Team 2 - Final Project Report

Agraj Srivastava
ai19btech11020@iith.ac.in

Mansi Nanavati
ee19btech11036@iith.ac.in

Soumi Chakraborty
es19btech11017@iith.ac.in

Tarun Ram Menta
ai19btech11004@iith.ac.in

Abstract

The original seam carving algorithm depends on the colour information and the gradients of the input image which sometimes lead to faulty outputs. In this paper, we present various methods to obtain energy maps: Major Blob Preservation, Deep-Learning based Saliency Maps, Foreground Preservation using Self-Attention Maps, and Combining Energy Maps. We also contribute towards Real-Time resizing by fastening the existing seam-carving algorithm. The combination of energy maps is determined as the best outcome for this experiment. Implementation can be found on our GitHub repo: https://github.com/peppermenta/seam-carving-project

1. Introduction

Conventional image resizing involves methods such as *cropping* and *interpolation*. The former works by trimming away information at the periphery of the image whereas the latter is based in the principle of using a weighted average of a pixel's neighbourhood in order to compress image information. Whilst these methods boast supreme efficiency in terms of computation speed, they can yield undesirable results such as a jagged resultant image, loss of information or over-softening of details, particularly when the aspect ratio of the original image is modified.

The drawbacks of the aforementioned algorithms lie in their inability to content-aware. Both interpolation and cropping fail to discriminate between different regions of the image based on their relative importance to the overall information contained in the photograph. Integrating the concept of *energy* in an image into the resizing protocol offers a promising, content-aware paradigm for intelligent resizing. *Energy* is a quantitative measure of the localized gradients with respect to brightness, colour etc. over different locations in the image.

Building on the idea of energy, seam carving is a content-aware image resizing algorithm that employs an energy function to map the importance of pixels in an image. Using the generated map, the algorithm aims to ”carve out seams” that minimize the loss of energy in the image. A seam is defined as an 8-connected path of low energy pixels crossing the image from top to bottom / left to right. The core principle at work is to find and remove the lowest collective energy seam so as to preserve image content. Rectangular structure of the images is maintained by concatenating the remaining pixel rows / columns together following seam removal.

The original seam carving algorithm deploys a simple energy map:

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

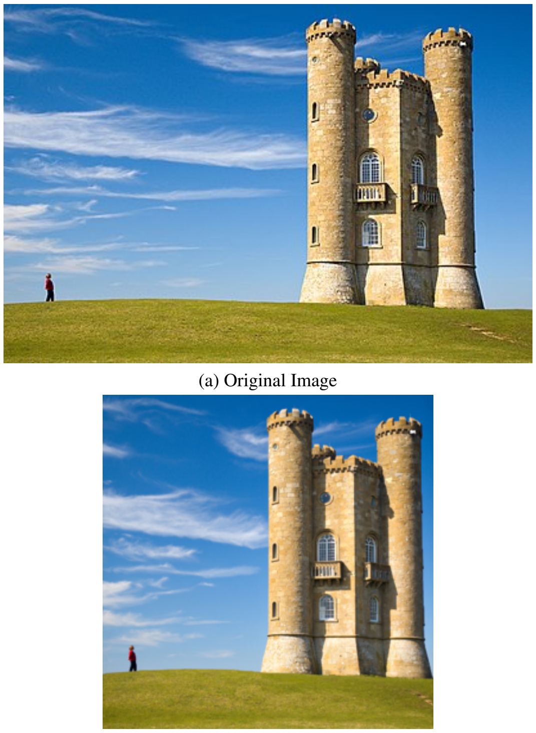
The authors suggest a re-computation of the energy map following removal of seams to account for the energy changes generated by new neighbouring pixel relationships.

The seam carving algorithm isn't restricted to the map mentioned above or its variants. Instead, we can deploy numerous differing energy maps in conjunction with seam carving to achieve different objectives during resizing. A brief discussion on the same can be found in the next section.

2. Problem Statement

Seam carving is entirely dependant on the formation of energy maps, as described in the Introduction. In this paper, we plan to explore various methods of creating energy maps, including but not limited to depth maps, saliency maps, attention maps, etc. The original algorithm proposed uses the colour gradients in the two-dimensional plane to calculate the energy. We will explore using newer algorithms to do the same without relying solely on the colour information of the picture.

