Improved Image Matting via Real-time User Clicks and Uncertainty Estimation Supplementary Material

Tianyi Wei[1], Dongdong Chen[2,†], Wenbo Zhou[1,†], Jing Liao[3], Hanqing Zhao[1], Weiming Zhang[1], Nenghai Yu[1]
[1]University of Science and Technology of China [2]Microsoft Cloud AI
[3]City University of Hong Kong
{bestwty@mail., welbeckz@, zhq2015@mail., zhangwm@, ynh@ }.ustc.edu.cn
cddlyf@gmail.com, jingliao@cityu.edu.hk

1. Quantitative Comparison on Original Size

In the paper, we keep the aspect ratio of all the test set images while resizing their long sides to 800 pixels for fair comparison with LF-Matting[9] (their model only supports this setting). In Table 1, we further give results for all methods except LF-matting on the original size, which performs consistently with the setting resizing their long sides to 800 pixels.

Methods	SAD	MSE	Grad	Conn
CS-Matting[7]	6.48	42.24	55.54	18.93
Closed-Form[5]	7.09	46.15	61.54	25.44
KNN Matting[1]	7.83	50.53	64.70	33.00
Shared Matting[2]	6.61	42.08	64.18	69.22
Global Matting[3]	7.20	44.61	59.42	26.07
DIM[8]	2.26	6.02	12.69	22.98
CA-Matting[4]	1.66	3.22	8.52	15.42
GCA-Matting[6]	1.55	3.37	7.87	14.20
Ours	1.86	3.79	9.72	16.20

Table 1. Quantitative comparison on the DIM dataset. The metrics SAD, MSE, Grad and Conn are scaled by 10^2, 10^3, 10^5 and 10^3, respectively.

2. Details about Network Architecture

The uncertainty decoder has the same network structure as the alpha matte decoder. In Table 2 and Table 3, we provide the details of the interactive matting network and the local refinement network respectively.

† Corresponding Author.

3. Visual Results of Uncertainty Estimation

In Figure 1, we show some output results from our uncertainty decoder on the DIM test dataset[8]. Obviously, the predicted uncertainty is a good indication of the potential alpha matte error calculated by using the ground truth alpha matte. On the other hand, the fact that most image regions already get pretty good matting results further justifies the local refinement design.

4. Results on Real-world Images

In Figure 2 and Figure 3, we provide more alpha matting results for real-world images. The high-quality alpha matte demonstrates the robustness of our method.

References

[1] Qifeng Chen, Dingzeyu Li, and Chi-Keung Tang. Knn matting. *CVPR 2012*, pages 869–876, 2013. 1

[2] Eduardo Simoes Lopes Gastal and Manuel Menezes de Oliveira Neto. Shared sampling for real-time alpha matting. *Comput. Graph. Forum*, 29:575–584, 2010. 1

[3] Kaiming He, Christoph Rhemann, C. Rother, X. Tang, and Jian Sun. A global sampling method for alpha matting. *CVPR 2011*, pages 2049–2056, 2011. 1

[4] Qiqi Hou and Feng Liu. Context-aware image matting for simultaneous foreground and alpha estimation. *ICCV 2019*, pages 4129–4138, 2019. 1

[5] A. Levin, Dani Lischinski, and Yair Weiss. A closed-form solution to natural image matting. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30:228–242, 2008. 1

[6] Yaoyi Li and Hongtao Lu. Natural image matting via guided contextual attention. In *AAAI 2020*, volume 34, pages 11450–11457, 2020. 1

[7] Ehsan Shahrian, Deepu Rajan, Brian L. Price, and Scott Cohen. Improving image matting using comprehensive sampling sets. *CVPR 2013*, pages 636–643, 2013. 1

