3D-SpLineNet: 3D Traffic Line Detection using Parametric Spline Representations - Supplementary material

Maximilian Pittner[1, 2], Alexandru Condurache[1, 2], Joel Janai[1]
[1]Bosch Mobility Solutions, Robert Bosch GmbH, 71229 Leonberg, Germany
[2]Institute of Signal Processing, University of Lübeck, 23562 Lübeck, Germany
{Maximilian.Pittner, AlexandruPaul.Condurache, Joel.Janai}@de.bosch.com

Abstract

This document serves as supplementary material of our paper "3D-SpLineNet: 3D Traffic Line Detection using Parametric Spline Representations". We provide details about our lane representation based on splines, the weighting function used for our proposed regression loss, as well as additional quantitative and qualitative results.

1. Spline representation

As described in Section 3.1 in the main paper, the line representation that we focus on in this work is based on B-Splines with additional offsets α_0, β_0, which simplifies to model mean shifts. Hence, the shape of a line $\boldsymbol{f_l}(t)$ is defined in the following way:

$$\boldsymbol{f_l}(t) = \begin{pmatrix} f_{\boldsymbol{l}_x}(t) \\ t \\ f_{\boldsymbol{l}_z}(t) \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{K_B} \alpha_k \cdot B_{k,d}(t) + \alpha_0 \\ t \\ \sum_{k=1}^{K_B} \beta_k \cdot B_{k,d}(t) + \beta_0 \end{pmatrix} . \quad (1)$$

We now consider a single one-dimensional spline function $f(t)$ in more detail, and ignore additional offset values for simplicity. The spline is defined as a linear combination of basis functions. Here we consider B-Splines as the basis functions, which can be imagined as piecewise defined recursive polynomials. Given K_B B-Spline basis functions $B_{k,d}$ of degree d we obtain

$$f(t) = \sum_{k=1}^{K_B} \theta_k \cdot B_{k,d}(t) \,, \quad (2)$$

where $\{\theta_k\}_{k=1}^{K_B}$ denotes the set of control points to weight each basis function, analogously to the control points $\{\alpha_k, \beta_k\}_{k=1}^{K_B}$ for the line x- and z-components.

The location and definition interval of the basis function depends on the pre-defined knots and the degree d. The expression of a recursively defined B-Spline basis function can be derived from the Cox-de Boor recursion formula [1] such that we obtain

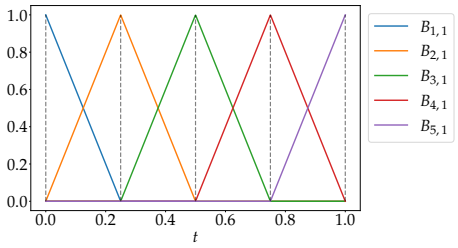
$$B_{k,d}(t) = \frac{t - t_k}{t_{k+d} - t_k} \cdot B_{k,d-1}(t) \quad (3)$$

$$+ \frac{t_{k+d+1} - t}{t_{k+d+1} - t_{k+1}} \cdot B_{k+1,d-1}(t) \quad (4)$$

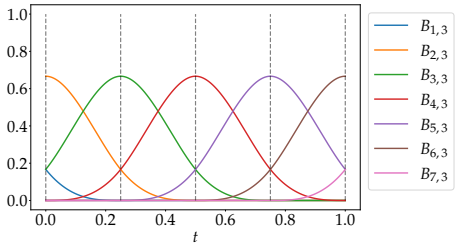
$$B_{k,0}(t) = \begin{cases} 1 & \text{if } t_i \leq t < t_{i+1}, \\ 0 & \text{else}\,. \end{cases} \quad (5)$$

As implied by the formula, each piecewise B-Spline basis function has a certain range along $d+1$ knots and can be computed recursively from its lower degree basis function, where the 0^{th} degree basis function is 1 along the range of knot k and 0 elsewhere. Since B-Splines and their linear combinations are fully differentiable we can use them as a continuous representation for the x- and z-line components of our lane detection head enabling us to train our full network end-to-end.

