Interactive Occlusion Boundary Estimation through Exploitation of Synthetic Data

Lintao XU, Chaohui WANG

LIGM, Univ. Gustave Eiffel, École des Ponts, CNRS, France

BMVC 2025

Introduction

OB — Fundamental geometry cue for visual scene understanding

- Semantics-free mathematically defined geometry
- Includes (non-closed) self-occlusion boundaries
- Full-image, pixel-level, all visible boundaries
- Excludes texture, illumination, reflections, etc.

Previous Occlusion Boundary (OB) datasets

Our synthetic and real-world benchmarks

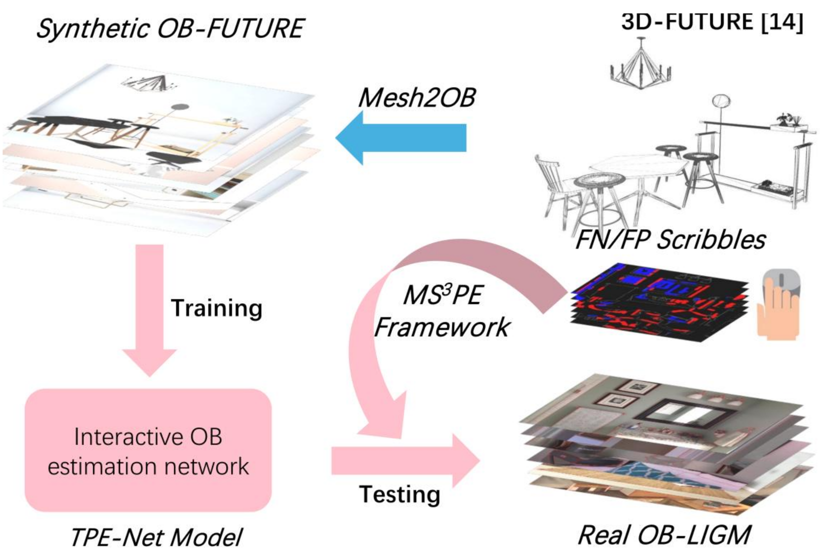
Benchmark visualization (**Red**: ground truth; **Blue**: regions containing missing/incomplete/erroneous OBs)

Motivations:

- Development of efficient methods for high-quality OB annotation
- Investigation of interactive methods for OBs, including both open and closed boundaries

Our contributions:

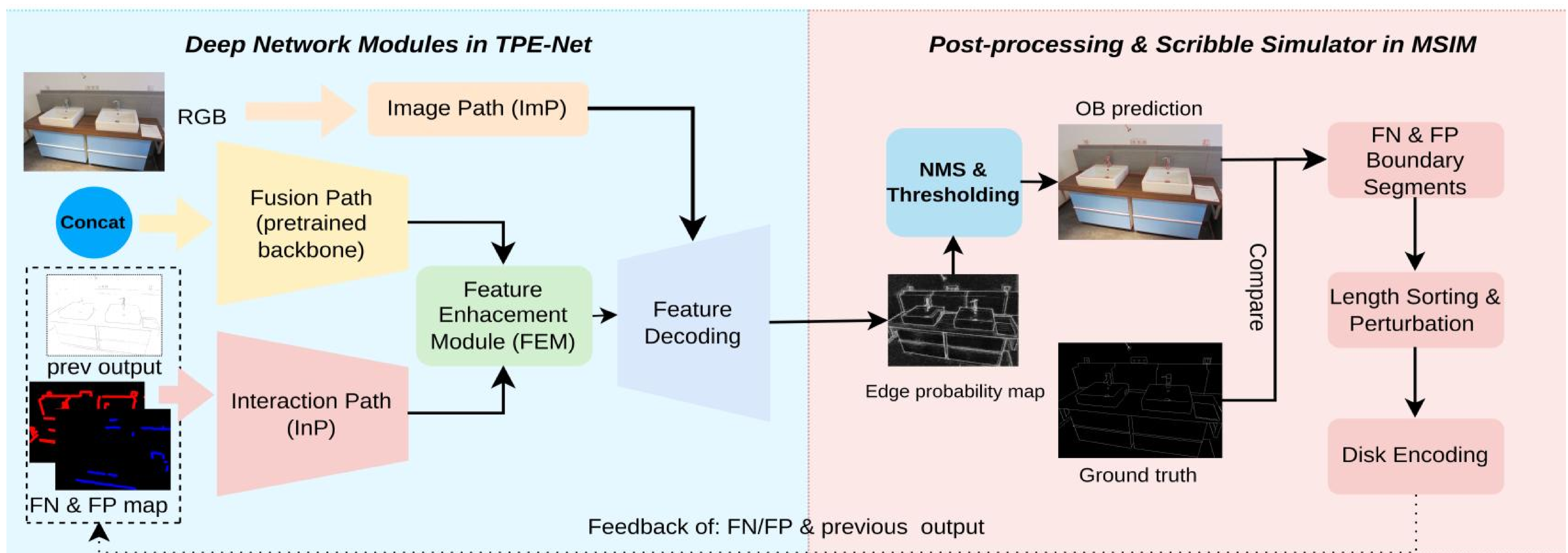
- Introducing ***MS³PE***, the first multi-scribble-guided framework for interactive OB estimation/annotation
- Proposing ***Mesh2OB***, an automated pipeline for generating high-quality synthetic OB data from 3D scenes
- Presenting ***OB-LIGM***, a novel real-world benchmark with pixel-level annotations for robust evaluation



