

# Supplementary Material for the Paper: RINDNet: Edge Detection for Discontinuity in Reflectance, Illumination, Normal and Depth

Mengyang Pu<sup>1\*</sup>, Yaping Huang<sup>1,†</sup>, Qingji Guan<sup>1</sup>, Haibin Ling<sup>2</sup>

<sup>1</sup>Beijing Key Laboratory of Traffic Data Analysis and Mining, Beijing Jiaotong University, China

<sup>2</sup>Department of Computer Science, Stony Brook University, USA

{mengyangpu, yphuang, qjguan}@bjtu.edu.cn; hling@cs.stonybrook.edu

In this supplementary material, we provide additional details, including our benchmark (BSDS-RIND), experiment analysis, visualization results, applications of different types of edges, and additional analysis. We discuss:

- Details of BSDS-RIND benchmark (§A);
- More experiments analysis on features of different layers, spatial cues, Decoder and different backbone (§B);
- More qualitative results on BSDS-RIND (§C);
- Experiments on other datasets (§D);
- Applications of different types of edges (§E).

## A. BSDS-RIND

### A.1. Type distribution on BSDS-RIND

BSDS-RIND is split into training and testing set, containing 300 and 200 images respectively. We report the number of pixels for each type on the BSDS-RIND training set and testing set, as shown in Fig. 1 and Table 2 (row 1 and row 2).

### A.2. Compared with original BSDS

BSDS-RIND is created by carefully labeling images from BSDS [1]. The number of edge pixels annotated in BSDS-RIND are twice as that in the original BSDS [1], as shown in rows 3 and 4 of the Table 2. Since edge detection is a pixel-wise task, the increased number of samples in BSDS-RIND brings strong supports for learning-based algorithms.

Fig. 2 provides some comparisons of edge annotations for BSDS-RIND and BSDS [1]. The first row (a2 – a6) and the third row (b2 – b6) show the ground-truth annotations given by multiple users for BSDS [1]; the second row (a7 – a12) and the fourth row (b7 – b12) show four types of

\*The work is partially done while the author was at Stony Brook University. † Corresponding author.

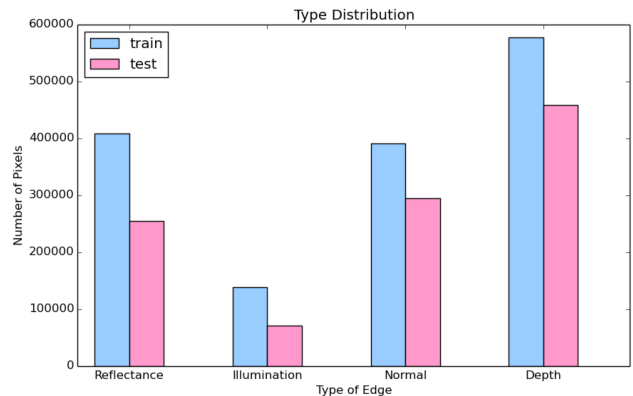


Figure 1. The distribution of pixels for each type of edges on BSDS-RIND training set and testing set.

Table 1. The average percentage of pixels labeled as edge or boundary in SBD, PASCAL Boundaries, NYUD, BSDS and our BSDS-RIND datasets.

Dataset	SBD	PASCAL	NYUD	BSDS	BSDS-RIND
Ave. Pixels	0.77%	1.45%	2.17%	1.81%	3.36%

\* Numbers of SBD, PASCAL and BSDS are from [4].

edges and all edges annotated by us. We can observe that our annotations are more dense and accurate than BSDS, making BSDS-RIND more challenging for edge detection.

We also summarize the percentage of pixels labeled as edges in several datasets presented in Table 1. It shows that more edge pixels per image (3.36%) are annotated in BSDS-RIND.

## B. More Details on Ablation Study

In Sec. 5.1, we explain the ablation studies in the proposed RINDNet and only give the average score due to space constraints. Therefore, we present more comprehensive results in this section for ablation studies.

Table 2. The number of pixels for each type of edge on BSDS-RIND.

Set \ Edge Type	Reflectance Edge	Illumination Edge	Normal Edge	Depth Edge	All Edge
Training Set (300 images)	408,128	138,994	390,980	577,618	1,515,720
Testing Set (200 images)	254,911	71,717	294,444	458,552	1,079,624
BSDS-RIND (500 images)	663,039	210,711	685,424	1,036,170	2,595,344
BSDS (500 images)	–	–	–	–	1,395,744

