-V2 Score (OLS) and ranked 2nd in the F1 score of centerline prediction in OpenLane Topology Challenge 2023. In comparison to the current state-of-the-art method, TopoNet, the proposed method has achieved similar performance in Frechet-based lane detection and outperformed TopoNet in Chamfer-based lane detection without utilizing its scene graph neural network.

1. Introduction

Accurate detection of stationary objects plays a vital role in downstream tasks like planning and control in autonomous driving. Static elements on the road include ground objects, such as lanes, road lines, curbs, and crosswalks as well as above-ground objects like traffic lights, traffic signs, and poles. Currently, High-Definition Maps (HDMaps) are utilized for the safe autonomous driving. These maps provide pre-computed static-object maps. However, HDMaps are costly to extract and are typically available for limited areas. They cannot capture recent changes on the road, necessitating continuous maintenance. Additionally, Global Navigation Satellite System (GNSS) receiver on the vehicle can introduce localization errors.

These errors, in turn, can cause discrepancies between the vehicle’s actual position and the information provided by the HDMaps, potentially leading to inaccuracies and drifts in the map-based guidance system.

Automatic HDMap extraction has gained significant attention in recent years for two primary reasons, as highlighted by several studies [1, 7, 10, 13]. Firstly, it has a potential to eliminate the reliance on HDMaps for autonomous driving. Secondly, it can reduce the cost of extraction and continuous maintenance of HDMaps, minimizing the required human effort. The introduction of Road Genome dataset [22], which introduces the centerline concept to represent lanes, is an important milestone with regards to these aspects. Centerlines provide a more natural representation of lanes and bring a number of advantages over lane markings. Firstly, centerlines eliminate the need for post-processing to match lane dividers with their corresponding counterparts, simplifying the lane detection process. Secondly, they provide a more intuitive representation of the direction of traffic flow. Thirdly, centerlines capture the relationships between lanes more effectively, enhancing the understanding of lane connections. An additional benefit of centerlines is their ability to address situations where lane markings are absent, such as in road intersections. In this dataset, association between lanes and traffic elements (traffic signs and lights) are also included into the problem definition which is significant for complete autonomous driving experience.

In this study, we have introduced a novel approach called TopoMask for representing centerlines, departing from the key-point based and parametric methods such as bezier and polynomial approaches. TopoMask leverages a transformer-based architecture combined with an instance-mask-based representation. A similar segmentation-based approach is employed in HDMapNet [7]. However, HDMapNet relies on semantic segmentation and requires additional instance embeddings and direction outputs, leading to a more complex post-processing step to merge the outputs. In contrast, our proposed formulation solely relies on the instance mask output with a novel direction label representation that denotes the flow of the centerlines: up,

down, left and right. This simplifies the overall process and avoids the need for extensive post-processing.

To the best of our knowledge, TopoMask achieves the state-of-the-art performance in OpenLane-V2 dataset comparing with the existing methods in the literature. Remarkably, our approach achieves comparable Frechet-based centerline detection performance to TopoNet with a score of 22.8, even without incorporating a relation improver block such as TopoNet’s Scene Graph Neural Network. In terms of Chamfer-based centerline detection performance, TopoMask significantly outperforms TopoNet with a score of 25.3.

In the OpenLane Topology Challenge 2023, our proposed architecture attains the 4th position with a OLS score of 39.2 and its F1-score of 45.9 is the 2nd highest score for centerline prediction.

2. Related Work

Studies in the field of lane detection can be classified into four main branches: perspective methods, 3D lane divider based methods, multi-camera BEV methods, and centerline based methods.

Perspective Methods: These methods focus on detecting lane dividers from the perspective view, followed by projecting them onto the ground using a homography matrix under the assumption of a flat surface. Although road dividers are different instances, semantic approaches are well-suited for this task due to the constant number of lane divider instances (side-left, ego-left, ego-right and side-right). SCNN [16] introduced a special module that applies sequential processes to the rows and columns of the feature map. The study in [18] aims to improve the inference speed, and converts the formulation from pixel-based to grid-based and proposes row-based anchor formulation. LaneATT [20] developed an anchor concept specifically for lanes (road lines), drawing inspiration from object detection studies. The emergence of the CurveLanes dataset [25] led to the adoption of instance-segmentation-based methods. CondLaneNet [11] introduced lane specific methodologies such as offset prediction and row-wise formulation. Some approaches such as Polynomial Order Structures [21] and Bezier Curve Approaches [5] used polynomial or bezier curve representations to reduce post-processing efforts and improve curve learning.

