SALIENT OBJECT DETECTION ON HYPERSPECTRAL IMAGES USING FEATURES LEARNED FROM UNSUPERVISED SEGMENTATION TASK

N. İmamoğlu[1,], G. Ding[1,2], Y. Fang[2], A. Kanezaki[1], T. Kouyama[1], R. Nakamura[1]*

[1]National Institute of Advanced Industrial Science and Technology (AIST),
Artificial Intelligence Research Center, Tokyo JAPAN
[2]Jiangxi University of Finance and Economics,
School of Information Technology, Nanchang CHINA

ABSTRACT

Various saliency detection algorithms from color images have been proposed to mimic eye fixation or attentive object detection response of human observers for the same scenes. However, developments on hyperspectral imaging systems enable us to obtain redundant spectral information of the observed scenes from the reflected light source from objects. A few studies using low-level features on hyperspectral images demonstrated that salient object detection can be achieved. In this work, we proposed a salient object detection model on hyperspectral images by applying manifold ranking (MR) on self-supervised Convolutional Neural Network (CNN) features (high-level features) from unsupervised image segmentation task. Self-supervision of CNN continues until clustering loss or saliency maps converges to a defined error between each iteration. Finally, saliency estimations is done as the saliency map at last iteration when the self-supervision procedure terminates with convergence. Experimental evaluations demonstrated that proposed saliency detection algorithm on hyperspectral images is outperforming state-of-the-arts hyperspectral saliency models including the original MR based saliency model.

***Index Terms*—** Hyperspectral image, Unsupervised learning, Convolutional Neural Networks, Manifold ranking, Salient object detection

1. INTRODUCTION

Reflected light from objects and radiance from a light source has information at various wavelengths both in visible and non-visible spectrum to human eye [1, 2, 3, 4, 5]. With the recent advancements on hyperspectral imaging systems, unlike the conventional cameras providing images with 1(monochrome images) or 3 channel (e.g. RGB or YCbCr

images), hyperspectral imaging systems enable researchers the opportunity to capture data from the observed scenes with high spatial and redundant spectral resolution (both visible and non-visible spectrum to human eye) of the observed scenes from the radiance or reflected light source from objects [2, 3, 4, 5]. These data have been used in many applications such as remote sensing, scene analysis or object detection [2, 6, 7, 8, 9, 10, 11], spectral estimation[11, 12, 13, 14], etc..

Visual attention modeling (saliency detection) [15, 16, 17, 18, 19, 20, 21] is a promising research field for practical applications, which may benefit many other applications on hyperspectral data processing stated prior. For instance, a few studies [7, 8, 9, 10] demonstrated that salient object detection can be achieved by using low-level features on hyperspectral images. In contrast to these models relying on low-level features or hand-crafted features to obtain saliency maps, higher-level features can be extracted and used in a self-supervised manner for hyper-spectral data, where each spectral bands' contribution to the representation can be learned with unsupervised neural network used for segmentation task [22]. In addition, works on hyperspectral saliency on natural scenes were mostly tested on dataset with a few hyperspectral images ([7] used 13 images and [8] used 17 images) collected and selected from various hyperspectral data. Moreover, these hyperspectral data was not collected and created for the purpose of salient object detection. And, quantitative evaluations of the models were mostly limited to Precision-Recall and F-measure metrics. Therefore, we believe that a dataset created specifically for salient object detection should be used for evaluating the models with various metrics.