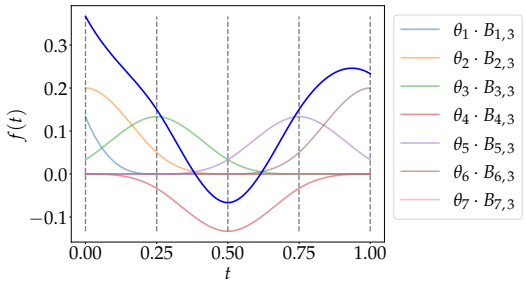
As discussed in Section 3.1 of the main paper, modeling the y-component as the scaled curve argument ($y(t) = \eta_y \cdot t$) restricts the representation to lanes that monotonously progress in driving direction. While the focus on this work does not lie on representing edge cases where lines do not progress monotonously in driving direction (horizontal stop lines, U-turns), we still want to experimentally investigate the capabilities of our representation. For this, we ran an experiment, in which our 3D-SpLineNet overfits to various challenging line geometries. We generated these lines synthetically to demonstrate the theoretical representation capabilities in critical scenarios, which are not included in the utilized dataset but might occur in more challenging real traffic environments. Fig. 2 shows that intersections with $90°$ turns as well as lane splits and merges can in theory be represented. Moreover, nearly horizontal lines are possible. The true limits of our representation can only be observed for practically horizontal lines of 0-$3°$.



(a) 1st degree B-Splines



(b) 3rd degree B-Splines



(c) Full spline function

Figure 1: B-Spline basis functions of 1st (a) and 3rd degree (b) with 5 knots located at positions indicated by the vertical dashed lines. (c) shows a full spline curve (*blue*) as a linear combination of 3rd degree basis functions weighted by the control points θ_k.

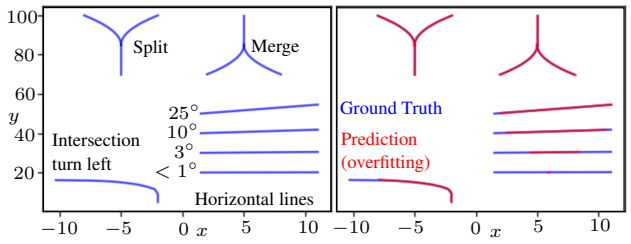
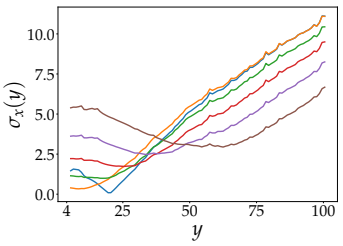
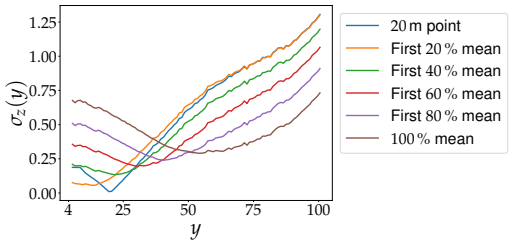


Figure 2: Results of an overfitting experiment (illustrated in top-view) to demonstrate the theoretical capabilities of our chosen lane representation.



(a) x-component



(b) z-component

Figure 3: Standard deviations of line x- and z-component from the reference along y-direction. We subtract the reference area from the line and associate each line point to a bin (corresponding to a certain y-interval) of a histogram. We then compute the standard deviations for all bins and obtain $\sigma_x(y)$ and $\sigma_z(y)$.

2. Weighting function

As described in Section 3.3 the shape regression loss

$$\mathcal{L}_s(\boldsymbol{l}, \hat{\boldsymbol{l}}) = \int_{\hat{t}_s}^{\hat{t}_e} \Big(w_x(t) \cdot \Big| f_{\boldsymbol{l}_x}(t) - \frac{1}{\eta_x}\hat{x}(t) \Big| \tag{6}$$

$$+ w_z(t) \cdot \Big| f_{\boldsymbol{l}_z}(t) - \frac{1}{\eta_z}\hat{z}(t) \Big| \Big)\, dt \tag{7}$$

includes a weighting function $\boldsymbol{w}(t) = [w_x(t), w_y(t), w_z(t)]^T$, which assigns a weight to each difference term depending on t. The weighting is required since the distribution of lines vary along t, i.e. along the direction of ego-motion (y-direction). More specifically, the distances from the ground truth association reference area show different standard deviations along y-direction as illustrated in Fig. 3.

