Learning Affinity-Aware Upsampling for Deep Image Matting Supplementary Material

Yutong Dai[1], Hao Lu[2], Chunhua Shen[1]
[1] The University of Adelaide [2] Huazhong University of Science and Technology

A. Network and Training Details of Image Reconstruction

We denote $C(k)$ to be a convolution layer with k-channel output and 3×3 filters (stride is 1 unless stated), followed by BatchNorm and ReLU, and denote D_r a downsampling operator with a ratio of r, and denote U_r an upsampling operator with a ratio of r. We build the network architecture as: $C(32)$-D_2-$C(64)$-D_2-$C(128)$-D_2-$C(256)$-$C(128)$-U_2-$C(64)$-U_2-$C(32)$-U_2-$C(1)$.

The image reconstruction experiments are implemented on the MNIST dataset [3] and Fashion-MNIST dataset [7]. They both include $60,000$ training images and $10,000$ test images. During training, the input images are resized to 32×32, and ℓ_1 loss is used. We use the SGD optimizer with an initial learning rate of 0.01. The learning rate is decreased by $\times 10$ at the 50-th, 70-th, and 85-th epoch, respectively. We update the parameters for 100 epochs in total with a batch size of 100. The evaluation metrics are Peak Signal-to-Noise Ratio (PSNR), Structural SIMilarity (SSIM), Mean Absolute Error (MAE) and root Mean Square Error (MSE).

B. Analysis of Complexity

Here we summarize the model complexity of different implementations of A^{2}U in Table 1. We assume that the encoding kernel size is $k \times k$, the upsampling kernel size is $s \times s$, and the channel number of feature map $\mathcal{X}$ is C. Since C is much larger than k and s, A^{2}U generally has the complexity: $\mathit{dynamic\ cw} > \mathit{hybrid\ cw} > \mathit{static\ cw} > \mathit{dynamic\ cs} > \mathit{hybrid\ cs} > \mathit{static\ cs}$.

Model	Type	# Params
static	cw	$4 \times s \times s + 2 \times k \times k \times C$
static	cs	$4 \times s \times s + 2 \times k \times k$
hybrid	cw	$4 \times s \times s \times C + 2 \times k \times k \times C$
hybrid	cs	$4 \times s \times s \times C + 2 \times k \times k$
dynamic	cw	$4 \times s \times s \times C + 2 \times C \times C$
dynamic	cs	$4 \times s \times s \times C + 2 \times C$

Table 1 – Analysis on the complexity of A^{2}U. 'cw': channel-wise, 'cs': channel-shared

C. Visualization of Upsampling Kernels

Here we visualize the learned upsampling kernel in a 'hybrid' model to showcase what is learned by the kernel. Two examples are illustrated in Fig. 1. We observe that, after learning, boundary details are highlighted, while flat regions are weakened.

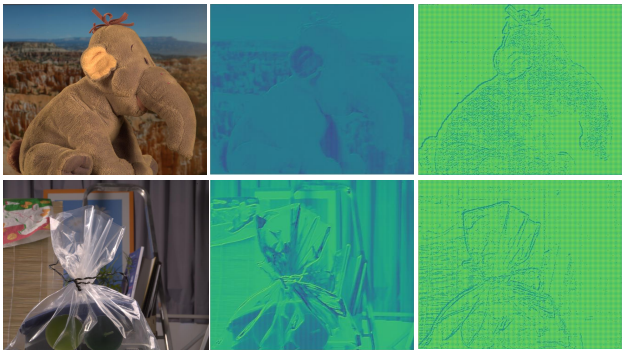

Figure 1 – Visualization of the upsampling kernel. The left is the randomly initialized kernel, and the right is the learned kernel.

D. Qualitative Results

We show additional qualitative results on the `alphamatting.com` benchmark [6] in Fig. 2. 4 top-performing methods are visualized here. Since all these methods achieve good performance, and their quantitative results on the benchmark are very close, it is difficult to tell the obvious difference in Fig. 2. It worth noting that, however, our method produces better visual results on detailed structures, such as gridding of the net, and leaves of the pineapple.

We also show qualitative results on the Distinction-646 test set [5] in Fig. 3. Since no implementation of other deep methods on this benchmark is publicly available, we only present the results of our baseline and our method here to

