Supplementary material

VIL-100: A New Dataset and A Baseline Model for Video Instance Lane Detection

Anonymous ICCV submission

Paper ID 2714

There are three parts in this supplementary material. The first part (part 1) presents additional visualization of dataset. The second part (part 2) presents additional comparison results against the following works:

Video object segmentation

• **GAM** Joakim Johnander, Martin Danelljan, Emil Brissman, Fahad Shahbaz Khan, and Michael Felsberg. A generative appearance model for end-to-end video object segmentation. In CVPR, pages 8953–8962, 2019.

• **RVOS** Carles Ventura, Miriam Bellver, Andreu Girbau, Amaia Salvador, Ferran Marques, and Xavier Giro i Nieto. RVOS: Endto- end recurrent network for video object segmentation. In CVPR, pages 5277–5286, 2019.

• **STM** Seoung Wug Oh, Joon-Young Lee, Ning Xu, and Seon Joo Kim. Video object segmentation using space-time memory networks. In ICCV, pages 9226–9235, 2019.

• **AFB-URR** Yongqing Liang, Xin Li, Navid Jafari, and Qin Chen. Video object segmentation with adaptive feature bank and uncertain-region refinement. In NIPS, 2020.

• **TVOS** Yizhuo Zhang, ZhirongWu, Houwen Peng, and Stephen Lin. A transductive approach for video object segmentation. In CVPR, pages 6949–6958, 2020.

Lane detection

• **LaneNet** Davy Neven, Bert De Brabandere, Stamatios Georgoulis, Marc Proesmans, and Luc Van Gool. Towards end-to-end lane detection: an instance segmentation approach. In IVS, pages 286–291, 2018.

• **SCNN** Xingang Pan, Jianping Shi, Ping Luo, Xiaogang Wang, , and Xiaoou Tang. Spatial as deep: Spatial cnn for traffic scene understanding. In AAAI, pages 7276–7283, 2018.

• **ENet-SAD** Yuenan Hou, Zheng Ma, Chunxiao Liu, and Chen Change Loy. Learning lightweight lane detection cnns by self attention distillation. In ICCV, pages 1013–1021, 2019. Nature 381(12), 1-213.

• **UFSA** Zequn Qin, Huanyu Wang, and Xi Li. Ultra fast structureaware deep lane detection. In ECCV. Springer, 2020.

• **LSTR** Ruijin Liu, Zejian Yuan, Tie Liu, and Zhiliang Xiong. Endto- end lane shape prediction with transformers. In WACV, pages 3694–3702, 2021.

The third part (part 3) presents the visualization of ablation study as follows:

• **MMA-Net:** final proposed methods.

• **Basic:** basic model.

• **Basic + LM:** add the attention mechanism to the local memory of basic model.

• **Basic + GM:** add the attention mechanism to the global memory of basic model.

• **Basic + LGM:** fuse the local memory and the global memory together into basic model.

Part 1. Visualization of dataset

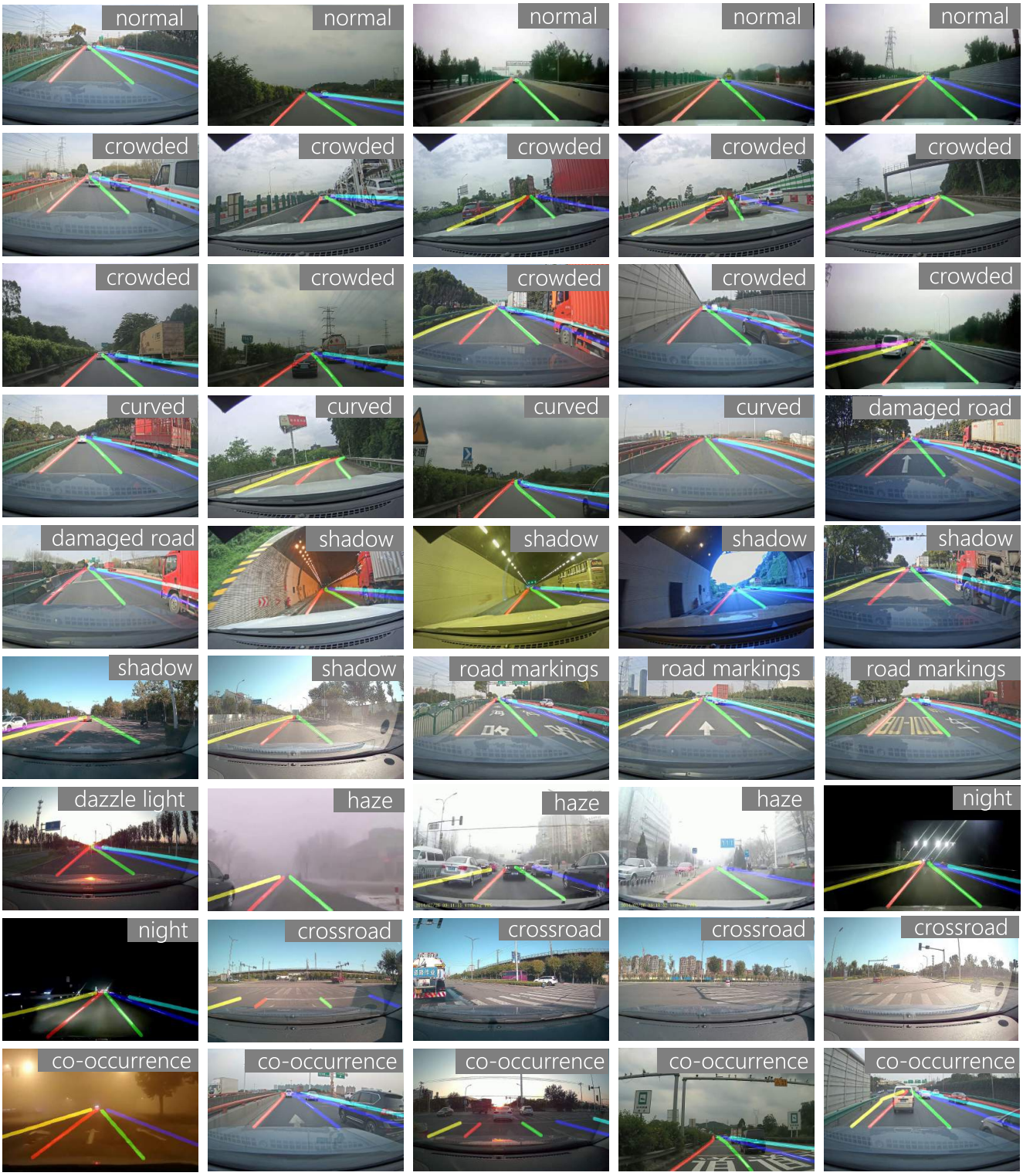
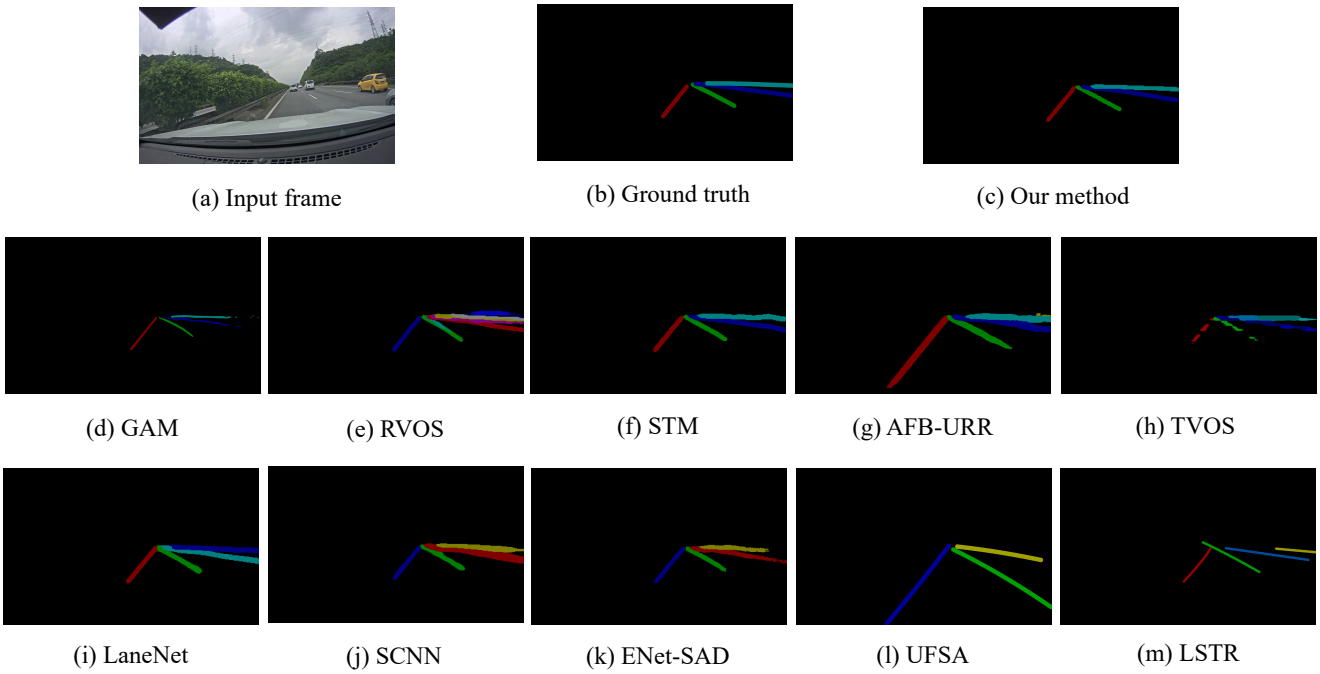


