Supplementary Material for the Paper: Tripartite Information Mining and Integration for Image Matting

Yuhao Liu[1,*], Jiake Xie[2,*], Xiao Shi[3], Yu Qiao[1], Yujie Huang[4], Yong Tang[2,†], Xin Yang[1,†]
[1] Dalian University of Technology, [2] PicUp.AI
[3] Agricultural Information Institute of CAAS, [4] Fudan University
{yuhaoLiu7456, coachqiao2018}@gmail.com, {jxie, yt}@picup.ai
sixiaosmile@outlook.com, 19112020091@fudan.edu.cn, xinyang@dlut.edu.cn

Methods	Time (s)	*320 × 320*	*512 × 512*
DIM [8]	0.265	4066.4	5666.5
IndexNet [4]	0.167	3594.5	4372.6
CAM [2]	15.120	-	-
GCA [3]	0.578	3535.8	4043.3
TIMI-Net (Ours)	0.624	3697.3	4431.3

Table 1: Runtime analysis and GPU memory (MB) comparison of different methods. - denotes the results could be fairly obtained due to the intrinsic mechanism in Tensorflow. *320* and *512* are specified resolution.

1. Overview

In this supplementary material, we first analyse and compare the runtime and GPU memory in Tab. 1. Then, we briefly introduce the online matting platform we have designed. Furthermore, we show some examples of our Human-2K dataset in Fig. 1. Finally, we present more comparisons to the SOTA approaches on Composition-1K [8], Distinctions-646 [5], Our Human-2K and Real-World images in Fig. 2, 3, 4, 5 and 6.

2. Runtime and Memory Statistics

We implement all the tests on a NVIDIA Tesla V100 GPU. As shown in Tab. 1, we calculate the total time in the Composition-1K [8] test set, and then divide it by the total number (*1000*) to obtain the average time. We abandoned the process of metric calculation and results saving, and the batch-size was set to 1 for all methods in all experiments. As for GPU memory calculation, we employ the method provided by this*. Derived from the light-weight model (*e.g.* SegNet [1], MobileNet v2 [7]), DIM [8] and IndexNet [4] can do the inference in a fast speed. Compared to GCA [3] with an single U-Net [6] structure, our RGB-Unit and Trimap-Unit inevitably increase the parameters and computational burden, leading to a fraction of slower speed and larger GPU memory. This may hinder the implementation and application of our approach to some speed-sensitive tasks.

*https://github.com/Oldpan/Pytorch-Memory-Utils

3. Online Matting Platform

We provide an online matting platform for users, the project page: https://wukaoliu.github.io/TIMI-Net. There are two options available. (i) We ask the users to upload an image and the corresponding trimap to predict an alpha matte. (ii) We develop and assemble a build-in trimap prediction into the image matting model. It only requires an image to do the inference, which largely facilitate novice users to use. For a fair comparison, all the model deployed on this platform are trained on the Composition-1K [8] dataset.

References

[1] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(12):2481–2495, 2017.

[2] Qiqi Hou and Feng Liu. Context-aware image matting for simultaneous foreground and alpha estimation. In *Proc. IEEE Int. Conf. Comput. Vis.*, pages 4129–4138, 2019.

[3] Yaoyi Li and Hongtao Lu. Natural image matting via guided contextual attention. In *Proc. AAAI Conf. Artif. Intell.*, pages 11450–11457, 2020.

[4] Hao Lu, Yutong Dai, Chunhua Shen, and Songcen Xu. Indices matter: Learning to index for deep image matting. In *Proc. IEEE Int. Conf. Comput. Vis.*, pages 3265–3274, 2019.

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

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation.

Figure 1: Examples of our Human-2K images and alpha mattes. Please zoom in to see the fine details.

