Disentangled Pre-training for Image Matting (Supplementary Material)

Yanda Li[1], Zilong Huang[2], Gang Yu[2], Ling Chen[1], Yunchao Wei[3], Jianbo Jiao[4]
[1]University of Technology Sydney, Australia [2]Tencent, China
[3]Beijing Jiaotong University, China [4]University of Birmingham, UK
{liyanda95,wychao1987}@gmail.com, {zilonghuang, skicyyu}@tencent.com
Ling.Chen@uts.edu.au, j.jiao@bham.ac.uk

1. Limitations

The training data we used is class-agnostic, and as a result no semantic information was considered explicitly (though maybe implicitly), which may restrict the potential of the proposed framework to be applied to other semantic-related tasks.

2. Comparison with the Dense-CL Method

To verify our effectiveness, we compare our DPT with DenseCL [2] based on GCA [1]. Since the backbone of GCA is different from that of DenseCL, for a fair comparison, we replace the backbone with the ResNet50, and perform fine-turning after loading pre-trained weights from DenseCL. As shown in Table 1, our DPT achieves better performance with fewer parameters.

Table 1. Compare with DenseCL based on Comp-1k.

Method	Init.	Backbone	SAD↓	MSE (10^{-3})↓
GCA	DenseCL	ResNet50	45.9	13.0
GCA	Supervised	ResNet34	35.3	9.1
GCA	DPT (Ours)	ResNet34	**33.0**	**8.3**

3. Ablation on losses

In our DPT, we adopt three kinds of losses, L1 regression loss, Composition loss and Laplacian loss. The impact of these loss functions (L_{l1}, L_{lap} and L_{comp}) on the final performance is shown in Table 2.

Table 2. Ablation on losses (50 epochs based on MatteFormer).

Method	Loss	SAD↓	MSE (10^{-3})↓
MatteFormer	L_{l1}	21.9	3.6
MatteFormer	$L_{l1} + L_{lap}$	22.5	3.3
MatteFormer	$L_{l1} + L_{comp}$	22.2	3.5
MatteFormer	$L_{l1} + L_{lap} + L_{comp}$	**21.0**	**3.2**

4. More fine-tuning qualitative results on Composition-1k

We fine-tune on Composition-1k with DPT initialization. As Fig. 1, Fig. 2, Fig. 3, Fig. 4 show, we provide more qualitative results of our DPT on the Composition-1k test set. From the results, it can be seen that our method performs well in various detailed foreground objects.

5. More fine-tuning qualitative results on Distinct-646

As Fig. 5 – Fig. 8 show, we further present more qualitative results of our DPT on Distinct-646 [1]. We first generate trimap by randomly dilating alpha mattes from the ground truth alpha matte with a threshold of 20 (following [3]). Then we use the fine-tuning model on Composition-1k to test directly on Distinct-646. We can see that our DPT achieves good performance, especially in detailed regions.

6. More fine-tuning qualitative results on Natural human images

To show the robustness of our method, we use a fine-tuning model on Composition-1k to test on 2k resolution natural human images as shown in Fig. 9. From left to right, we provide the image, generated alpha matte α, and the synthetic foreground generated by α and image. We can see that our method performs well at extracting the boundary of the human body, such as the hair part.

References

[1] Yu Qiao, Yuhao Liu, Xin Yang, Dongsheng Zhou, Mingliang Xu, Qiang Zhang, and Xiaopeng Wei. Attention-guided hierarchical structure aggregation for image matting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13676–13685, 2020. 1

[2] Xinlong Wang, Rufeng Zhang, Chunhua Shen, Tao Kong, and Lei Li. Dense contrastive learning for self-supervised visual