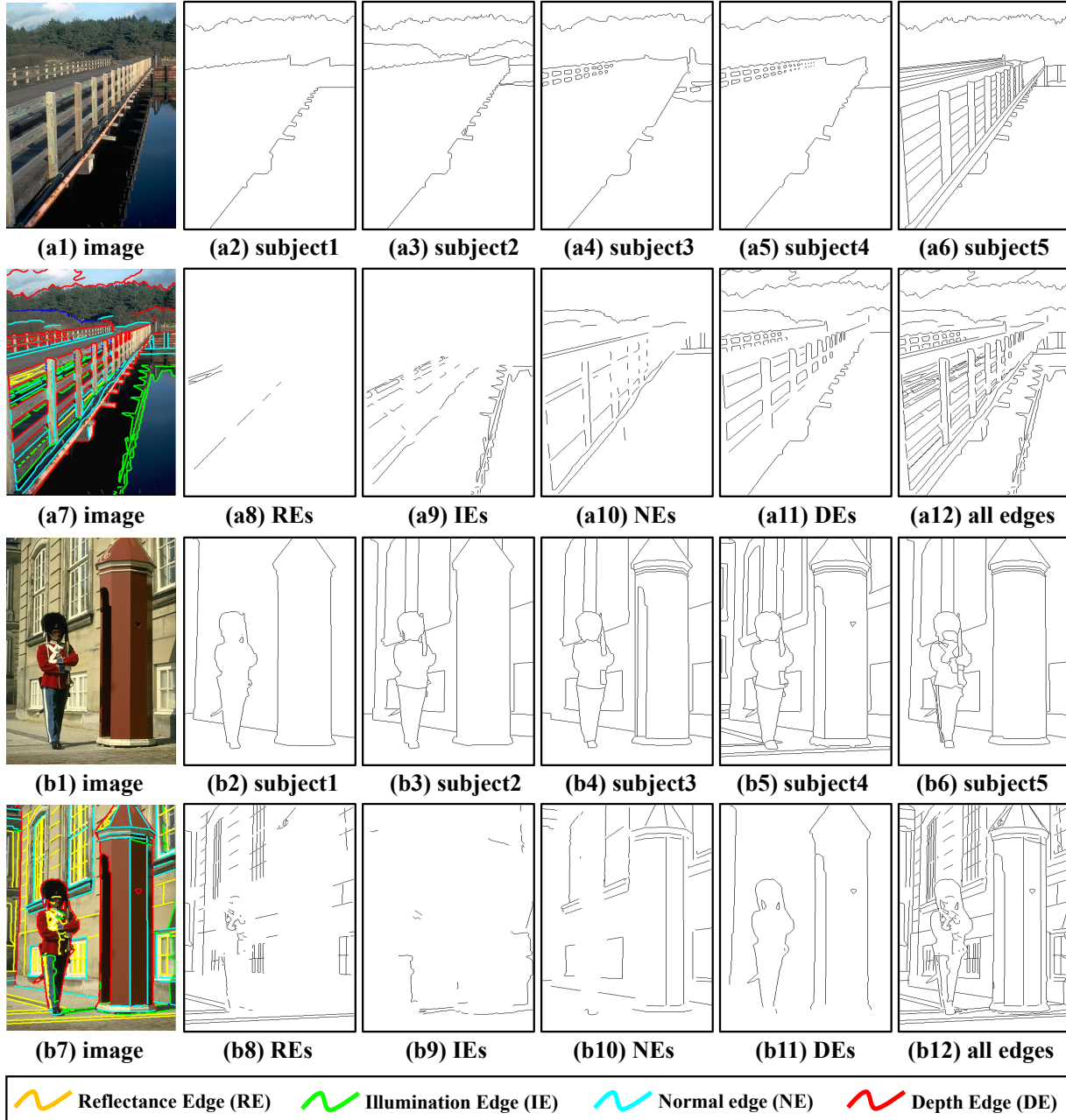


Figure 2. The examples of BSDS and BSDS-RIND (ours), (a1) and (b1) are the images from the BSDS benchmark; (a2 – a6) and (b2 – b6) show the ground-truth annotations of all edges annotated by multiple users for BSDS; (a7) and (b7) are the annotation examples from the proposed BSDS-RIND, (a7 – a11) and (b7 – b11) are four types of edges annotations from BSDS-RIND, and (a12) and (b12) are the ground-truth annotations of all edges from BSDS-RIND.

Table 3. Ablation study on the choices of features from different layers for the proposed RINDNet.

R&I	N&D	Reflectance			Illumination			Normal			Depth			Average		
		ODS	OIS	AP	ODS	OIS	AP	ODS	OIS	AP	ODS	OIS	AP	ODS	OIS	AP
$res_{1-3}$	$res_5$	0.478	0.521	0.414	0.280	0.337	0.168	0.489	0.522	0.440	0.697	0.724	0.705	0.486	0.526	0.432
$res_{1-3}$	$res_{1-3}$	0.472	0.522	0.415	0.278	0.294	0.181	0.459	0.495	0.408	0.657	0.683	0.674	0.467	0.499	0.422
$res_5$	$res_5$	0.463	0.491	0.397	0.223	0.245	0.138	0.481	0.513	0.354	0.687	0.708	0.695	0.464	0.489	0.396
$res_5$	$res_{1-3}$	0.448	0.490	0.387	0.225	0.233	0.132	0.467	0.511	0.313	0.669	0.694	0.690	0.452	0.482	0.381

Table 4. Ablation study on the spatial cues from different layers for the proposed RINDNet.

R&I	N&D	Reflectance			Illumination			Normal			Depth			Average		
		ODS	OIS	AP	ODS	OIS	AP	ODS	OIS	AP	ODS	OIS	AP	ODS	OIS	AP
$f_{sp}^{1-3}$	$f_{sp}^{1-5}$	0.478	0.521	0.414	0.280	0.337	0.168	0.489	0.522	0.440	0.697	0.724	0.705	0.486	0.526	0.432
$f_{sp}^{1-3}$	$f_{sp}^{1-3}$	0.478	0.521	0.421	0.274	0.303	0.167	0.466	0.506	0.414	0.668	0.686	0.662	0.472	0.504	0.416
$f_{sp}^{1-5}$	$f_{sp}^{1-5}$	0.466	0.519	0.394	0.285	0.317	0.190	0.475	0.501	0.388	0.687	0.709	0.698	0.478	0.512	0.418
w/o $f_{sp}$	w/o $f_{sp}$	0.472	0.521	0.401	0.283	0.323	0.191	0.478	0.515	0.401	0.682	0.705	0.686	0.478	0.516	0.420