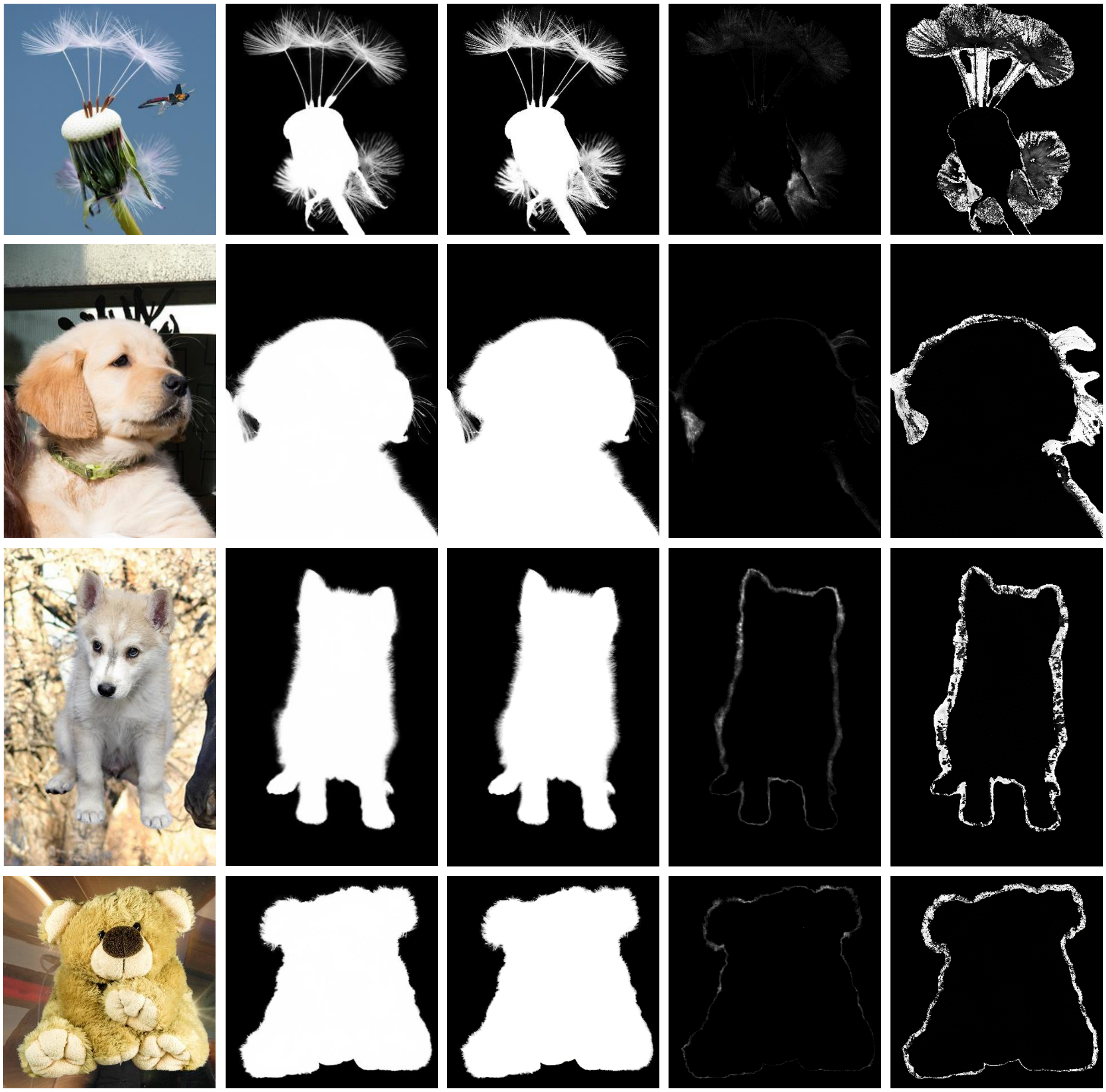
Stage Name	Layer Name	Kernel	Str.	Ch I/O	Normalization	Activation	Input
Encoder Stage1	conv1	3 x 3	2	4/32	BN	ReLU	Images
	conv2	3 x 3	1	32/32	BN	ReLU	ES1conv1
	conv3	3 x 3	2	32/64	BN	ReLU	ES1conv2
Skip Stage1	conv1	3 x 3	1	4/32	BN	ReLU	Images
	conv2	3 x 3	1	32/32	BN	ReLU	SS1conv1
Skip Stage2	conv1	3 x 3	1	32/32	BN	ReLU	ES1conv2
	conv2	3 x 3	1	32/32	BN	ReLU	SS2conv1
Encoder Stage2	res_block1	3 x 3	1	64/64	BN	ReLU	ES1conv3
	res_block2	3 x 3	1	64/64	BN	ReLU	ES2res_block1
	res_block3	3 x 3	1	64/64	BN	ReLU	ES2res_block2
Skip Stage3	conv1	3 x 3	1	64/64	BN	ReLU	ES2res_block3
	conv2	3 x 3	1	64/64	BN	ReLU	SS3conv1
Encoder Stage3	res_block1	3 x 3	2	64/128	BN	ReLU	ES2res_block3
	res_block2	3 x 3	1	128/128	BN	ReLU	ES3res_block1
	res_block3	3 x 3	1	128/128	BN	ReLU	ES3res_block2
	res_block4	3 x 3	1	128/128	BN	ReLU	ES3res_block3
Skip Stage4	conv1	3 x 3	1	128/128	BN	ReLU	ES3res_block4
	conv2	3 x 3	1	128/128	BN	ReLU	SS4conv1
Encoder Stage4	res_block1	3 x 3	2	128/256	BN	ReLU	ES3res_block4
