Image Alpha Matting via Residual Convolutional Grid Network

Huizhen Zhang, Yang Zhou, Lei Chen, and Jiying Zhao

School of Electrical Engineering and Computer Science, University of Ottawa, Canada

Problem

Alpha matting aims at estimating the opacity information of the foreground objects (alpha matte) from a natural image. Mathematically, it can be expressed using:

$I_i = \alpha_i F_i + (1 - \alpha_i) B_i, \alpha_i \in [0, 1]$

Because only the RGB channels of I_i are known in the matting equation above, alpha matting is an inherently under-constrained problem with no exact solution.

Conventional matting methods solve matting problems mainly relying on color information, spatial position information, affinity between neighboring pixels and some uncertain assumptions.
They may fail to comprehend the semantic information of the images. Hence, when the foreground objects are transparent, semi-transparent, perforated or hairy, the matting results will decay.

Contributions

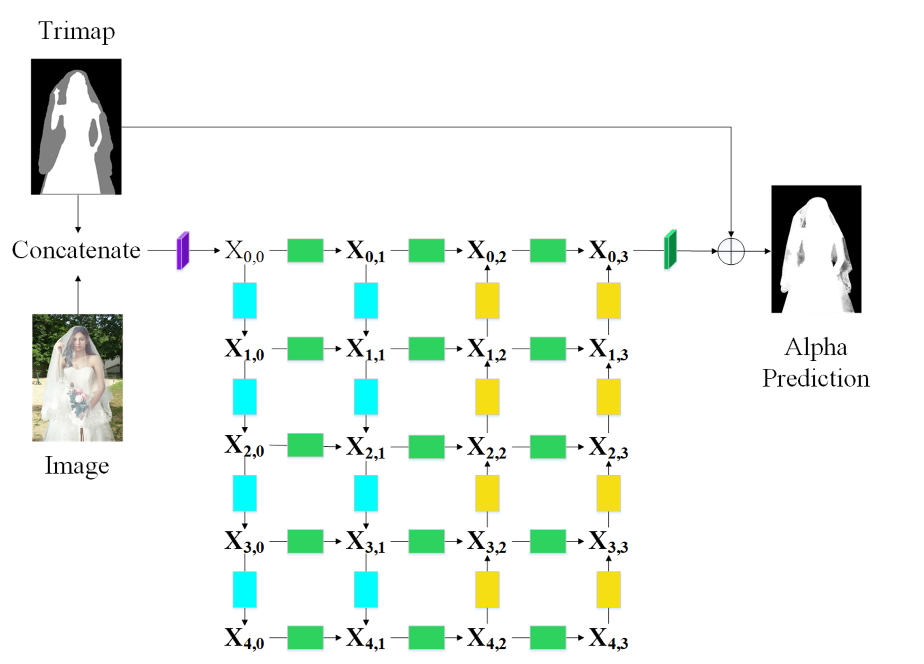
In this paper, we proposed a full convolutional neural network for alpha matting, which directly learns an end-to-end mapping, with little pre/post-processing beyond the optimization. There are two main contributions:

1. We introduced grid network to alpha matting for the first time. The matting results have demonstrated the superiority of this network.
2. The proposed matting method is more efficient than, and has the performance comparable to the best deep learning based matting method.

