Supplementary Material for Reconstruct from Top View: A 3D Lane Detection Approach based on Geometry Structure Prior

Chenguang Li* [1], Jia Shi*† [1,2], Ya Wang† [1,3], Guangliang Cheng‡ [1,4]
[1]SenseTime Research [2]Robotics Institute, Carnegie Mellon University
[3]University of Tuebingen [4]Shanghai AI Laboratory
{lichenguang, wangya}@senseauto.com jiashi@andrew.cmu.edu guangliangcheng2014@gmail.com

A. Point Pair Searching Algorithm

In this section, we present the searching algorithm on finding lane point pairs to impose geometry supervision, which is implemented as a greedy matching algorithm with linear time complexity. As shown in Algorithm 1, the input of this method is two lists of points on the corresponding neighbour lane boundaries, and the output is the matched key-value pairs of lane keypoints from the shorter lane boundary to the longer one.

B. Derivation of the 2D Geometry Constraint

In this section, we derive the 2D geometry prior constraint proposed in main paper. First we define the 3D Euclidean distance D_{3D} and 2D Euclidean distance D_{flat} in Equation 1,

$$\begin{aligned} D_{3D}(P_i^{3D}, P_j^{3D}) &= |\overrightarrow{P_i^{3D}P_j^{3D}}| \\ &= \sqrt{(x_i^{3D}-x_j^{3D})^2 + (y_i^{3D}-y_j^{3D})^2 + (z_i^{3D}-z_j^{3D})^2} \\ D_{flat}(P_i^{2D}, P_j^{2D}) &= |\overrightarrow{P_i^{2D}P_j^{2D}}| \\ &= \sqrt{(x_i^{2D}-x_j^{2D})^2 + (y_i^{2D}-y_j^{2D})^2} \end{aligned} \tag{1}$$

As illustrated in Figure 1, we define point P_A^{3D} and P_E^{3D} in 3D space, and derive the corresponding target points P_B^{3D} and P_F^{3D} from the formula of parallel curves in 2D parametric representation under the assumption of equal height in the z-axis between corresponding point pairs, as shown in Equation 2. Let c represent the constant distance between two parallel curves,

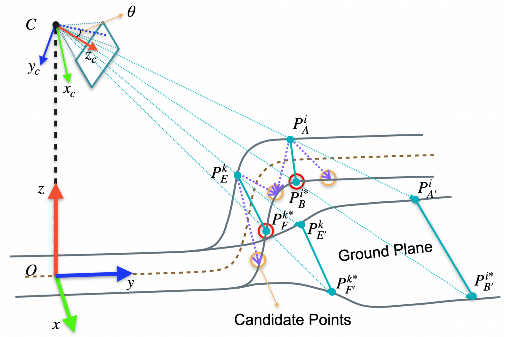

Figure 1. **Geometry prior of 3D lanes.**

$$\begin{aligned} P_A^{3D} &= \begin{pmatrix} x_A \\ y_A \\ z_A \end{pmatrix}; P_B^{3D} = \begin{pmatrix} x_A \pm \frac{c}{\sqrt{x_A'^2 + y_A'^2}} \cdot y_A' \\ y_A \pm \frac{c}{\sqrt{x_A'^2 + y_A'^2}} \cdot x_A' \\ z_A \end{pmatrix} \\ P_E^{3D} &= \begin{pmatrix} x_E \\ y_E \\ z_E \end{pmatrix}; P_F^{3D} = \begin{pmatrix} x_E \pm \frac{c}{\sqrt{x_E'^2 + y_E'^2}} \cdot y_E' \\ y_E \pm \frac{c}{\sqrt{x_E'^2 + y_E'^2}} \cdot x_E' \\ z_E \end{pmatrix} \end{aligned} \tag{2}$$

According to the virtual top view projection [2] in Equation 3, the 3D point is projected onto the ground plane g w.r.t. the camera height h and lane point height z. Thus we have the projection of 3D points P_A^{3D} and P_B^{3D} to the ground as P_A^{2D} and P_B^{2D} in Equation 4.

$$\begin{aligned} P^{2D} &= \overrightarrow{CP^{3D}} \cap \Pi_g \\ \begin{pmatrix} x^{2D} \\ y^{2D} \end{pmatrix} &= \frac{h}{h-z} \cdot \begin{pmatrix} x^{3D} \\ y^{3D} \end{pmatrix} \end{aligned} \tag{3}$$

*Both authors contributed equally.