Obviously, points in the y-range close to the reference show small deviations and points far off the reference show larger deviations. Consequently, without a weighting along the y-range the difference values with more significant deviations would make a higher contribution to the overall loss. In order to treat all distances along the y-axis evenly, we need to divide the difference term by the standard deviation while considering the normalization factors. Thus, the weighting functions for the x- and z-component are defined

as

$$w_x(t) = \frac{\eta_x}{\sigma_x(t)} \text{ and } w_z(t) = \frac{\eta_z}{\sigma_z(t)} \text{ with } t = \frac{y}{\eta_y}\,. \quad (8)$$

Since the y-component does not contribute to the loss, we set $w_y(t) = 0$.

Without using the weighting function we observed a less balanced minimization of the near-range error and a final increase of $+2.9\,\text{cm}$ for the x-near error and $+0.1\,\text{cm}$ for the z-near error.

3. Quantitative results

In this section, we provide additional quantitative results for experiments of our ablation studies from Section 4.2 of the main paper as well as additional results regarding the state-of-the-art comparison.

Table 1 shows the results of our ablation study analyzing all considered association strategies. The detection scores on both datasets confirm our argumentation that using mean matching leads to better detection performance than using fixed point matching. On the Standard dataset, varying the range for mean matching does not lead to significant changes in the performance metrics. Still, we obtain the best results for $40\,\%$ on the Rare Scenes test set and $40\,\% - 60\,\%$ on the Standard test set. Regarding the geometric accuracy, the results show that fixed point matching achieves similar performance compared to mean matching in the range of $40\,\% - 60\,\%$. A possible reason for this is that our chosen straight line initializations fit well for the fixed point ground truth association. We expect that using more sophisticated initializations will lead to lower geometric errors for the mean based matching and will therefore outperform fixed point matching on all evaluation metrics.

Table 2 provides a comparison of the considered representations on the Standard test set. The comparison shows similar relations among the results of the representations as on the Rare Scenes test set. While our chosen representation based on B-Splines with 15 knots and 3rd degree shows the best geometric accuracy, it achieves second best results on the detection scores. As discussed in Section 4.2 of the main paper, the representations based on B-Splines with 3rd degree and a lower number of knots (e.g. 3rd degree, 5 knots) show similar performance and thus provide a good alternative with a lower amount of required parameters.

During our experiments we realized a large gap in the number of training epochs compared to Gen-LaneNet [2]. To guarantee a fair comparison, we also trained Gen-LaneNet for the same 300 epochs (like ours). The results for the two configurations of Gen-LaneNet evaluated on the Standard dataset are shown in Table 3. The results indicate only small performance improvements apart from an increasing z-far error. In contrast, we observed that training our network results in a decreasing regression error beyond 200 epochs. We conclude that training our full architecture end-to-end on the detection task leverages the full capacity of the backbone, while solely training the detection head of Gen-LaneNet leads to earlier saturation with respect to the detection performance. This further confirms the superiority of 3D-SpLineNet over Gen-LaneNet.

4. Qualitative results

In this section, we show additional examples from all three test datasets to compare the predictions of our 3D-SpLineNet to Gen-LaneNet [2]. Apart from cases where our model detects lanes accurately, we also show failure cases, for which we see potential for improvements.

One suggestion to improve our method is to use a larger variety of line initializations. This could improve the prediction in scenarios like depicted in Fig. 5 (last row), where our method tries to model the straight rotated road as a curve instead of rotated straight lines, which could be added as initializations. Other scenarios, where better suiting intitializations might lead to benefits, are shown in Fig. 4 (second last row) and Fig. 5 (second last row), where lines deviate significantly from straight line initializations in both x- and z-direction.