3D Lane Divider methods: Recent advancements in lane detection have shifted focus towards directly predicting the 3D locations of lane instances, eliminating the need for post-processing on perspective outputs. This shift gained momentum with the introduction of 3D lane datasets such as such as Once-3DLanes [27], OpenLane [3] and Apollo 3D Synthetic Lane dataset [6].

Persformer [3] utilizes a deformable attention-based decoder that utilizes reference points obtained from Inverse

Projection Mapping (IPM). It adapts the 2D anchor concept to 3D lane points. In the BEV-LaneDet method [23], a grid-based approach is employed, where positive points belonging to the same instance are unified using an embedding concept with a triplet-based loss. It also incorporates a virtual camera concept to unify different extrinsics from multiple camera positions for the View Relation Module (VRM) [15]. PETRV2 [14] extends its sparse query design for lane detection. M²-3DLaneNet explores the fusion of lidar and camera sensors and shows that these sensors are complementary in the 3D lane detection task.

Multi-Camera BEV methods: HDMapNet [7], one of the first studies in this branch, aim to detect lane dividers, road dividers and pedestrian crossings by utilizing three different heads: semantic segmentation, instance embedding and direction head. However, the use of three heads necessitates substantial post processing to convert them into polyline or polygon output. VectorMapNet [13] adopts a two-stage coarse-to-fine approach. In the first stage, sparse anchor keypoints are detected. In the second stage, dense keypoints are detected by utilizing the outputs of the first stage in an autoregressive manner. Instead of following an autoregressive approach, MapTR [10] predicts the points on the polyline or polygon directly with a permutation-invariant Hungarian matcher. InstaGraM [19] models this problem as a graph and models vertices and edges as heatmap and distance transform, respectively in 8×8 patches. From each patch, an embedding is extracted by considering the vertex positions and local directional information obtained from the distance transform. Subsequently, a graph neural network structure is constructed using an attention-based method. The Sinkhorn algorithm is employed to iteratively normalize the exponential scores between the vertices.

Centerline Concept: STSU [1] is one of the earlier studies that predicts the centerlines instead of lane markings. An extension of this study introduces the concept of minimal cycles, which aims to reduce the area between two intersections [2]. CenterLineDet [26] focuses on predicting centerlines from multiple cameras by leveraging temporal information. Additionally, it explores the impact of camera-LiDAR sensor fusion on centerline prediction. LaneGAP [9] considers centerlines as path-wise rather than piece-wise and instead of directly predicting centerline portions and their connections, the paths can be obtained by preprocessing the centerline graph. Subsequently, the model predicts the paths and converts them back into the piece-wise graph structure. TopoNet [8] is a pioneering study that proposes an architecture which is specifically designed for the Road Genome project [22]. It introduces a specific relation model that creates a graph neural network connecting centerlines and traffic elements. The predictions of centerlines are generated at the end of the topology network, which enhances centerline prediction accuracy. The incorporation of a dedi-

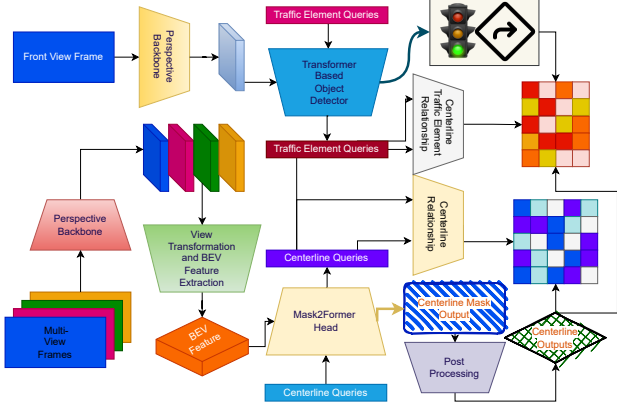


Figure 1. Overview of the TopoMask architecture, featuring two main branches: one for detecting traffic elements and the other for detecting centerlines. The Mask2Former architecture treats the BEV feature as 2D image features, eliminating the need for additional structure. Queries from both transformer architectures are treated as embeddings and used in the relationship blocks.

cated traffic element embedding also improves the detection of centerline elements.

