Boosting Robustness of Image Matting with Context Assembling and Strong Data Augmentation: Supplementary Material

Yutong Dai[1], Brian Price[2], He Zhang[2], Chunhua Shen[3]
[1]The University of Adelaide [2]Adobe Inc. [3]Zhejiang University

1. List of Content

This supplementary material includes the following contents:

- Measure of GFLOPs.
- Robustness to trimap precision.
- Details on $\mathcal{SA}$ strategies.
- Ablation study and additional verification on $\mathcal{SA}$.
- More visual results on real-world images and benchmarks, including Composition-1k [10], Distinction-646 [7], SIMD$_{our}$ [9], AIM-500 [4].
- Results on the `alphamatting.com` [8] benchmark.
- Failure cases.

2. Measure of GFLOPs

We measure GFLOPs of SIM [9], FBA [2] and our method. Results are reported in Table.1.

Method	GFLOPs
SIM [9]	48.30
FBA [2]	30.47
M3‡	16.62
M7‡	22.22

Table 1. GFLOPs measured on a 224×224 input.

3. Robustness to Trimap Precision

We conduct evaluations on the AIM-500 with different trimap dilation distances. Methods in comparison are IndexNet [6], GCA [5], A^{2}U [1], SIM [9], FBA [2] and our M7‡. In detail, we generate 4 sets of trimaps using random dilation distances within $[11, 20]$, $[21, 30]$, $[31, 40]$, $[41, 50]$, respectively. We denote them as 20, 30, 40, 50 accordingly in Fig. 1. As shown in Fig. 1, our method is obviously more robust to varying trimap precision on all the metrics.

4. Details on the Strong Data Augmentation Strategies

To supplement the content in the main text, we further detail the $\mathcal{SA}$ strategies in our experiments here.

If AF or AFB is applied alone, we set the possibility as 0.5 and keep the ground truths unmodified; if they are combined, possibility of each is changed to 0.25; further, if AC is added on, we set its possibility as 0.1 when AF and AFB do not happen.

Specifically, in AF and AFB, linear pixel-wise augmentation, nonlinear pixel-wise augmentation and region-wise augmentation happen with a probability of 0.8, 0.1 and 0.1, respectively. In AC, linear pixel-wise augmentation, nonlinear pixel-wise augmentation and region-wise augmentation happen with a probability of 0.2, 0.4 and 0.4, respectively. All the augmentations are randomly selected from the options list in the main text during each operation.

5. Ablation study and extensive verification on Strong Data Augmentation

We report ablation study results on $\mathcal{SA}$ on the Composition-1k in Table. 2. In consistent with results in the main text, our $\mathcal{SA}$ produces comparable results on the synthetic benchmarks.

Method	SAD	MSE	Grad	Conn
N3	25.86	0.0046	9.69	21.16
N3+AF	25.86	0.0045	9.82	21.27
N3+AFB	26.21	0.0048	9.91	21.43
N3+AF+AFB	26.55	0.0050	10.45	21.93
N3+AF+AFB+AC	26.46	0.0049	9.98	21.72

Table 2. Ablation on $\mathcal{SA}$ on the Composition-1k.

We also show additional verification of $\mathcal{SA}$ on the A^{2}U [1] on AIM-500 in Table. 3.

6. More Visual Results

Here we show more visual results. The visualized methods include IndexNet [6], CA [3], GCA [5], A^{2}U [1],

