VALSE 2021

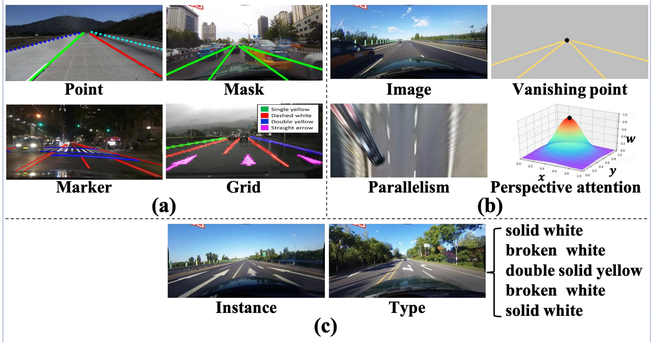
August 18 – 20, 2021

Structure Guided Lane Detection

Jinming Su, Chao Chen, Ke Zhang, Junfeng Luo, Xiaoming Wei and Xiaolin Wei

IJCAI 2021

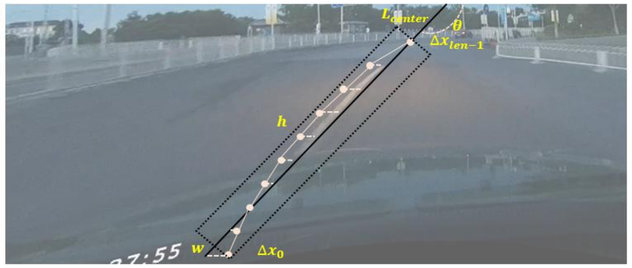
Motivation



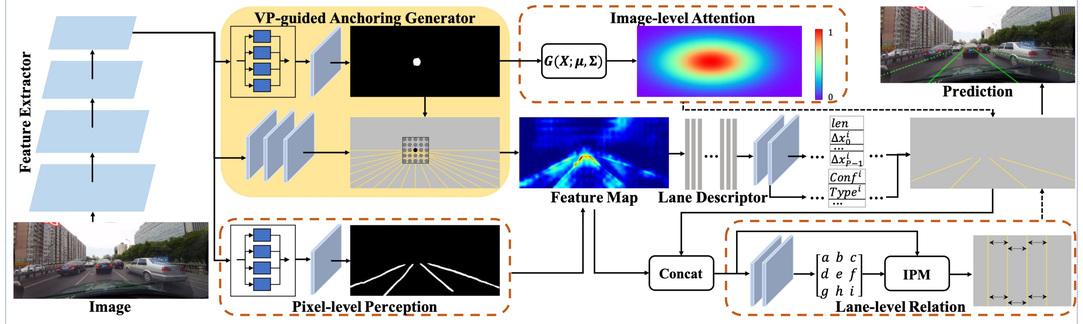
- (a) There lacks a unified and effective lane representation. There are many definition including point, mask and so on, which are quite different in form for demands of scenarios.
- (b) It is difficult to model the structural relationship between scenes and lanes, which is very useful, but there is no scheme to describe it.
- (c) While predicting lanes, it is also important to predict other attributes including instance, type and so on, but it is not easy to extend these for existing methods.

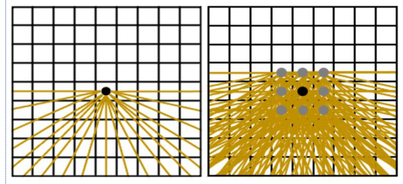
Lane Representation

- We propose a box-line based method for lane instance based on L_{center} and ΔX to represent L_{lane}.



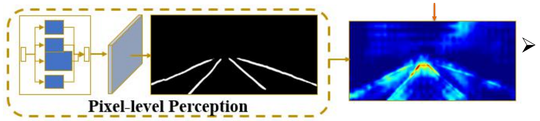
Network Architecture





Vanishing Point Guided Anchoring

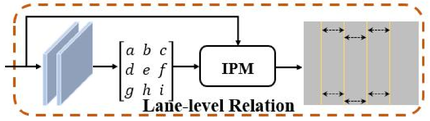
- Dense lines radiated from vanishing points (VPs) can theoretically cover all lanes in the image.
- We generate anchors (golden lines) from VP (black points) with an interval of fixed degree and set candidates (gray points) of VP for dense prediction.