Figure 1. Visual results of VIL-100. Please refer to 'VIL-100.mp4' for more details.

Part 2. Visual Comparison of State-of-the-art Methods

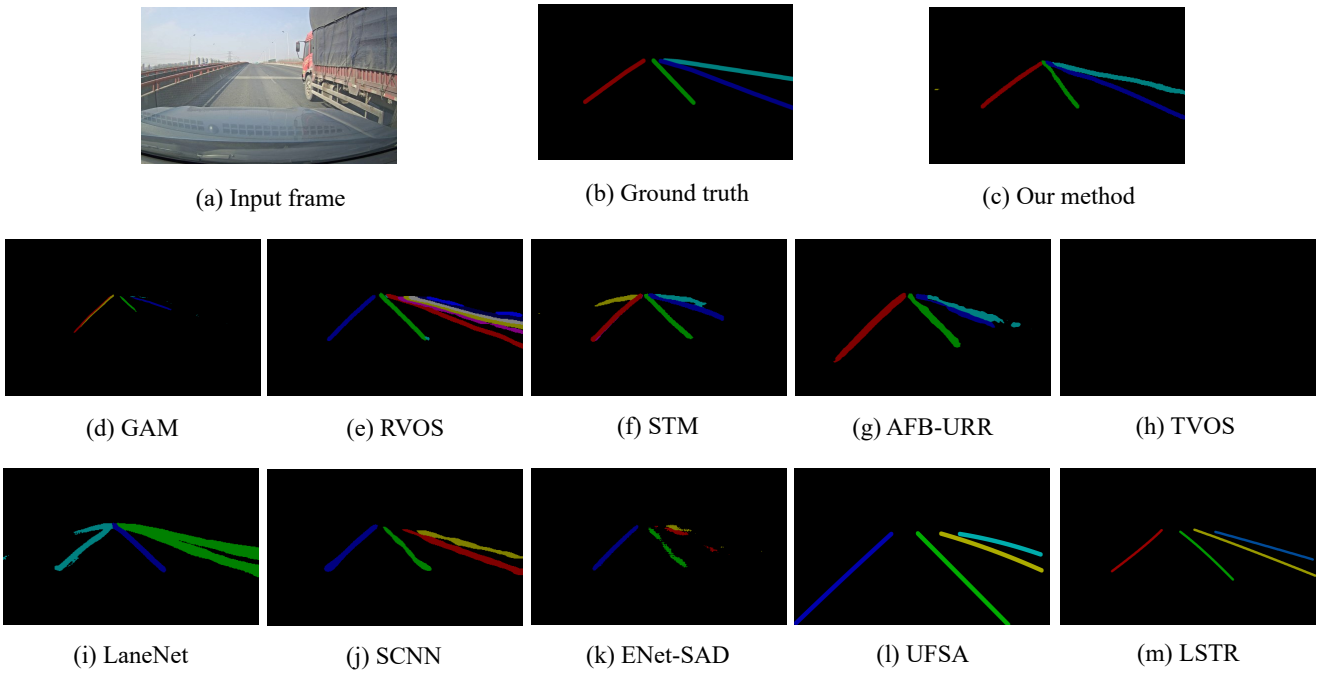


(a) Input frame (b) Ground truth (c) Our method

(d) GAM (e) RVOS (f) STM (g) AFB-URR (h) TVOS

(i) LaneNet (j) SCNN (k) ENet-SAD (l) UFSA (m) LSTR

Figure 2. More visual comparison results #1 (each color indicates a lane instance and black indicates non-lane pixels).

