Journal of Discrete Mathematical Sciences & Cryptography
ISSN 0972-0529 (Print), ISSN 2169-0065 (Online)
Vol. 26 (2023), No. 3, pp. 939–949
DOI : 10.47974/JDMSC-1780

Neural network based image object detection and tracking for security and surveillance

Nishu Sethi *
Shalini Bhaskar Bajaj †
Department of Computer Science and Engineering
Amity School of Engineering and Technology
Amity University
Haryana
India

Abstract

The saliency map obtained from the source image determines the efficacy of the traditional seam carving process. The importance map proposed in this paper is used to highlight the shadows and important objects in the images. It combines the saliency map, shadow map, and gradient map acquired from the image to discover the image's prominent regions. The proposed map, when compared to others, highlights more distinct details with the state-of-the-art. The improved seam carving technique is paired with cropping and warping image retargeting operators in the suggested hybrid sequence. By labelling a picture with a class label and object localisation, the coordinates of the objects are generated using R-CNN object detection techniques. This will help in identifying the non-salient objects from the image for security and surveillance purposes with pin-point accuracy.

Subject Classification: *Primary 93A30, Secondary 49K15.*

Keywords: *Importance map, Seam carving, Image retargeting, Gradient map, Object detection.*

1. Introduction

As the Internet and technology improve, more people are using photos to share their knowledge and thoughts. Traditional image retargeting techniques are being improved by researchers in order to preserve the image's important portions. The following are the four types

* *E-mail:* **sethinishu@gmail.com (Corresponding Author)**
† *E-mail:* **sbhaskarbajaj@ggn.amity.edu**

of content-aware picture retargeting techniques: Cropping [1, 2, 3] Image retargeting based on segmentation [4,5], non-homogeneous warping [6], and seam carving [7]. The seam carving method delivers promising retargeting results with minimum image distortions. The saliency map produced by these techniques primarily emphasises the object's structural features based on colour and texture. The shadows of the items are vital in making the scene realistic and attracting the viewers' attention for a better understanding of the situation. The presence of shadow allows the user to notice important details such as object depth, light source angle, and other image characteristics. Existing methods [8] should be paired with enhanced saliency detection algorithms to improve the effectiveness of image retargeting. After integrating three various maps, such as saliency, shadow, and energy, a significance map is created. It is integrated with additional state-of-the-arts such as cropping, warping, and scaling to improve the efficiency of the suggested seam carving process. The suggested approach and various sequences of image retargeting operators are tested on various types of images.

Data support for aerospace activities such as satellite pointing and tracking can be provided via accurate picture coordinate detection. We researched the imaging characteristics of a satellite picture captured in space and suggested a method for obtaining the satellite component coordinates in the image to enable precise satellite component detection. This technology, which is based on a "regional-based convolutional neural network" (R-CNN), allows for accurate image coordinate detection.

2. Literature Review

The image's derivate can be computed using a variety of existing methods. The complexity of the picture features and variations in the color information determine the computing time of the image's salient features. Individual maps such as the salient map, shadow map, and energy map can be used to highlight the image's prominent sections.

Itti et. al., [9] suggested an approach through which bottom up saliency of the image can be computed based on low level features. For each low level feature of the image the individual saliency map is generated. The computed saliency map is further combined to obtain single saliency map.

Ji et. al., [10] multiple saliency maps are combined using low-level and abjectness properties in this architecture. The saliency map is optimised in the last step of the proposed technique by incorporating multiple features using multilayer cellular automata.

The content-aware seam carving technique was first proposed by Avidan and Shamir [11] which became very popular in image resizing. The energy of neighbouring pixels is compared to determine the best seam, and the pixel with the least energy is chosen for the seam. Their approach is based on intensity variation, which offers the significance map, and hence removes pixel paths with little intensity variation.

For refining the image scaling method, Achanta et. al., [12] suggested a novel saliency detection scheme and integrated it with the seam carving technique. An improved energy function was recommended to protect the image's normal structures which improves the existing seam carving technique. Liu et. al., [13] produced a multi-scale contrast based adaptive significance map that overcomes the seam carving technique's total limitations. They created a reserving ratio map with a mapping and re-sampling scheme. The significant objects can be preserved and the visual arrangement can be maintained utilising a continuous seam carving technique.

