Journal of Information Technology Management

Print ISSN: 2008-5893

Online ISSN: 2423-5059

Assessing the performance of Co-Saliency Detection method using various Deep Neural Networks

Anuj Mangal *

*Corresponding Author, Department of Computer Engineering & Applications, GLA University, Mathura. E-mail: anuj.mangal@gla.ac.in

Hitendra Garg

Department of Computer Engineering & Applications, GLA University, Mathura. E-mail: hitendra.garg@gla.ac.in

Charul Bhatnagar

Department of Computer Engineering & Applications, GLA University, Mathura. E-mail: charul@gla.ac.in

Abstract

Co-Saliency object detection is the process of identifying common and repetitive objects from the group of images. Earlier studies have looked over several state-of-art deep neural network methodologies for co-saliency detection approach. The Deep CNN approaches rely heavily on co-saliency detection due to their potent feature extraction capabilities both deep and wide. This article assess the performance of several state-of-art deep learning model (VGG19, Inceptionv3, modifiedResNet, MobileNetV2 and PoolNet) for the purpose of co-saliency detection among images from benchmark datasets. All the models were trained on 70% part of the dataset and remaining were used for testing purpose. Experimental results show that modified ResNetmodel outperforms getting 96.53% accuracy as compared to other state-of-the-art deep neural network models.

Keywords: CNN, Co-Saliency detection, SGDM, ADAM, RMS, VGG19, Inceptionv3, ResNet, MobileNet and PoolNet.

Journal of Information Technology Management, 2023, Vol. 15, Special Issue, pp. 23- 34
Published by University of Tehran, Faculty of Management
doi: https://doi.org/ 10.22059/jitm.2023.95243
Article Type: Research Paper

Received: July 06, 2023
Received in revised form: August 24, 2023
Accepted: November 09, 2023
Published online: December 24, 2023

Introduction

Everywhere Co-saliency has been utilized to support the understanding of visual substance in an assortment of uses because of its higher versatility picture/video co-division, object co-limitation, content mindful pressure. Researchers anticipate that by using saliency detection models (Sun et al., 2018) the human visual attention process can be imitated, as machine will be able to automatically pick the most exciting visual components in natural images. Researchers have focused on a fundamental difficulty that emerges in a computer vision application, such as; content-based image retrieval (Zhang et al., 2016), salient object segmentation (Yang et al., 2022), semantics segmentation (Bai et al., 2018), and scene categorization (Zhang et al., 2016). Saliency detection can be further classified into co-saliency detection, video saliency detection (Wang et al., 2020), and RGBD saliency detection (Jeong et al., 2018).

Object co-segmentation (Fan et al., 2020), object co-recognition (Fu et al., 2013), and weakly supervised localization (Song et al., 2017) are some of the applications of co-saliency detection, which focuses on the common salient objects in a series of relevant images. For decades, researchers have been working on how to determine the saliency of a single image. In (Ullah et al., 2020) a unique framework for saliency detection that generates high-quality regional and pixel-wise saliency maps were presented. In order to generate a collection of suggestion groups, the researchers devised an optimization technique (Han et al., 2017) that yielded several ideas with a high degree of shared saliency in the original picture. Co-saliency maps are created for each proposal group and then combined using a low-rank technique to create the final image's co-saliency map.

Methodology

Semantic segmentation

The role of semantic segmentation plays a very important part in understanding the image properties which helps in the task related to analysis work. Semantic segmentation is the process of identifying an image class and isolating it from the other image classes by covering a segmentation mask over it. Basically it is a three step process: (a) classification of object, (b) localizing the objects within the image, (c) masking grouped pixels in a localized part of image (segmentation). In order to obtain relatable correlation of the input image and effectively reducing the noise, there is a need of semantic segmentation which extracts the features and various representation of the inputted image.

Pre-trained Model

A pre-trained model has previously undergone training on benchmarked, sizeable dataset, used for a wide ranging image-classification task. A pre-trained model has weights and biases pre-loaded into it because it was trained on a specific set of data. These weights and biases show the specific attributes that are unique to that dataset.

