# Combining image segmentation and seam carving for automated object removal

Jeremy Day
*University of South Carolina*
Columbia SC, USA
jadaytime@gmail.com

Noel Raley
*Unversity of South Carolina*
Columbia SC, USA
noelraley@gmail.com

***Abstract*—This research investigates the effectiveness of combining seam carving with state of the art image segmentation tools in performing automatic feature (object) removal in images. Seam carving is an effective way to remove undesirable pixels from an image, which can either be unimportant pixels, or pixels that belong to an object that is to be removed. The state of the art in image segmentation (Mask R-CNN) can produce masks of objects, which can be used to specify low energy areas in energy maps to mark areas for removal during seam carving.**

***Index Terms*—seam carving,object removal,image segmentation,mask r-cnn**

## I. Introduction

Seam carving was first introduced by Shai Avidan and Ariel Shamir in 2007 [1]. They demonstrated the effectiveness of seam carving as a tool to achieve content aware image resizing, as well as a tool to perform object removal. The original software developed by Avidan and Shamir allowed for a user to specify masks over portions of the image to mark the pixels as important (meaning seams would not run through them) or unimportant (meaning seams would always run through them). Marking these unimportant pixels allowed users to draw masks over objects that they desired to remove from the image. Avidan and Shamir also demonstrated the use of face detectors to automatically provide masks over parts of the image in which distortion or removal was undesirable (faces).

Mask R-CNN is an improvement over Faster R-CNN [2] which, in addition to object detection, classification, and bounding box detection, also produces a high accuracy mask specifying which pixels of the image belong to the detected object [3]. We are using an instance of Mask R-CNN trained to detect 81 different classes of objects. Mask R-CNN can process images very quickly (5 fps). This makes it a powerful and computationally feasible tool for generating masks for object removal.

We combine the output masks of detected objects from Mask R-CNN and an energy map generated for seam carving. By multiplying the intensity of the mask and subtracting it from the energy map, we can guarantee that all pixels marked by the mask get removed by the seam carving operation. We have developed a script to automate this process: to detect all objects in an image and remove each one one at a time. This allowed us to quickly test our development and proposed method, as well as find situations in which our approach fails to produce desirable results. Using the techniques we have developed, it is easy to remove a single detected object from an image, or an entire class of objects (like all people, or all cars, for example).

## II. Background

Seam carving's original intention was to provide a tool to perform content-aware resizing. With the huge surge of popularity of mobile phones, tablets, laptops, and other devices with smaller screens, it is an important part of web development to produce responsive content that can be viewed on such devices. New tools have largely solved this problem for static text content, but images are more difficult to display in a responsive way. If the image is cropped to fit on smaller screens, important content may be lost, reducing the effectual impact of the image. If the image is resized, the aspect ratio may be such that the image is nearly impossible to see on small screens. If the aspect ratio is modified, the image will be skewed in a potentially undesirable way.

Seam carving solves these problems by describing a way to remove seams of unimportant pixels from anywhere in an image. A seam is an 8-connected path through an image from top to bottom (vertical seam) or from left to right (horizontal seam). The importance of a pixel (the pixel's energy) is described by an energy map. The energy map can be generated in a variety of ways. Avidan and Shamir investigated a number of algorithms to generate energy maps. An effective strategy identified to compute enerpy maps is to compute the gradient of the image. Avidan and Shamir proposed the following energy function:

$$e_1(\mathbf{I}) = |\frac{\partial}{\partial x}\mathbf{I}| + |\frac{\partial}{\partial y}\mathbf{I}| \tag{1}$$

Research into using neural networks to detect objects in images was rekindled in 2012 by Krizhevsky, Sutskever, and Hinton with their promising results in using ImageNet, a deep convolutional neural network, to classify images [4]. In 2014, their research into image classification was applied for object recognition by researchers at UC Berkeley who built R-CNN: a convolutional neural network that localizes objects within an image [5]. Their results were very promising, and in 2017 their research was improved upon with Faster R-CNN [2], which was able to detect objects in an image at a reasonable framerate