†This work was done during internship at SenseTime Research.

‡Guangliang Cheng is the corresponding author.

Algorithm 1: Search point pairs between two lane boundaries using greedy matching with a sliding window

```
Require: lane boundary $L1$ with 3D points$\{(x_{i1}, y_{i1}, z_{i1}), i = \{1, 2, .., len(L1)\}\}$, lane boundary $L2$ with 3D points$\{(x_{j2}, y_{j2}, z_{j2}), j = \{1, 2, .., len(L2)\}\}$, constant window size $\eta$, constant threshold $\theta_{dist}$
Ensure: $dict\{(i : j)\}$ as a dictionary of indices for all matched pairs between lane boundary $L1$ and $L2$
$N1 \leftarrow len(L1)$
$N2 \leftarrow len(L2)$ if $N1 > N2$, swap($L1, L2$) // to ensure $L1$ is not larger than $L2$, i.e. we can always find a matched point in $L2$ with the corresponding point in $L1$ as a pair; Always start searching point pairs from the shorter lane boundary
$mid1 \leftarrow \frac{N1}{2}$ // the index of the middle number in $L1$, as the start point for searching
$mid2 \leftarrow index(Y_{ref}^{mid1})$ // generate the search start point in $L2$ from the identical y-reference of $mid1$
$mid2 \leftarrow [\frac{mid2-\eta}{2}, \frac{mid2+\eta}{2}]$ and $\underset{mid2}{argminDist_{3D}}(L1[mid1], L2[mid2])$ // to find the pair $(mid1, mid2)$ between two lane boundaries
// search backward:
for $(i = mid1 - 1; i > 1; i - -)$ do
    $j \leftarrow [mid2 - 1, mid2 - \eta]$ and $\underset{j}{argminDist_{3D}}(L1[i], L2[j])$
    $dict \leftarrow (i, j)$ // to find pair$(i, j)$ and add it to $dict$
    if $|minDist_{3D}(L1[i], L2[j]) - minDist_{3D}(L1[i - 1], L2[j - 1])| > \theta_{dist}$ then
        return NULL // to ensure the difference of lane width is not large locally
    end if
end for
// search forward:
for $(i = mid1 + 1; i < N1; i + +)$ do
    $j \leftarrow [mid2 + 1, mid2 + \eta]$ and $\underset{j}{argminDist_{3D}}(L1[i], L2[j])$
    $dict \leftarrow (i, j)$ // to find pair$(i, j)$ and add it to $dict$
    if $|minDist_{3D}(L1[i], L2[j]) - minDist_{3D}(L1[i + 1], L2[j + 1])| > \theta_{dist}$ then
        return NULL // to ensure the difference of lane width is not large locally
    end if
end for
return $dict$
```

$$P_A^{2D} = \begin{pmatrix} x_A \cdot \frac{h}{h-z_A} \\ y_A \cdot \frac{h}{h-z_A} \end{pmatrix}$$
$$P_B^{2D} = \begin{pmatrix} (x_A \pm \frac{c}{\sqrt{x_A'^2+y_A'^2}} \cdot y'_A) \cdot \frac{h}{h-z_A} \\ (y_A \pm \frac{c}{\sqrt{x_A'^2+y_A'^2}} \cdot x'_A) \cdot \frac{h}{h-z_A} \end{pmatrix} \tag{4}$$

Following the 3D geometry constraint presented in the paper, we assume the 3D lane has constant width c for each lane point pairs, as described in Equation 5,

$$D_{3D}(P_A^{3D}, P_B^{3D}) = D_{3D}(P_E^{3D}, P_F^{3D}) = c \tag{5}$$

thus we have the 2D Euclidean distance of point pairs $\{P_A^{2D}, P_B^{2D}\}$ and $\{P_E^{2D}, P_F^{2D}\}$as

$$\begin{aligned} &D_{flat}(P_A^{2D}, P_B^{2D}) \\ &= \sqrt{(\frac{h}{h - z_A})^2(\pm\frac{c \cdot y'_A}{\sqrt{x_A'^2 + y_A'^2}})^2 + (\frac{h}{h - z_A})^2(\pm\frac{c \cdot x'_A}{\sqrt{x_A'^2 + y_A'^2}})^2} \\ &= \sqrt{(\frac{h}{h - z_A})^2 \cdot (\frac{(c \cdot y'_A)^2 + (c \cdot x'_A)^2)}{x_A'^2 + y_A'^2})} \\ &= \sqrt{(\frac{h}{h - z_A})^2 \cdot (\frac{(c^2 \cdot (y_A'^2 + x_A'^2))}{x_A'^2 + y_A'^2})} \\ &= \sqrt{(\frac{h}{h - z_A})^2 \cdot c^2} \\ &= \boxed{\frac{h}{h - z_A} \cdot c} \iff h > z_A \end{aligned}$$