Proposed work and contributions: In this work, we propose a salient object detection model (see Figure 1) on hyperspectral images by applying manifold ranking [17] to self-supervised Convolutional Neural Network (CNN) features (high-level features) learned by an unsupervised image segmentation task [22]. Self-supervision of CNN continues until clustering loss or saliency map computed from CNN features converges to a defined error difference between each iteration. Then, saliency map at the last iteration is used as the
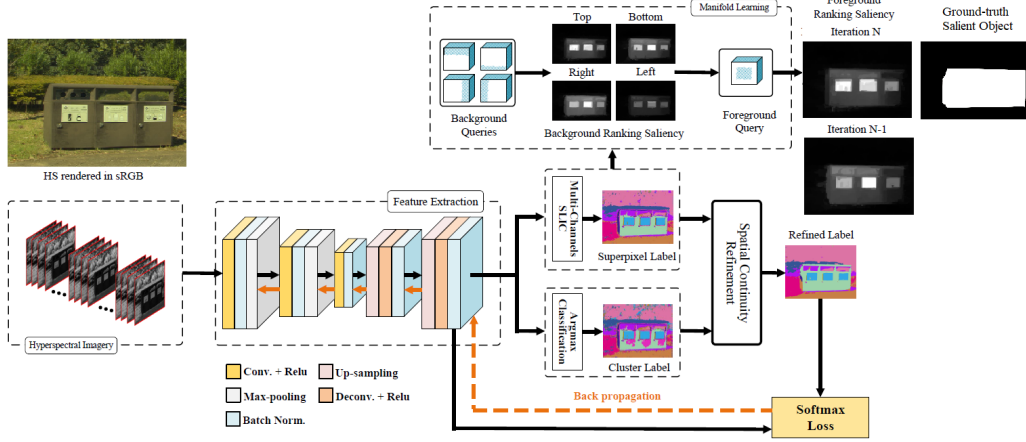
Fig. 1. Proposed hyperspectral salient object detection model with unsupervised deep features .

Table 1. The detailed configuration of the CNN model. Note that *BN* represents batch normalization operation.

layer	# filters	kernel	stride
Conv. + Relu + BN	64	3×3	1×1
Maxpooling	-	2×2	2×2
Conv. + Relu + BN	64	3×3	1×1
Maxpooling	-	2×2	2×2
Conv. + Relu + BN	64	3×3	1×1
Upsampling	-	2×2	2×2
Deconv. + Relu + BN	64	3×3	1×1
Upsampling	-	2×2	2×2
Deconv. + Relu + BN	64	1×1	1×1

result of proposed model when the self-supervision procedure terminates.

To the best of our knowledge, there are not any works on hyperspectral salient object detection for natural scenes as a self-supervised approach, which combines unsupervised segmentation using CNN and salient object detection task on the scene. Regarding the approach, contributions or differences of the proposed model can be explained as: First, unsupervised segmentation task used in previous paper [22] takes advantage of cluster refinement process based on the superpixels obtained by the input color image. However, in this work, we apply the refinement process based on the superpixels obtained by the high-level features of the CNN (see Fig.1) that takes hyperspectral image as input. Interestingly, this process resulted in better saliency detection performance and it seemed faster convergence regarding the segmentation task. Second, in contrast to the saliency model with manifold ranking (MR) in [17] using low-level features, we utilized self-supervised deep-features with higher order semantics, which seems to improve the saliency detection performance drastically compared to study in [17]. Then, unlike the CNN model

used in [22], we included max-pooling for down-sampling and we replaced last two CNN layers with deconvolution layers as in Table 1. Finally, self-supervision of CNN model does not need to finalize until a defined maximum iteration because we check clustering loss and saliency map for termination; in addition, saliency results of proposed model seems to converge faster than the segmentation task in most cases while using self-supervised deep features on manifold ranking based saliency detection. Experiments demonstrated that proposed saliency detection algorithm on hyperspectral images is outperforming state-of-the-arts hyperspectral saliency models including the original MR [17] saliency model.

2. SELF-SUPERVISED SALIENT OBJECT DETECTION ON HYPERSPECTRAL IMAGES

To achieve salient object detection goal in Fig. 1, we propose to use an unsupervised backpropagation semantic segmentation method [22] to learn high-level visual features that will be used in the manifold raking algorithm [17] for saliency computation. Given k channels hyperspectral imagery $I = \{H^k\}_{n=1}^k$ as input to our model, first, all the pixel values are normalized to [0,1]. Then, we adopt a CNN model to extract p-dimension feature maps $\{x_n\}$ from the Batch-Normalization (BN) output of the last Deconvolution layer of model getting hyperspectral imagery I as input. The detailed configuration of the CNN model is shown in Table 1. Note that the spatial resolution of output feature map and input hyperspectral image I are identical. After normalizing the learned response maps via batch normalization as in [22], we obtain cluster label $\{c_n\}$ by using argmax classification to the feature maps to classify each pixel by choosing the dimension that has the maximum value as $\{y_n\}$ [22]. Then, we apply refinement process on $\{c_n\}$ based on the superpixels obtained by the high-level features of the CNN (see Fig. 1) in contrary to [22] using superpixels based on the input data (e.g. hyper-