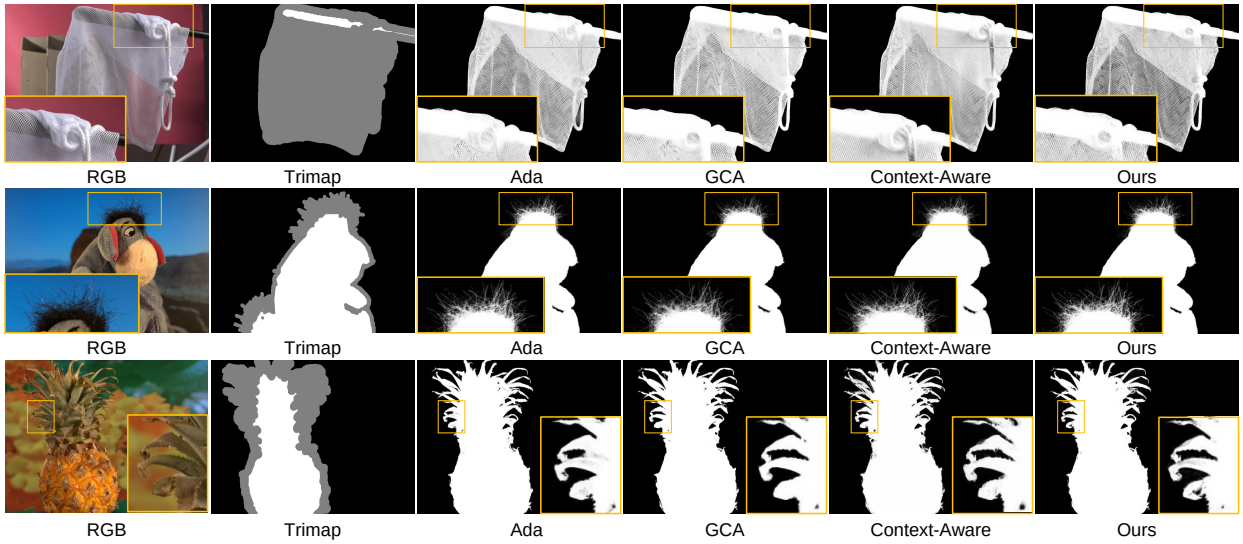

Figure 2 – Qualitative results on the `alphamatting.com` test set. The methods in comparison include AdaMatting [1], GCA Matting [4], Context-Aware Matting [2], and our method.

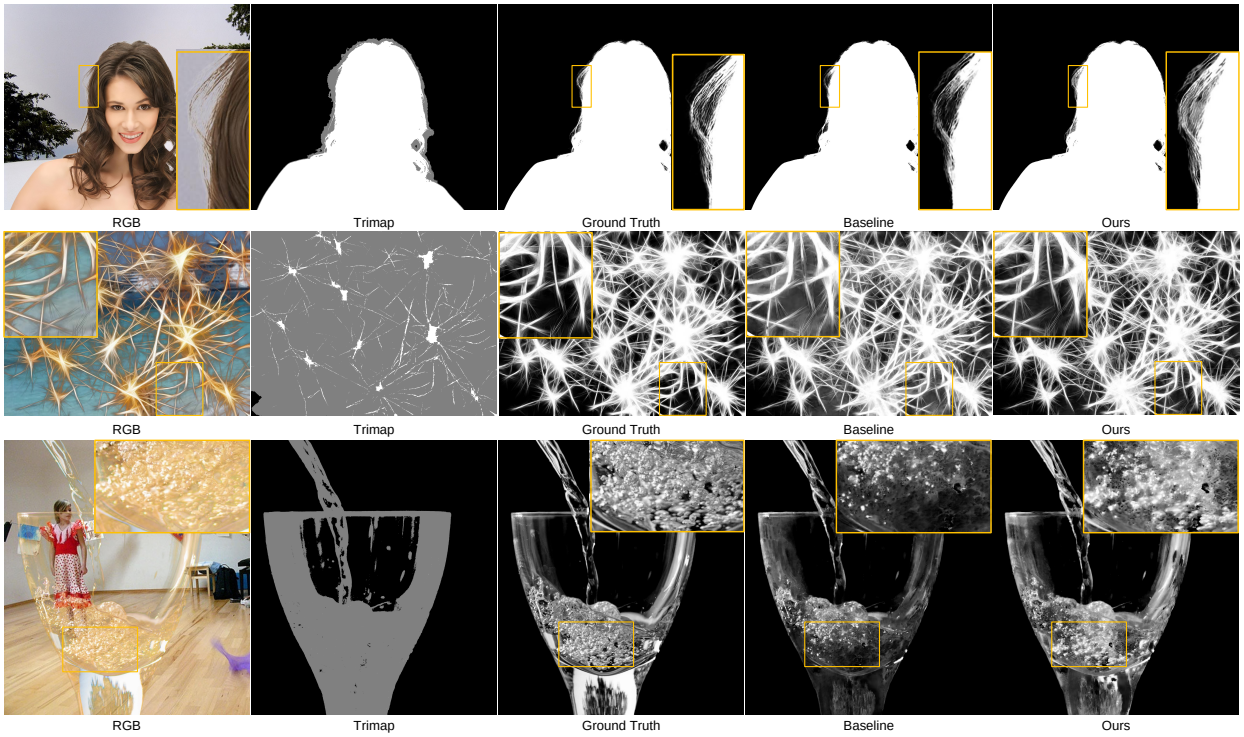

Figure 3 – Qualitative results on the Distinction-646 test set. The methods in comparison include the baseline and our method.

show the relative improvements. According to Fig. 3, our method produces clearly better predictions on highly transparent objects such as the bubbles.

References

[1] Shaofan Cai, Xiaoshuai Zhang, Haoqiang Fan, Haibin Huang, Jiangyu Liu, Jiaming Liu, Jiaying Liu, Jue Wang, and Jian Sun. Disentangled image matting. In *Proc. IEEE Int. Conf. Comp. Vis.*, pages 8819–8828, 2019. 2

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

[3] Yann LeCun. The mnist database of handwritten digits. *http://yann. lecun. com/exdb/mnist/*, 1998. 1

[4] Yaoyi Li and Hongtao Lu. Natural image matting via guided

contextual attention. In *Proc. AAAI Conf. Artificial Intell.*, volume 34, pages 11450–11457, 2020. 2

[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. Comp. Vis. Patt. Recogn.*, pages 13676–13685, 2020. 1

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

[7] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017. 1