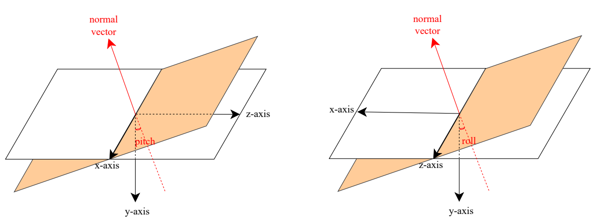
Homography Guided Temporal Fusion for Road Line and Marking Segmentation Appendix

Shan Wang[1,2] Chuong Nguyen[1] Jiawei Liu[2] Kaihao Zhang[2] Wenhan Luo[3]
Yanhao Zhang[2] Sundaram Muthu[1] Fahira Afzal Maken[1] Hongdong Li[2]
[1]Data61, CSIRO [2]Australian National University [3]Sun Yat-sen University

1. Derivation of Jacobian

In this section, we present the detailed derivations for the Equation. 9, Equation. 10 and Equation. 11 of the main paper, which calculate the Jacobian of Homography Transformation w.r.t. the pitch θ and the roll ϕ angle. The pitch angle θ and the roll angle ϕ represent the rotation of the road surface w.r.t. the plane constructed with the x-axis and z-axis of camera coordinates, as shown in Fig. 1.



(a) Pitch: rotation around x-axis (b) Roll: rotation around z-axis

Figure 1: Pitch and Roll of Road Surface Normal Vector Components

Derivation of Equation. 9 of the Main Paper is the Jacobian of Homography Transformation (Equation. 1 of the Main Paper) w.r.t. road surface normal $\mathbf{n}$.

$$\begin{aligned}\frac{\partial p_i^j}{\partial \mathbf{n}} &= \frac{\partial(\mathbf{K}(\mathbf{R}_i - \frac{\mathbf{t}_i\mathbf{n}^\top}{d})\mathbf{K}^{-1}(p_t \oplus 1))}{\partial \mathbf{n}} \\ &= \frac{\partial(\mathbf{K}\mathbf{R}_i\mathbf{K}^{-1}(p_t \oplus 1))}{\partial \mathbf{n}} - \frac{\partial(\mathbf{K}(\frac{\mathbf{t}_i\mathbf{n}^\top}{d})\mathbf{K}^{-1}(p_t \oplus 1))}{\partial \mathbf{n}} \\ &= 0 - \frac{1}{d}\frac{\partial(\mathbf{K}\mathbf{t}_i\mathbf{n}^\top\mathbf{K}^{-1}(p_t \oplus 1))}{\partial \mathbf{n}} \\ &= -\frac{1}{d}\mathbf{K}\mathbf{t}_i(\mathbf{K}^{-1}(p_t^j \oplus 1))^\top,\end{aligned} \tag{1}$$

Derivation of Equation. 10 of the Main Paper is the Jacobian of road surface normal w.r.t. the pitch angle θ.

$$\begin{aligned}\frac{\partial \mathbf{n}}{\partial \theta} &= (\frac{\partial(-\sin\phi\cos\theta)}{\partial\theta}, \frac{\partial(-\cos\phi\cos\theta)}{\partial\theta}, \frac{\partial(\sin\theta)}{\partial\theta}) \\ &= (\sin\phi\sin\theta, \cos\phi\sin\theta, \cos\theta) \\ &= (\frac{\sin\phi\cos\theta\sin\theta}{\cos\theta}, \frac{\cos\phi\cos\theta\sin\theta}{\cos\theta}, \cos\theta) \\ &= (-\frac{\mathbf{n}_1\mathbf{n}_3}{\cos\theta}, -\frac{\mathbf{n}_2\mathbf{n}_3}{\cos\theta}, \cos\theta) \\ &= (-\frac{\mathbf{n}_1\mathbf{n}_3}{\sqrt{1-\mathbf{n}_3^2}},\ -\frac{\mathbf{n}_2\mathbf{n}_3}{\sqrt{1-\mathbf{n}_3^2}},\ \sqrt{1-\mathbf{n}_3^2}),\end{aligned} \tag{2}$$