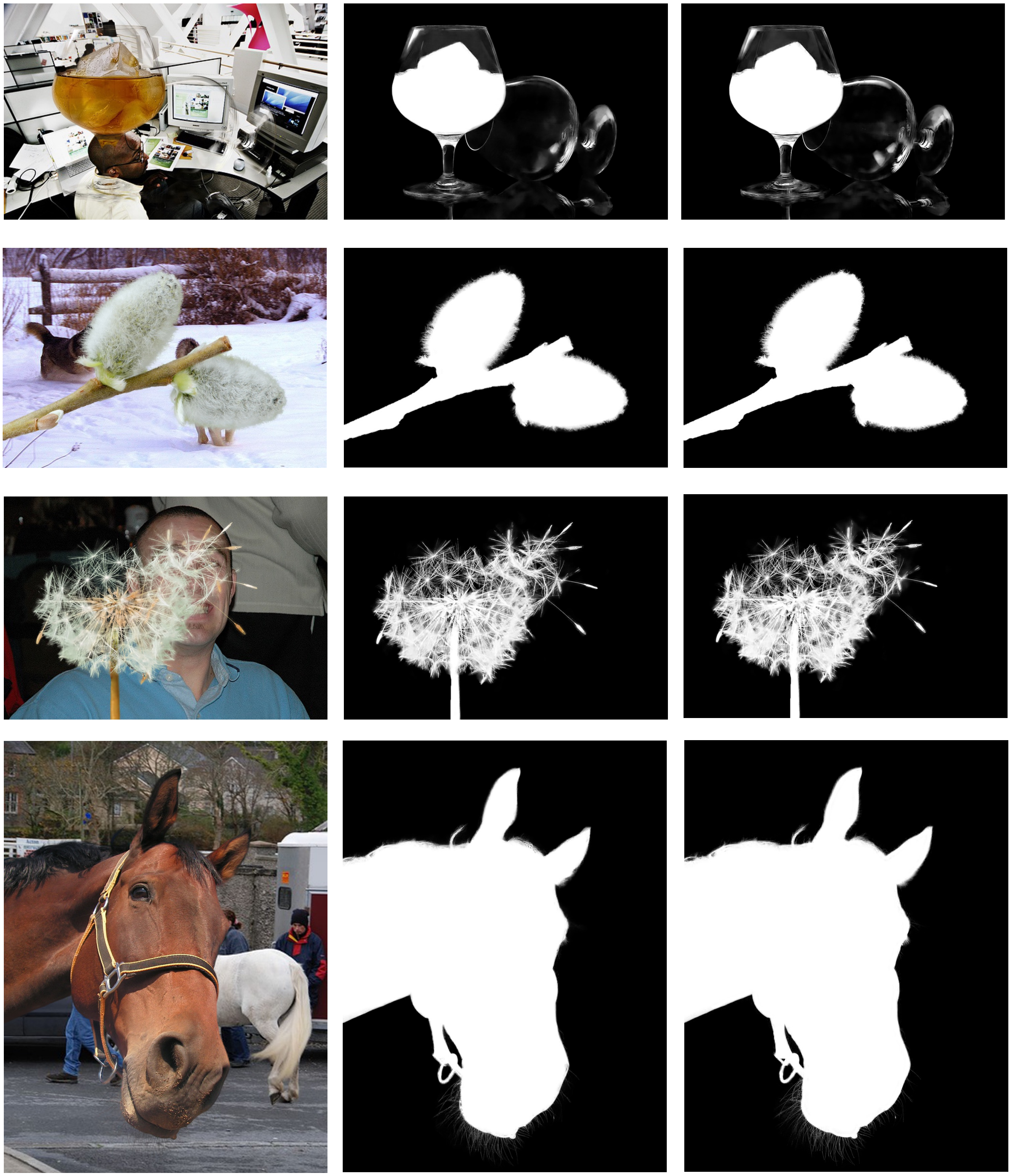

Image Ours GT

Figure 1. Qualitative results of our method on the Composition-1k test set.