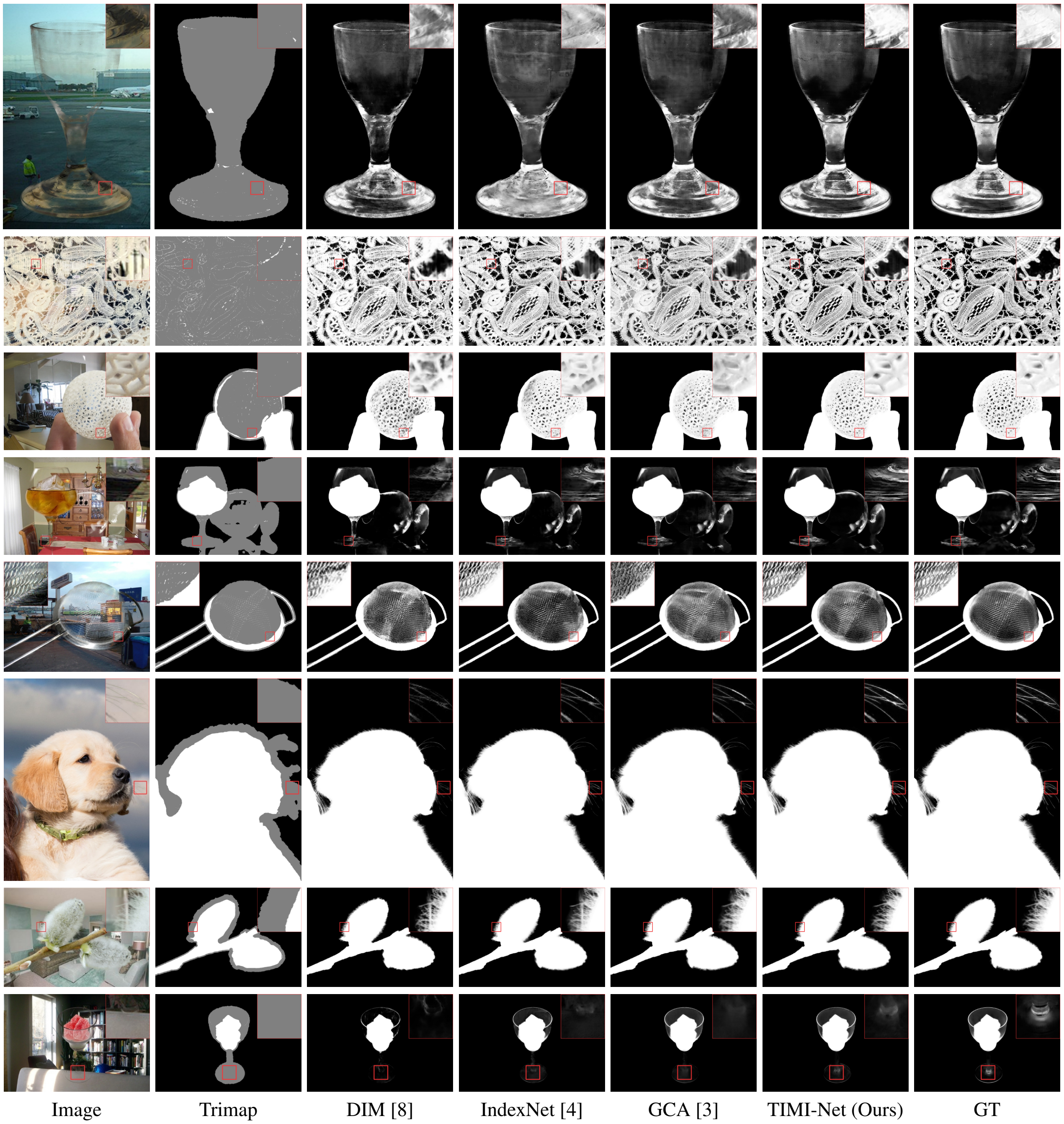

Figure 2: Additional visual comparison of the Composition-1K [8] test set. The results of DIM [8] and IndexNet [4] are obtained based on the released model by IndexNet [4]. We also load the released model by GCA's author for GCA [3].

In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241, 2015.

[7] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 4510–4520, 2018.

[8] Ning Xu, Brian Price, Scott Cohen, and Thomas Huang. Deep image matting. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 311–320, 2017.

