International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Saliency Detection Based on Fourier Single Pixel Imaging and Deep Learning

Ji Zhaoyuan

College of Communication and Art Design, University of Shanghai for Science and Technology, China

Abstract: ***Single pixel imaging (SPI) is new type of computational imaging technology, which can obtain target object information with only one single pixel detector. It has the characteristics of low cost, wide imaging range and wide application range. The SPI using the Fourier base pattern as the illumination pattern has been proved to be a technology with both high imaging quality and high imaging efficiency. However, in complex scenes, it is still very time-consuming and resource-consuming task to obtain all the object information. Saliency detection has the property of quickly capturing regions containing important information visually from complex scenes. Therefore, we adopt the method of combining SPI with saliency detection, and propose this study of saliency detection based on Fourier single pixel imaging and deep learning models.***

Keywords: single pixel imaging, saliency detection, deep learning

1. Introduction

In recent years, Fourier has been proved to be a technology with both high imaging quality and high imaging efficiency0. In medium, high quality is often traded for multiple measurements. Inefficiency due to multiple measurements is major factor affecting single-pixel applications. The problem of SPI system is mainly caused by the interference of redundant information in the object image. Therefore, there is no shortage of reasonable use of detection resources to reduce the amount of measurement, which is an effective solution. Saliency detection is to use computers to simulate the human attention mechanism to quickly find the salient regions that people are interested in. As one of the research hotspots in the field of computer vision, saliency detection is widely used in image and video compression[2], object detection[3], image indexing[4], etc., and has important research significance and application value.

With the development of deep learning theory and the improvement of computing power of hardware devices, more and more network models have been proposed. The Amulet network [5]proposed by Zhang et al. utilizes convolutional features from multiple layers as saliency cues for salient object detection. With the popularity of Convolutional Neural Network (CNN), CNN-based models are gradually becoming the mainstream direction of salient object detection. Li [6]demonstrated the high efficiency of CNN applied to saliency detection.

In this paper, we adopt the method of combining SPI with saliency detection, and propose a saliency detection scheme based on Fourier single pixel imaging and deep learning. In this scheme, the advantage of compressed sensing in the data acquisition stage is adopted by Fourier single pixel imaging technology, the low-frequency information of the image is used for reconstruction, and the result of single pixel imaging is used for saliency detection using a fully convolutional network. From the simulation results, the reconstructed images at different or even extremely low sampling rates still have the ability to detect salient regions.

2. Method

2.1 Fourier single pixel imaging

When acquiring the real and imaginary parts of the Fourier coefficients, project two mutually inverse Fourier base patterns respectively, and then subtract the obtained single pixel measurement values, which can effectively suppress the effects of background lighting or circuits in the circuit. noise, so as to improve the quality of the reconstructed image. In this paper, the four-step phase shift method[7] is used to obtain the spatial frequency corresponding to the Fourier coefficient. The four-step phase-shifting method uses four Fourier base patterns with the same frequency (f_x, f_y) but with an initial phase level of $2k\pi/4$ (k=0, 1, 2, 3) respectively as the spatial light modulation pattern. The spatial light modulation pattern of the four-step phase shift method is

$$P_\phi(x, y; f_x, f_y) = a + b \cdot \cos(2\pi f_x x + 2\pi f_y y + \phi) \quad (1)$$

where a is the average intensity, b is the contrast, x, y are the Cartesian coordinates of the plane where the target object is located, f_x and f_y are the spatial frequencies, respectively, and ϕ is the initial phase. Then the light response value measured by the singlepixel detector is:

$$D_\varphi(f_x, f_y) = D_n + \beta \times \iint_s R(x, y) \times P(x, y; f_x, f_y, \varphi)\mathrm{d}x\mathrm{d}y \quad (2)$$

In the formula, D_n is the light response value caused by the background illumination at the location of the detector, and β is a factor related to the magnification of the singlepixel detector and the spatial relationship between the detector and the object. Where the four-step phase-shift Fourier spectrum of the image is recorded as $\tilde{I}_4$:

$$\tilde{I}_4 = (D_0 - D_\pi) + j \cdot (D_{\pi/2} - D_{3\pi/2}) \quad (3)$$

2.2 Residual Refinement Network

In this paper, we adopt R^{2}Net (Residual Refinement Network) [8]to solve the problem of saliency detection. The network introduces a new residual learning strategy for saliency detection. Residual learning is divided into a process, from coarse small scales to fine large scales, until the best prediction results consistent with GT (Ground Truth) are generated. It uses the DCPP (Dilated Convolutional Pyramid Pooling) module to generate coarse predictions based on contextual semantics. The generated rough predictions are then gradually refined with the help of the ARM (Attentional Residual Modules) module to make the saliency map closer to the real situation.