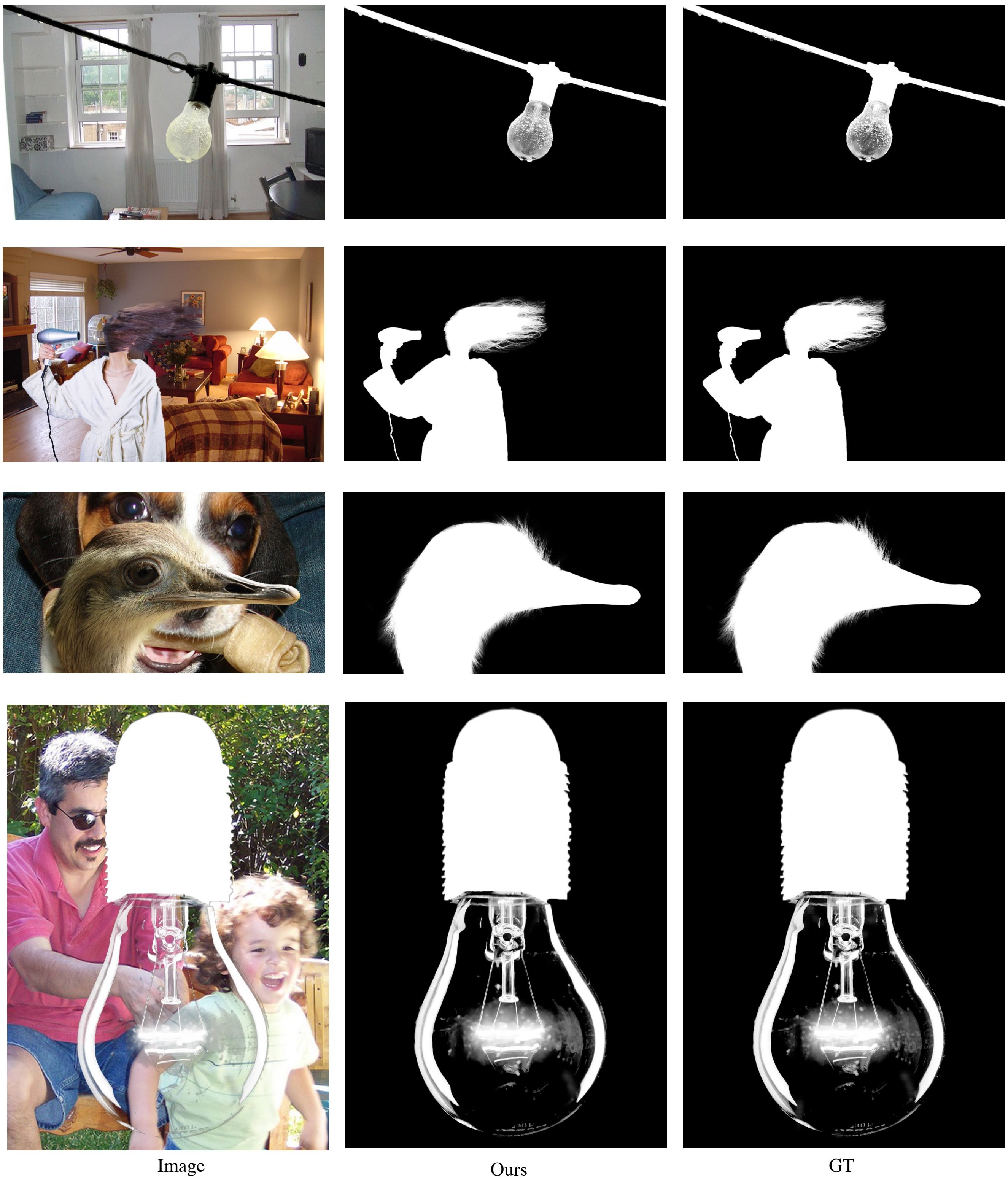

Figure 2. Qualitative results of our method on the Composition-1k test set.

