Self-Distilled Depth Refinement with Noisy Poisson Fusion

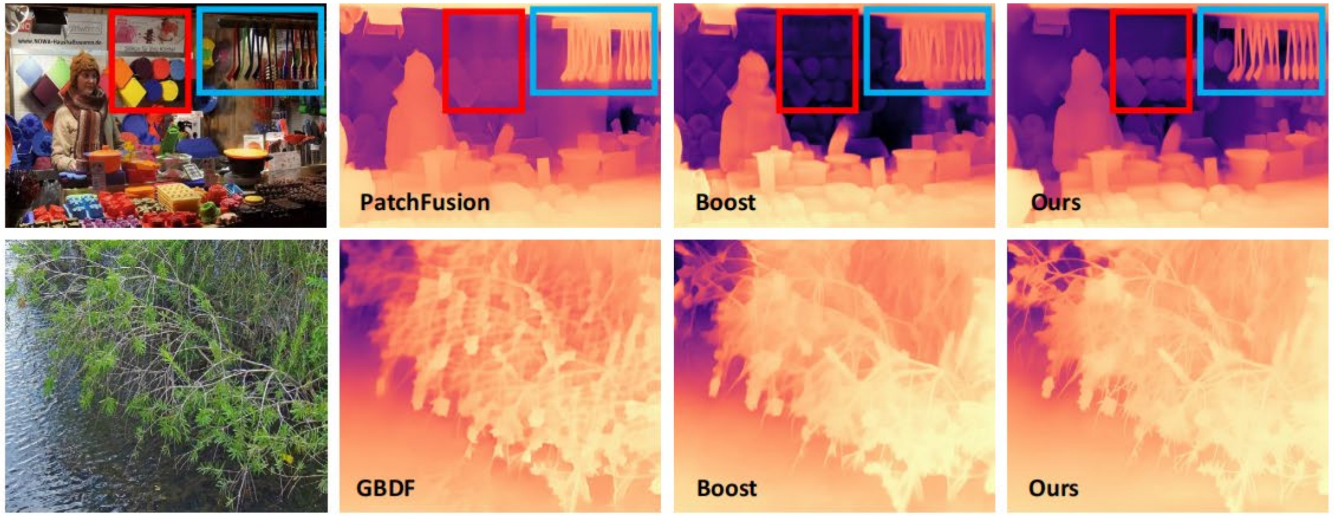
Jiaqi Li[1,*], Yiran Wang[1,*], Jinghong Zheng[1], Zihao Huang[1], Ke Xian[2],
Zhiguo Cao[1,†], Jianming Zhang[3]

[1] School of AIA, Huazhong University of Science and Technology
[2] School of EIC, Huazhong University of Science and Technology [3] Adobe Research

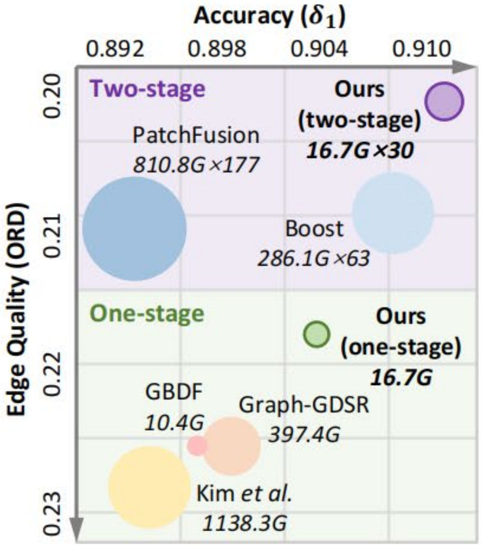
https://github.com/lijia7/SDDR

*Equal contribution; †Corresponding author.

- Depth refinement aims to infer high-resolution depth with fine-grained edges and details, refining low-resolution results of depth estimation models.



(a) Visualization of Depth Refinement Approaches



(b) Performance and Efficiency

- The prevailing methods adopt tile-based manners by merging numerous patches, which lacks efficiency and produces inconsistency.
- Besides, prior arts suffer from fuzzy depth boundaries and limited generalizability.
- Most of prior arts are trained only on synthetic datasets.