$$\text{Similarly, } D_{flat}(P_E^{2D}, P_F^{2D}) = \boxed{\frac{h}{h - z_E} \cdot c} \tag{6}$$

therefore

$$\begin{aligned} D_{flat}(P_E^{2D}, P_F^{2D}) \cdot (h - z_E)& \\ &= h \cdot c \\ &= h \cdot D_{3D}(P_E^{3D}, P_F^{3D}) \\ &= h \cdot D_{3D}(P_A^{3D}, P_B^{3D}) \\ &= D_{flat}(P_A^{2D}, P_B^{2D}) \cdot (h - z_A) \end{aligned} \tag{7}$$

Thus,we define the D_{2D} as the 2D Euclidean distance D_{flat} weighted by camera height h and the same lane height z. Thus

$$\begin{aligned} D_{2D}(P_E^{2D}, P_F^{2D}) &:= D_{flat}(P_E^{2D}, P_F^{2D}) \cdot (h - z_E) \\ D_{2D}(P_A^{2D}, P_B^{2D}) &:= D_{flat}(P_A^{2D}, P_B^{2D}) \cdot (h - z_A) \end{aligned} \tag{8}$$

therefore under the assumption of shared lane height between points, we have

$$D_{2D}(P_E^{2D}, P_F^{2D}) = D_{2D}(P_A^{2D}, P_B^{2D}) \tag{9}$$

C. Transformation matrix of augmentation

We show the transformation matrix for imposing lane augmentation on pitch (x), roll (y) and yaw (z) axes respectively as

$$\mathbf{R}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\phi & -sin\phi \\ 0 & sin\phi & cos\phi \end{bmatrix} \tag{10}$$

$$\mathbf{R}_y = \begin{bmatrix} cos\psi & 0 & sin\psi \\ 0 & 1 & 0 \\ -sin\psi & 0 & cos\psi \end{bmatrix} \tag{11}$$

$$\mathbf{R}_z = \begin{bmatrix} cos\theta & -sin\theta & 0 \\ sin\theta & cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{12}$$

D. Detailed Experimental Results

In this section, we first provide detailed quantitative results of lane lines in Table 1 and center lines in Table 2 on the data split of balanced scenes, rarely observed, visual variations and extra-long range following the same setting in the main paper. The joint metric results are evaluated on all methods within the identical data split in the same table, thus cross-split and cross-table comparison of the joint metric is meaningless. Table 3 presents the ablation study results on center lines, similar to the one for lane lines presented in our main paper.

For our proposed framework, the camera pose for view projection can be estimated by an optional pose regression branch jointly trained with the top view mask in the feature extraction network. As a result, we conduct a comparison for the 3D lane detection under predicted camera pose. We also report the top view mask intersection-over-union (IoU) for a better illustration of the effect on joint training of 2D lane mask and camera pose. As shown in Table 4, when using predicted camera pose, the IoU of predicted lane mask drops for roughly 3%, however, our proposed augmentation can greatly ease the accuracy drop under unstable camera pose prediction and result in only a slightly drop in the accuracy of the 3D lane detection.

Also, we conduct an ablation study of the proposed augmentation method. As shown in Table 5, the removal of any augmentation method will result in an accuracy drop on both segmentation and lane detection, which proves the effectiveness of all proposed augmentations. Specifically, the roll augmentation should be considered significant for the refinement of the segmentation mask by efficiently generating enriched lane patterns, and pitch augmentation is critical for generating new data with greater fluctuation of lane height which will boost the results of the offset on the far side.

E. Qualitative comparison

We present a detailed qualitative comparison of our proposed method in Figure 4, 5 and 6. Compared to previous methods, our proposed solution makes tremendously progress in noise elimination, outlier rejection, and structure preservation for 3D lanes. Result shows that our method remains robust even under extreme lightness and strong occlusion. Besides, we provide extra examples on top view mask predictions of [2] and ours in 3, which proves that our method could tremendously resolve the problem of feature confusion especially in the far end.

