ISBN: 978-1-948012-08-9

Contents List available at VOLKSON PRESS

Computer Science and Artificial Intelligence(CSAI)

DOI : http://doi.org/10.26480/iscsai.01.2017.33.35

COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE
VOLUME 1 ISSUE 3
INTERNATIONAL SYMPOSIUM ON COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE (ISCSAI 2017)

Saliency Detection Via Dense Convolution Network

Zheng Fang[a], Tieyong Cao[a*] and Yibo Xing[a]

[a]Institute of Command Information System, Army Engineering University, Muxuyuan Street, Nanjing, China
****Corresponding Author: 542050417@qq.com***

ARTICLE DETAILS

Article History:

Received 02 october 2017
Accepted 06 october 2017
Available online 11 october 2017

Keywords:

Saliency detection, Convolution Neural Network, Dense Convolution Network, computer vision, end-to-end

ABSTRACT

Saliency detection is a fundamental problem in the field of image processing and computer vision. The convolutional model has been also used in saliency detection for its outstanding performance on image classification and localization task. In this paper, we propose a novel way to detect the salient object by modifying the Dense Convolution Network (DenseNet). We replace the fully-connected layer and the final pooling layer in DenseNet into a convolution layer and a deconvolution layer to fit the saliency detection task. And our network ends up with a squared Euclidean loss layer for saliency regression. Our network is end-to-end architecture which outputs saliency maps directly. Experimental results demonstrate that our approach is competitive in comparison with the state-of-the-art approaches.

1. Introduction

Saliency detection aims to mimic the human visual system which naturally separates predominant objects of a scene from the rest of the image. Several applications benefit from saliency detection including image and video compression, context aware image re-targeting, scene parsing, image resizing, object detection and segmentation [1-6].

In most traditional methods, the salient objects were derived by the features extracted from pixels or regions, images were usually decomposed into several superpixel regions and final saliency maps consisted of these regions with their saliency scores [7-10]. The performance of these models rely on the segmentation methods and the selection of the feature. When facing images with multiple salient objects or low-contrast contents, these approaches can't produce satisfied results.

Recently, due to the outstanding performances on image classification and location tasks, Convolution Neural Network (CNN) is introduced in saliency detection and has substantially improved the performance of saliency detection. With CNNs, the saliency problem has been redefined as a labeling problem where feature selection between salient and non-salient objects is done automatically through gradient descent [11]. A CNN can't be directly used to train a saliency model, one way to exploit CNN in saliency detection is alleviating the problem by extracting a square patch around each pixel and use the patch to predict the center pixel's class [12,13]. Patches are often taken from different resolutions of the input image to capture global information. Another way is adding upsample layers into CNN. The modified CNN is called Fully Connected Network (FCN) which is first proposed in **Error! Reference source not found.** for semantic segmentation. Most saliency detection CNN model follow this way for it can capture more global and local information [15-17].

In this paper, we proposed a saliency model based on the Dense Convolution Network (DenseNet) by conducting a new FCN [18]. The DenseNet is a very efficient network which outperforms the current state-of-the-art results on most of the benchmark tasks but requires much less parameters. For saliency detection, we replace the fully-connected layer and the final pooling layer into a 1*1 kernel size convolution layer and a deconvolution layer. And a sigmoid layer is applied to get the saliency maps. In training process, the saliency network ends with a squared Euclidean loss layer for saliency regression. The architecture of our saliency network is shown in Figure 1 (the Dense block is defined in Figure 2).

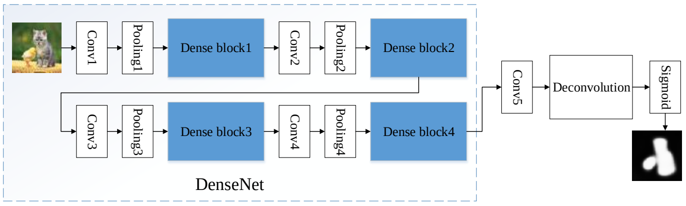

***Figure 1:** Our saliency model*

2. DENSE CONVOLUTION NETWORK

DenseNet is the latest CNN model proposed this year. It mainly consists of Dense blocks and transition layers. As shown in Figure 1, the transition layer is conducted by a convolution layer and a pooling layer and connects 2 Dense blocks. The architecture of the Dense block is shown in Figure 2.