ResNet-50 model

Residual Network is referred to as ResNet50. He et al. (2016) first proposed this novel neural network in computer vision research paper. This model was incredibly effective, as evidenced by the fact that it took first place in the 2015 ILSVRC classification competition with a mere 3.57 percent error. There are 48 convolutional layers in the ResNet50 model, as well as a Max Pool layer and one average pooling layer. It can deal with 3.8x 109 floating point operations.

MobileNetV2 model

Convolutional neural networks are used in the MobileNetV2 architecture (Han et al., 2017) to make it run well on mobile devices. This network is of 54 layers and takes input image of 224 x 224 sizes. This approach works successfully because the bottleneck layers are linked by an inverted residual structure. To filter features, the middle expansion layer leverages lightweight deep convolutions giving non-linearity model. In overall design of the MobileNetV2, a 32 filter convolution is used trailed by 19 bottleneck layers. MobileNetV1 includes layer 1 and 2. The first layer is referred to as a depth wise convolution, and it performs light filtering by applying a single convolutional filter for each input channel in the first layer. The second layer (1 x 1) convolution is used to develop extra features and attributes by linearly combining the input channels. There are two different kinds of blocks in MobileNetV2. First one is residual block with stride 1. Second one is used for downsizing with stride 2.

PoolNet model

The PoolNet model (Ren et al., 2018) proves to give excel result when tested against several robust models (like ResNet, VGG etc.), produces one of the most cutting-edge salient object identification outcomes. This model incorporates pooling operation which is beneficial for deep and shallow features illustrations. This network comprises of two main modules working together for the salient area's feature representation named as Feature Aggregation Module (FAM) and the Global Guidance Module (GGM). One of the modules is responsible for extracting deep features and the other one shallow feature.

Inception-v3 model

Inception-v3 convolutional neural network (Li et al., 2018) design uses factorized 7x7 convolution factorization, label smoothing, and the use of an extra classifier for label propagation (additionally, layers in the side head are batch-normalized). It has total 42 layers making it more deep architecture and producing minimal error ratio compared with earlier versions.

Dataset

The CoSal 2015 dataset (Wei et al., 2017) and CoSOD3k dataset (Ye et al., 2017) are used with our model's performance comparison. The CoSOD3k is a dataset with more realistic settings. The complexity, challenge, and scalability of CoSOD3k represent a considerable advance for associated visual tasks. The CoSal2015 is a great dataset to start if the work needs to be faced with challenging factors. The proposed method is tested using these three different co-saliency benchmark datasets.

Optimizers

Optimizers are the programs or methods that are responsible for changing the properties of neural network, like theirs weights and learning rate, for minimizing the loss and give more accurate possible results. By decreasing the loss function, optimizers can resolve optimization challenges. The weights are initially assigned with some pre-initialized values and with each epoch their weights are adjusted in accordance with the updated equation.

$$Wn = Wo - \alpha * \partial L / \partial W \quad (1)$$

Where Wn is new weight, WO is old weight, α is leaning rate, $\partial L/\partial W$ is partial derivate of loss function. The equation mentioned above is used for weight updation in order to produce more accurate result. By utilizing various optimization techniques or so-called optimizers, the best outcome can be obtained. In below section some of the optimizers are used for training purpose for producing best result and minimize the errors.

Stochastic Gradient Descent with Momentum (SGDM) optimizer

SGDM optimizer and its variants can handle a wide range of difficult learning tasks, including those requiring deep learning (Ullah et al., 2020). The use of momentum (Li et al., 2018) is an approach that can be used to both accelerate SGD and reduce oscillations by denoising the gradients. It is equal to 0.9. The updates in weights are highly dependent on noisy derivatives and by de-noising the derivatives; the convergence time can be shorten. As a result, the amount of fluctuation that happens in momentum is reduced. The goal is to use exponential weighting average to de-noise the derivative by which new updates values are accepted over old ones at time momentum t which accelerates to convergence.