### B.1. Ablation study on features of different layers

Our method is motivated by many empirically validated generic edge detectors. For instance, HED [9] and RCF [3] predict edge maps on different scales of CNN features. Similarly, in our work, REs and IEs are mainly due to photometric reasons (lower layer features  $res_{1-3}$ ), while NEs and DEs reflect geometry changes in object surfaces or depth discontinuity (higher layer features  $res_5$ ). We perform an experiment to verify the effectiveness of features of different layers for detecting Reflectance Edges and Illumination Edges (REs/IEs) and Normal Edges and Depth Edges (NEs/DEs).

The quantitative results are reported in Table 3. As shown in the first row, choosing  $res_{1-3}$  for REs/IEs and  $res_5$  for NEs/DEs achieves the best scores in terms of all metrics. However, if we use the low-level features  $res_{1-3}$  instead of the high-level features  $res_5$  to infer NEs and DEs (row 2), the scores of NEs decrease by 3.0%, 2.7% and 3.2% in terms of ODS, OIS and AP, and the performances of DEs also decline by 4.0%, 4.1% and 3.1% accordingly. The performance drop suggests the effectiveness of high-level features for detecting NEs and DEs. Besides, as shown in the third row of Table 3, the performance decreases when using  $res_5$  to reason REs and IEs, showing that the low-level features are beneficial to infer REs and IEs. Finally, we utilize  $res_5$  for REs/IEs and  $res_{1-3}$  for NEs/DEs, as shown in row 4, the scores for all metric terms are much lower than our best choices (row 1). The above experimental results demonstrate that it is importance to choose appropriate features for detecting different types of edges, according to the attributes of different edges.

The high-level tasks mainly work on top-layers where features encode high-level information, which are commonly used in other methods (e.g., [11]). Additionally, we test to include  $res_4$  but observe little performance change

and dramatically increase computational burden. Therefore,  $res_4$  is only used to generate spatial cues for NEs/DEs.

### B.2. Ablation study on different spatial cues

We also explore the impact of choosing different spatial cues for REs/IEs and NEs/DEs, and we report the quantitative results in Table 4. From the first and last rows, we observe that spatial cues are important and bring performance improvement in terms of all metrics. Especially, the performance of NEs and DEs drops significantly without using spatial cues  $f_{sp}^{4-5}$  in row 2. This is due to the fact that lacking the spatial cues of objects is harmful for predicting NEs and DEs. As shown in row 3,  $f_{sp}^{4-5}$  is used for REs and IEs, which leads the performance improvement for the ODS of IEs. However, NEs and DEs could influence the parameters of Spatial Layers in the training process, which results in the performance drop of NEs and DEs. Especially the average scores drop by 0.8%, 1.4% and 1.4% in three metrics, respectively.

### B.3. Ablation study on Decoder

We also carefully design ablation experiments (Table 5) to study the effectiveness of each stream in Decoder. The two-stream design of Decoder combining with share weights (referred to as SW) for NEs/DEs performs better together (row 1) than using either of them separately (rows 3 – 4). More specifically, without ‘‘Share Weight’’ (row 4), the scores of ODS drop 2.1% and 1.7% for NEs and DEs, respectively. Besides, using only the first stream of RE-Decoder and IE-Decoder (row 3), the performance on almost all types of edges is also reduced. Moreover, as shown in row 2, if we remove the second stream and only adopt one stream in each Decoder, the average scores are the worst among four combinations.

Table 5. Ablation study to verify the effectiveness of Decoder for the proposed RINDNet. SW refers to ‘‘Share Weight’’, 1<sup>st</sup> and 2<sup>nd</sup> refer to the first stream and the second stream in Decoder.

RE&IE-Decoder		NE&DE-Decoder		Reflectance			Illumination			Normal			Depth			Average		
1 <sup>st</sup>	2 <sup>nd</sup>	1 <sup>st</sup>	2 <sup>nd</sup>	ODS	OIS	AP	ODS	OIS	AP	ODS	OIS	AP	ODS	OIS	AP	ODS	OIS	AP
✓	✓	✓	w/ SW	0.478	0.521	0.414	0.280	0.337	0.168	0.489	0.522	0.440	0.697	0.724	0.705	0.486	0.526	0.432
✓	×	✓	×	0.480	0.521	0.422	0.268	0.318	0.162	0.437	0.467	0.364	0.641	0.661	0.643	0.457	0.492	0.398
✓	×	✓	w/ SW	0.471	0.528	0.414	0.277	0.317	0.176	0.476	0.508	0.380	0.681	0.704	0.688	0.476	0.514	0.415
✓	✓	✓	w/o SW	0.466	0.515	0.386	0.283	0.336	0.176	0.468	0.515	0.386	0.680	0.702	0.684	0.474	0.517	0.408

Table 6. The ablations with multiple backbone networks for the proposed RINDNet.