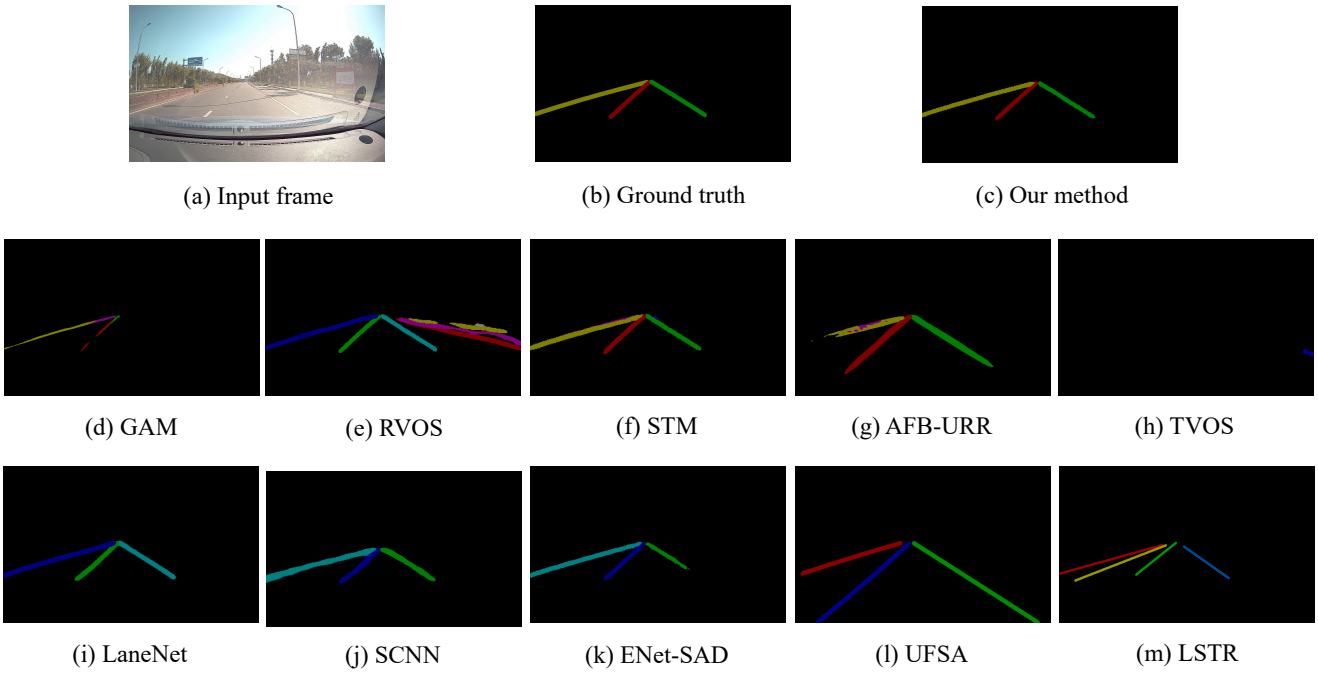


(a) Input frame (b) Ground truth (c) Our method

(d) GAM (e) RVOS (f) STM (g) AFB-URR (h) TVOS

(i) LaneNet (j) SCNN (k) ENet-SAD (l) UFSA (m) LSTR

Figure 3. More visual comparison results #2 (each color indicates a lane instance and black indicates non-lane pixels).

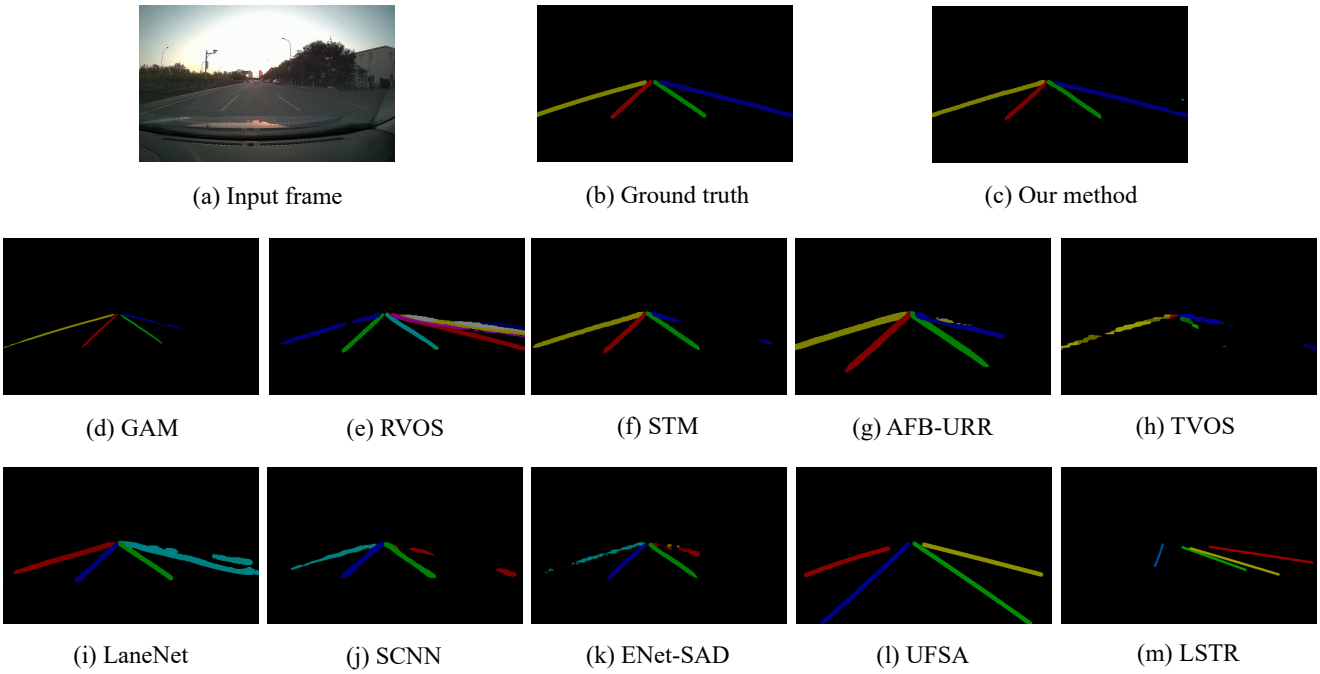


(a) Input frame (b) Ground truth (c) Our method

(d) GAM (e) RVOS (f) STM (g) AFB-URR (h) TVOS

(i) LaneNet (j) SCNN (k) ENet-SAD (l) UFSA (m) LSTR

Figure 4. More visual comparison results #3 (each color indicates a lane instance and black indicates non-lane pixels).

