Disentangled Image Mattings

Shaofan Cai[1*], Xiaoshuai Zhang[1,2*], Haoqiang Fan[1] Haibin Huang[1],
Jiangyu Liu[1], Jiaming Liu[1], Jiaying Liu[2], Jue Wang[1], and Jian Sun[1]
[1]Megvii Technology
[2]Institute of Computer Science and Technology, Peking University
{caishaofan, fhq, huanghaibin, liujiangyu, liujiaming, wangjue, sunjian}@megvii.com
{jet, liujiaying}@pku.edu.cn

1. Network Architectures

In this section, we present the network architectures of the AdaMatting in detail. The architecture of the encoder and the decoder is in Fig. 1. Shortcut connections are linked between two encoder-decoder modules of each line. The two decoders (for alpha estimation and trimap adaptation(shared the same architectures except for the output layers.same)

In order to prove the effectiveness of each modules (namely the sub-pixel convolutions and the global convolutions), we perform ablation experiments for these modules. The results are presented in Tab. 1. Obviously all these designed technique contributes to the network performance.

Table 1. Ablation study of each component. SP for sub-pixel convolutions, GC for global convolutions and PU for propagation unit. The gradient loss is scaled by 10^3.

Model	Grad	SAD	MSE
Ours-w/o-SP-w/o-GC-w/o-PU	25.18	51.45	0.0139
Ours-w/o-SP-w/o-PU	23.93	47.32	0.0124
Ours-w/o-GC-w/o-PU	20.77	45.71	0.0117
Ours-w/o-T-Decoder	21.50	46.68	0.0129
Ours-w/o-PU	17.86	44.13	0.0111
Ours	**16.89**	**41.70**	**0.0102**

2. More Analysis on Multi-Task Loss

We include more results for the multi-task loss in this section. The curve of σ_1, σ_2 of Eq. 4 during training is presented in Fig. 2. It can be observed from the figure that the two weights stably converges to a fixed weights (*i.e.* $\sigma_1 \approx 0.0995, \sigma_2 \approx 0.0878$). However, if we use this weights initially and fix them during training, the results are not as good. The quantitative results are in Tab. 2, where "D" represents for dynamically weighted, and "F" represents for using fixed $\sigma_1 = 0.0995, \sigma_2 = 0.0878$ Obviously. the dynamically weighted loss lead to better results. This phenomenon further indicates that dynamically adjust the importance for each task is of great help for the multi-task learning.

Table 2. Analysis of multi-task loss. AdaMatting-F is the model trained using fixed linearly combined loss ($\sigma_1 = 0.0995, \sigma_2 = 0.0878$), and AdaMatting-D is the model trained with dynamically weighted loss.

Model	Grad	SAD	MSE
AdaMatting-F	18.35	43.54	0.0118
AdaMatting-D	**16.89**	**41.70**	**0.0102**

3. More Qualitative Results

In this section, we present more results produced by the AdaMatting.

3.1. More Results on Composition-1k

Four results tested on the Composition-1k are presented in Fig. 3 and Fig. 4. As can be observed from these results, our AdaMatting generates more vivid details while clearly separating the foreground and background objects. Especially for the first image "lace", benefiting from the structural semantics learned from trimap adaptation, the AdaMatting could easily distinguish the foreground from the background with low-quality trimap inputs, while the DIM produces tainted alpha on the background.

3.2. More Results on Real Image

More results on real-world images are presented in Fig. 5. For clearer demonstration, we paste the extracted foreground onto a new background. Specially, the input trimap

*Equal Contributors. This work is supported in part by NSFC under grant #61271269 and #61321061.