As shown in Figure 1, the network is mainly divided into three modules: R-VGG, DCPP and ARM. The R-VGG module is modified from the VGG-16[9]. This network optimizes the network structure compared to the previous model, and each convolutional layer adopts the same convolution kernel parameters. The DCPP module uses four dilated convolutional layers. Except for the different rate parameters, the four dilated convolutional layers are implemented using atrous convolution. The purpose of using atrous convolution is to enlarge the receptive field without losing spatial resolution. After obtaining a rough saliency map from DCPP, it is necessary to further improve the prediction results according to ARM. As shown in Figure 1, an ARM has three input messages. After obtaining these three input information, ARM calculates a rough saliency map and GT residual prediction at the same scale by adding the corresponding elements of the saliency map and residual prediction.

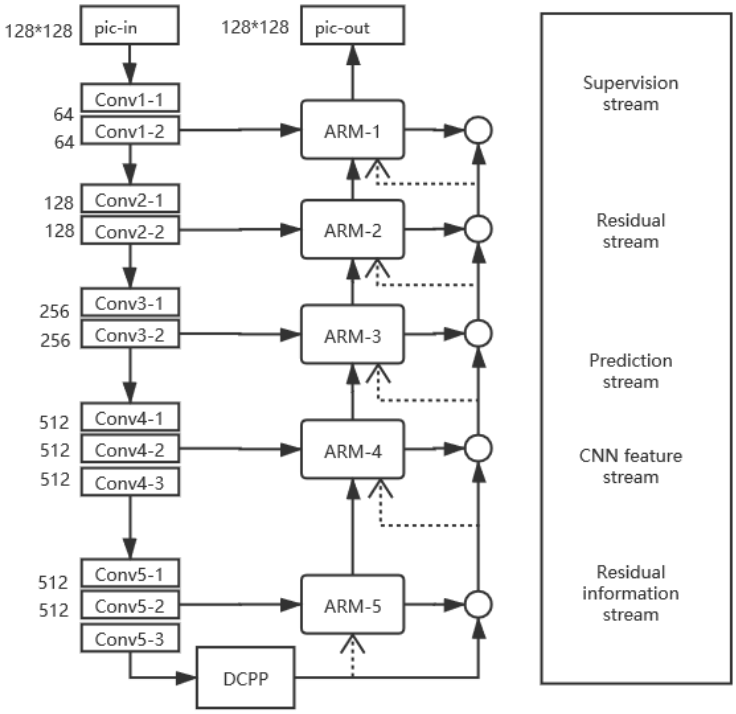

Figure 1: Residual Refinement Network

2.3 Loss Function

Loss function is one of the most basic and key elements in deep learning, so it is very important to choose an appropriate loss function to determine the quality of the model. Cross-entropy, which is commonly used to solve classification problems, is used in R^{2}Net as a loss function. To adapt the model to a single-pixel imaging system, we use the Hybrid_E_{loss}[10][10]function proposed by Cheng et al. It has the effect of enabling deep networks to learn pixel-level features, including weighted binary cross-entropy loss function ($\mathcal{L}_{\text{ce}}^{\text{w}}$), enhanced matching loss function ($\mathcal{L}_{\text{e}}$), and weighted intersection ratio loss function ($\mathcal{L}_{\text{iou}}^{\text{w}}$).This loss function show as:

$$\text{Hybrid\_}E_{\text{loss}} = \mathcal{L}_{\text{ce}}^{\text{w}} + \mathcal{L}_{\text{iou}}^{\text{w}} + \mathcal{L}_{\text{e}} \quad (4)$$

3. Experiments

We select images from DUTS[11] as the training set, and use three different datasets, ECSSD[12], HKU-IS[13], and PASCAL-S[14], as the test set. First, each of the 10553 images in the DUTS training set is resized to 128 × 128 pixels. Each image is then converted to a single-channel grayscale image. The processed image is then Fourier transformed, resulting in a new dataset. When generating images through Fourier transform, different sampling rates (100%, 80%, 50%, 25%, 16%, 1%) were used for processing, so a total of 73871 images were obtained as our training set. Similar to the case of extending our training set, the test set is also processed by FSPI（fourier single pixel imaging） at different sampling rates (100%, 80%, 50%, 25%, 16%, 1%) to obtain six batches of test sets.

We evaluate the test set by using four evaluation criteria, E-measure[15], S-measure[16], weightedF-measure[17] and MAE[18]. E-measure(E_ϕ),E_ϕ is proposed based on cognitive vision research to obtain image-level statistical information and its local pixel matching information. S-measure(S_α), in order to capture the importance of structural information in an image, S_α is used to evaluate the structural similarity between region-awareness and object-awareness. Weighted F-measure (F_β^w),F_β^w can provide more reliable evaluation results than the traditional F_β. MAE, for the pixel-wise mean absolute error between the predicted saliency map and the ground truth.