Interactive Framework *MS³PE*

MS³PE **-** A multi-scribble-guided deep learning framework

- ***MSIM***: intuitive multi-scribble interaction mechanism
- ***TPE-Net***: three-path encoder with multi-scale strip convolutions
- **Training**: synthetic *OB-FUTURE* → **Evaluation**: real-world *OB-LIGM*



***MS³PE* framework:** integration of the deep network *TPE-Net* into the interaction mechanism *MSIM*

MSIM highligts:

- Two-stage training scheme
- Scribble simulator
- Boundary-scribble interaction
- One-shot interaction strategy

Two-step inference:

- **Step 1:** RGB + 3 zero maps → Initial OB prediction
- **Step 2:** Initial OB output + FN & FP scribbles → Refined OB prediction

Dataset Comparison & *Mesh2OB*

Datasets	General Information				OB GT's Characteristics		
	Data Volume	Image Resolution	Scene Type	GT Annotation	Self-occlusion Boundary	Object-contour Completeness	Full-image Annotation
CMU [62]	30	(640, 480)	real	manual	ignored	✓	×
BSDS Ownership [58]	200	(481, 321)	real	manual	ignored	✓	✓
PIOD [70]	10,000	(475, 390)	real	manual	ignored	✓	×
NYUv2-OC++ [56]	654	(592, 440)	real	manual	incomplete	×	×
iBims1_OR [54]	100	(640, 480)	real	syn	incomplete	×	×
InteriorNet_OR [54]	10,000	(640, 480)	syn	syn	incomplete	×	×
our *OB-FUTURE*	19,186	(1080, 1080)	syn	syn	**complete**	✓	✓
our *OB-DIODE*	50	(1024, 768)	real	manual	**complete**	✓	✓
our *OB-EntitySeg*	70	(982, 882)	real	manual	**complete**	✓	✓

Comparison of OB generation quality: our *Mesh2OB* vs. previous 2D depth-based OB generation method P2ORM. *Mesh2OB*'s result: the union of those OBs in white & red; P2ORM's result: only those OBs in white.

Mesh2OB: first 3D mesh–based OB generation (handles self-occlusions)
OB-FUTURE: 19k+ high-quality synthetic indoor scenes with high-quality OB annotations

Method Analysis and Ablation Study

Real user annotation time & quality of *MS³PE*

Dataset	Time cost↓ (mins)	ODS↑	OIS↑	AP↑
OB-DIODE	**5±1**	**88.8**	**89.2**	**90.7**
OB-EntitySeg	**6±1.5**	**84.8**	**86.1**	**86.8**

Manual annotation — 6–10× slower and far more meticulous than *MS³PE*-based annotation

Analysis on the component benefits of *MS³PE*

InP: **In**teraction **P**ath
ImP: **Im**age **P**ath

†: the corresponding model trained using the conventional **I**nteractive **S**egmentation **I**nteraction **M**echanism (ISIM)

InP	ImP	FEM	Interaction	ODS↑	OIS↑	AP↑
×	×	×	scribbles	48.8	50.2	41.8
×	✓	×	scribbles	76.1	78.1	80.6
✓	✓	×	scribbles	72.9	75.4	77.4
×	✓	✓	scribbles	**80.0**	**81.2**	**85.4**
✓	✓	✓	clicks †	46.1	52.8	39.5
✓	✓	✓	scribbles †	43.5	52.0	36.7
✓	✓	✓	clicks	71.4	73.4	73.9
✓	✓	✓	scribbles	<u>**81.3**</u>	<u>**82.5**</u>	<u>**86.4**</u>