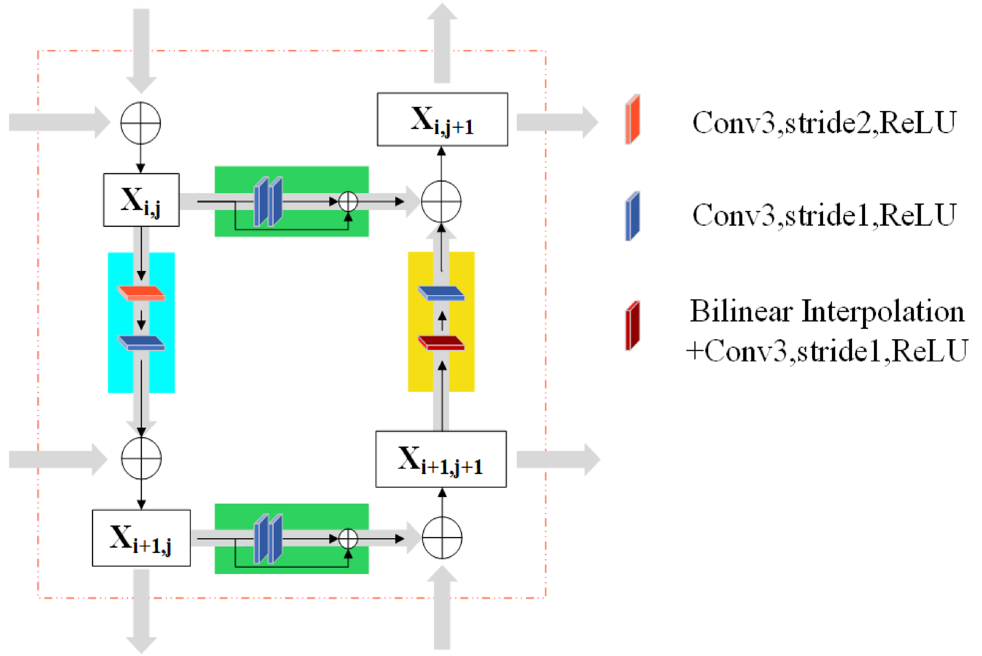
Proposed Method

A. Network architecture:

We introduce a residual convolutional grid network to realize an end-to-end matting system. The grid network with 5 rows and 4 columns performs the best under the current environment.



Overall network architecture of our method



Detailed configuration of our grid network

B. Matting dataset:

In our method we use the dataset provided by the deep image matting [1].

➢ Training dataset contains 43100 images, producing by compositing every one of the 431 unique foreground objects onto the 100 different background images from MS COCO.

➢ Testing dataset, which is called Composition-1k test set, is created by compositing every one of the 50 unique foreground images onto the 20 different background images from Pascal VOC.

C. Loss function:

We use a weighted sum of two different loss functions to train our network. The two loss functions are alpha-prediction loss and compositional loss, which are proposed by [1].

The alpha-prediction loss function is defined as:

$$L1^{alpha} = \sum_{i=1}^{n} \sqrt{\left(\alpha_{gt}^{i} - \alpha_{pre}^{i}\right)^2 + \sigma^2} \quad (\sigma=10^{-6})$$

The compositional loss function is defined as:

$$L1^{comp} = \sum_{i=1}^{n} \sqrt{\left(I_{gt}^{i} - I_{pre}^{i}\right)^2 + \sigma^2} \quad (\sigma=10^{-6})$$

The overall loss function is defined as:

$L1^{grident} = 0.5L1^{alpha} + 0.5L1^{comp}$

Experimental Results

Quantitative results of various methods on the Composition-1k dataset:

Our method outperforms the other matting methods and matches the deep image matting method (with refinement).

