The Treasure Beneath Multiple Annotations: An Uncertainty-aware Edge Detector Supplementary Material

Caixia Zhou[1], Yaping Huang[1]*, Mengyang Pu[2], Qingji Guan[1], Li Huang[1], Haibin Ling[3]
[1]Beijing Key Laboratory of Traffic Data Analysis and Mining, Beijing Jiaotong University, China
[2] School of Control and Computer Engineering, North China Electric Power University, China
[3]Department of Computer Science, Stony Brook University, USA
{cxzhou, yphuang, qjguan,20112044}@bjtu.edu.cn, mengyang.pu@ncepu.edu.cn, hling@cs.stonybrook.edu

In this supplementary material, we provide additional details, including the details of the encoder-decoder network, the details of different uncertainty estimation methods, more ablation studies and visualization results on BSDS500 [1] and Multicue [6] dataset.

A. Details of the Encoder-decoder Network

The encoder of the proposed UAED is EfficientNet [11], whose details can be found in Table 1. It contains eight stages, and involves five up-sampling operations. The size of the final feature map is reduced to $1/32$ of the input image size. We store the first, the third, the fourth, the sixth, and the eighth as multi-scale features for objects with different sizes that are further fed into the following decoders. The structure of the decoder is UNet++ which is the same as the initial design [15].

B. Different Uncertainty Estimation Methods

We explore different uncertainty estimation methods including MC Dropout [3], RBUE [12], generative model based methods [2, 13], probabilistic embeddings [8], and our proposed UAED. The baseline model is a deterministic encoder-decoder structure shown in Figure 1 (a), containing an encoder, a decoder and a prediction head.

*Corresponding author.

MC dropout [3] captures the epistemic (model) uncertainty by sampling from the Bernoulli distribution with a defined probability to decide whether a neuron is valid and operates dropout [10] both the training and test process. By randomly sampling valid neurons, the model acquires different predictions.

RBUE [12] also models epistemic (model) uncertainty. Since not all networks contain Dropout layers, RBUE [12] adds a weight randomly sampling from a uniform distribution for the case when the value is lower than zero in the ReLU activation function. RBUE is easy to implement and does not bring learnable weights.

Generative model based methods, including CVAE-based models [9] and EBM-based models [2], learn low-level latent space to capture randomness caused by the data. As shown in Figure 1 (b), those methods sample features from the latent space learned by the generative models such as CVAE and EBM, and concatenate the features from the label space and encoder. Specifically, CVAE-based model [13] constructs a prior network learning from the im-

Table 1. The detailed network structure of the encoder EfficientNet.

Stage	Layer Name	Kernel	Stride	Channel Input→ Output	Normalization	Activation
1	conv_stem	3×3	2	3→64	BN	-
2	MBConvBlock0 MBConvBlock1-3	3×3; 1×1; 1×1; 1×1 3×3; 1×1; 1×1; 1×1	1 1	64→64→16→64→32 32→32→8→32→32	BN BN	Swish Swish
3	MBConvBlock4 MBConvBlock5-10	1×1; 3×3; 1×1; 1×1; 1×1 1×1; 3×3; 1×1; 1×1; 1×1	1; 2; 1; 1; 1 1	32→192→192→8→192→48 48→288→288→12→288→48	BN BN	Swish Swish
4	MBConvBlock11 MBConvBlock12-17	1×1; 5×5; 1×1; 1×1; 1×1 1×1; 5×5; 1×1; 1×1; 1×1	1; 2; 1; 1; 1 1	48→288→288→12→288→80 80→480→480→20→480→80	BN BN	Swish Swish
5	MBConvBlock18 MBConvBlock19-27	1×1; 3×3; 1×1; 1×1; 1×1 1×1; 3×3; 1×1; 1×1; 1×1	1; 2; 1; 1; 1 1	80→480→480→20→480→160 160→960→960→40→960→160	BN BN	Swish Swish
6	MBConvBlock28 MBConvBlock29-37	1×1; 5×5; 1×1; 1×1; 1×1 1×1; 5×5; 1×1; 1×1; 1×1	1 1	160→960→960→40→960→224 224→1344→1344→56→1344→224	BN BN	Swish Swish
7	MBConvBlock38 MBConvBlock39-50	1×1; 5×5; 1×1; 1×1; 1×1 1×1; 5×5; 1×1; 1×1; 1×1	1; 2; 1; 1; 1 1	224→1344→1344→56→1344→384 384→2304→2304→96→2304→384	BN BN	Swish Swish
8	MBConvBlock51 MBConvBlock52-54	1×1; 3×3; 1×1; 1×1; 1×1 1×1; 3×3; 1×1; 1×1; 1×1	1 1	384→2304→2304→96→2304→640 640→3840→3840→160→3840→640	BN BN	Swish Swish