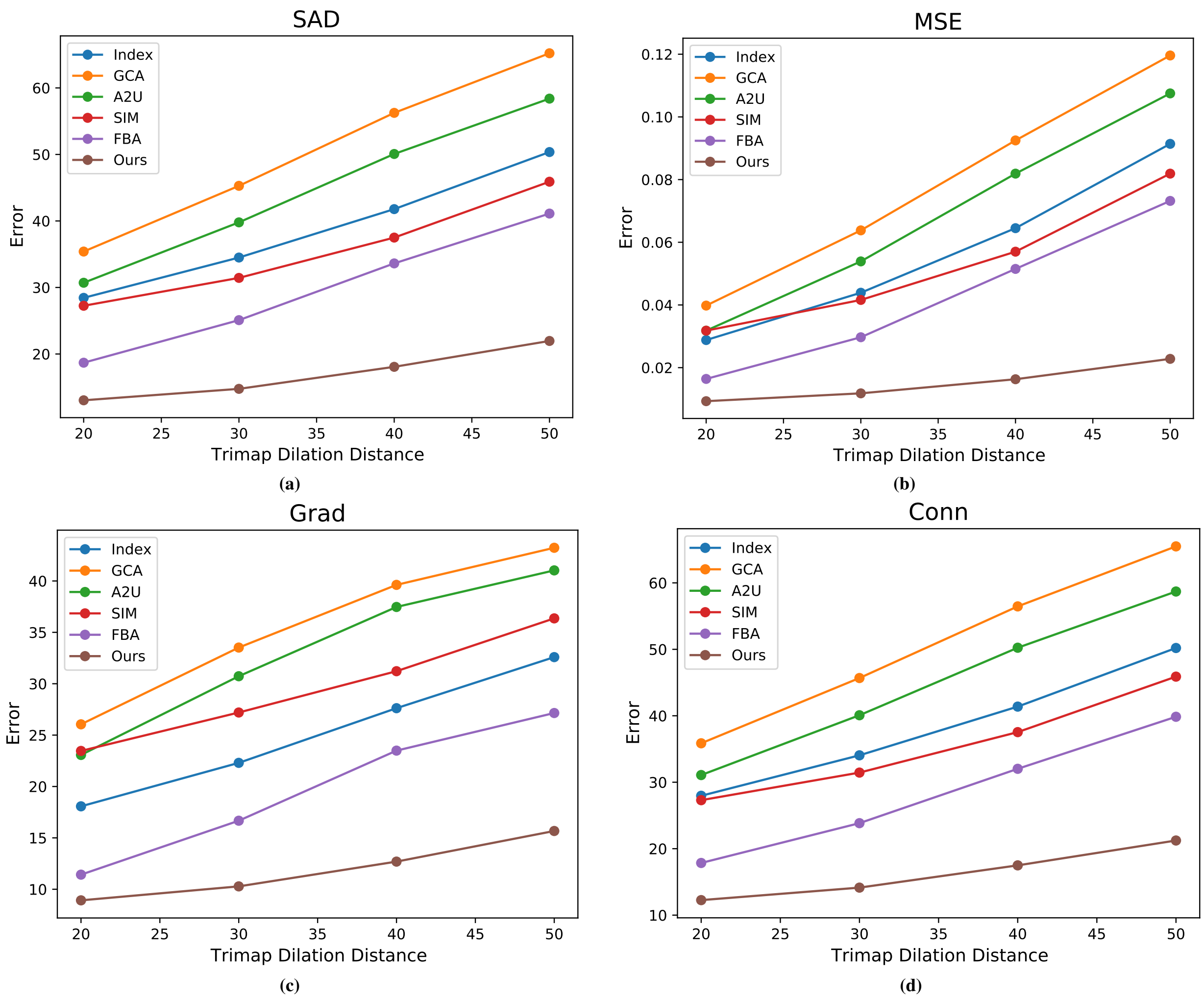

Figure 1. Robustness to trimap precision on the AIM-500.

Method	SAD	MSE	Grad	Conn
A^{2}U [1]	30.38	0.0307	22.60	30.69
A^{2}U+AF+AFB+AC	19.55	0.0165	15.02	19.10

Table 3. Results of $\mathcal{SA}$ on the A^{2}U [1] on AIM-500.

SIM [9], FBA [2] and our M7‡.

Visual results on real-world images are present in Fig. 2 and 3, where Fig. 3 further shows results on coarse trimaps. Our method is more robust in these real-world test cases with coarse-to-fine trimaps.

Visual results on the AIM-500 [4] are exhibited in Fig. 4. Our method achieves better results on structures such as leaves and net.

Visual results on the Composition-1k [10], Distinction-646 [7] and SIMD$_{our}$ [9] are displayed in Fig.5, 6 and 7, respectively. The top-performing methods in comparison all demonstrate appealing results on these synthetic benchmarks, but our method still performs better at background suppression (e.g. Fig. 7) and foreground structure modeling (e.g. Fig. 5 and 6).

