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

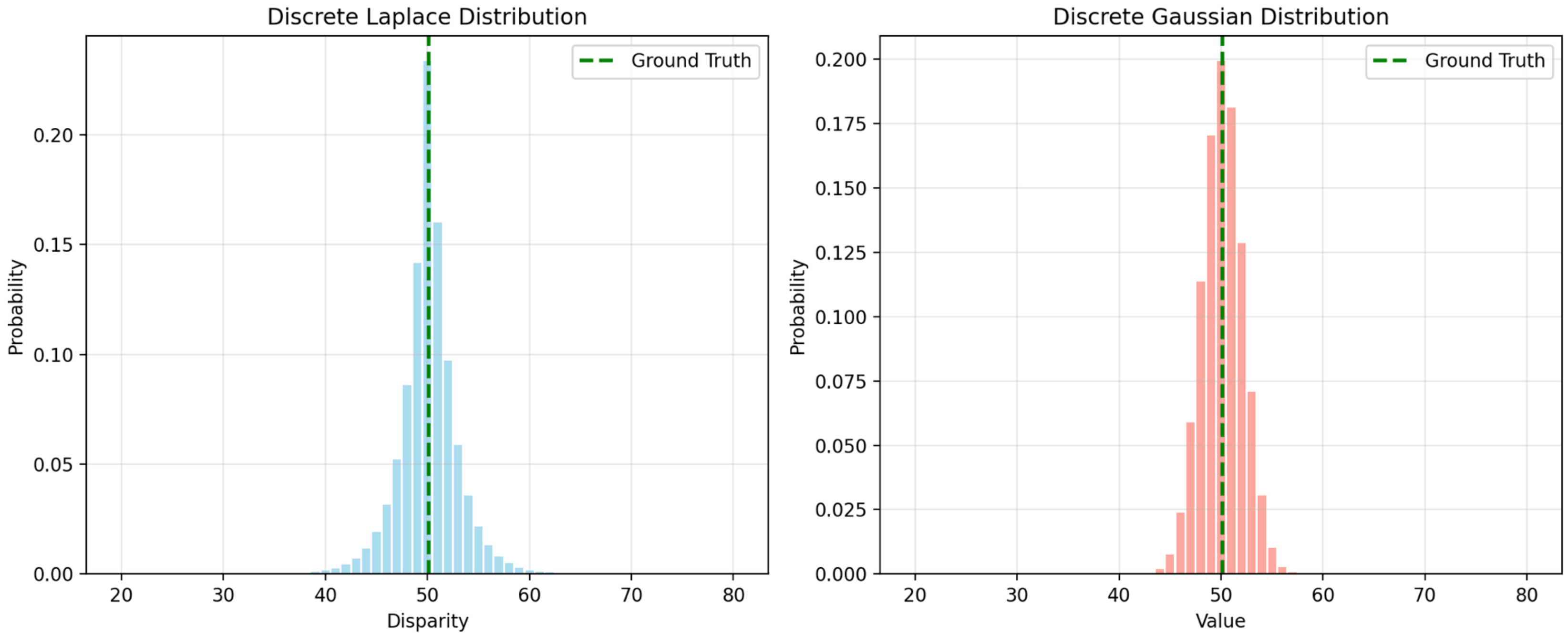
Peng Xu   Zhiyu Xiang   Jingyun Fu   Tianyu Pu   Hanzhi Zhong   Eryun Liu

Date: Oct.20, 2025

1. Motivation

- 3D convolution–based stereo networks output **discrete disparity probability distributions**, making them naturally suitable for supervision with cross-entropy loss.

