EDTER: Edge Detection with Transformer
Supplementary Material

Mengyang Pu[1,3], Yaping Huang[1]*, Yuming Liu[2], Qingji Guan[1], Haibin Ling[3]
[1]Beijing Key Laboratory of Traffic Data Analysis and Mining, Beijing Jiaotong University, China
[2]Shenzhen Urban Transport Planning Center Co.,Ltd., China
[3]Department of Computer Science, Stony Brook University, USA
{mengyangpu, yphuang, qjguan}@bjtu.edu.cn; liuyuming@sutpc.com; hling@cs.stonybrook.edu

In this supplementary material, we provide additional details, including more experiment results on NYUDv2 [12] as well as visualization results on BSDS500 [1], NYUDv2 [12], and Multicue [11].

A. More Results on NYUDv2

In addition to RGB images, NYUDv2 [12] also provides depth images. Most works [2, 8, 9, 13, 15, 16] leverage the depth information to improve performance. Although our method is designed for RGB images, we still provide results of HHA and RGB-HHA. The depth information is encoded into three channels: horizontal disparity, height above ground, and angle with gravity [6]. Following previous works [2, 8, 9, 13, 15, 16], we regard the HHA feature as color images to train the model. The predictions of RGB-HHA are generated by averaging the outputs of the RGB model and the HHA model.

The quantitative results are shown in Table 1, and Fig. 1 shows Precision-Recall curves of most methods. The HHA features are generated from the depth images, thereby only reflecting geometric information. However, our model focuses on modeling context information in RGB images. Thus, EDTER can bring significant gains for RGB inputs, but the performance is slightly worse on HHA. For the RGB-HHA, EDTER still produces competitive results and achieves 78.0%, 79.7%, and 81.4% in terms of ODS, OIS, and AP, respectively.

B. Qualitative Results

In this section, we report qualitative results on BSDS500 [1], NYUDv2 [12], and Multicue [11]. Specifically, we provide more qualitative results of different stages in EDTER on BSDS500, as shown in Fig. 2. The predicted

*Corresponding author.

Table 1. Quantitative comparisons on NYUDv2 [12]. The best two results are highlighted in **red** and **blue**, respectively. All results are computed with a single scale input.

	Method	Pub.'Year	RGB			HHA			RGB-HHA		
			ODS	OIS	AP	ODS	OIS	AP	ODS	OIS	AP
Traditional	gPb-ucm [1]	PAMI'11	0.632	0.661	0.562	-	-	-	-	-	-
	Silberman *et al.* [12]	ECCV'12	0.658	0.661	-	-	-	-	-	-	-
	gPb+NG [4]	CVPR'13	0.687	0.716	0.629	-	-	-	-	-	-
	SE [3]	PAMI'14	0.695	0.708	0.679	-	-	-	-	-	-
	SE+NG+ [5]	ECCV'14	0.706	0.734	0.738	-	-	-	-	-	-
	OEF [7]	CVPR'15	0.651	0.667	-	-	-	-	-	-	-
	SemiContour [17]	CVPR'16	0.680	0.700	0.690	-	-	-	-	-	-
CNN-based	HED [15]	ICCV'15	0.720	0.734	0.734	0.682	0.695	0.702	0.746	0.761	0.786
	COB [10]	ECCV'16	-	-	-	-	-	-	**0.784**	**0.805**	**0.825**
	RCF [9]	CVPR'17	0.729	0.742	-	0.705	0.715	-	0.757	0.771	-
	AMH-Net [16]	NeurIPS'17	0.744	0.758	0.765	**0.717**	**0.729**	**0.734**	0.771	0.786	0.802
	LPCB [2]	ECCV'18	0.739	0.754	-	0.707	0.719	-	0.762	0.778	-
	BDCN [8]	CVPR'19	**0.748**	**0.763**	**0.770**	0.707	0.719	**0.731**	0.765	0.781	0.813
	PiDiNet [13]	ICCV'21	0.733	0.747	-	**0.715**	**0.728**	-	0.756	0.773	0.813
	EDTER(Ours)	-	**0.774**	**0.789**	**0.797**	0.703	0.718	0.727	**0.780**	**0.797**	**0.814**