7. Results on the `alphamatting.com`

We report results of M7‡ on the `alphamatting.com` online benchmark in Table. 4. The methods in comparison are SIM [9], A^{2}U [1], GCA [5], CA [3], IndexNet [6]. There are only 8 test images in this online benchmark. It worth noting that, SIM is trained with the SIMD training set, which has 736 foregrounds (including 360 foregrounds from DIM) in the training set, while DIM only has 431 foregrounds in the training set; GCA and A^{2}U retrain their models with the whole DIM dataset (including both training set and test set, 481 foregrounds in total) for this benchmark. Our result is directly reported from M7‡ trained with the DIM training set without using extra data or fine-tuning the model, but it still achieves top-performing ranks.

Figure 2. Visual results on real-world images. Best viewed by zooming in.

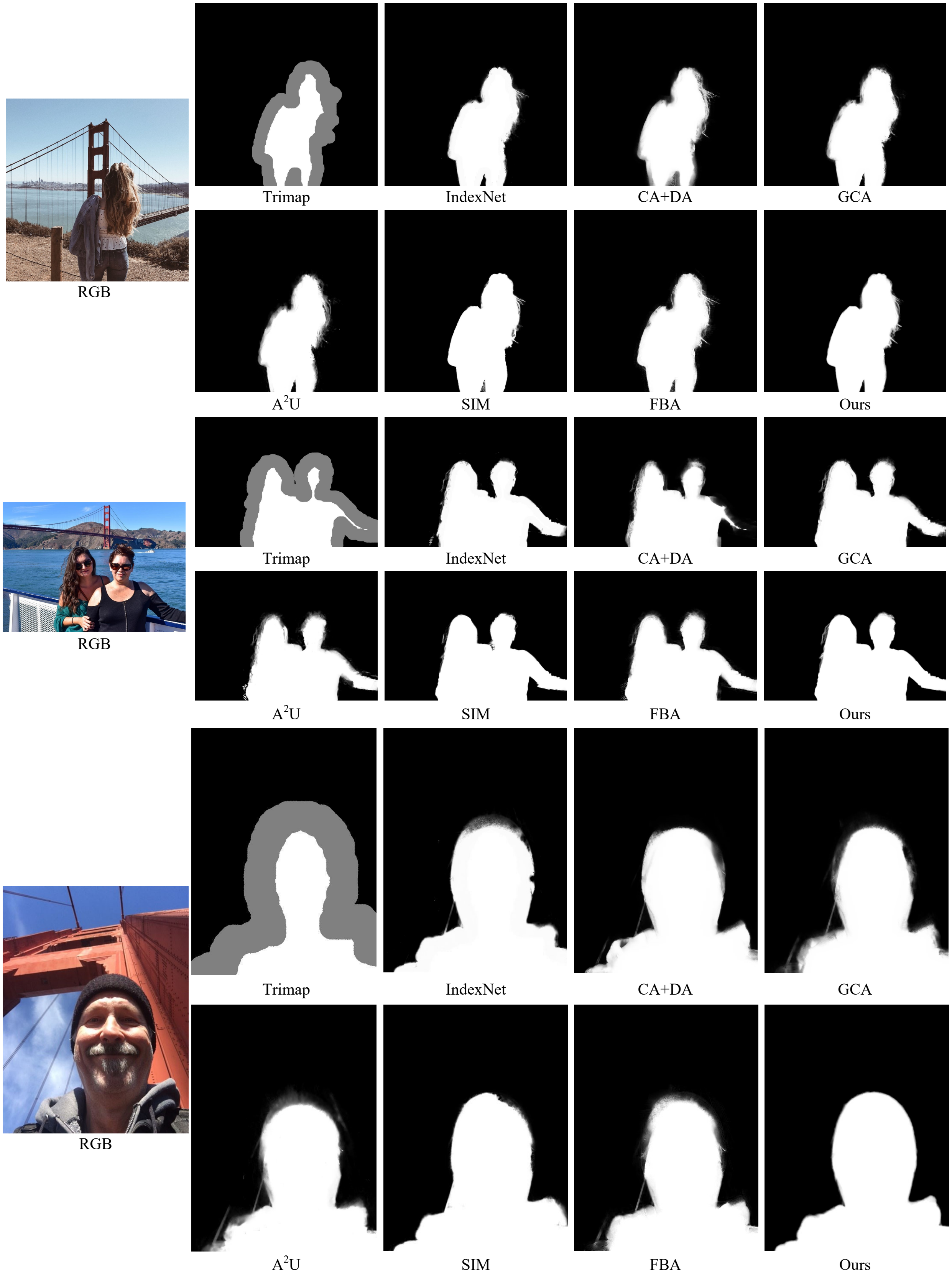

Figure 3. Visual results on real-world images with coarse trimaps.

