Instance Segmentation based Semantic Matting for Compositing Applications

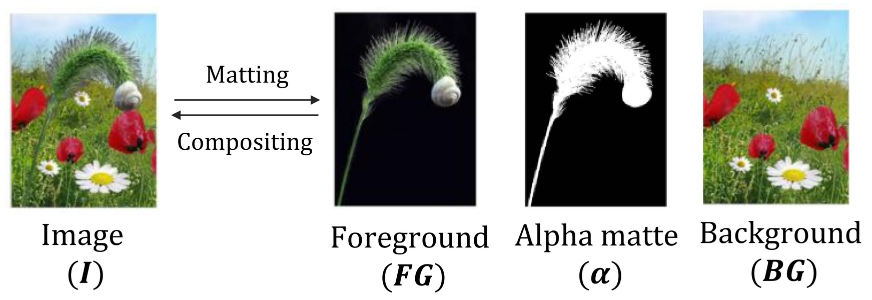
Guanqing (Bonnie) Hu and James J. Clark   Centre for Intelligent Machines, McGill University, Montréal, Canada

Introduction

Objective

Automatically generate **alpha mattes (α)** for detected objects in **natural scenes.**

An alpha matte is the opacity map of the target object and can be used in Image Compositing applications.



$$I = \alpha \cdot FG + (1 - \alpha) \cdot BG$$

Image Matting and Compositing illustration [1]

Our approach

- Our approach can be seen both as a **refinement** of existing **instance segmentation** algorithms and as a fully **automated semantic image matting** method.
- The proposed deep learning-based approach consists of three steps.
  - **First**, the input image is fed into a Mask R-CNN network [2] where the object bounding box and instance mask are generated.
  - **Then**, using these two results, a trimap is estimated for each detected object in the image.
  - **Finally**, the trimap and the original RGB image are used together as inputs to the Deep Image Matting network [3] to generate alpha matte.

Background

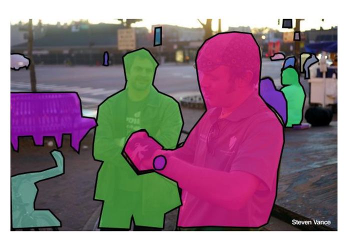
Chroma keying

Existing automatic image compositing approach (chroma keying/green screening) does not work in natural scenes.

Green screening
<https://www.youtube.com/watch?v=mUlFZ8wCnvM>

Instance Segmentation

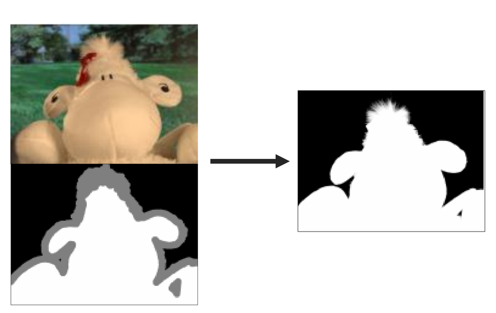
- Instance Segmentation can automatically detect and mask objects in natural scenes, but results lack boundary details.



Instance segmentation example
<http://cocodataset.org/#explore>

Natural Image Matting

- Natural Image Matting estimates an opacity map (alpha matte) for a target object in natural scene. Alpha matte segments the foreground object precisely. But high-quality Image Matting approaches are usually not automatic, requiring extra user input (e.g. trimap).