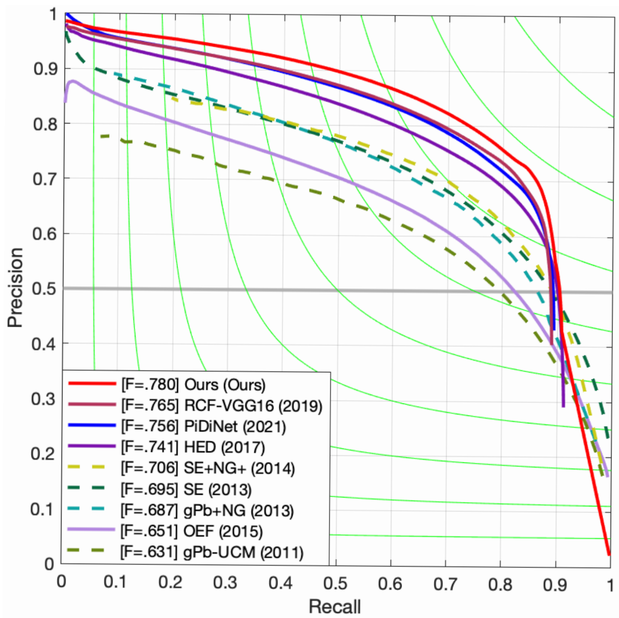


Figure 1. The precision-recall curves on NYUDv2.

edges of Stage II are more clear and crisp, which shows the effectiveness of the two-stage strategy for edge detection. In Fig. 3, we also present more visualizations of the proposed EDTER on BSDS500 [1]. Moreover, Fig. 4 shows the visual comparisons for BSDS500 [1]; Fig. 5 illustrates our visual examples on NYUDv2 [12]; and Fig. 6 depicts qualitative results for Multicue Edge and Multicue Boundary [11].

References

[1] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):898–916, 2010. 1, 2

[2] Ruoxi Deng, Chunhua Shen, Shengjun Liu, Huibing Wang, and Xinru Liu. Learning to predict crisp boundaries. In *Eur. Conf. Comput. Vis.*, pages 562–578, 2018. 1

[3] Piotr Dollár and C Lawrence Zitnick. Fast edge detection using structured forests. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(8):1558–1570, 2014. 1

[4] Saurabh Gupta, Pablo Arbelaez, and Jitendra Malik. Perceptual organization and recognition of indoor scenes from rgb-d images. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 564–571, 2013. 1

[5] Saurabh Gupta, Ross Girshick, Pablo Arbeláez, and Jitendra Malik. Learning rich features from rgb-d images for object detection and segmentation. In *Eur. Conf. Comput. Vis.*, pages 345–360. Springer, 2014. 1

[6] Saurabh Gupta, Ross B. Girshick, Pablo Andrés Arbeláez, and Jitendra Malik. Learning rich features from RGB-D images for object detection and segmentation. In *Eur. Conf. Comput. Vis.*, pages 345–360, 2014. 1

[7] Sam Hallman and Charless C Fowlkes. Oriented edge forests for boundary detection. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1732–1740, 2015. 1

[8] 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, 5

[9] 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. 1, 5

[10] Kevis-Kokitsi Maninis, Jordi Pont-Tuset, Pablo Arbeláez, and Luc Van Gool. Convolutional oriented boundaries. In *Eur. Conf. Comput. Vis.*, pages 580–596. Springer, 2016. 1

[11] David A Mély, Junkyung Kim, Mason McGill, Yuliang Guo, and Thomas Serre. A systematic comparison between visual cues for boundary detection. *Vis. Res.*, 120:93–107, 2016. 1, 2

[12] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *Eur. Conf. Comput. Vis.*, pages 746–760, 2012. 1, 2