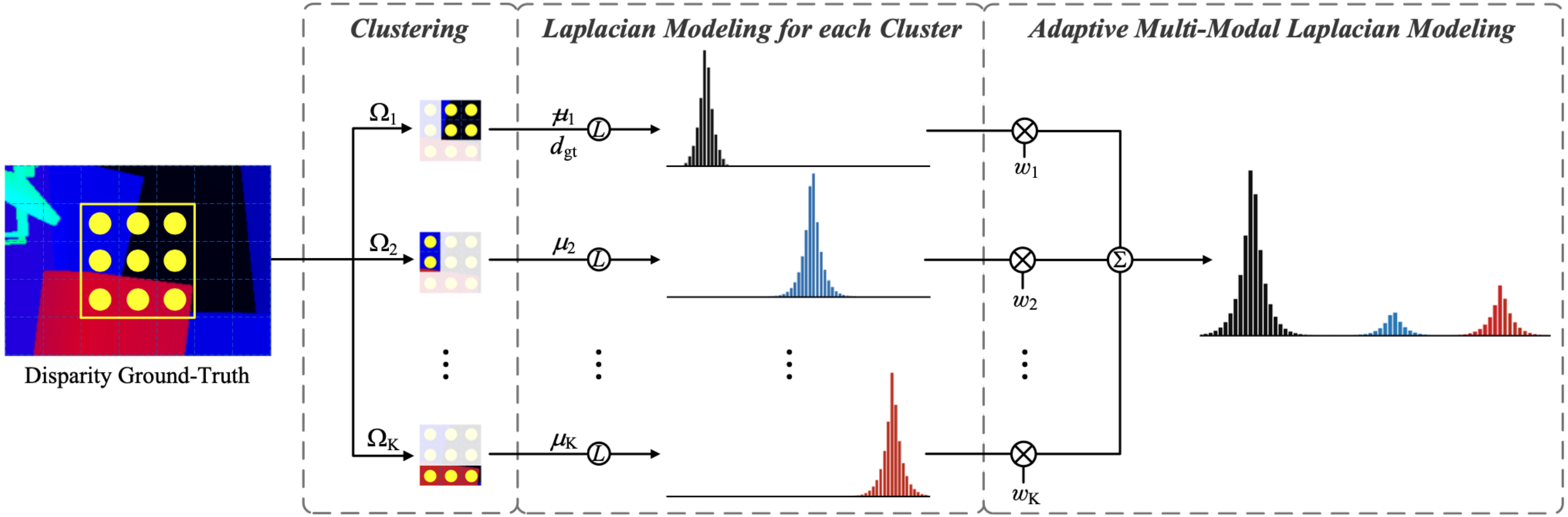
- Previous works modeled stereo ground truth as a **uni-modal Laplacian [1]** or **Gaussian [2]**, centered on the disparity ground truth.



[1] Tulyakov S, Ivanov A, Fleuret F. Practical deep stereo (pds): Toward applications-friendly deep stereo matching. NeurIPS 2018.
[2] Chen C, Chen X, Cheng H. On the over-smoothing problem of cnn based disparity estimation. ICCV 2019.

1. Motivation

- Later work (ADL) extended this to multi-modal modeling, but only for **edge** pixels.



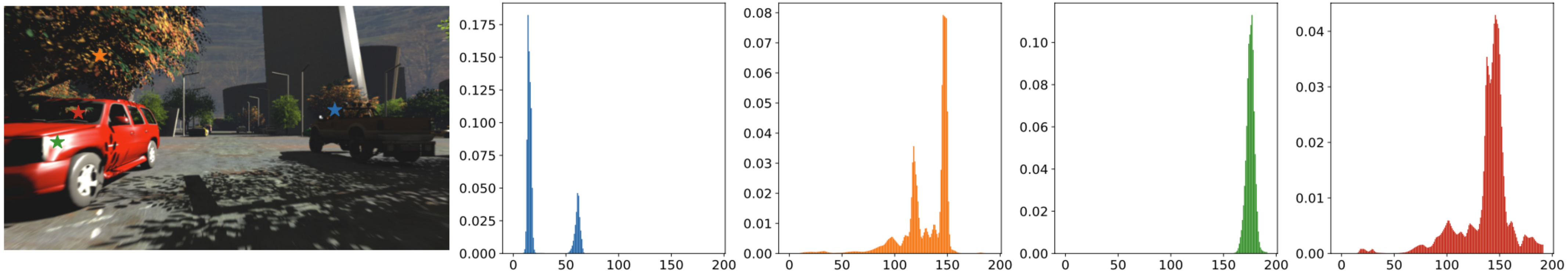
Xu P, Xiang Z, Qiao C, et al. Adaptive multi-modal cross-entropy loss for stereo matching. CVPR 2024.