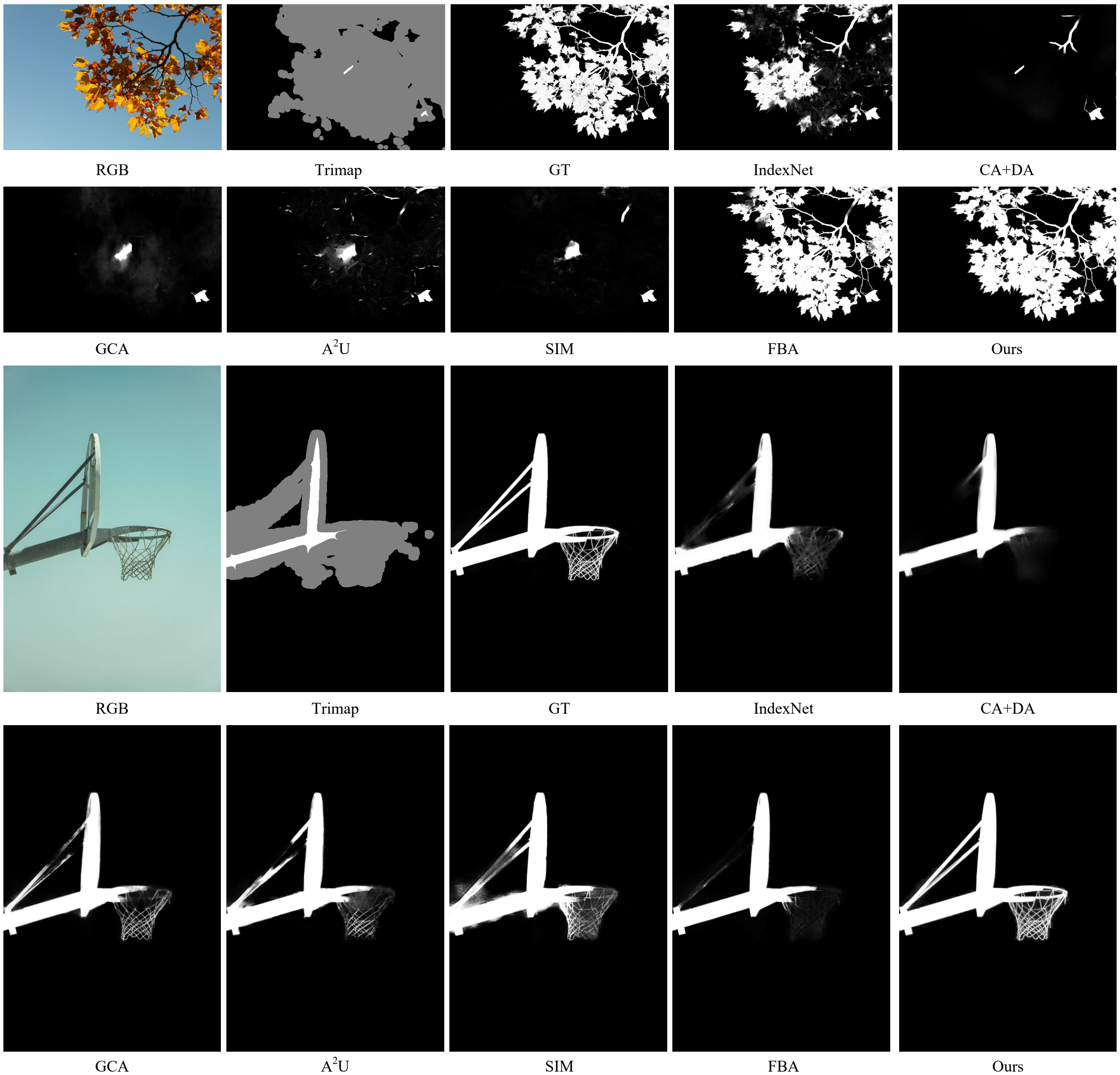

Figure 4. Visual results on the AIM-500.