S. Dhyakesh et. al., (2021) [14] employed remote sensing satellite data to obtain information about the terrain. The term "object detection" refers to the problem of identifying items within an image. Object detection presents a number of difficulties, including non-trivial noise, blurring, and lower resolutions compared to aerial photos.

Shubhankar Rawat (2019) [15] employed deep learning models and RCNN to detect objects in aeroplanes, resulting in their widespread application. It is well established that the performance of deep learning models is proportional to the amount of training data. As a result of data augmentation, additional images are generated for training, which improves the model's performance.

Jakaria Rabbi et. al., (2020) [16] developed a three-component architecture consisting of an ESRGAN, an EEN, and a detection network. They employed residual-in-"residual dense blocks (RRDB) for the ESRGAN and EEN, and a faster region-based convolutional network (FRCNN) (two-stage detector) and a single-shot multibox detector (SSD) for the detector networks (one stage detector)."

Nidhi Bansal (2020) [17] Satellite imaging was utilised to track marine movements. Increased ship mobility, the possibility of ship accidents, and illegal activities have all prompted a number of organisations to keep a closer eye on the open seas.

3. Methodology

The steps of proposed model are shown in Figure 1.

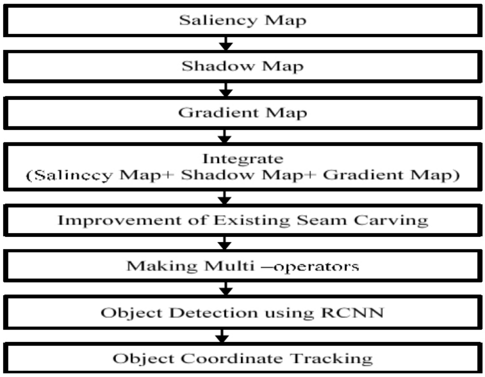

Figure 1

The Process Flow of Complete Research Methodology

3.1 *Saliency Map Extraction*

To compute an efficient saliency map, the segmentation based gPb-owt-ucm (Arbelaez et. al., [18]) method is applied. To obtain the level of the saliency correspond to each segment above three properties are combined. To obtain the final result of the saliency level of each segment *k*, eq. (1) is used.

$$S_i^k = \Pi_{j \in (c, b_0, b_1)} S_j^k \tag{1}$$

3.2 *Extraction of Shadow Mask using Region Growing & Gap filling*

We must conduct three primary processes to detect the shadow map of the input RGB image: pre-processing, shadow detection, and region filling utilising image morphological procedures.

The c 3 band was found to be suitable for recognising the shadow regions present in the image after examining multiple images in the SBU-shadow dataset (Vicente at al. [19]). The c 3 component has been found to be noisy in testing, which may cause misunderstanding between pixels in shadow and non-shadow regions. To address this problem, the c 3 component is convolved with the 33% average filter to decrease noise. The RGB to HSI conversion procedure[20] is used to extract the HSI components from the image.

3.3 *Gradient Map Extraction using Canny Operator*

To reduce noise and detect strong edges, the Gradient Map is created using the Canny Operator.

A filter with a defined dimension is applied to the image to create this style of image. Forward and reverse mapping are used to determine the relative locations of pixels from the source image to the target image. To map all the pixel locations from source image to target image eq (2). is used.

$$\begin{aligned} IMG'_x &= forward\_mapping(IMG_x) \\ IMG'_y &= forward\_mapping(IMG_y) \\ IMG'^{(x,y)} &= \mathrm{Re}\,sample \\ w1 &\approx 1/S_{scale}(IMG_x, IMG_y) \end{aligned} \tag{2}$$

3.4 *Integrate (Saliency Map + Shadow Map+ Gradient Map)*

All three maps, namely the saliency map, shadow map, and gradient map, are integrated to produce an accurate importance map. The produced significance map indicates the image's most prominent area where the distortion is plainly visible to the naked eye.

3.5 *Improvement of Existing Seam Carving*

Due to the intersection of optimal seams in the existing seam carving method, the contour of significant line structures is crushed. This is accomplished by retargeting the lines[21], then locating the intersection point and detecting the lines that correspond to it. After determining the lines and points, the energy is modified, and the updated procedure, which is now the enhanced seam carving process, is implemented.