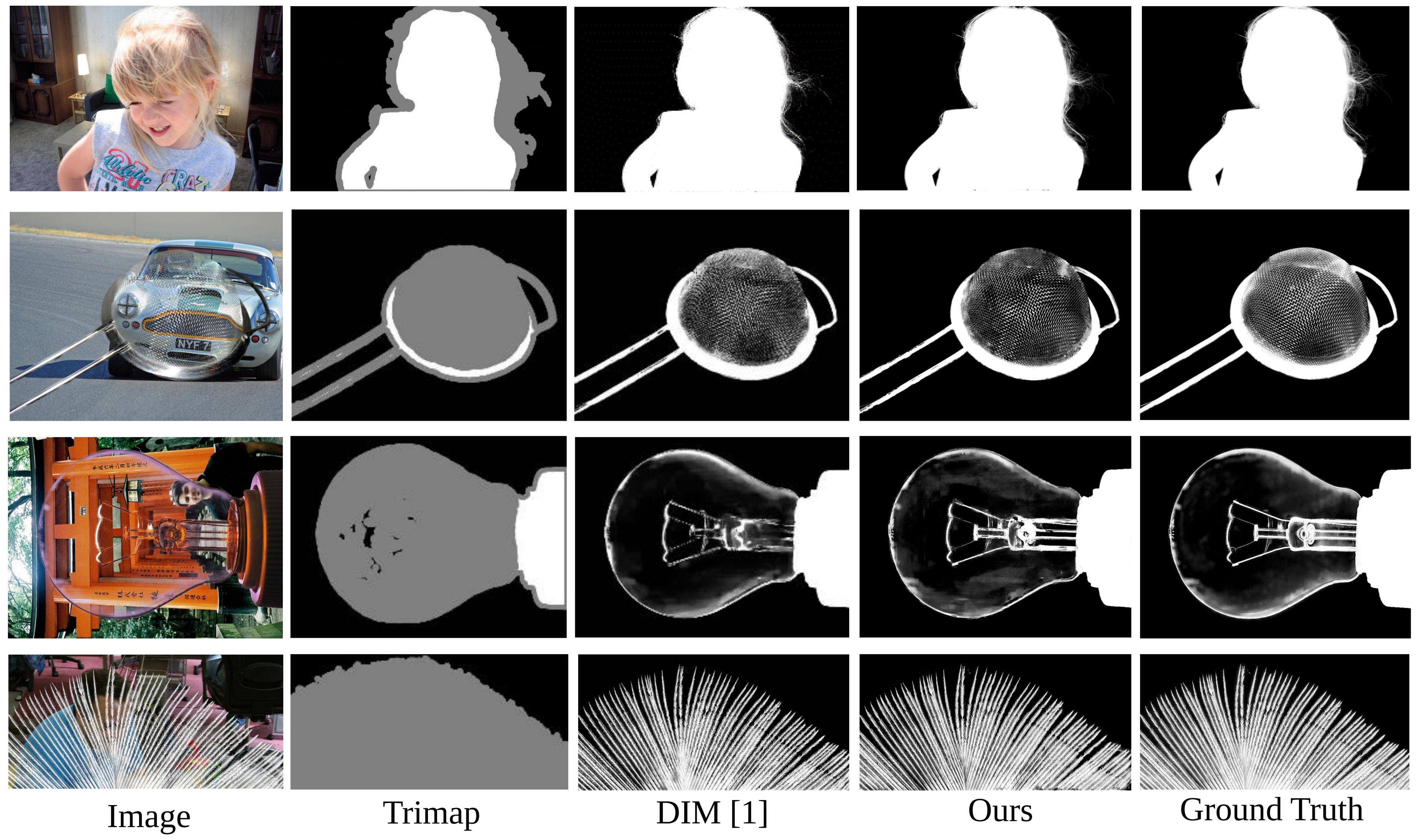
Deep image matting:
➢The best matting method in natural image matting.
➢49,300 images as training dataset.
➢26 million trainable parameters before adding refinement stage.