$W_n = W_o - \alpha * \partial L/\partial$
$V_n = \alpha * V_o - \alpha * \partial L/\partial W$ (2)
$W_n = W_o - V_n$
$0 < \alpha < 1$

Where Vn = sum of square of past gradients. To obtain new weights, the gradient is subtracted from the old weight making result to move in one direction.

ADAM optimizer

ADAM optimizer algorithm originates name from adaptive moment estimation. It combines the features of Adaptive gradient descent optimizer (Zhang et al., 2020) and RMS prop algorithms (Su et al., 2022). In this approach, the gradient descent algorithm is speed up by using an 'exponentially weighted average' of the gradients. This is done to shorten the time required to perform the algorithm. With the help of averaging, the process of determining the best response can be accomplished more rapidly. This algorithm is easy to implement, quick run, use less memory and need fewer adjustments.

$W_{t+1} = W_t - \alpha * g_t$ (3)
Where,
$m_t = \beta * g_{t-1} + (1 - \beta)[\frac{\delta L}{\delta w}]$

Here g_t denotes gradient collection at current time t , g_{t-1} denotes gradient collection at previous time aggregates, Wt denotes weight at t, Wt+1 denotes weight at t+1, αt denotes learning rate, δL denotes loss function derivatives, ∂W denotes weights derivatives i.e. sum(∂L/∂Wt-1)) at time t and β represents moving average parameter.

RMSprop Optimizer

RMSprop also known as Root Mean Square prop optimizer is analogous to the gradient descent technique with momentum. It is feasible to limit vertical oscillations by using the RMSprop optimizer. It introduced the exponential moving average of the squared gradients, making learning rate significantly increased and the algorithm made larger horizontal jumps in order to get faster convergence. RMSprop and gradient descent are distinguished by the technique in which gradients are computed.

$$w_{t+1} = w_t - \frac{\alpha_t}{(v_t+\varepsilon)^{1/2}} * \left[\frac{\delta L}{\delta w_t}\right]$$
$$v_t = \beta v_{t-1} + (1-\beta) * \left[\frac{\delta L}{\delta w_t}\right]^2 \quad (4)$$

Here Wt denotes weight at t, Wt+1 denotes weight at t+1, αt denotes learning rate, δL denotes derivative of loss function, ∂Wt denotes derivative of weights i.e. sum(∂L/∂Wt-1)) at time t, v_t denotes square of past gradients, β denotes moving average parameter (0.9), ε denotes small constant (10-8).

Results

Some of the latest deep CNN architectures, such as VGG19, InceptionV3; ResNet50, MobileNet and PoolNet are used for the co-saliency assessment. Various dataset are used on these latest deep CNN architectures. The section below discuss about the model training and testing outcomes. This section evaluates the performance of suggested model for co-saliency identification. To get a decent detection performance, the consistency and difference of single-saliency and co-saliency image features must be considered. The images are annotated by placing them into different categories folders to train different convolutional neural network. The dataset was sampled 70% of all images from each class and include them in the training set. And the 30% of remaining images are placed in the test set. All images are sized into 224 × 224 pixels. The model's initial learning rate was fixed at 1e-3 with a batch size of 15. Network was trained for 20 iterative epochs. The proposed work details about training accuracy with different model, dataset and optimizer shown in in table 1.