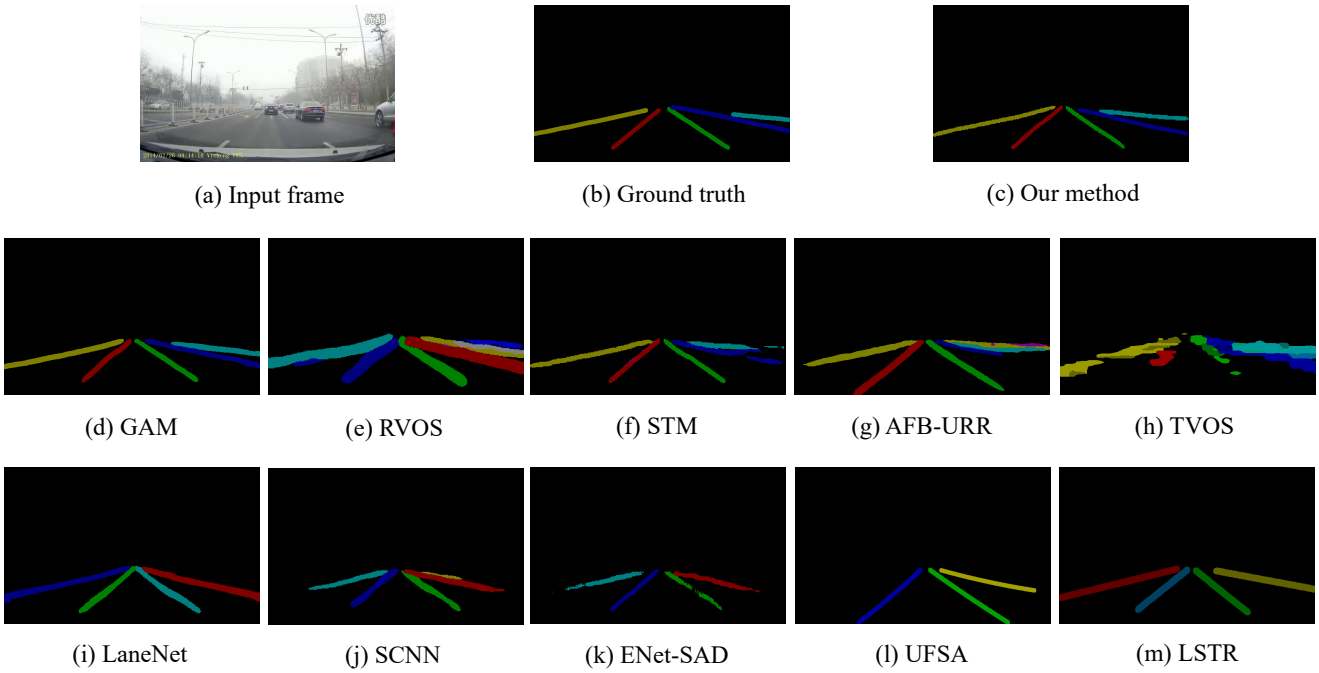


(a) Input frame (b) Ground truth (c) Our method

(d) GAM (e) RVOS (f) STM (g) AFB-URR (h) TVOS

(i) LaneNet (j) SCNN (k) ENet-SAD (l) UFSA (m) LSTR

Figure 5. More visual comparison results #4 (each color indicates a lane instance and black indicates non-lane pixels).

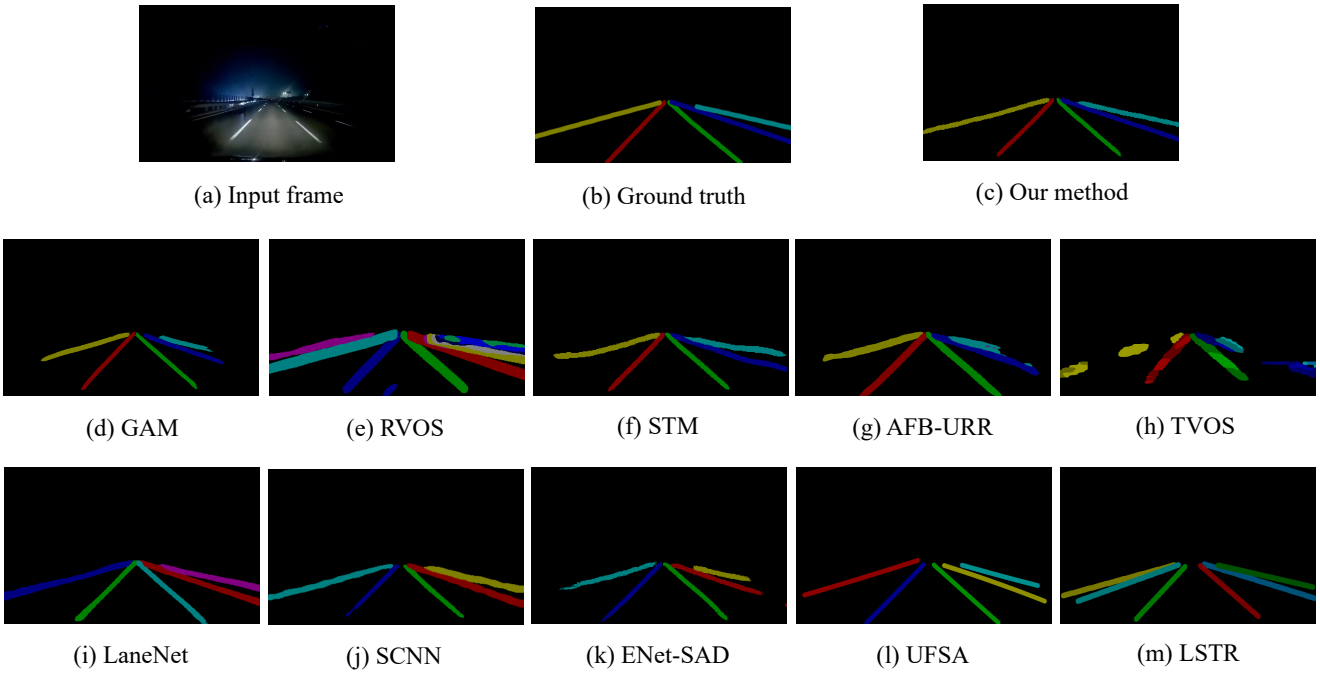


(a) Input frame (b) Ground truth (c) Our method

(d) GAM (e) RVOS (f) STM (g) AFB-URR (h) TVOS

(i) LaneNet (j) SCNN (k) ENet-SAD (l) UFSA (m) LSTR

Figure 6. More visual comparison results #5 (each color indicates a lane instance and black indicates non-lane pixels).