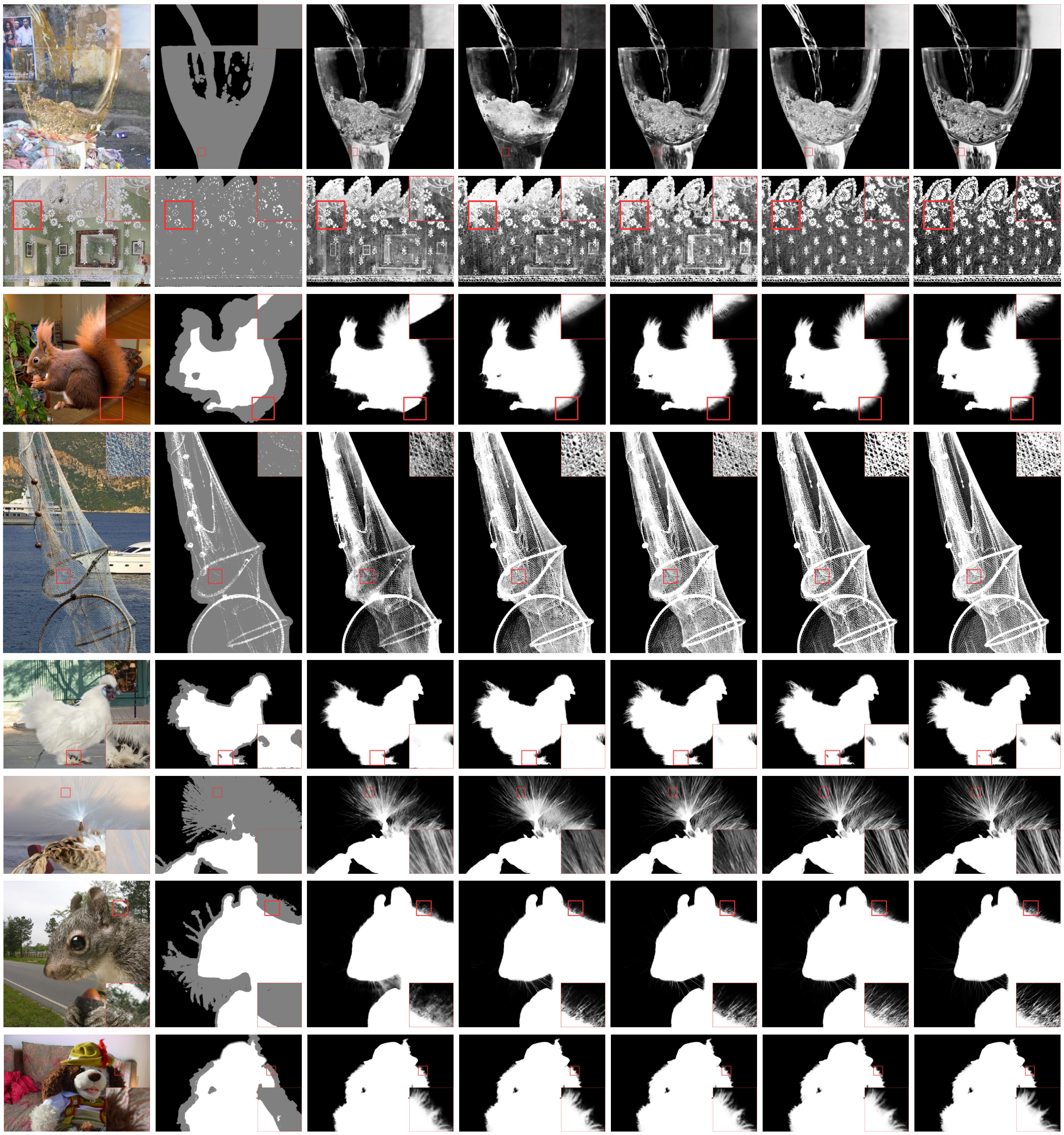

Figure 3: Additional visual comparison of the Distinctions-646 [5] test set. We retrained the models of DIM [8], IndexNet [4] and GCA [3] by their officially released code.