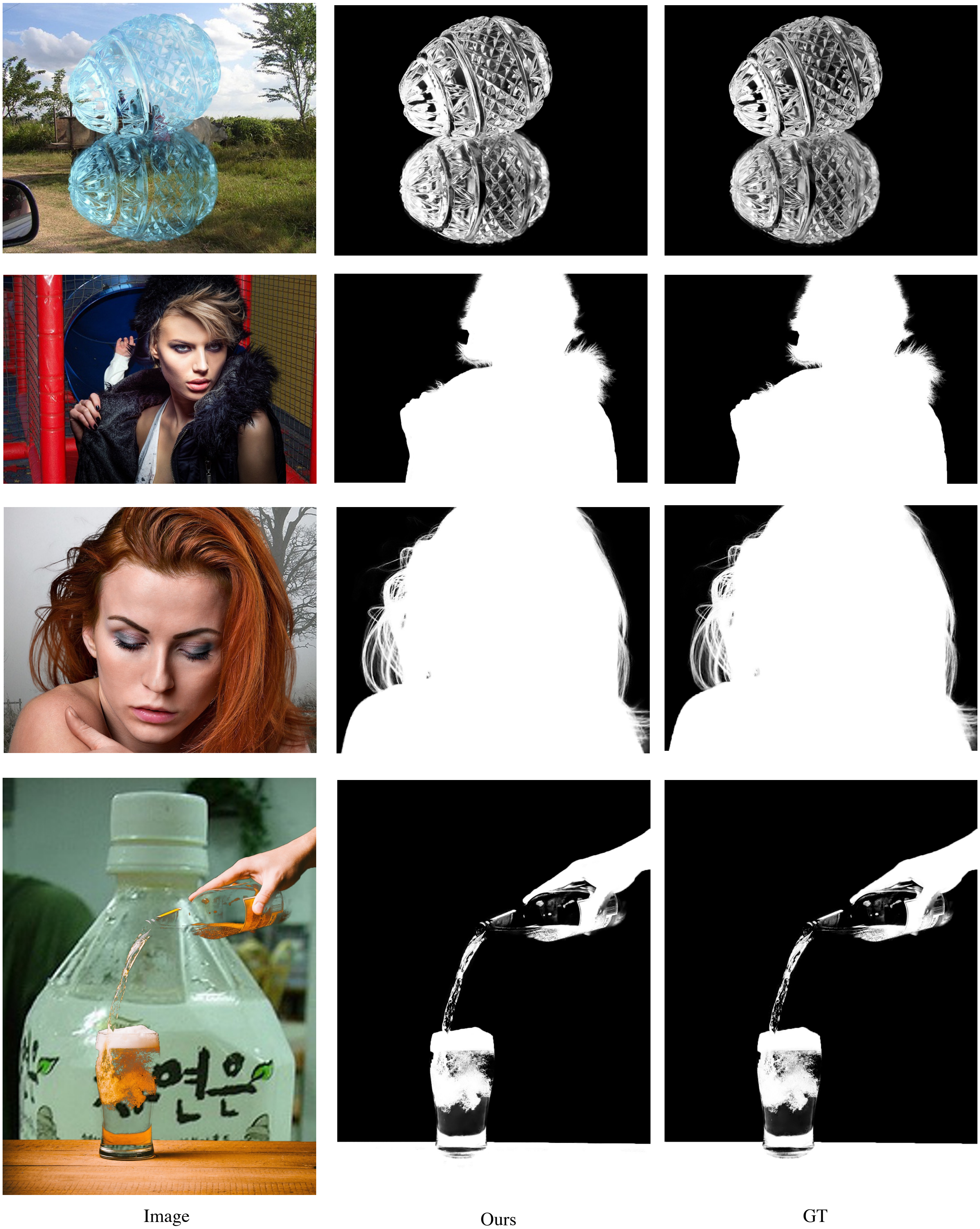

Figure 3. Qualitative results of our method on the Composition-1k test set.