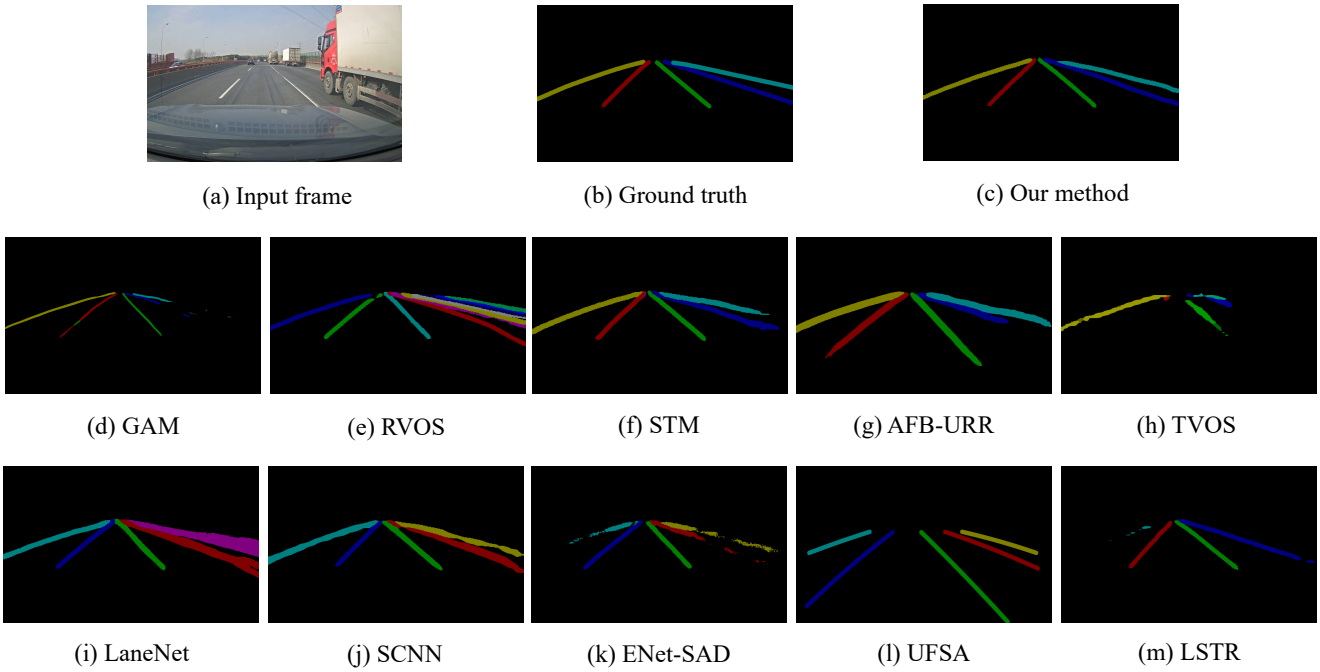


(a) Input frame (b) Ground truth (c) Our method

(d) GAM (e) RVOS (f) STM (g) AFB-URR (h) TVOS

(i) LaneNet (j) SCNN (k) ENet-SAD (l) UFSA (m) LSTR

Figure 7. More visual comparison results #6 (each color indicates a lane instance and black indicates non-lane pixels).

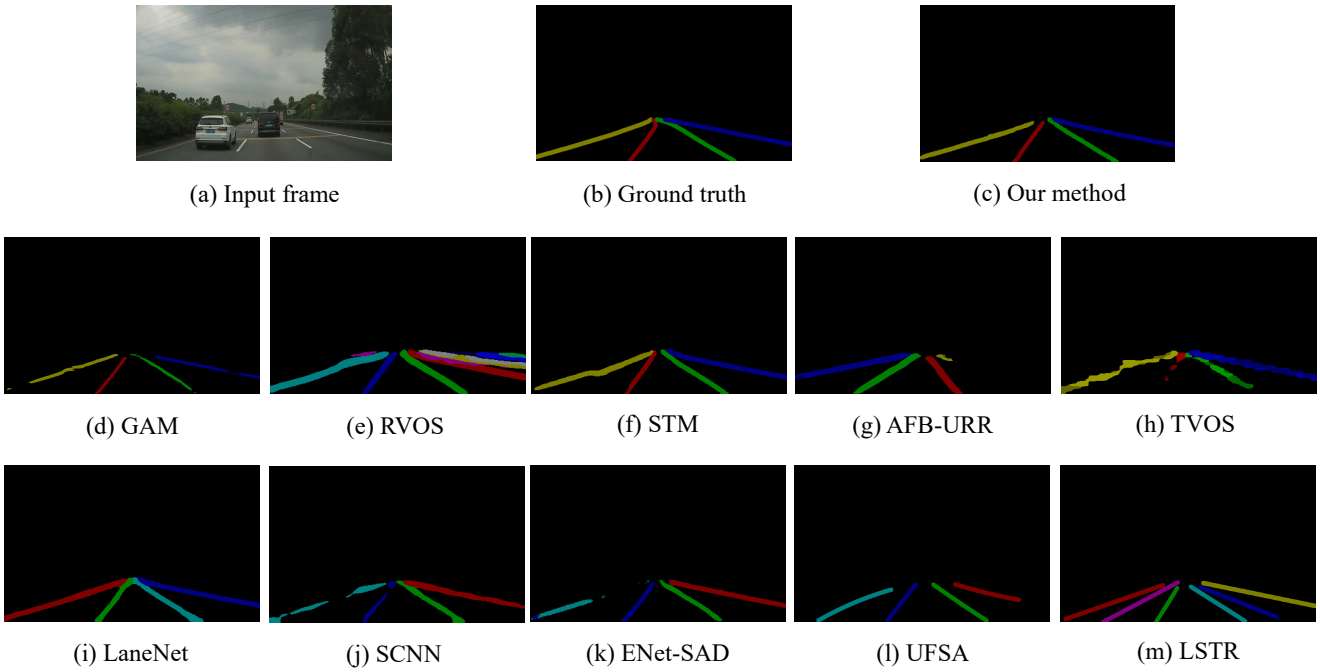


(a) Input frame (b) Ground truth (c) Our method

(d) GAM (e) RVOS (f) STM (g) AFB-URR (h) TVOS

(i) LaneNet (j) SCNN (k) ENet-SAD (l) UFSA (m) LSTR

Figure 8. More visual comparison results #7 (each color indicates a lane instance and black indicates non-lane pixels).

