Random Forest with Data Ensemble for Saliency Detection

Seungjun Nah and Kyoung Mu Lee
Department of Electrical and Computer Engineering, ASRI, Seoul National University, Seoul, Korea
E-mail: seungjun.nah@gmail.com, kyoungmu@snu.ac.kr

***Abstract*—Saliency detection is one of the most active research area in computer vision. Since L.Itti *et al.* [1] suggested computational model of visual attention, numerous detection algorithms have been proposed. However, most of modern saliency detection methods are based on superpixels which make detection results have abrupt edges inside the salient part. In this paper, we propose pixel-wise detection algorithm that makes more natural detection result. It makes our algorithm excel in describing detailed part of salient objects. Furthermore, we utilize the ensemble of not only random forest but also the data itself. Our algorithm achieves comparable performance with state of the art detection results.**

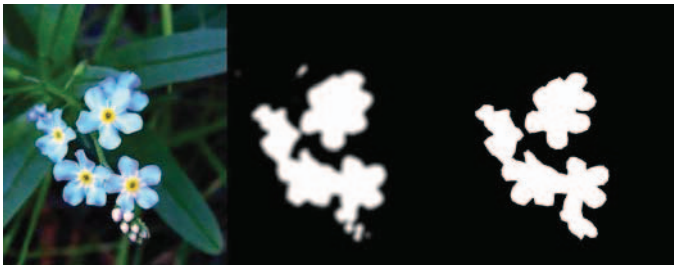
I. Introduction

Within every moment, rich visual information comes through human eyes. Visual data is instantly processed in brain so that visual attention is concentrated on the most interesting part of the vision [7]. Saliency detection model is for a lot of use. It can be used for video compression, advertisement design, adaptive super-resolution, video tracking. A lot of research has been done to mimic human attention as computational models based on Feature Integration Theory [3].

Most of the performance gain was led by integration of newly suggested features [2]. Judd *et al.* [4] introduced low, mid, high-level features and combined them to produce saliency map. Another flow was to introduce prior knowledge like center prior or boundary prior. In [5], [8], saliency models were constructed on the assumption that boundary area tends to be included in background and is not likely to be salient.

Recently, machine learning methods were popular in saliency detection. Shen and Zhao [9] used multi-layer sparse network to predict saliency map. In [10], multi-kernel learning was employed to learn webpage saliency. Random forest was used to regress saliency map by Jiang *et al* [6]. However, quite many of the machine learning techniques exploit superpixels, resulting in unnatural saliency map.

In this paper, we propose pixel-wise regression method with relaxed random forest classification. We label each pixel in image and make vote for saliency value with neighboring pixels. Since we are dealing with pixels, we excel in describing detailed parts compared to superpixel based approaches. It is shown on figure 2. After random forest test result, we apply test result ensemble to refine our result. By making vote with similar data, we can reject outliers and make predicition more accurate. Since we constrict the data to be ensembled with the distance limit, we can prevent saliency map structure from being messed up.



(a) (b) (c)

Fig. 1: Saliency detection example:
(a) input image, (b) our detection result, (c) groundtruth

In short, the contribution of this paper follows: the ensemble of relevant test data to predict saliency in pixel-level, and making the result robust against outliers still maintaining structure.

II. Related Works

Many works[11], [12] have suggested effective features for saliency detection. Judd *et al.*[4] suggested low, mid, high-level features, for saliency detection. However, the selection of high-level and mid-level features like horizontal line and human faces were without enough justification, and the true impact of suggested features were not much discovered. Horizontal line and human face may not exist in a lot of images, even though there may be significant salient object. Even worse, horizontal striped shirt is not guaranteed to be salient in all cases.

On the other line of research, there were several efforts to involve machine learning techniques into saliency prediction result. [4] used linear SVM to train their model. However, linear SVM is weak against outliers, suffering from one outlier degrading whole prediction result. More recently, Shen *et al.* [9] suggested neural network architecture to learn saliency. Jiang *et al.*[5] proposed random forest approach based on superpixel segmentation. In Jiang's work, the image is oversegmented, and then merged to generate several levels of segmentation. And then, for each level, segment's feature becomes the input to random forest. Testing images are tested in the same way.

