Seam Carving using Object Detection Deep Learning Algorithms

Vrishabh Lakhani
Rochester Institute of Technology
1 Lomb Memorial Dr, Rochester, NY 14623
val3917@rit.edu

Abstract

There has been various methods and approaches developed to formulate image resizing.[5] [6] Improvising over this, we propose a model which combines the Seam Carving method for image resizing to be combined with modern deep learning algorithms. While the current Seam Carving method is dependent upon energy values of the image, we ought to find better context if we used a modern deep learning algorithm for object detection in the imagery. Since the context-aware resizing image is dependent upon basically finding the object of interest, traditional Seam Carving methods use image features like magnitude, edges and second differentials of edges, and sometimes even the luminance of the image pixel, it fails to work in all cases since these method are at best, a guess based on very rough features of what would be defined as an object, therefore it is at most a very crude object detection algorithm. Thus using deep learning for finding the objects and giving those objects in a bounding box fashion to the Seam Carving algorithm, where it may increase the weights of the energy values of pixels in contain the object and thus we will be able to resize the image in a better method, creating an advancement over the traditional Seam Carving approach. Our expectation is that this method should outperform all other types of context aware resizing methods for image resizing.

1. Introduction

One of the most challenging task in Image processing is finding the context of the image. To expand this, coming up with a context-aware algorithm without giving any context to the algorithm and expecting it to retrieve context from the image is one of the biggest challenges in computer vision. Since it's important that resizing of image is not dependent upon just the geometric restraint, we need to know the context of the image while doing so. Traditional Seam carving approaches are used for image resizing which includes reducing as well as expanding the size of the imagery.

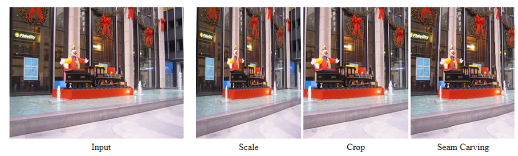

Figure 1. Seam Carving using traditional energy function map

The deep learning algorithm which we are using is a pre-trained YOLO v3 Real-Time Object Detection algorithm which is trained on the COCO dataset. This algorithm can do real-time object detection at the rate of 30 frame per second. Not only is this algorithm more time efficient but it also gives better results than other deep learning methods which are used for object detection. Thus inspired from the success of this model and the importance of having objects as the fundamental units for Seam Carving [**?**], it seems like a good approach to combine it with the state of art context aware models for image resizing.

2. Background work

2.1. YOLOv3 Real-Time Object Detection Algorithm

Typical deep learning system prior to YOLO used classifiers to detect objects in images. In order to do so, they applied classification model over the image or a subset of an image over which they tried to classify that it into a category. The category getting the highest probability would be detected as an object that is part of the image.
YOLOv3 [2] uses a completely different approach to the problem of object detection. Instead of dividing the entire image into many parts and applying a classification algorithm, what it does is that it uses a single neural network which decides which part of the image may contain the object. This approach is better since it reduces the computational task of sub dividing the entire image into many sub part of different scales and geometry. Thus this neural

comparison.PNG

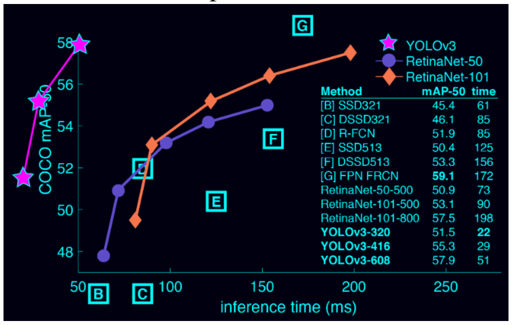

Figure 2. YOLOv3 Real Time Object Detection algorithm comparision with other models.

results.PNG

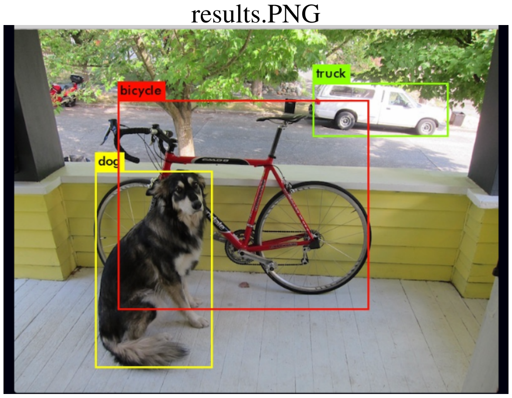

Figure 3. YOLOv3 Real Time Object Detection algorithm in action

network divides the image into different regions which helps in predicting the bounding boxes and probabilities for each of the regions. Later these are sent over to the entire model and each of these bounding boxes are assigned weights by the predicted probabilities.