Method	SAD				MSE				Grad				Conn			
	overall	S	L	U	overall	S	L	U	overall	S	L	U	overall	S	L	U
Ours-M7‡	<u>6.7</u>	**5.9**	**5.8**	<u>8.5</u>	**6.8**	**5.9**	**5.5**	<u>9.1</u>	**4.7**	**4.8**	**3.8**	**5.5**	<u>12.4</u>	<u>16.4</u>	<u>13.7</u>	**7.6**
SIM [9]	**6.5**	<u>7</u>	**5.8**	**6.6**	<u>7</u>	<u>8.1</u>	**5.5**	**7.4**	<u>6.9</u>	<u>8.5</u>	<u>5.9</u>	<u>6.5</u>	**10**	**10**	**8.9**	<u>11.3</u>
A^{2}U [1]	13.3	12.3	10.6	17	15.5	13	12.6	20.8	12.3	11.3	9.4	16.1	27.3.	30.1	28	24.3
GCA [5]	14.5	15.3	12.4	16	15.3	15.1	14.5	16.4	13.7	13.6	12.5	15	22.5	26	20.1	21.4
CA [3]	22.9	26.9	20.9	21	17.6	20.9	18.6	13.3	14.6	15.8	15.5	12.6	25.9	28	24.6	25
IndexNet [6]	19.4	21.5	18.1	18.6	22.9	25.3	21.5	22	18.6	17.3	17.3	21.4	25.5	24.1	26.4	26

Table 4. Results on the `alphamatting.com` online benchmark.