Derivation of Eq. 11 of the Main Paper is the Jacobian of road surface normal w.r.t. the roll angle ϕ.

$$\begin{aligned}\frac{\partial \mathbf{n}}{\partial \phi} &= (\frac{\partial(-\sin\phi\cos\theta)}{\partial\phi}, \frac{\partial(-\cos\phi\cos\theta)}{\partial\phi}, \frac{\partial(\sin\theta)}{\partial\phi}) \\ &= (-\cos\phi\cos\theta, \sin\phi\cos\theta), 0) \\ &= (\mathbf{n}_2,\ -\mathbf{n}_1,\ 0),\end{aligned} \tag{3}$$

2. Performance with Estimated Camera Extrinsic

One limitation of our method is that it requires additional camera extrinsic information. Although it is easy to obtain in a real vehicle system, it may not be available for some datasets. To explore our performance under extrinsic unknown settings, we estimated the camera extrinsic using COLMAP [9] and used the estimated extrinsic to evaluate the proposed method. However, the ApolloScape dataset's larger camera elevation angle caused the key points obtained by COLMAP to be further away from the vehicle, resulting in noisy camera poses. Despite these less accurate poses, our method still achieved plausible performances ("w/ COLMAP Extrinsic" in Tab. 1).

3. Performance with Estimated Homography

We also use FindHomography function from OpenCV package to directly estimate the homography transformation matrix. The estimations are less accurate than our HomoGuide, leading to inferior lane segmentation results ("w/ OpenCV Homography" of Tab. 1).

Table 1: Comparison of Estimated Extrinsic and OpenCV Homography on ApolloScape Dataset

	18 mIoU↑	36 mIoU↑
HomoFusion (ours)	**59.3**	**35.9**
w/ COLMAP Extrinsic	56.2	34.8
w/ OpenCV Homography	47.9	29.4

4. Challenging Scenarios

In this section, we evaluate the performance of our method on challenging scenarios. As we utilize the LM algorithm for road surface normal vector estimation, there is a convex range of the algorithm that can result in failure due to excessive re-projection errors. To further investigate this scenario, we selected the top 100 inputs with the highest residuals after optimization. We compared our approach with and without RSNE and found that incorporating the RSNE led to improved performance (18 mIoU: 53.3 and 52.5, respectively), even in cases where the optimization has a high possibility of being non-convex.

Additionally, we investigate whether our method successfully estimates the road surface normal vector under uphill/downhill road scenarios. Fig. 2 demonstrates that the normal is accurately estimated in uphill scenarios, resulting in good alignment of road marks across multiple frames.

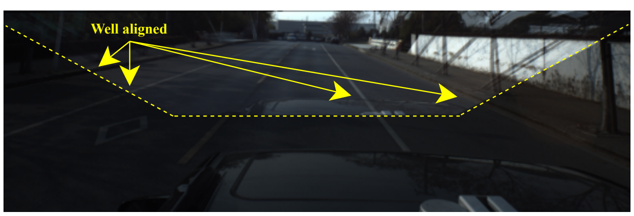


Figure 2: Correctly fused uphill frames using estimated normal.