Backbone	Reflectance			Illumination			Normal			Depth			Average		
	ODS	OIS	AP	ODS	OIS	AP	ODS	OIS	AP	ODS	OIS	AP	ODS	OIS	AP
ResNet-50	0.478	0.521	0.414	0.280	0.337	0.168	0.489	0.522	0.440	0.697	0.724	0.705	0.486	0.526	0.432
ResNet-101	0.484	0.538	0.404	0.291	0.332	0.191	0.483	0.515	0.390	0.691	0.716	0.679	0.487	0.526	0.416
ResNeXt-50	0.468	0.521	0.385	0.265	0.304	0.162	0.474	0.514	0.348	0.678	0.702	0.672	0.471	0.510	0.391
ResNeXt-101	0.491	0.540	0.406	0.285	0.328	0.181	0.492	0.530	0.403	0.701	0.723	0.693	0.492	0.530	0.420

#### B.4. Ablation study with different backbones

We build several variants of RINDNet by adapting different backbone nets, such as ResNet-101, ResNeXt-50 and ResNeXt-101, achieving better performance compared to our original model based on ResNet-50. As shown in Table 6, adopting ResNeXt-101 instead of ResNet-50 significantly improves RINDNet performance, *e.g.*, our ResNeXt-101 based model attains 49.2% and 53.0% in terms of ODS and OIS, compared to 48.6% and 52.6% of our RINDNet ResNet-50 based variant.

### C. Qualitative Results

In this section, we report qualitative results of HED, CED, RCF, DexiNed, BDCN, CASENet, DFF, DeepLabV3+, DOOBNet, OFNet and proposed RINDNet on BSDS-RIND. Fig. 4 shows the visual comparisons for reflectance edges; Fig. 5 depicts qualitative results for illumination edges; Fig. 6 presents some qualitative examples of normal edges; Fig. 7 illustrates the visual examples of depth edges; and qualitative comparisons for generic edges are demonstrated in Fig. 8.

Moreover, in order to intuitively compare the prediction results made across different edge types on an image, we present the visualizations of all types of edges on the same images for our RINDNet in Fig. 9 and Fig. 10.

### D. Experiments on other Datasets

#### D.1. Results on the original BSDS dataset.

To verify the general ability of the proposed RINDNet, we train and test RINDNet on the original BSDS dataset. The results are reported in Table 7, where RINDNet performs competitive with state-of-the-art methods.

Table 7. Results on the original BSDS dataset.

Method	ODS	OIS	AP
HED	0.788	0.808	0.840
CED	0.794	0.811	0.847
RCF	0.798	0.815	-
BDCN	0.806	0.826	0.847
RINDNet (Ours)	0.805	0.821	0.813

#### D.2. Performance on PASCAL VOC and NYUD.

Since the proposed network is trained using the finer annotations provided by BSDS-RIND, we evaluate our RINDNet on other edge detection datasets, such as PASCAL VOC and NYUD, to verify the scalability of our method. The average scores in ODS, OIS, AP are reported in Table 8.

Table 8. The results of RINDNet on the PASCAL VOC dataset and the NYUD dataset.

Method	ODS	OIS	AP
PASCAL VOC	0.577	0.607	0.497
NYUD	0.647	0.662	0.568

### E. Applications of Different Types of Edges

Since edges play an important role in many vision tasks, many recent applications that rely directly on these fine-grained edges have shown promising results. For example, reflectance edges (*e.g.*, crack) and illumination edges (*e.g.*, shadow edges) provide hints for the intelligent transportation [10] or path detection [8]. Besides, depth edges and normal edges can also benefit depth estimation [5, 7]. We design two experiments on depth estimation and shadow removal to demonstrate that the fine-grained edges can help downstream tasks.

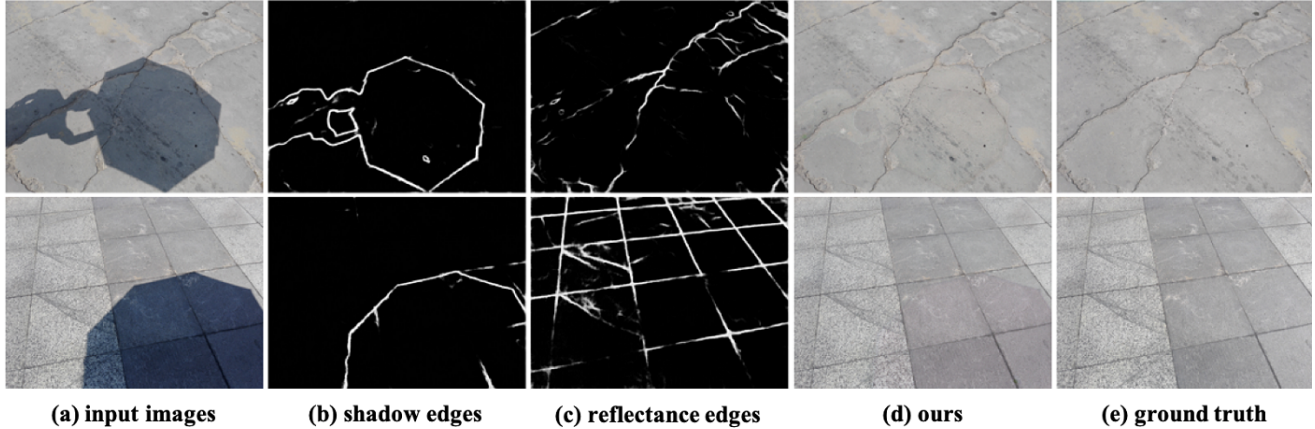


Figure 3. Qualitative results for **shadow removal**. (a) input image, (b) and (c) are shadow edges and reflectance edges detected from the proposed RINDNet that trained on BSDS-RIND, (d) ours shadow removal result, and (e) ground truth.