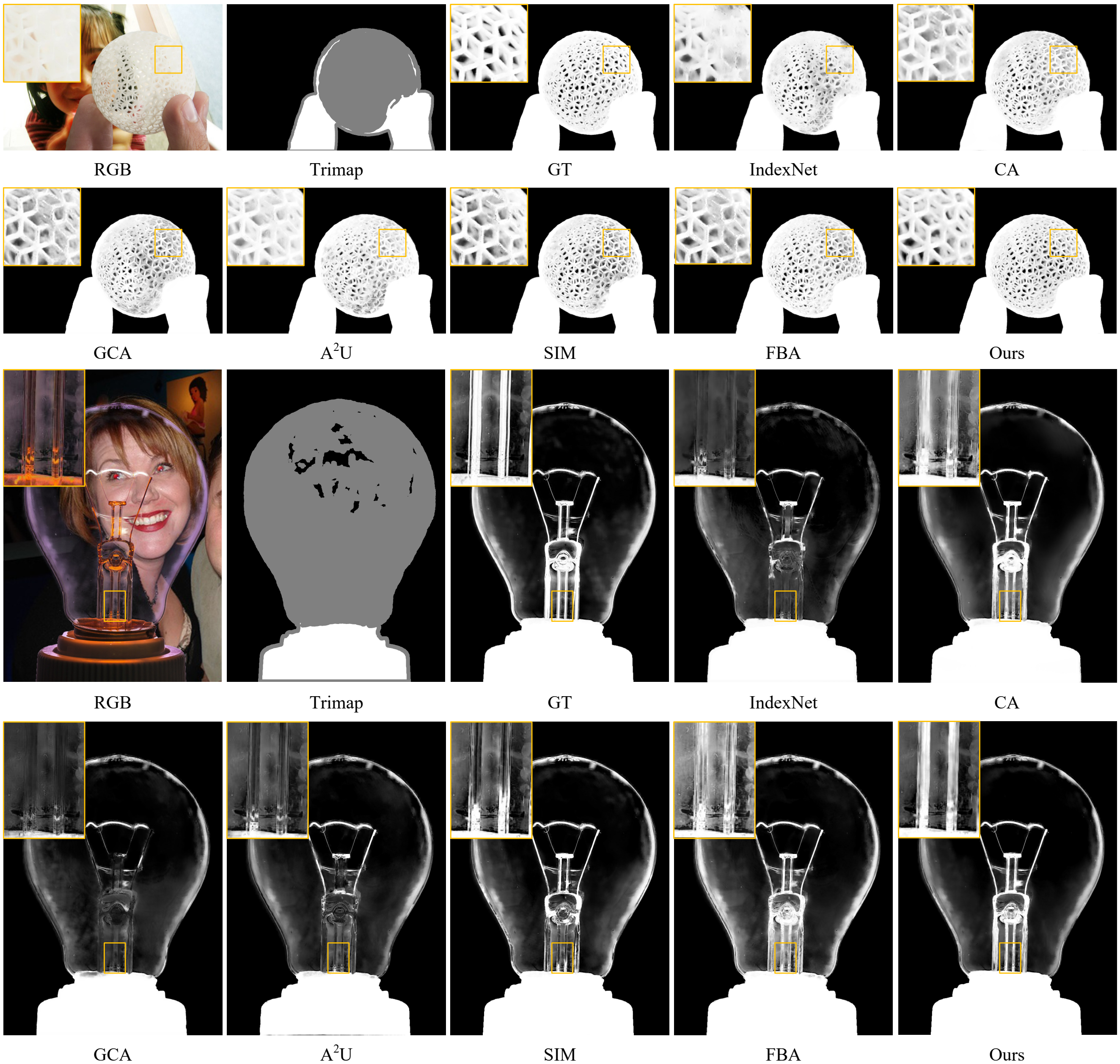

Figure 5. Visual results on the Composition-1k. Best viewed by zooming in.