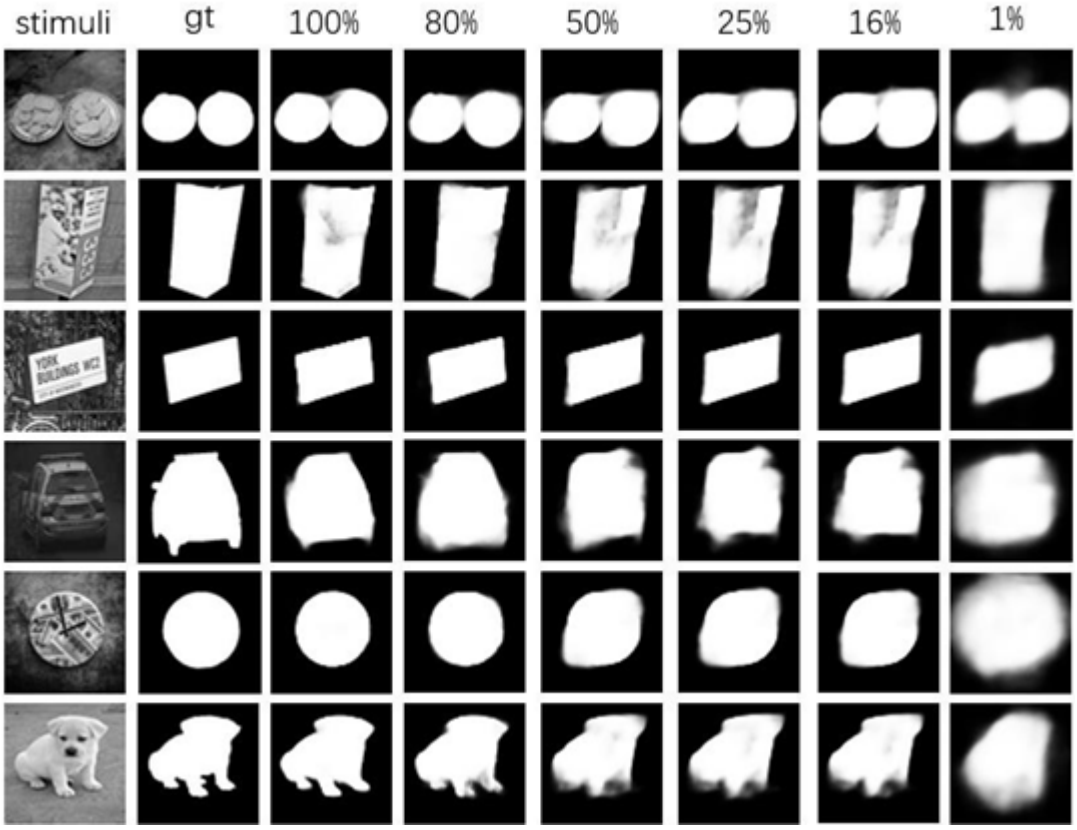

Figure 2: Visually detected results

As shown in Figure 2, the saliency object detection effect decreases as the sampling rate decreases. When the sampling rate is 100%, 80% and 50%, the edge contour changes as the sampling rate decreases. This shows that the validity of the saliency results is guaranteed while saving measurement resources at the loss of a certain accuracy. When the sampling rate is 25%, it can still be recognized normally. When the sampling rate is 16%, the image after reconstruction by FSPI（fourier single pixel imaging）has become very blurred, and the rough shape can still be obtained. Even at a sampling rate of 1%, the reconstructed image is almost unrecognizable, and an approximate location can be obtained. It can be seen from the results that when the sampling rate is too small, the distortion of the reconstructed image is large, and the accuracy of salient target detection is significantly reduced. But in some applications, the requirement for salient object detection accuracy is lower. For example in visual tracking, the contour information of objects that are constantly changing in complex scenes is not important. Therefore, there is no need to use more measurements to reconstruct the less distorted images used to locate salient objects in the reconstruction. According to the needs of the application in different situations, the resources consumed by unnecessary measurements are saved.