5. Dataset Selection Criteria

Our method relies on borrowing information from adjacent frames, and therefore datasets that do not include continuous frames with a common field of view, such as CeyMo [6], CULane [10], and VPGNet [8], are unsuitable for our purposes. These datasets do not provide the necessary information for our method to work effectively, as they lack the continuity required for the information to be meaningfully borrowed across frames. In addition, although vehicle pose is free information that we can exploit, transitional lane mark detection datasets, such as , LLAMAS [2], SDLane [7] and VIL-100 [14], do not provide camera intrinsic and extrinsic information, nor do they provide GPS/IMU data that could be used to calculate camera global extrinsic information. These datasets are unsuitable for our method because they lack the necessary information for our approach to work effectively. Furthermore, our method is a segmentation task, and as such, datasets that do not provide segmentation masks, such as OpenLane [3], Tusimple [10] are also unsuitable. Without segmentation masks, it is impossible to accurately determine the boundaries of lane markings in the image, making it difficult to apply our method effectively. While the Waymo Open Dataset [12] primarily targets general panoptic segmentation, our lane marker segmentation task focuses on only two categories—lane markers and road markers—limiting its utility. Our comparison solely involves CFFM [11] on this dataset, highlighting our approach's enhanced efficacy in Table 2.

Table 2: Comparison on Waymo Open Dataset

	CFFM [11]	Ours
3 mIoU↑	63.22	**66.45**

Given these limitations, our primary experimentation is carried out on the ApolloScape dataset [5], as it offers the necessary information for our approach. In addition, we created an artificial dataset called ApolloScape Night from the ApolloScape dataset using a cross-domain generation network [1]. This dataset allows us to evaluate the effectiveness of our method under challenging lighting conditions. We conducted experiments on these datasets to demonstrate the effectiveness of our method across various domains, including daytime and nighttime driving scenarios. The results show that our approach achieves SOTA performance on these datasets, confirming its effectiveness and suitability for real-world lane detection applications.

6. Visualization on Real Night/Rain Scenes

We verify the effectiveness of our proposed method in challenging real-world environments, *e.g.* dark and rainy situations. Fig. 3 and Fig. 4 present the segmentation results of our proposed method on real images taken at night and in rainy conditions, respectively. Despite being trained only on the artificial ApolloScape Night dataset, which simulates road conditions at night, our proposed method successfully segments various road lines and markings in real images.

7. Visualization on HomoGuide

This section complements Section. 4.6 of the Main Paper on the "Impact of HomoFusion". Fig. 7 compares our proposed method with and without Homo Guide. It illustrates

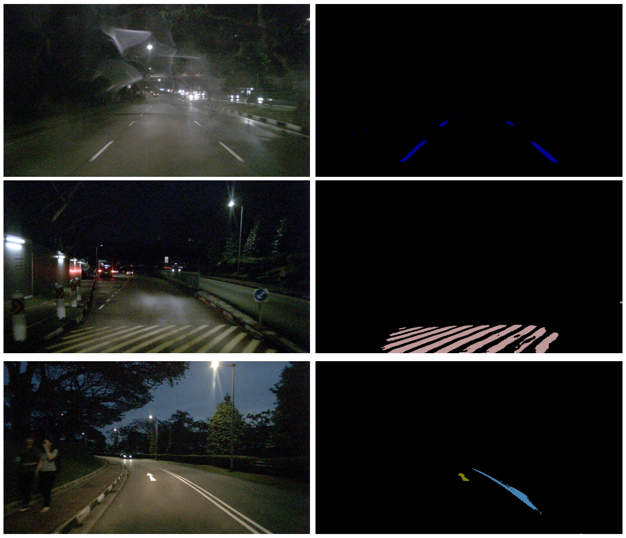


Figure 3: Segmentation results (Right) of real night frames (Left) using the model trained on the artificial ApolloScape Night dataset. Our proposed method accurately segments various road lines and markings, such as dot lines, solid lines, crosswalks, and straight arrows.

Figure 4: Segmentation results (Right) of real rainy day frames (Left) using the model trained on the artificial ApolloScape Night dataset. The proposed method accurately segments various road lines and markings, including dot lines, solid lines, double yellow lines, stop lines, and crosswalks.

that incorporating HomoGuide enables our method to accurately classify road lines and markings even under adverse conditions such as occlusion, road reflection, and poor light conditions.