Table 1. Training Accuracy and Loss with different model, dataset and optimizer

Pre-Trained Model	Optimizer	CoSOD3k		Cosal2015		COCO		CoSOD3k	Cosal2015	COCO
		Training Accuracy	Training loss	Training Accuracy	Training loss	Training Accuracy	Training loss	Time (min., sec.)		
VGG-19	ADAM	98.23	0.012	97.21	0.025	98.76	0.051	1423,23	1256,11	1125,36
	SGDM	98.44	0.051	98.24	0.014	97.11	0.047	1325.12	1458,36	1526,58
	RMS	98.56	0.014	97.21	0.025	98.74	0.026	1236,14	1256,11	1236,58
Inception-V3	ADAM	97.87	0.084	97.23	0.026	97.25	0.025	1369,15	1025,36	1478,11
	SGDM	97.85	0.089	98.21	0.014	98.25	0.069	1693,12	1526,15	1256,12
	RMS	98.78	0.015	98.22	0.025	97.81	0.036	1456,36	1458,51	1458,14
MobileNetV2	ADAM	96.56	0.067	97.14	0.055	97.51	0.045	1569,15	1581,36	1258,14
	SGDM	98.96	0.026	96.48	0.089	98.21	0.074	1325,12	1693,23	1428,69
	RMS	97.78	0.023	98.92	0.045	98.02	0.048	1425,15	1561,21	1425,26
ResNet-50	ADAM	99.62	0.058	98.23	0.014	99.21	0.15	1456,25	1698,88	1256,14
	SGDM	98.53	0.002	97.25	0.018	99.36	0.0174	1458,36	1475,25	1258,28
	RMS	98.21	0.025	99.36	0.025	99.32	0.026	1526,25	1569,12	1472,14
PoolNet	ADAM	98.58	0.369	98.56	0.056	97.24	0.026	1369,25	1425,23	1256,89
	SGDM	97.56	0.041	97.00	0.078	98.36	0.035	1256,14	1896,23	1526,69
	RMS	90.58	0.0732	96.26	0.061	97.25	0.036	1259,36	1742,14	1258,89

ModelTesting

In testing phase, we chose representative datasets to assess every trained model. On inputting test image, the co-salient object was detected from the final layer of the deep neural network. Both training and testing was performed onNvidiaGeforce GTX TITAN X GPU.

Qualitative Result

This section represents various qualitative results obtained from the three benchmark datasets CoSOD3k, CoSal2015 represented by figure 1, figure 2respectively. These figures represent co-saliency maps produced by various state-of-the-art pre-trained models i.e. VGG19, InceptionV3, MobileNetV2, ResNet50 and PoolNet. Among all of them, it can be visually noticed that modified ResNet produces output very close to the ground truth. It can be further noticed that the poorest result was obtained from MobileNetV2 model.

Quantitative Result

This section assesses the performance of state-of-the-art deep learning models that employ the semantic segmentation technique for CoSOD. All of the above mentioned pre-trained models (VGG-19, Inception-V3, MobileNetV2, ResNet-50 and PoolNet) are used for recognition of co-saliency detectionThe CoSOD3k is a dataset with more realistic settings. The complexity, challenge, and scalability of CoSOD3k represent a considerable advance for associated visual tasks. The CoSal2015 is a great dataset to start if the work needs to be faced with challenging factors. The proposed method is tested using these different co-saliency benchmark datasets to evaluate the different parameters like Accuracy, Precision, Recall, F-Score, MAE etc., showing accuracy, loss and time with different models, datasets and optimizers. Amongst all the table used for performance evaluation (table 2, table 3); the modified ResNet model has shown better performance for Accuracy, Precision, Recall, MAE, Dice, Jacardcoff. Specificity etc. AndMobileNet model has shown consistently lowest performance in all the tables. The reason for MobileNet degraded performance is over-fitting problem which is mainly faced by deep learning neural network. Figure 3 and 4 are graphically representing models performance on various evaluation metrics with their corresponding dataset.