The seam carving algorithm was originally proposed in the year of 2007, and the fact that the algorithm tries to be ”content-aware” essentially means that it tries to locate objects of interest in the input image, and carve seams around it. Since then, algorithms involving object detection using



(a) Original Image

(b) Resized image using seam carving

Figure 1. Seam Carving

deep learning have progressed leaps and bounds. Thus, we aim to propose a novel energy map creation method which doesn't rely on gradients and utilises deep learning for detecting the objects in the images. [10] illustrates using the graph cut algorithm to find an optimal seam to stitch two pictures together. To do so, they use the CNN-based method YOLACT [2] to find overlapping objects between the left and the right frames, and assign high object energy to these overlapping areas so as to segment the image around these objects. Drawing inspiration from this paper, we intend to use CNN-based algorithms to detect features of interest in the input image and generate an energy map based on the results returned.

We then compare these algorithms and study the effect of different energy maps on the final output. Moreover, we will compare these algorithms based on different parameters and try to draw conclusions about the correlation between the energy maps and input images. For example, we will compare the algorithms based on their run time and their ability to perform better on images with comparatively lower gradient changes.

3. Literature Review

Seam Carving for Content-Aware Image Resizing. Introduced in 2007, Seam Carving is an algorithm that effectively resizes images by considering the geometric constraints as well as the image content. A seam can be either vertical or horizontal, where optimality is defined by an image energy function 1. Seams are either removed or inserted from an image based on the application. Aspect ratio modification, image re-targeting, content amplification, and object removal are all examples of image manipulations that can be done with this operator.
The authors tested L1 and L2-norm of the gradient, saliency measure [11], Harris-corners measure [7], and eye gaze measurement [5] as Image Energy Functions and found that no single energy function performs well across all images but in general they accommodate a similar range for resizing.

Energy Maps. The energy maps are a central part of the seam carving algorithm, which bring about the content-awareness. A poorly chosen energy map would not be able to capture important parts of the image, which should be preserved. The original seam carving paper [1] used simple choices for their energy maps, as detailed above. As mentioned in the Problem Statement, there have been various works which have studied the use of different energy maps with the seam carving algorithm. [14] proposed an energy map which utilizes both the magnitude and the orientation of the gradient, since an energy map based on magnitude alone would assign low scores to foreground objects, if they are flat and homogeneous. [17] used an energy map which involves 3 components: Magnitude of the gradients, a user-defined depth map, and a saliency map. [4] focuses on the utilization of the energy map, by propagating the importance of a removed pixel to its neighbors, to reduce artifacts in the final output. [12] proposes an improvement on the seam carving algorithm itself. The main aim of an energy map would be to isolate 'important' parts of an image. The aforementioned methods all have suggested energy maps based on different factors, taking into consideration a variety of aspects, since it is difficult to optimally assign an importance score to every pixel. Such a problem has been widely studied in modern Deep-Learning based CV, under object detection and semantic segmentation. We aim to utilize heatmaps from various saliency methods for CNNs [16, 19, 20], object detection models [6, 15], and semantic segmentation [8, 18] models. Combining these newer developments with a thorough analysis of the performance and effects of existing energy map choices, we aim to develop an improved energy map for seam carving.

Emerging Properties in Self-Supervised Vision Transformers. [3] develops self-supervised Vision Transformer for object segmentation. The model automatically learns an interpretable representation and separates the main object from the background clutter. Self-attention layers in the model build the representation of a spatial location by "attending" to other locations. The network generates a high-level understanding of the picture by

”looking” at other bits of the image. The model learns a feature space by detecting object pieces and similar properties across the image. Attention maps are generated which pose as potential energy maps for seam carving as they hold important data of the image.