1. Motivation

- Other ambiguous regions, such as repeating textures and transparency, intuitively should also be modeled as multi-modal distributions.

- Our goal is to **model the ground-truth distribution for all regions**, enabling the network to learn more generalizable matching principles.

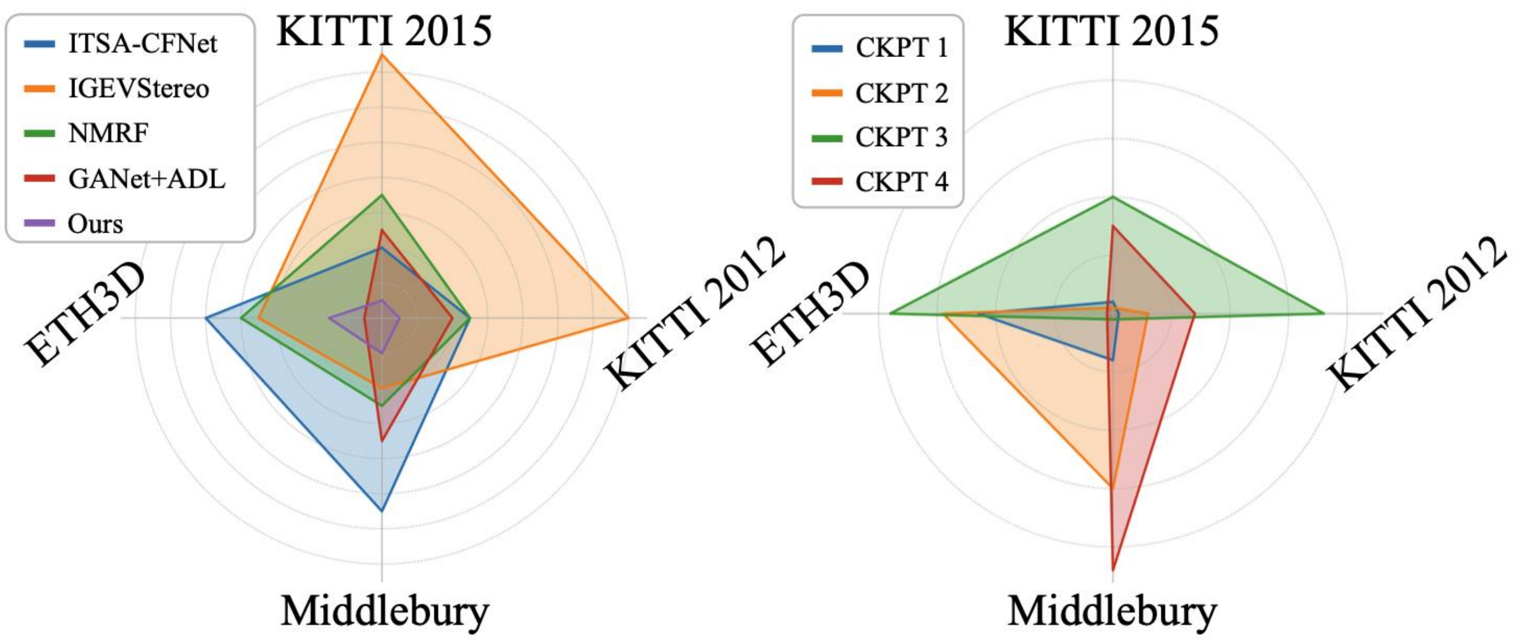
- **Dark knowledge**: stereo networks trained with a unimodal distribution spontaneously learn dark knowledge such as similarity and uncertainty.



