Occlusion Boundary and Depth: Mutual Enhancement via Multi-Task Learning

Lintao XU[1], Yinghao WANG[2], Chaohui WANG[1]

[1]LIGM, Univ. Gustave Eiffel, École des Ponts, CNRS, France

[2]LTCI, Télécom Paris, Institute Polytechnique de Paris, France

WACV 2026
Tucson, Arizona • Mar 6 - 10, 2026

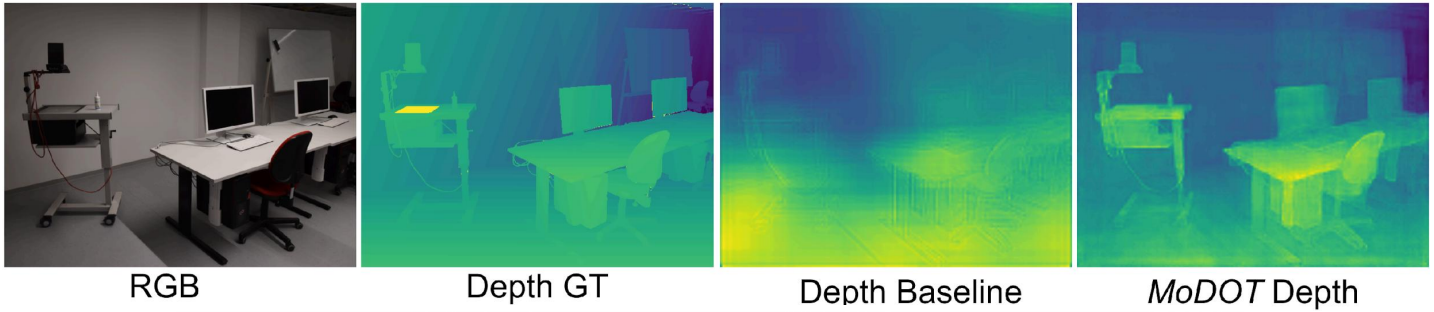
Introduction

- **Monocular Depth Estimation:**
  - A continuous geometric regression task
  - Infers depth information from a single image
- **Occlusion Boundary (OB) Estimation:**
  - A sparse structural prediction task
  - Identifies boundaries caused by inter-object occlusions and self-occlusions

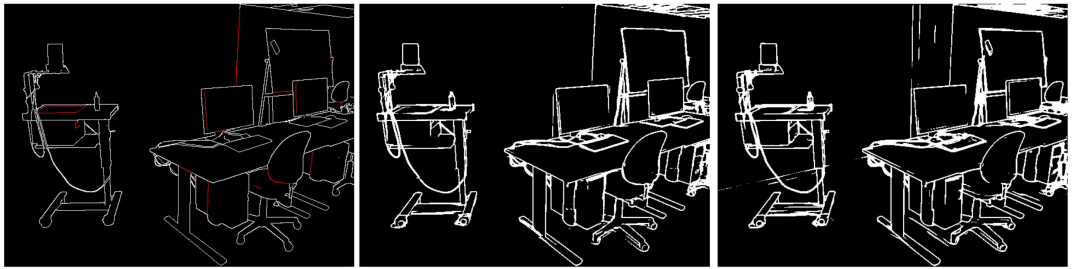
- **The challenge** - inherent ill-posedness of both tasks due to single-view ambiguity
- **The opportunity** - mutual benefits between the two tasks:
  - Critical geometric cues from OBs for resolving depth ambiguities
  - Depth information for refining occlusion reasoning

Introduction

Zero-shot results on iBims-1



RGB | Depth GT | Depth Baseline | *MoDOT* Depth



OB GT | OB Baseline | *MoDOT* OB

Joint estimation → more accurate & structurally consistent depth maps, especially around OBs

Parameters → only 11M additional (4% of single-task depth baseline)