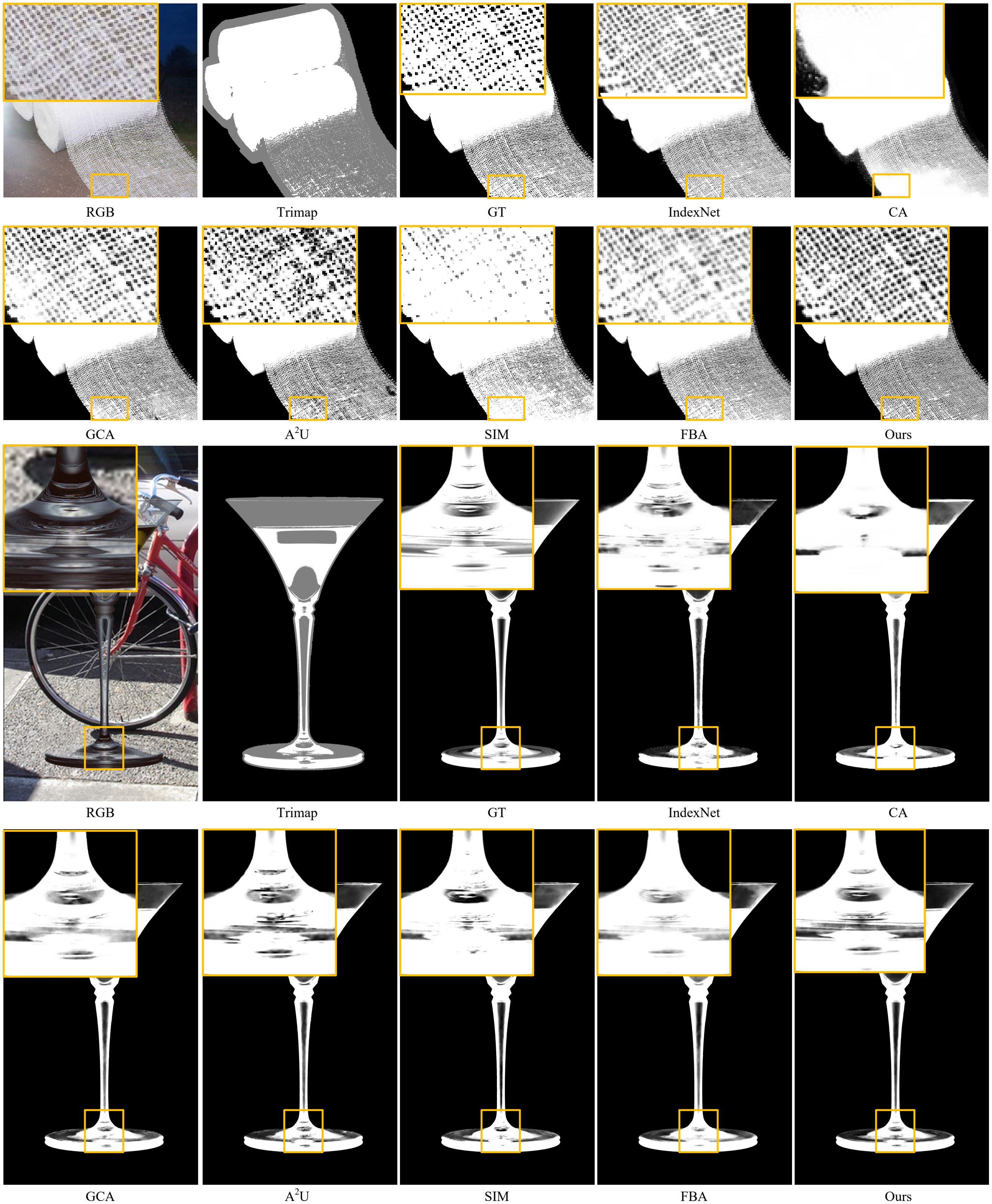

Figure 6. Visual results on the Distinction-646. Best viewed by zooming in.

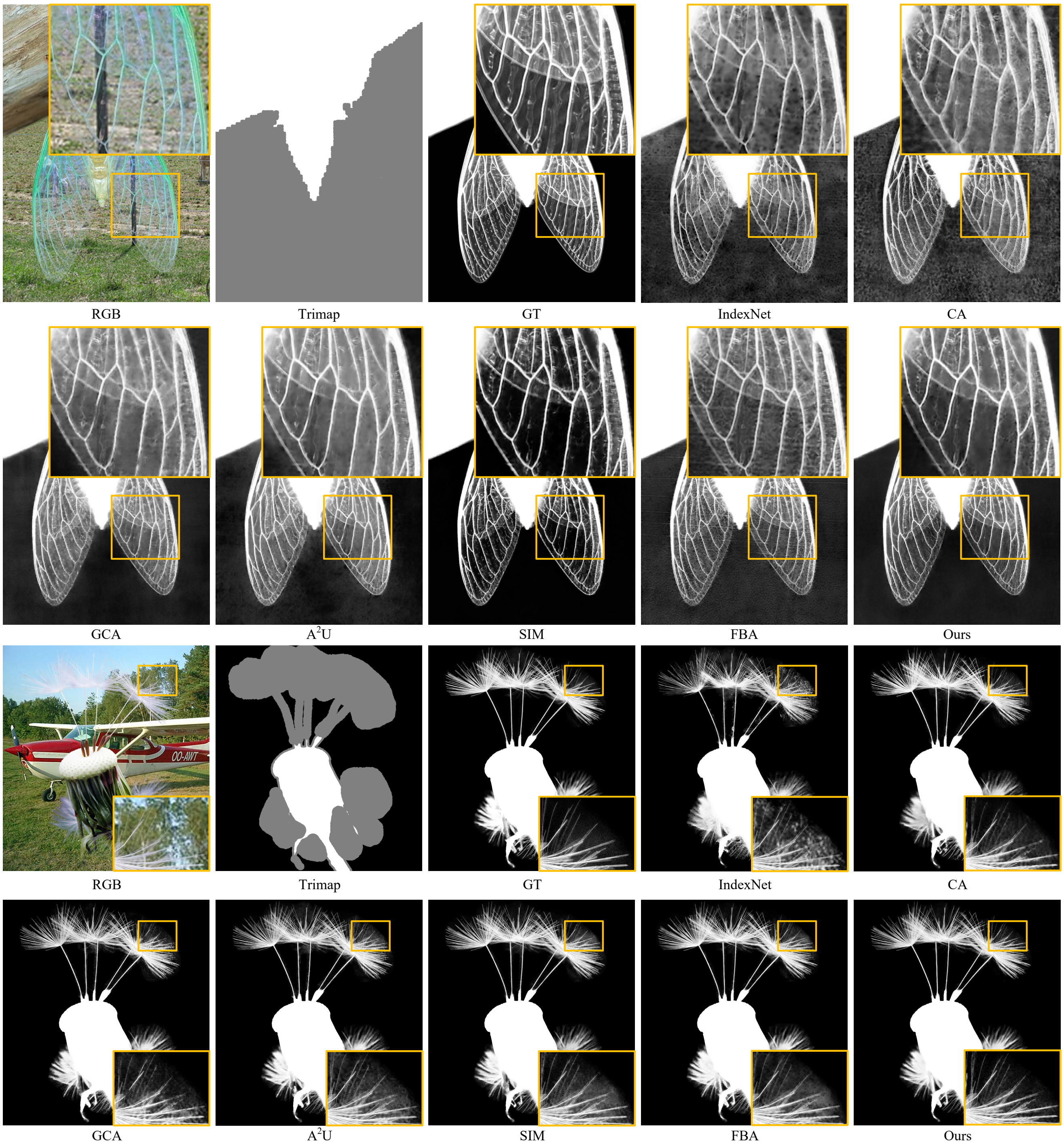

Figure 7. Visual results on the SIMD$_{our}$. Best viewed by zooming in.

