MIDAS: Modeling Ground-Truth Distributions with Dark Knowledge for Domain Generalized Stereo Matching

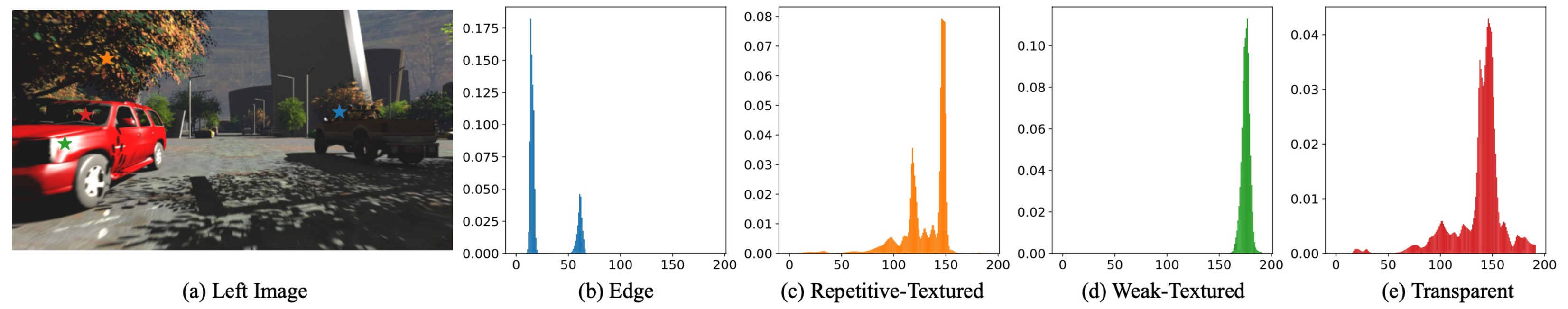
Peng Xu Zhiyu Xiang Jingyun Fu Tianyu Pu Hanzhi Zhong Eryun Liu

Zhejiang University

ICCV HONOLULU HAWAII OCT 19-23, 2025

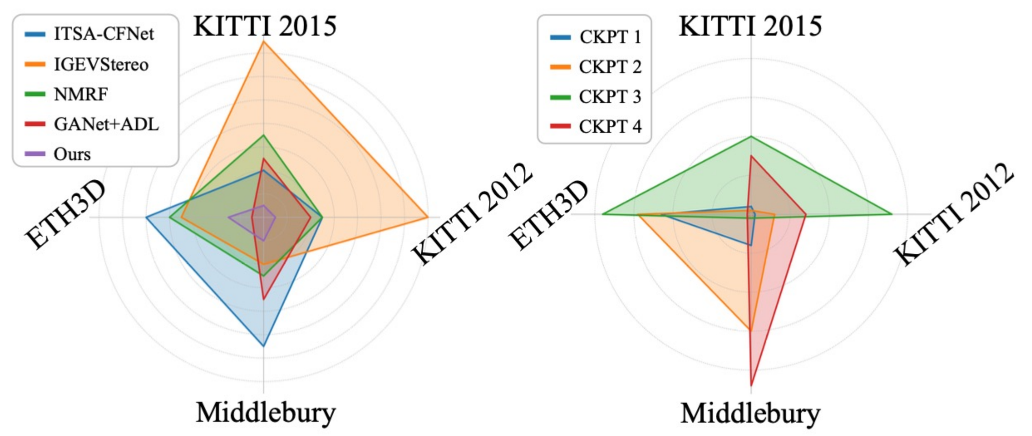
① Motivation

- Previous work modeled multi-modal ground truth for **edge pixels** with matching ambiguity.
- An elegant way to simultaneously model multi-modal distributions for other ambiguous regions, such as **repetitive textures and transparency**, is still missing.
- Stereo networks can **spontaneously** learn and output multi-modal distributions, implicitly capturing **similarity and uncertainty**.