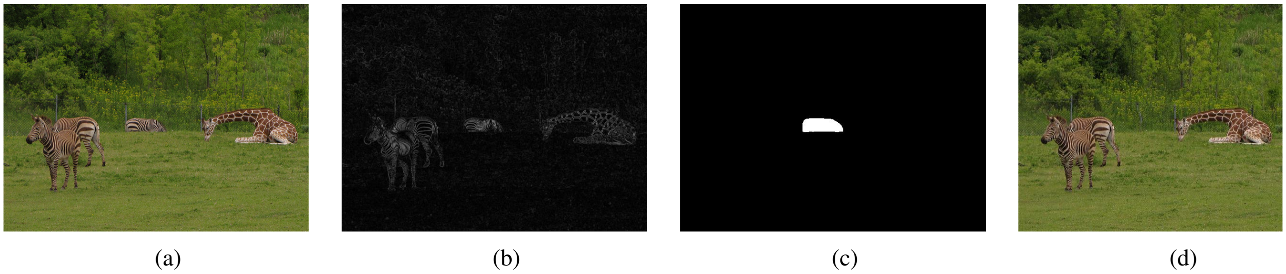


Fig. 1: The original image is in (a). (b) shows the energy map generated with the Sobel operators. In (c) we see the processed mask output from Mask R-CNN, masking the back most zebra. Finally in (d) we see the image after seam carving, with the zebra removed.

(5 fps). This was enhanced in 2017 by the Facebook AI Research team in their development of Mask R-CNN, which is cabaple of detecting objects within an image and describing which pixels are members of that object (masks) [3]. This additional step only imposes a small overhead over the original Faster R-CNN, and the mask output provides accurate image segmentation, useful for a variety of applications.

## III. Proposed Method

We propose a multi step process for an image to perform automated object removal. First, the image is fed into Mask R-CNN. The output of Mask R-CNN includes detected object's classifiers, bounding boxes, masks, and prediction confidence. We then use the object's mask and bounding box as supplementary input to a custom seam carving algorithm. This seam carving algorithm perfoms normal energy map generation (using the Sobel operators), and then subtracts the detected object's mask multiplied by a factor $b$ from the energy map. If $h$ and $w$ are the respective height and width of the input image, we define $b$ to be:

$$b = \max(h, w) \tag{2}$$

This definition of $b$ is to ensure that each pixel in the masked area of the energy map has an intensity so negative that the pixel must be removed in some seam. We calculate the number of vertical seams to remove from the image by using the bounding box of the detected object to calculate the width of the mask. This number provides an upper bound on the number of seams we must remove to ensure the entire object mask is eliminated.

We make an assumption that if Mask R-CNN detects an object, this object is probably important to the image. In general, the parts of an image that are least important tend to be background pieces, landscape, ground in between subjects, etc. Mask R-CNN does not detect these things. It only detects objects of certain classes that have reasonable visual representation in the image, such as clearly visible people or cars. By adding the energy map and all the masks of the detected objects that we do not wish to remove, we can attempt to preserve the visual saliency of these features that were detected. This is similar to Adivan and Shamir's original idea to feed the output of face detectors as preservation masks for seam carving [1].

## IV. Experiment

Our experiment produced promising results under certain circumstances. Seam carving only works in some specific conditions. It fails to produce desirable results in several situations: when the image is very dense in energy (meaning there are few unimportant pixels in the image); or when the object to be removed is close to other objects for which preservation is desired. For energy dense images, there are few seams that can be removed without causing distortion of the image. For example: on images of dense crowds, removing one person is very difficult to do without removing another. When an object that is to be removed has important pixels on two adjacent cardinal directions (e.g. top and left or bottom and right), then neither vertical nor horizontal seams can be removed from the image without removing pixels for which preservation is desired. See figure 2 for a visual depiction of this phenomenon.

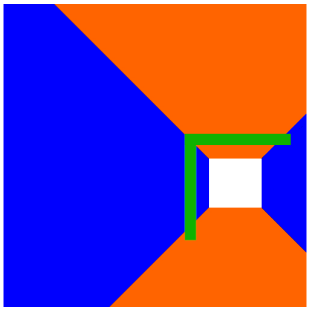


Fig. 2: Because a seam is an 8-connected path, given a mask marked for removal there are only certain pixels that can be removed by vertical or horizontal seams. In this image, the white square represents the mask for removal. The blue region shows the pixels that can be removed by horizontal seams. The orange region shows pixels that can be removed by vertical seams. If we have some region (a single mask or combination of masks) marked for preservation that goes across the entire blue and orange region, there is no vertical or horizontal seam from the image that can be removed without removing one or more pixels from the preservation region.