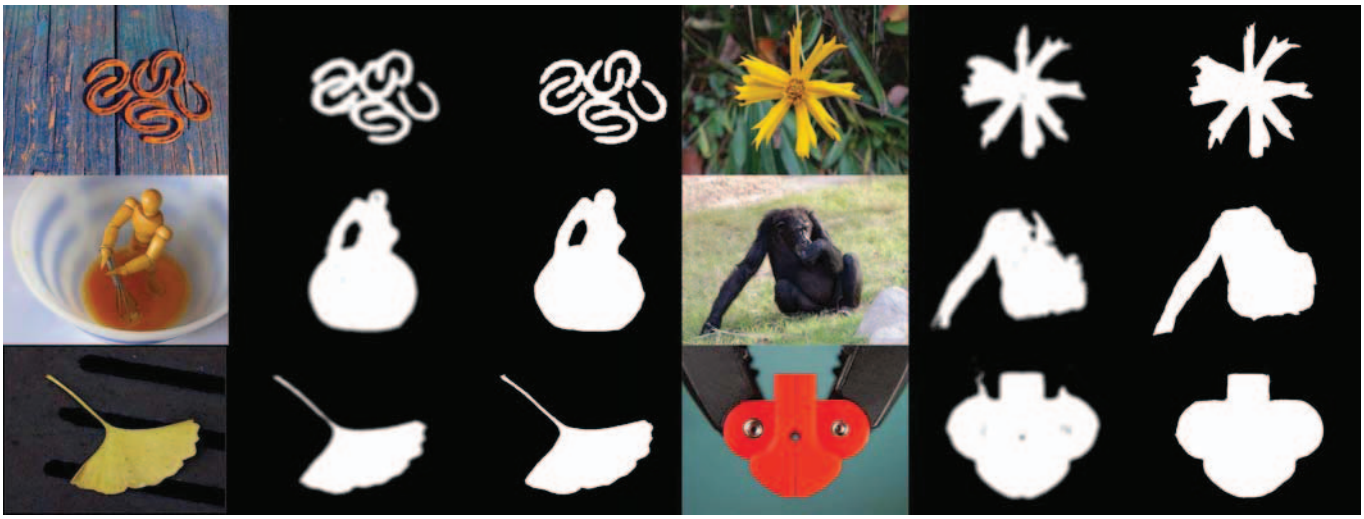


Fig. 2: Our Saliency detection result of various images. Our model detects complex salient region. (a) test image, (b) detected map, (c) groundtruth, (d) test image, (e) detected map, (f) groundtruth

III. Proposed Method

Given an image patch, our method utilizes classification random forest to classify whether the patch is salient or not. By using patches, there are two advantages over using superpixels. The flow of algorithm is described below. The trick of using patches for pixel-wise result is described in test part. We used MSRA-B dataset, which has pixel-wise annotated 5000 images, for both training and test. Half of the images were chosen to train, the others were used for testing.

A. Training

Our random forest is used for classifying a patch is salient or not. Given a training set, fixed-size patches are randomly sampled from the images. To avoid overfitting, only N_p patches are sampled per image. To label patches, if more than half of the pixels in a patch is salient, then the patch is labelled as salient. If not, it is labelled as not salient. We empirically set $N_p = 50$. From the sampled patches we extracted features similar to [6]. Modified regional property, regional contrast, boundary contrast features for image patches were used. The other details are described in Table I. Then, the classification forest is trained.

B. Testing

For testing, every possible patches from each test image is tested with trained random forest. The patch label is transferred to the pixels included. If a patch is salient, the patch pixels are labeled as 1. Since we evaluate every patches, there are a lot of overlaps between neighboring patches. After processing all the patches, pixel label is finalized by averaging the relaxed labels computed. There are three benefits of taking mean of labels.

First, even with the square patches, averaging pixel values can represent the structure of salient objects. Second, within the size of patch, the saliency value changes smoothly. Therefore, even if there are outliers with overlapping inlier patches, the saliency value will hurt little. Third, the test data ensemble is made within similar input data. It is natural to hypothesize that similar input to random forest should result in similar output. A patch and its one-pixel slided patch shares most of the part, only differing in two columns or rows. Therefore, it is highly likely for them to have same class label. By taking mean value, the test accuracy can be more stable. This can be considered as the way of injecting another ensemble in random forest.