There are many advantage of using such an algorithm over the classifier-based approach, one of the major one's being that the neural network is able to understand the global context of the image. And secondly, since we are only running this model over a single image instead of converting a single image into thousands of image prior to putting it into a model like the RCNN, it is very computationally efficient. Since for our purpose of image resizing, we need a pre-trained version of YOLOv3 Real time object detection algorithm, we will be using a pre-trained version which is trained over the COCO dataset.

This method was invented by Redmon, Joseph et al for real time object detection which can detect objects at the rate of 30 FPS for videos.

Figure 4. Here we see how a Seam Carving algorithm fails when the image is highly textured.

2.2. COCO Dataset

COCO dataset is a large-scale object detection, segmentation and captioning dataset which is an industry standard for pre-training deep learning models for the purpose of object detection. It is one of the best dataset out there for this purpose because it contains variety of advantages over other datasets like it has Object segmentation, labelling for recognition in context of the entire image, Superpixel stuff segmentation, etc.
To understand the size of this dataset, here is some information about it. It has 330K images with more than 200K images which are labelled. It has 1.5 million object instances in all of it's images and 80 different object categories. Apart from object categories, it has 91 stuff categories for the context of the image. It has 5 captions per image to describe each of the image and 250000 people with keypoints.

2.3. Seam Carving Algorithm

Seam Carving was invented by Shai Avidan et al. at Mitsubishi Electric Research Labs. It takes raw input images as input and created Seam energy maps which are used to find relevant object points in the image and resizes the image accordingly. This energy map is based on important image features and regions based on gradient features like Sobel filter, Scharr, entropy, etc.
In order for it to work, it establishes a number of Seams which is figures out from the energy map and based on whether we want to contract or expand the image, removes the most useless seams or inserts some seams to extend it. Another functionality of Seam Carving is that it allows to manually define areas where we don't want to change the Seam numbers and thus make it a feature where we have the ability to retain an object or remove an object completely from an image. Each of these seams are assigned some weight which is retrieves based on the energy function.

2.4. Using Deep Learning for Image Resizing

There has been some research [3] which has been done before for using deep learning for some sort of application in Seam Carving. In this method, the researcher try to cmobine Convolutional Neural Network along with Local Binary Patterns to recognize whether an object has been modified using Seam Carving or not. They achieve an accuracy of upto [81% - 98%] depending upon how severely the images have been tampered. Thus, it is evident that using deep learning for the purpose of Seam Carving definitely has some scope to look forward to.
Apart from that, another research [1] tries to to Seam Carving using deep learning algorithm taking a raw input image and the aspect ratio as input and trying out output the resulting resized image. They use a model with weakly and self-supervised deep convolutional neural network. This generative deep learning approach shows some great results and we expect that our model should therefore work because not only it takes in the context of the entire image, with the help of modern object detection algorithm we find that we are giving additional information regarding the object which can help it find the Seams to care about and therfore reducing the number of artifacts.

3. Method and Approach

3.1. Energy Function

There are many possible energy functions to use for the purpose of Seam Carving, but previous research [?] have compares performance of energy functions based on L1 and L2-norm based gradient, saliency measures, Harris-corners measures and eye gaze measurement. As expected, none of the energy function outperformed each other but give results based on the subjectivity of individual image and it's property. The variation that was observed was based on how the artifacts were created and what sort of defects in the image were visible. However, the found that energy function which is based on entropy which is computed over a range of 9x9 pixels to work better in most of the cases.