Multi-level Structure Constraints

- **Pixel-level Perception**

  For the sake of improving the detail perception, we introduce lane segmentation branch to locate lanes and promote pixel-level unary details.



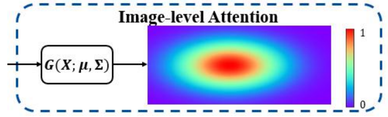
- **Lane-level Relation**

  L'_{Lane1} on bird's eye view: $a_1 * x + b_1 * y + c_1 = 0$;

  L'_{Lane2} on bird's eye view: $a_2 * x + b_2 * y + c_2 = 0$;

  $\because L'_{Lane1}$ and L'_{Lane2} are parallel $\therefore a_1 * b_2 = a_2 * b_1$

  $L_{\mathcal{L}} = \sum_{i=0,j=0,i\neq j}^{L-1} L1(a_i b_j - a_j b_i)$



- **Image-level Relation**

  Distant objects are small after projection, so we control the attention of different regions according to distance between lanes and VP.

  $L_g = \sum_{L_{lane}=0}^{L-1} \sum_{p=0}^{P-1} L1(\Delta X_p^{L_{lane}}, G\Delta x_p^{Lane}) \cdot \left(1 + \left|E\left(\Delta x_p^{Lane}, G\Delta x_p^{Lane}\right)\right|\right)$

Results

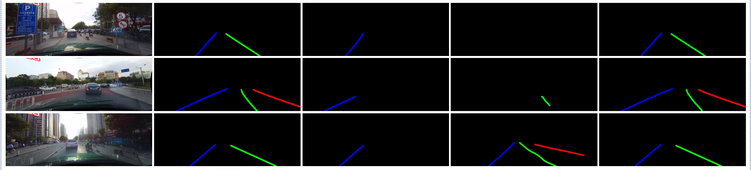
- Our method outperforms 12 SOTA algorithms on CULane dataset.

	Total	Normal	Crowd	Dazzle	Shadow	Noline	Arrow	Curve	Cross	Night	FPS
DeepLabV2	66.70	87.40	64.10	54.10	60.70	38.10	79.00	59.80	2505	60.60	-
SCNN	71.60	90.60	69.70	58.50	66.90	43.40	84.10	64.40	1990	66.10	8
FD	-	85.90	63.60	57.00	59.90	40.60	79.40	65.20	7013	57.80	-
Enet-SAD	70.80	90.10	68.80	60.20	65.90	41.60	84.00	65.70	1998	66.00	75
PointLane	70.20	88.00	68.10	61.50	63.30	44.00	80.90	65.20	1640	63.20	-
RONELD	72.90	-	-	-	-	-	-	-	-	-	-
PINet	74.40	90.30	72.30	66.30	68.40	49.80	83.70	65.60	1427	67.70	25
ERFNet-E2E	74.00	91.00	73.10	64.50	74.10	46.60	85.80	71.90	2022	67.90	-
IntRA-KD	72.40	-	-	-	-	-	-	-	-	-	98
UltraFast-18	68.40	87.70	66.00	58.40	62.80	40.20	81.00	57.90	1743	62.10	323
UltraFast-34	72.30	90.70	72.20	59.50	69.30	44.40	85.70	69.50	2037	66.70	175
CurveLanes	74.80	90.70	72.30	67.70	70.10	49.40	85.80	68.40	1746	68.90	-
Ours-Res18	76.12	91.42	74.05	66.89	72.17	50.16	87.13	67.02	1164	70.67	117
Ours-Res34	77.27	92.07	75.41	67.75	74.31	50.90	87.97	69.65	1373	72.69	92

- Ablation study for VPG-Anchoring and multi-level structures.

	VPG-Anchoring	Pixel-level	Lane-level	Image-level	Total
Base					71.98
Base+V	√				74.27
Base+V+P	√	√			76.30
Base+V+P+L	√	√	√		76.70
SGNet (Ours)	√	√	√	√	77.27

- Qualitative comparisons of the state-of-the-art algorithms and our approach.



For more details, please contact Jinming Su (sujinming@meituan.com).