	res_block2	3 x 3	1	256/256	BN	ReLU	ES4res_block1
	res_block3	3 x 3	1	256/256	BN	ReLU	ES4res_block2
	res_block4	3 x 3	1	256/256	BN	ReLU	ES4res_block3
Skip Stage5	conv1	3 x 3	1	256/256	BN	ReLU	ES4res_block4
	conv2	3 x 3	1	256/256	BN	ReLU	SS5conv1
Encoder Stage5	res_block1	3 x 3	2	256/512	BN	ReLU	ES4res_block4
	res_block2	3 x 3	1	512/512	BN	ReLU	ES5res_block1
Decoder Stage1	res_block1	4 x 4	2	512/256	BN	ReLU	ES5res_block2
	res_block2	3 x 3	1	256/256	BN	ReLU	DS1res_block1
Decoder Stage2	res_block1	4 x 4	2	256/128	BN	ReLU	DS1res_block2+SS5conv2
	res_block2	3 x 3	1	128/128	BN	ReLU	DS2res_block1
	res_block3	3 x 3	1	128/128	BN	ReLU	DS2res_block2
Decoder Stage3	res_block1	4 x 4	2	128/64	BN	ReLU	DS2res_block3+SS4conv2
	res_block2	3 x 3	1	64/64	BN	ReLU	DS3res_block1
	res_block3	3 x 3	1	64/64	BN	ReLU	DS3res_block2
Decoder Stage4	res_block1	4 x 4	2	64/32	BN	ReLU	DS3res_block3+SS3conv2
	res_block2	3 x 3	1	32/32	BN	ReLU	DS4res_block1
Decoder Stage5	transpose_conv1	4 x 4	2	32/32	BN	ReLU	DS4res_block2+SS2conv2
	conv1	3 x 3	1	32/1	-	Tanh	DS5trans_conv1+SS1conv2