Figure 1 above demonstrates a successful object removal using our proposed method. We use the bounding box of the zebra marked for removal to determine how many vertical seams we must remove in order to remove the entire object. This bounding box can be seen in the output of Mask R-CNN, shown below.

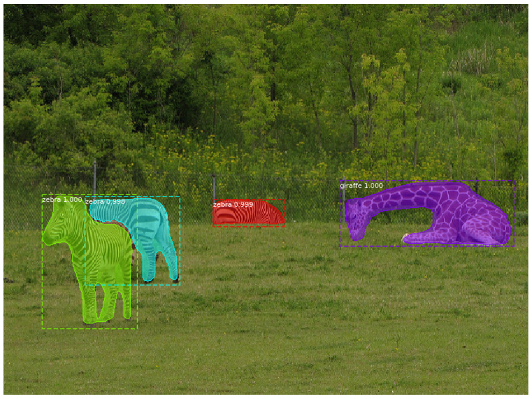
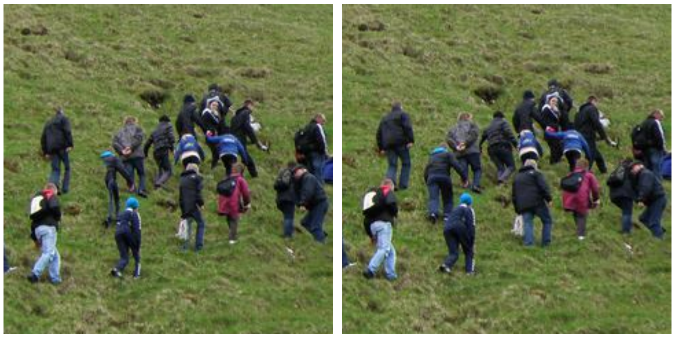


Fig. 3: Output of Mask R-CNN on the image from Figure 1.

For certain types of images, we found that adding the masks of detected objects that were *not* marked for removal produced significantly better results. For most images, the objects detected by Mask R-CNN are important to the image. The pixels that we wish to remove are background pixels, or pixels of large neutral areas of the image (like the ground, for example). Mask R-CNN does not detect these features, so if we mark the features that Mask R-CNN detects as regions to preserve we can keep these important regions during seam carving.

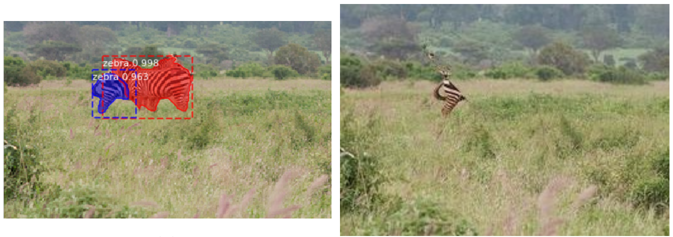


(a)

(b)

Fig. 4: (a) shows the result without adding in the masks of detected objects. (b) shows the result with the detected object masks included. Notice that in (a) many seams are carved through important objects (people), causing undesirable distortion.

Mask R-CNN is extremely effective, and can produce masks of partially overlapping objects. When we attempt to remove an object that is occluding another one, we are usually left with an unfavorable result. We can successfully remove the object closer to the camera, but then the remaining object looks cut off. The once occluded part of the object is still missing, but the foreground object has been removed. An example of this can be seen below.



(a)

(b)

Fig. 5: (a) shows the masks generated by Mask R-CNN of the two zebras. (b) shows the result of removing the zebra closer in the foreground. What was visible of the zebra occluded by the zebra closer to the foreground is still in the resulting image, but the zebra now clearly looks cut off and out of place.