Table 2. The performance of different saliency detection methods are given as values with Red, Green, and Blue colors indicate the best three results in respective order. Note that larger AUC_{Borji}, CC, F_β, $maxF_\beta$, $aveF_\beta$, Precision, Recall, NSS values, and smaller KLdiv values means better performance.

Model	AUC_{Borji}	CC	F_β	$maxF_\beta$	$aveF_\beta$	Precision	Recall	NSS	KLdiv
Itti *et al.* [15, 7]	0.7774	0.3536	0.3530	0.3754	0.1674	0.1909	0.2329	1.3636	2.3186
SED [7]	0.7691	0.2797	0.3082	0.3420	0.1541	0.3479	0.1207	1.3498	2.3426
SAD [7]	0.7707	0.3034	0.2635	0.2662	0.1397	0.1547	0.2314	1.1767	2.2719
GS [7]	0.7781	0.3403	0.3004	0.3694	0.1861	0.2753	0.2275	1.5637	2.1944
SED-OCM-GS [7]	0.8021	0.3730	0.3260	0.3634	0.1708	0.2757	0.2209	1.5908	2.1707
SED-OCM-SAD [7]	0.8108	0.3882	0.2635	0.2662	0.1397	0.1547	0.2314	1.5301	2.1601
SGC [8]*	0.8252	0.5012	0.1891	0.2214	0.1815	0.2344	0.2822	1.4739	2.2154
HS-MR [17]**	0.7369	0.3492	0.3636	0.4308	0.3638	0.4397	0.3587	1.2702	3.7637
SUDF$_{HS-Slic}$	0.8509	0.5563	0.4580	0.5355	0.4430	0.5346	0.4449	2.1938	1.7853
SUDF$_{HF-Slic}$	0.8602	0.5829	0.4671	0.5654	0.4668	0.5436	0.4834	2.1200	1.7241

*SGC [10] The codes were not available for [10] so the implementation was done by the authors in Matlab based on the paper [10]

**HS-MR [17] saliency detection is originally for color images; however, published codes by the authors can be used for hyperspectral data for MR and superpixel methods in the code.

spectral image). Refinement process is achieved by assigning all pixels same cluster label based on the highest frequency of label in the superpixel area [22].

Iteration 10   Iteration 20   Iteration 30   Iteration 40   Iteration 50
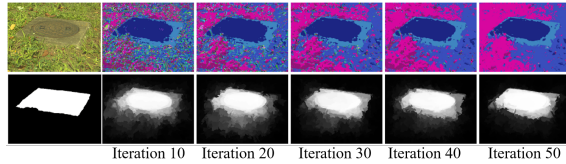
Fig. 2. (top) sample image and segmentation results at different iterations during self-supervision, (bottom) ground-truth image and saliency maps from different iterations.

Similar with the supervised learning, we use the softmax cross entropy loss between the network responses $\{y_n\}$ and the refined cluster labels $\{c_n\}$ at iteration n [22]. Using this error with back-propagation, the parameters of convolutional and deconvolution filters are updated by utilizing gradient-descent with momentum [22]. As in [22], Glorot and Bengio method [28] is employed for the parameter initialization, which uses uniform distribution normalized according to the input and output layer size. While self-supervising the network for unsupervised segmentation task, at each iteration, $\{x_n\}$ is used to obtain saliency map by employing MR [17] with multi-channel. For the model, we use two main termination conditions as:

$$\mathcal{L}_{i+1} - \mathcal{L}_i \leq \varepsilon_1 \qquad (1)$$

$$S_{i+1} - S_i \leq \varepsilon_2 \qquad (2)$$

where $\mathcal{L}_{i+1}$ and $\mathcal{L}_i$ denote the cross-entroppy losses of step $(i + 1)$ and (i), S_{i+1} and S_i denote the predicted saliency maps of step $(i + 1)$ and (i), ε_1 and ε_2 are defined small non-zero constants to terminate the training process. Also, when

the training step N achieve maximum value $\kappa = 200$, it will stop the process. It should be noted that self-supervised learing on each image treated independent of the other remaining images on the dataset by initiliazing the parameters of the model and learning procedures from the begining for each image sample. In Fig. 2, unsupervised segmentation outputs and computed saliency maps are shown for different iterations of self-supervised learning. It can be seen that saliency map results are converging even though clustering through self-supervision is not giving optimal segmentation result yet.

3. EXPERIMENTAL RESULTS

Metrics: We made extensive evaluation (see Table 2) based on various performance metrics [24, 25, 26, 27] such as Area Under Curve (AUC_{Borji}), Cross Correlation (CC), Normalized Scanpath Saliency (NSS), Kullback-Leibler divergence (KLdiv), Precision, Recall, F-measure (F_β, $maxF_\beta$, $aveF_\beta$) with Precision-Recall or Precision-Recall Curves.

Dataset: We made evaluation of the model on hyperspectral salient object detection (HS-SOD) dataset [5] consisting of 60 hyperspectral images with their respective binary ground-truth images referring to salient objects. The dataset details can be seen in [5], and is available on [28]. For each image, spatial resolution is 768x1024 pixels, and there are 81 spectral channels covering the wavelengths between 380-780nm (visible spectrum) with 5nm intervals [5].

Evaluation: We selected [7] and [8] for comparison as being the hyperspectral salient object detection models for natural scenes. In work [7], various approaches were tested on hyperspectral data so we also apply the approaches tested in [7] on HS-SOD dataset [5] for comparison. i) spectral distances between each spatial region for saliency computation by using spectral Euclidean distance (SED) and spectral
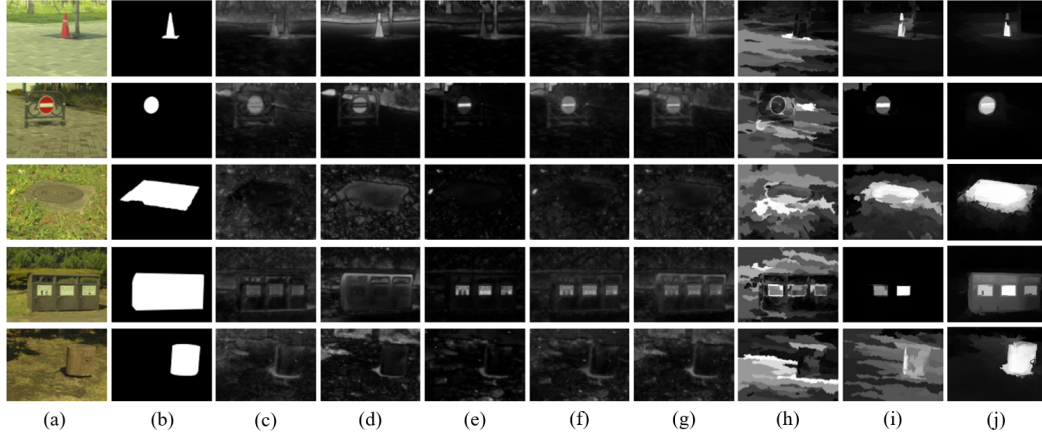
Fig. 3. (a) Sample scenes of the the hyperspectral data rendered in sRGB with its respective (b) ground-truth salient objects, and saliency map results of the compard models: (c) Itti *et al.* [15, 7], (d) SAD [7], (e) SED [7], (f) SED-OCM-GS [7], (g) SED-OCM-SAD [7], (h) SGC [8], (i) HS-MR [17], (j) **Proposed SUDF**$_{HF-Slic}$