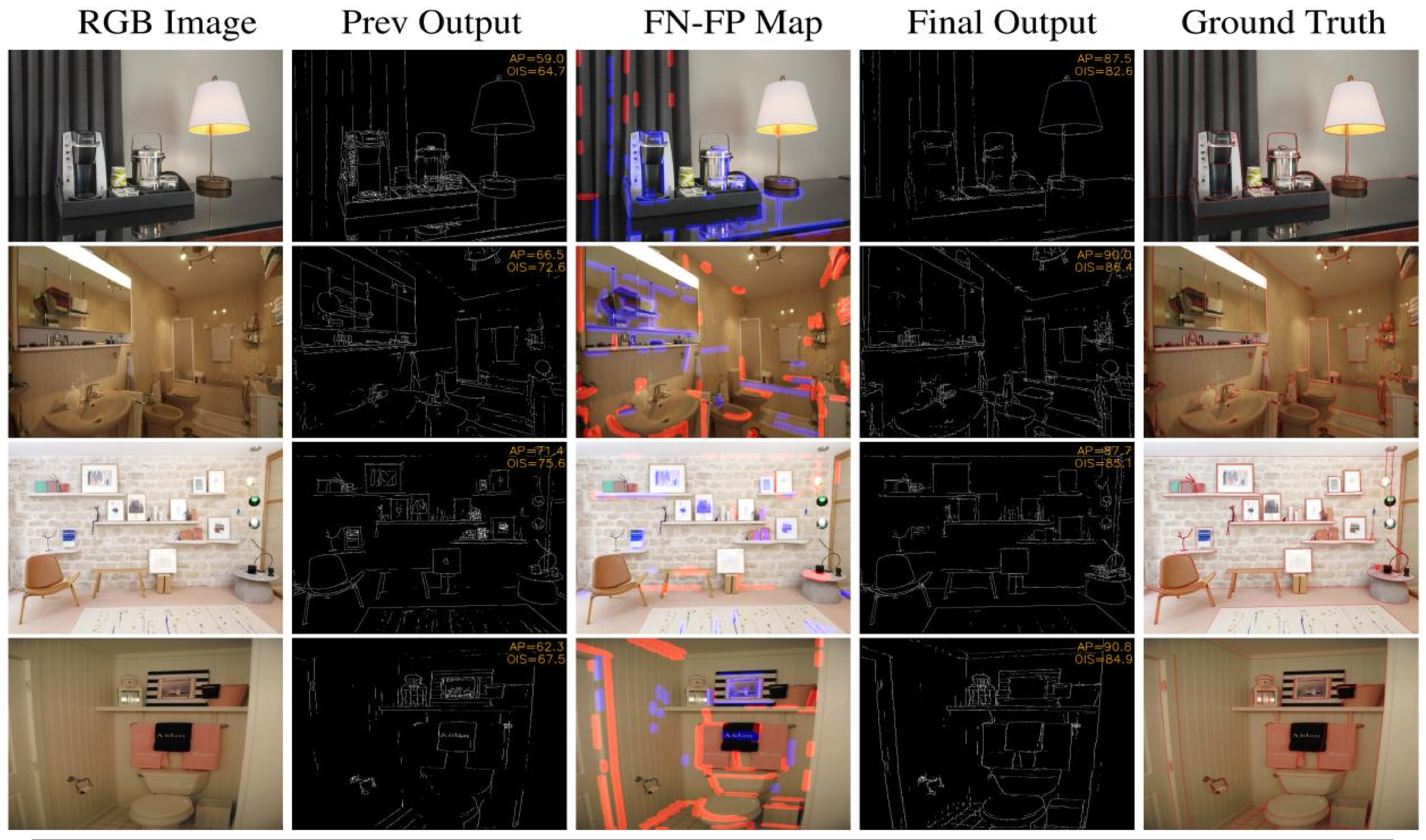
Feature heat maps before & after *FEM*

Visualized channel-wise averaged feature maps (fusion path output & *FEM*-enhanced outputs) overlaid on images

Resources

- *More experimental details in the Supplementary Materials*
- *Code & data: https://github.com/xul-ops/IOBE*
- *MoDOT*: https://arxiv.org/abs/2505.21231

Qualitative & Quantitative Results



Method	*OB-DIODE*			*OB-EntitySeg*		
	ODS↑	OIS↑	AP↑	ODS↑	OIS↑	AP↑
OPNet [13]	72.8	74.0	61.8	50.7	57.7	36.8
RITM [61] ‡	61.9	**70.3**	55.6	24.6	30.1	11.3
CDNet [7] ‡	35.1	44.5	22.2	34.3	38.3	24.3
TOS-Net [38] ‡	58.4	68.7	46.1	**57.5**	**60.3**	**50.4**
FCA-Net [40] ‡	58.9	66.5	55.7	49.7	52.2	40.3
FocalClick [8] ‡	60.6	68.9	**59.4**	47.6	51.8	44.5
AdaptiveClick [39] ‡	61.3	69.1	55.8	47.8	49.5	39.2
SimpleClick [41] ‡	**64.4**	69.4	59.3	50.7	54.3	44.0
RITM [61]	75.0	78.0	76.0	32.8	35.1	13.7
CDNet [7]	67.6	71.2	65.6	43.2	50.7	35.3
TOS-Net [38]	74.8	81.8	60.4	69.0	70.9	54.9
FCA-Net [40]	**77.1**	**82.6**	**81.9**	**71.8**	**75.2**	**73.7**
FocalClick [8]	73.9	79.0	76.9	63.3	69.0	68.6
AdaptiveClick [39]	75.8	79.8	79.0	66.2	69.7	64.2
SimpleClick [41]	73.6	78.5	78.2	69.5	71.4	70.0
Ours Swin-S	84.2	87.0	89.4	81.2	82.1	<u>**86.4**</u>
Ours Swin-B	85.5	<u>**87.9**</u>	90.0	<u>**81.3**</u>	<u>**82.5**</u>	<u>**86.4**</u>
Ours Swin-L	<u>**86.3**</u>	87.8	<u>**91.4**</u>	79.3	80.9	84.8

Second block: The results (marked with ‡) of interactive segmentation methods trained/tested using the conventional ISIM with boundary click.

Third block: The performance of the same interactive segmentation methods trained with our proposed *MSIM*.