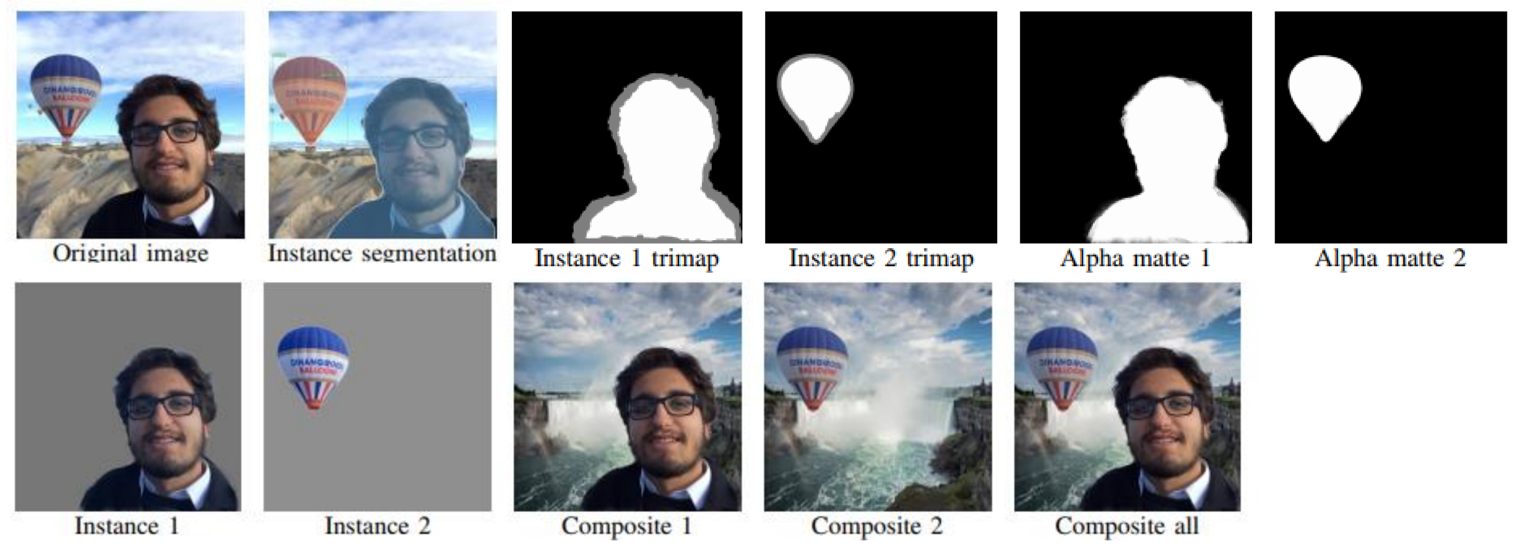
Trimap

Trimap is a user-labeled input indicating where the matting algorithm needs to estimates in grey color. A precise trimap is favored by the matting algorithm as it imposes stronger constraints.

Soft Segmentation

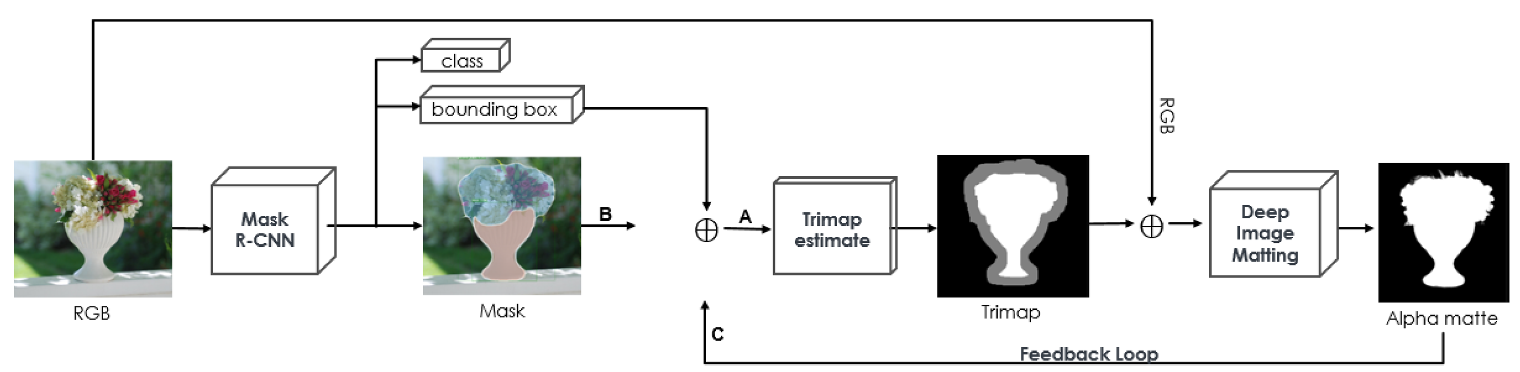
Soft Segmentation estimates more precise object boundary transactions than Instance Segmentation. Semantic Soft Segmentation [4] achieves good performance on automatic soft segmentation by using high-level semantic features to combine low-level texture and color features. However, it is not capable of instance-level segmentation.

Quick Demo



- Our approach first generates coarse instance masks which are used to create trimaps.
- Then, the trimaps and the original image are fed to the Image Matting network to produce an alpha matte for each object.
- Finally, the alpha mattes are used for image compositing task.