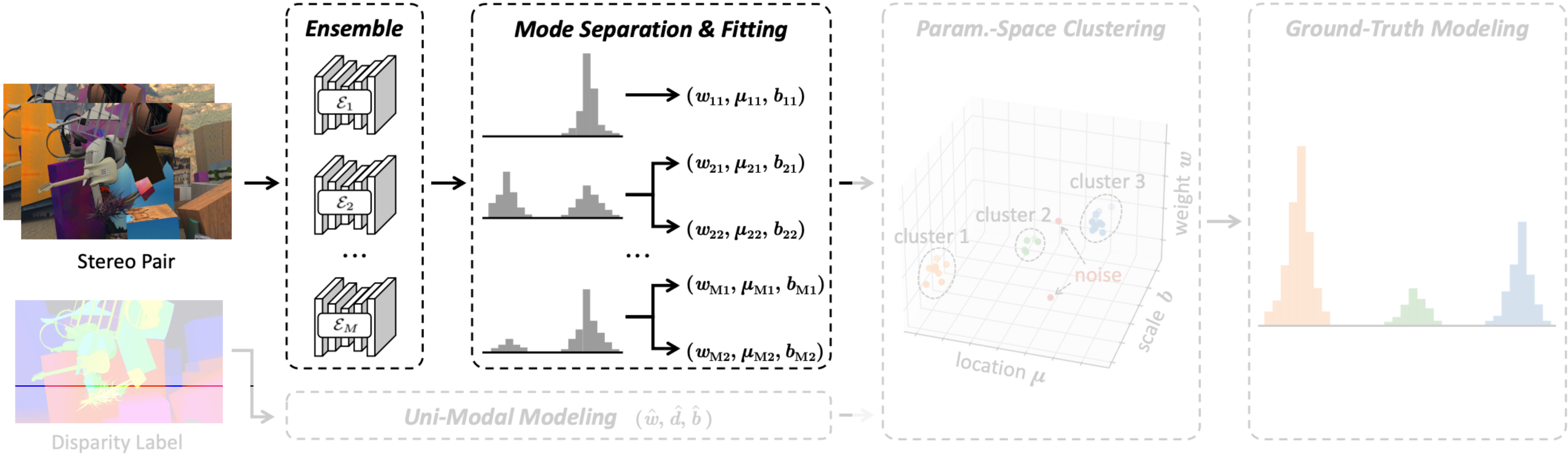
(a) Left Image (b) Edge (c) Repetitive-Textured (d) Weak-Textured (e) Transparent

1. Motivation

- Pretrained networks with different **architectures**, and even different **checkpoints** of the same architecture, all show **domain preferences** for different test sets.



2. MIDAS

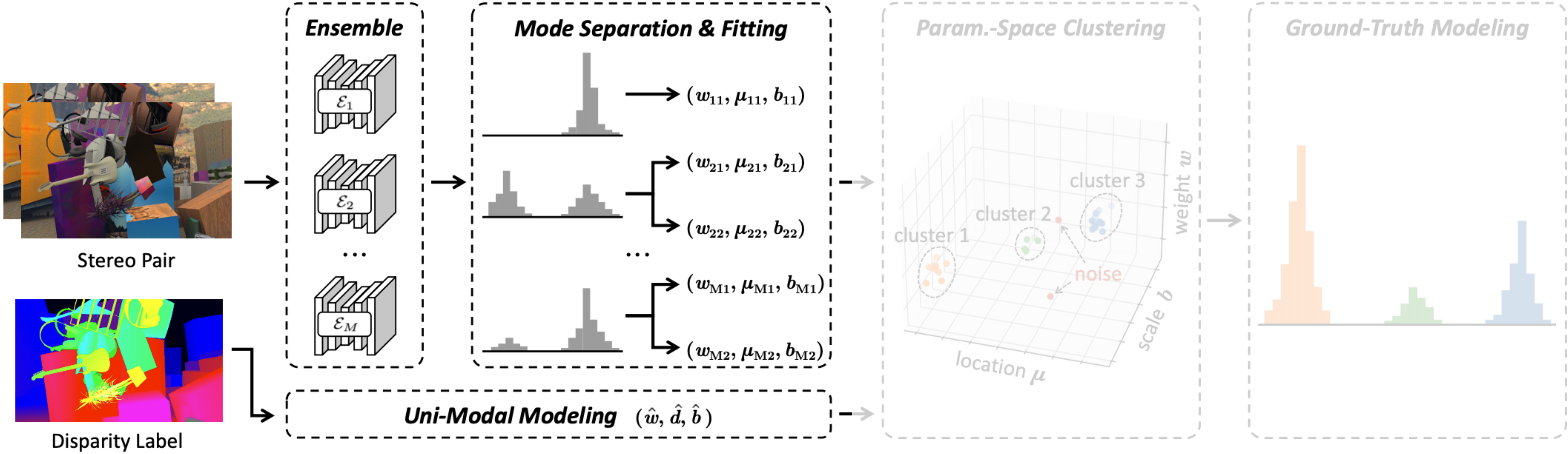


- For each pixel, the network ensemble predicts M multi-modal probability distributions.
- Individual modes are separated from these distributions and fitted as **parameterized Laplacians** (w, μ, b).