8. Visualization of Water Hazard Detection

This section complements Section. 4.6 of the Main Paper on the "Application to Another Task", we provide a visualization of the predicted segmentation and the input frames transformed by Homography in Fig. 5 and Fig. 6, respectively. These figures demonstrate that our proposed HomoFusion can align ground surfaces well enough to improve the performance of detection of flat objects on the ground.

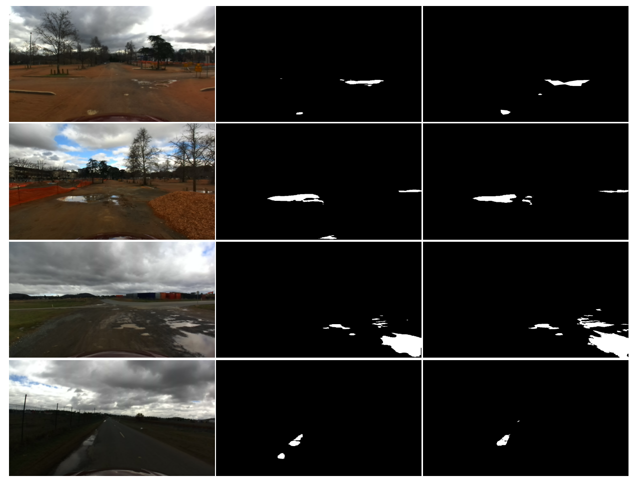


Figure 5: Current frame, prediction and ground truth masks of water puddle segmentation. The 1^{st} and 2^{nd} rows are from the Off-road dataset, and the 3^{nd} and 4^{th} rows are from the On-road dataset.

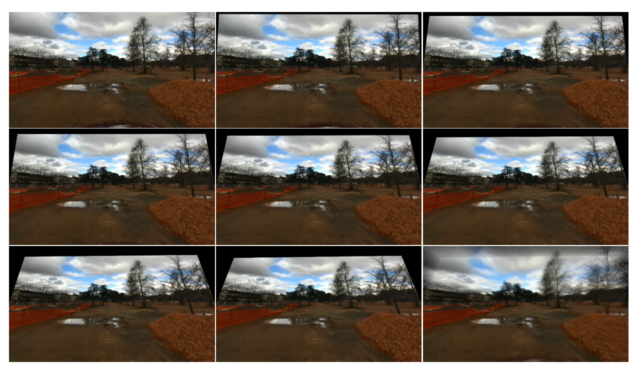


Figure 6: Input frames transformed to the current frame by homography using the estimated road surface normal vectors and camera movement. The fused image of all eight frames is shown in the bottom right corner.

9. Extra Qualitative Examples

Fig. 8 demonstrates that our proposed method can accurately segment the road lines and markings of various categories.