Angle distances (SAD) [7, 5], ii) color opponency method in [15, 7] is replaced by spectral grouping rather than Red-Green and Blue-Yellow differences, in which Euclidean distance between spectral group (GS) vectors by dividing spectral bands into four groups (G1,G2,G3,G4) [7, 5]. iii) In [5], spectral distance based saliency also combined with orientation based saliency with combinations such as SED-OCM-GS and SED-OCM-SAD. iv) saliency maps from Itti et al. [15] were also provided for hyperspectral saliency comparison in [7, 5] as a baseline model. As a more recent work, we also tested saliency from spectral gradient contrast (SGC) proposed by [8]. In [8], local region contrast is computed from the superpixels obtained by spatial and spectral gradients, which is used to calculate spectral gradient contrast (SGC) for saliency detection [8, 5].

In addition, saliency detection by graph-based manifold ranking (MR) is also applied on hyperspectral dataset, referred as HS-MR [17] to compare with the proposed model **SUDF**$_{HF-Slic}$ (SUDF: Saliency from Unsupervised Deep Features), which uses higher-level features for both MR based saliency and cluster refinement compared with the original approaches in [17] and [22]. In addition, to demonstrate the performance improvement on saliency detection when the cluster refinement is done on high-level fetures, we also implemented and compared saliency computation when cluster refinement is done based on input hyperspectral image that is referred as **SUDF**$_{HS-Slic}$.

As it can be seen in Table 2, proposed **SUDF**$_{HF-Slic}$ performs better than other approaches in all metrics except being second best on NSS. However, although the performance difference is very close with **SUDF**$_{HF-Slic}$, best performing model in NSS metric is also a variation of the proposed approach, **SUDF**$_{HS-Slic}$ which applies cluster refinement based on the superpixels obtaind from hyperspectral image directly as in the original work [22]. In addition, proposed

SUDF$_{HF-Slic}$ demonstrated that using higher-level features even learned from self-supervision is more beneficial to saliency computation using manifold-ranking since proposed **SUDF**$_{HF-Slic}$ outperformed HS-MR [17] manifold-ranking using low-level features in all evaluation metrics. In Fig. 3, some sample scenes for hyperspectral data are given rendered in sRGB for visualization with their respective grount-truth images for salient objects, and saliency maps results of various approaches on these scenes are also given to demonstrate the performance of proposed model **SUDF**$_{HF-Slic}$ with respect to other models. It can be seen from the saliency maps that our model performs better qualitatively too.

4. CONCLUSION

In this work, we demonstrated hyperspectral salient object detection approach based on self-supervised deep features in a multi-task model. Paramater update of the CNN model is done based on cross-entropy loss of clustering performance, and saliency is computed by the learned features of the unsupervised segmentation task, in which saliency convergence is the termination criteria for the self-supervised learning procedure. Evaluation on the HS-SOD dataset [5] demonstrates promising results for salient object detection with the proposed approach. As a future work, we would like to investigate how to improve representation of hyperspectral image during self-supervision process (e.g. adding sparsity loss, orthogonality constraint, decoder based image generation loss, etc.) to improve the accuracy for salient object detection and also to increase the convergence on clustering and saliency map results. Moreover, we would like to investigate other options for saliency computation compared to MR model since it assumes boundary prior for background regions.

5. REFERENCES

[1] D. H. Sliney, "What is light? The visible spectrum and beyond," *Eye (Nature)*, Vol.30, pp.222-229, 2016.

[2] M. Borengasser, W. Hungate, R.Watkins, "Hyperspectral Remote Sensing: Principles and Applications," CRC Press, Boca Raton FL, 2008.

[3] A. Chakrabarti and T. Zickler, "Statistics of Real-World Hyperspectral Image"s, *in Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[4] R. B. Smith, Introduction to "Hyperspectral Imaging with TNT-mips", online available: http://www.microimages.com

[5] N. Imamoglu, Y. Oishi, X. Zhang, G. Ding, Y. Fang, T. Kouyama, R. Nakamura, "Hyperspectral Image Dataset for Benchmarking on Salient Object Detection", *10th International Conference on Quality of Multimedia Experience (QoMEX)*, 2018.