Table 9. Evaluation of **depth estimation** on NYUD.

Method	<i>higher is better</i>			<i>lower is better</i>			
	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$	Abs Rel	Sq Rel	RMSE	RMSE log
BTS (baseline)	0.685	0.921	0.982	0.196	0.164	0.640	0.227
BTS-RINDNet (ours)	0.717	0.931	0.983	0.188	0.148	0.597	0.216

In the **depth estimation** experiment, we train and evaluate our model on the NYUD dataset [6]. Besides, we adopt BTS [2] as the baseline model, since it shows outstanding accuracy on depth estimation. Note that, different from the original BTS [2], we use labeled dataset instead of raw dataset, because labeled data has been preprocessed to fill in missing depth labels. We follow the official train and test split as NYUD [6], using 796 images for training and 654 images for testing. The quantitative results are reported in Table 9. We first train and evaluate BTS [2] model on the labeled data, as shown in row 1.

In order to improve the depth estimation performance utilizing the depth edge and normal edge cues, we combine our RINDNet and BTS in a simple way: we first extract feature representations of depth edges and normal edges from RINDNet that trained on BSDS-RIND, and then the feature representations are concatenated with the encoded features generated by the Encoder of BTS [2]. Afterward, the concatenated features are fed directly into the Decoder of BTS [2] to estimate depth maps. Our method achieves promising results with a significant improvement in both of the inlier measures (*i.e.*,  $\delta_1$ ,  $\delta_2$ ,  $\delta_3$ ) and accuracy metrics (*i.e.*, Abs Rel, Sq Rel, RMSE, RMSE log).

We also show examples of **shadow removal** in Fig. 3, in which our shadow edge (illumination edge) results and reflectance edge maps are applied to improve the shadow removal. Moreover, Fig. 3 (c) shows the reflectance edges detected from the proposed RINDNet (only trained on BSDS-RIND), which correspond to cracks or patterns on the

ground. We can observe that, in the process of shadow removal, detected reflectance edges can be applied to appearance constraints that encourage the consistency of appearance between the non-shadow and the shadow regions.

In the future, we will further explore these tasks. At the same time, we expect these fine-grained edges to help facilitate more related studies.

## References

- [1] Jianzhong He, Shiliang Zhang, Ming Yang, Yanhu Shan, and Tiejun Huang. Bi-directional cascade network for perceptual edge detection. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3828–3837, 2019. 1
- [2] Jin Han Lee, Myung-Kyu Han, Dong Wook Ko, and Il Hong Suh. From big to small: Multi-scale local planar guidance for monocular depth estimation. *arXiv preprint arXiv:1907.10326*, 2019. 5
- [3] Yun Liu, Ming-Ming Cheng, Xiaowei Hu, Kai Wang, and Xiang Bai. Richer convolutional features for edge detection. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3000–3009, 2017. 3
- [4] Vittal Premachandran, Boyan Bonev, Xiaochen Lian, and Alan Yuille. Pascal boundaries: A semantic boundary dataset with a deep semantic boundary detector. In *IEEE Winter Conf. Appl. Comput. Vis.*, pages 73–81, 2017. 1
- [5] Michaël Ramamonjisoa, Yuming Du, and Vincent Lepetit. Predicting sharp and accurate occlusion boundaries in monocular depth estimation using displacement fields. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 14648–14657, 2020. 4