Moreover, 3D-SpLineNet faces problems with sharp changes in z-direction (last row in Fig. 6). We see potentials for improvements for such cases by using other basis functions for the splines that are more suitable to model sharp edges in the z-profile.

In summary, we see a larger variety of initializations and the investigation of other spline basis functions as promising research directions for future work.

References

[1] Carl de Boor. On calculating with b-splines. *Journal of Approximation Theory*, 6(1):50–62, 1972.

[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 *Proc. of the European Conf. on Computer Vision (ECCV)*, 2020.

Dataset	Reference	F-Score	AP	x-error		z-error	
				near	far	near	far
Standard	20 m	95.7 %	97.9 %	**3.4**	33.5	**0.7**	21.3
	First 20 %	96.2 %	98.1 %	3.5	34.4	0.8	22.1
	First 40 %	**96.3 %**	98.1 %	3.7	32.4	0.9	21.3
	First 60 %	**96.3 %**	**98.3 %**	4.2	33.0	0.8	21.5
	First 80 %	96.1 %	98.0 %	5.1	**31.8**	1.1	**21.1**
	100 %	95.9 %	98.0 %	5.3	31.9	1.1	21.3
Rare Scenes	20 m	90.7 %	92.5 %	**6.8**	70.7	1.9	56.8
	First 20 %	92.1 %	94.1 %	7.0	71.2	**1.8**	57.5
	First 40 %	**92.9 %**	**94.8 %**	7.7	69.9	2.1	56.2
	First 60 %	92.5 %	94.3 %	9.2	69.4	**1.8**	55.9
	First 80 %	92.4 %	94.6 %	10.5	**67.4**	2.7	**55.7**
	100 %	91.2 %	93.4 %	12.1	70.1	2.6	56.8

Table 1: Comparison of detection scores and geometrical accuracy for different ground truth association methods on the Standard and Rare Scenes test sets. All distance metrics are provided in cm.

Representation	d	N	F-Score	AP	x-error		z-error	
					near	far	near	far
Polynomial	2	-	95.2 %	97.4 %	5.9	39.1	1.0	22.0
	3	-	95.8 %	97.7 %	5.3	35.1	1.1	22.0
	5	-	96.1 %	98.1 %	4.2	34.8	**0.9**	22.1
B-Splines	1	3	93.5 %	95.7 %	14.0	57.9	1.2	22.4
	3	3	95.8 %	97.9 %	4.7	34.0	1.0	21.5
	1	5	96.0 %	97.8 %	7.0	36.2	1.0	21.5
	3	5	**96.5 %**	**98.3 %**	4.0	32.9	**0.9**	21.5
	1	10	96.1 %	98.1 %	4.2	34.2	**0.9**	21.6
	3	10	96.2 %	98.0 %	4.0	34.1	**0.9**	21.4
	1	15	96.3 %	97.9 %	4.8	33.9	**0.9**	**21.3**
	3	15	96.3 %	98.1 %	**3.7**	**32.4**	**0.9**	**21.3**

Table 2: Comparison of geometrical accuracy and F-Score resulting from different representations on the Standard test set. d denotes the degree and N the number of knots of the B-Splines. All distance metrics are provided in cm.

Method	Number of Epochs	F-Score	AP	x-error		z-error	
				near	far	near	far
Gen-LaneNet [2]	30	88.1 %	90.1 %	6.1	49.6	1.2	21.4
	300	89.0 %	91.1 %	4.6	47.7	1.0	23.9

Table 3: Evaluation of Gen-LaneNet trained for different numbers of epochs on the Standard test set. All distance metrics are provided in cm.