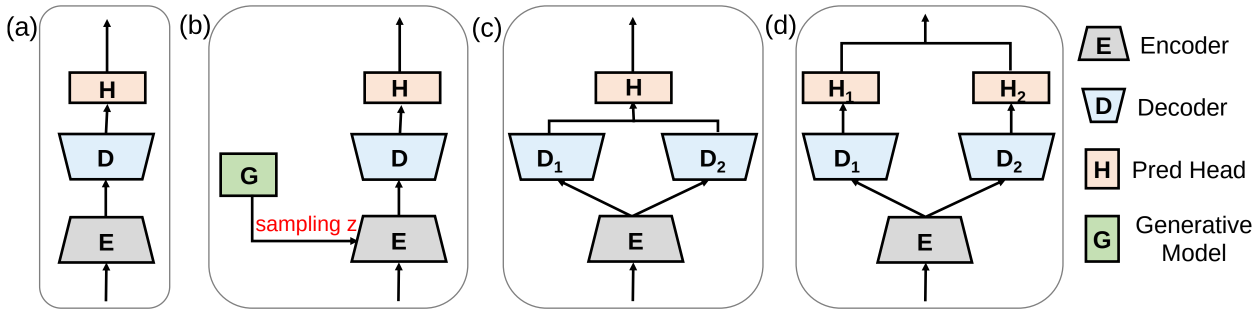


Figure 1. The structures for different uncertainty estimation methods. (a) Baseline. (b) Generative model based method. (c) Probabilistic embeddings. (d) Our proposed UAED.

age and a posterior network learning from the image-label pairs. EBM-based model [14] estimates the prior and the posterior distribution by an energy function. The energy function is learned by a neural network which constitutes several fully connected layers. The sampling from the latent space brings randomness and results in different predictions.

The structure of probabilistic embedding is shown in Figure 1 (c), where two separate decoders are fused into a single prediction header. Compared to our proposed UAED in Figure 1 (d), the probabilistic embedding regards the decoded features as Gaussian distribution rather than the labels in label space.

C. More Ablation Studies

In this section, to further understand the performance gain of our proposed UAED, we conduct more ablation studies to test different design variants.

First, to validate the effectiveness of regarding the labels as distributions, we simply use the averaged probability map as a soft and continuous label ranging $[0, 1]$ for training BCE loss and achieve a score of 0.792 (ODS), 0.807 (OIS), and 0.842 (AP) under the single-scale setting. The performance is much lower that the corresponding encoder-decoder model, which can demonstrate that treating the prediction as a learnable Gaussian distribution can capture the label ambiguity efficiently.

Moreover, instead of using two decoders to estimate the mean and variance of predicted labels separately, we design to use a single decoder by doubling the number of channels of the output layer for predicting the mean and variance. The result is 0.828 (ODS), 0.845 (OIS), and 0.890 (AP), which is slightly lower than our UAED (ODS=0.829, OIS=0.847, AP=0.892). Moreover, compared with single-decoder design, our UAED adds only a negligible GPU memory consumption (from 11.8G to 12G), and slows down the inference only slightly (from 19 FPS to 17 FPS), so we choose two separate decoders for better performance.

D. More Visualization Results

In this section, we report more qualitative results on BSDS500 [1] and Multicue [6] dataset. In Figure 2, we present more visualizations of the proposed UAED on BSDS500 [1]. Figure 3 shows the visual results compared with other approaches for BSDS500 [1]. Moreover, Figure 4 depicts qualitative results for Multicue edge and Multicue boundary [6].

Besides, our method has the sampling ability, which can be found in Figure 5. From the left to right, we can observe that each prediction has a slightly different but reasonable appearance.