The time that this process takes depends on a lot of factors. Mask R-CNN takes longer on large images. The time it takes to remove a feature is dependent on the size of the feature: larger features require more seams to be removed which requires more iterations of seam carving. We were still able to process images fairly quickly. On modest hardware (i7-920@2.67 GHz, NVIDIA GeForce GTX 77-) it took 8.13 seconds to process a $1024 \times 768$ image.

## V. Future Work

We have identified several major areas for future work to take place:

1) Currently we are always performing vertical seam carving. For many objects or images, this is not ideal. There can be important pixels above and below an object for which removal is desired, but not to the left or right. This means that horizontal seam carving could likely do a better job of preserving the important features of the image. Performing horizontal *and* vertical seam carving, then computing which resultant image has higher energy would likely be a decent way to determine which direction produces a better result. In some situations, it may be optimal to utilize some combination of horizontal and vertical seam carving in the removal of a single object.
2) We are using the width of the detected object's bounding box to calculate how many seams to remove from the image. This is a rather naive way to determine how many seams to remove from the image, and causes undesirable results in cases like the one presented in figure 6. By using a smarter method of calculating the number of seams to remove (for example, using the maximum count of masked pixels in a single row) we can improve our method when applied to cases where the mask does fill the bounding box.
3) Avidan and Shamir also showed seam insertion in their original paper [1]. Seam insertion allows for vertical or

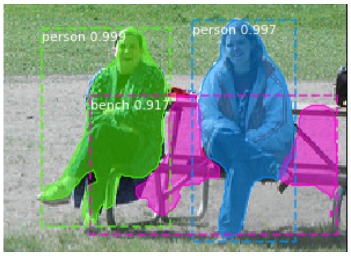


Fig. 6: The mask of the bench is much smaller than the width of the bounding box.

horizontal seams to be inserted into an image to make the image larger. After removing an object with our technique, the same number of seams that were removed could then be inserted to maintain the original size of the image. This provides object removal without the potentially negative side effect of producing a smaller image.

4) Our current implementation feeds the output masks from Mask R-CNN into our seam carving energy map construction with minimal preprocessing. If we introduce a preprocessing step to our masks, we may be able to produce more desirable results. For example, if two objects that we wish to preserve have overlapping masks, our seam carving implementation may need to remove some pixels from this preservation region. It is unlikely that every pixel in a preservation mask is equally important. We may choose, for example, that the center of the preservation masks are usually more important than the edges. We can then perform some processing step on our input masks (like a Gaussian blur) to reflect this property. In some situations, this may produce better results.
5) The script we have developed to perform our object removal is rudimentary. There are many features intrinsic to Mask R-CNN that we could leverage in the development of an API or even GUI to better iterate on experiments, and to make the technology accessible to more individuals.

## VI. Conclusion

Combining state of the art image segmentation techniques with seam carving has produced a robust and effective way to perform automated object removal for a variety of objects in a variety of images. We use the masks generated by Mask R-CNN to mark areas of an image for removal (removal masks) and then subtract these masks from an energy map to ensure these pixels are removed by the application of seam carving to the image. Our base energy map is generated by computing the gradient of the image using the Sobel operators. We experimented with a variety of images, and identified many circumstances that produced positive and negative results. We determined that adding the masks of objects that Mask R-CNN detected to the energy map allowed us to preserve important features of the image.

We had some great results on certain classes of images, but in some circumstances our method produced unfavorable results. Important areas were removed from the image, and some distortion occured as a result of seam carving. In many cases however, our algorithm works as it is currently implemented. We identified a number of ways to potentially improve the results in the future results section.

## References

[1] S. Avidan and A. Shamir, “Seam carving for content-aware image resizing,” *ACM Trans. Graph.*, vol. 26, no. 3, p. 10, 2007.

[2] S. Ren, K. He, R. B. Girshick, and J. Sun, “Faster R-CNN: towards real-time object detection with region proposal networks,” *CoRR*, vol. abs/1506.01497, 2015. [Online]. Available: http://arxiv.org/abs/1506.01497

[3] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, “Mask R-CNN,” *CoRR*, vol. abs/1703.06870, 2017. [Online]. Available: http://arxiv.org/abs/1703.06870

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[5] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Computer Vision and Pattern Recognition*, 2014.