3. Methodology

The general architecture of the TopoMask is shown in Figure 1. In this formulation, both traffic element and centerline detection branches use transformer-based architectures. The centerline detection part of the TopoMask architecture relies on the instance-mask formulation. In this approach, each centerline instance is represented as a mask instance. By applying appropriate post-processing to each mask instance, it becomes possible to extract a point set for each proposal.

However, the point set extracted from the mask instances does not have a specific order. In the problem definition of road topology, the flow information of each centerline is also crucial. In order to enrich the point sets with flow information, we propose a direction label representation (Figure 2) in TopoMask. As depicted in the leftmost image of the figure which represents the semantics of the direction information, there are four directions: up, down, left and right. The labels are assigned according to the *dominant monotonicity* between the two axes of the ground truth point set. In the post-processing stage, according to the predicted direction label representation, point sets are ordered with respect to one axis (x axis in vehicle coordinate system) for up and down direction labels, and ordered with respect to the other axis for left and right direction labels (y axis in vehicle coordinate system).

4. Experiments

Dataset and Metrics: Argoverse 2 (AV2) [24] dataset includes 7 cameras from 6 different cities and Subset-A release of the OpenLane-V2 dataset [22] (also known as Road Genome) used in this study which is built on top of the AV2.



Figure 2. Label representation of TopoMask: The leftmost image displays the semantic information of directions (up, down, left, and right) represented by distinct colors. The subsequent images depict the instances for each direction, showcasing the labels for up, down, left, and right, respectively.

Subset-A contains 22480 training, 4808 validation and 4816 test samples. For centerline and traffic element evaluation, the area considered is within +50 and -50 meters in the forward direction (x direction in the vehicle coordinate system) and +25 and -25 meters in the side direction (y direction in the vehicle coordinate system).

Both centerline detection and traffic element recognition utilize mAP (mean Average Precision) metrics for evaluation. True positive samples are determined using different distance measures depending on the task. For centerlines, the Frechet distance with thresholds of 1, 2, and 3 meters and Chamfer distance with 0.5, 1, and 1.5 meters are used. The Frechet distance takes into account both the distance and direction information between the predicted and ground truth centerlines while Chamfer distance ignores the direction information. For traffic elements, the Intersection over Union (IoU) metric is employed with a threshold of 0.75.

In addition to these metrics, a specific mAP metric has been proposed to evaluate topology reasoning in the graph domain. This metric takes into account the connectivity and relationships between centerlines and traffic elements. Note that, in order to assume connectivity (edge) as true positive, both vertices should be detected correctly regarding Frechet distance and IoU.

F1-score is also calculated, but it is not included in the OLS metric. The F1-score is commonly used in the evaluation of predicted 3D lane dividers, such as in Gen-LaneNet [6]. According to this metric, if 75% of the predicted points are within 1.5 meters error compared to the ground truth points of a corresponding instance, the sample is considered correct.

$$OLS = \frac{1}{4} \left[DET_l + DET_t + f(TOP_{lt}) + f(TOP_{tl}) \right] \quad (1)$$

The evaluation system utilizes OpenLane-V2 Score (OLS) as the overall metric, which is calculated as the average of multiple task metrics (Eq. 1): centerline prediction with Frechet-based mAP (DET_l), traffic element prediction with IOU based mAP (DET_t), and topology relations between centerlines and traffic elements (TOP_{lt} and TOP_{tl}). However, OLS does not take F1 score and Chamfer distance based mAP into account.

Method	DET _l ↑	DET _{l, chamfer} ↑	TOP _{lt} ↑	DET _t ↑	TOP _{lt} ↑	OLS↑
TopoNet [8]	22.1	20.2	2.7	59.1	14.9	34.0
TopoMask	22.1	23.8	5.8	58.2	15.5	36.0

Table 1. Comparison of TopoMask and TopoNet architectures. Both architectures use ResNet-50 model as the backbone and both were trained for 24 epochs. The key distinction is that TopoMask does not utilize a shared feature extractor.