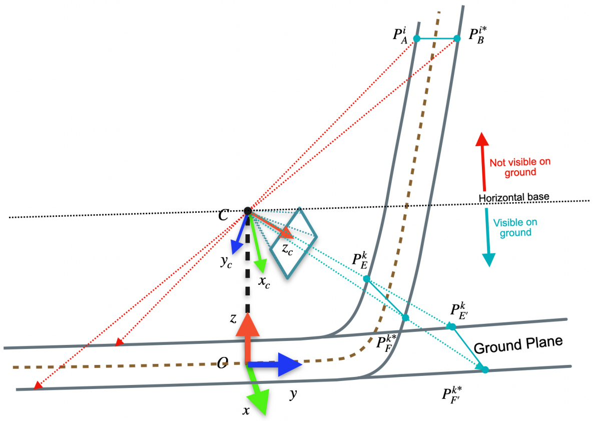

Figure 2. **An extreme case of lane height**

F. Limitation

In this paper, the proposed pipeline of 3D lane detection is based on 2D-3D lane reconstruction from the top view lane representation. Compared with previous image feature based 3D-LaneNet [1], the lane-mask based method could tremendously reduce the number of network parameters by extracting the lane height information from the flat ground lane representation under the virtual top view projection. However, one underlying limitation of such method is the detection range on lane height. For the virtual top view projection utilized in Gen-LaneNet [2] and our paper, the lane representation on flat ground is generated by projecting 3D lanes via a ray start from the camera center. As a result, as shown in Figure 2, for the cases when part of the lane exceeds the camera height, the 3D lane would be projected to the backside of the camera and become invisible from the top view. In this situation, the network could only make implicit prediction on the out-of-range lanes by following the geometry structure of the visible parts. Even though such hard cases exist, our method could make a certain improvement over Gen-LaneNet [2] on these cases by involving geometry prior.

For the vast majority of lane lines in reality, the lane height would not exceed the camera height within the detection range, which make it a trivial problem in most of the cases. Thus we choose not to propose a detailed solution in this paper for this problem and leave it to the future work.

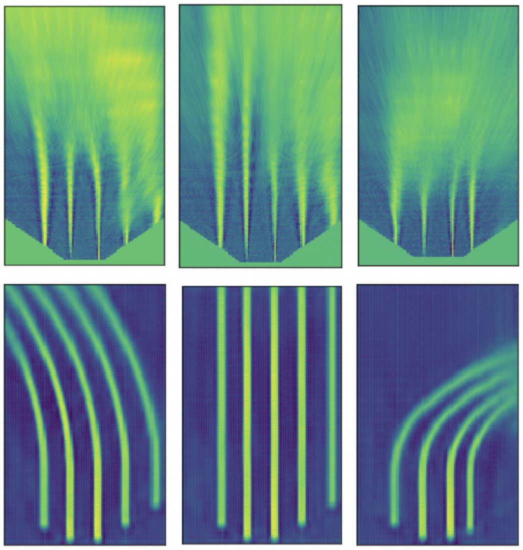

Figure 3. **Examples of front view supervision with mask projection [2] (top) and our top view supervision with feature projection (bottom).**

G. Future Work

For the problem mentioned in the limitation section, we consider to address this problem by two solutions in the future work. First, utilizing multi-view feature fusion for a combination of the full lane visibility in front view and the accurate mask representation in top view. Second, involving the “virtual camera view”, where multiple virtual cameras are utilized to simulate different installation height for the ensemble of detection results from various view projection. Thus, the edge case of extra-height lanes would be fully considered and the existing pipeline can be more robust under extreme cases.

References

[1] Noa Garnett, Rafi Cohen, Tomer Pe’er, Roee Lahav, and Dan Levi. 3d-lanenet: end-to-end 3d multiple lane detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2921–2930, 2019. 3, 5, 6

[2] 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 *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*, pages 666–681. Springer, 2020. 1, 3, 4, 5, 6

Table 1. Evaluation results on lane lines.

