Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 44 (2023), No. 6, pp. 1035–1042
DOI : 10.47974/JIOS-1331

Modified ResNet50 model and semantic segmentation based image co-saliency detection

Anuj Mangal *
Hitendra Garg †
Charul Bhatnagar §
Department of Computer Engineering & Applications
GLA University
Mathura
Uttar Pradesh
India

Abstract

Co-salient object identification is a new and emerging branch of the visual saliency detection technique that tries to find salient pattern appearing in several image groups. The proposed work has the potential to benefit a wide variety of important applications including the detection of objects of interest, more robust object recognition and animation synthesis, handling input image query, 3D object reconstruction, object co-segmentation etc. To build modified ResNet50 model, the hyperparameters are adjusted in the current work to increase accuracy while minimizing loss. The modified network is trained on HOG features to mine more significant features along with their corresponding ground truth images. For a more streamlined outcome, the proposed system was built using the SGDM optimizer. During testing among the relevant and irrelevant image the network generates appropriate co-saliency map of relevant images. Integrating the associated and prominent characteristics of the image yields the appropriate ground truth for each image. The proposed method reports better F1 value 98.7% and MAE score 0.089 value when compared with SOTA model.

Subject Classification: *68T07*

Keywords: *Semantic segmentation, Co-saliency, CoSOD3k, ResNet model, HOG features, Semantic Segmentation.*

* *E-mail:* **anuj.mangal@gla.ac.in (Corresponding Author)**
† *E-mail:* **hitendra.garg@gla.ac.in**
§ *E-mail:* **charul@gla.ac.in**

1. Introduction

People tend to focus their visual attention on places and things that are captured through their eyes and hold their attention for a while to acquire information [1]. In order to comprehend the scenario, CoSOD model replicates the human visual system and searches for corresponding prominent objects [2] in a collection of image set. The CoSOD [3] model is used for various image processing task. In order to precisely describe the most important details of an image, important representative characteristics must be extracted. Additionally, a quick and efficient computational methodology must be created in order to produce and identify common patterns. Convolutional neural networks (CNNs)can capture far-reaching correlations by expanding fields of receptivity by overlaying convolution layers over low-resolution input data. On the other hand, multi-hop dependency modeling is made more difficult by recurrent convolution approaches [16]. The framework of the proposed article is as outlined below: The proposed work is described in section 2 which details about the dataset pre-processing phase and optimization of the proposed model. The comparison study with other models over dataset is presented in Section 3. The intended work is concluded in Section 4.

2. Proposed Work

The co-saliency approach's principal goal is to separate recurrent and salient events from background noise. The dataset images [13] are split into 70%:30% ratio along with their corresponding ground truth images. The suggested method pre-trains the model using the retrieved HOG features from every group of dataset sample. Every sample image along

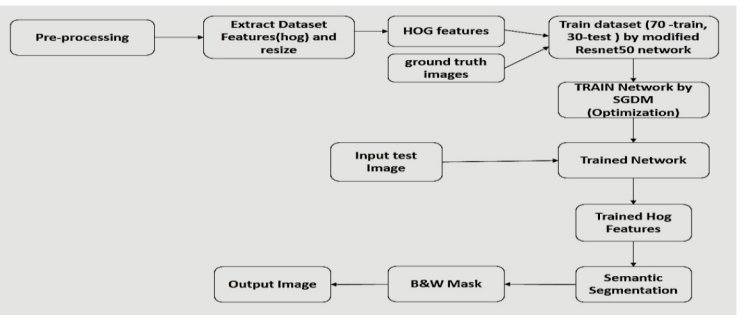

Figure 1

The abstract view of the proposed model

with their ground truths are rescale to 224 x 224 to match the input for existing ResNet 50 [15] network. By including HOG characteristics and the ground truth of the related source image, the precision and effectiveness of the suggested technique are increased. In the suggested model, hyperparameters are also altered to increase accuracy while minimizing loss. The network preserves the output parameters of the original fully linked layer while reducing the total no. of system elements. The concept model of the proposed approach is depicted in Figure 1.