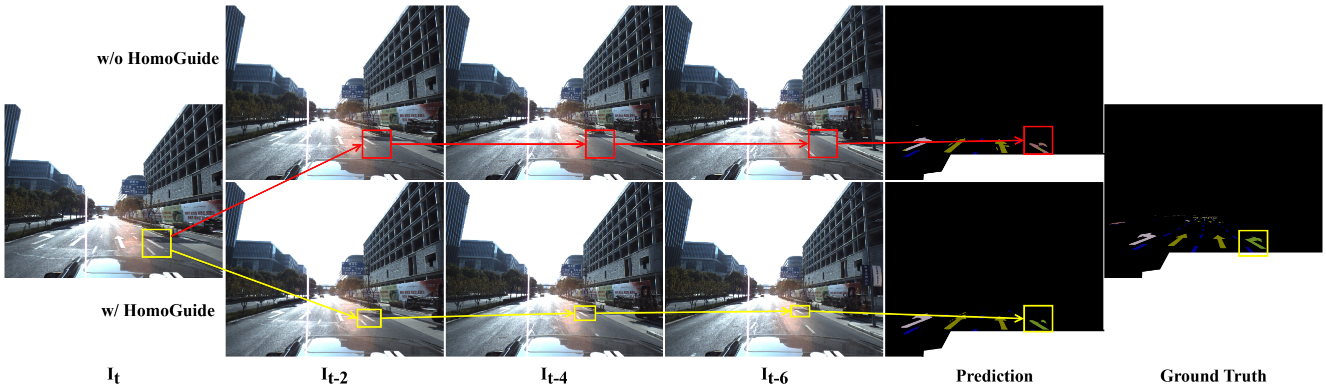


Figure 7: Impact of HomoGuide on segmentation results. Yellow boxes indicate the correspondence across frames with HomoGuide and Red boxes indicate the correspondence without HomoGuide. HomoGuide facilitates the correct classification of partially occluded road markings, as shown by the example of the partially occluded right turn arrow.

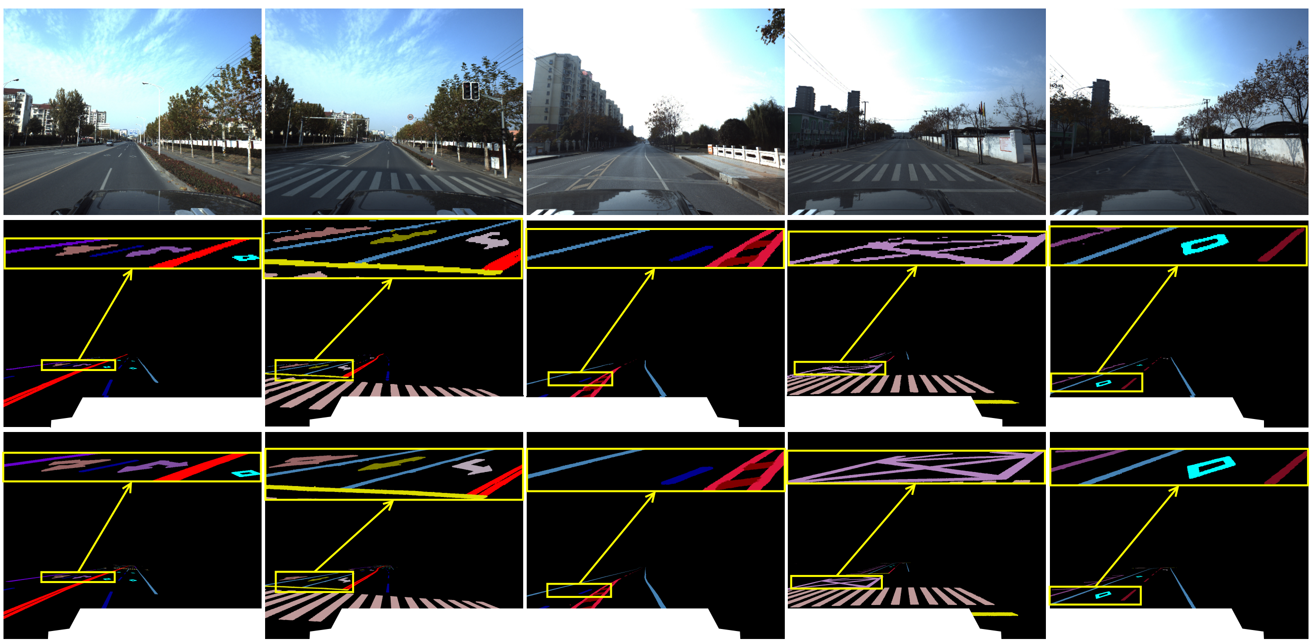


Figure 8: Examples of prediction on various road lines and markings. From top to bottom are input images, prediction of our model, and ground truth. Best viewed in color.