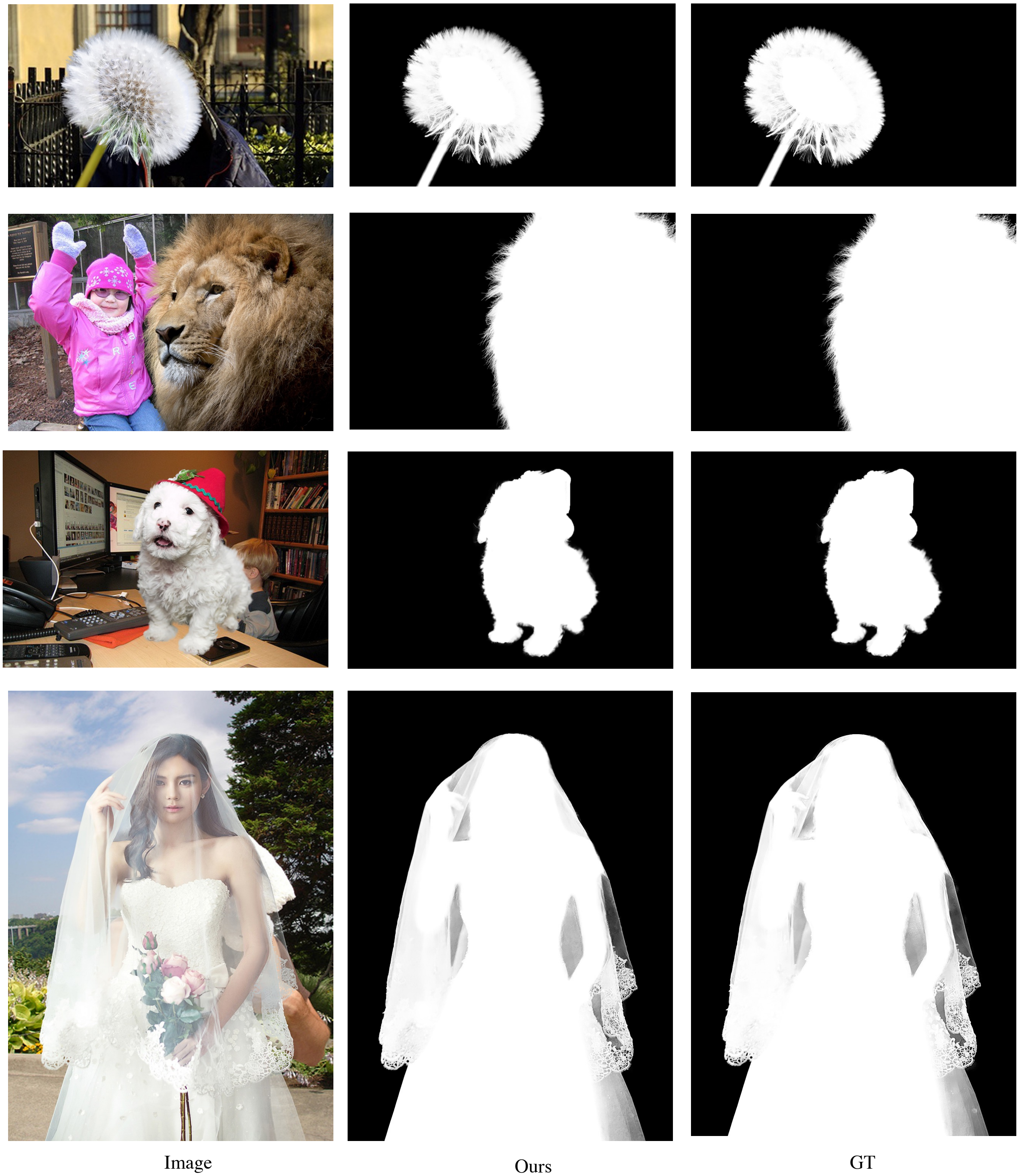

Figure 4. Qualitative results of our method on the Composition-1k test set.