The Vertical and Horizontal Seams are calculated by the formula below:

$$IV(i,j) = \begin{cases} \text{if } I(i,j) = 2DArray(i,j) = 1 \\ \text{otherwise, } y = y+1,\ x = x-1 \\ x = 1,2,3,4\ldots\ldots n \\ y = 1,2,3\ ,4\ldots\ldots.n \end{cases} \tag{3}$$

$$IH(i,j) = \begin{cases} \text{if } I(i,j) = 2DArray(i,j) = 1 \\ \text{otherwise, } y = y-1,\ x = x+1 \\ x = 1,2,3,4\ldots\ldots n \\ y = 1,2,3,4\ldots\ldots.n \end{cases} \tag{4}$$

3.6 *Object Detection using R-CNN Architecture*

Ross Girshick, Shaoqing Ren, Kaiming He, and Jian Sun introduced RCNN, a convolutional neural network-based object detection architecture, in 2015. Along with YOLO (You Look Only Once) and SSD, it is one of the most well-known object identification architectures that makes use of convolutional neural networks (Single Shot Detector). Its architecture is shown in Figure 2.

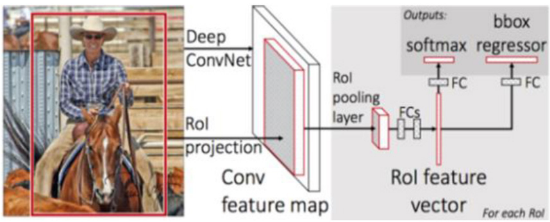

Figure 2

R-CNN Architecture (Source: Girshick, 2015)

4. Results and Discussion

4.1 *Proposed Importance Map (PIM)*

From the Figure 3 the contours of the objects, the various boundaries of the structure, and their shadows may all be seen in a single significance map. When the proposed importance map is used in combination with the proposed seam carving approach, the structure of the significant regions can be preserved while the image is retargeted.

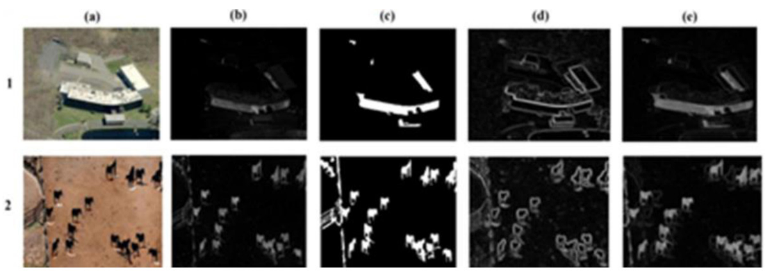

Figure 3

Comparison of (a) Original Image (b) Saliency Map (c) Shadow Map (d) Energy Map (e) Proposed Importance Map (PIM)

Figure 4 shows the results of a similar experiment on other images to produce a saliency map and a significance map. The proposed technique was used to create the importance map. The obtained saliency map in Figure 4 highlights few prominent structures of the image.

4.2 *Improved Seam Carving (ISC)*

In figure 5(b) the line structures on the building's roof have been warped and changed into curved shaped items, as can be seen. This type of distortion can arise in the traditional seam carving process due to the meeting of ideal seams in both the horizontal and vertical directions. When the ideal seam intersects on prominent line

structures and removes the low energy pixels from the line structures, the shape of the lines starts to convert into curved shaped objects in the traditional seam carving procedure. To solve this challenge, the suggested technique uses the canny line detection operator to first detect the significant line structures. In figure 5(c), the straight lines seen on the building's roof have been totally preserved. Only a few distortions can be seen on the ground line segments that are not immediately visible to the naked eye.

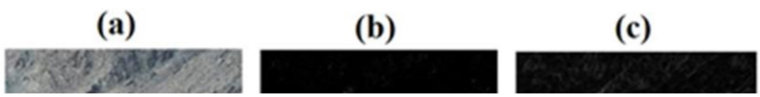

Figure 4

Comparison of (a) Original Image (b) Saliency Map (c) PIM

Figure 5

Comparison of (a) Original Image (b) Conventional Seam Carving (c) Improved Seam Carving

4.3 *Comparison with other existing operators*

The findings of the improved seam carving (ISC) technique were compared to existing state-of-the-art techniques such as shift-Map(SM) and stretch and scale (SNS) and displayed in Figure 6,Various photos, including a range of line structures, curved shaped objects, and shadow regions, were captured to test the efficiency of the improved seam carving approach. From figures 6(b), 6(c), and 6(d), in comparison to the SNS and the improved seam carving technique, the image retargeting results obtained by the SM cause a greater amount of distortions on the line segments. When compared to the SM and SNS retargeting operators, the suggested approach produces visually pleasing results because image distortion is very low.