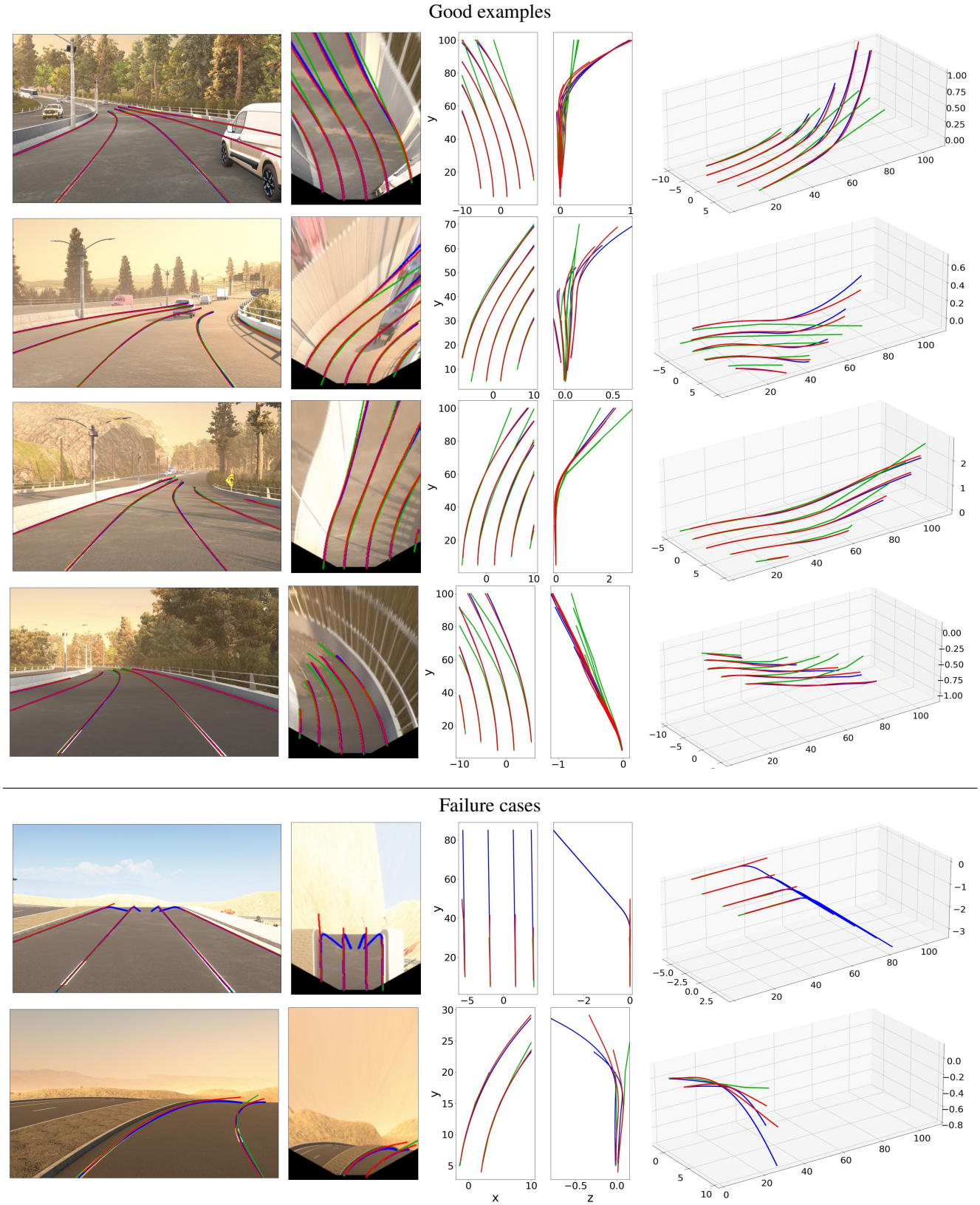


Figure 4: Predictions (*3D-SpLineNet* vs. *Gen-LaneNet* [2]) and *ground truth* from the *Standard* test set illustrated in front-view, top-view, x- and z-profile and 3D space.