Dataset Split	Method	F-Score	AP	x error near (m)	x error far (m)	**joint** x error far (m)	z error near (m)	z error far (m)	**joint** z error far (m)
balanced scenes	3D-LaneNet [1]	86.4	89.3	0.068	0.477	0.454	0.015	0.202	0.186
	Gen-LaneNet [2]	88.1	90.1	0.061	0.496	0.480	0.012	0.214	0.196
	Ours	**91.9**	**93.8**	0.049	0.387	**0.340**	0.008	0.213	**0.175**
	Gen-LaneNet(GT mask) [2]	91.8	93.8	0.054	0.412	0.361	0.011	0.226	0.180
	Ours (GT mask)	**92.8**	**94.7**	0.044	0.360	**0.306**	0.007	0.219	**0.172**
rarely observed	3D-LaneNet [1]	72.0	74.6	0.166	0.855	0.906	0.039	0.521	0.551
	Gen-LaneNet [2]	78.0	79.0	0.139	0.903	0.950	0.030	0.539	0.570
	Ours	**83.7**	**85.2**	0.126	0.903	**0.870**	0.023	0.625	**0.567**
	Gen-LaneNet(GT mask) [2]	84.7	86.6	0.117	0.839	0.785	0.024	0.611	0.548
	Ours (GT mask)	**87.8**	**89.5**	0.101	0.791	**0.719**	0.017	0.605	**0.526**
visual variations	3D-LaneNet [1]	72.5	74.9	0.115	0.601	0.546	0.032	0.230	0.175
	Gen-LaneNet [2]	85.3	87.2	0.074	0.538	0.444	0.015	0.232	0.161
	Ours	**89.9**	**92.1**	0.06	0.446	**0.331**	0.011	0.235	**0.156**
	Gen-LaneNet(GT mask) [2]	90.2	92.3	0.073	0.502	0.385	0.013	0.249	0.150
	Ours (GT mask)	**91.3**	**93.2**	0.055	0.435	**0.309**	0.010	0.249	**0.155**
extra-long range	3D-LaneNet [1]	60.1	63.2	0.106	0.559	0.544	0.014	0.139	0.123
	Gen-LaneNet [2]	68.5	69.2	0.064	0.524	0.503	0.010	0.112	0.088
	Ours	**83.6**	**85.3**	0.039	0.290	**0.250**	0.007	0.087	**0.072**
	Gen-LaneNet(GT mask) [2]	80.7	82.5	0.052	0.335	0.300	0.011	0.097	0.084
	Ours (GT mask)	**87.2**	**89.1**	0.032	0.242	**0.221**	0.006	0.054	**0.047**

Table 2. Evaluation results on center lines.

Dataset Split	Method	F-Score	AP	x error near (m)	x error far (m)	**joint** x error far (m)	z error near (m)	z error far (m)	**joint** z error far (m)
balanced scenes	3D-LaneNet [1]	89.5	91.4	0.066	0.456	0.433	0.015	0.179	0.160
	Gen-LaneNet [2]	90.8	92.6	0.055	0.457	0.444	0.011	0.176	0.167
	Ours	**94.6**	**96.9**	0.046	0.346	**0.306**	0.007	0.185	**0.149**
	Gen-LaneNet (GT mask) [2]	94.5	96.8	0.050	0.372	0.325	0.010	0.190	0.149
	Ours (GT mask)	**95.0**	**97.2**	0.038	0.317	**0.273**	0.006	0.180	**0.140**
rarely observed	3D-LaneNet [1]	77.0	80.0	0.162	0.883	0.927	0.040	0.557	0.587
	Gen-LaneNet [2]	79.5	80.6	0.121	0.885	0.937	0.026	0.547	0.606
	Ours	**84.1**	**85.7**	0.127	0.887	**0.851**	0.024	0.625	**0.575**
	Gen-LaneNet (GT mask) [2]	85.9	87.7	0.105	0.845	0.812	0.022	0.622	0.576
	Ours (GT mask)	**87.7**	**89.7**	0.086	0.785	**0.743**	0.015	0.616	**0.559**
visual variations	3D-LaneNet [1]	75.5	77.7	0.120	0.636	0.578	0.030	0.227	0.174
	Gen-LaneNet [2]	88.2	90.0	0.072	0.438	0.430	0.015	0.187	0.143
	Ours	**92.2**	**94.3**	0.055	0.411	**0.300**	0.010	0.213	**0.131**
	Gen-LaneNet (GT mask) [2]	92.3	94.2	0.071	0.467	0.356	0.013	0.234	0.135
	Ours (GT mask)	**93.7**	**96.1**	0.055	0.397	**0.281**	0.009	0.222	**0.130**
extra-long range	3D-LaneNet [1]	62.2	64.0	0.106	0.559	0.526	0.014	0.139	0.071
	Gen-LaneNet [2]	69.4	70.1	0.058	0.507	0.465	0.009	0.082	0.040
	Ours	**85.8**	**87.6**	0.037	0.264	**0.204**	0.007	0.067	**0.032**
	Gen-LaneNet (GT mask) [2]	83.0	84.3	0.050	0.313	0.251	0.011	0.076	0.035
	Ours (GT mask)	**88.6**	**90.4**	0.031	0.205	**0.147**	0.007	0.054	**0.028**

Table 3. Ablation study on center line prediction.