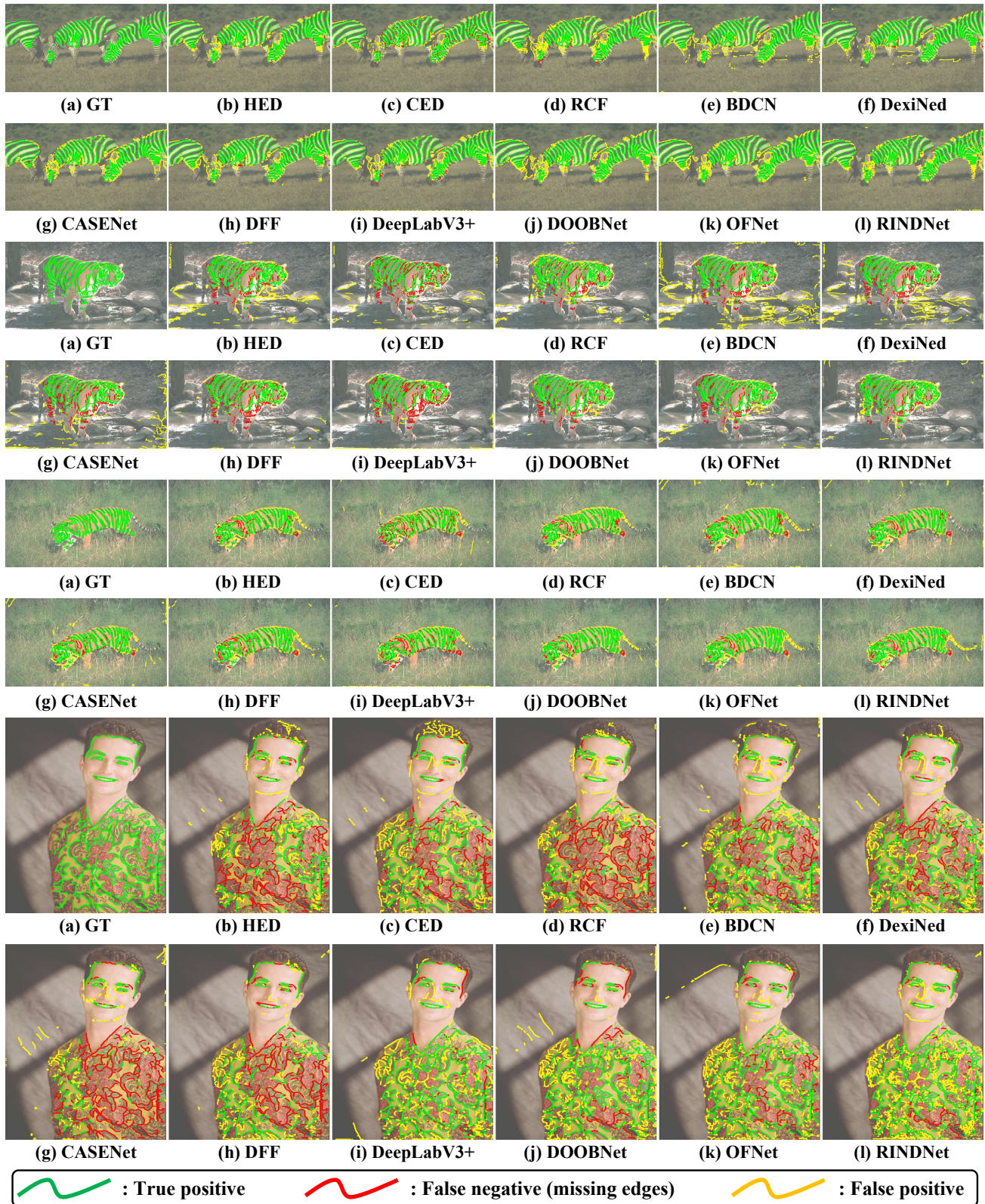


Figure 4. Qualitative comparison for **Reflectance Edges** (best viewed in color: “green” for true positive, “red” for false negative (missing edges), and “yellow” for false positive).



Figure 5. Qualitative comparison for **Illumination Edges** (best viewed in color: “green” for true positive, “red” for false negative (missing edges), and “yellow” for false positive).



Figure 6. Qualitative comparison for **Normal Edges** (best viewed in color: “green” for true positive, “red” for false negative (missing edges), and “yellow” for false positive).