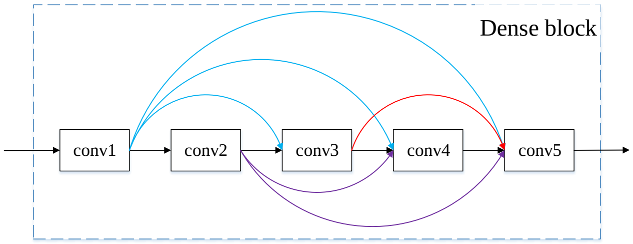

***Figure 2:** The structure of a Dense block which contains 5 conv layers*

Unlike most traditional CNN (Alex-Net, Google-Net and VGG-Net) which connect the layers in order from front to back, DenseNet connects all layers directly, each layer obtains additional inputs from all preceding layers and passes on its own feature-maps to all subsequent layers [19-21]. All features maps are concatenating. So the l^{th} layer has l inputs and a L-layer Dense block has L(L+1) connections. The l^{th} layer's output x_l is calculated by the following formula

$$x_l = H_l([x_0, x_1, x_2, ..., x_{l-1}]) \qquad (1)$$

Where $[x_0, x_1, x_2, ..., x_{l-1}]$

refers to the concatenation of the feature-maps produced in layers *0,1,...,l-*

Cite the article: Zheng Fang, Tieyong Cao and Yibo Xing (2017). Saliency Detection Via Dense Convolution Network ,
Topics in Intelligent Computing and Industry Design, 1(3) :33-35.

1, $H_l(\bullet)$ can be a can be a composite function of operations such as Batch Normalization (BN), rectified linear units (ReLU), Pooling or convolution(Conv). DenseNet has several advantages: it can alleviate the vanishing-gradient problem, strengthen feature propagation, encourage feature reuse, and substantially reduce the number of parameters [18].

3. SALIENCY MODEL BASED ON THE DENSENET

We choose the DenseNet-161 to construct our saliency model [18]. The original DenseNet ends up with a global average pooling layer and a softmax layer. The network predicts a class score for every input image. But saliency detection aims at predicting every pixel's saliency score of the input image. To obtain a saliency map for the input image, we replace the global average pooling layer and softmax layer into a 1*1 conv layer (conv5) and a deconvolution layer as shown in Figure 1. Then we add a sigmoid layer to get the saliency maps.

The deconvolution layer is used to resize the output images to match up with input images. The deconvolution is the inverse operator to the convolution and its output image's size is calculated by the following formula:

$$\begin{aligned} O_w &= (I_w - 1)\bullet stride - 2pad + \ker nelsize \\ O_h &= (I_h - 1)\bullet stride - 2pad + \ker nelsize \end{aligned} \qquad (2)$$

Where O_w and O_h indicate the width and height of the output image, I_w, I_h indicate the input's width and height. The *stride*, *pad* and *kernelsize* are the parameters of deconvolution layer. In our network, the input size is set as 500*500. Thus, using the model in Figure 1 to do a forward calculation, we will get a 16*16 saliency map before the deconvolution layer. Following equation (2), we set *stride=31, pad=14, kernelsize=63* to resize saliency maps from 16*16 to 500*500.

In training process, the saliency network ends with a squared Euclidean loss layer for saliency regression. The model is trained by minimizing the following cost functions:

$$J(Z) = \frac{1}{N}\sum_{i=1}^{N} \| M_i - f(Z_i) \|_F^2 \qquad (3)$$

Where *Z= {Z_i}(i=1,...,N)* denote a set of training images, *M_i(i=1,...,N)* denote the corresponding ground-truth binary maps.

4. EXPERIMENT RESULTS

We fine-tune the pretrained DenseNet-161 to train our saliency model. Our training set consists of 3900 images which are randomly selected from 6 saliency public dataset: ECSSD, SOD, HKU-IS, MSRA, and ICOSEG [22-26]. Our saliency network is implemented in Caffe toolbox [27]. The input images and ground-truth maps are resizing to 500*500 for training, the momentum parameter is chosen as 0.99, the learning rate is set to 10^{-10}, and the weight decay is 0.0005. The SGD learning procedure is accelerated using a NVIDIA GTX TITAN X GPU device, it takes about one day in 200,000 iterations.

To evaluate the performance of our saliency model, we choose 10 state-of-the-art saliency models: DRFI, IDRFI, BL, CGVS, DW, HDCT, MS, PR-GL, RRWR and SPMP to compare with our model [10, 28-36]. In experiments, we utilize 2 metrics for quantitative performance evaluations including Precision and Recall curve(PR) and F-measure. The PR-curve reflects the object retrieval performance in precision and recall by binarizing the final saliency map using different thresholds (usually ranging from 0 to 255). The F-measure characterizes the balance degree of object retrieval between precision and recall such that

$$F_\eta = \frac{(1+\eta^2)\,precision * recall}{\eta^2 * precision + recall} \qquad (3)$$

where η^2 is usually set to 0.3.