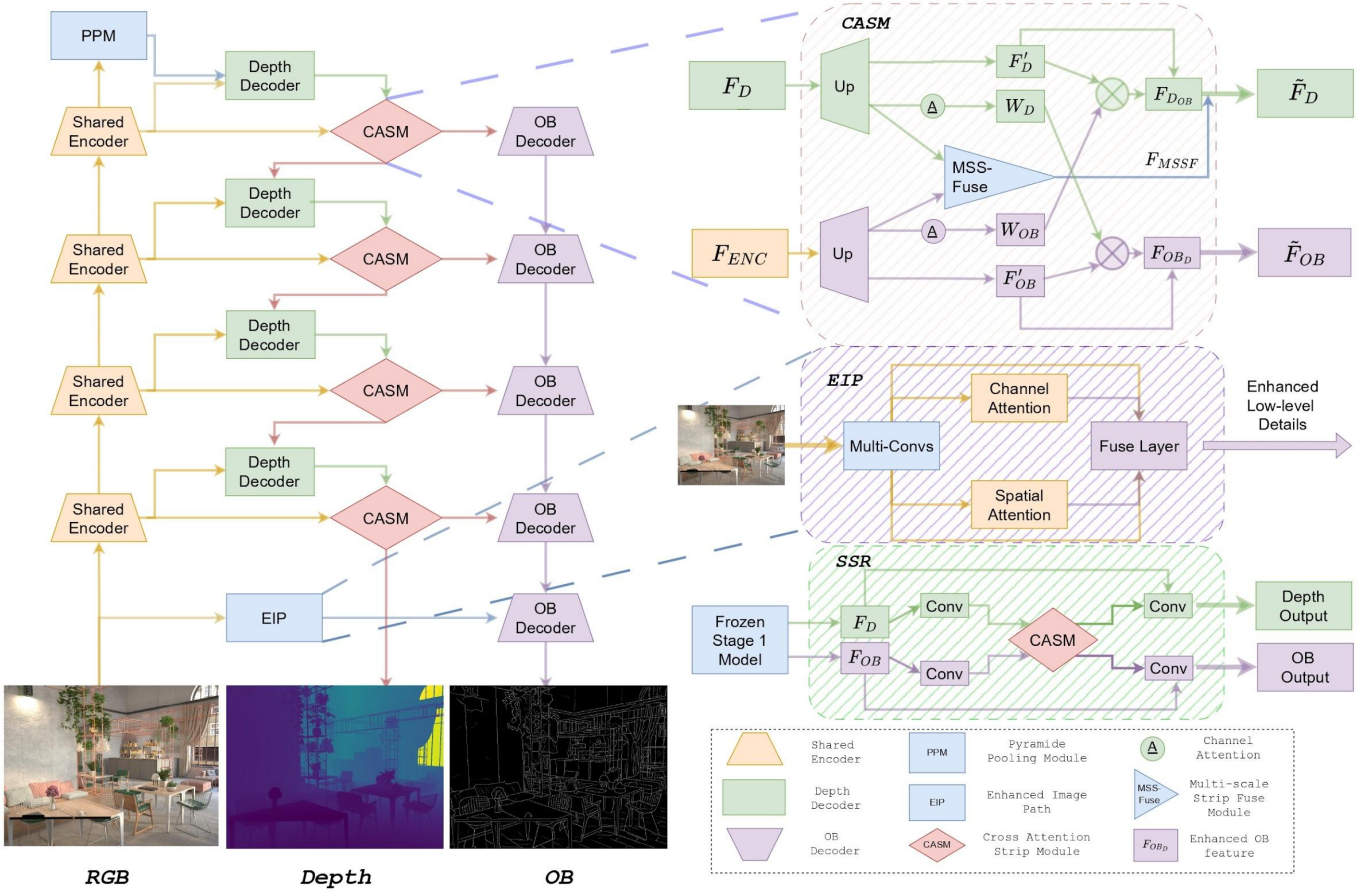
Our Contributions

- ***MoDOT*:** the first multi-task framework that simultaneously optimizes depth and OBs, achieving significantly better performance than competitors

- **OBDCL loss** & **CASM module:** effectively capture geometric correlations between OBs and depth discontinuities

