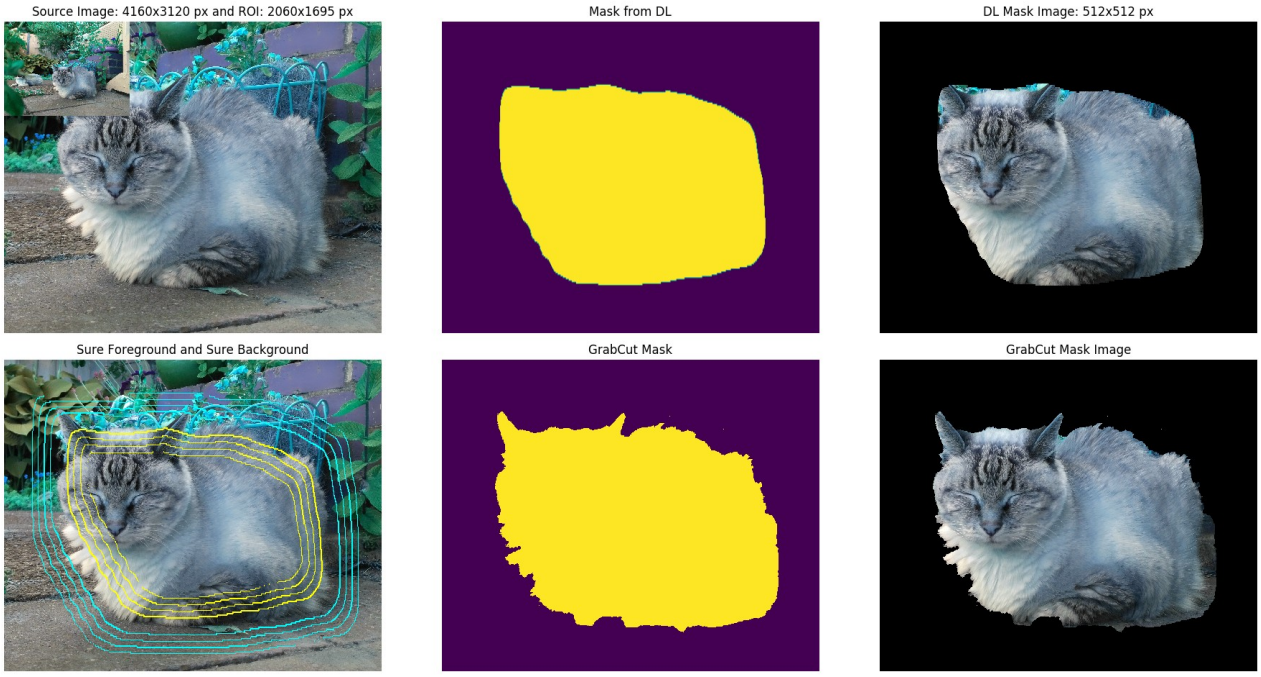
“DeepCat” - Cat Foreground Extraction Using MRCNN and GrabCut

Stephen Meschke



Abstract

The problem of fully automated cat/background segmentation in high-resolution still images is of great practical importance in image editing. Classical image segmentation tools like GrabCut work well, but require manual annotation of sure foreground and sure background. Recently, an approach based on deep learning has been developed which successfully generates segmentation masks of cats. This deep learning approach is called MRCNN. The masks are not pixel perfect, but the masks provide insight into what is surely the foreground and what is surely the back ground. I show how to combine MRCNN segmentation masks and the GrabCut algorithm to achieve fully automated pixel-perfect cat/background segmentation.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation – Display algorithms; I.4.6 [Image Processing and Computer Vision]: Segmentation - Cat classification;

Keywords: Image Segmentation, GrabCut, MRCNN, Image Editing, Foreground Extraction, Alpha Matting, Cats

1 Introduction

This paper addressees the problem of fully automated cat/background segmentation in a complex environment whose background cannot be trivially subtracted. The resulting foreground object (the cat) is an alpha-matte which reflects the proportion of the cat and the background. The aim is to achieve high performance with no manual annotation from the user. High performance in this task includes: accurate segmentation of cat from background, subjectively convincing alpha values, in response to blur, mixed pixels and transparency; clean foreground color, and free of color bleeding from the source background to the cat.