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] Yilun Du and Igor Mordatch. Implicit generation and generalization in energy-based models. *arXiv preprint arXiv:1903.08689*, 2019. 1

[3] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016. 1

[4] 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. 5

[5] 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. 5

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

[7] Mengyang Pu, Yaping Huang, Yuming Liu, Qingji Guan, and Haibin Ling. Edter: Edge detection with transformer. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1402–1412, 2022. 5

[8] Yichun Shi and Anil K Jain. Probabilistic face embeddings. In *Int. Conf. Comput. Vis.*, pages 6902–6911, 2019. 1

[9] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. *Adv. Neural Inform. Process. Syst.*, 28, 2015. 1

[10] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014. 1

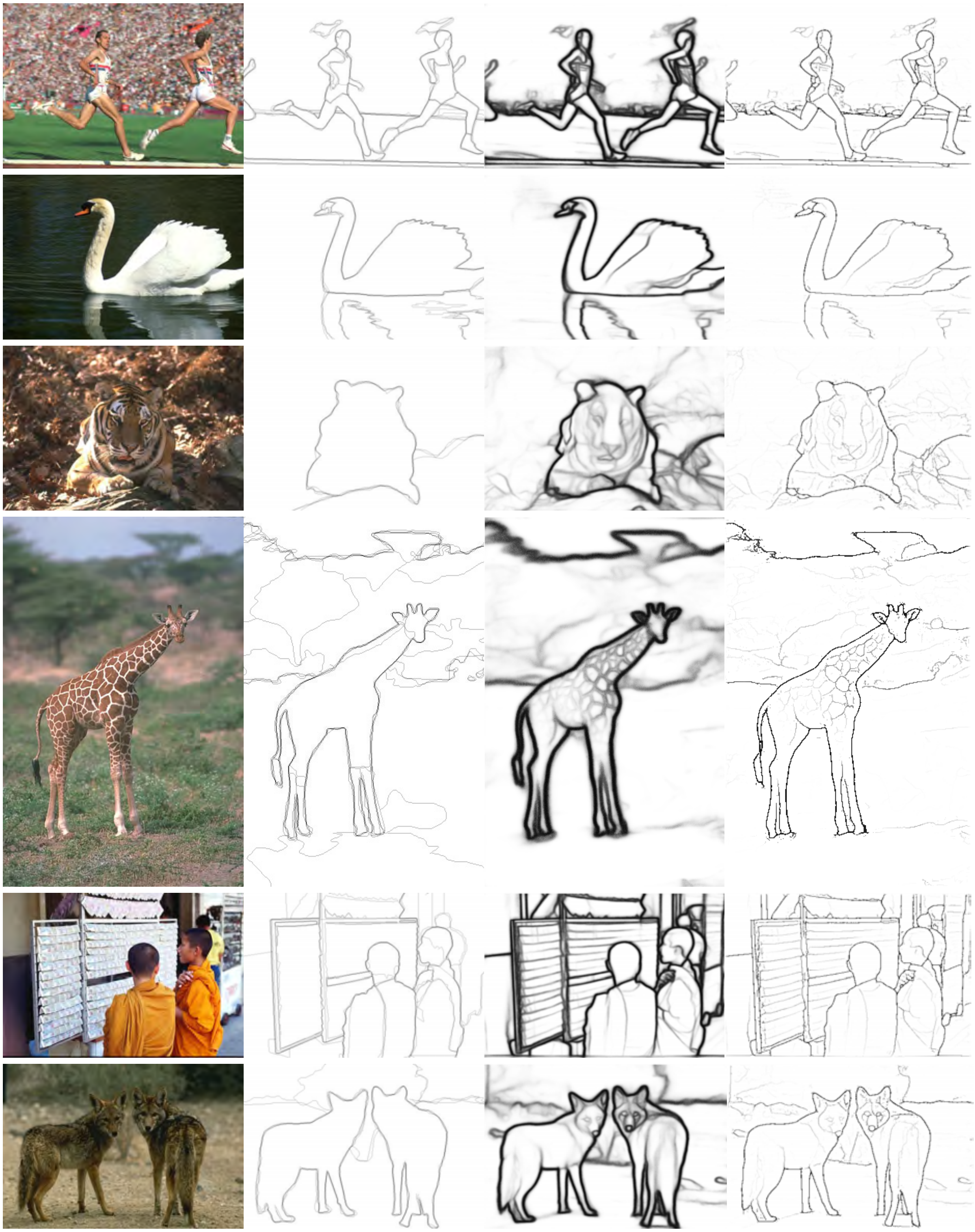
[11] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019. 1

[12] Yufeng Xia, Jun Zhang, Zhiqiang Gong, Tingsong Jiang, and Wen Yao. Rbue: A relu-based uncertainty estimation method of deep neural networks. *arXiv preprint arXiv:2107.07197*, 2021. 1