Table 2. The detailed network structure of the interactive matting network.

Stage Name	Layer Name	Kernel	Str.	Ch I/O	Normalization	Activation	Input
Refinement Stage1	conv1	3 x 3	1	4/64	BN	ReLU	Patches
Refinement Stage2	res_block1	3 x 3	1	64/64	BN	ReLU	RS1conv1
	res_block2	3 x 3	1	64/64	BN	ReLU	RS2res_block1
	res_block3	3 x 3	1	64/64	BN	ReLU	RS2res_block2
	res_block4	3 x 3	1	64/64	BN	ReLU	RS2res_block3
Refinement Stage3	conv1	3 x 3	1	64/1	-	Tanh	RS2res_block4

Table 3. The detailed network structure of the local refinement network.

[8] Ning Xu, Brian L. Price, Scott Cohen, and Thomas S. Huang. Deep image matting. *CVPR 2017*, pages 311–320, 2017. 1



Input Image A-Decoder Output Ground Truth U-Decoder Output Oracle

Figure 1. Visual results of our uncertainty estimation ("U-Decoder Output") on the DIM test dataset. Oracle denotes the ℓ_1-distance between predicted alpha matte and the corresponding ground-truth. Both Oracle and the predicted uncertainty have been normalized for visualization.

[9] Yunke Zhang, Lixue Gong, Lubin Fan, Peiran Ren, Qixing Huang, Hujun Bao, and Weiwei Xu. A late fusion cnn for digital matting. *CVPR 2019*, pages 7461–7470, 2019. 1

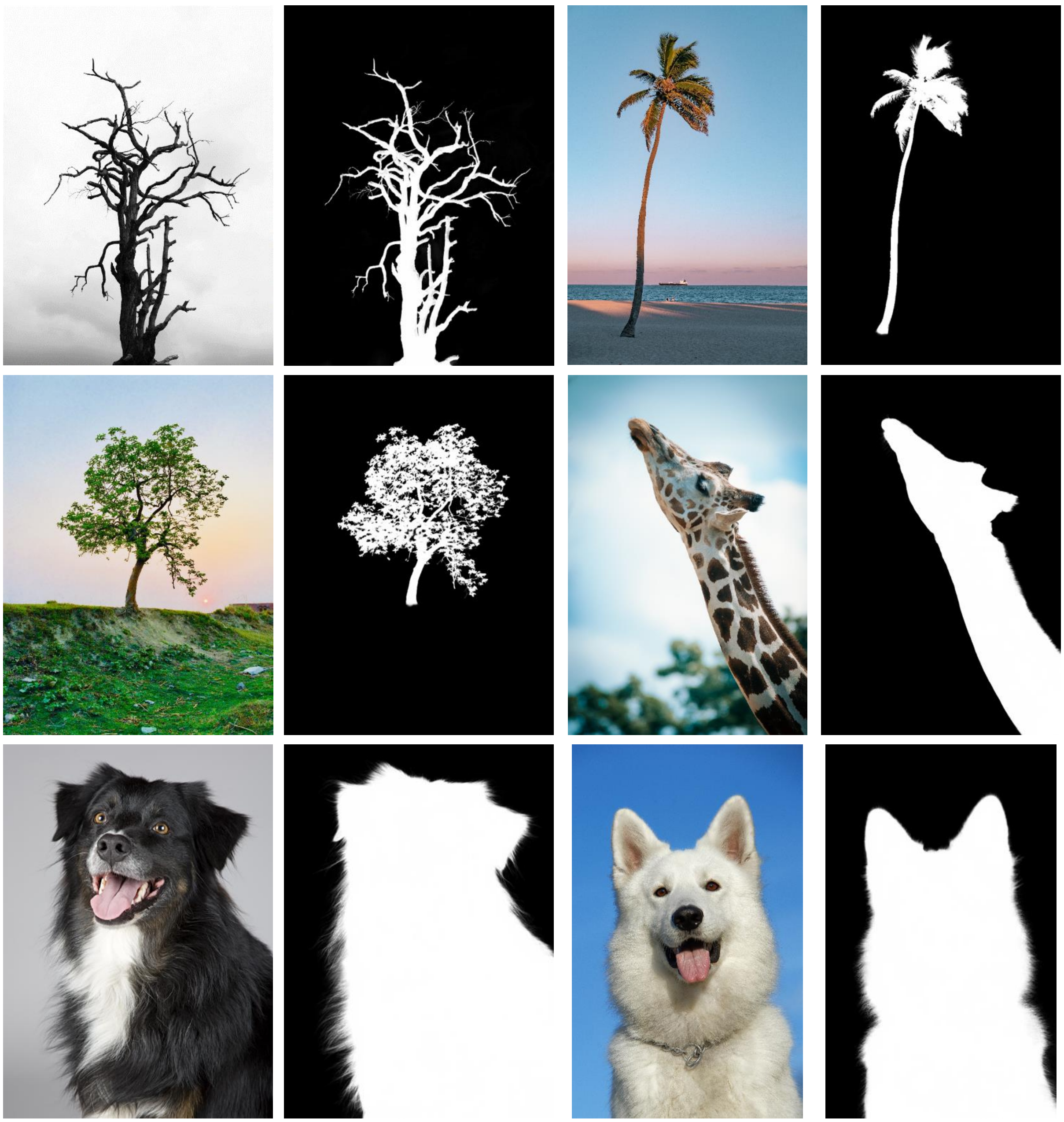


Figure 2. Real-world image matting results.

Figure 3. Real-world image matting results.