4. Results & Discussions

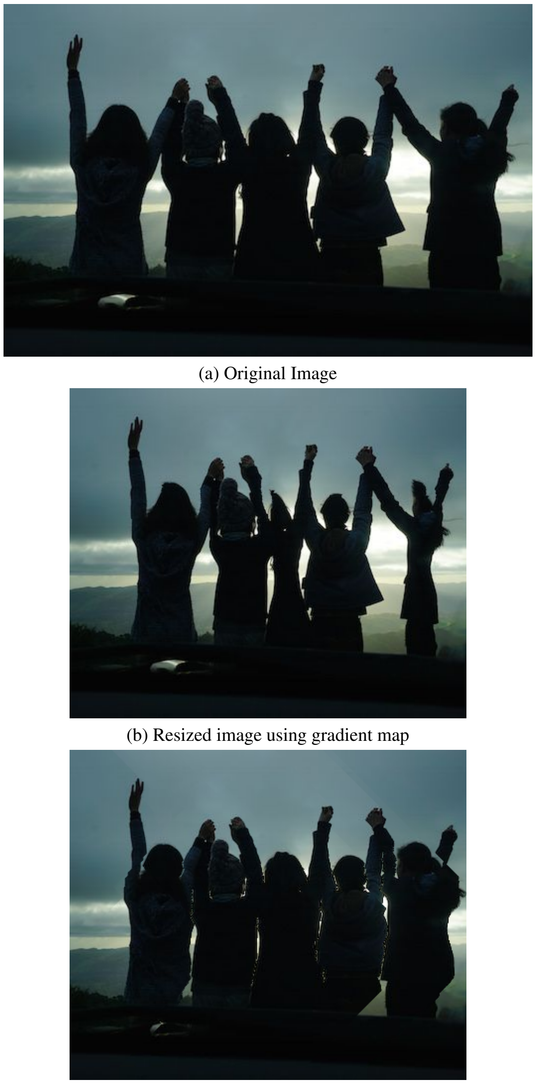
As mentioned above, [1] uses a gradient based energy map that gives precedence to conventional points of interest such as corners and edges and preserves the same, trimming away blobs. This energy map delivers underwhelming results in scenarios where the high-gradient regions may not be important to the overall content of the image. In this section we explore such scenarios and provide suitable alternative energy maps for resizing.

4.1. Major Blob Preservation

Consider the original image in figure 2. The silhouette of the people in the image is the information-dense region of the image. However, the gradients in this blob are low and high gradient regions are located in the sky background that is relatively poor in information. When we deploy the original energy map in conjunction with the seam carving algorithm, we can see that the results are far from ideal. The sky background gains priority on preservation and the shape of the people, particularly the person at the extreme right of the image, is greatly distorted.

To counter this effect, we make use of major blob detection algorithm to preserve the silhouette. To generate this energy map, we take the grayscale version of the image, followed by an otsu binarization of the same. This binarized image is then fed to a depth-first-search based major blob detection algorithm that connects regions using 8-neighbour connectivity. At the end of blob detection, pixel locations part of the major blob are awarded high energy value with the rest of the locations set to 0 energy.

Clearly, the major blob energy map is successful in preserving the information-rich silhouette of the people in the image, trimming away at the background-heavy gaps between the people.