[13] Jing Zhang, Deng-Ping Fan, Yuchao Dai, Saeed Anwar, Fatemeh Sadat Saleh, Tong Zhang, and Nick Barnes. Uc-net: Uncertainty inspired rgb-d saliency detection via conditional variational autoencoders. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 8582–8591, 2020. 1

[14] Jing Zhang, Jianwen Xie, Nick Barnes, and Ping Li. Learning generative vision transformer with energy-based latent space for saliency prediction. *Adv. Neural Inform. Process. Syst.*, 34:15448–15463, 2021. 2

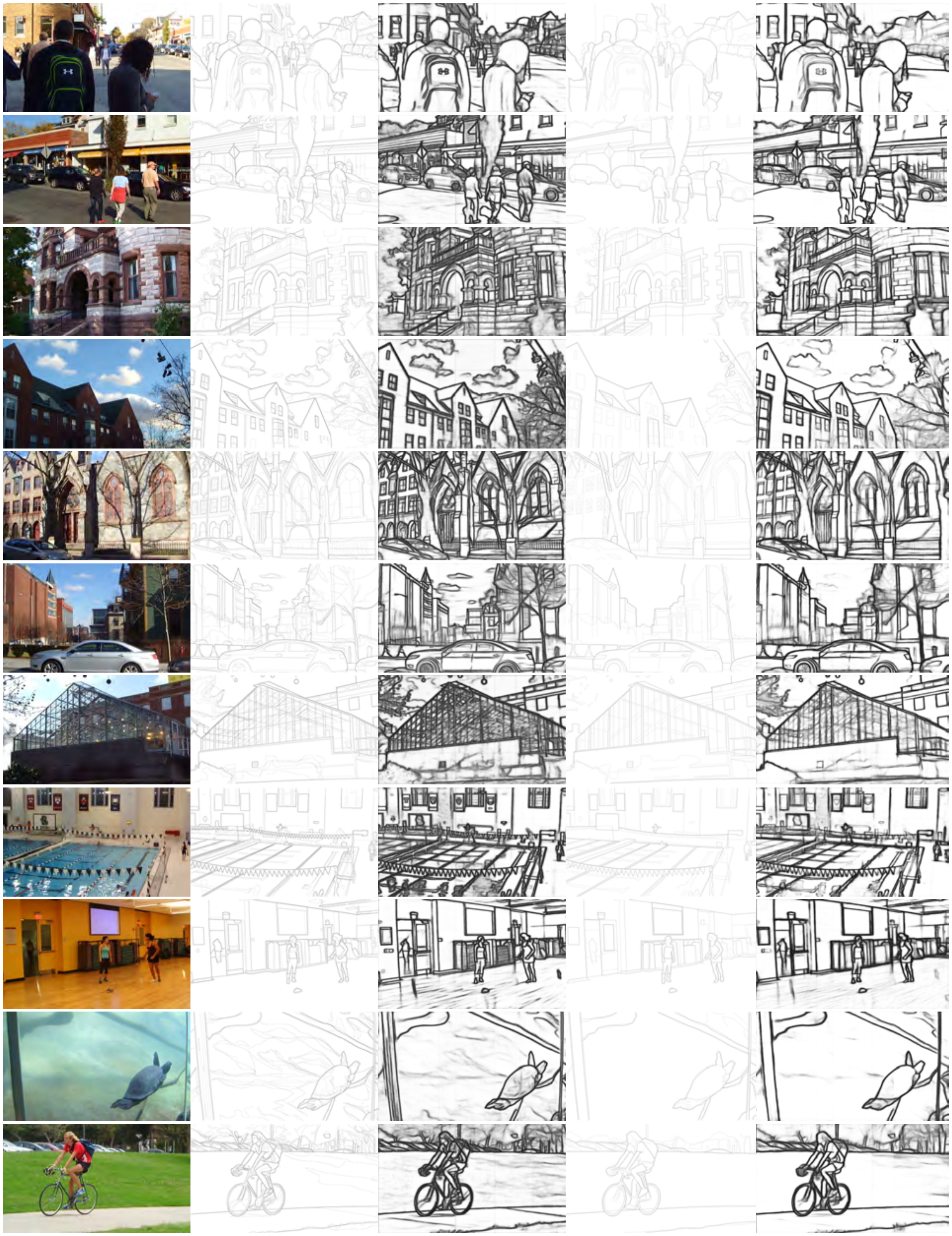
[15] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 3–11. Springer, 2018. 1