1.1 Previous approaches to cat/background segmentation

In the following I briefly describe the most common cat/background segmentation tools. Each has it's strengths and weaknesses. Before these segmentation tools were developed, segmentation masks of cats were obtained manually using techniques like Magic Wand, Intelligent Scissors, Bayes Matting, or Knockout 2.

GrabCut was developed by Carsten Rother, Vladimir Kolmogorov, and Andrew Blake at Microsoft Research Cambridge, UK in the 2000's. GrabCut requires the user to label parts of the image that are surely cat and surely background. The algorithm then applies iterated graph cuts to find the border between cat and background. The GrabCut algorithm works better at higher resolutions.

MRCNN is a model that generates bounding boxes and segmentation masks for cats in an image. It's based on Feature Pyramid Network (FPN) and a ResNet101 backbone. This model was trained on the COCO dataset which has over 330K images. MRCNN generates the masks without any manual annotation, but it is limited to a resolution of 512 px square.

1.2 Proposed system: Deep Cat

Ideally, a matting tool should be able to produce continuous alpha values over the entire inference region of the unsure pixels, without any hard constraint that the alpha values may only be 0 or 1. In that way, problems involving whiskers, hair, tail, etc., could be dealt with appropriately in an automatic way.

First, using the MRCNN, I obtain a 'general location' segmentation and bounding box. This tells me where the cat is in the image. I am only interested in the cat, so the bounding box containing the cat is my region of interest. After the ROI of the cat is found, the the MRCNN again finds the segmentation of the roi.

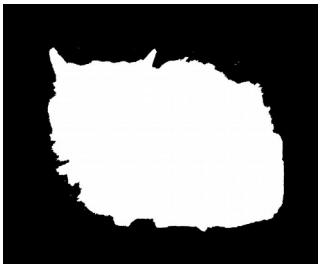
The mask from the MRCNN is then resized to match the size of the source image. The sure background is obtained by making the mask bigger using the morphological transformation dilation. The mask pixels that for sure belong to the cat is obtained by eroding the mask from the MRCNN. The areas that are neither sure foreground or sure background are left to the GrabCut algorithm to be segmented.

Finally, the GrabCut algorithm takes the sure-cat and sure-background masks generated by the MRCNN and finds that pixel-purrfect alpha mask.

MRCNN First Pass

MRCNN Second Pass

Application of GrabCut



2. Comparison

In this section I compare Deep Cat to the MRCNN results.

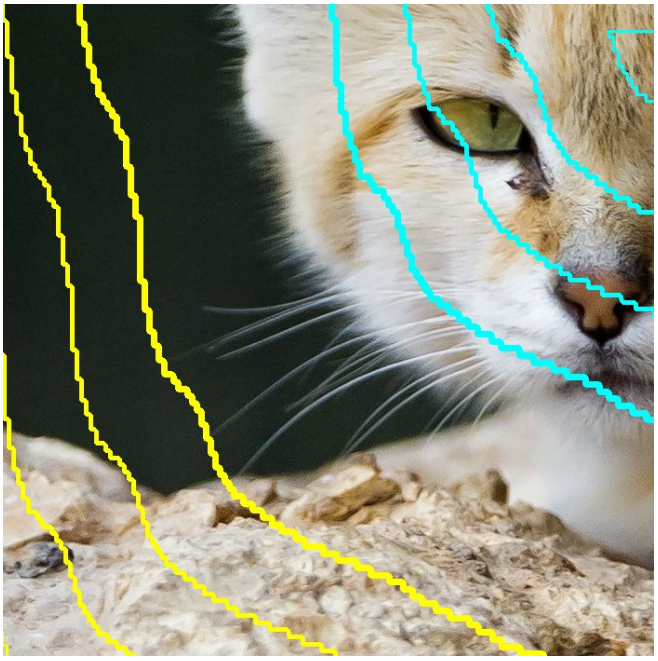
2.1 Disadvantages of Deep Cat

Speed. Each pass of the MRCNN takes 7 seconds. Running with 4 iterations, the GrabCut algorithm takes 12 seconds.