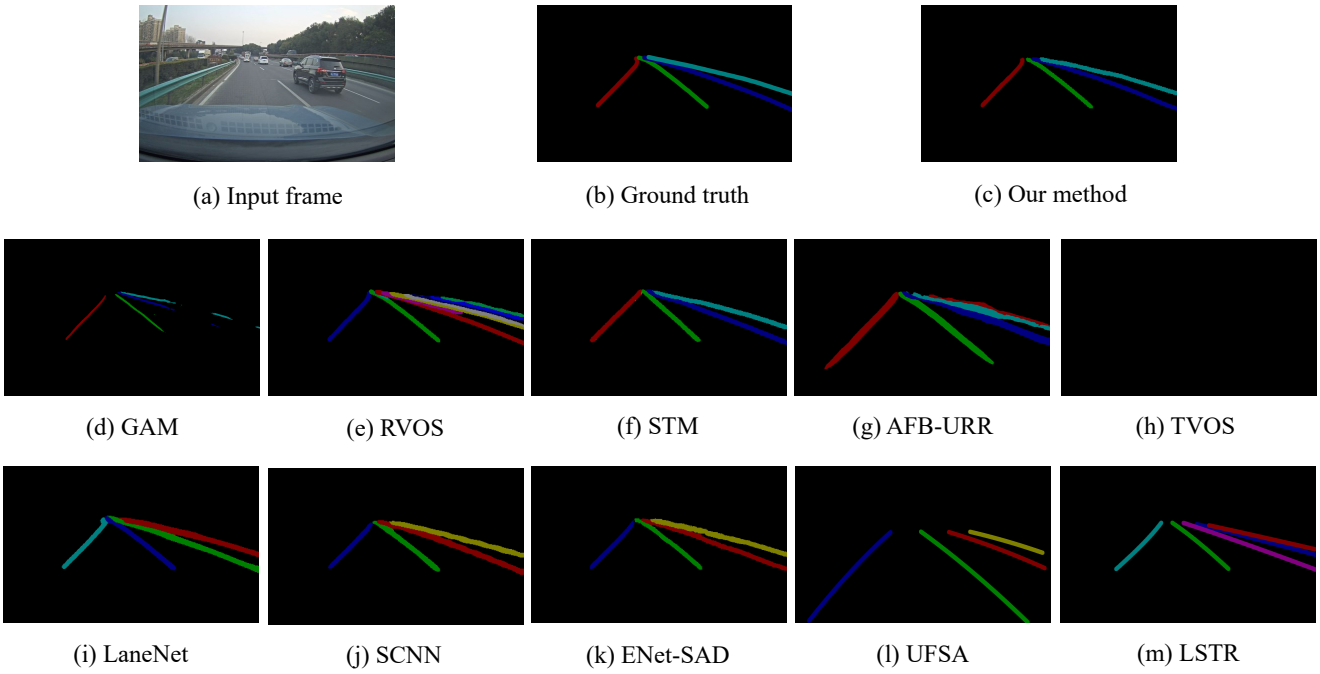


(a) Input frame (b) Ground truth (c) Our method

(d) GAM (e) RVOS (f) STM (g) AFB-URR (h) TVOS

(i) LaneNet (j) SCNN (k) ENet-SAD (l) UFSA (m) LSTR

Figure 9. More visual comparison results #8 (each color indicates a lane instance and black indicates non-lane pixels).

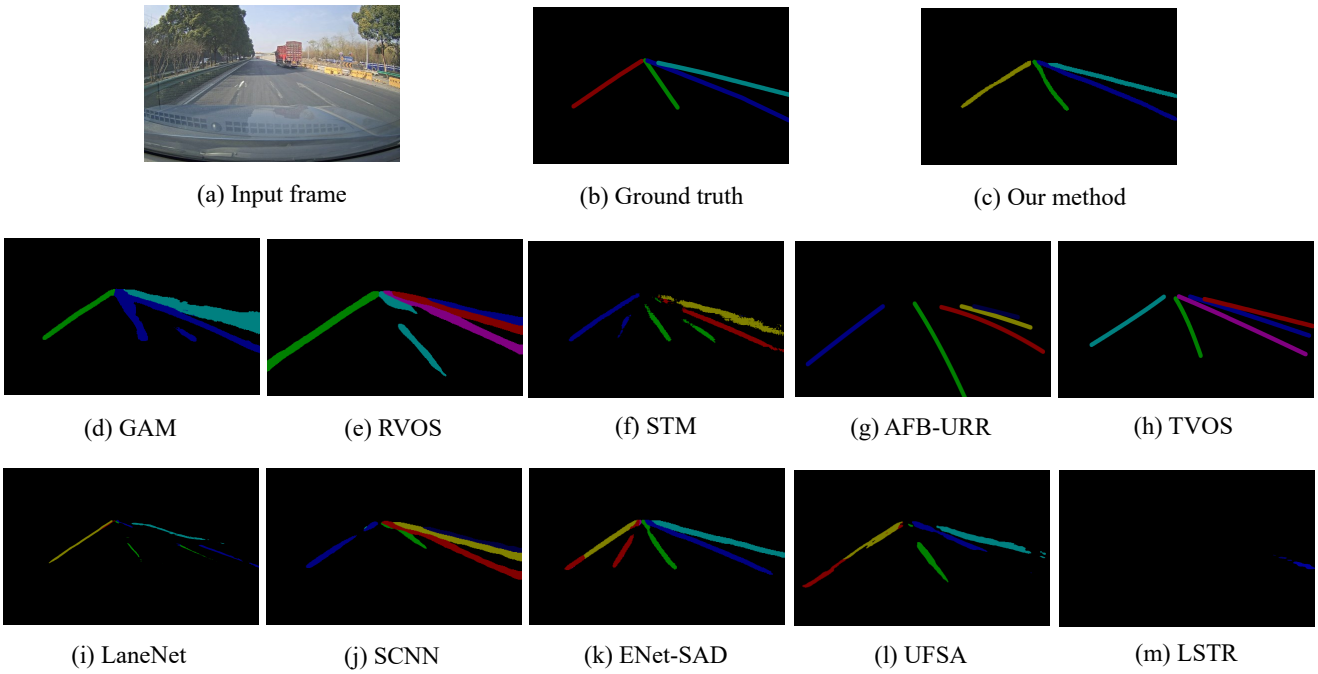


(a) Input frame (b) Ground truth (c) Our method

(d) GAM (e) RVOS (f) STM (g) AFB-URR (h) TVOS

(i) LaneNet (j) SCNN (k) ENet-SAD (l) UFSA (m) LSTR

Figure 10. More visual comparison results #9 (each color indicates a lane instance and black indicates non-lane pixels).