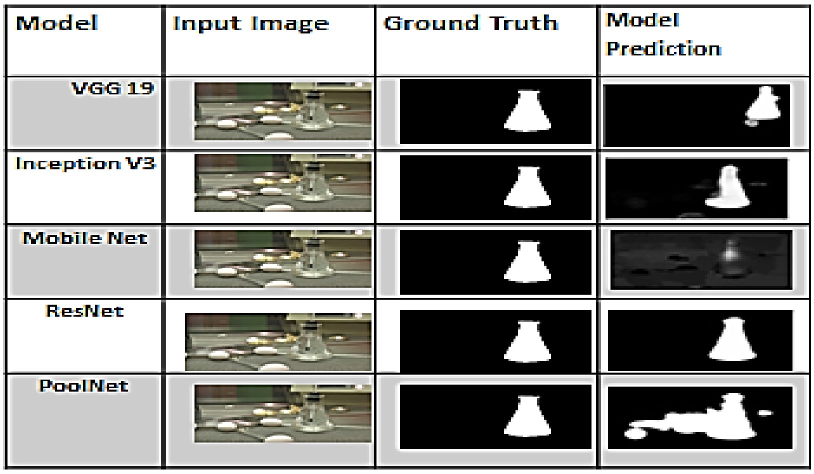

Figure 1. Co-Saliency maps for CoSOD3k datasets

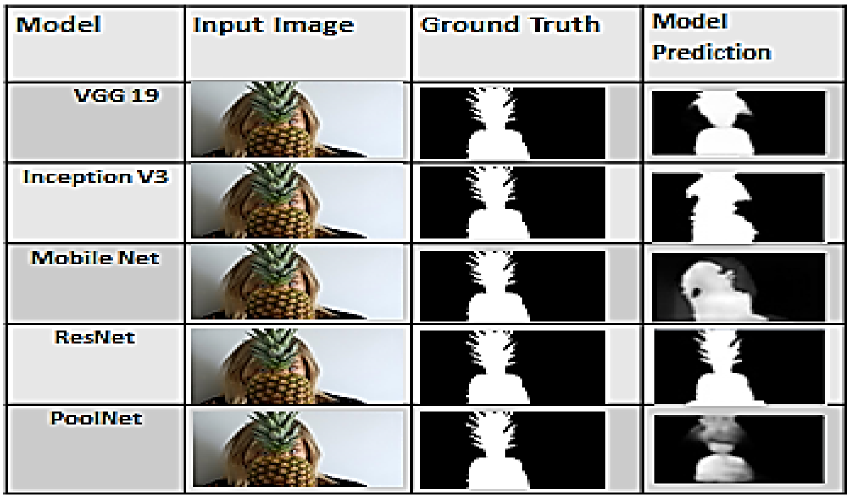

Figure 2. Co-Saliency maps for Cosal2015 datasets

Table 2. Performance parameter of Pre-Trained Models for CoSOD3k Dataset

Pre-trained Model	Optimizer	Sensitivity	F-1	Precision	MCC	Dice	Jaccard Coff.	Specificity	Recall	MAE
VGG-19	ADAM	91.89	90.12	91.23	92.23	91.12	92.23	90.56	90.3	0.125
	SGDM	92.78	91.36	91.15	92.54	92.45	91.96	91.89	92.9	0.147
	RMS	91.96	92.25	90.55	90.9	93.47	93.63	92.45	92.89	0.236
Inception -V3	ADAM	91.89	94.36	93.55	90.47	92.58	90.89	90.58	90.56	0.254
	SGDM	92.96	91.64	92.14	90.6	91.78	91.96	91.87	92.56	0.28
	RMS	91.89	90.64	91.56	90.45	90.47	93.23	90.69	91.58	0.125
MobileNetV2	ADAM	90.96	87.89	89.44	88.54	87.58	87.56	89.41	89.89	0.654
	SGDM	89.58	90.19	88.66	91.45	87.56	90.47	90.58	91.78	0.325
	RMS	89.48	90.99	88.44	87.63	87.66	88.56	90.78	91.85	0.2147
Resnet-50	ADAM	95.91	95.87	94.13	95.58	91.64	95.58	95.23	95.36	0.112
	SGDM	95.35	96.53	94.41	95.54	93.78	93.96	95.63	96.45	0.012
	RMS	94.78	95.44	94.65	95.45	92.58	95.89	96.87	95.96	0.145
PoolNet	ADAM	90.69	91.23	91.66	89.58	91.69	90.56	90.85	91.78	0.358
	SGDM	90.87	90.74	90.45	90.36	91.56	89.12	90.56	91.58	0.336
	RMS	90.87	90.14	90.6	90.89	90.3	93	94.97	92.32	0.458