3.2. Seam Creation from the energy function

Seams are either horizontal or vertical oriented. For the image resizing part, the question remains of what order to use for seam removal, should it be vertical first or horizontal? This is optimized using a dynamic programming approach. This dynamic programming approach minimizes the error which is given by a polynomial which is based on the energy value of each of the Seam.
In order to reduce the size of the image and resize it, all the seams are sorted in order of their energy value, run through the dynamic programming algorithm and the one's with the lease energy value are removed. This is how the pixels which contain the least amount of edges, the most bland pixels are removed since the assumption is that such a pixel won't be containing any particularly important object.
In order to enlarge an image, we have to input artificial seams into the image. In order to do this, an inverse of seam removal algorithm is used. This 'inversion' is working in such a way that first the optimal vertical seam is computed and the color of the pixels within each seams are computed by averaging the pixels to the left and right of this pixel. This is basically a sort of context aware interpolation of pixels in an image.
This works only when we don't stretch the image by a lot and hence, a better approach is used whereby instead of taking one single seam, first k important seams according to the energy functions are chosen and the above steps is run on all of them. The number of seams chosen is based on how much we want to expand the image. Thus even though this approach will work better than the former one, extreme stretching of the pixels will still research into some visible artifacts.

3.3. Changing the weights/density/energy according to the deep learning algorithm

Since the energy functions defined by properties measured above are still not accurate, we use deep learning algorithm to get some better results.[4] Using YOLOv3 we get a number of objects that are detected inside the image. We target to choose only the first k objects, where k is a number chosen as a hyperparameter for our deep learning based Seam Carving algorithm. Our deep learning algorithm gives coordinates of the bounding box of each of the detected objects as the output. We use these coordinates to find the bounding box and compare it to our energy map. Since a vector of Seam is already created with weightage given to each of the Seam in the image, we increase the weights of the seams which intersect with this bounding box.

3.4. Final model

Thus in our final model, we are using the algorithm of the traditional Seam Carving algorithm and using deep learning algorithm to optimize the weights that are given to the Seams in the image. This approach should ideally give better performance in retaining or complete removal of the object using the Seam Carving algorithm.
However, it is expected that this algorithm may have some minor artifacts that can be created due the fact that we are using the bounding box instead of just the pixels of the object that are detected. For the future work, we wish to apply State of The Art object detection algorithm where we will use only the pixels instead of the bounding box and thereby

reducing the number of artifacts significantly.

4. Experiments to be performed

4.1. Experiment

As described above, we are using a pre-trained model for the deep learning object detection part since training the model on a dataset is not feasible. Apart from this, the results from the pre-trained are good enough for our application of modified Seam Carving. We first have to check whether our original energy function is working properly or not. If it is working properly, we will apply the new weights that are assigned from the deep learning model. Otherwise, we will have to change the energy function to a function which gives us good enough results. This is important because otherwise no matter how good our deep learning algorithm is, we won't be getting any satisfying result.

In order to change the energy maps of the image, we have two approaches to do so, first method is the method is which we try tweak the section of the dynamic programming code to figure out the seams. In the code, when a pixel chooses the least weighted pixels from the 3 pixels above it, we are given higer priority to the pixels which are outside the bounding boxes. Thus the seams will naturally only form for pixels which are outside the bounding box. In case all the pixels are inside the bounding box only then we go the traditional route and choose according to their energy values.

Second method involves given higer energy values to the energy maps prior to the construction of the seams, thus in this case, the minimum seam finding algorithm remains the same but we are only changing the values in the energy maps. Since our deep learning model is only given bounding box, we are giving higher weights to the pixels in the bounding box. This could be done in various ways like if it is a gray scale images of edges as the energy map, we will mark all the pixels in the bounding box as the white. Doing so will increase the total weights of all the seams that pass through the pixels in the bounding box and thus these seams will be deleted later on.

To start the evaluation part of the experiment, we will have to choose different test cases for image expansion and reduction. If our algorithm works for both of the approach we can finally approve that is working ideally and as needed.

4.2. Evaluation Metrics

For evaluating the performance of our model as compared to the baseline model, we are going to compare the results for a number of test cases. These test cases include various types of images over which both the models, the baseline and the deep learning model can fail to work as intended.