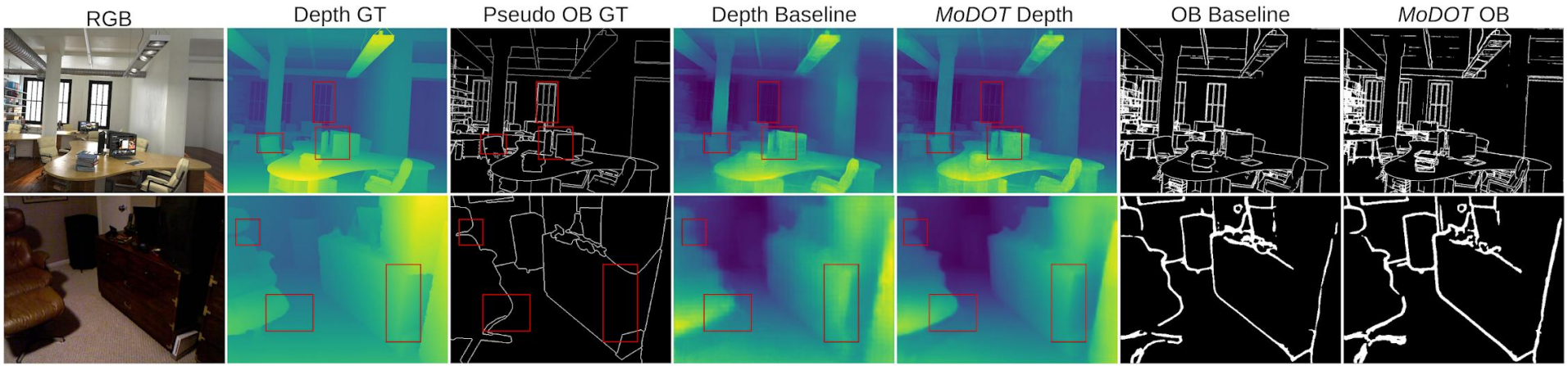
- **OB-Hypersim:** a photorealistic dataset with pixel-perfect annotations for both tasks across diverse indoor scenes

MoDOT Pipeline



Qualitative Examples

Qualitative results on OB-Hypersim (synthetic) and NYUD-v2 (real)



Quantitative Results

Quantitative results on two synthetic datasets

Method	OB-FUTURE				*OB-Hypersim*			
	RMSE↓	Abs Rel↓	$\delta < 1.25$ ↑	OB-Recall↑	RMSE↓	Abs Rel↓	$\delta < 1.25$ ↑	OB-Recall↑
Depth Baseline [51]	0.4524	0.1011	0.9215	-	0.6948	0.3123	0.4759	-
OB Baseline (ours)	-	-	-	0.7655	-	-	-	0.8099
SharpNet [29]	0.9535	0.1890	0.5966	**0.9571**	0.8551	0.4422	0.3561	0.7342
MTAN [20]	0.5576	0.1218	0.8523	0.9238	0.8050	0.3765	0.4023	0.7430
PAD-Net [43]	0.5447	0.1188	0.8627	0.9022	0.8404	0.4322	0.3866	0.6732
MTI-Net [37]	0.5064	0.1106	0.8891	0.9125	0.7560	0.3746	0.4365	0.7490
InvPT [49]	0.9335	0.2371	0.6122	0.3228	0.9018	0.5106	0.3555	0.8004
DenseMTL [23]	0.5217	0.1106	0.8818	0.8927	0.7475	0.4095	0.4410	0.8520
Ours	0.3963	0.0901	0.9427	0.9090	0.6583	0.2963	0.5167	0.8670
Ours + *SSR*	**0.3809**	**0.0843**	**0.9518**	0.9486	**0.6537**	**0.2954**	**0.5148**	**0.8732**

Quantitative Results

Quantitative comparisons on real-world NYUD-v2

Details	RMSE↓	$RMSE_{log}$ ↓	Abs Rel↓	Sq Rel↓	log10↓	$\delta < 1.25$ ↑	OB-Recall↑	OB-Fscore↑
Depth Baseline [51]	0.4619	0.1684	0.1363	0.0876	0.0558	0.8438	-	-
OB Baseline (ours)	-	-	-	-	-	-	0.5509	0.1736
SharpNet [29]	0.6188	0.2215	0.1888	0.1523	0.0777	0.7107	0.5072	0.1642
MTAN [20]	0.5499	0.2043	0.1664	0.1241	0.0685	0.7655	0.5439	0.1544
PAD-Net [43]	0.6404	0.2361	0.2009	0.1706	0.0810	0.6964	0.5681	0.1493
MTI-Net [37]	0.5630	0.2004	0.1674	0.1303	0.0687	0.7650	0.5049	0.1596
InvPT [49]	0.6574	0.2389	0.2087	0.1871	0.0830	0.6998	0.5746	0.1443
DenseMTL [23]	0.5158	0.1852	0.1502	0.1081	0.0631	0.8006	**0.6331**	0.1616
TaskPrompter [50]	0.4446	0.1548	0.1228	0.0758	0.0525	0.8591	0.2906	0.1626
MLoRE [47]	0.4576	0.1601	0.1285	0.0819	0.0544	0.8493	0.5059	0.1616
Ours	0.4174	0.1475	0.1169	0.0692	0.0498	0.8741	0.6287	0.1729
Ours + *SSR*	**0.4137**	**0.1460**	**0.1155**	**0.0679**	**0.0492**	**0.8767**	0.6244	**0.1863**