Furthermore, we provide additional qualitative examples of our proposed method and state-of-the-art (SOTA) algorithms, including (a) IntRA-KD [4], (b) SegFormer [13], and (c) CFFM [11] and (d) MMA-Net [14], on ApolloScape [5] in Fig.9 and ApolloScape Night in Fig.10. Our method demonstrates better true-positive predictions and fewer false-positive predictions in both scenarios. Additionally, our method accurately predicts more precise boundaries even in challenging glare and poor lighting conditions.

References

[1] Vinicius F Arruda, Thiago M Paixão, Rodrigo F Berriel, Alberto F De Souza, Claudine Badue, Nicu Sebe, and Thiago Oliveira-Santos. Cross-domain car detection using unsupervised image-to-image translation: From day to night. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019. 2

[2] Karsten Behrendt and Ryan Soussan. Unsupervised labeled lane markers using maps. In *Proceedings of the IEEE International Conference on Computer Vision*, 2019. 2

[3] Li Chen, Chonghao Sima, Yang Li, Zehan Zheng, Jiajie Xu, Xiangwei Geng, Hongyang Li, Conghui He, Jianping Shi,

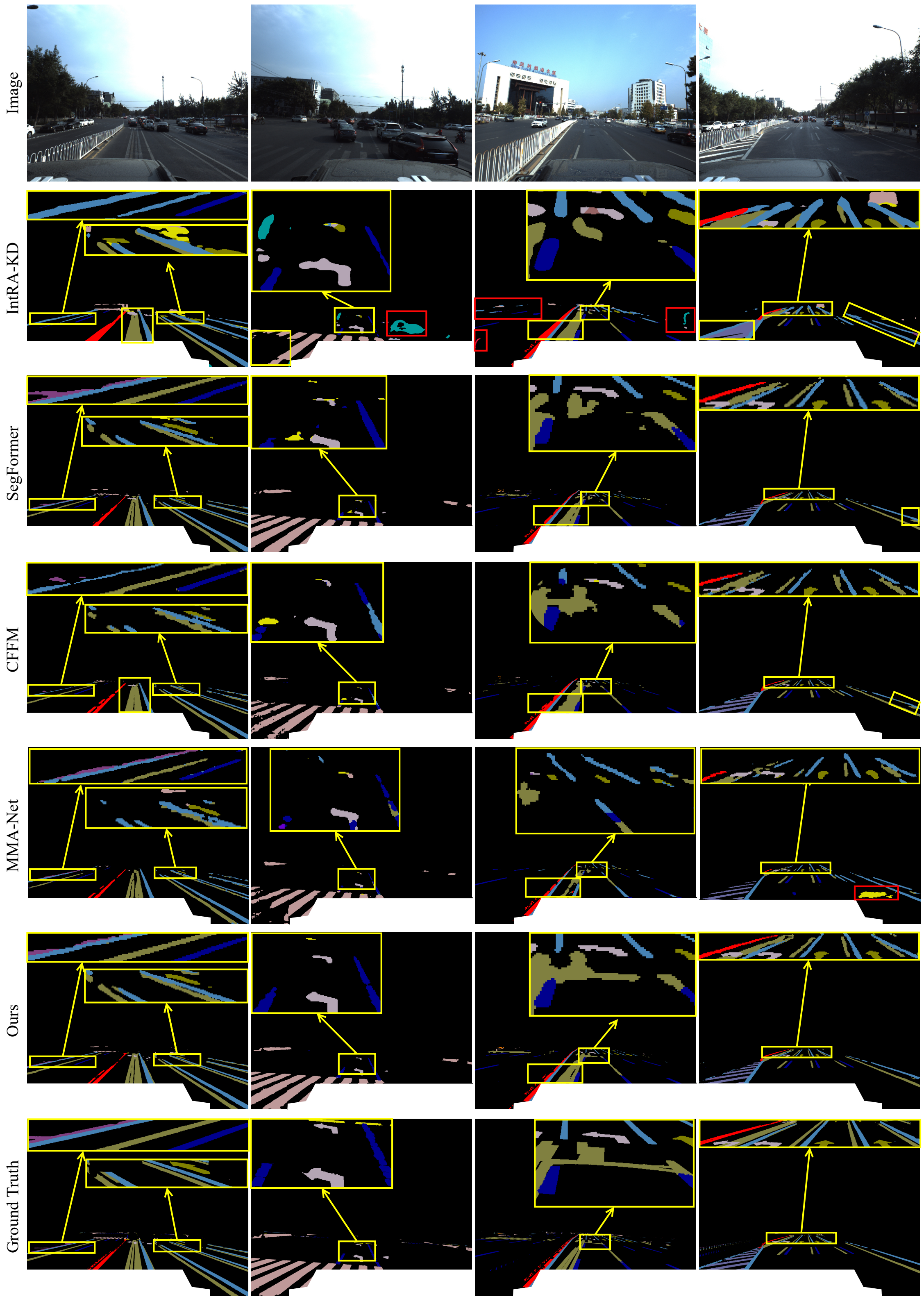


Figure 9: Extra qualitative comparison of our proposed method with SOTA methods on the ApolloScape [5] dataset. Yellow boxes highlight the differences and enlarge the target area for better visibility. Red boxes indicate false-positive segmentation predictions.