2.1 Key Advantages of Deep Cat

Deep Cat gets fine detail that makes cat identification certain. Whiskers and ears are two features that are distinct to cats. The MRCNN tends to smooth out the ears. Whiskers are never distinct in MRCNN masks.

This image shows a part of a cat's face. The yellow contours mark areas that are sure background. The cyan contours mark areas that are sure foreground. Those areas were obtained by dilating and eroding the mask from the MRCNN.

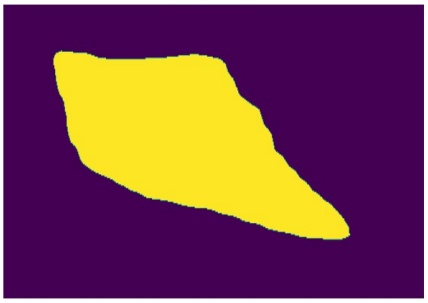


This image shows the same area of detail in the mask generated by the MRCNN. Notice how the MRCNN mask doesn't show any distinct whiskers:

Applying the GrabCut algorithm preserves the detail in the whiskers:

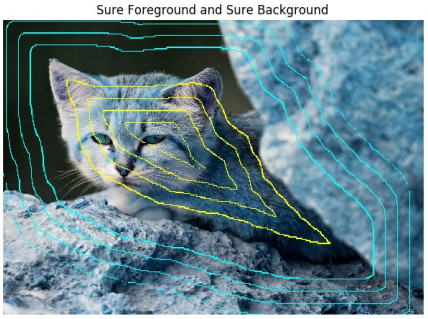
Source Image: 3938x2625 px and ROI: 2585x1808 px

Mask from DL



DL Mask Image: 512x512 px

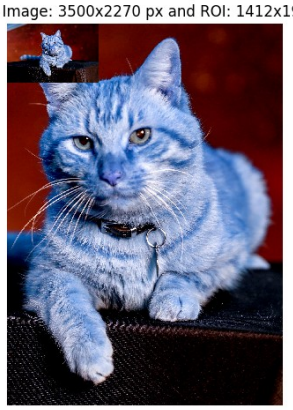
Sure Foreground and Sure Background



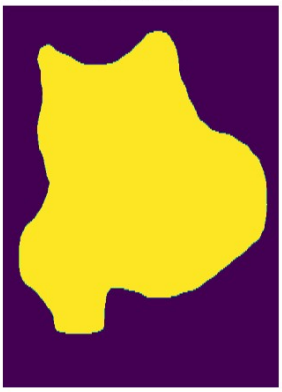
GrabCut Mask

GrabCut Mask Image

Source Image: 3500x2270 px and ROI: 1412x1953 px

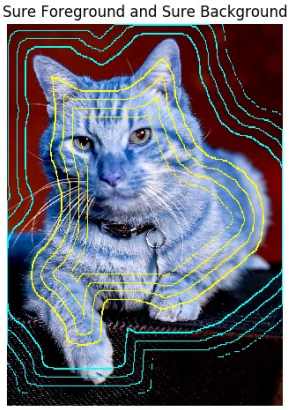


Mask from DL



DL Mask Image: 512x512 px

Sure Foreground and Sure Background



GrabCut Mask

GrabCut Mask Image

Acknowledgements. Thanks to everyone who helps others because they have interesting problems.

References

matterport, MRCNN, 2019, https://github.com/matterport/Mask_RCNN

Rosebrock Adrian, Keras Mask R-CNN, 2019,
https://www.pyimagesearch.com/2019/06/10/keras-mask-r-cnn/

Rother, Kolmogorov, Blake, “GrabCut”, 2000’s,
https://cvg.ethz.ch/teaching/cvl/2012/grabcut-siggraph04.pdf

Bradsky Gary, OpenCV, 1999,
https://docs.opencv.org/trunk/d6/d00/tutorial_py_root.html

Szeliski, Computer Vision, 2010,
http://szeliski.org/Book/