(a) Original Image

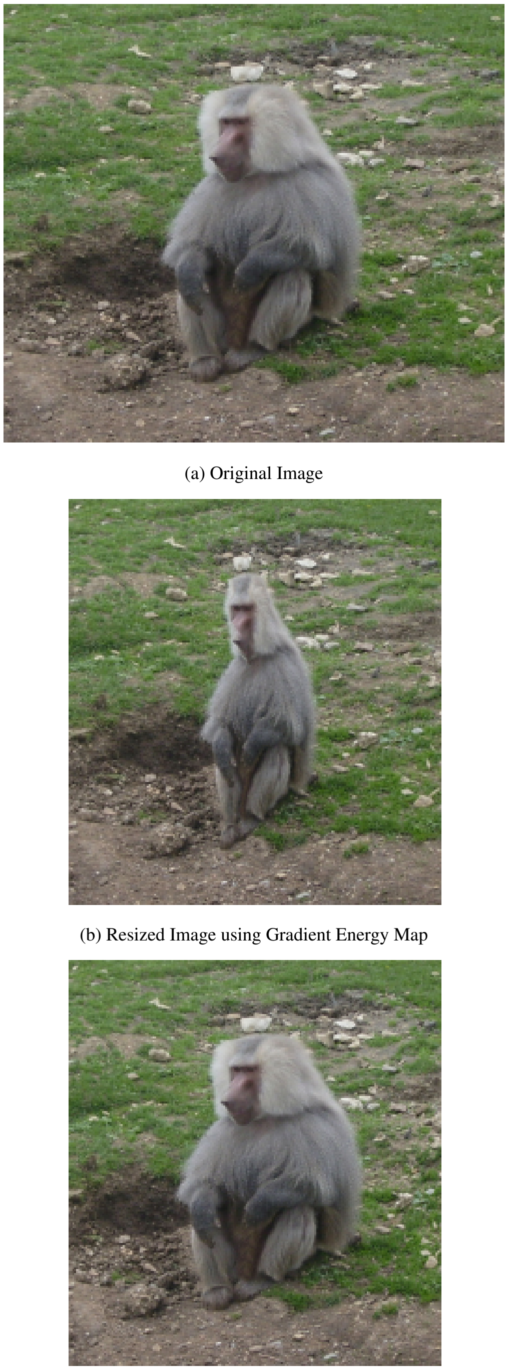
(b) Resized image using gradient map

(c) Resized image using major blob map

Figure 2. Major blob energy map

4.2. Traditional Deep-Learning based Saliency Methods

Deep neural networks (DNNs) have shown state-of-the-art performance in diverse application domains, including complex tasks such as image recognition, video synthesis, speech-to-text conversion and autonomous navigation, to name a few. Interpretability of the decisions made by neural networks is of particular significance, and there exist a diverse set set of approaches for interpretability. In particular, for convolutional neural networks (CNNs) trained for image classification, pixel-based attribution methods have been very popular. These methods produce an importance score for each pixel in the input image, highlighting the contribution made to the final decision of the CNN. One such method is Integrated Gradients [20]. We used the output heat-map of this method as the energy map for seam carving. However, this method does not scale well to all types of images, as it is dependent on the dataset which the model has been trained on. When utilized on images that are outside the distribution that the model has been trained on, the results are unsatisfactory. The process of generating these saliency maps is also time-consuming, and too slow to use the original seam-carving process on. These drawbacks led us to look for other Deep-Learning based solutions, and modify the seam-carving algorithm for faster execution. We explain this modification in our Results and Approach Section.



(a) Original Image

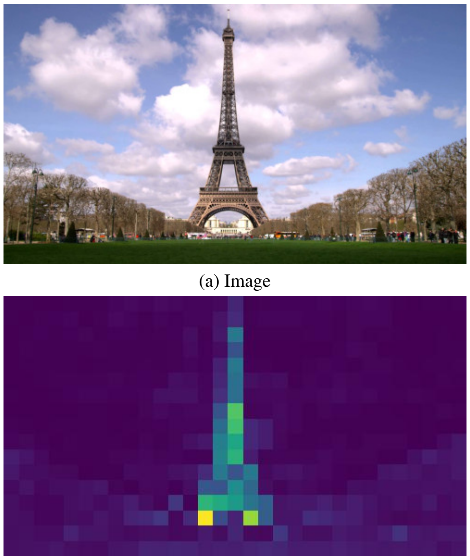
(b) Resized Image using Gradient Energy Map

(c) Resized Iamge using Integrated Gradients Energy Map

Figure 3. Seam Carving using Integrated Gradients Energy Map

4.3. Foreground Preservation using Self-Attention

As mentioned in [3], the DINO self-attention model allows for discovery and segmentation of objects in an image or a video with absolutely no supervision. Figure 4b is an example of self-attention heat-map of the Image 4a. In Figure 4b, it is clear that the main object of the image has the highest gradient compared to the rest. Based on this observation, we decided to use DINO energy maps and plug it in the seam carving algorithm used.