Methods	Efficiency	Performance
Tile-based	😞	😃 *(synthetic)*
One-stage	😃	😞 *(blur edge)*
Ours	😃	😃 *(robust)*

➢ Analyzing the fundamental reasons for these limitations, we model depth refinement as a noisy Poisson fusion problem with local inconsistency and edge deformation noises.

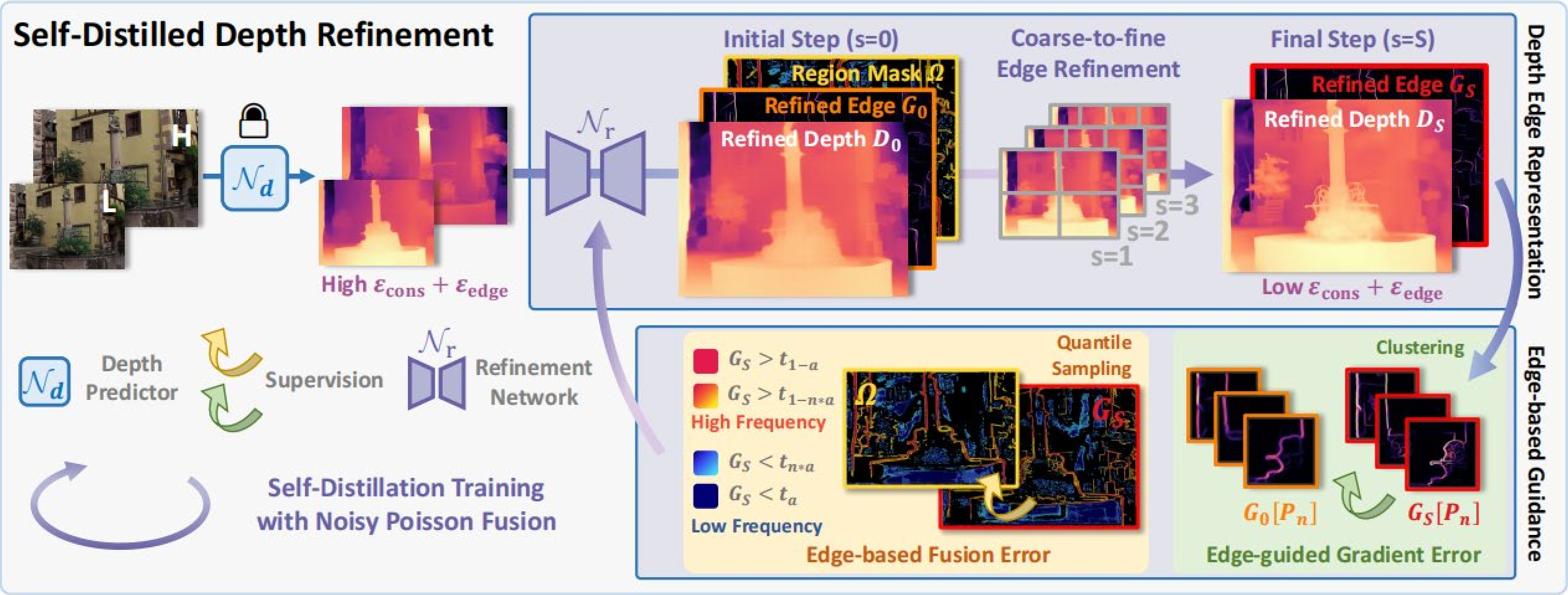
$$D_0 = \mathcal{N}_r(\mathcal{N}_d(L), \mathcal{N}_d(H))\,,$$
$$\text{s.t.} \min_{D_0,\Omega} \iint_{\Omega} |\nabla D_0 - \nabla D^*|\, \partial\Omega + \iint_{I-\Omega} |D_0 - D^*|\, \partial\Omega\,.$$

$$D_s^w \approx D^* + \epsilon_{\text{cons}} + \epsilon_{\text{edge}}\,,$$
$$\min_{G_s} \sum_w \iint_{\Omega_s^w} |G_s^w - \nabla D_s^w|\, \partial\Omega_s^w\,.$$
$$\min_{D0,\Omega} \iint_{\Omega} |\nabla D_0 - G_S|\, \partial\Omega + \iint_{I-\Omega} |D_0 - D_{gt}^*|\, \partial\Omega\,.$$