HOG is popularly known as a feature descriptor technique for object identification in image processing and computer vision field. This descriptor method counts the number of times in an image that a detection window or region of interest has a gradient orientation (ROI). The detailed description with algorithm is mentioned below.

1. Preprocess the Input Image Ti
   a. Resize the image to 64 x 128 and in 8 x 16 grid.
   b. Every block of 8 x 8 pixel.

1. Calculating the gradient of the image
   a. Calculate the G_x and G_y of each pixel:

      $G_x(a,b) = I(a, b + 1) - I(a, b - 1)$; $G_y(a, b) = I(a - 1, b) - I(a + 1,b)$

      where G_x and G_y are the gradient of a pixel in x and y direction and a, b as row and column respectively.
   b. Calculate magnitude and angle of each pixel of the block :

$$\text{Magnitude } (\mu) = \sqrt{G_x^2 + G_y^2}$$

$$\text{Angle } (\theta) = \left| tan^{-1} - 1\left(\frac{G_y}{G_x}\right)\right|$$

2. Obtain Histogram of gradient
   a. Develop 9 bin histogram separated by 20 degree representing intensity of gradient in that bin.
   b. Bin in histogram are assigned with some values, calculated from the matrices of magnitude and angle.
   c. Create a new block of 2 x 2 size.
   d. Keep stride of 8 pixels for grouping purpose.

e. Concatenate cells of block to form 36 feature vector.

3. Normalize the histogram to minimize the effect of contrast variation in the image.

4.1 Calculate p and f_{bi} value:

$$p = \sqrt{b_1^2 + b_2^2 + b_3^2 + \ldots + b_{36}^2}$$

$$f_{bi} = \left[\left(\frac{b1}{p}\right), \left(\frac{b2}{p}\right), \left(\frac{b3}{p}\right) \ldots\ldots \left(\frac{b36}{p}\right)\right]$$

4. Calculate the cumulative HOG feature vector

a. Collect all 36 point vectors from each block.

b. Concatenate them into one giant vector which will be total 3780 features.

The SGDM optimizer is used throughout the training phase in order to get the network to produce the least amount of loss. A quicker convergent time and improved outcomes are obtained by accelerating gradient vectors in the desired orientations using SGD optimizer. The SGDM method alters the weights and biases of the network in order to reduce the overall loss and ahead to local minima. The resulting formula produces an entirely new vector V using exponentially weighted average:

$$V_t = \beta V_{t-1} + (1-\beta) S_t \tag{1}$$

Where V stands for vector, where V_{t-1} preceding assigned data fact, S_t is newly assigned data facts and β denotes hyper-parameter value between 0 - 0.9.

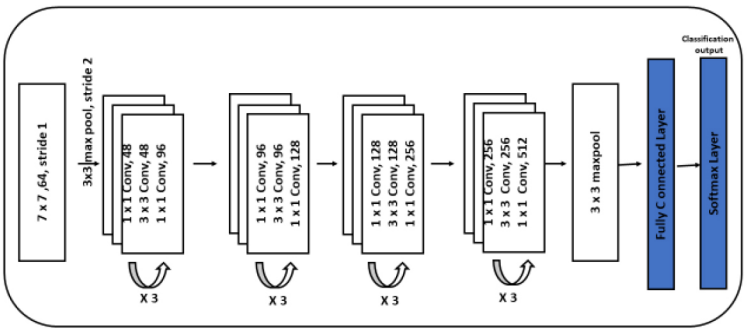

Figure 2

The details of proposed model architecture

$$W_n = W_o - \alpha V_t \tag{2}$$

Where W_n denotes updated weight, W_o is old weight, α learning rate which is assigned with 0010 value. We have kept less parameters than existing ResNet-50 model which is mentioned in Figure 2.