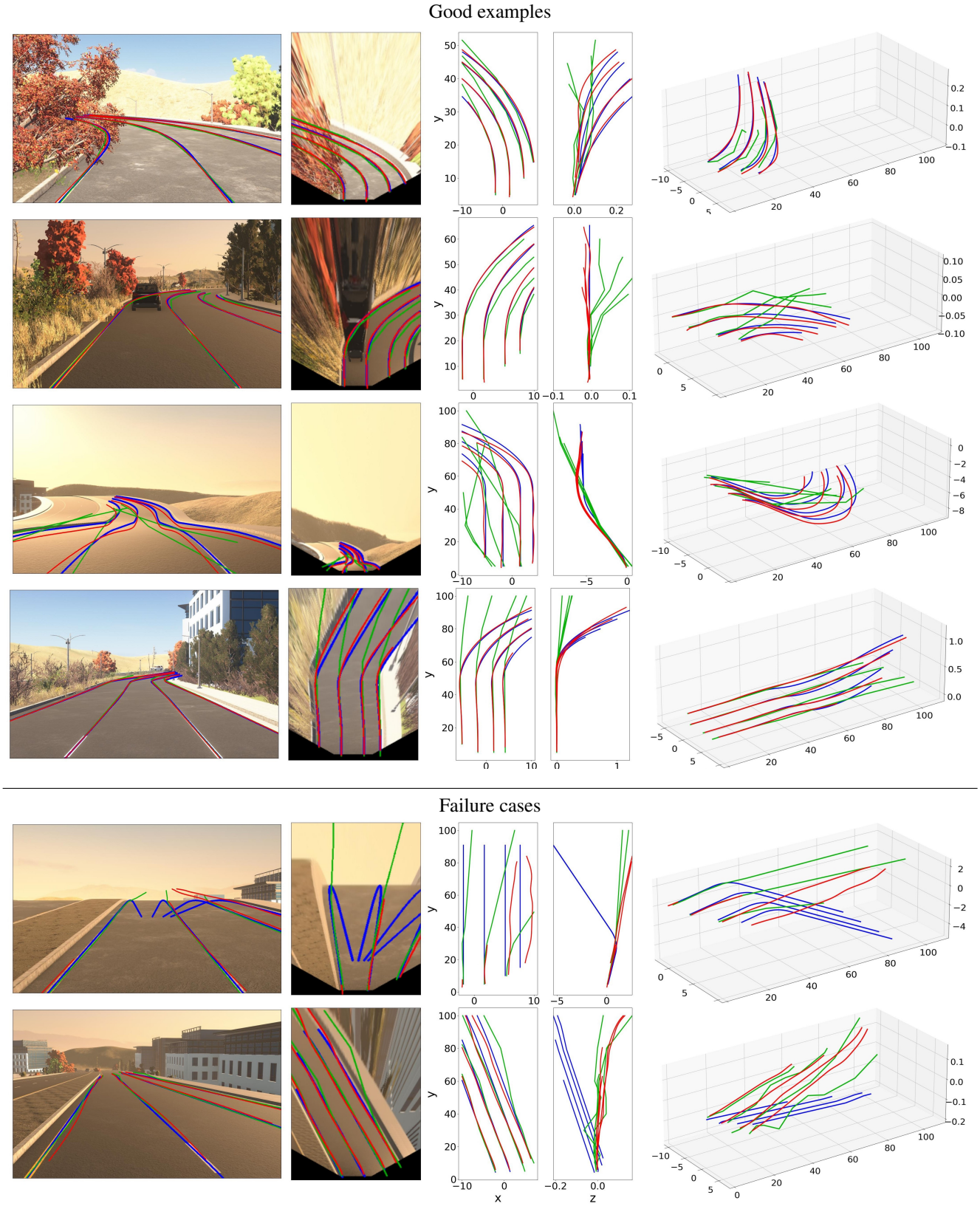


Figure 5: Predictions (*3D-SpLineNet* vs. *Gen-LaneNet* [2]) and *ground truth* from the *Rare Scenes* test set illustrated in front-view, top-view, x- and z-profile and 3D space.