Encoder			Decoder		
Component	Output Size	Filter Size	Component	Output Size	Filter Size
Conv1	320×320	2×2 max pooling	GC1_Skip	10×10	$\begin{bmatrix}5\times1,\ 512\\1\times5,\ 512\end{bmatrix}+\begin{bmatrix}1\times5,\ 512\\5\times1,\ 512\end{bmatrix}$
			Upsample1	20×20	1×1, 2048 Sub-pixel 2x
Res-2	80×80	$\begin{bmatrix}1\times1,\ 64\\3\times3,\ 64\\1\times1,\ 192\end{bmatrix}\times3$	GC2_Skip	20×20	$\begin{bmatrix}5\times1,\ 512\\1\times5,\ 512\end{bmatrix}+\begin{bmatrix}1\times5,\ 512\\5\times1,\ 512\end{bmatrix}$
			Upsample2	40×40	1×1, 2048 Sub-pixel 2x
Res-3	40×40	$\begin{bmatrix}1\times1,\ 128\\3\times3,\ 128\\1\times1,\ 384\end{bmatrix}\times4$	GC3_Skip	40×40	$\begin{bmatrix}5\times1,\ 256\\1\times5,\ 256\end{bmatrix}+\begin{bmatrix}1\times5,\ 256\\5\times1,\ 256\end{bmatrix}$
			Upsample3	80×80	1×1, 1024 Sub-pixel 2x
Res-4	20×20	$\begin{bmatrix}1\times1,\ 256\\3\times3,\ 256\\1\times1,\ 1024\end{bmatrix}\times6$	GC4_Skip	80×80	$\begin{bmatrix}5\times1,\ 128\\1\times5,\ 128\end{bmatrix}+\begin{bmatrix}1\times5,\ 128\\5\times1,\ 128\end{bmatrix}$
			Upsample4	160×160	1×1, 512 Sub-pixel 2x
Res-5	10×10	$\begin{bmatrix}1\times1,\ 512\\3\times3,\ 512\\1\times1,\ 1024\end{bmatrix}\times3$	GC5_Skip	160×160	$\begin{bmatrix}5\times1,\ 64\\1\times5,\ 64\end{bmatrix}+\begin{bmatrix}1\times5,\ 64\\5\times1,\ 64\end{bmatrix}$
			Upsample5	320×320	1×1, 256 Sub-pixel 2x
			Output	320×320	3×3, 1 + Sigmoid(alpha) 3×3, 3 + Softmax(Trimap)

Figure 1. Architecture of the Multi-task AutoEnocoder.

is genreated by portrait segmentation model followed by eroding the boundary for fixed pixels. As can be observed, our AdaMatting produces more natural and robust results compared to other state-of-the-arts.

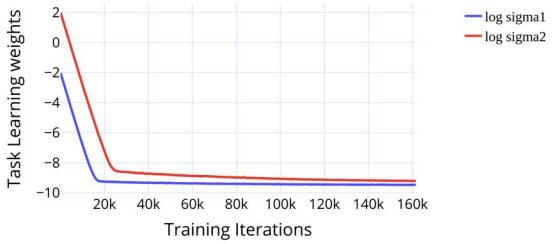

Figure 2. Training plots showing convergence of learning weights.