$$\mathrm{Laplacian}(\mathbf{d}; w, \mu, b) = w \cdot \frac{\exp(-\frac{|\mathbf{d}-\mu|}{b})}{\sum_{d \in \mathbf{d}} \exp(-\frac{|d-\mu|}{b})}$$

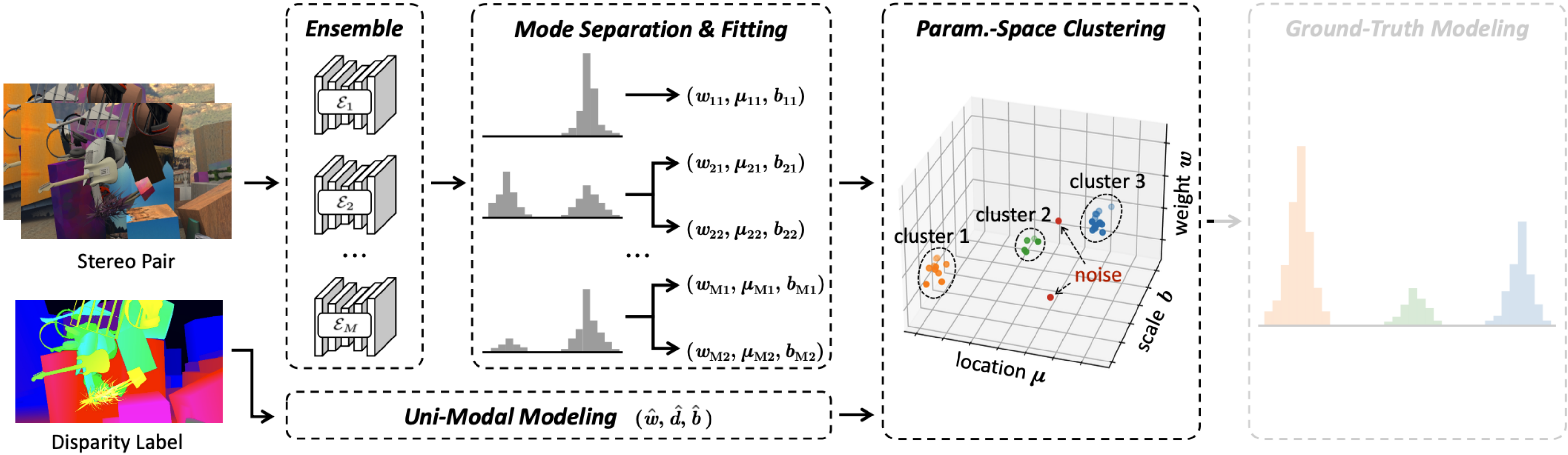
$$w \leftarrow \textstyle\sum_{d=l}^{r} \mathbf{p}[d]$$
$$\mu \leftarrow \textstyle\sum_{d=l}^{r} (\mathbf{p}[d]/w) \cdot d$$
$$b \leftarrow \textstyle\sum_{d=l}^{r} (\mathbf{p}[d]/w) \cdot |d - \mu|$$