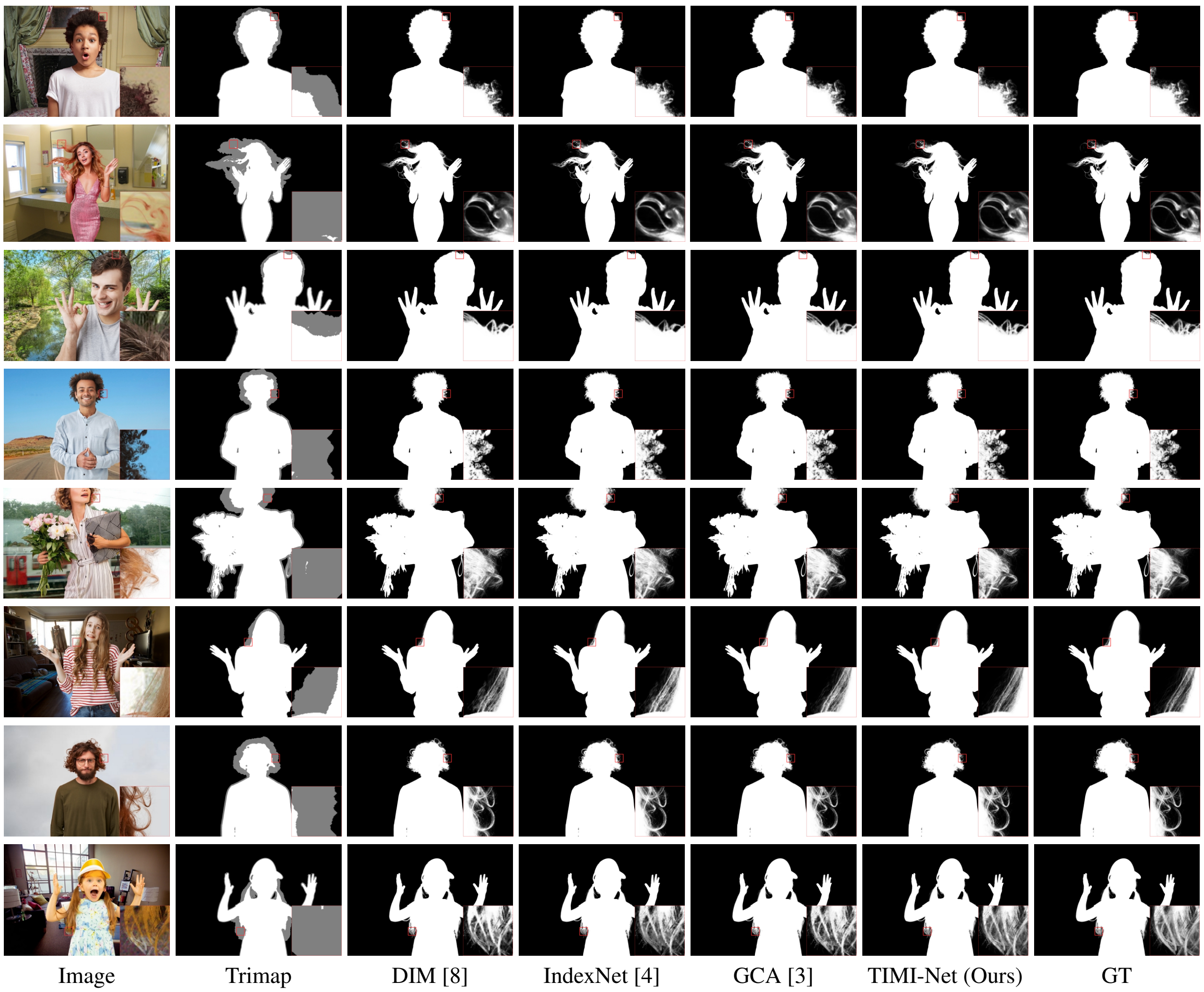

Figure 4: Additional visual comparison of our Human-2K test set. We retrained the models of DIM [8], IndexNet [4] and GCA [3] by their officially released code.