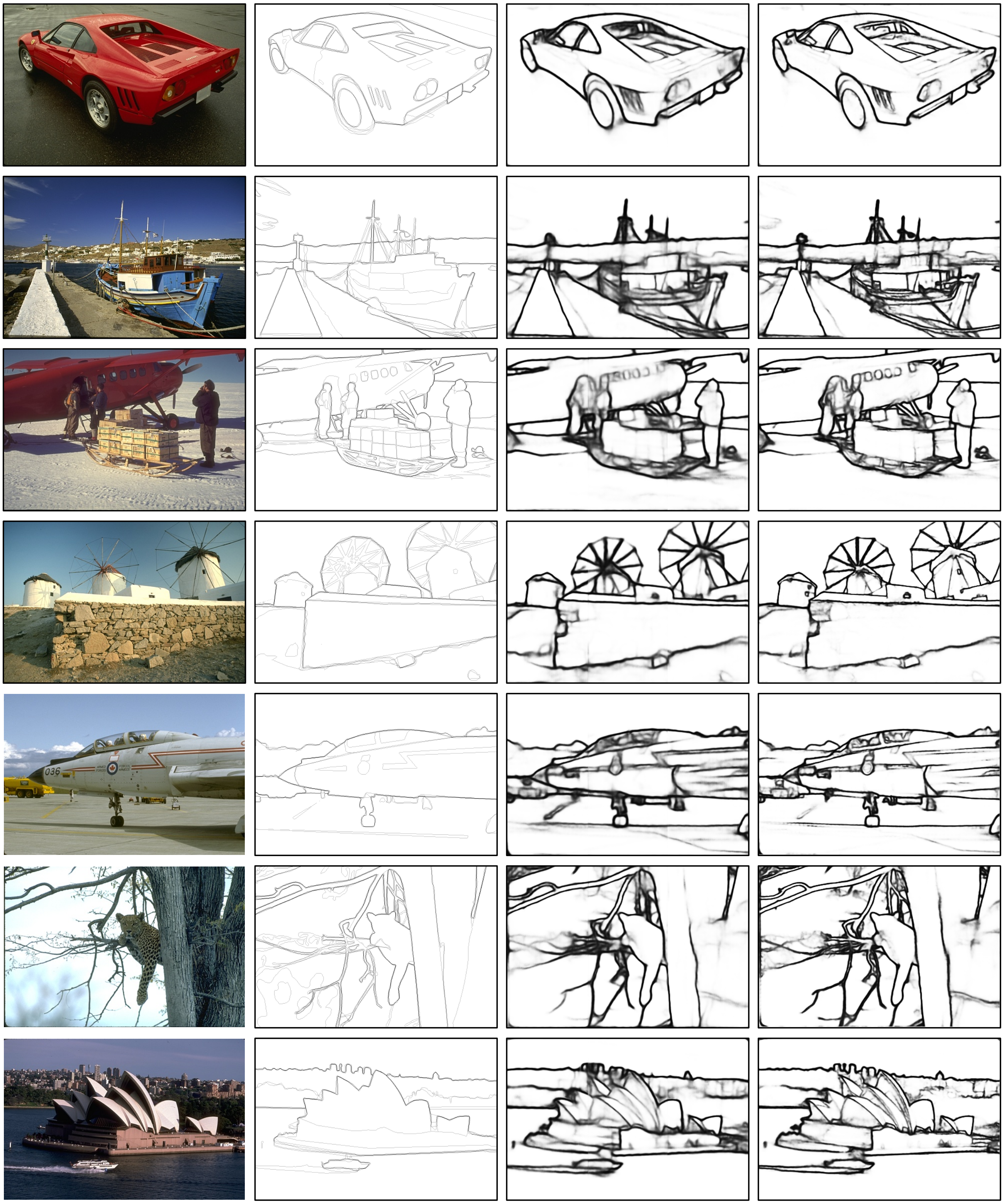
[13] Zhuo Su, Wenzhe Liu, Zitong Yu, Dewen Hu, Qing Liao, Qi Tian, Matti Pietikainen, and Li Liu. Pixel difference networks for efficient edge detection. In *Int. Conf. Comput. Vis.*, pages 5117–5127, October 2021. 1

[14] Yupei Wang, Xin Zhao, and Kaiqi Huang. Deep crisp boundaries. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3892–3900, 2017. 5

[15] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Int. Conf. Comput. Vis.*, pages 1395–1403, 2015. 1

[16] Dan Xu, Wanli Ouyang, Xavier Alameda-Pineda, Elisa Ricci, Xiaogang Wang, and Nicu Sebe. Learning deep structured multi-scale features using attention-gated crfs for contour prediction. In *Adv. Neural Inform. Process. Syst.*, pages 3961–3970, 2017. 1

[17] Zizhao Zhang, Fuyong Xing, Xiaoshuang Shi, and Lin Yang. Semicontour: A semi-supervised learning approach for contour detection. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 251–259, 2016. 1



(a) Image (b) Ground truth (c) EDTER-Stage I (d) EDTER-Stage II

Figure 2. Qualitative comparison of different stages in EDTER on BSDS500. From left to right are the input images, the results of EDTER-Stage I and EDTER-Stage II, respectively.