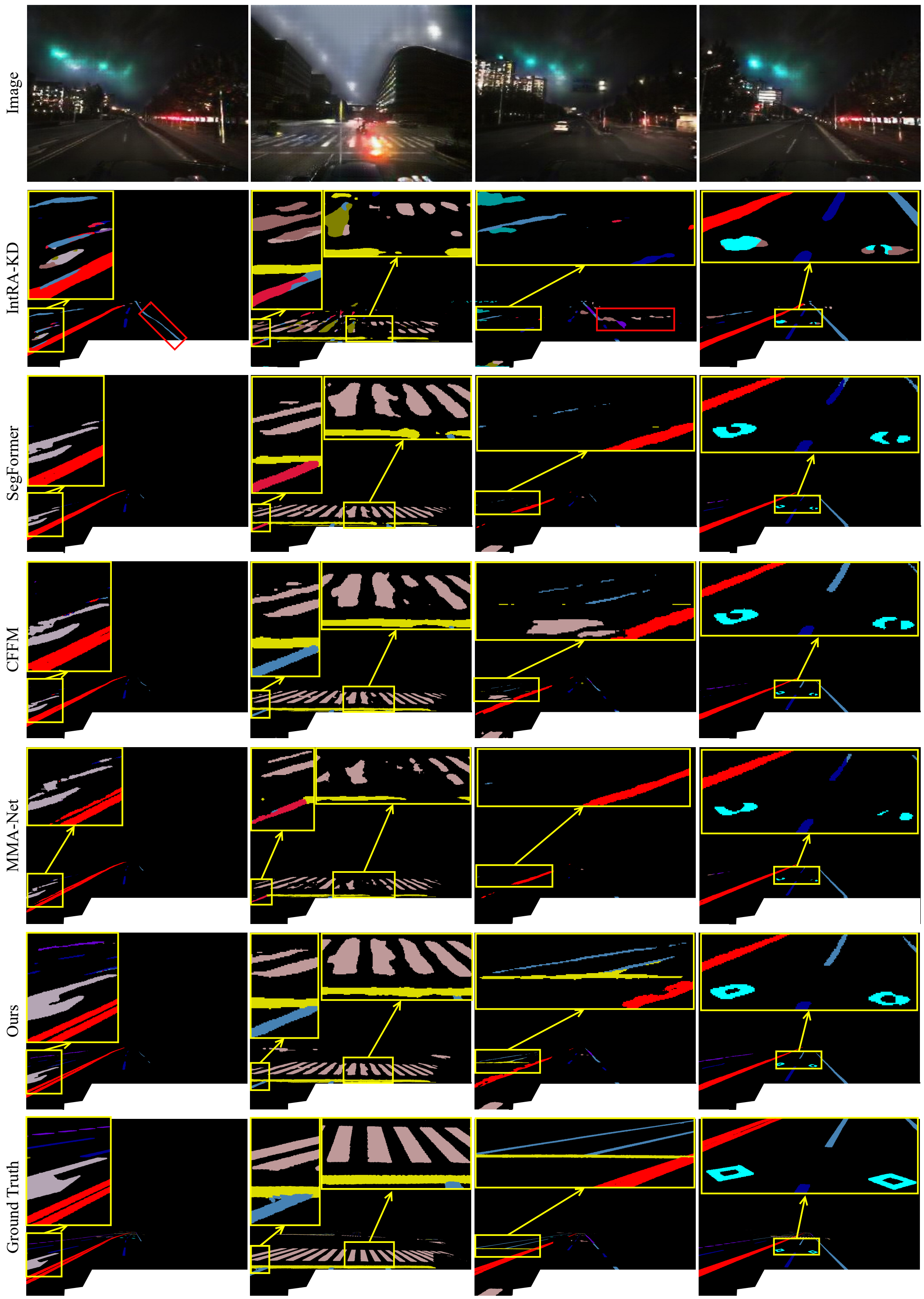


Figure 10: Extra qualitative comparison of our proposed method with SOTA methods on the ApolloScape Night dataset. Yellow boxes highlight the differences and enlarge the target area for better visibility. Red boxes indicate false-positive segmentation predictions.

Yu Qiao, et al. Persformer: 3d lane detection via perspective transformer and the openlane benchmark. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXVIII*, pages 550–567. Springer, 2022. 2

[4] Yuenan Hou, Zheng Ma, Chunxiao Liu, Tak-Wai Hui, and Chen Change Loy. Inter-region affinity distillation for road marking segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12486–12495, 2020. 4

[5] Xinyu Huang, Xinjing Cheng, Qichuan Geng, Binbin Cao, Dingfu Zhou, Peng Wang, Yuanqing Lin, and Ruigang Yang. The apolloscape dataset for autonomous driving. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 954–960, 2018. 2, 4, 5

[6] Oshada Jayasinghe, Sahan Hemachandra, Damith Anhettigama, Shenali Kariyawasam, Ranga Rodrigo, and Peshala Jayasekara. Ceymo: See more on roads-a novel benchmark dataset for road marking detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3104–3113, 2022. 2