Comparison with Depth Anything

Ablation on Canny thresholds for extracting coarse OBs from depth maps for OB evaluation

		T (0.05, 0.15)		T (0.15, 0.30)		T (0.25, 0.50)	
Method	**Dataset**	**OB-Recall↑**	**OB-Fscore↑**	**OB-Recall↑**	**OB-Fscore↑**	**OB-Recall↑**	**OB-Fscore↑**
Depth Anything Large	OB-FUTURE	0.0846	0.1276	0.0581	0.0928	0.0386	0.0646
Depth Anything v2 Large	OB-FUTURE	**0.1136**	**0.1630**	0.0873	0.1323	0.0642	0.1015
Depth Anything Large	*OB-Hypersim*	0.0294	0.0446	0.0198	0.0341	0.0133	0.0238
Depth Anything v2 Large	*OB-Hypersim*	**0.0595**	**0.0936**	0.0491	0.0809	0.0402	0.0681

Comparison with depth-only methods on OB-Hypersim (metrics depth)

Method	OB-Recall↑	OB-Fscore↑	RMSE↓	$RMSE_{log}$ ↓	Abs Rel↓	Sq Rel↓	log10↓	$\delta < 1.25$ ↑
Depth Anything Large	0.0294	0.0446	1.1461	0.6358	0.4118	0.5581	0.2439	0.2677
Depth Anything v2 Large	0.0595	0.0936	1.1269	0.5934	0.3604	0.5260	0.2205	0.3433
Depth Anything v2 Base	0.0585	0.0911	1.1287	0.5984	0.3683	0.5287	0.2229	0.3307
Depth Anything v2 Small	0.0530	0.0824	1.1308	0.6034	0.3775	0.5346	0.2252	0.3249
Ours	0.8670	**0.5163**	0.6583	0.3463	0.2963	0.2279	**0.1223**	**0.5167**
Ours + *SSR*	**0.8732**	0.5109	**0.6537**	**0.3456**	**0.2954**	**0.2266**	0.1243	0.5148

Method Analysis

Depth performance with different auxiliary tasks (Bold: best performance)

Auxiliary GT	RMSE↓	$RMSE_{log}$ ↓	Abs Rel↓	Sq Rel↓	log10↓	$\delta < 1.25$ ↑
None	0.6948	0.3739	0.3123	0.2481	0.1342	0.4759
Occlusion boundary	**0.6583**	**0.3463**	0.2963	**0.2279**	**0.1235**	**0.5167**
Instance contour	0.6736	0.3580	0.3195	0.2496	0.1278	0.4991
Semantic contour	0.6798	0.3613	0.3131	0.2453	0.1294	0.5013
Edge	0.6786	0.3557	0.2988	0.2409	0.1273	0.5082
Semantic segmentation	0.6699	0.3591	**0.2883**	0.2315	0.1281	0.5058
Surface normal	0.6729	0.3603	0.3025	0.2377	0.1275	0.5108