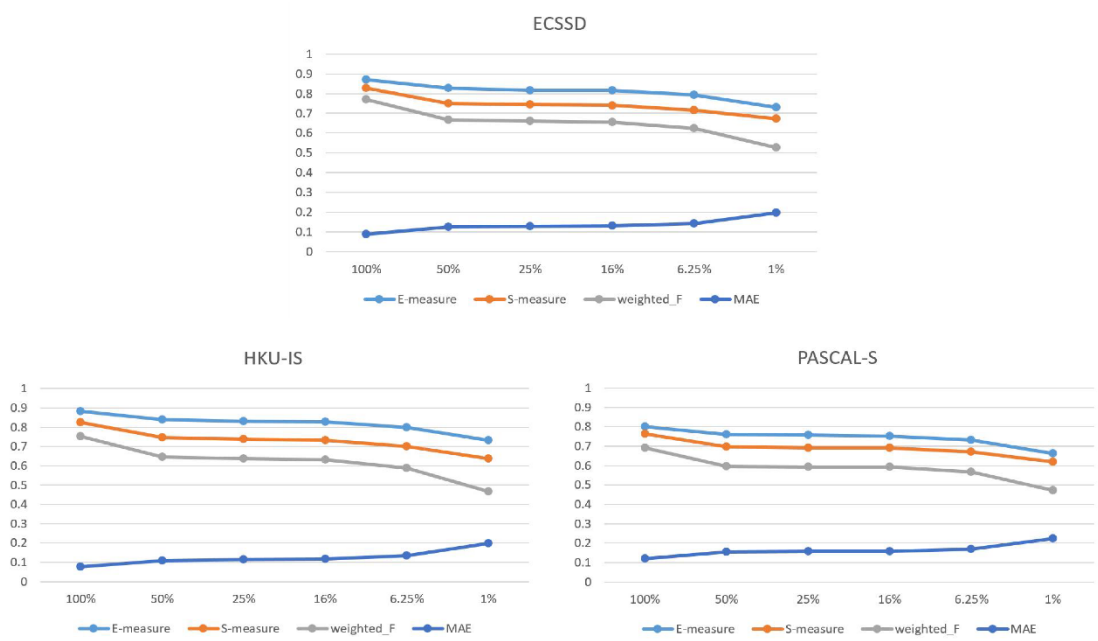

Figure 3: Results under the test datasets

4. Conclusion

In this paper, we propose a scheme for saliency detection based on Fourier single pixel imaging and deep learning models by combining single pixel imaging with saliency detection. Through simulation, it can be seen that saliency detection improves the efficiency of single-pixel imaging in complex scene tasks. And by using different sampling rates, the scheme reflects the adaptability to different application scenarios. In summary, our scheme has the ability to rapidly generate saliency objects in complex scenes. It can help vision tasks operate more efficiently in complex scenes.

References

[1] Zhang Zibang,Wang Xueying,Zheng Guoan,Zhong Jingang. Hadamard single-pixel imaging versus Fourier single-pixel imaging.[J]. Optics express,2017,25(16).
[2] Hadizadeh Hadi,Bajić Ivan V. Saliency-aware video compression.[J]. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society,2014,23(1).
[3] Wang Xuehao,Li Shuai,Chen Chenglizhao,Hao Aimin,Qin Hong. Depth Quality-aware Selective Saliency Fusion for RGB-D Image Salient Object Detection[J]. Neurocomputing,2020.
[4] Liang Zheng,Shengjin Wang,Ziqiong Liu,Qi Tian. Fast Image Retrieval: Query Pruning and Early Termination.[J]. IEEE Trans. Multimedia,2015,17(5).
[5] Zhang P , Wang D , Lu H , et al. Amulet: Aggregating Multi-level Convolutional Features for Salient Object Detection[C]// IEEE Computer Society. IEEE Computer Society, 2017.
[6] Li G, Yu Y. Visual Saliency Based on Multiscale Deep Features[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015.
[7] Bracewell R . The Fourier transform and its applications [C]// American Association of Physics Teachers. American Association of Physics Teachers, 2002.
[8] Feng Mengyang,Lu Huchuan,Yu Yizhou. Residual Learning for Salient Object Detection.[J]. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society,2020.
[9] Karen Simonyan,Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition.[J]. CoRR,2014,abs/1409.1556.
[10] FanDengPing , Ji GePeng , Qin Xuebin and Cheng MingMing. Cognitive Vision Inspired Object Segmentation Metric and Loss Function.[J]. SCIENTIA SINICA Informationis,2021,51(09):1475-1489.
[11] Wang L, Lu H, Wang Y, et al. Learning to Detect Salient Objects with Image-Level Supervision[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2017.
[12] Yan Q, Li X, Shi J, et al. Hierarchical Saliency Detection[C]// Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. IEEE, 2013.
[13] Li G , Yu Y . Visual Saliency Based on Multiscale Deep Features[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015.
[14] Li Y , Hou X , Koch C , et al. The Secrets of Salient Object Segmentation[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2014.
[15] Fan D P , Cheng G , Yang C , et al. Enhanced-alignment Measure for Binary Foreground Map Evaluation[C]// Twenty-Seventh International Joint Conference on Artificial Intelligence {IJCAI-18. 2018.
[16] Fan DP, Cheng MM, Liu Y, et al. Structure-Measure: A New Way to Evaluate Foreground Maps[C]// 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017.
[17] Ran M , Zelnikmanor L , Tal A . How to Evaluate Foreground Maps[C]// IEEE. IEEE, 2014.

[18] KrP , henbü. Saliency filters: Contrast based filtering for salient region detection[C]// Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2012.