(a) Image

(b) Self-Attention heat-map of the image

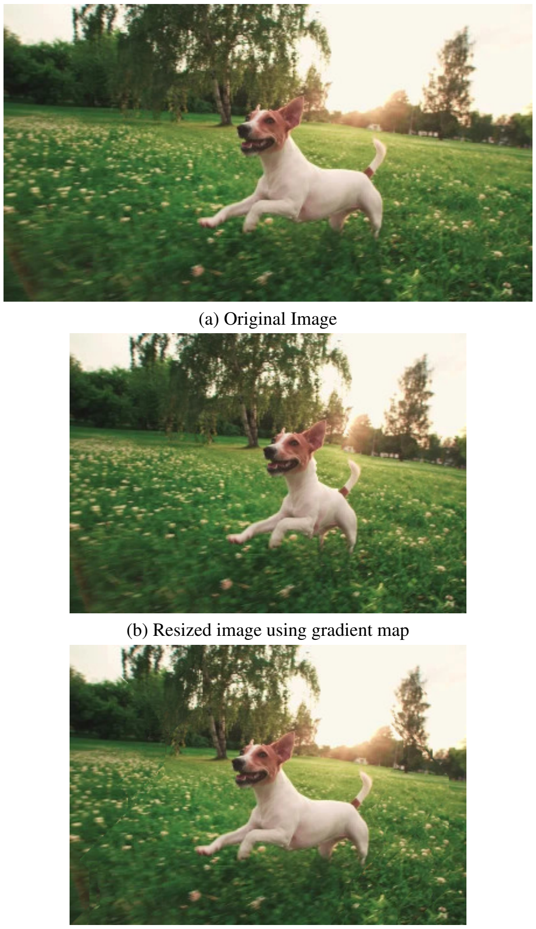
Figure 4

In this experiment, the model renders multi-head attention maps (totally 5) and the output from the third attention head is considered as it relatively gave better differentiation between foreground and background. Figure 5b is obtained using the regular gradient map method and it is noticed that the main object from the original image 5a is heavily distorted. The DINO self-attention heat-maps come to rescue by preserving the main subject of the image, i.e, the dog, and results in successful seam carving.

Foreground preservation is best performed when DINO self-attention heat-maps are employed when compared to the conventional gradient energy maps. Moreover, Vision Transformer (ViT) is going to be the future and seam carving is a good application for it. The overall computation time of the self-attention based seam carving turns out to be lesser than the time taken for the Major Blob method.

4.4. Combining Energy Maps

[17] makes use of three types of energy maps to build the final energy map. They've made use of the regular gradients energy map along with a saliency map (built using [9]) and a user-defined depth map. We explored the saliency map built in [9] and decided to build our own because albeit robust, the algorithm has now become a little outdated and can be improved upon by using deep-learning based algorithms. Moreover, [17] makes use of a depth map in addition to



(a) Original Image

(b) Resized image using gradient map

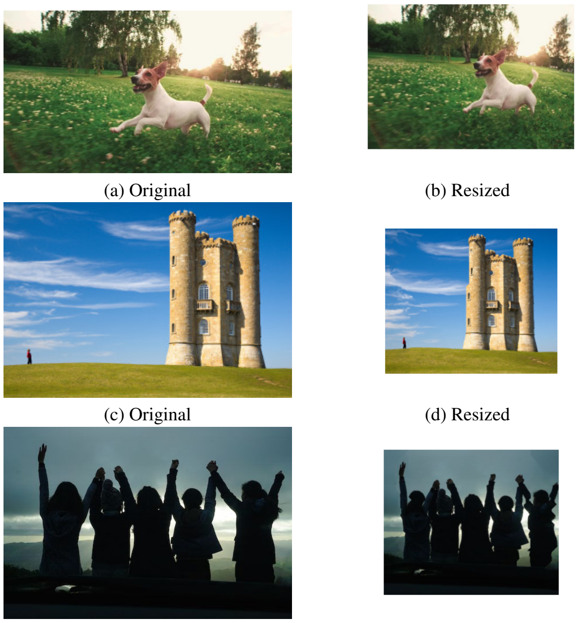
(c) Resized image using DINO heat-maps

Figure 5. Self-Attention Energy Maps

the saliency map, and generating a map using deep-learning eliminates the need of both these maps since it takes care of all the issues that the two maps were generated for. However, while working on this project, we realized that self-attention maps work better than saliency maps, hence we decided to not use saliency maps for combination of energy maps. To build the final energy map, we use the following equation:

$$EM_{ij} = \alpha EM^1_{ij} + (1 - \alpha) max(GM_{ij}, EM^2_{ij}) \quad (2)$$

Where EM refers to the final energy map, EM^1 & EM^2 refer to the energy maps rendered by DINO and major blob detection respectively, and GM refer to gradient maps. The subscripts refer to the pixel values at the (i, j) coordinate. α is a hyper-parameter that determines the weightage/importance given to each group of energy maps. We found that $\alpha = 0.85$ gives the best results across various settings. All the components energy maps (EM^1, GM & EM^2, as well as the final resultant map EM were normalized to lie in the range $[0, 1]$