C. Used features

Similarly to [6], three kinds of features are used. Modified regional property feature, regional contrast feature and boundary contrast features are computed. Originally proposed features were for superpixel, so some unrelated features for fixed-size square patches were removed. Instead, some additional features were added. Regional property feature contains statistical property of the patch itself. It describes the distribution of pixels inside the patch. Regional contrast feature has the information of the difference of local patch and its neighborhood. It describes how much the patch is distinctive relative to it neighborhood patches, considering that salient region should have close relevance with distinctiveness. The boundary contrast feature has same elements as regional contrast feature, computing contrast of patch and each 4 boundaries. Using boundary contrast feature is semantically more natural than introducing center prior or boundary prior. We consider boundary contrast for each 4 edges since background is not guaranteed to be uniform. Since random forest can learn from the difference in feature space instead of uniform spatial difference by image, relation between patch and boundary is effectively valued. It is described in table I.

Category	Features
Regional Property	patch center location, variance of RGB, Lab, HSV
Regional Contrast	difference of mean of RGB, Lab, histogram χ^2 distance of Lab, H, S
Boundary Contrast	difference of mean of RGB, Lab, histogram χ^2 distance of Lab, H, S

TABLE I: Used features. Total 30 dimension.

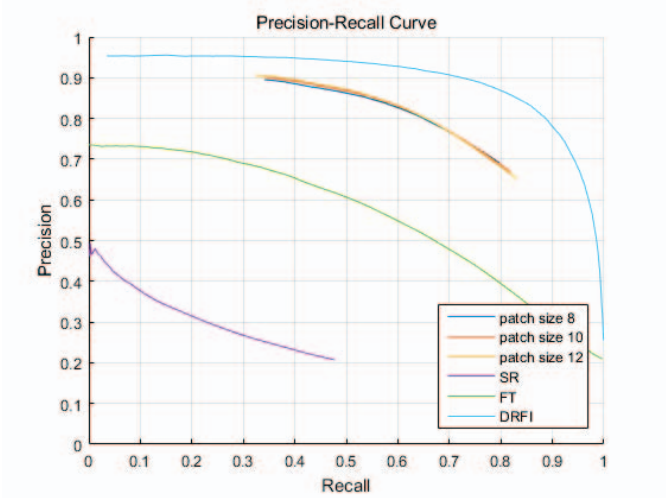


Fig. 3: Precision-Recall curve on MSRA datset

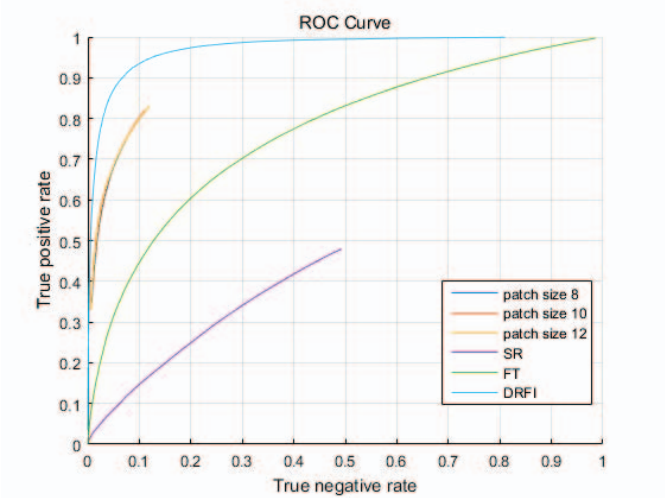


Fig. 4: ROC curve on MSRA dataset

IV. Experimental Result

We compared our method with several other saliency detection algorithms - SR [16], FT [4], DRFI [6]. The performance evaluation is done with FT and MSRA-B [15] dataset. FT and MSRA-B dataset are each composed of 1000, 5000 images. Every groundtruth images are pixel-wise binary labeled. For each experiment test, half of the dataset were used to train our random forest model, and the other half were used as test set. For training set images, 50 patches per image of width 8, 10, 12 were sampled, which is less than 0.1%. After training the forest, it is applied to every possible patches in the test set images.