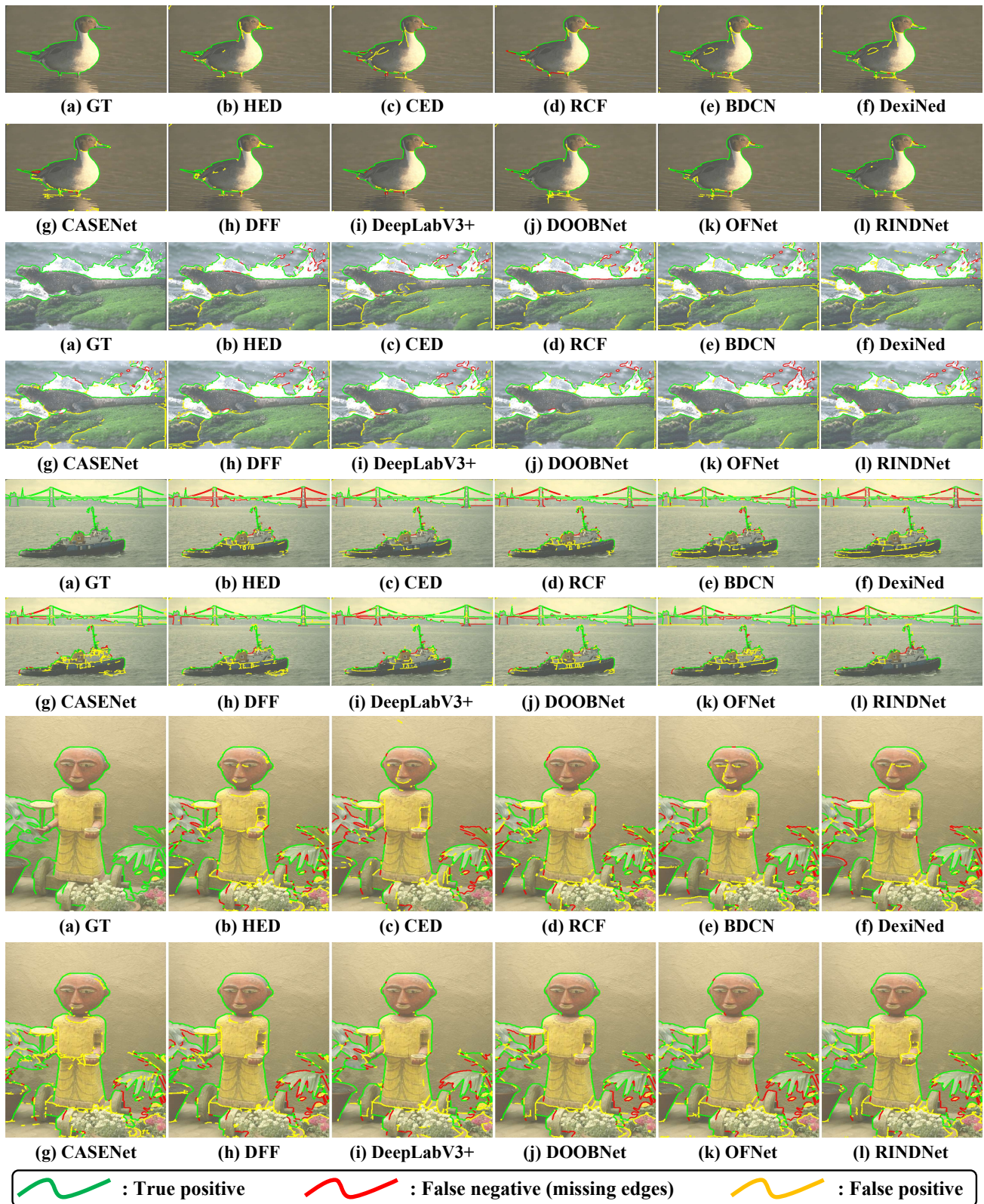


Figure 7. Qualitative comparison for **Depth Edges** (best viewed in color: “green” for true positive, “red” for false negative (missing edges), and “yellow” for false positive).



Figure 8. Qualitative comparison for **Generic Edges** (best viewed in color: “green” for true positive, “red” for false negative (missing edges), and “yellow” for false positive).

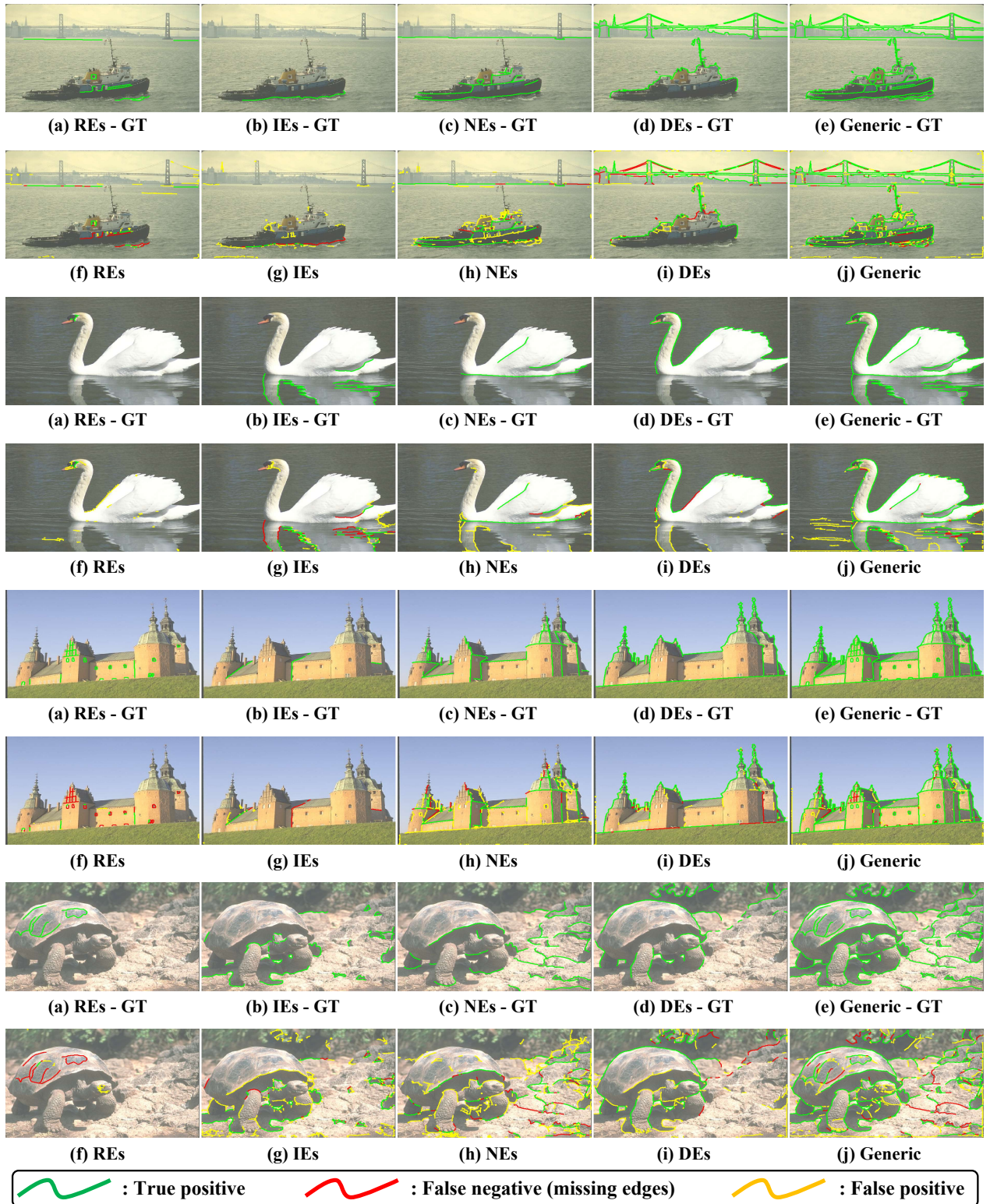


Figure 9. The visualizations of the proposed RINDNet on the same images, for Reflectance Edges (REs), Illumination Edges (IEs), Normal Edges (NEs), Depth Edges (DEs) and Generic Edges. Rows 1, 3, 5, and 7 are the ground-truth annotations including four types of edges and generic edges; Rows 2, 4, 6, and 8 are the corresponding predictions of our method. (best viewed in color: "green" for true positive, "red" for false negative (missing edges), and "yellow" for false positive)