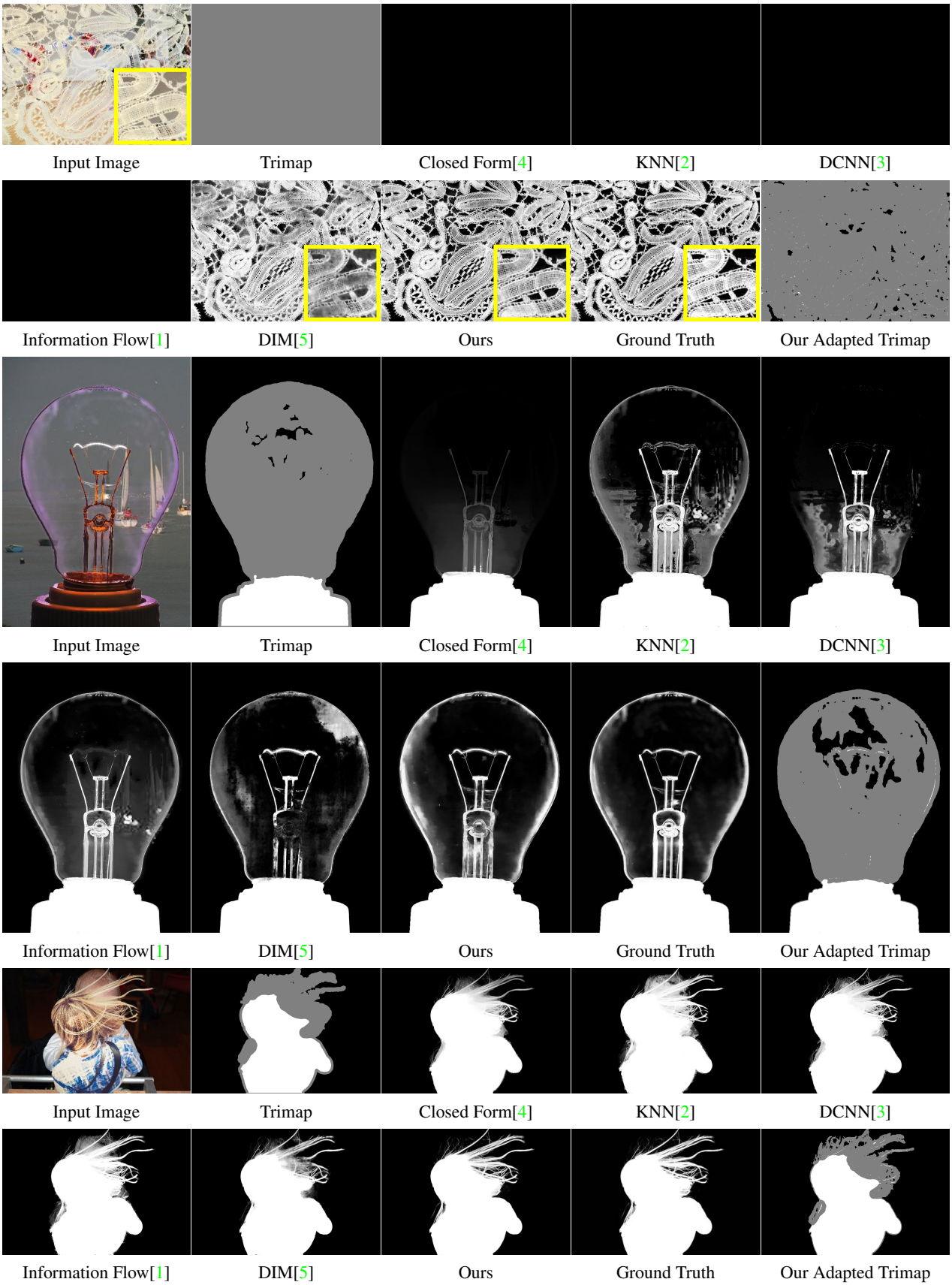

Figure 3. Qualitative comparisons on the Adobe Composition-1k test set.