[6] D. Manolakis, R. Lockwood, T. Cooley, "Hyperspectral Imaging Remote Sensing", Cambridge University Prss, Cambridge, United Kingdom, 2016.

[7] J. Liang J. Zhou, X. Bai, Y. Qian, "Salient object detection in hyperspectral images", *IEEE Int. Conf. on Image Processing (ICIP)*, pp.2393-2397,2013.

[8] H. Yan, Y. Zhang, W. Wei, L. Zhang, Y. Li, "Salient object detection in hyperspectral imagery using spectral gradient contrast", *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp.1560-1563,2016.

[9] Y. Cao, J. Zhang, Q. Tian, L. Zhuo, Q. Zhou, "Salient target detection in hyperspectral images using spectral saliency", *IEEE ChinaSIP*, pp.1086-1090, 2015.

[10] S. Le Moan, A. Mansouri, J. Hardeberg and Y. Voisin, "Saliency in spectral images", *in Proc. of the 17th Scandinavian Conference on Image Analysis*, pp. 114-123, 2011.

[11] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural RGB images", *in Proc. of the Europian Conference on Computer Vision (ECCV)*, pp.19-34, 2016.

[12] R. Kawakami, J. Wright, T. Yu-Wing, Y. Matsushita, M. Ben-Ezra, and K. Ikeuchi, "High resolution hyperspectral imaging via matrix factorization", *IEEE Conf. on Computer Vision and Pattern Recognition (IEEE CVPR)*, 2011.

[13] B. Arad and O. Ben-Shahar, "Filter selection for hyperspectral estimation", *in Proc. of the IEEE International Conference on Computer Vision (ICCV)*, pp.3153-3161, 2017.

[14] H. Kwon and Y. W. Tai, "RGB-guided hyperspectral unsampling", *in Proc. of the IEEE International Conference on Computer Vision (ICCV)*, pp.307-315, 2016.

[15] 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*, vol.20, no11, pp.1254-1259, 1998.

[16] A. Borji and L. Itti, "State-of-the-art in visual attention modeling", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.35, no.1, pp.353-367, 2013.

[17] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency Detection via Graph-Based Manifold Ranking", *IEEE Conf. on Computer Vision and Pattern Recognition (IEEE CVPR)*, 2013.

[18] N. Imamoglu, C. Zhang, W. Shimoda, Y. Fang, and B. Shi, "Saliency detection by forward and backward cues in deep CNNs", *IEEE International Confetrence on Image Processing (ICIP)*, 2017.

[19] T. Liu, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect salient object", *in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.353-367, 2007.

[20] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach", *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-8, 2007.

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

[22] A. Kanezaki, "Unsupervised Image Segmentation by Backpropagation", *IEEE Conf. on Computer Vision and Pattern Recognition (IEEE CVPR)*, 2013.

[23] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks", *in AISTATS*, 2010.

[24] A. Borji, H. R. Tavakoli, D. N. Sihite, and L. Itti, "Analysis of scores, datasets, and models in visual saliency prediction", *IEEE International Conference on Computer Vision (ICCV)*, 2013.

[25] M Kmmerer, T Wallis, M Bethge, "Information-theoretic model comparison unifies saliency metrics", *PNAS*, 112(52), 16054-16059, 2015.

[26] Z Bylinskii, T Judd, A Oliva, A Torralba, F Durand, "What do different evaluation metrics tell us about saliency models?" *arXiv preprint arXiv:1604.03605*, 2016.

[27] S Rahman, N Bruce, "Visual Saliency Prediction and Evaluation across Different Perceptual Tasks", *PloS one*, vol.10 , no.9, 2015.

[28] Hyperspectral Salient Object Detection (HS-SOD) dataset, online available: *https://github.com/gistairc/HS-SOD*