(a) Original

(b) Resized

(c) Original

(d) Resized

(e) Original

(f) Resized

Figure 6. Results using the Combination of Energy Maps

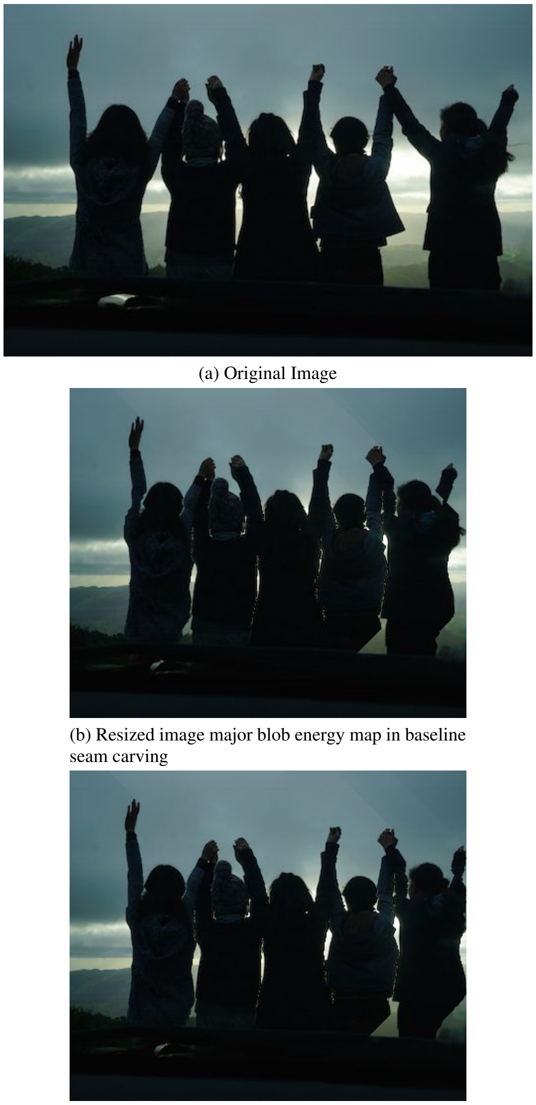
Figure 6 shows how well the Combined Energy Maps work over various images.

4.5. Towards Real-Time Resizing

In our discussions above, we have experimented with various energy maps used in conjunction with the seam-carving algorithm. Although we have been able to achieve improvements on the baseline algorithm in multiple situations using different notions of energy, these methods have a significantly heavier computation workload than the vanilla, gradient-based resizing method. Take the case of the major blob energy map for example, wherein the simple resizing of 3c took over two minutes to execute on a standard computer. Clearly, this approach doesn't scale to higher resolution pictures.

For a more responsive use-case of user-interactive resizing we propose an alternative implementation of the seam-carving algorithm. In the *fast* implementation, we compute the energy map for the image to be resized only once at the start of the resizing. Then, for every seam deleted from image, we apply the equivalent deletion mask to the image's energy map to trim away energy values of the deleted pixel locations and use the newly formed image and its corresponding energy map for the removal of the next seam. This approach saves computation that would be expended in recomputing energy maps for the image from scratch at each step in the baseline algorithm.

While the *fast* approach does lead to some loss in quality of the image owing to a negligence of alterations to energy on



(a) Original Image

(b) Resized image major blob energy map in baseline seam carving

(c) Resized image using fast seam carving

Figure 7. Effects of fast seam carving

removal of seams (ex: new edges being formed on removing a row or edges being deleted), from our observations the loss in quality is at an acceptable level for the computational gains as can bee seen in 7.
A middle ground to the two versions of seam carving outlined above would be a hybrid model that recomputed the energy map from scratch at set interval points while using the masking technique in the iterations between for an optimal trade-off between final image quality and speed.