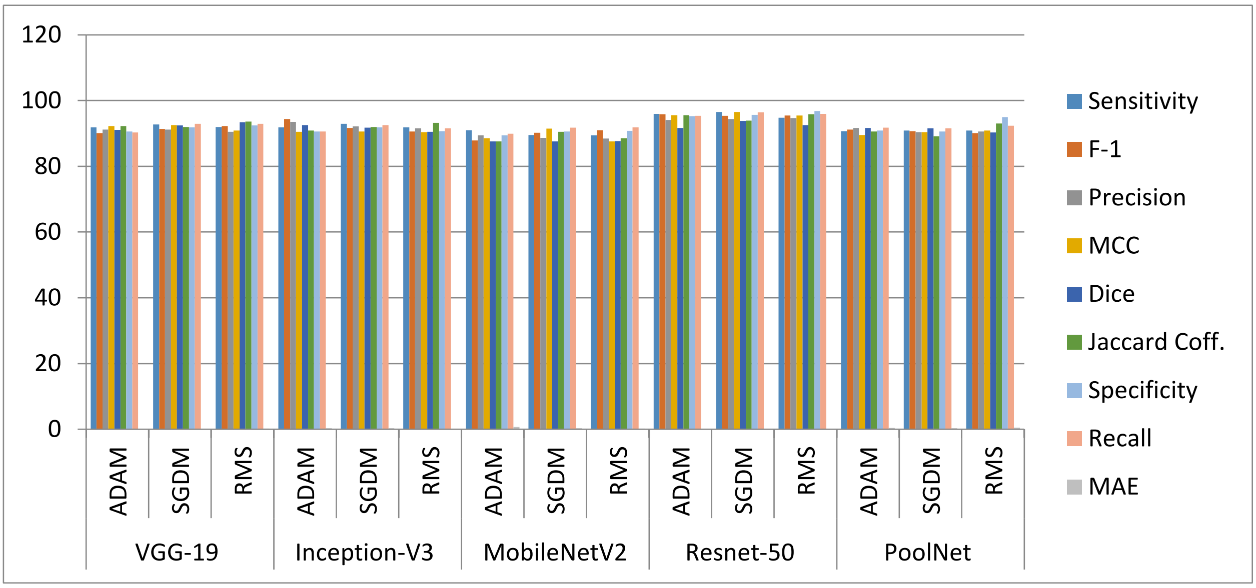

Figure 3. Performance Graph of Pre-Trained Models for Cosod3k Dataset

Table 3. Performance Parameter of Pre-Trained Models forCosal2015 Dataset

Model	Optimizer	Sensitivity	F-1	Precision	MCC	Dice	JaccardCoff.	Specificity	Recall	MAE
VGG-19	ADAM	92.96	92.32	90.25	90.23	92.21	91.12	92.56	91.15	0.125
	SGDM	90.23	91.56	92.89	91.45	90.56	91.56	90.01	90.23	0.145
	RMS	90.56	92.56	90.25	90.48	92.58	91.26	91.45	92.56	0.236
Inception-V3	ADMA	91.56	89.56	90.48	91.45	91.45	90.12	91.47	90.85	0.244
	SGDM	92.56	90.23	90.14	90.56	90.56	90.63	90.78	92.56	0.342
	RMS	91.56	91.25	93.69	90.74	92.44	92.89	91.56	92.56	0.125
MobileNet V2	ADAM	87.56	87.28	89.56	87.63	86.98	86.56	87.58	86.56	0.416
	SGDM	89.63	90.89	86.89	89.58	88.85	88.57	84.56	85.23	0.654
	RMS	89.63	88.12	87.54	86.78	89.96	87.15	86.45	89.56	0.447
Resnet-50	ADAM	94.23	93.21	93.63	93.63	93.89	93.15	94.25	92.47	0.126
	SGDM	**94.26**	**95.15**	**94.54**	**95.56**	**95.36**	**94.45**	**95.12**	**93.25**	**0.032**
	RMS	93.56	95.85	94.58	94.58	94.89	94.48	95.26	93.25	0.112
PoolNet	ADAM	91.96	90.12	89.45	90.58	91.78	90.25	90.23	90.52	0.235
	SGDM	90.63	91.45	90.47	90.57	90.15	89.89	90.25	91.56	0.212
	RMS	89.45	90.25	89.45	90.36	90.63	90.23	91.12	90.58	0.336