2. MIDAS



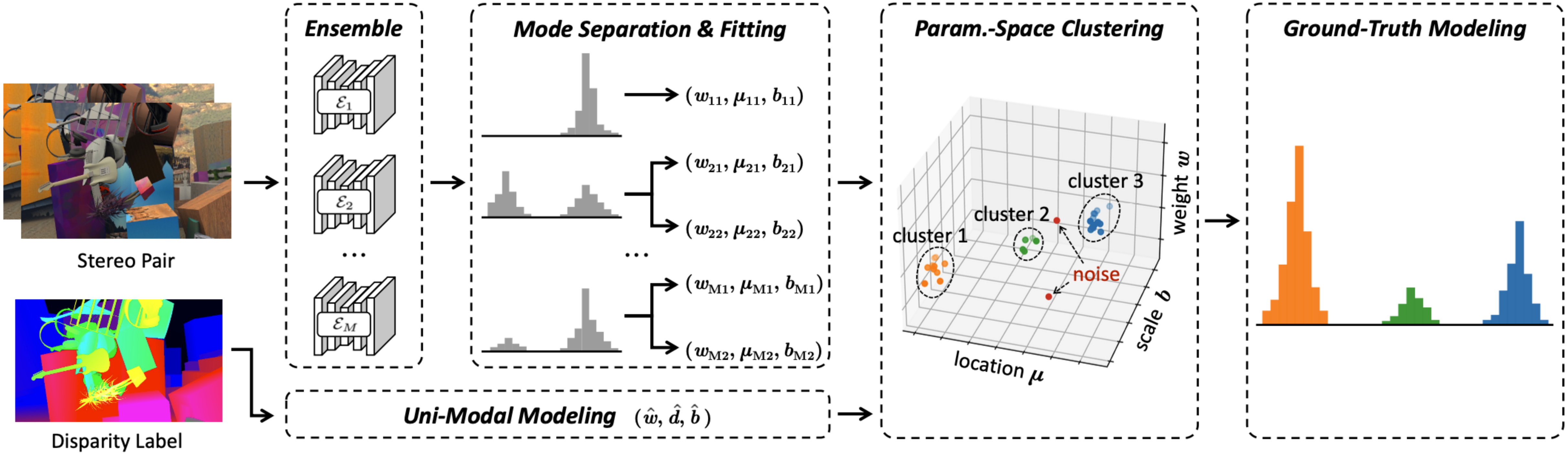
- For each pixel, the network ensemble predicts M multi-modal probability distributions.
- Individual modes are separated from these distributions and fitted as **parameterized Laplacians** (w, μ, b).
- The disparity ground-truth is also modeled as the uni-modal Laplacian with coordinate $(\hat{w}, \hat{d}, \hat{b})$.

2. MIDAS



- 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 $(\widehat{w}, \hat{d}, \hat{b})$.
- We cluster the points in the parameter space to distinguish the **objective knowledge** (effective clusters) from the **biased knowledge** (noise).