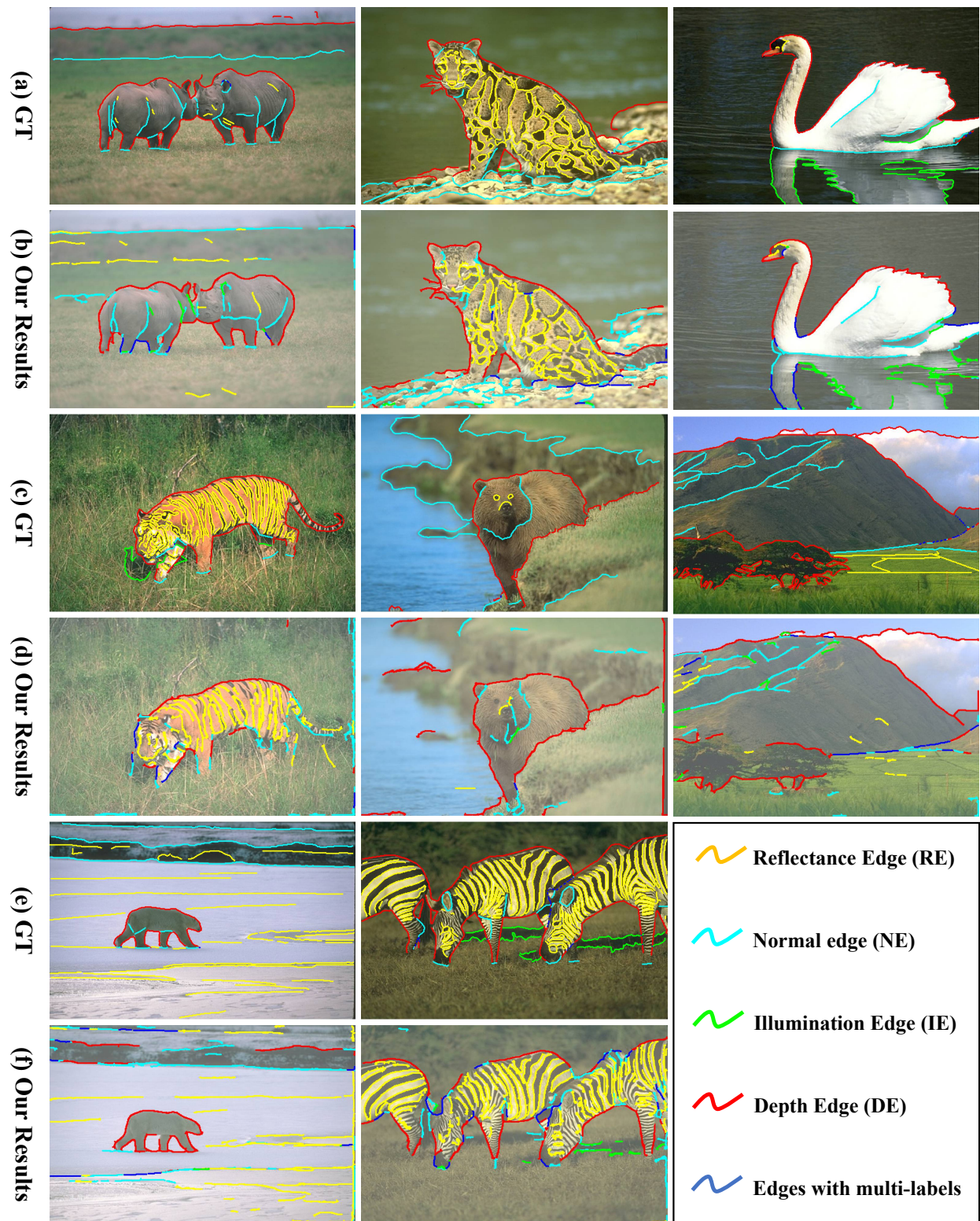


Figure 10. The visualizations of the results of our RINDNet. Rows (a), (c), and (e) are the ground-truth annotations including four types of edges; Rows (b), (d), and (f) are our results, in which the results of four types are visualized into one image. (best viewed in color: “yellow” for Reflectance Edges, “green” for Illumination Edges, “cyan” for Normal Edges, and “red” for Depth Edges. Moreover, some edges have multiple type labels in the ground-truth annotations or our predicted results, thus we use “blue” to indicate them.)

- [6] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgb-d images. In *Eur. Conf. Comput. Vis.*, pages 746–760, 2012. 5
- [7] Xiaolong Wang, David F. Fouhey, and Abhinav Gupta. Designing deep networks for surface normal estimation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 539–547, 2015. 4
- [8] Qi Wu, Wende Zhang, and B. V. K. Vijaya Kumar. Strong shadow removal via patch-based shadow edge detection. In *International Conference on Robotics and Automation*, pages 2177–2182, 2012. 4
- [9] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Int. Conf. Comput. Vis.*, pages 1395–1403, 2015. 3
- [10] Fan Yang, Lei Zhang, Sijia Yu, Danil Prokhorov, Xue Mei, and Haibin Ling. Feature pyramid and hierarchical boosting network for pavement crack detection. *IEEE Trans. Intell. Transp. Syst.*, 21(4):1525–1535, 2019. 4
- [11] Zhiding Yu, Chen Feng, Ming-Yu Liu, and Srikumar Ramalingam. Casenet: Deep category-aware semantic edge detection. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5964–5973, 2017. 3