8. Failure Cases

Failure examples are visualized in Fig. 8. Our method may fail if there is strong light in the background or there are tiny objects overlapping with the foreground object. A possible solution is to learn the structure of the foreground objects. We leave it as future work.

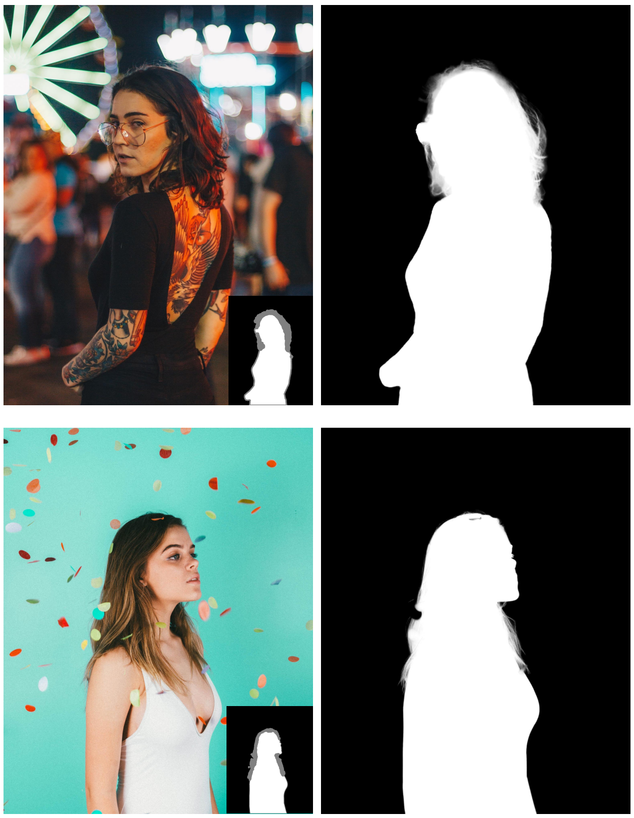

Figure 8. Failure cases.

References

[1] Yutong Dai, Hao Lu, and Chunhua Shen. Learning affinity-aware upsampling for deep image matting. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 6841–6850, 2021. 1, 2, 5

[2] Marco Forte and François Pitié. f, b, alpha matting. *CoRR*, 2020. 1, 2

[3] Qiqi Hou and Feng Liu. Context-aware image matting for simultaneous foreground and alpha estimation. In *Int. Conf. Comput. Vis.*, pages 4130–4139, 2019. 1, 2, 5

[4] Jizhizi Li, Jing Zhang, and Dacheng Tao. Deep automatic natural image matting. *Int. Joint Conf. Artificial Intell.*, 2021. 1, 2

[5] Yaoyi Li and Hongtao Lu. Natural image matting via guided contextual attention. In *Proc. AAAI Conf. Artificial Intell.*, volume 34, pages 11450–11457, 2020. 1, 2, 5

[6] Hao Lu, Yutong Dai, Chunhua Shen, and Songcen Xu. Indices matter: Learning to index for deep image matting. In *Int. Conf. Comput. Vis.*, pages 3266–3275, 2019. 1, 2, 5

[7] Yu Qiao, Yuhao Liu, Xin Yang, Dongsheng Zhou, Mingliang Xu, Qiang Zhang, and Xiaopeng Wei. Attention-guided hierarchical structure aggregation for image matting. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 13676–13685, 2020. 1, 2

[8] Christoph Rhemann, Carsten Rother, Jue Wang, Margrit Gelautz, Pushmeet Kohli, and Pamela Rott. A perceptually motivated online benchmark for image matting. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1826–1833. IEEE, 2009. 1

[9] Yanan Sun, Chi-Keung Tang, and Yu-Wing Tai. Semantic image matting. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 11120–11129, 2021. 1, 2, 5

[10] Ning Xu, Brian Price, Scott Cohen, and Thomas Huang. Deep image matting. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2970–2979, 2017. 1, 2