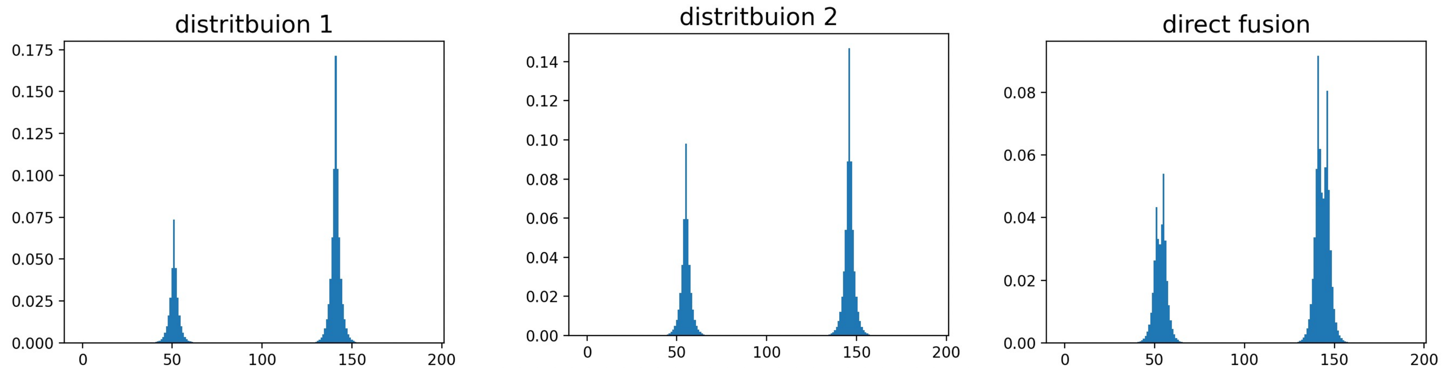


② Challenges

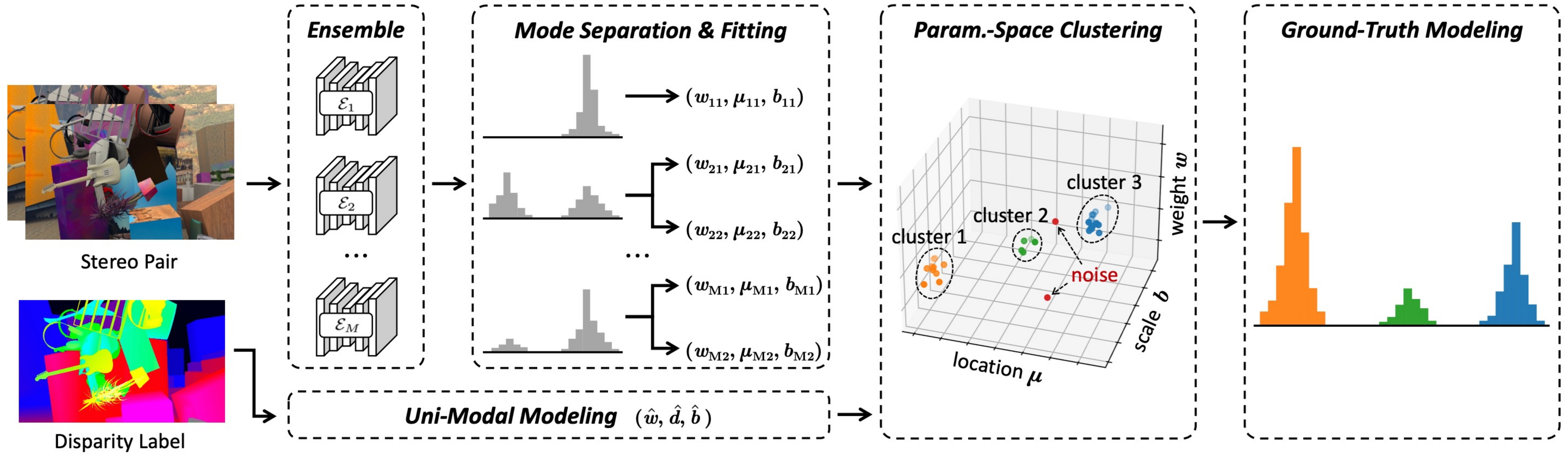
- Cross-domain preferences of different network architectures (left) and different checkpoints of the same network (right).



- Directly fusing the outputs of the network ensemble can disrupt the unimodal property of each mode.