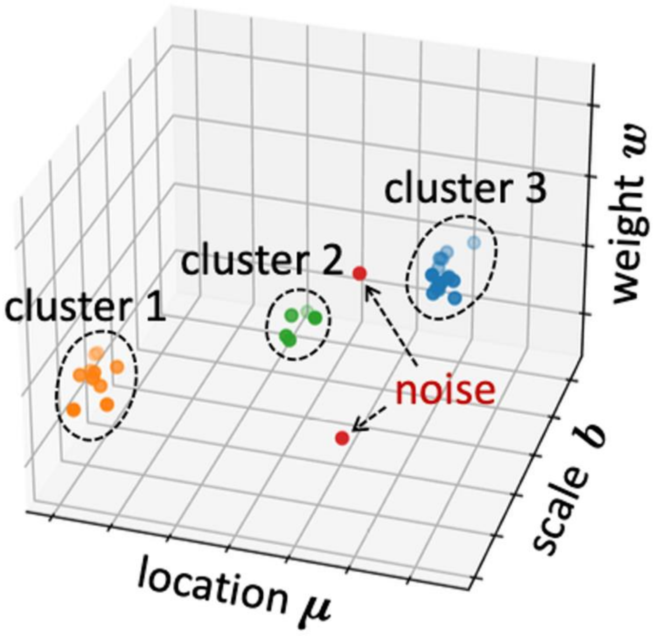
2. MIDAS

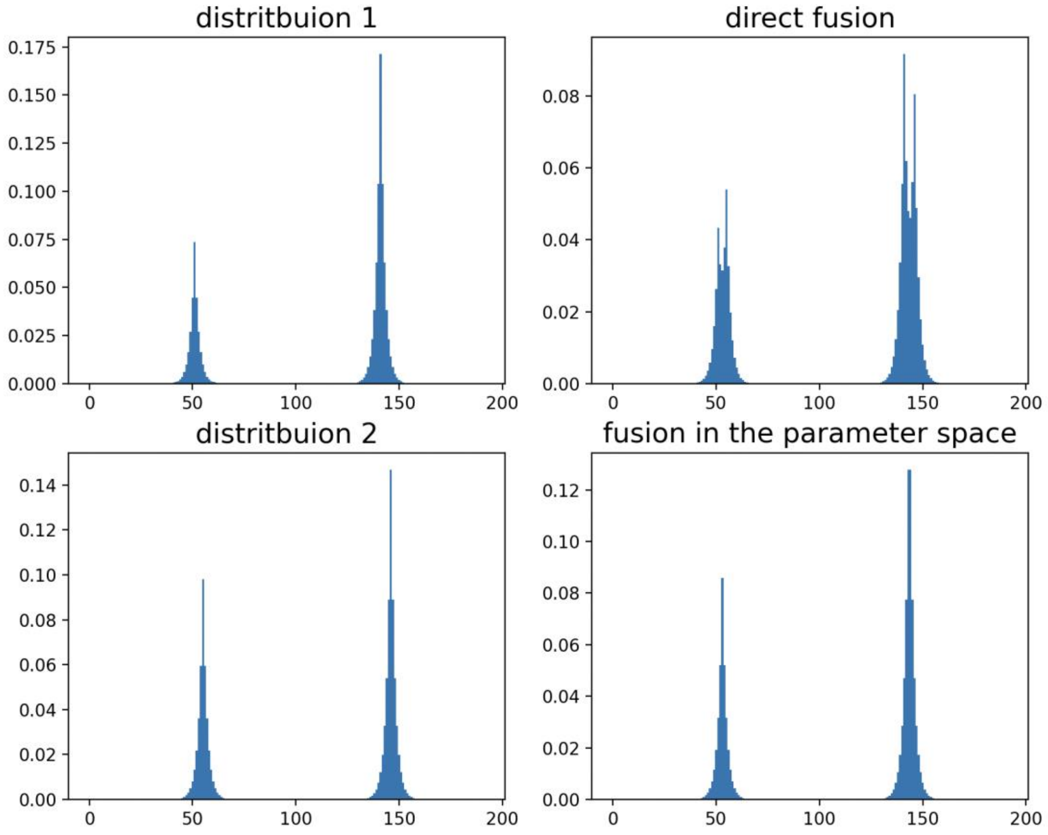


➢ The elements within each cluster are **fused** and **re-modeled** as a formulated mode in the final ground-truth distribution.

$$\hat{\mathbf{p}} = \sum_{k=1}^{K} \mathrm{Laplacian}(\mathbf{d}; w_k, \mu_k, b_k) = \sum_{k=1}^{K} w_k \cdot \frac{\exp(-\frac{|\mathbf{d}-\mu_k|}{b_k})}{\sum_{d\in\mathbf{d}} \exp(-\frac{|d-\mu_k|}{b_k})}$$

