Visual Recognition

Udith Sai. M
IMT2018081

Assignment-2

May 21, 2021

Overview

We need to compute the fore-ground segmentation mask for the images. In digital image processing and computer vision, image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as image objects). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. There are many algorithms with which we can get a fore-ground segmentation mask of images. Some of the Algorithms are OpenCV GrabCut, Superpixel Methods, Spectral Clustering, CNN Based Object Detection.

In this Assignment, we have implemented two models for fore-ground segmentation.

Models

1. **OpenCV GrabCut:** This is a simple model for fore-ground segmentation. In this model, the boundary box is given manually by us for all images.
2. **R-CNN + OpenCV GrabCut:** This is an improvement to the above model when we automatically find the boundary box of an object using object detection using CNN, then we apply the grabCut to the output boundary box.

Input Images



OpenCV GrabCut

Prior to deep learning and instance/semantic segmentation networks such as Mask R-CNN, U-Net, etc. GrabCut was the method to accurately segment the foreground of an image from the background. Firstly We take an image as input then three are two ways to apply grabcut. In the first way, we take the boundary box manually and give that as input to grabCut and the other one is where we give the mask as the input to grabCut and it gives the output removing the fore-ground using the mask. As we want a mask as output we used the first option. The function performs iteratively three steps those are:

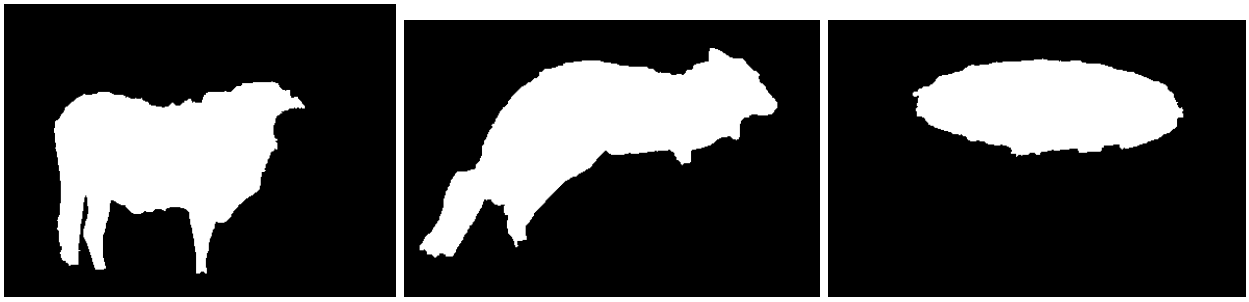
- Estimating the color distribution of the foreground and background via a Gaussian Mixture Model (GMM)
- Constructing a Markov random field over the pixels labels (i.e., foreground vs. background)
- Applying a graph cut optimization to arrive at the final segmentation

In this model, we need to manually give the values of the boundary box of images. In the given Images using the properties of image i.e height and width of image, we calculated approximately and got the output.

In this model, Advantage is that it can detect objects better compared to other model because here we manually give the boundary box. But at the same time, cons are that

sometimes we can't manually keep bounding boxes to every image and also sometimes accuracy may not be good.

Results using OpenCV GrabCut:



Mask R-CNN + OpenCV GrabCut

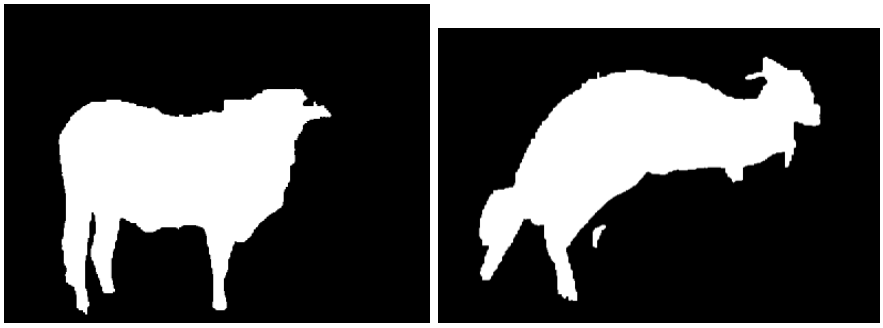
Mask R-CNN is a state-of-the-art deep neural network architecture used for image segmentation. In the before model we manually needed to give the boundary box of the objects. As we are approximately giving the boundary box, that is affecting the accuracy of the model. So in this model we initially take a pre-trained R-CNN model which helps us in detecting objects. Then we pass this mask as input to grabCut which gives us the output mask of the image. Here the pre-trained model is trained on COCO dataset to detect objects. In this we do not need to manually give the values of the boundary box. Mask R-CNN, can automatically predict both the bounding box and the pixel-wise segmentation mask of each object in an input image. The downside is that masks produced by Mask R-CNN aren't always "clean" — there is typically a bit of background that "bleeds" into the foreground segmentation. In this implementation we have three stages, those are:

- Loads an input image from disk.
- Computes a pixel-wise segmentation mask for each object in the input image.
- Applies GrabCut to the object via the mask to improve the image segmentation

In this model, pros are that it automatically figures out the mask of the object and because of that accuracy of the model is better compared to just GrabCut. But at the same time, the cons is that it all depends on the pre-trained model because sometimes, the model may not detect an

object. Like for example in our 3rd input image, this model couldn't recognize the object so it couldn't produce output.

Results using Mask R-CNN + OpenCV GrabCut:



Conclusion

After comparing the results, we can see that it's better to use only the GrabCut model for images like our 3rd input image so that we will get a proper output. But when we have some familiar objects in input images like cows, sheeps, cars, cycles, human..etc we can use the Mask R-CNN + GrabCut because we can get a more accurate mask compared to just GrabCut. And also the value of iterations for GrabCut gets considered in the quality of output. When there are more iterations, the more accurate the mask will be produced but at the same time it takes more time to produce the output when there are more iterations. We have used 100 iterations in both models.

References

- https://www.pyimagesearch.com/2020/07/27/opencv-grabcut-foreground-segmentation-and-extraction/
- https://www.pyimagesearch.com/2020/09/28/image-segmentation-with-mask-r-cnn-grabcut-and-opencv/
- https://dahtah.github.io/imager/foreground_background.html
- https://www.analytixlabs.co.in/blog/what-is-image-segmentation/

Acknowledgements

We would like to thank our Professor G Viswanth sir for giving us deep insights in this world of Visual Recognition.