Methodology



Instance Segmentation Stage

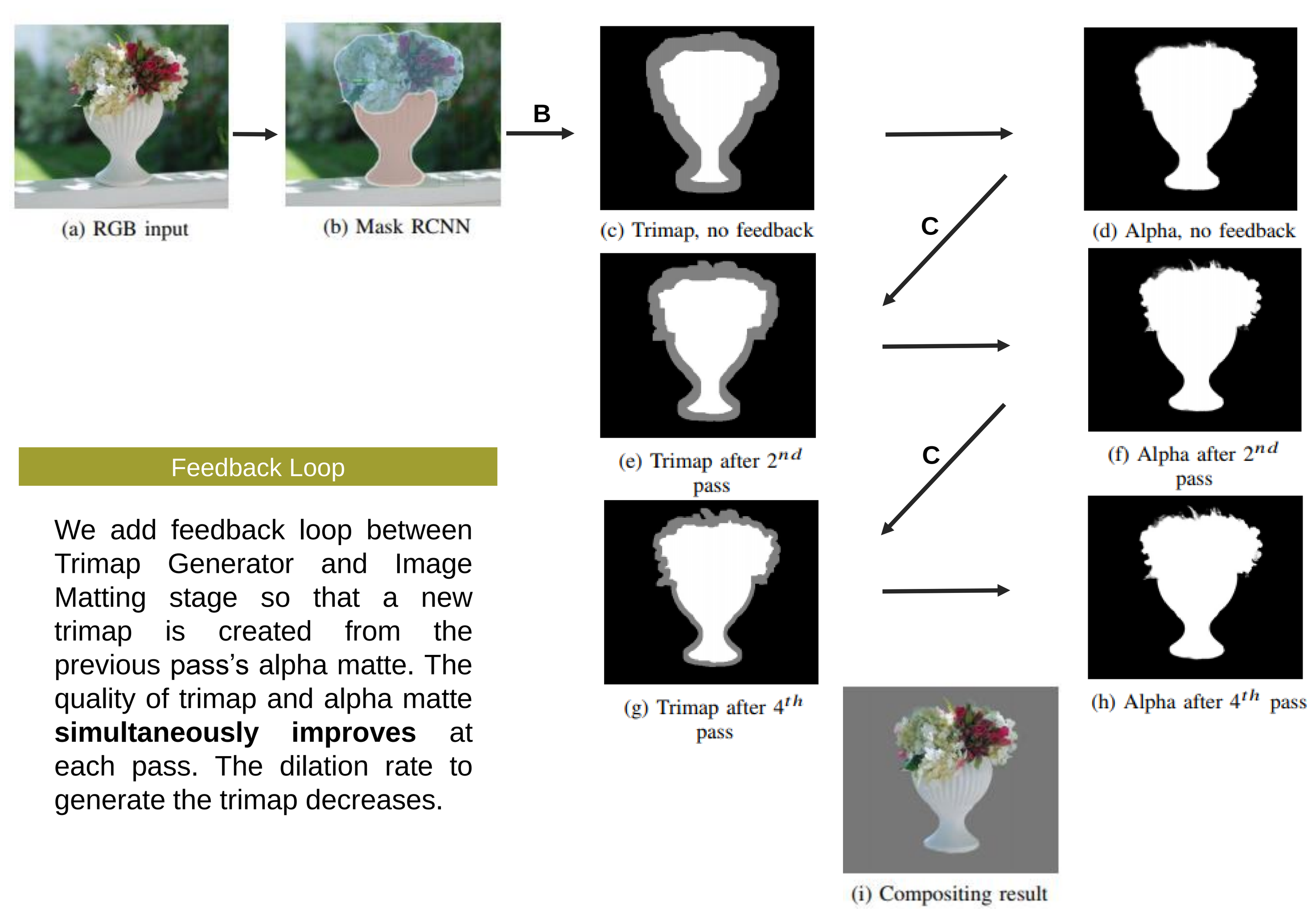
Mask-RCNN is used to generate solid instance segmentation mask, object class, and bounding box.

Trimap Generator

A certain region dilated from the segmentation mask or alpha matte is defined as the unknown area for the next stage to work on.

Image Matting Stage

Deep Image Matting [3] is used in this stage for its stable performance on low-quality trimaps and color similarities.