We test our model and the 10 compared models in 3 public datasets: ECSSD, HKU-IS, MSRA. Figure 3 is a bar graph of the F-measure results and the PR curves are plotted in Figure 4. For an intuitive illustration, Figure 5 shows some saliency results compared with other methods.

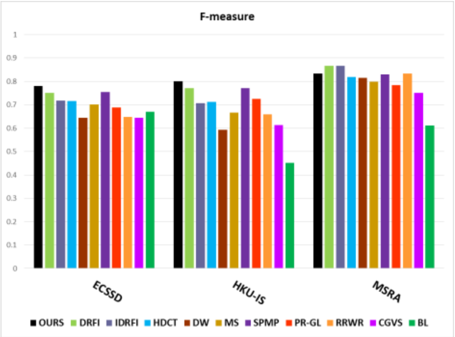

Figure 3: *F-measure results*

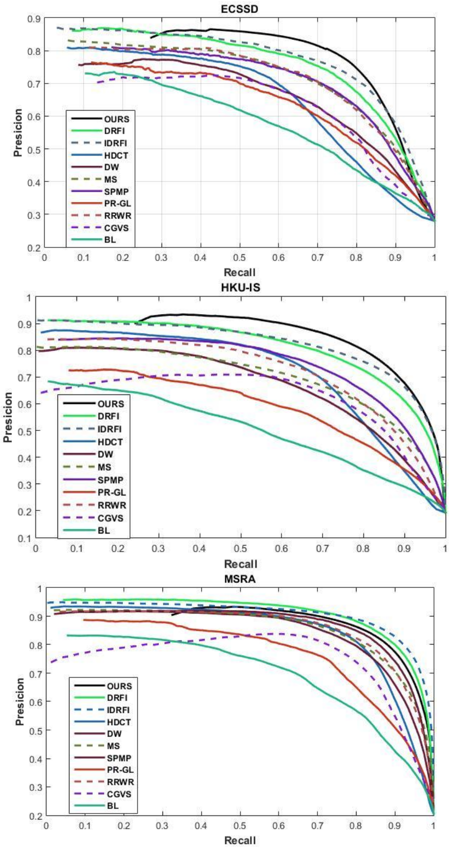

Figure 4: *PR-curve results*

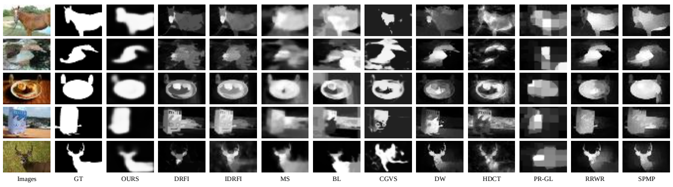

Figure 5: *Some saliency results*

It can be clearly seen in Figure 3 and Figure 4 that our model outperforms the other algorithms in 2 datasets: ECSSD, HKU-IS. And the result in MSRA is also competitive compared with other methods. Results in Figure 5 demonstrate that our method is more robust when facing complicated images or low contrast images.

5. CONCLUSION

In this paper, we propose a saliency model based on the DenseNet. We replace the fully-connected layer and the final pooling layer into a 1*1 kernel size convolution layer and a deconvolution layer to train a saliency model. Experimental results demonstrate that our approach is competitive in comparison with the state-of-the-art approaches.

REFERENCES

[1] Guo, C., Zhang, L. 1988. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. IEEE Transactions on Image Processing, 3 (5), 523.

[2] Li, G., Yu, Y. 2015. Visual saliency based on multiscale deep features.

Computer Vision and Pattern Recognition, 5455-5463. IEEE.

[3] Zhao, R., Ouyang, W., Li, H., Wang, X. 2015. Saliency detection by multi-context deep learning. Computer Vision and Pattern Recognition, 1265-1274. IEEE.

[4] Achanta, R., Süsstrunk, S. 2010. Saliency detection for content-aware image resizing. IEEE International Conference on Image Processing, 4542 (35), 1005-1008. IEEE.

[5] Zha, H., Feng, J., Zeng, G., Wang, J., Wang, P., Li, S. 2012. Salient object detection for searched web images via global saliency. Computer Vision and Pattern Recognition, 157 (2), 3194-3201. IEEE.

[6] Mehrani, P. 2010. Saliency segmentation based on learning and graph cut refinement, 1-12.

[7] Cheng, M.M., Zhang, G.X., Mitra, N.J., Huang, X., Hu, S.M. 2011. Global contrast based salient region detection. IEEE Conference on Computer Vision and Pattern Recognition, 37 (2), 409-416. IEEE Computer Society.

[8] Shi, J., Yan, Q., Li, X., Jia, J. 2014. Hierarchical image saliency detection on extended cssd. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38 (4), 717.