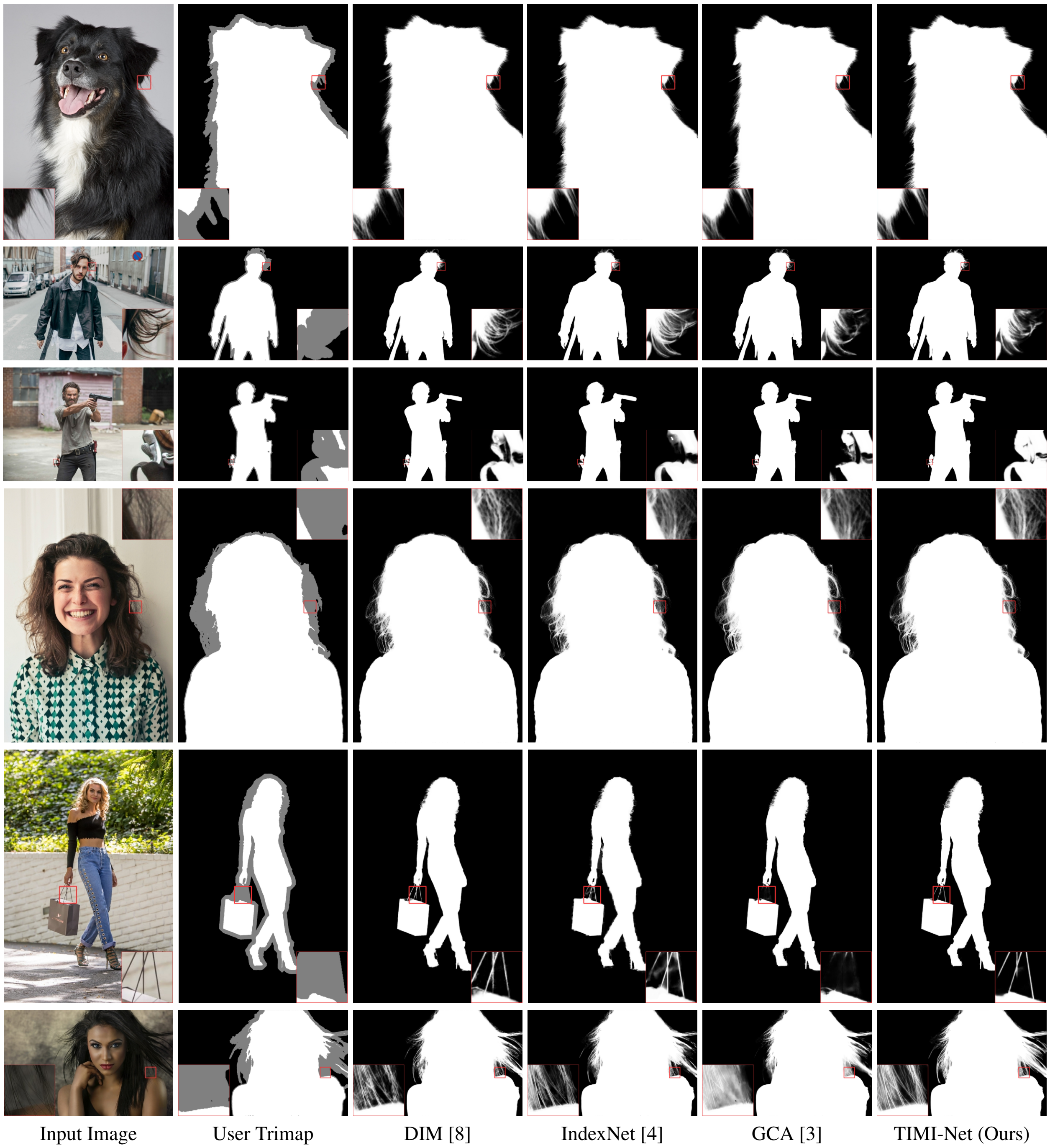

Figure 5: Additional visual comparison of Real-World images. For a fair comparison, all the results are obtained from their model trained on Composition-1K [8] dataset. As for DIM [8] and IndexNet [4], the results are obtained based on the officially released model by IndexNet [4]. We also load the released model by the GCA's authors for GCA [3].