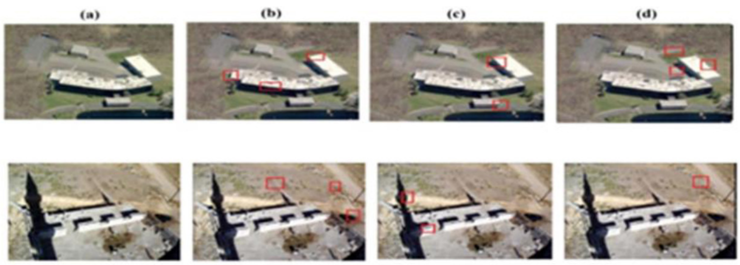

Figure 6

Comparison of (a) Original Image (b) SM (c) SNS (d) Improved Seam Carving

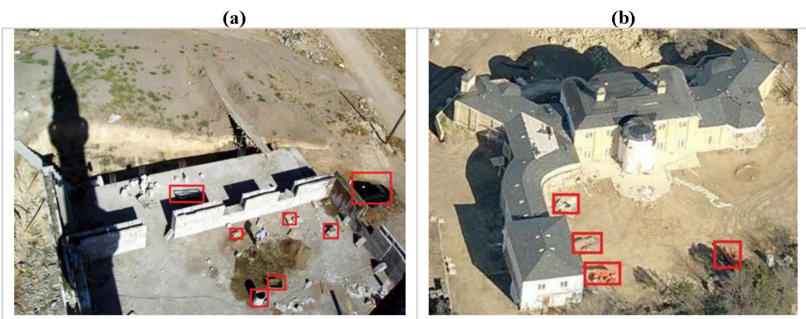

Figure 7

Object Detection using RCNN

Coordinate Retarget of Vehicle to Common Google Earth Images

We collected images from Google Earth and applied R-CNN inception. The images were collected and some objects are identified with

Figure 8

Training Loss and Accuracy Curve of the images

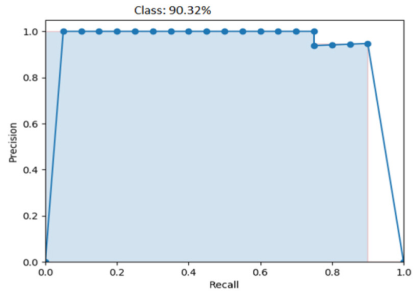

Figure 9

R-CNN Model – Curve for Accuracy of location retargeting

common Google earth images (Figure 7 (a) and (b)). More accurate results are shown in Figure 8 because it has minimum loss. These red boxes would solve the security application of the work.

5. Conclusions & Outcomes

The saliency map, which highlights the image's most important portions, determines how effective the image retargeting process is. The traditional saliency detection method simply highlights the most important things and ignores shadow as a prominent structure. The observer can comprehend the vital contents such as the strength of light incident on the object, the depth of the object, the angle of the incident light, and other significant elements of the objects due to the presence of shadows. The proposed technique offers considerable results in terms of reduced distortion percentage when compared to the state of the art. The typical images were retargeted and categorised, R-CNN is used to find the coordinates of images. In terms of object detection and coordinate information of retargeted images of satellite views, the vehicle marked images and curve presented in figure 9 were discovered to be the most accurate. This pin-point accuracy is very much useful for security and surveillance purposes for the identification of the infiltration activities occurring at the border regions of any country.

References

[1] B. Suh, H. Ling, B. B. Bederson, and D. W. Jacobs, "Automatic thumbnail cropping and its effectiveness," *UIST Proc. Annu.*

ACM Symp. User Interface Softaware Technol., pp. 95-104 (2003), doi: 10.1145/964696.964707.

[2] Santella, M. Agrawala, D. DeCarlo, D. Salesin, and M. Cohen, "Gaze-based interaction for semi-automatic photo cropping," *Conf. Hum. Factors Comput. Syst. - Proc.*, vol. 2, pp. 771–780 (2006), doi: 10.1145/1124772.1124886.