pre-training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3024–3033,

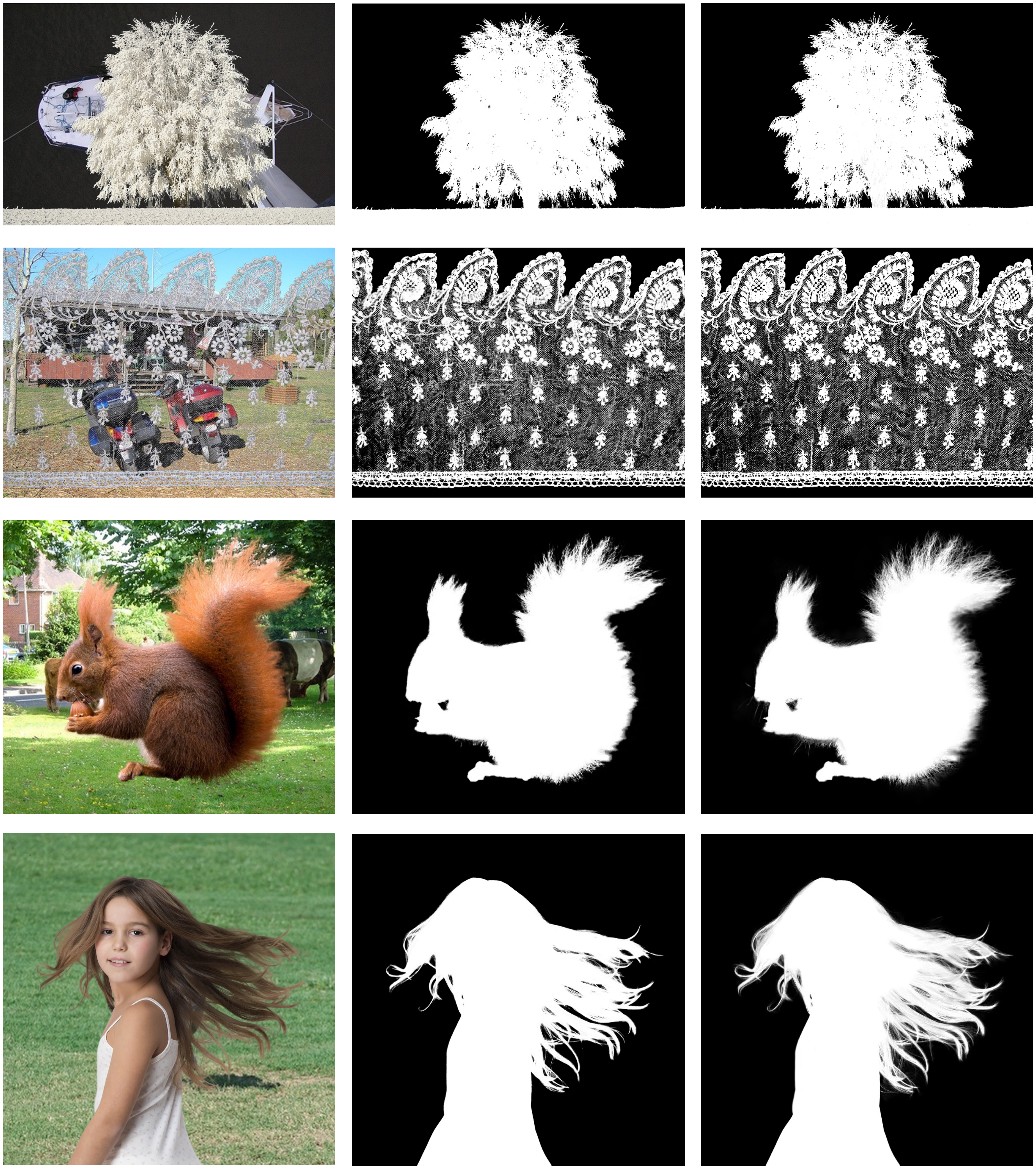

Image Ours GT

Figure 5. Qualitative results of our method on the Distinct-646 test set.

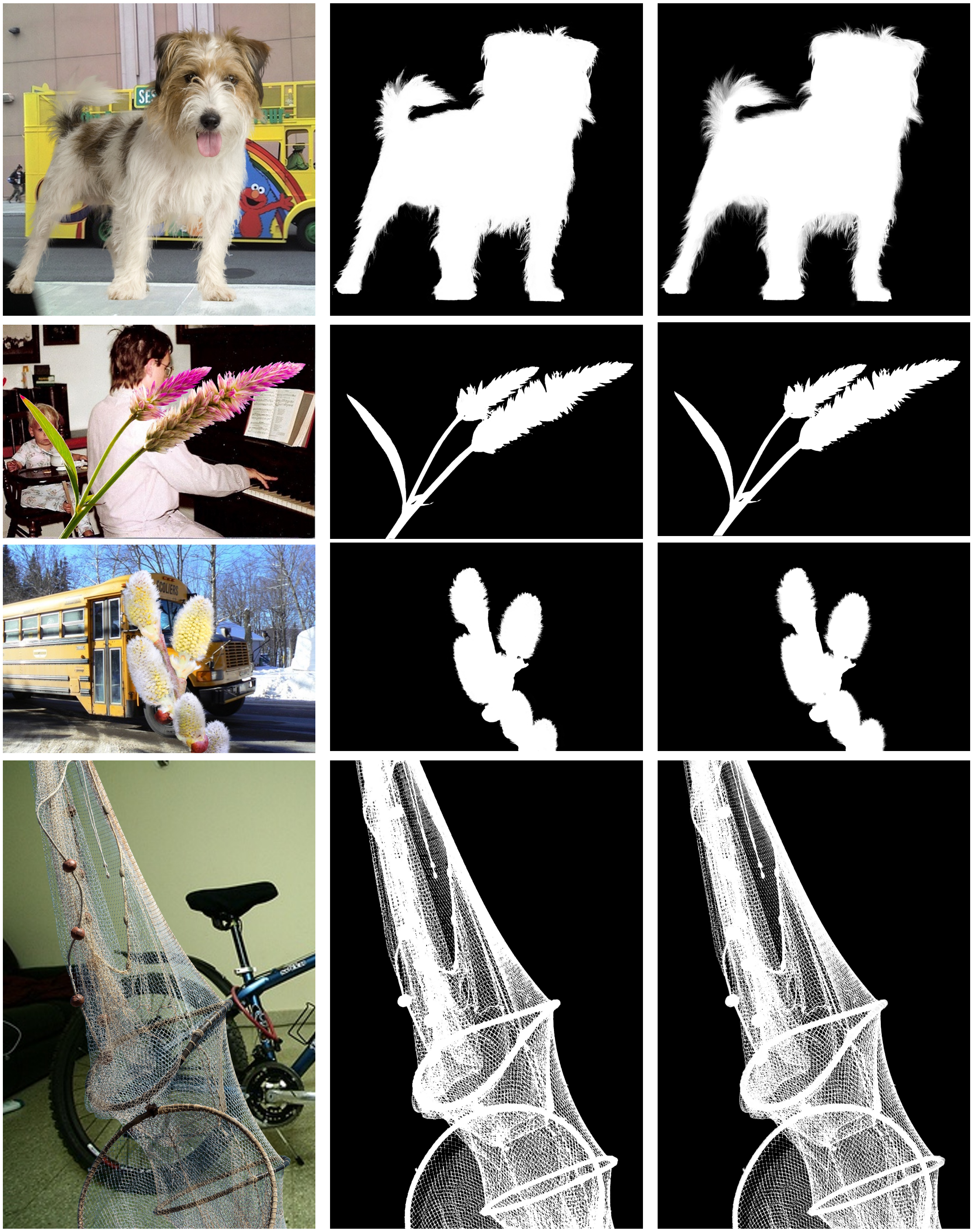

Image Ours GT

Figure 6. Qualitative results of our method on the Distinct-646 test set.