4.6. Evaluation Metric

Thus far, we have been judging the results of each method discussed subjectively with a naked eye test. But we wanted to have a more concrete evaluation metric set up to better judge the results of each energy map. To do so, we implemented the Mean Area Ratio metric proposed in [13]. This metric essentially returns a score based on the amount of area of salient objects that is successfully preserved after seam carving. We ran all our tests on the MSRA 10K Salient Object dataset. This dataset contains images and a corresponding binary mask of the most salient object in the image. To compute the mean area ratio between the original and seam carved image, we removed seams from the salient object mask alongside the original image, and then computed the ratio of the number of white pixels in the original mask and the resized mask. We computed the average of the ratios over a 1000 random images from the dataset for the gradient, major blob and saliency maps and the results can be found in Table 1. We planned to evaluate the DINO energy maps, and the combined energy maps as well but were unable to do so due to a lack of computational resources.

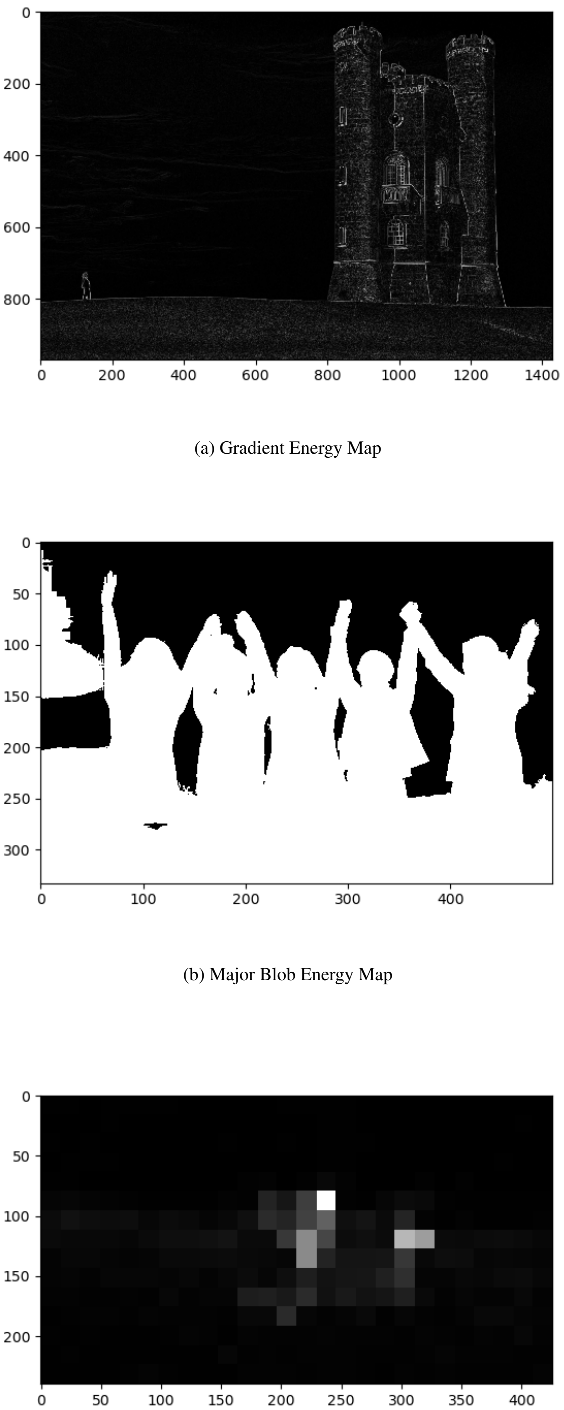
The results show that though the Major Blob Map works very well in some situations like when there are only minor changes in the gradients of an image, the gradient map outperforms it in most cases.