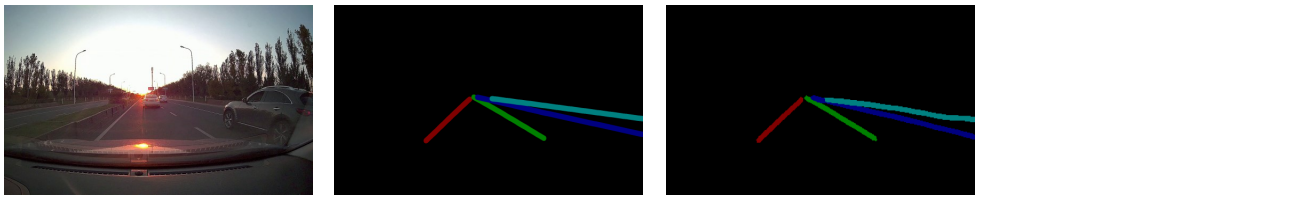
(a) Input frame (b) Ground truth (c) Our method

(d) GAM (e) RVOS (f) STM (g) AFB-URR (h) TVOS

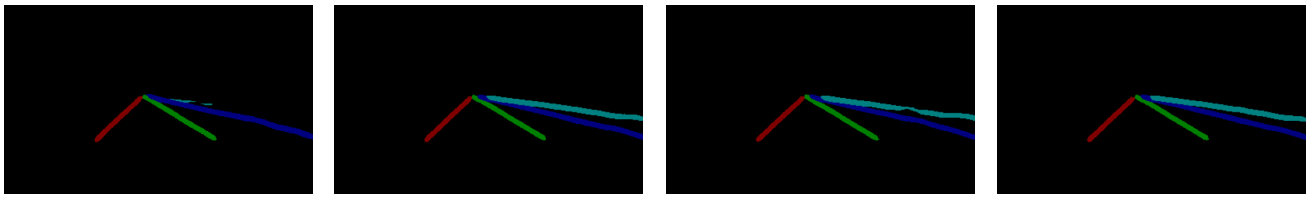
(i) LaneNet (j) SCNN (k) ENet-SAD (l) UFSA (m) LSTR

Figure 11. More visual comparison results #10 (each color indicates a lane instance and black indicates non-lane pixels).

Part3. Visual Comparison of Ablation Study

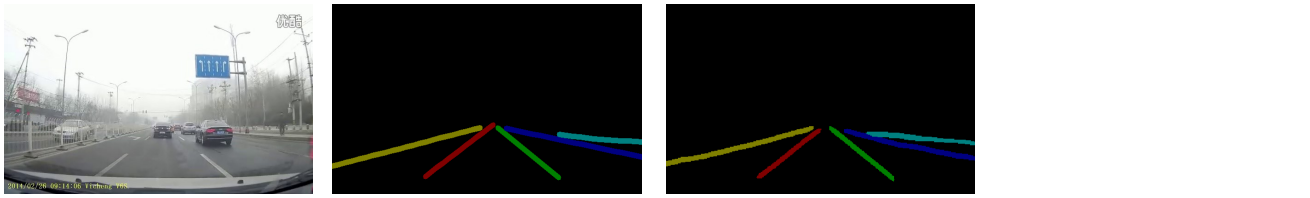


(a) Input frame (b) Ground truth (c) MMA-Net (ours)

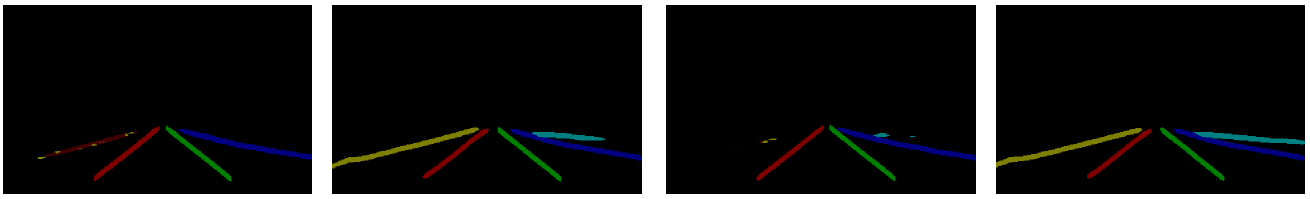


(d) Basic (e) Basic + LM (f) Basic + GM (g) Basic + LGM

Figure 12. Visualizations of ablation study #1 (each color indicates a lane instance and black indicates non-lane pixels).

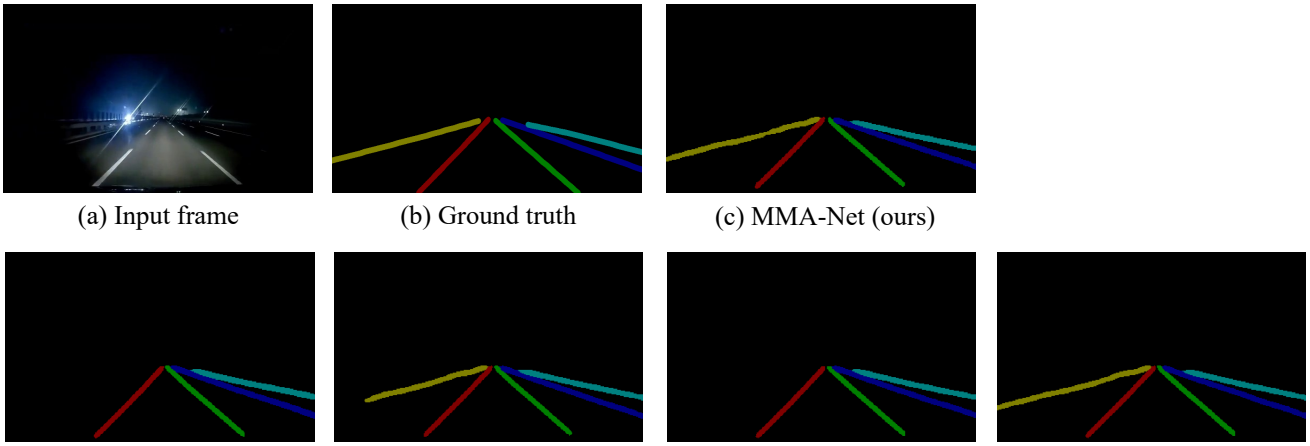


(a) Input frame (b) Ground truth (c) MMA-Net (ours)