Config Type	BEV Backbone	2D Backbone	# Epochs
Config1	ResNet50	ResNet50	25
Config2	RegNetY-800mf	ResNet50	40
Config3	RegNetY-800mf	ConvNextBase (OpenAi)	40
Config3*	RegNetY-800mf	ConvNextBase (OpenAi)	40 + 30 trainval

Table 2. Configs that are utilized in the ablations of Table 3.

Experimental Details: The study utilizes DAB-DETR [12] for the 2D object detection and LSS [17] for the BEV feature map creation with 200×104 BEV resolution. On top of the BEV feature, Mask2Former architecture [4] is utilized as instance mask prediction structure by preserving its original settings. For the relationship part, Sinkhorn iterations has been utilized as in InstaGraM study [19]. To establish a directed graph for centerline relationships, the start and end points of the predictions are combined with the centerline embeddings in the form of sinusoidal positional encodings. The image resolution is 640×896 in BEV branch and a batch size of 16 is utilized. AdamW with a learning rate of 0.0003 is utilized as the optimizer with 0.1 multiplication on backbone gradients. As a post processing, expectation based location estimation is implemented as row-wise formulation for the up/down directions, and column-wise formulation for the left/right directions similar to [18]. Then, second order polynomial fit is applied on top of the selected points and 11 points are sampled. During the mask prediction, bezier curves are also predicted with 5 control points in which start and end points are fixed. During bezier prediction, the 3D coordinates of centerlines are not normalized.

Experimental Results and Ablation Studies: We comparatively evaluate TopoMask with the current SOTA architecture, TopoNet, and report the results in Table 1. It should be noted that TopoMask does not have a relation improver method that enriches centerline embeddings with traffic element embeddings and implements second stage centerline prediction with these enriched embeddings. Despite this fact, TopoMask obtains on-par Frechet-based centerline detection performance and compares favorably in terms of Chamfer-based centerline detection and F1-scores. The results indicate that there is potential for improvement in the direction information of the TopoMask architecture. Additionally, the Frechet distance metric is negatively affected by sparse, highly erroneous detections among the 11-point predictions, particularly at intersections.

TopoMask contains both mask prediction and parametric prediction in Bezier format. Therefore, potential fusion

Method	DET _l ↑	DET _{l, chamfer} ↑	TOP _{lt} ↑	TOP _{lt} ↑	OLS↑	F-score↑
Config1 Bezier	14.4	14.7	4.2	11.4	31.7	36.1
Config1 Mask	22.1	23.8	5.8	15.5	36.0	45.2
Config1 Fusion	21.2	23.2	5.5	14.7	35.4	43.1
Config2 Bezier	17.7	20.1	4.9	12.1	33.1	40.1
Config2 Mask	22.8	25.3	6.0	15.0	35.9	47.3
Config2 Fusion	22.7	25.1	6.0	15.1	36.0	45.9
Config3 Bezier	19.4	20.9	5.2	13.2	35.0	41.3
Config3 Mask	22.8	24.0	6.0	15.8	37.1	47.5
Config3 Fusion	22.6	23.8	6.1	15.9	37.1	46.4
Config3* Mask	21.6	NA	5.8	15.6	38.9	46.1
Config3* Fusion	22.1	NA	6.0	15.7	39.2	45.9

Table 3. Comparison of bezier parametric structure, mask structure and fusion strategies. Configurations vary in terms of backbones and training epochs and shown in Table 2. In some configurations, fusion options might increase the performance. * denotes the test results from the challenge. Test servers does not yield chamfer-based centerline detection performance for the test set, indicated as NA.

strategies are also explored. Table 3 shows the comparison of instance-mask and bezier based methodologies and their possible fusion. For the fusion, it is observed that accepting prediction for left/right direction from bezier predictions and up/down directions from mask predictions for some configurations is helpful. According to the results, masks have a superiority over bezier representation.

5. Conclusion and Future Work

In this work we introduced a novel instance-mask based methodology called TopoMask as an alternative to keypoint-based and parametric approaches commonly used in centerline detection. The performance of TopoMask was evaluated in comparison to the TopoNet architecture, and the results showed that TopoMask performs comparably to TopoNet in terms of Frechet-based centerline detection performance, while outperforming TopoNet in Chamfer-based centerline detection performance.