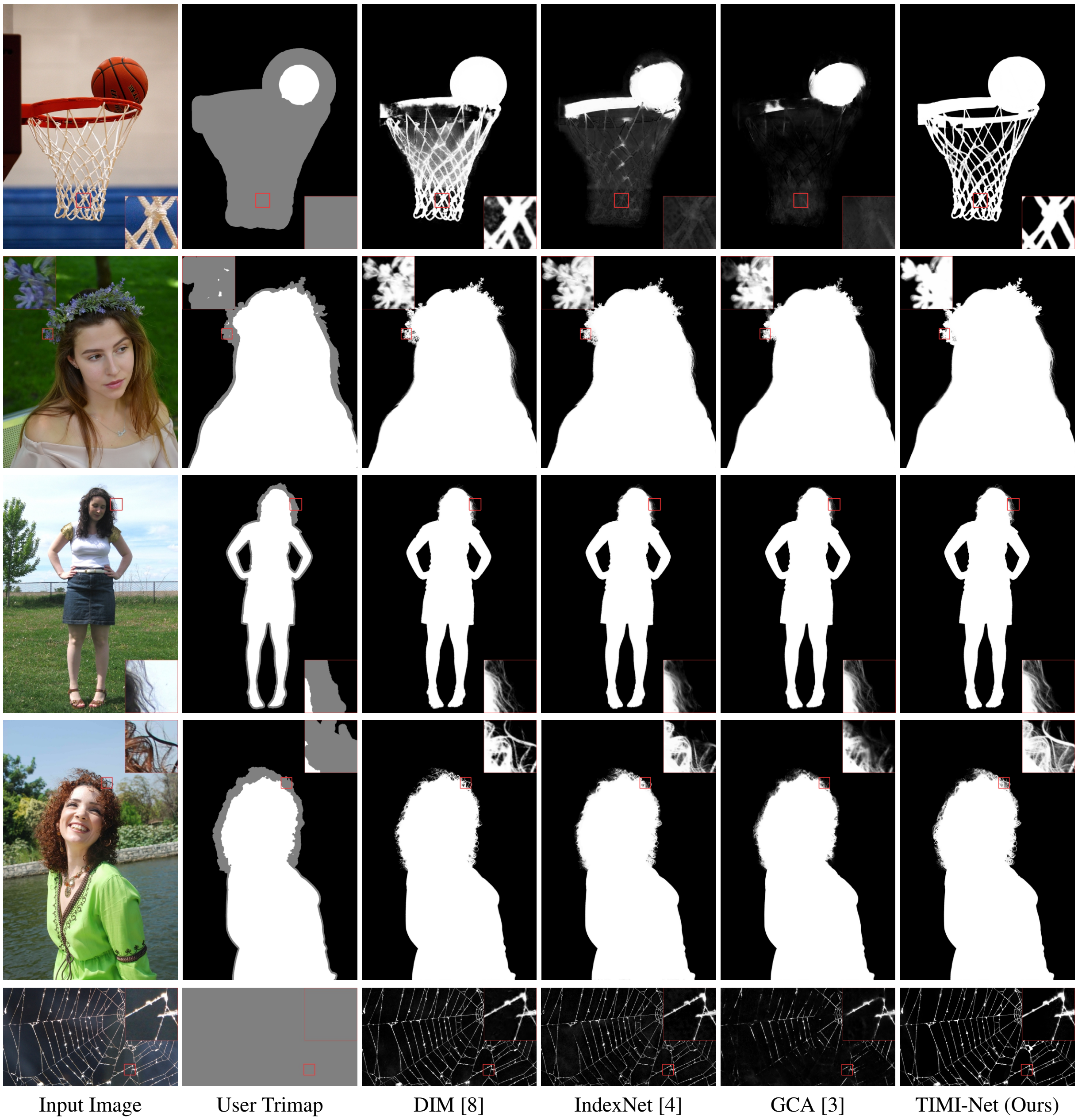

Figure 6: Additional visual comparison of Real-World images. For a fair comparison, all the results are obtained from their model trained on Composition-1K [8] dataset. As for DIM [8] and IndexNet [4], the results are obtained based on the officially released model by IndexNet [4]. We also load the released model by the GCA's authors for GCA [3].