RGB | Ideal Depth D^* | $D^* + \epsilon_{\text{cons}} + \epsilon_{\text{edge}}$ | Prediction D | $\epsilon_{\text{cons}} + \epsilon_{\text{edge}}$ | Depth Error $D - D^*$

- Predictions of $\mathcal{N}_d$ are noisy and low-noise gradient domain optimization objectives ∇D^* are not available in realistic datasets. This is the key to previous methods limited by synthetic data.
- Qualitatively and quantitatively, it is demonstrated that the combination of two types of noise can characterize the overall degradation. A coarse-to-fine framework suppresses both types of noise in the predicted results.

Overview of Self-Distilled Depth Refinement

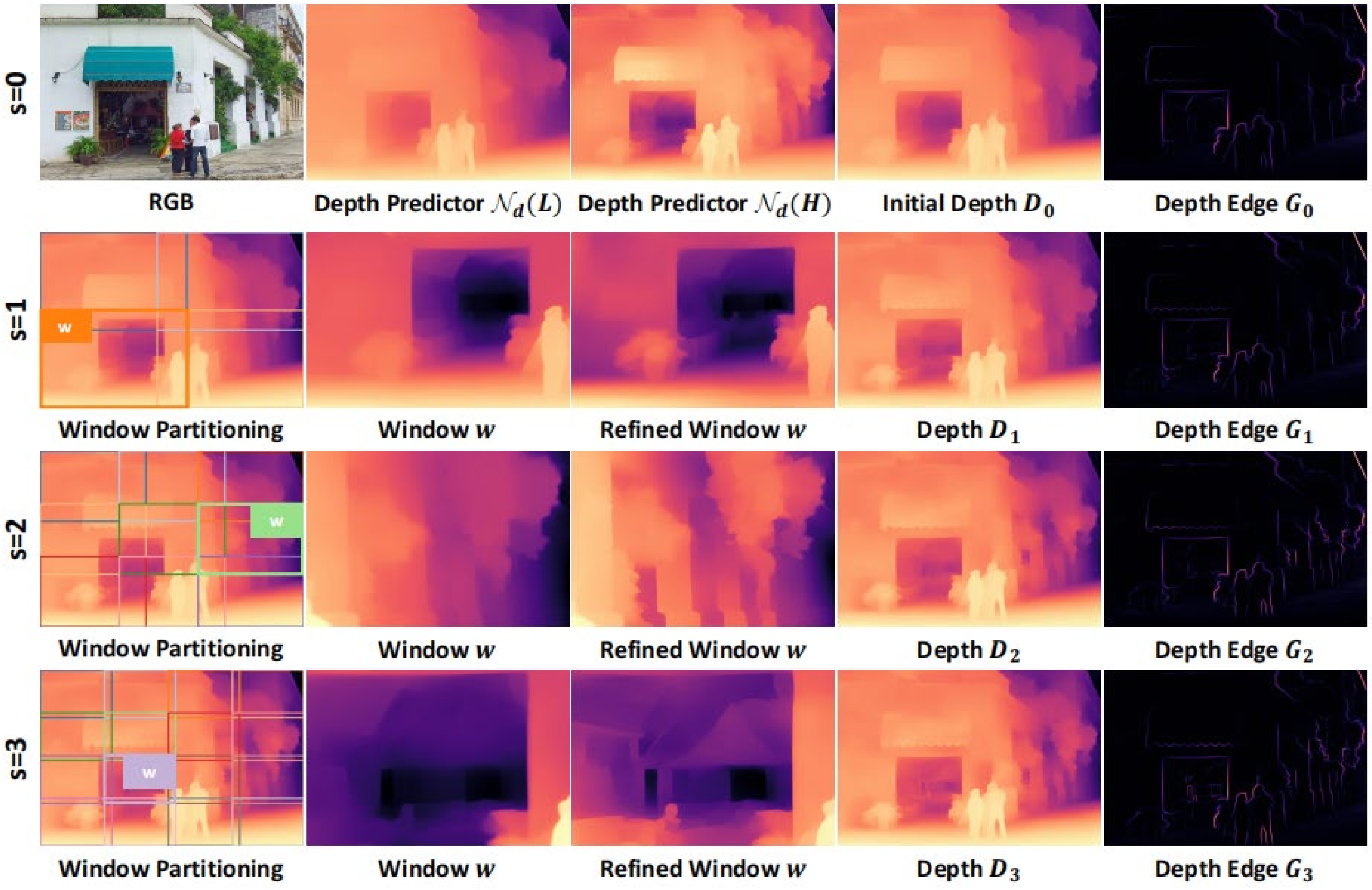


1) Refinement network $\mathcal{N}_r$ produces initial refined depth D_0, edge representation G_0, and learnable regional soft mask Ω.

*2) The final depth edge representation G_S is updated **from coarse to fine as pseudo-labels**.*

3) Edge-guided gradient loss and edge-based fusion loss supervises $\mathcal{N}_r$ to achieve consistent structures and fine-grained edges.

Overview of Self-Distilled Depth Refinement



*1) **Robustness:** Self-distilling framework generates plausible gradient pseudo-labeling in natural scenarios and can be trained across datasets to improve robustness*

*2) **Efficiency:** There is a coarse-to-fine generation process that can be used both for training to generate pseudo-labels and for two-stage reasoning, significantly improving efficiency.*

*3) **Accuracy:** Accurate pseudo-labeling makes the model after convergence from self-distillation training already have good accuracy for one-stage inference and even better accuracy for two-stage inference.*