Figure 3 and figure 4 illustrate averaged precision-recall plots and ROC plots for each methods on MSRA dataset. Our method did not reach the state of the art in terms of precision and recall, but still it has some advantages over other methods. The lower bound for recall is about 0.35, higher than most of the methods grounding to 0. This means that our model finds salient area with higher confidence than other methods. In other words, our method is robust against outliers as expected by ensemble of patches. This high confidence tendency is well shown on figure 5. Also, the property helps our model describe detailed boundary of salient objects compared to other methods. In terms of patch size, there were neglegible difference in performance.

To measure accuracy of proposed method, AUC score and F-measure is reported. AUC score is popular for saliency detection, which is the area under the ROC curve. Proposed method scored 0.8971. Also, F-measure is widely used. It is basically harmonic mean of precision and recall. To give more weight for precision, F_α measure, $(1+\alpha)/(\alpha \times precision + recall)$ is used. Here, α is set to 0.3.

dataset	AUC	F_α measure
FT	0.8971	0.8738
MSRA-B	0.8634	0.8283

TABLE II: AUC score and F_α measure, patch size 10x10

V. Conclusions

Our saliency detection algorithm utilizes the effect of ensemble of test data itself as well as the ensemble in random forest. We achieved high-confidence saliency detection algorithm which is robust to outliers. With much less features and training data compared to [6], we reached qualitatively better result.

However, several future work remains. First, scale invariance issue. Although proposed method works quite well with both small and large salient object, the effect of patch size is not fully explored. Also, the ensemble of different sized patches could be considered for more ensemble. The use of decision jungles [13] is another option. Decision jungle was reported to reach higher prediction accuracy in classification task. Since saliency detection problem is binary labeling problem, decision jungle could improve performance.

References

[1] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI),* vol. 20, pp.1254-1259, 1998