[7] Dongkwon Jin, Wonhui Park, Seong-Gyun Jeong, Heeyeon Kwon, and Chang-Su Kim. Eigenlanes: Data-driven lane descriptors for structurally diverse lanes. In *CVPR*, 2022. 2

[8] Seokju Lee, Junsik Kim, Jae Shin Yoon, Seunghak Shin, Oleksandr Bailo, Namil Kim, Tae-Hee Lee, Hyun Seok Hong, Seung-Hoon Han, and In So Kweon. Vpgnet: Vanishing point guided network for lane and road marking detection and recognition. pages 1947–1955, 2017. 2

[9] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 1

[10] Suvarna Shirke and R Udayakumar. Lane datasets for lane detection. In *2019 International Conference on Communication and Signal Processing (ICCSP)*, pages 0792–0796. IEEE, 2019. 2

[11] Guolei Sun, Yun Liu, Henghui Ding, Thomas Probst, and Luc Van Gool. Coarse-to-fine feature mining for video semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision and Patern Recognition (CVPR)*, 2022. 2, 4

[12] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2446–2454, 2020. 2

[13] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34:12077–12090, 2021. 4

[14] Yujun Zhang, Lei Zhu, Wei Feng, Huazhu Fu, Mingqian Wang, Qingxia Li, Cheng Li, and Song Wang. Vil-100: A new dataset and a baseline model for video instance lane detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15681–15690, 2021. 2, 4