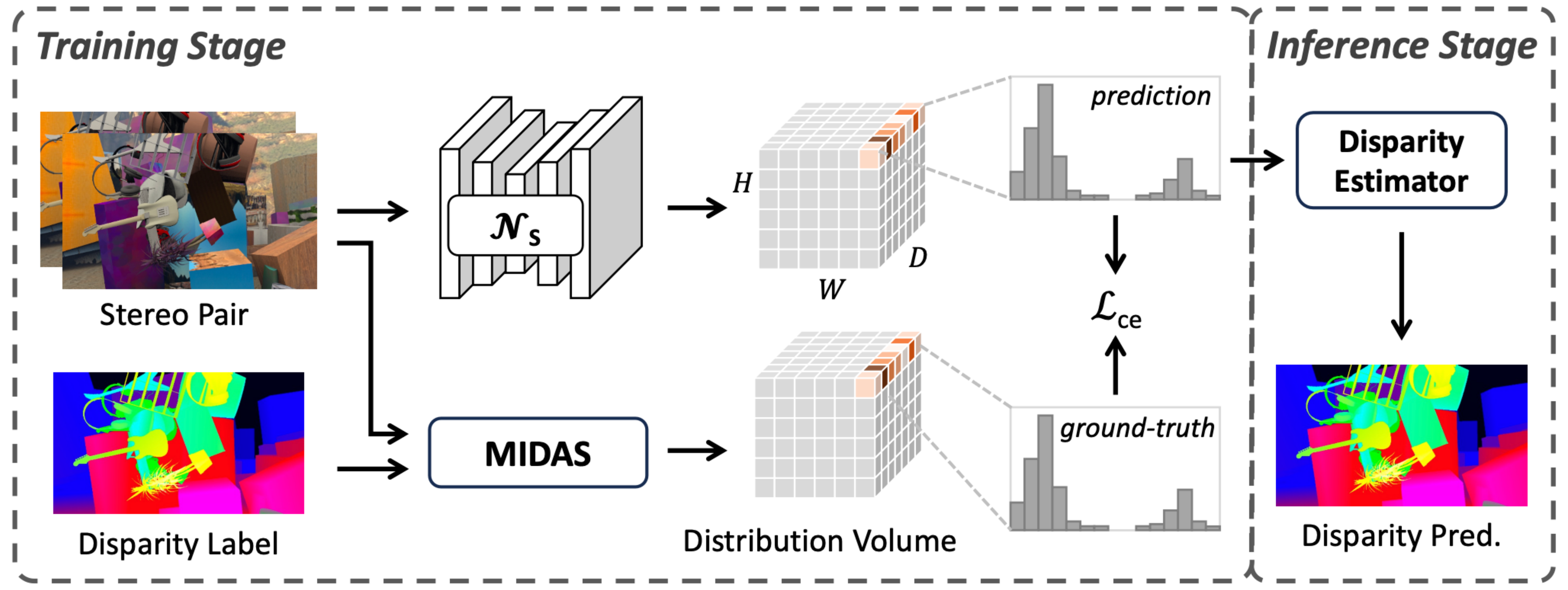


③ Ground-truth Distribution Modeling



- For each pixel, the network ensemble predicts M multi-modal probability distributions.
- Individual modes are separated from these distributions and fitted as **parameterized Laplacians** (w, μ, b).
- The disparity label is also modeled as the uni-modal Laplacian with coordinate $(\hat{w}, \hat{d}, \hat{b})$.
- We cluster the points in the parameter space to distinguish the **objective knowledge** (effective clusters) from the **biased knowledge** (noise).
- The elements within each cluster are **fused** and **re-modeled** as a formulated mode in the final ground-truth distribution.

④ Overall Pipeline



⑤ Ablations

#Arch.	#CKPT	KT15	KT12	MB	ETH3D
0	0	4.73	4.64	9.76	4.18
1	1	4.67	4.41	9.00	4.02
1	2	4.57	3.87	8.47	3.34
2	1	4.64	3.89	8.27	3.64
2	2	4.59	3.82	8.01	3.40
3	3	**4.49**	**3.72**	**7.95**	**3.17**

Method	KT15	KT12	MB	ETH3D
PSMNet [2] + Ours	**4.49**	**3.72**	**7.95**	**3.17**
w/o BKF	4.57	3.81	8.47	3.40

⑥ Quantitative Results

- Our method significantly enhances the backbone's performance and surpasses previous state-of-the-art methods.

Method	Publication	KITTI 2015 >3px	KITTI 2012 >3px	Middlebury >2px	ETH3D >1px	Mean Rank
PSMNet [2]	CVPR 2018	16.30^{18}	15.10^{18}	25.10^{18}	23.80^{18}	18.00
GwcNet [15]	CVPR 2018	12.80^{17}	11.70^{17}	18.10^{16}	9.00^{16}	16.50
GANet [47]	CVPR 2019	11.70^{16}	10.10^{16}	20.30^{17}	14.10^{17}	16.5
DSMNet [48]	ECCV 2020	6.50^{15}	6.20^{15}	13.80^{13}	6.20^{14}	14.25
CFNet [33]	CVPR 2021	5.80^{12}	4.70^{11}	15.30^{14}	5.80^{12}	12.25
Mask-CFNet [30]	CVPR 2023	5.80^{12}	4.80^{12}	13.70^{12}	5.70^{11}	11.75
Raft-Stereo [24]	3DV 2021	5.70^{11}	5.20^{14}	12.60^{11}	3.30^{6}	10.50
FC-GANet [50]	CVPR 2022	5.30^{9}	4.60^{10}	10.20^{9}	5.80^{12}	10.00
PCWNet [34]	ECCV 2022	5.60^{10}	4.20^{5}	15.77^{15}	5.20^{10}	10.00
IGEV-Stereo [40]	CVPR 2023	6.03^{14}	5.18^{13}	7.27^{3}	3.60^{7}	9.25
Graft-GANet [25]	CVPR 2022	4.90^{6}	4.20^{5}	9.80^{8}	6.20^{14}	8.25
ITSA-CFNet [9]	CVPR 2022	4.70^{4}	4.20^{5}	10.40^{10}	5.10^{9}	7.00
StereoRisk [26]	ICML 2024	5.19^{8}	4.43^{9}	9.32^{7}	2.41^{2}	6.50
NMRF [14]	CVPR 2024	5.10^{7}	4.20^{5}	7.50^{4}	3.80^{8}	6.00
GANet + ADL [41]	CVPR 2024	4.84^{5}	3.93^{4}	8.72^{6}	2.31^{1}	4.00
PSMNet + Ours	——	4.49^{3}	3.72^{2}	7.95^{5}	3.17^{5}	3.75
GwcNet + Ours	——	4.16^{2}	3.74^{3}	7.23^{2}	2.91^{4}	2.75
PCWNet + Ours	——	**3.96**^{1}	**3.57**^{1}	**7.20**^{1}	2.72^{3}	**1.50**

⑦ Qualitative Results

- Our method demonstrates excellent reliability on weak textures, repetitive textures, object edges, and strong glare.