However, one limitation of the TopoMask architecture is its inability to predict the z information. Assigning zero values to the ground truth z leads to a Frechet-based detection score of approximately 0.92, indicating the potential for further improvement. Additionally, in the proposed architecture, some of the point sets of the centerlines are not monotonic, but the direction labels are assigned based on the dominant monotonicity of each axis. It is important to address these cases differently in future research. By applying the direction label strategy of TopoMask to the ground truths, a Frechet-based detection score of 0.92 is achieved, suggesting that there is still room for possible enhancement. These findings highlight the potential for refining the TopoMask methodology and addressing the mentioned limitations to further improve centerline detection performance.

ACKNOWLEDGMENTS

The numerical calculations reported in this paper were performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources)

References

- [1] Yigit Baran Can, Alexander Liniger, Danda Pani Paudel, and Luc Van Gool. Structured Bird’s-Eye-View Traffic Scene Understanding from Onboard Images. *Proceedings of ICCV*, pages 15641–15650, oct 2021. 1, 2
- [2] Yigit Baran Can, Alexander Liniger, Danda Pani Paudel, and Luc Van Gool. Topology preserving local road network estimation from single onboard camera image. In *Proceedings of CVPR*, pages 17263–17272, 2022. 2
- [3] Li Chen, Chonghao Sima, Yang Li, Zehan Zheng, Jiajie Xu, Xiangwei Geng, Hongyang Li, Conghui He, Jianping Shi, Yu Qiao, and Junchi Yan. Persformer: 3d lane detection via perspective transformer and the openlane benchmark. In *European Conference on Computer Vision (ECCV)*, 2022. 2
- [4] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *Proceedings of CVPR*, pages 1290–1299, 2022. 4
- [5] Zhengyang Feng, Shaohua Guo, Xin Tan, Ke Xu, Min Wang, and Lizhuang Ma. Rethinking efficient lane detection via curve modeling. In *Proceedings of CVPR*, pages 17062–17070, 2022. 2
- [6] Yuliang Guo, Guang Chen, Peitao Zhao, Weide Zhang, Jinghao Miao, Jingao Wang, and Tae Eun Choe. Gen-lanenet: A generalized and scalable approach for 3d lane detection. In *Proceedings of ECCV*, pages 666–681, 2020. 2, 3
- [7] Qi Li, Yue Wang, Yilun Wang, and Hang Zhao. Hdmapnet: An online hd map construction and evaluation framework. In *Int. Conf. Robotics and Automation (ICRA)*, pages 4628–4634. IEEE, 2022. 1, 2
- [8] Tianyu Li, Li Chen, Xiangwei Geng, Huijie Wang, Yang Li, Zhenbo Liu, Shengyin Jiang, Yuting Wang, Hang Xu, Chun-jing Xu, et al. Topology reasoning for driving scenes. *arXiv preprint arXiv:2304.05277*, 2023. 2, 4
- [9] Bencheng Liao, Shaoyu Chen, Bo Jiang, Tianheng Cheng, Qian Zhang, Wenyu Liu, Chang Huang, and Xinggang Wang. Lane graph as path: Continuity-preserving path-wise modeling for online lane graph construction. *arXiv preprint arXiv:2303.08815*, 2023. 2
- [10] Bencheng Liao, Shaoyu Chen, Xinggang Wang, Tianheng Cheng, Qian Zhang, Wenyu Liu, and Chang Huang. Maptr: Structured modeling and learning for online vectorized hd map construction. *arXiv preprint arXiv:2208.14437*, 2022. 1, 2
- [11] Lizhe Liu, Xiaohao Chen, Siyu Zhu, and Ping Tan. Condlanenet: a top-to-down lane detection framework based on conditional convolution. In *Proceedings of CVPR*, pages 3773–3782, 2021. 2