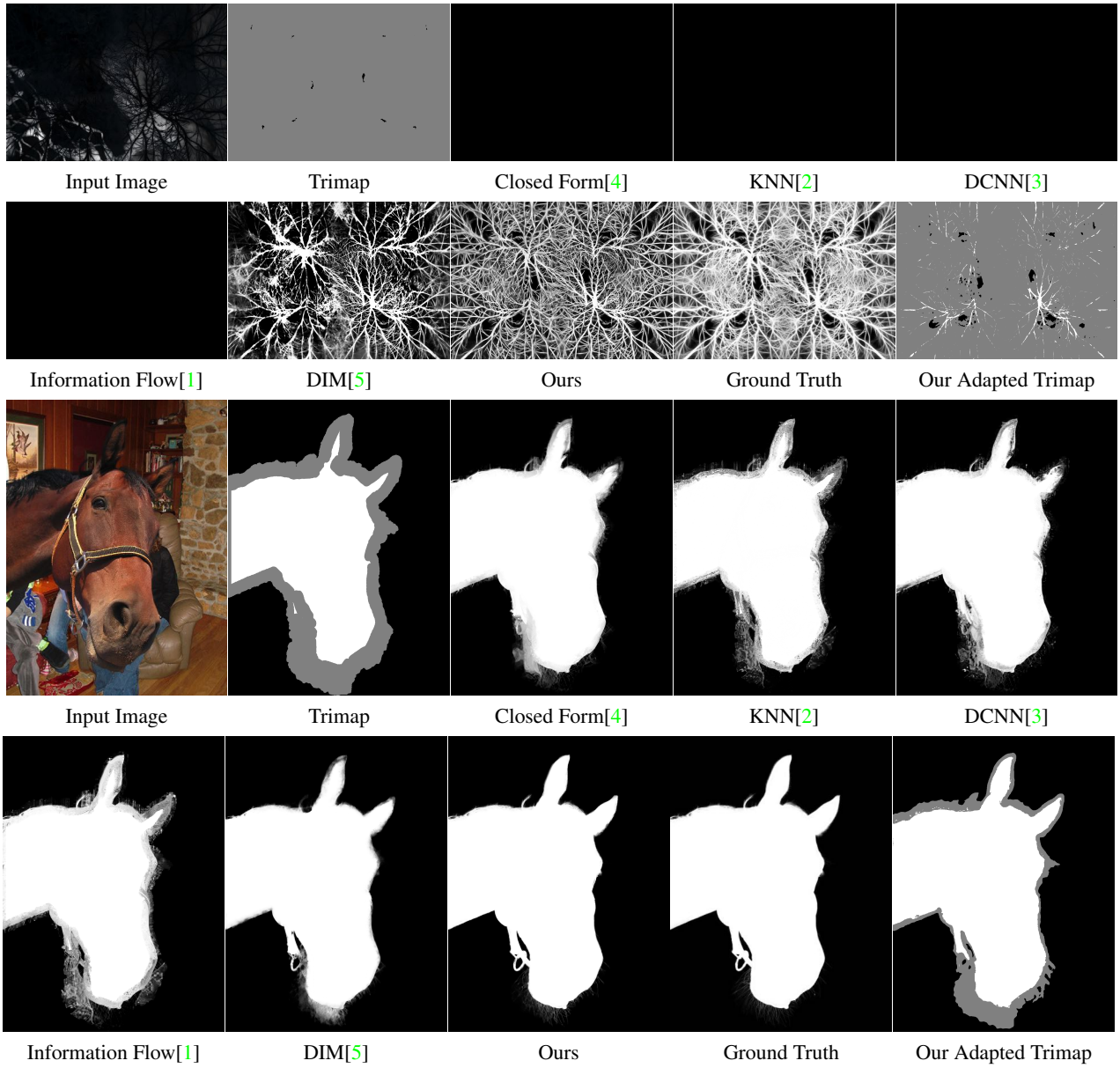

Figure 4. Qualitative comparisons on the Adobe Composition-1k test set.

Figure 5. Comparisons of background replacement on a real-world image.

References

[1] Yagız Aksoy, Tunç Ozan Aydın, Marc Pollefeys, and ETH Zürich. Designing effective inter-pixel information flow for natural image matting. In *Computer Vision and Pattern Recognition (CVPR)*, 2017. 3, 4

[2] Qifeng Chen, Dingzeyu Li, and Chi-Keung Tang. Knn matting. *IEEE transactions on pattern analysis and machine intelligence*, 35(9):2175–2188, 2013. 3, 4

[3] Donghyeon Cho, Yu-Wing Tai, and Inso Kweon. Natural image matting using deep convolutional neural networks. In *European Conference on Computer Vision*, pages 626–643. Springer, 2016. 3, 4

[4] Anat Levin, Dani Lischinski, and Yair Weiss. A closed-form solution to natural image matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):228–242, 2008. 3, 4

[5] Ning Xu, Brian Price, Scott Cohen, and Thomas Huang. Deep image matting. In *Computer Vision and Pattern Recognition (CVPR)*, 2017. 3, 4