[9] Hwang, I., Lee, S. H., Park, J.S., Cho, N.I. 2017. Saliency detection based on seed propagation in a multilayer graph. Multimedia Tools and Applications, 76 (2), 2111-2129.

[10] Jiang, H., Wang, J., Yuan, Z., Wu, Y., Zheng, N., Li, S. 2013. Salient Object Detection: A Discriminative Regional Feature Integration Approach. Computer Vision and Pattern Recognition, 123 (2), 2083-2090. IEEE.

[11] Luo, Z., Mishra, A., Achkar, A., Eichel, J., Li, S., Jodoin, P. M. 2017. Non-Local Deep Features for Salient Object Detection. In IEEE CVPR.

[12] Liu, N., Han, J., Zhang, D., Wen, S., Liu, T. 2015. Predicting eye fixations using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 362-370.

[13] Zhao, R., Ouyang, W., Li, H., Wang, X. 2015. Saliency detection by multi-context deep learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1265-1274.

[14] Long, J., Shelhamer, E., Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3431-3440.

[15] Bruce, N.D., Catton, C., Janjic, S. 2016. A deeper look at saliency: feature contrast, semantics, and beyond. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 516-524.

[16] Pan, J., Sayrol, E., Giro-i-Nieto, X., McGuinness, K., O'Connor, N. E. 2016. Shallow and deep convolutional networks for saliency prediction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 598-606.

[17] Li, G., Yu, Y. 2016. Deep contrast learning for salient object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 478-487.

[18] Huang, G., Liu, Z., Weinberger, K. Q., van der Maaten, L. 2016. Densely connected convolutional networks. arXiv preprint arXiv:1608.06993.

[19] Krizhevsky, A., Sutskever, I., Hinton, G. E. 2012. ImageNet classification with deep convolutional neural networks. International Conference on Neural Information Processing Systems, 25 (2), 1097-1105. Curran Associates Inc.

[20] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D. 2015. Going deeper with convolutions. Computer Vision and Pattern Recognition, 1-9. IEEE.

[21] Simonyan, K., Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. Computer Science.

[22] Yan, Q., Xu, L., Shi, J., Jia, J. 2013. Hierarchical saliency detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1155-1162.

[23] Movahedi, Vida, Elder, J.H. 2010. Design and perceptual validation of performance measures for salient object segmentation. Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE Computer Society Conference on. IEEE.

[24] Guanbin, L., Yu, Y. 2015. Visual saliency based on multiscale deep features. Proceedings of the IEEE conference on computer vision and pattern recognition.

[25] Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.Y. 2011. Learning to detect a salient object. IEEE Transactions on Pattern analysis and machine intelligence, 33 (2), 353-367.

[26] Batra, D., Kowdle, A., Parikh, D., Luo, J., Chen, T. 2010. Icoseg: Interactive co-segmentation with intelligent scribble guidance. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. 3169-3176. IEEE.

[27] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Darrell, T. 2014. Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM international conference on Multimedia, 675-678. ACM.

[28] Zhou, X., Liu, Z., Sun, G., Ye, L., Wang, X. 2016. Improving saliency detection via multiple kernel boosting and adaptive fusion. IEEE Signal Processing Letters, 23 (4), 517–521.

[29] Tong, N., Lu, H., Ruan, X., Yang, M.H. 2015. Salient object detection via bootstrap learning. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1884–1892.

[30] Yang, K.F., Li, H., Li, C.Y., Li, Y.J. 2016. A unified framework for salient structure detection by contour-guided visual search. IEEE Transactions on Image Processing, 25 (8), 3475–3488.

[31] Li, H., Lu, H., Lin, Z., Shen, X., Price, B. 2015. Inner and inter label propagation: salient object detection in the wild. IEEE Transactions on Image Processing, 24 (10), 3176–3186.

[32] Kim, J., Han, D., Tai, Y.W., Kim, J. 2014. Salient region detection via high-dimensional color transform. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 883–890.

[33] Tong, N., Lu, H., Zhang, L., Ruan, X. 2014. Saliency detection with multiscale superpixels. IEEE Signal processing letters, 21 (9), 1035–1039.

[34] Tong, N., Lu, H., Zhang, Y., Ruan, X. 2015. Salient object detection via global and local cues. Pattern Recognition, 48 (10), 3258–3267.

[35] Li, C., Yuan, Y., Cai, W., Xia, Y., Feng, D.D. 2015. Robust saliency detection via regularized random walks ranking. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2710–2717.

[36] Hwang, I., Lee, S.H., Park, J.S., Cho, N. I. 2017. Saliency detection based on seed propagation in a multilayer graph. Multimedia Tools and Applications, 76 (2), 2111–2129.