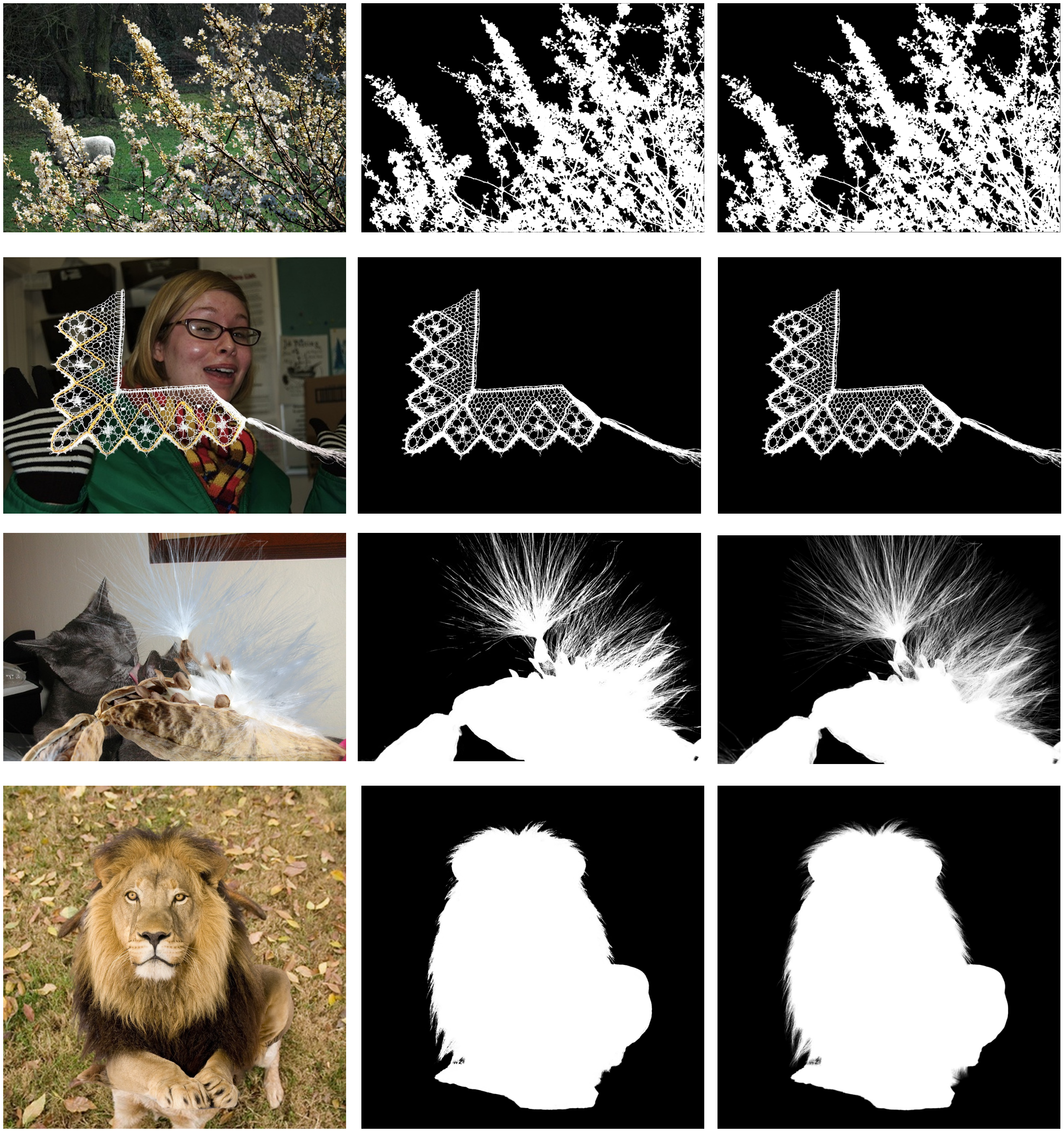

Figure 7. Qualitative results of our method on the Distinct-646 test set.