3. Results and discussion

The CoSOD3k dataset [26] used in the proposed model is one of the biggest dataset. The below metrics used for system performance are defined below:

$$\text{Pre} = \frac{\text{Tp}}{\text{Tp}+\text{Fp}} \quad \text{Re} = \frac{\text{Tp}}{\text{Tp}+\text{Fn}} \quad \text{F1}-\text{score} = 2 * \frac{\text{Pre} * \text{Re}}{\text{Pre} + \text{Re}} \tag{3}$$

$$\text{MAE} = \frac{1}{H \times W} \sum_{x=1}^{H} \sum_{y=1}^{W} \left| S(x,y) - G(x,y) \right| \tag{4}$$

where H is height and W is width of image and Tp, Fp, Fn, and Tn stand for true-positive, false-positive, false-negative, and true-negative values, respectively. With a batch size of 20, the proposed network's starting learning rate was set at 1e-4. Nineteen epochs were used to train the network. The network was built with the SGDM optimizer to more efficiently reduce the loss function for a more optimal result. The suggested model's training accuracy and loss were 99.39% and 0.04%, respectively.

4.2 *Qualitative Results*

The proposed model's output visible outcomes are contrasted with representative SOTA models, such as those in Figure 3.

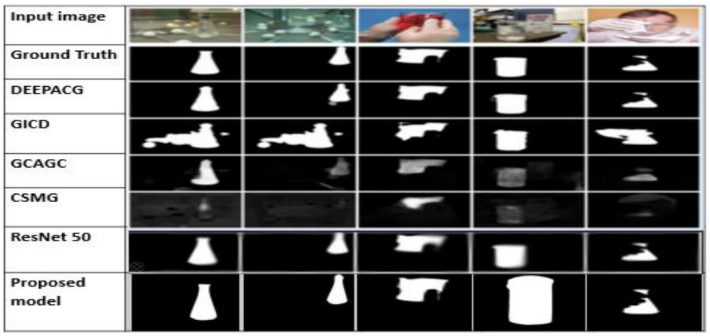

Figure 3

Visual outcome of proposed network compared with SOTA model on beaker

Table 1

Quantitative comparison table

Ref.	Model	F1- SCORE	MAE
[5]	BASNet	0.736	0.122
[4]	EGNet	0.782	0.119
[8]	PoolNet	0.797	0.12
[6]	SCRN	0.806	0.118
[7]	CBCD	0.589	0.228
[9]	ESMG	0.615	0.239
[16]	CODR	0.645	0.229
[10]	DIM	0.458	0.327
[11]	CSMG	0.645	0.157
[12]	SSNM	0.675	0.12
[14]	GCAGC	0.74	0.1
[13]	GICD	0.743	0.189
[15]	ResNet-50	0.927	0.122
MODEL	Mod. ResNet	0.987	0.089

As we can see from the result Table no. 1 the network we suggest performs better than the original ResNet-50 model when compared. Figure 4 displays a performance graph contrasting the suggested network with several standard SOTA models, clearly demonstrating the high F1-score and lowest MAE value when compared to others.

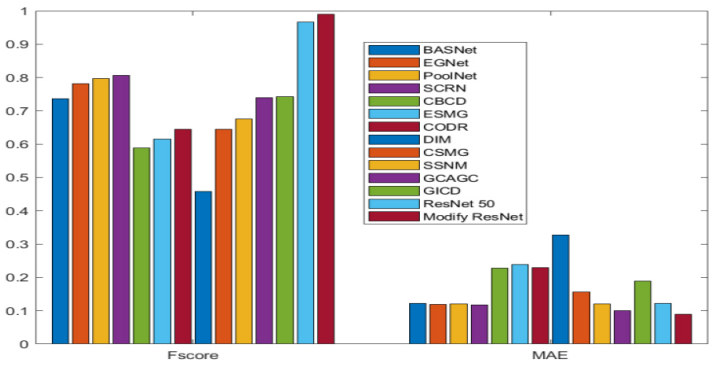

Figure 4

Progress graph compared with SOTA models

5. Conclusion

The suggested article uses a modified version of the ResNet network model to accomplish co-saliency identification process. This study proposes a new deep learning method for strong co-saliency detection task. The suggested work is evaluated using the ground truth images from the standard CoSOD3k dataset. We calculate HOG features of the images from the standard dataset and train the modified ResNet-50 model along with their corresponding ground truth value. The proposed work is being tested on 30% of the dataset images, followed by their HOG feature extraction. The proposed method comes with 98.7% F1-score and 0.089 MAE value which is quite good with state of art methods. In future additional visual characteristics can be employed in the existing model and research on motion features for video co-saliency identification can be carried out.