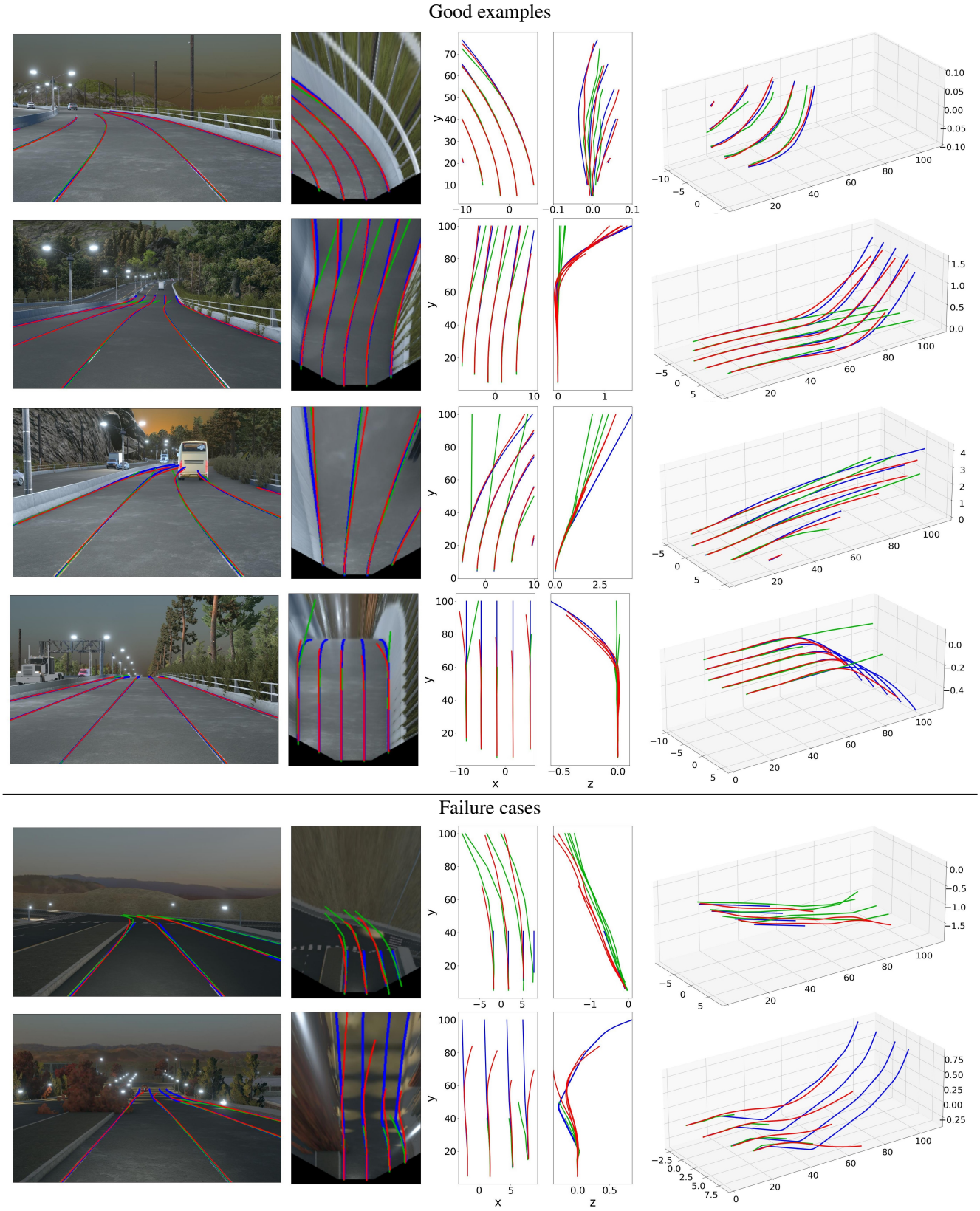


Figure 6: Predictions (*3D-SpLineNet* vs. *Gen-LaneNet* [2]) and *ground truth* from the *Visual Var.* test set illustrated in front-view, top-view, x- and z-profile and 3D space.