Method	F-Score	AP	x error near (m)	x error far (m)	**joint** x error far (m)	z error near (m)	z error far (m)	**joint** z error far (m)
3D-LaneNet [1]	89.5	91.4	0.066	0.456	0.427	0.015	0.179	0.157
Gen-LaneNet [2]	90.8	92.6	0.055	0.457	0.435	0.011	0.176	0.164
Ours w/ GS	92.3	94.2	0.047	0.415	0.372	0.008	0.186	0.150
Ours w/ TVS	92.7	94.7	0.059	0.411	0.360	0.012	0.208	0.158
Ours w/ GS + TVS	93.7	95.9	0.062	0.377	0.333	0.008	0.187	0.150
Ours w/ GS + TVS + Aug	**94.6**	**96.9**	0.046	0.346	**0.300**	0.007	0.185	**0.146**
Gen-LaneNet (GT mask) [2]	94.5	96.8	0.050	0.372	0.318	0.010	0.190	0.145
Ours (GT mask)	**95.0**	**97.2**	0.038	0.317	**0.268**	0.006	0.180	**0.137**

Table 4. Ablation study on camera pose and augmentation.

Task	Method	Top-view mask IoU	F-Score	AP	x error near (m)	x error far (m)	**joint** x error far (m)	z error near (m)	z error far (m)	**joint** z error far (m)
Lane line	Pred cam pose w/o Aug	91.7	90.1	92.2	0.064	0.447	0.426	0.017	0.243	0.222
	Pred cam pose w Aug	92.9	91.0	93.2	0.066	0.444	0.407	0.019	0.240	0.222
	GT cam pose w/o Aug	94.7	91.2	93.2	0.065	0.415	0.394	0.009	0.220	0.207
	GT cam pose w Aug	**96.4**	**91.9**	**93.8**	0.049	0.387	**0.363**	0.008	0.213	**0.200**
Center line	Pred cam pose w/o Aug	91.7	92.9	94.9	0.058	0.419	0.394	0.017	0.213	0.196
	Pred cam pose w Aug	92.9	93.3	95.4	0.063	0.408	0.369	0.018	0.212	0.194
	GT cam pose w/o Aug	94.7	93.7	95.9	0.062	0.377	0.354	0.008	0.187	0.173
	GT cam pose w Aug	**96.4**	**94.6**	**96.9**	0.046	0.346	**0.322**	0.007	0.185	**0.169**

Table 5. Ablation study on 3D lane augmentation on different axes.

Task	Method	Top-view mask IoU	F-Score	AP	x error near (m)	x error far (m)	z error near (m)	z error far (m)
Lane line	No Aug	94.7	91.2	93.2	0.065	0.415	0.009	0.220
	Pitch + Yaw (w/o Roll)	95.0	91.3	93.3	0.051	0.402	0.012	0.219
	Pitch + Roll (w/o Yaw)	95.4	91.4	93.4	0.049	0.414	0.010	0.217
	Yaw + Roll (w/o Pitch)	95.4	91.2	93.3	0.060	0.438	0.011	0.228
	Pitch + Yaw + Roll (Ours)	**96.4**	**91.9**	**93.8**	0.049	0.387	0.008	0.213
Center line	No Aug	94.7	93.7	95.9	0.062	0.377	0.008	0.187
	Pitch + Yaw (w/o Roll)	95.0	94.2	96.6	0.050	0.359	0.010	0.188
	Pitch + Roll (w/o Yaw)	95.4	93.7	95.8	0.047	0.365	0.008	0.187
	Yaw + Roll (w/o Pitch)	95.4	93.5	95.7	0.057	0.386	0.010	0.191
	Pitch + Yaw + Roll (Ours)	**96.4**	**94.6**	**96.9**	0.046	0.346	0.007	0.185