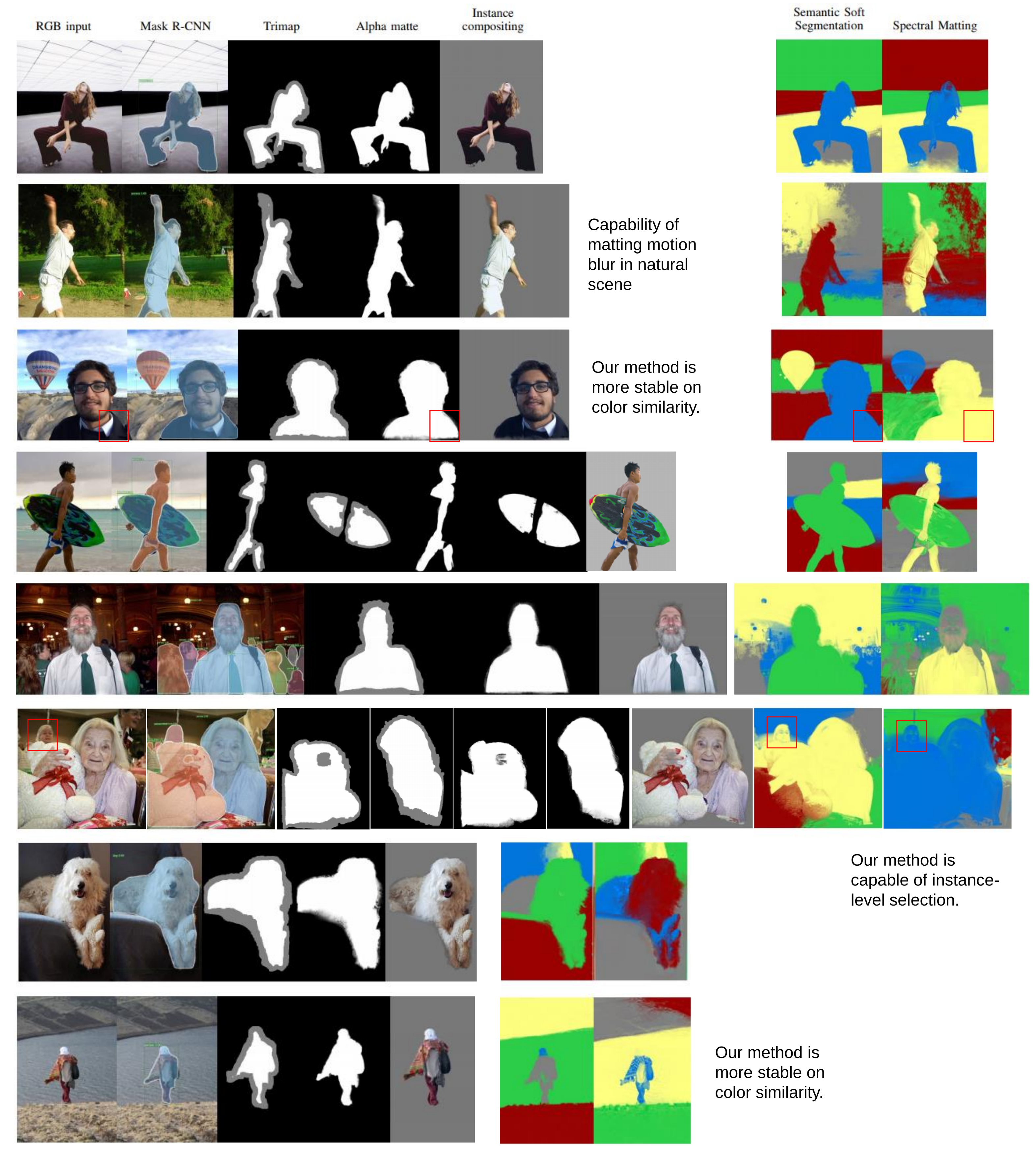


(a) RGB input (b) Mask RCNN (c) Trimap, no feedback (d) Alpha, no feedback (e) Trimap after 2^{nd} pass (f) Alpha after 2^{nd} pass (g) Trimap after 4^{th} pass (h) Alpha after 4^{th} pass (i) Compositing result

Feedback Loop

We add feedback loop between Trimap Generator and Image Matting stage so that a new trimap is created from the previous pass's alpha matte. The quality of trimap and alpha matte **simultaneously improves** at each pass. The dilation rate to generate the trimap decreases.

Experimental Analysis



Conclusion

Our approach extends automatic image compositing techniques to scenes with complex natural backgrounds without the need for any kind of user interaction. Our approach performs as good as existing approaches with an advantage in instance-level selections. Our approach could help non-experts in image compositing tasks and accelerate the image editing process.

Acknowledgements

The authors would like to thank the support of Digital District Canada, especially from Dr. Jonathan Bouchard, and the scholarship support from McGill University for GH. Also thanks go to Nvidia for the donation of GPU boards for this project, and to our colleague Ibtihel Amara for her comments on the paper.

References

[1] Q. Zhu, P. A. Heng, L. Shao, and X. Li, "What's the role of image matting in image segmentation?," *2013 IEEE Int. Conf. Robot. Biomimetics, ROBIO 2013*, no. December, pp. 1695–1698, 2013.
[2] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," Proc. IEEE Int. Conf. Comput. Vis., vol. 2017-Octob, pp. 2980- 2988, Mar. 2017.
[3] N. Xu, B. Price, S. Cohen, and T. Huang, "Deep Image Matting," Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 311-320, Mar. 2017.
[4] Y. Aksoy, T.-H. Oh, S. Paris, M. Pollefeys, and W. Matusik, "Semantic Soft Segmentation," ACM Trans. Graph., vol. 37, no. 4, pp. 1-13, 2018.