[3] M. Nishiyama, T. Okabe, Y. Sato, and I. Sato, "Sensation-based photo cropping," *MM'09 - Proc. 2009 ACM Multimed. Conf. with Co-located Work. Symp.*, pp. 669-672 (2009), doi: 10.1145/1631272.1631384.

[4] Y. S. Wang, C. L. Tai, O. Sorkine, and T. Y. Lee, "Optimized scale-and-stretch for image resizing," *ACM Trans. Graph.*, vol. 27, no. 5, pp. 1-8 (2008), doi: 10.1145/1409060.1409071.

[5] V. Setlur, S. Takagi, R. Raskar, M. Gleicher, and B. Gooch, "Automatic image retargeting," *ACM SIGGRAPH 2004 Sketches, SIGGRAPH'04* (2004), doi: 10.1145/1186223.1186228.

[6] Y. Niu, F. Liu, X. Li, and M. Gleicher, "Image resizing via non-homogeneous warping," *Multimed. Tools Appl.*, vol. 56, no. 3, pp. 485-508 (2012), doi: 10.1007/s11042-010-0613-0.

[7] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1-9 (2008), doi: 10.1145/1360612.1360615.

[8] Garg, Ankit, Ashish Negi, and Prakhar Jindal. "Structure preservation of image using an efficient content-aware image retargeting technique." *Signal, Image and Video Processing* (2020): 1-9.

[9] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 11, pp. 1254-1259 (1998).

[10] Y. Ji, H. Zhang, K. Tseng, T. W. Chow, and Q. J. Wu, "Graph model-based Salient object detection using objectness and multiple saliency cues," *Neurocomputing*, vol. 323, pp. 188-202 (2019).

[11] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *Proc. ACM SIGGRAPH Conf. Comput. Graph.* (2007), doi: 10.1145/1275808.1276390.

[12] R. Achanta and S. Süsstrunk, "Saliency detection for content-aware image resizing," *Proc. - Int. Conf. Image Process. ICIP*, no. Ic, pp. 1005–1008 (2009), doi: 10.1109/ICIP.2009.5413815.

[13] Z. Liu, H. B. Yan, L. Q. Shen, et. al., "Adaptive image retargeting using saliency-based continuous seam carving," Optical Engineering, vol. 49, no. 1, pp. 017002 (2010).

[14] Dhyakesh, S., et. al., Mask R-CNN for instance segmentation of water bodies from satellite image. In *EAI/Springer Innovations in Communication and Computing,* pp. 301-307 (2021). Springer International Publishing.

[15] Rawat, S., *Airplanes Detection for Satellite using Faster RCNN*. Towards Data Science (2019). https://towardsdatascience.com/airplanes-detection-for-satellite-using-faster-rcnn-d307d58353f1.

[16] Rabbi, J., Ray, N., Schubert, M., Chowdhury, S., & Chao, D. Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network. *Remote Sensing*, 12(9), 1432 (2020).

[17] *Bansal, N., Ship detection in satellite images using mask R-CNN*. Spell.Ml. Retrieved August 5, 2021 (2020), from https://spell.ml/blog/ship-detection-in-satellite-images-using-mask-r-cnn-XyMA4xEAACUAb9G1

[18] Arbelaez, Pablo, Michael Maire, Charless Fowlkes, and Jitendra Malik. "Contour detection and hierarchical image segmentation." *IEEE transactions on pattern analysis and machine intelligence* 33, no. 5, 898-916 (2010).

[19] Vicente, Tomás F. Yago, Le Hou, Chen-Ping Yu, Minh Hoai, and Dimitris Samaras. "Large-scale training of shadow detectors with noisily-annotated shadow examples." in *European Conference on Computer Vision*, pp. 816-832 (2016). Springer, Cham.

[20] Sharma, Gajanand, et. al., "Video tempering detection assessment in full reference mode using difference matrices." *Journal of Discrete Mathematical Sciences and Cryptography* 22.4, 645-659 (2019).

[21] Kumar, Ankit, et. al., "Improving the visual quality of a size deterministic visual cryptography scheme for Grayscale Images." *Journal of Discrete Mathematical Sciences and Cryptography* 25.4, 1113-1123 (2022).