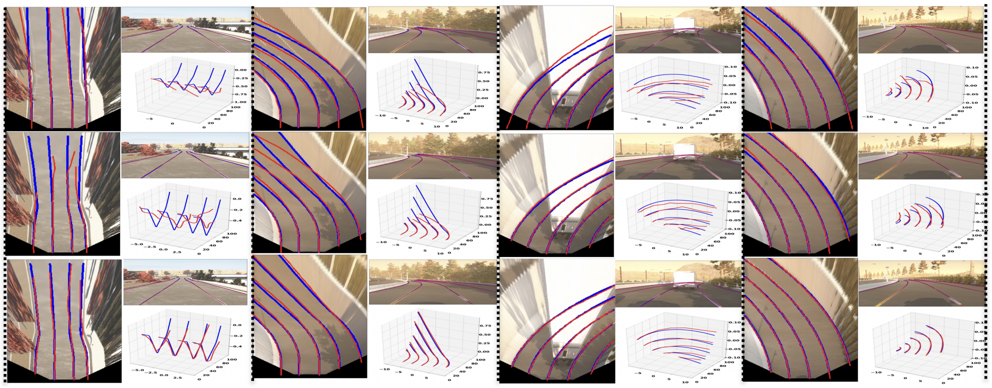

Figure 4. **Qualitative comparison results of proposed method on lane lines with a maximum range of 100m.** First row: 3D-LaneNet [1]; Second row: Gen-LaneNet [2]; Third row: Our proposed method.

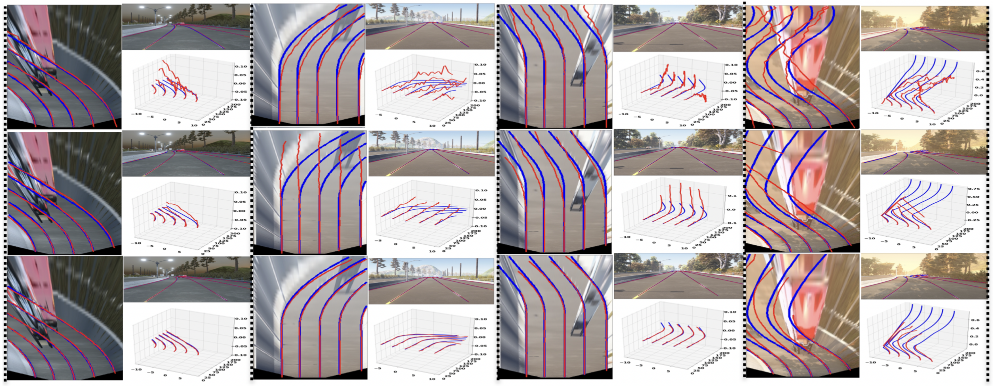

Figure 5. **Qualitative comparison results of proposed method on lane lines with a maximum range of 200m.** First row: 3D-LaneNet [1]; Second row: Gen-LaneNet [2]; Third row: Our proposed method.

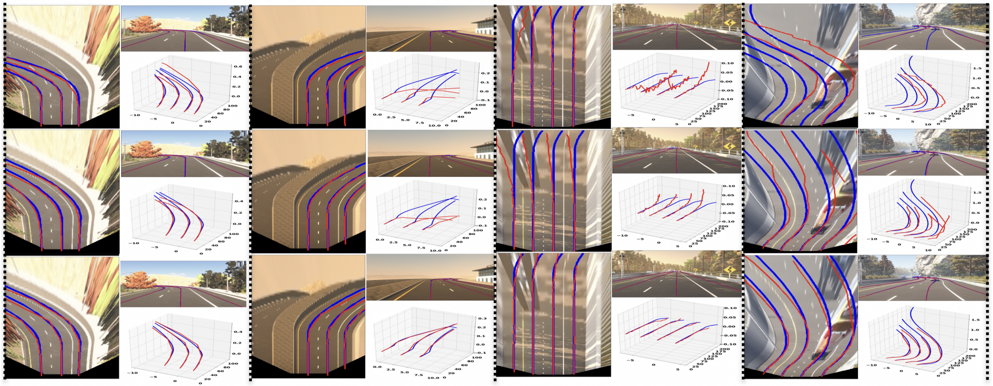

Figure 6. **Qualitative comparison results of proposed method on center lines.** First row: 3D-LaneNet [1]; Second row: Gen-LaneNet [2]; Third row: Our proposed method.