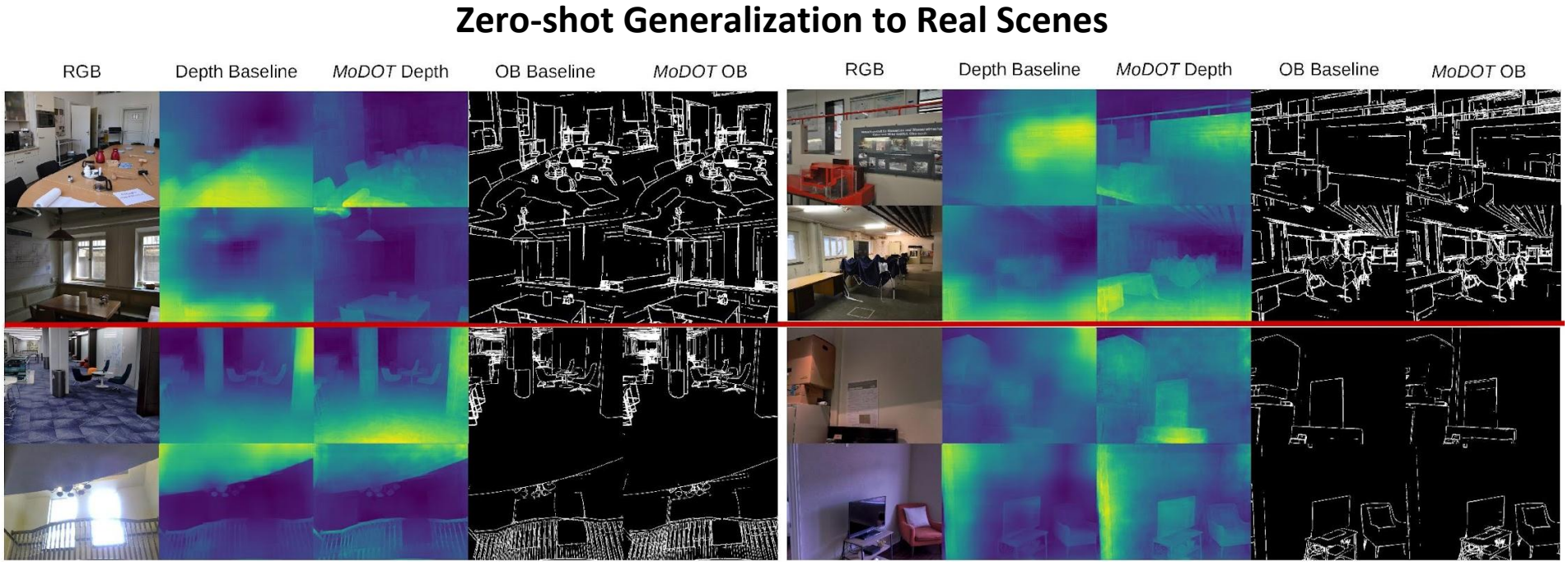
(Self-occlusion-aware) OBs → the most significant depth estimation improvements

Method Analysis

Component-wise ablation study of *MoDOT* on OB-FUTURE

Method	RMSE↓	$RMSE_{log}$ ↓	Abs Rel↓	Sq Rel↓	log10↓	$\delta < 1.25$ ↑	OB-Recall↑	OB-Fscore↑
Depth Baseline [51]	0.4524	0.1149	0.1011	0.0655	0.0327	0.9215	-	-
OB Baseline (ours)	-	-	-	-	-	-	0.7655	0.5634
Shared encoder + **T**wo decoder (ST)	0.4376	0.1117	0.0983	0.0616	0.0414	0.9265	0.9154	0.6042
ST + *CASM*	0.4051	0.1038	0.0907	0.0536	0.0385	0.9415	0.9388	0.5621
ST + *CASM* + *OBDCL*	0.3960	0.1007	0.0883	0.0503	0.0372	0.9494	0.9287	0.5895
Ours (ST + *CASM* + *OBDCL* + *EIP*)	0.3963	0.1020	0.0901	0.0523	0.0380	0.9427	0.9090	**0.6131**
Ours (w/o *EIP*) + *SSR*	0.3863	0.0983	0.0845	0.0469	0.0363	0.9516	0.9437	0.5510
Ours + *SSR*	**0.3809**	**0.0974**	**0.0843**	**0.0468**	**0.0361**	**0.9518**	**0.9486**	0.5415

More Zero-Shot Examples



Evaluation on IBims-1 and DIODE, where both single-task baselines and ***MoDOT*** are trained exclusively on the proposed synthetic OB-Hypersim dataset.

Future Work

- A module that consistently improves performance across all OB metrics
- Incorporating more tasks (i.e., surface normals, segmentation) into *MoDOT*
- Extending occlusion–depth modeling to outdoor and in-the-wild scenes

Paper and Code

GitHub

arXiv

Thank you for your attention

Q&A