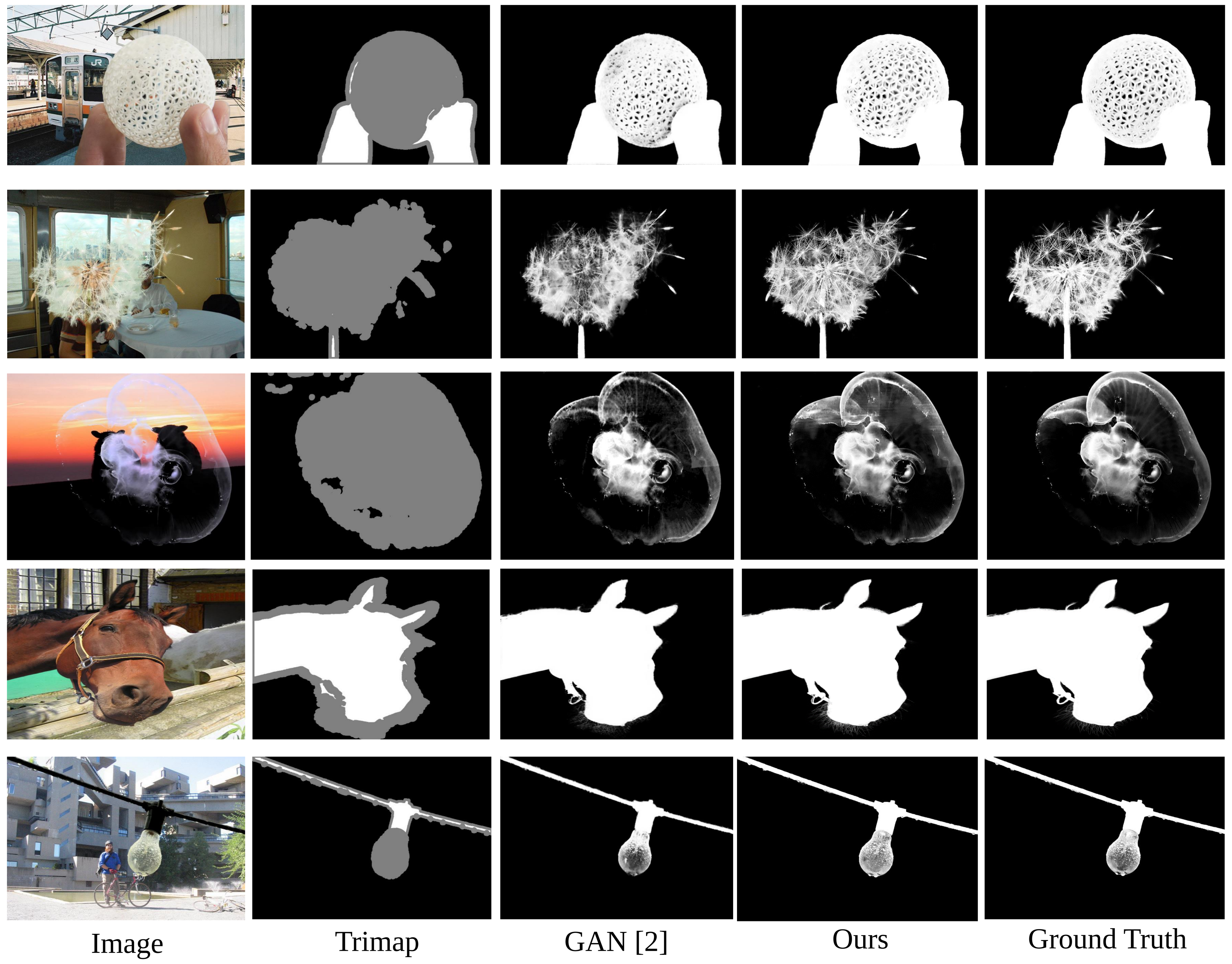
Our method:
➢43,100 images as training dataset.
➢8 million trainable parameters in the network.

The number of trainable parameters can affect computation complexity and memory usage. With the same experiment environment, the more the parameters the longer the training time. Our matting method trades off the computation complexity, memory usage and matting performance well.

Methods	SAD	MSE	Gradient(10^3)	Connectivity(10^3)
Shared Sampling Matting (SSM)	128.9	0.091	126.5	135.3
Comprehensive Sampling Matting (CSM)	143.8	0.071	102.2	142.7
Global Sampling Matting (GSM)	133.6	0.068	97.6	133.3
Closed-Form Matting (CFM)	168.1	0.091	126.9	167.9
KNN Matting (KNNM)	175.4	0.103	124.1	176.4
Learning Based Matting (LBM)	113.9	0.048	91.6	122.2
DCNN Matting (DCNNM)	161.4	0.087	115.1	161.9
GAN Matting (GANM)	68.8	0.032	51.4	60.2
Deep Image Matting (without refinement)	54.6	0.017	36.7	55.3
Deep Image Matting (DIM)	**50.4**	**0.014**	**31.0**	**50.8**
Ours	51.4	**0.014**	**31.0**	51.0

Visual results with top two deep learning based matting methods: the deep image matting [1] and the GAN matting [2]:





References

[1] N. Xu, B. Price, S. Cohen, and T. Huang, “Deep image matting,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017, pp. 311–320.
[2] S. Lutz, K. Amplianitis, and A. Smolic, “Alphagan: Generative adversarial networks for natural image matting,” in Proceedings of the British Machine Vision Conference (BMVC). 2018, p. 259, BMVA Press.