2021. 1

[3] Qihang Yu, Jianming Zhang, He Zhang, Yilin Wang, Zhe Lin, Ning Xu, Yutong Bai, and Alan Yuille. Mask guided matting via progressive refinement network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1154–1163, 2021. 1

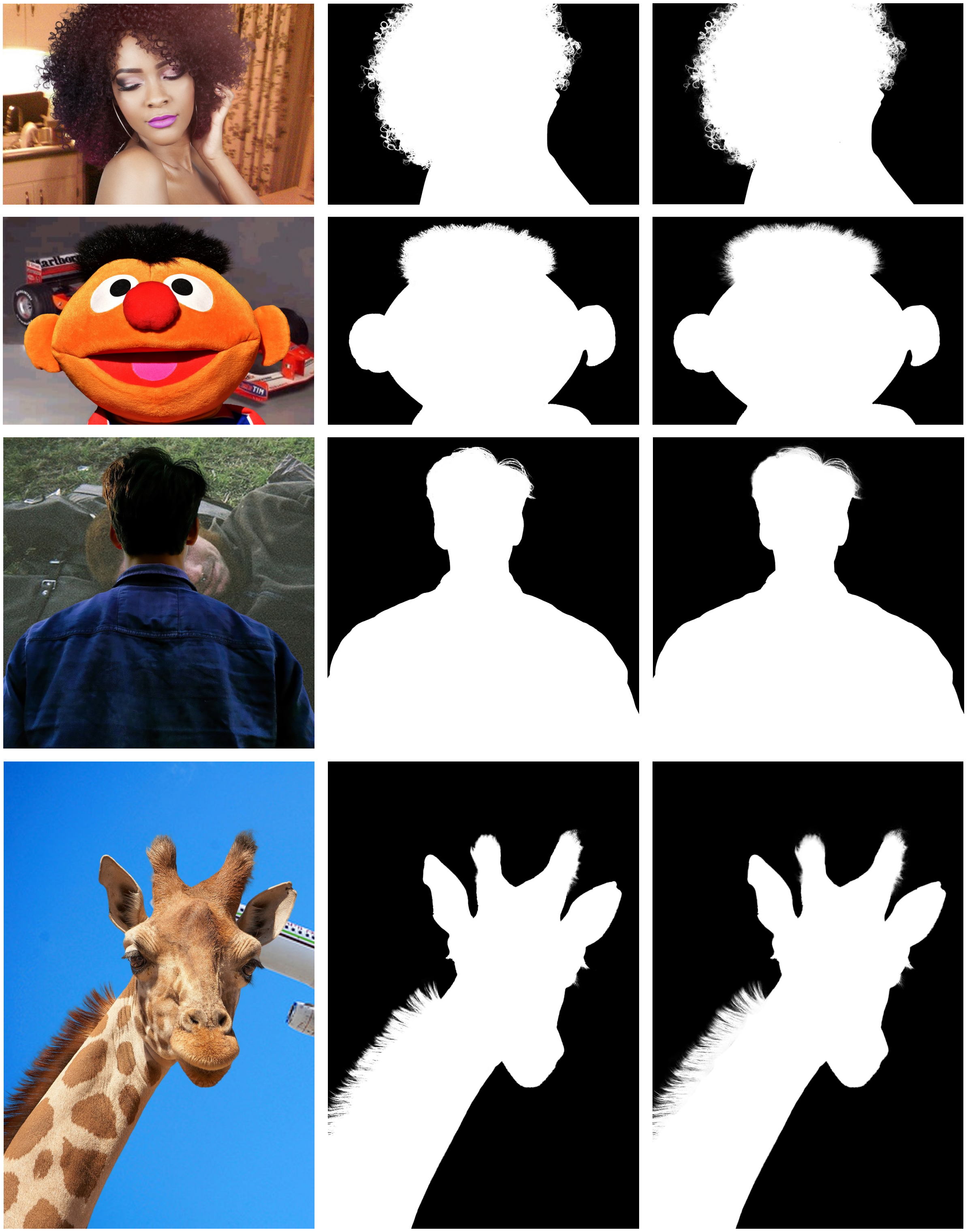

Image Ours GT

Figure 8. Qualitative results of our method on the Distinct-646 test set.

Image Ours Synthetic foreground

Figure 9. Qualitative results of our method on natural human images.