- [12] Shilong Liu, Feng Li, Hao Zhang, Xiao Yang, Xianbiao Qi, Hang Su, Jun Zhu, and Lei Zhang. DAB-DETR: Dynamic anchor boxes are better queries for DETR. In *International Conference on Learning Representations*, 2022. 4
- [13] Yicheng Liu, Yue Wang, Yilun Wang, and Hang Zhao. Vectormapnet: End-to-end vectorized hd map learning. *arXiv preprint arXiv:2206.08920*, 2022. 1, 2
- [14] Yingfei Liu, Junjie Yan, Fan Jia, Shuailin Li, Qi Gao, Tiancai Wang, Xiangyu Zhang, and Jian Sun. PETRv2: A Unified Framework for 3D Perception from Multi-Camera Images. *arXiv preprint arXiv:2206.01256*, 2022. 2
- [15] Bowen Pan, Jiankai Sun, Ho Yin Tiga Leung, Alex Andonian, and Bolei Zhou. Cross-view semantic segmentation for sensing surroundings. *IEEE Robotics and Automation Letters*, 5(3):4867–4873, 2020. 2
- [16] Xingang Pan, Jianping Shi, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Spatial as deep: Spatial cnn for traffic scene understanding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018. 2
- [17] Jonah Philion and Sanja Fidler. Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3d. In *Proceedings of the European Conference on Computer Vision*, 2020. 4
- [18] Zequn Qin, Huanyu Wang, and Xi Li. Ultra fast structure-aware deep lane detection. In *Proceedings of ECCV*, pages 276–291. Springer, 2020. 2, 4
- [19] Juyeb Shin, Francois Rameau, Hyeonjun Jeong, and Dong-suk Kum. Instagram: Instance-level graph modeling for vectorized hd map learning. *arXiv preprint arXiv:2301.04470*, 2023. 2, 4
- [20] Lucas Tabelini, Rodrigo Berriel, Thiago M Paixao, Claudine Badue, Alberto F De Souza, and Thiago Oliveira-Santos. Keep your eyes on the lane: Real-time attention-guided lane detection. In *Proceedings of CVPR*, pages 294–302, 2021. 2
- [21] Lucas Tabelini, Rodrigo Berriel, Thiago M Paixao, Claudine Badue, Alberto F De Souza, and Thiago Oliveira-Santos. Polylanenet: Lane estimation via deep polynomial regression. In *Int. Conf. Pattern Recognition (ICPR)*, pages 6150–6156. IEEE, 2021. 2
- [22] Huijie Wang, Zhenbo Liu, Yang Li, Tianyu Li, Li Chen, Chonghao Sima, Yuting Wang, Shengyin Jiang, Feng Wen, Hang Xu, et al. Road genome: A topology reasoning benchmark for scene understanding in autonomous driving. *arXiv preprint arXiv:2304.10440*, 2023. 1, 2, 3
- [23] Ruihao Wang, Jian Qin, Kaiying Li, and Dong Cao. Bev lane det: Fast lane detection on bev ground. *arXiv preprint arXiv:2210.06006*, 2022. 2
- [24] Benjamin Wilson, William Qi, Tanmay Agarwal, John Lambert, Jagjeet Singh, Siddhesh Khandelwal, Bowen Pan, Ratnesh Kumar, Andrew Hartnett, Jhony Kaesemodel Pontes, Deva Ramanan, Peter Carr, and James Hays. Argoverse 2: Next generation datasets for self-driving perception and forecasting. In *NeurIPS Datasets and Benchmarks*, 2021. 3
- [25] Hang Xu, Shaoju Wang, Xinyue Cai, Wei Zhang, Xiaodan Liang, and Zhenguo Li. Curvelane-nas: Unifying lane-sensitive architecture search and adaptive point blending. In *Proceedings of ECCV*, pages 689–704, 2020. 2
- [26] Zhenhua Xu, Yuxuan Liu, Yuxiang Sun, Ming Liu, and Lujia Wang. Centerlinedet: Road lane centerline graph detection with vehicle-mounted sensors by transformer for high-definition map creation. *arXiv preprint arXiv:2209.07734*, 2022. 2
- [27] Fan Yan, Ming Nie, Xinyue Cai, Jianhua Han, Hang Xu, Zhen Yang, Chaoqiang Ye, Yanwei Fu, Michael Bi Mi, and Li Zhang. Once-3dlanes: Building monocular 3d lane detection. In *Proceedings of CVPR*, pages 17143–17152, 2022. 2