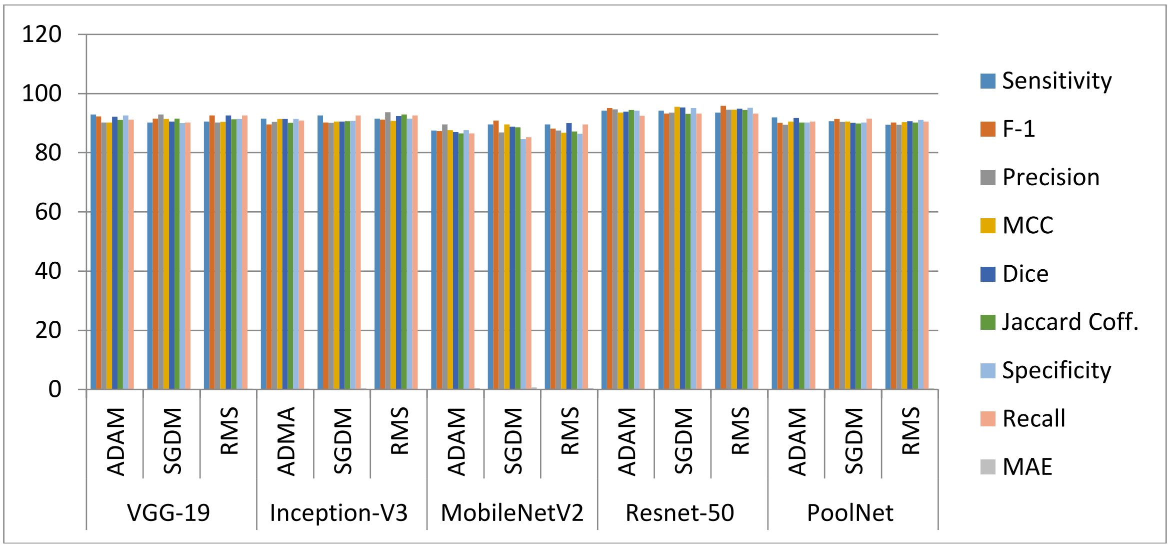

Figure 4. Performance of Pre-Trained Models for Cosal2015 Dataset

Conclusion

This paper evaluated the performance of pre-trained deep learning models such as VGG19, Inceptionv3, ResNet50, MobileNet and PoolNet used for recognition of the co-saliency detection for robust co-saliency detection. Our method captures the concept-level features of the co-salient items by examining deeply. By looking broadly, common backgrounds are suppressed in the image group by modeling the background with cross-group information. The results reveal that the modified ResNet model obtained satisfactory result with 96.53 percent which is highest score amongst all other state-of-the–art deep neural models. The proposed work is evaluated on three co-saliency benchmark Cosal2015, CoSOD3k dataset. The future work will be to detect co-salient objects in video sequences and to overcome the issues of existing neural network.

Conflict of interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Sun, M., Zhou, Z., Hu, Q., Wang, Z., & Jiang, J. (2018). SG-FCN: A motion and memory-based deep learning model for video saliency detection. *IEEE transactions on cybernetics*, 49(8), 2900-2911.

Zhang, L., Yang, C., Lu, H., Ruan, X., & Yang, M. H. (2016).Ranking saliency. *IEEE transactions on pattern analysis and machine intelligence,* 39(9), 1892-1904.