(d) Basic (e) Basic + LM (f) Basic + GM (g) Basic + LGM

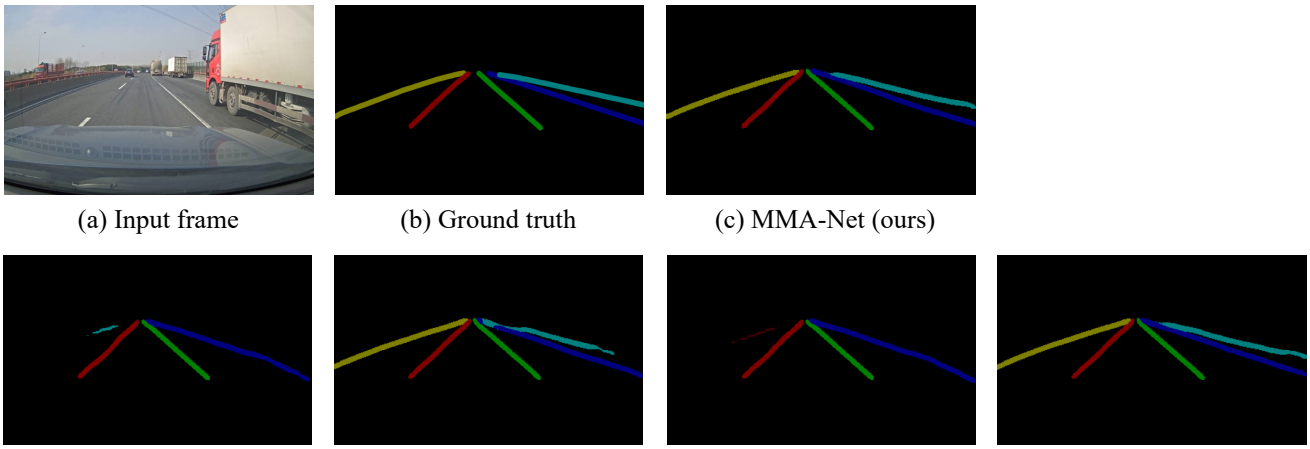
Figure 13. Visualizations of ablation study #2 (each color indicates a lane instance and black indicates non-lane pixels).



(a) Input frame (b) Ground truth (c) MMA-Net (ours)

(d) Basic (e) Basic + LM (f) Basic + GM (g) Basic + LGM

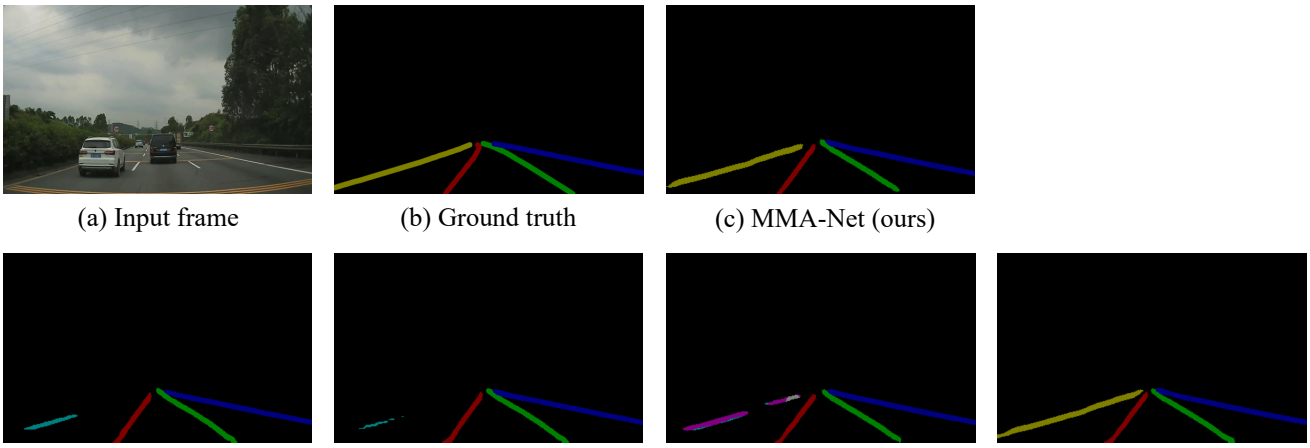
Figure 14. Visualizations of ablation study #3 (each color indicates a lane instance and black indicates non-lane pixels).



(a) Input frame (b) Ground truth (c) MMA-Net (ours)

(d) Basic (e) Basic + LM (f) Basic + GM (g) Basic + LGM

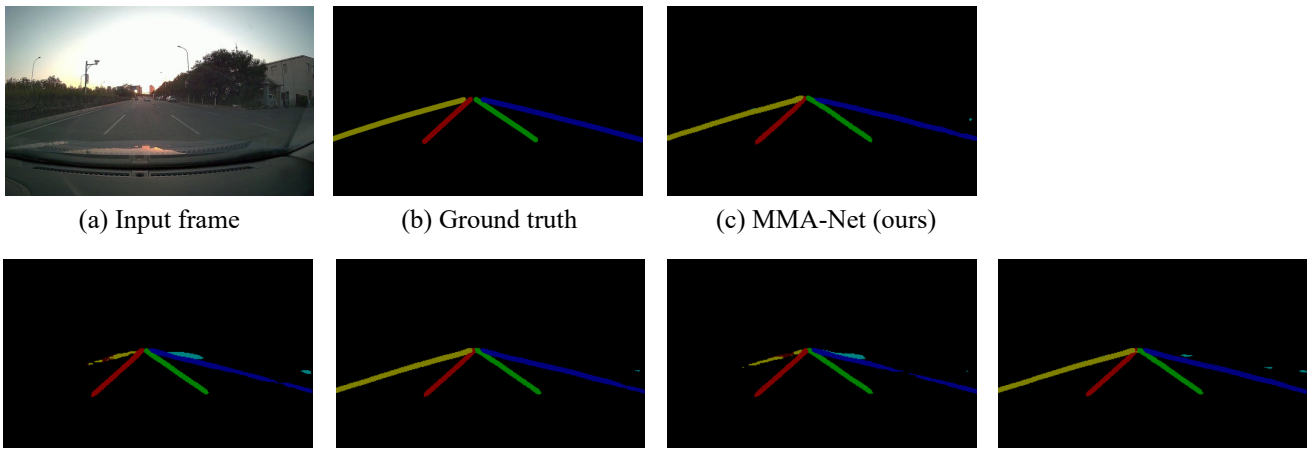
Figure 15. Visualizations of ablation study #4 (each color indicates a lane instance and black indicates non-lane pixels).



(a) Input frame (b) Ground truth (c) MMA-Net (ours)

(d) Basic (e) Basic + LM (f) Basic + GM (g) Basic + LGM

Figure 16. Visualizations of ablation study #5 (each color indicates a lane instance and black indicates non-lane pixels).



(a) Input frame (b) Ground truth (c) MMA-Net (ours)

(d) Basic (e) Basic + LM (f) Basic + GM (g) Basic + LGM

Figure 17. Visualizations of ablation study #6 (each color indicates a lane instance and black indicates non-lane pixels).