Gradient Map	Major Blob Map	Saliency Map
0.9089	0.7741	0.91

Table 1. Ratios of Salient Area Preserved by each Energy Map

5. Conclusion

The baseline version of the seam carving algorithm presented with numerous drawbacks, mainly pertaining to the narrow scope of the gradient-based energy map. To make the algorithm more robust, we experimented with custom energy maps starting with the Major Blob Map, which emphasised object preservation and helped handle images which had low gradients in information rich zones. However, this map focused on a very niche set of images. For more generic methods, we turned to saliency methods for CNNs. While they serve a more generic use case, the saliency methods are hindered by their database-specific characteristics that make it difficult to generalize the energy map for varied image sets. Self-Attention based energy maps provide better results as the map helps retain foreground details in the image while seam carving. Combining the best of all the worlds, energy maps obtained in 4.4 have considerable information on the main object in the image along with preservation of background. In conclusion, we present an improved energy map method: Combination of Energy Maps, as our final product along with a faster implementation of the seam-carving algorithm to support real-time resizing.



(a) Gradient Energy Map

(b) Major Blob Energy Map

(c) DINO Energy Map

Figure 8. Energy Map Screenshots

References

[1] Shai Avidan and Ariel Shamir. Seam carving for content-aware image resizing. *ACM Trans. Graph.*, 26(3):10–es, July 2007. 2, 3

[2] Daniel Bolya, Chong Zhou, Fanyi Xiao, and Yong Jae Lee. YOLACT: real-time instance segmentation. *CoRR*, abs/1904.02689, 2019. 2

[3] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers, 2021. 2, 4

[4] Sunghyun Cho, Hanul Choi, Yasuyuki Matsushita, and Seungyong Lee. Image retargeting using importance diffusion. In *Proceedings of the 16th IEEE International Conference on Image Processing*, ICIP'09, page 973–976. IEEE Press, 2009. 2

[5] Doug DeCarlo and Anthony Santella. Stylization and abstraction of photographs. *ACM Trans. Graph.*, 21(3):769–776, July 2002. 2

[6] Ross Girshick, Ilija Radosavovic, Georgia Gkioxari, Piotr Dollár, and Kaiming He. Detectron. `https://github.com/facebookresearch/detectron`, 2018. 2

[7] Chris Harris and Mike Stephens. A combined corner and edge detector. In *In Proc. of Fourth Alvey Vision Conference*, pages 147–151, 1988. 2

[8] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn, 2018. 2

[9] Huaizu Jiang, Jingdong Wang, Zejian Yuan, Tie Liu, Nanning Zheng, and Shipeng Li. Automatic salient object segmentation based on context and shape prior. *British Machine Vision Conference*, 6, 01 2011. 4

[10] Taeha Kim, Seongyeop Yang, Byeongkeun Kang, Heekyung Lee, Jeongil Seo, and Yeejin Lee. Segmentation-based seam cutting for high-resolution 360-degree video stitching. *IEEE Access*, 9:93018–93032, 2021. 2

[11] D K Lee, L Itti, C Koch, and J Braun. Attention activates winner-take-all competition among visual filters. *Nature Neuroscience*, 2(4):375–381, Apr. 1999. 2

[12] Alex Mansfield, Peter Gehler, Luc Van Gool, and Carsten Rother. Visibility maps for improving seam carving. In Kiriakos N. Kutulakos, editor, *Trends and Topics in Computer Vision*, pages 131–144, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. 2

[13] Tam V. Nguyen and Guangyu Gao. Novel evaluation metrics for seam carving based image retargeting, 2017. 6

[14] Hyeonwoo Noh and Bohyung Han. Seam carving with forward gradient difference maps. In *Proceedings of the 20th ACM International Conference on Multimedia*, MM '12, page 709–712, New York, NY, USA, 2012. Association for Computing Machinery. 2

[15] Joseph Redmon and Ali Farhadi. Yolo9000: Better, faster, stronger, 2016. 2

[16] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, 128(2):336–359, Oct 2019. 2

[17] Fahime Shafieyan, Nader Karimi, Ebrahim Nasr Esfahani, and Shadrokh Samavi. Image seam carving based on content importance and depth maps. 05 2014. 2, 4

[18] Evan Shelhamer, Jonathan Long, and Trevor Darrell. Fully convolutional networks for semantic segmentation, 2016. 2

[19] Avanti Shrikumar, Peyton Greenside, and Anshul Kundaje. Learning important features through propagating activation differences, 2019. 2

[20] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks, 2017. 2, 3