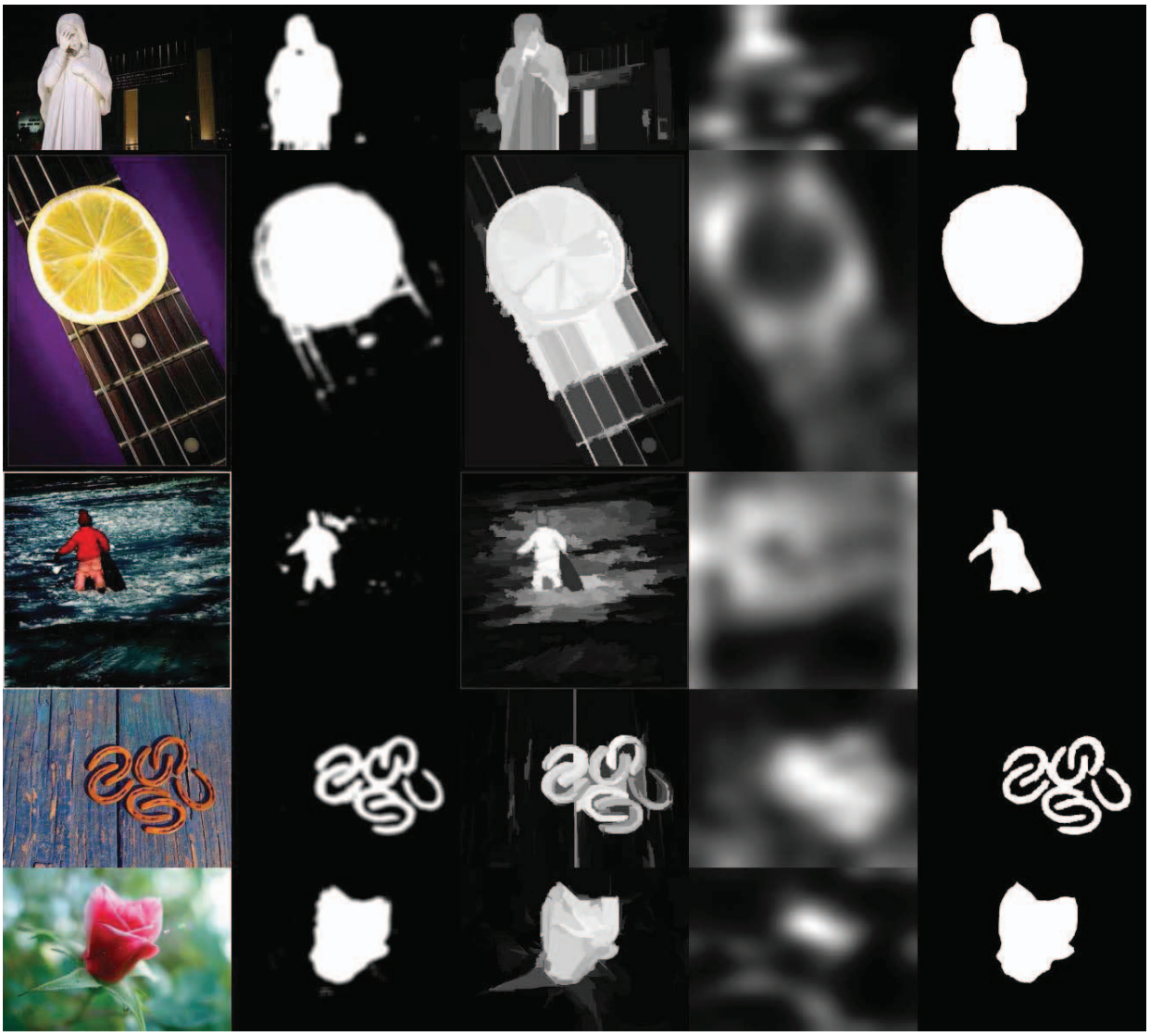


Fig. 5: Qualitative comparison with other methods. (a) Input Image, (b) Our result, (c) DRFI, (d) SR, (e) Groundtruth

[2] A. Borji, D. Sihite, and L. Itti, "Salient object detection: A benchmark," In *Proc. European Conference on Computer Vision (ECCV),* 2012

[3] Treisman, Anne M, Gelade, Garry, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, Elsevier, 1980, pp. 97-136

[4] T. Judd, K. Ehinger, and F. Durand, and A. Torralba, "Learning to Predict Where Humans Look," In *Proc. International Conference on Computer Vision (ICCV),* 2009

[5] B. Jiang, L. Zhang, H. Lu, C. Yang, and MH Yang, "Saliency detection via absorbing markov chain," In *Proc. International Conference on Computer Vision (ICCV),* 2013

[6] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," In *Proc. Computer Vision and Pattern Recognition (CVPR),* 2013

[7] K. Koch, J. McLean, R. Segev, M. A. Freed, M. Berry II, Michael J, V. Balasubramanian, P. Sterling, "How Much the Eye Tells the Brain," *Current Biology,* vol. 16, Elsevier, pp. 1428 - 1434, 2006

[8] X. Li, H. Lu, L. Zhang, X. Ruan, and MH Yang, "Saliency detection via dense and sparse reconstruction," In *Proc. International Conference on Computer Vision (ICCV),* 2013

[9] C. Shen and Q. Zhao, "Learning to predict eye fixations for semantic contents using multi-layer sparse network, " *Neurocomputing*, Elsevier, vol. 138, pp. 61 - 68, 2014

[10] C. Shen and Q. Zhao, "Webpage Saliency," In *Proc. European Conference on Computer Vision (ECCV),* 2014

[11] R. Margolin, A. Tal and Zelnik-Manor, "What Makes a Patch Distinct?," In *Proc. Computer Vision and Pattern Recognition (CVPR),* 2013

[12] X. Sun, H. Yao, R. Ji, and X-M. Liu, "Towards Statistical Modeling of Saccadic Eye Movements and Visual Saliency," *IEEE Transactions on Image Processing,* vol. 23, IEEE, pp.4649 - 4662, 2014

[13] J. Shotton, T. Sharp, P. Kholi, S. Nowozin, J. Winn and A. Criminisi, "Decision Jungles: Compact and Rich Models for Classification," In *Advances in Neural Information Processing Systems,* 2013

[14] R. Achanta, S. Hemami, F. Estrada and S. Susstrunk, "Frequency-tuned salient region detection," In *Proc. Computer Vision and Pattern Recognition (CVPR),* 2009

[15] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang and HY Shum, "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI),* pp. 353 - 367, 2011

[16] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," In *Proc. Computer Vision and Pattern Recognition (CVPR),* 2007