(a) Image (b) Ground truth (c) UAED (Ours) (d) UAED after NMS

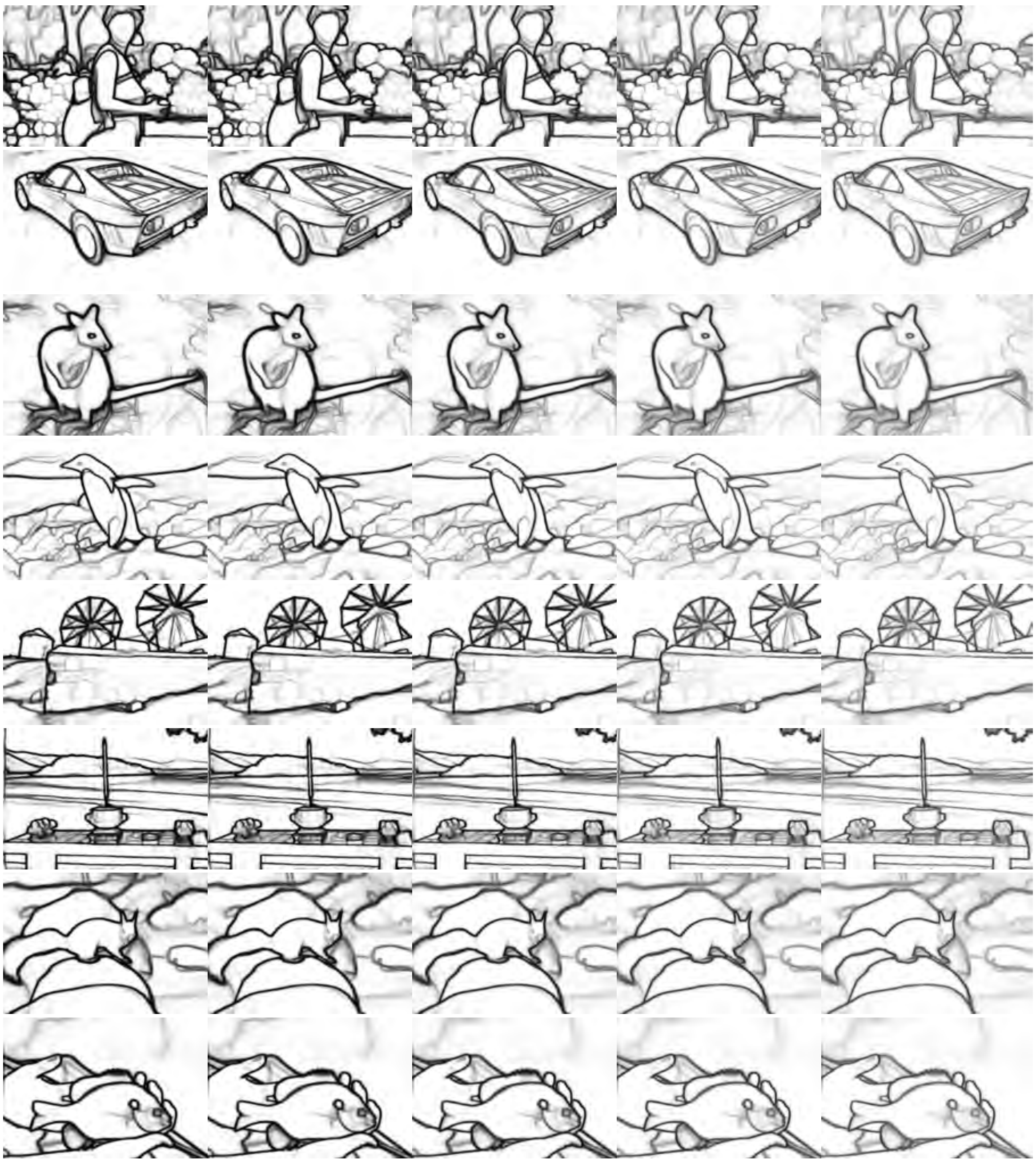
Figure 2. Qualitative results of proposed UAED on BSDS500.

Figure 3. Qualitative comparisons on the testing set of BSDS500.



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

Figure 4. Qualitative results on Multicue Edge and Multicue Boundary.



(a) $\mu + 3\sigma$ (b) $\mu + 2\sigma$ (c) μ (e) $\mu - 2\sigma$ (d) $\mu - 3\sigma$

Figure 5. Different samplings on the testing set of BSDS500.