Yang, X., Li, S., Ma, J., Yang, J. Y., & Yan, J. (2022). Co-saliency-regularized correlation filters for object tracking. Signal Processing: *Image Communication*, 103, 116655.

Bai, C., Chen, J. N., Huang, L., Kpalma, K., & Chen, S. (2018). Saliency-based multi-feature modeling for semantic image retrieval. *Journal of Visual Communication and Image Representation*, 50, 199-204.

Zhang, D., Meng, D., & Han, J. (2016).Co-saliency detection via a self-paced multiple-instance learning framework. *IEEE transactions on pattern analysis and machine intelligence*, 39(5), 865-878.

Zhang, D., Fu, H., Han, J., Borji, A., & Li, X. (2018). A review of co-saliency detection algorithms: fundamentals, applications, and challenges. *ACM Transactions on Intelligent Systems and Technology* (TIST), 9(4), 1-31.

Wang, Q., Zhang, L., Li, Y., &Kpalma, K. (2020).Overview of deep-learning based methods for salient object detection in videos. *Pattern Recognition*, 104, 107340.

Jeong, D. J., Hwang, I., & Cho, N. I. (2018). Co-salient object detection based on deep saliency networks and seed propagation over an integrated graph. *IEEE Transactions on Image Processing*, 27(12), 5866-5879.

Fan, D. P., Lin, Z., Ji, G. P., Zhang, D., Fu, H., & Cheng, M. M. (2020).Taking a deeper look at co-salient object detection.In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 2919-2929).

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

Song, H., Liu, Z., Du, H., Sun, G., Le Meur, O., &Ren, T. (2017).Depth-aware salient object detection and segmentation via multiscale discriminative saliency fusion and bootstrap learning. *IEEE Transactions on Image Processing*, 26(9), 4204-4216.

Ullah, I., Jian, M., Hussain, S., Guo, J., Yu, H., Wang, X., & Yin, Y. (2020).A brief survey of visual saliency detection. *Multimedia Tools and Applications*, 79(45), 34605-34645.

Han, J., Cheng, G., Li, Z., & Zhang, D. (2017).A unified metric learning-based framework for co-saliency detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(10), 2473-2483.

He, K., Zhang, X., Ren, S., & Sun, J. (2016).Deep residual learning for image recognition.In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Han, J., Quan, R., Zhang, D., &Nie, F. (2017).Robust object co-segmentation using background prior. *IEEE Transactions on Image Processing*, 27(4), 1639-1651.

Ren, J., Liu, Z., Zhou, X., Sun, G., &Bai, C. (2018).Saliency integration driven by similar images. *Journal of Visual Communication and Image Representation*, 50, 227-236.

Li, L., Liu, Z., & Zhang, J. (2018). Unsupervised image co-segmentation via guidance of simple images. *Neurocomputing*, 275, 1650-1661.

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

Ye, L., Liu, Z., Li, L., Shen, L., Bai, C., & Wang, Y. (2017).Salient object segmentation via effective integration of saliency and objectness. *IEEE Transactions on Multimedia*, 19(8), 1742-1756.

Gupta, A. K., Seal, A., Prasad, M., Khanna, P. (2020). Salient object detection techniques in computer vision—a survey. *Entropy*, 22(10), 1174.

Zhang, K., Li, T., Shen, S., Liu, B., Chen, J., & Liu, Q. (2020). 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).

Su, Y., Deng, J., Sun, R., Lin, G., & Wu, Q. (2022). A Unified Transformer Framework for Group-based Segmentation: Co-Segmentation, Co-Saliency Detection and Video Salient Object Detection.

Bibliographic information of this paper for citing:

Mangal, Anuj; Garg, Hitendra & Bhatnagar, Charul (2023). Assessing the performance of Co-Saliency Detection method using various Deep Neural Networks. *Journal of Information Technology Management,* 15 (Special Issue), 23-34. https://doi.org/10.22059/jitm.2023.95243