Quantitative Comparison

Predictor	Method	Middlebury2021			Multiscopic			Hypersim		
		$\delta_1\uparrow$	REL↓	ORD↓	$\delta_1\uparrow$	REL↓	ORD↓	$\delta_1\uparrow$	REL↓	ORD↓
MiDaS	MiDaS [29]	0.868	0.117	0.384	0.839	0.130	0.292	0.781	0.169	0.344
	Kim *et al.* [14]	0.864	0.120	0.377	0.839	0.130	0.293	0.778	0.175	0.344
	Graph-GDSR [4]	0.865	0.121	0.380	0.839	0.130	0.292	0.781	0.169	0.345
	GBDF [3]	0.871	0.115	0.305	0.841	0.129	0.289	0.787	0.168	0.338
	Ours	**0.879**	**0.112**	**0.299**	**0.852**	**0.122**	**0.267**	**0.791**	**0.166**	**0.318**
LeReS	LeReS [49]	0.847	0.123	0.326	0.863	0.111	0.272	0.853	0.123	0.279
	Kim *et al.* [14]	0.846	0.124	0.328	0.860	0.113	0.286	0.850	0.125	0.286
	Graph-GDSR [4]	0.847	0.124	0.327	0.862	0.111	0.273	0.852	0.123	0.281
	GBDF [3]	0.852	0.122	0.316	0.865	0.110	0.270	0.857	0.121	0.273
	Ours	**0.862**	**0.120**	**0.305**	**0.870**	**0.108**	**0.259**	**0.862**	**0.120**	**0.273**
ZoeDepth	ZoeDepth [1]	0.900	0.104	0.225	0.896	0.097	0.205	0.927	0.088	0.198
	Kim *et al.* [14]	0.896	0.107	0.228	0.890	0.099	0.204	0.923	0.091	0.204
	Graph-GDSR [4]	0.901	0.103	0.226	0.895	0.096	0.208	0.926	0.089	0.199
	GBDF [3]	0.899	0.105	0.226	0.897	0.096	0.207	0.925	0.089	0.199
	Ours	**0.905**	**0.100**	**0.218**	**0.904**	**0.092**	**0.199**	**0.930**	**0.086**	**0.191**

Table 1: **Comparisons with one-stage methods.** As prior arts [14, 4, 3], we conduct evaluations with different depth predictors [29, 49, 1]. For each predictor, we report the initial metrics and results of refinement methods. Best performances with each depth predictors [29, 49, 1] are in boldface.