(a) Image (b) Ground truth (c) DETER(Ours)

Figure 3. Qualitative results of EDTER on BSDS500.

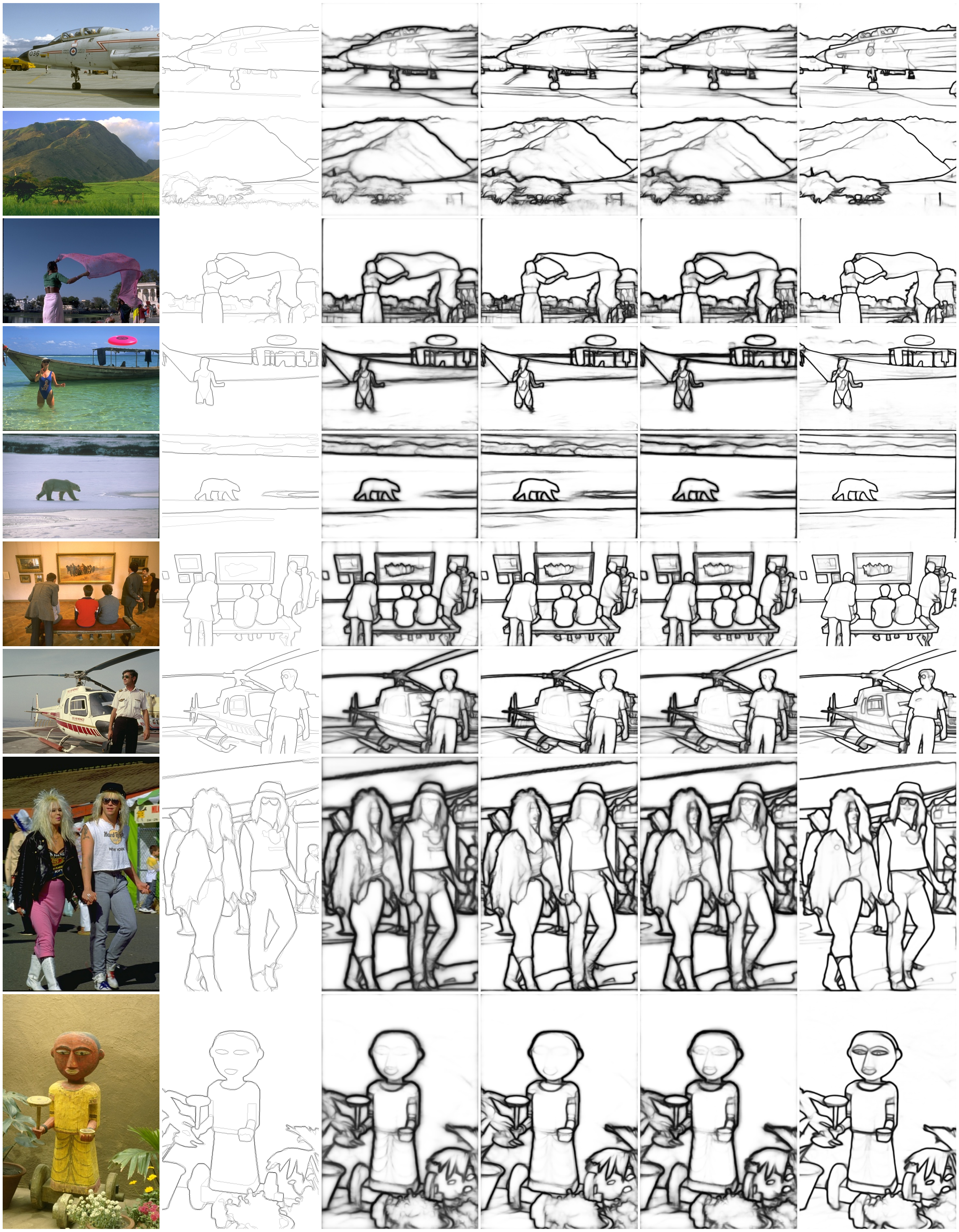
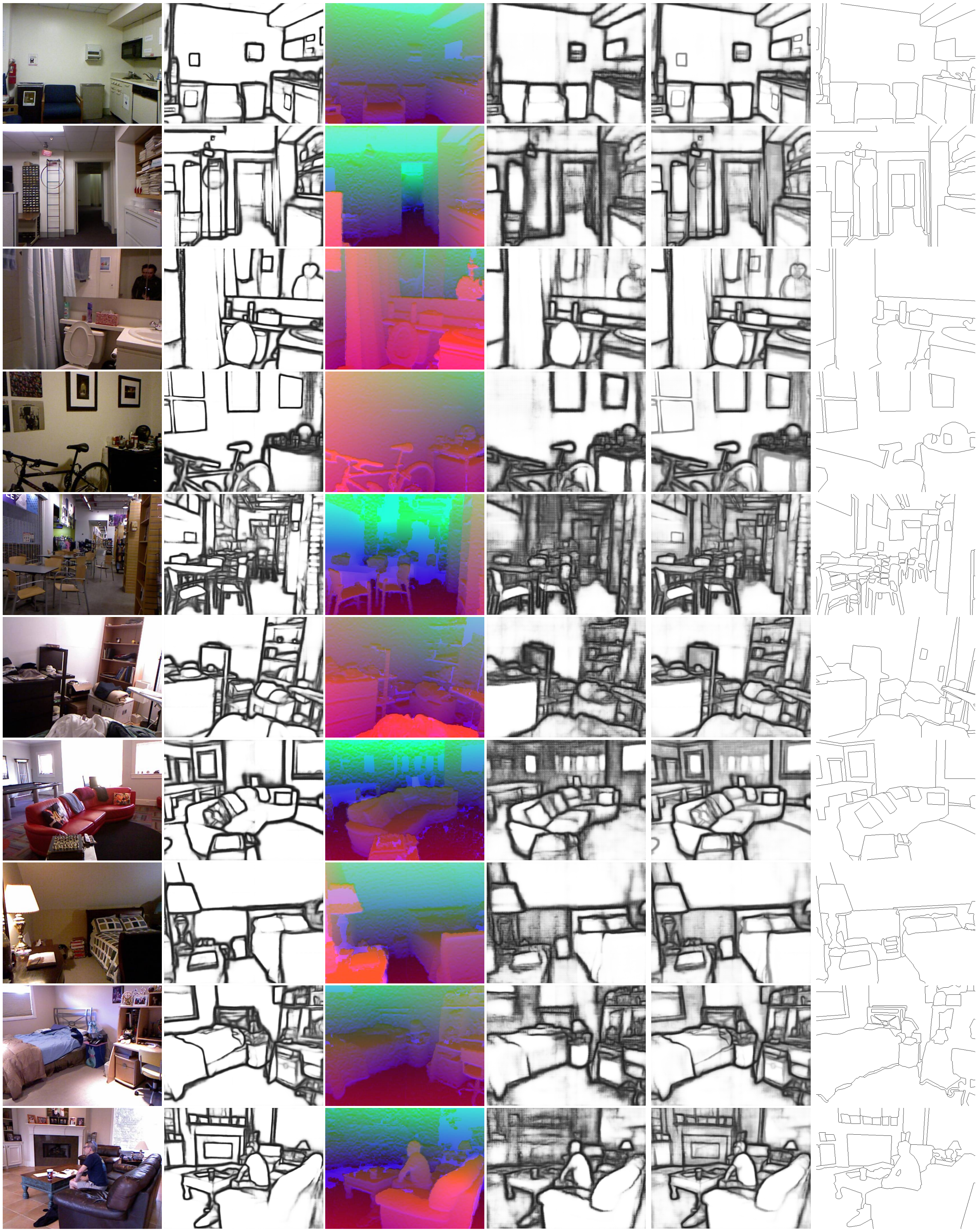