Conflicts of Interest

The authors declare that they have no competing interests.

References

[1] Zhang, K., Dong, M., Liu, B., Yuan, X. T., & Liu, Q., DeepACG: Co-Saliency Detection via Semantic-aware Contrast Gromov-Wasserstein Distance. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13703-13712 (2021).

[2] Sharma, P., Sharma, M. S. P., & Tomar, R. S. : A new approach for image segmentation using improved k-means and ROI saliency map. *Journal of Information and Optimization Sciences*, 38(6), 927-935 (2017).

[3] Qian, X., Zeng, Y., Wang, W., & Zhang, Q., Co-saliency Detection Guided by Group Weakly Supervised Learning. IEEE Transactions on Multimedia (2022).

[4] Tsai, C. C., Hsu, K. J., Lin, Y. Y., Qian, X., & Chuang, Y. Y., Deep co-saliency detection via stacked autoencoder-enabled fusion and self-trained cnns. *IEEE Transactions on Multimedia*, 22(4), 1016-1031 (2019).

[5] Gao, G., Zhao, W., Liu, Q., & Wang, Y., Co-saliency detection with co-attention fully convolutional network. IEEE Transactions on Circuits and Systems for Video Technology, 31(3), 877-889 (2020).

[6] Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., & Jagersand, M., Basnet: Boundary-aware salient object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 7479-7489 (2019).

[7] Ye, L., Liu, Z., Li, J., Zhao, W. L., & Shen, L., Co-saliency detection via co-salient object discovery and recovery. IEEE Signal Processing Letters, 22(11), 2073-2077 (2015).

[8] Fu, H., Cao, X., & Tu, Z., Cluster-based co-saliency detection. IEEE Transactions on Image Processing, 22(10), 3766-3778 (2013).

[9] Wei, L., Zhao, S., Bourahla, O. E. F., Li, X., & Wu, F., Group-wise deep co-saliency detection (2017). arXiv preprint arXiv:1707.07381.

[10] Qian, X., Zeng, Y., Wang, W., & Zhang, Q., Co-saliency Detection Guided by Group Weakly Supervised Learning. IEEE Transactions on Multimedia (2022).

[11] Zhang, D., Han, J., Han, J., & Shao, L., Cosaliency detection based on intrasaliency prior transfer and deep intersaliency mining. IEEE transactions on neural networks and learning systems, 27(6), 1163-1176 (2015).

[12] Wei, L., Zhao, S., Bourahla, O. E. F., Li, X., & Wu, F., Group-wise deep co-saliency detection (2017). arXiv preprint arXiv:1707.07381.

[13] Zhang, Z., Jin, W., Xu, J., & Cheng, M. M., Gradient-induced co-saliency detection. In European Conference on Computer Vision, pp. 455-472. Springer, Cham (2020, August).

[14] Zhang, K., Li, T., Shen, S., Liu, B., Chen, J., & Liu, Q., Adaptive graph convolutional network with attention graph clustering for co-saliency detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 9050-9059 (2020).

[15] Simonyan, K., & Zisserman, A., Very deep convolutional networks for large-scale image recognition (2014). arXiv preprint arXiv:1409.1556.

[16] Li, T., Zhang, K., Shen, S., Liu, B., Liu, Q., & Li, Z., Image co-saliency detection and instance co-segmentation using attention graph clustering based graph convolutional network. IEEE Transactions on Multimedia (2021).

[17] Sahoo, D. K., Das, A., Mohanty, M. N., & Mishra, S., Brain tumor detection using in painting and deep ensemble model. *Journal of Information and Optimization Sciences*, 43(8), 1925-1933 (2022).