Predictor	Method	Middlebury2021			Multiscopic			Hypersim		
		$\delta_1\uparrow$	REL↓	ORD↓	$\delta_1\uparrow$	REL↓	ORD↓	$\delta_1\uparrow$	REL↓	ORD↓
MiDaS	MiDaS [29]	0.868	0.117	0.384	0.839	0.130	0.292	0.781	0.169	0.344
	Boost [24]	0.870	0.118	0.351	0.845	0.126	0.282	0.794	0.161	0.332
	Ours	**0.871**	**0.115**	**0.303**	**0.858**	**0.120**	**0.263**	**0.799**	**0.154**	**0.322**
LeReS	LeReS [49]	0.847	0.123	0.326	0.863	0.111	0.272	0.853	0.123	0.279
	Boost [24]	0.844	0.131	0.325	0.860	0.112	0.278	**0.865**	**0.118**	0.272
	Ours	**0.861**	**0.123**	**0.309**	**0.870**	**0.109**	**0.268**	0.858	0.123	**0.271**
ZoeDepth	ZoeDepth [1]	0.900	0.104	0.225	0.896	0.097	0.205	0.927	0.088	0.198
	Boost [24]	0.911	0.099	0.210	**0.910**	0.094	0.197	0.926	0.089	0.193
	PatchFusion [20]	0.887	0.102	0.211	0.908	0.095	0.212	0.881	0.116	0.258
	Ours	**0.913**	**0.096**	**0.202**	0.908	**0.091**	**0.197**	**0.933**	**0.083**	**0.189**

Table 2: **Comparisons with two-stage methods.** PatchFusion [20] only adopts ZoeDepth [1] as the fixed baseline, while other approaches are pluggable for different depth predictors [29, 49, 1].

Qualitative Comparison

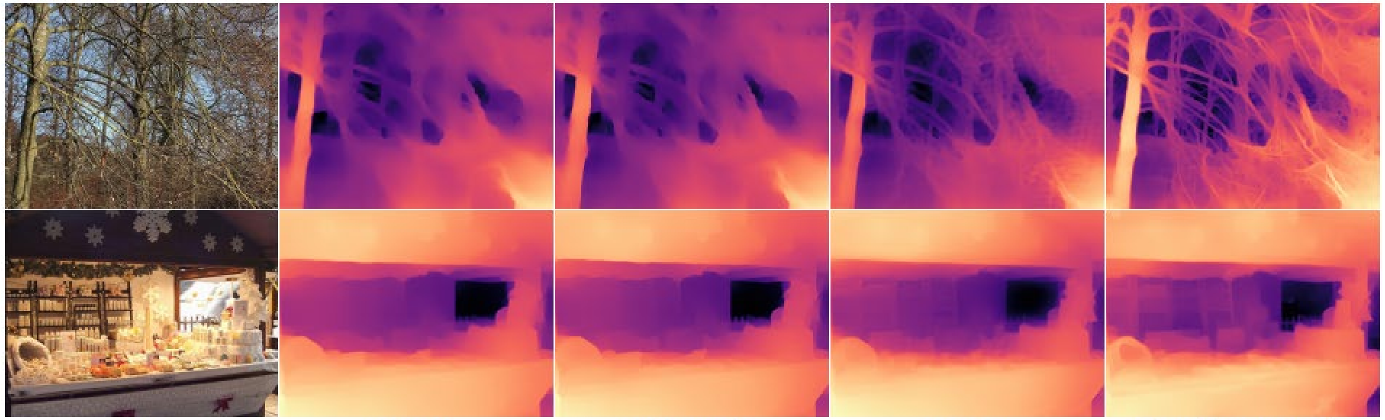


Figure 5: **Qualitative comparisons of one-stage methods on natural scenes.** LeReS [49] is used as the depth predictor. SDDR predicts sharper depth edges and more meticulous details than prior arts [3, 14], *e.g.*, fine-grained predictions of intricate branches. Better viewed when zoomed in.