2. MIDAS





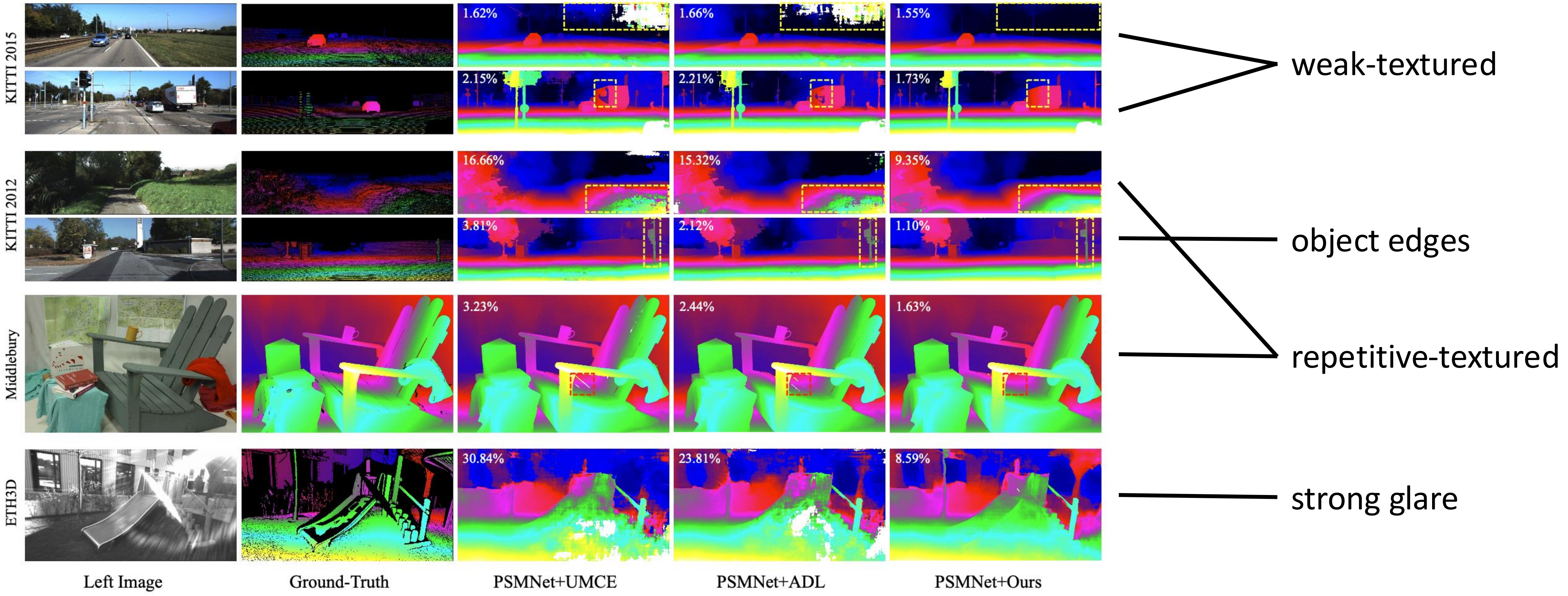
- filter out biased knowledge
- preserves unimodal property

3. Experiments

➢ Trained on **synthetic** dataset SceneFlow and evaluated on four **real-world** datasets

Method	Publication	KITTI 2015 >3px	KITTI 2012 >3px	Middlebury >2px	ETH3D >1px	Mean Rank
PSMNet [1]	CVPR 2018	16.30^{18}	15.10^{18}	25.10^{18}	23.80^{18}	18.00
GwcNet [14]	CVPR 2018	12.80^{17}	11.70^{17}	18.10^{16}	9.00^{16}	16.50
GANet [45]	CVPR 2019	11.70^{16}	10.10^{16}	20.30^{17}	14.10^{17}	16.5
DSMNet [46]	ECCV 2020	6.50^{15}	6.20^{15}	13.80^{13}	6.20^{14}	14.25
CFNet [32]	CVPR 2021	5.80^{12}	4.70^{11}	15.30^{14}	5.80^{12}	12.25
Mask-CFNet [29]	CVPR 2023	5.80^{12}	4.80^{12}	13.70^{12}	5.70^{11}	11.75
Raft-Stereo [23]	3DV 2021	5.70^{11}	5.20^{14}	12.60^{11}	3.30^{6}	10.50
FC-GANet [48]	CVPR 2022	5.30^{9}	4.60^{10}	10.20^{9}	5.80^{12}	10.00
PCWNet [33]	ECCV 2022	5.60^{10}	4.20^{5}	15.77^{15}	5.20^{10}	10.00
IGEV-Stereo [39]	CVPR 2023	6.03^{14}	5.18^{13}	7.27^{3}	3.60^{7}	9.25
Graft-GANet [24]	CVPR 2022	4.90^{6}	4.20^{5}	9.80^{8}	6.20^{14}	8.25
ITSA-CFNet [8]	CVPR 2022	4.70^{4}	4.20^{5}	10.40^{10}	5.10^{9}	7.00
StereoRisk [25]	ICML 2024	5.19^{8}	4.43^{9}	9.32^{7}	2.41^{2}	6.50
NMRF [13]	CVPR 2024	5.10^{7}	4.20^{5}	7.50^{4}	3.80^{8}	6.00
GANet + ADL [40]	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**

3. Experiments



Thanks!

Date: Oct.20, 2025