These test cases include things like, images with a lot of objects, images with not objects, images with highly textured objects, images with untextured objects, images with high contrast, images with high contrast within each object, images with objects which are not pre-trained on YOLO RCNN model, etc.

Based on this test cases, we can figure out how each of these models hold for them and where they fail. Our expectation is such that the deep learning model should be as good of a result as the baseline model even in images which contain objects which are not pre-trained on the YOLO RCNN model. This is because if no object is detected by the deep learning model, then there is no additional weight giving to the Seams in the energy map function and thus it should work exactly as good as the baseline Seam Carving algorithm. Thus, we are reducing the chance of error by wrong detection of objects and thus reducing the number of false positives altogether.

5. Results and discussion

In the following figure, we see some examples of how our deep learning model works in image resizing when compared to the baseline Seam Carving algorithm and non context-aware image resizing algorithm. On the extreme right column, we have the results of our model and on the middle we have the baseline Seam Carving algorithm, on the extreme left is the ground truth image and right to that is the image which is resized without a context-aware algorithm.

We notice that our model gets the context better than the baseline model, it is more aesthetically pleasing and the pixels of the image don't seem to smear that much into each other, especially for the objects that are of importance with the background. Even though the baseline model works for most of the cases, it gives multiple defects when the object is highly textured and there isn't much difference between the luminance values of the object and the background, these things aren't that strong in our model.

We still notice however, that there are still many defects in our model. This is mostly because we are still using the energy function in our final model for objects which and pixels which are not classified in our deep learning model. These pixels are still prone to all the issues that the baseline model has and does removing these defects is difficult unless we chuck the concept of energy function altogether. This might be a good idea for creating a deep learning model which is based on encoder and decoder networks which act like a blackbox to resize the image. Even then, another issue would arise of creating a seperate model for

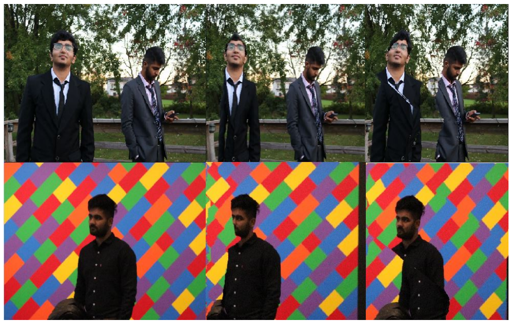

Figure 5. Results of comparing the ground truth with the baseline model and the deep learning model

different sizes of resizing.

6. Conclusion and future work

Through this experiment we have shown that deep learning can be successfully used for effective and potentially superior approach for context-aware image resizing. In particular, we have shown that using object detection algorithm like YOLO RCNN v2 which is traditionally only used for object detection in an image can be used for improved assignment of weights while resizing image using Seam Carving algorithm, and thus there seems to be a potential promise for further development in this direction for image resizing using deep learning approaches.
Therefore, our work lays a foundation for such future works. Further, we can improve the usability of such a method by using even advanced deep learning pipelines to give weights to only specific objects in the scene, this can be done with the help of Seq2Seq methods used as encoder and decoder networks, or ensembling a NLP deep learning network to get the context of which object the user is trying to assign higher importance to and resize accordingly. Further, using techniques like attention mechanism for improving the accuracy of such an NLP task or even increasing the accuracy of the object detection algorithm can be foreseen on the future potential horizon.

References

[1] Tae-Hyun Oh Yu-Wing Tai In So Kweon Donghyeon Cho, Jinsun Park. Weakly- and self-supervised learning for content-aware deep image retargeting. 2017.

[2] Ali Farhadi Joseph Redmon. Yolov3: An incremental improvement. 2018.

[3] Kelton A. P. Da Costa Luiz Fernando da Silva Cieslak, Joao Paulo Papa. Seam carving detection using convolutional neural network. 2018.

[4] Pyimagesearch.com. https://www.pyimagesearch.com/2017/01/23/seam-carving-with-opencv-python-and-scikit-image/.

[5] Ariel Shamir Shai Avidan. Seam carving for content-aware image resizing. 2012.

[6] Haibin Yu Xueyi Ye Li Zhang Guojin Ma Zhiwei He, Mingyu Gao. A new improved seam carving content aware image resizing method. 2013.