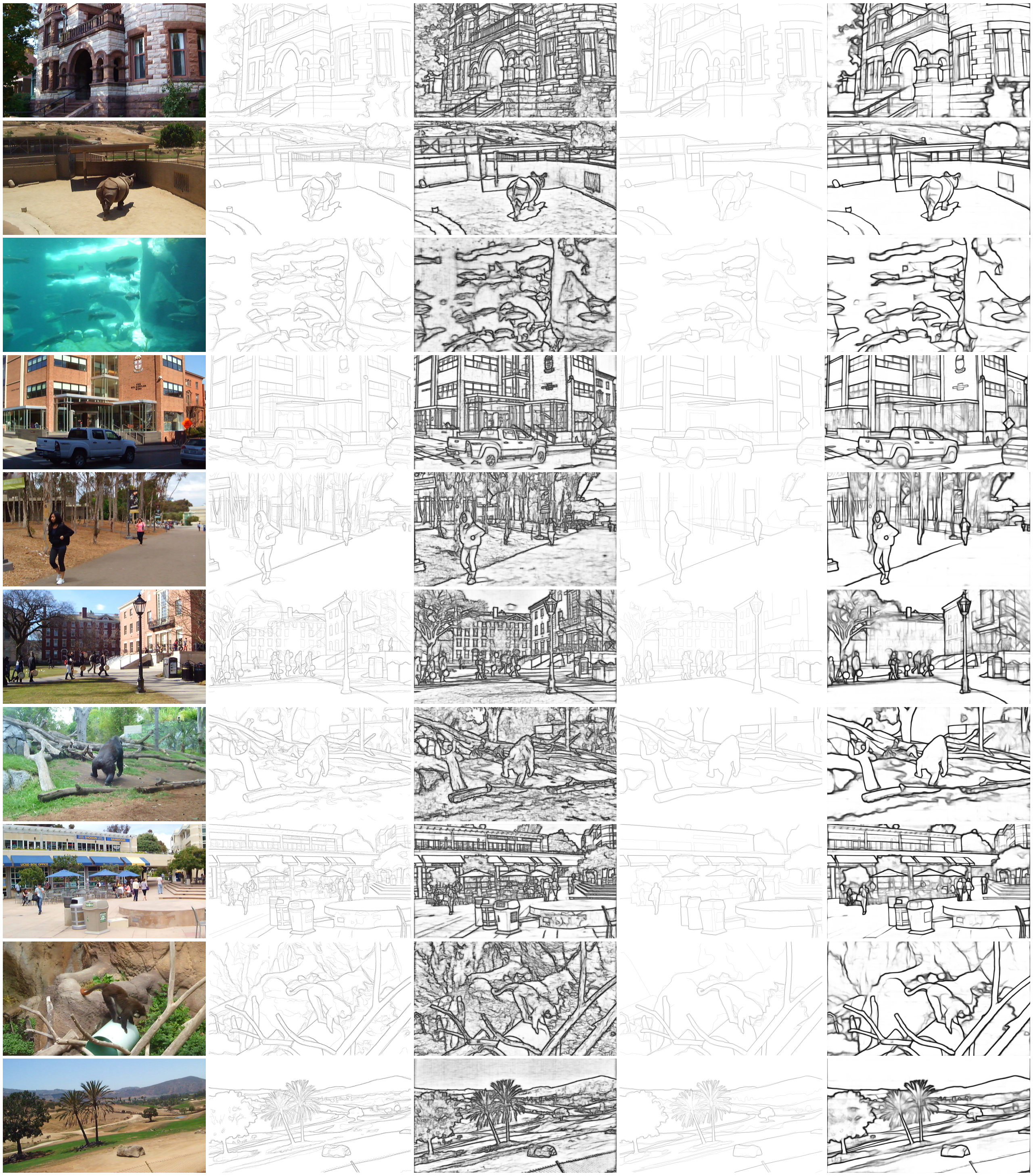


Figure 4. Qualitative comparisons on the testing set of BSDS500.



(a) RGB (b) EDTER (RGB) (c) HHA (d) EDTER (HHA) (e) EDTER (RGB-HHA) (f) GT

Figure 5. Qualitative results on NYUDv2.



(a) Input (b) GT-Edge (c) EDTER-Edge (d) GT-Boundary (e) EDTER-Boundary

Figure 6. Qualitative results on Multicue Edge and Multicue Boundary.