Figure 6: **Qualitative comparisons of two-stage methods on natural scenes.** ZoeDepth [1] is adopted as the depth predictor. The SDDR with coarse-to-fine edge refinement can predict more accurate depth edges and more consistent spatial structures than the tile-based methods [20, 24].

Ablation Studies

Method	DIML				DIODE			
	$\delta_1\uparrow$	REL↓	ORD↓	$D^3R\downarrow$	$\delta_1\uparrow$	REL↓	ORD↓	$D^3R\downarrow$
LeReS [49]	0.902	0.101	0.242	0.284	0.892	0.105	0.324	0.685
Kim *et al.* [14]	0.902	0.100	0.243	0.301	0.889	0.105	0.325	0.713
Graph-GDSR [4]	0.901	0.101	0.243	0.300	0.890	0.104	0.326	0.690
GBDF [3]	0.906	0.100	0.239	0.267	0.894	0.105	0.322	0.673
Boost [24]	0.897	0.108	0.274	0.438	0.892	0.105	0.343	0.640
Ours	**0.926**	**0.098**	**0.221**	**0.220**	**0.900**	**0.098**	**0.293**	**0.637**

Table 3: **Comparisons of model generalizability.** We conduct zero-shot evaluations on DIML [15] and DIODE [39] datasets with diverse in-the-wild scenarios to compare the generalization capability. We adopt LeReS [49] as the depth predictor for all the compared methods in this experiment.

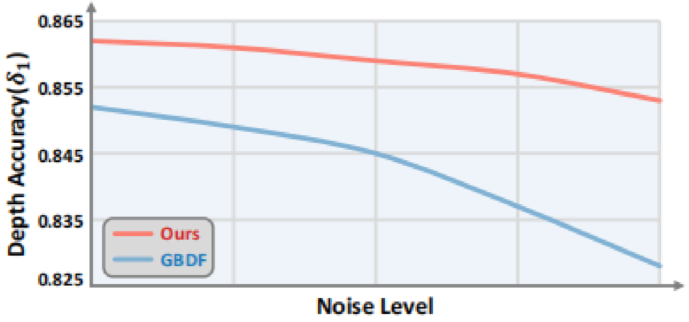


Figure 7: **Robustness against noises.** X-axis shows noise level of $\epsilon_{cons} + \epsilon_{edges}$. With higher noises, our SDDR is more robust with less performance degradation than the prior GBDF [3].

Method	$\delta_1\uparrow$	REL↓	ORD↓	$D^3R\downarrow$
$S=0$	0.859	0.125	0.313	0.235
$S=1$	0.860	0.122	0.309	0.223
$S=2$	0.860	0.120	0.307	0.219
$S=3$	**0.862**	**0.120**	**0.305**	**0.216**

(a) Coarse-to-fine Edge Refinement

$\mathcal{L}_{gt}$	$\mathcal{L}_{grad}$	$\mathcal{L}_{fusion}$	$\delta_1\uparrow$	REL↓	ORD↓	$D^3R\downarrow$
✓			0.854	0.124	0.313	0.240
✓	✓		0.858	0.122	0.307	0.220
✓		✓	0.859	0.120	0.311	0.229
✓	✓	✓	**0.862**	**0.120**	**0.305**	**0.216**

(b) Edge-based Guidance

Method	Training Data	$\delta_1\uparrow$	REL↓	ORD↓	$D^3R\downarrow$
GBDF [3]	HRWSI [47]	0.852	0.122	0.316	0.258
Ours	HRWSI [47]	**0.860**	**0.121**	**0.309**	**0.222**

(c) Effectiveness

Method	$\delta_1\uparrow$	REL↓	ORD↓	$D^3R\downarrow$
GBDF [3]	0.852	0.122	0.316	0.258
GBDF (w/G_S)	**0.858**	**0.122**	**0.307**	**0.230**

(d) Transferability

Table 4: **Ablation Study.** All ablations are on Middlebury2021 [33] with depth predictor LeReS [49].

Thanks for watching!

Paper, code, and videos are available:
https://github.com/lijia7/SDDR
Feel free to contact
lijiaqi.work@outlook.com