MATERIAL SEGMENTATION IN HYPERSPECTRAL IMAGES WITH A SPATIO-SPECTRAL TEXTURE DESCRIPTOR

Yu Zhang[⋆] *Cong Phuoc Huynh*[ξ] *Nariman Habili*[†,ξ] *King Ngi Ngan*[⋆,‡]

[⋆] Department of Electronic Engineering, The Chinese University of Hong Kong
[ξ]The Australian National University, Canberra ACT 2601, Australia
[†] CSIRO Data61, Locked Bag 8001, Canberra ACT 2601, Australia
[‡]University of Electronic Science and Technology of China

ABSTRACT

In this paper, we address the problem of ground-based hyperspectral image segmentation by combining pixel-level and region-level classification with a region boundary refinement approach. To this end, we represent the spatio-spectral feature of image regions by a descriptor based on Vector of Locally Aggregated Descriptors (VLAD). Further, the region boundaries are refined by minimizing the total region perimeter. Experimental results on a ground-based hyperspectral image dataset clearly demonstrate the advantage of the proposed method over recent prior works, based on several metrics.

***Index Terms*—** material segmentation, hyperspectral images, region texture descriptor, minimal region perimeters

1. INTRODUCTION

The classification and segmentation of materials is essential to the understanding of the constituent substances in a scene. This is useful in many applications. For example, in interior design, inferring the materials of objects in an image would inform the selection of compatible materials from online stores. Specific to hyperspectral images, material recognition has found use in anomalous object detection from aerial images [1], resource mapping [2] and biometrics [3].

Traditional classification methods designed for remote sensing, such as those described in [4–7] only exploit imaging spectra as the exclusive feature for recognition purposes. This is most often due to the limited spatial resolution of remote sensing images. The advent of ground-based hyperspectral sensors [8, 9] with high spatial resolution paves the way for new ideas of exploiting the spatial information for material classification.

A popular way of utilizing spatial information is to post-process the result of a pixel-wise material classifier using a region partitioning method such as partitional clustering [10], watershed transformation [11] and minimal spanning forest [12, 13]. In these approaches, the pixel-wise material map resulting from an SVM classifier is refined by majority voting, which means all pixels in a region are assigned the most frequent label within the region. In [14], the pixel-wise material label map is further refined by an edge-preserving filter to reduce noise and preserve edge information with the first few principle components as the guidance image. An alternative approach is to impose spatial structure on pixel-wise material maps by graph-based optimization techniques [15, 16]. In [15], spatial context provides information to refine the pixel-wise classification by Support Vector Machines using a Markov Random Field regularization.

On the other hand, the explicit modeling of texture is of high relevance and importance to material segmentation. The Bag-of-Features (BoF) method [17] characterizes texture by the responses of pixel values to a set of filter banks. Later, it has been demonstrated that superseding the filter responses by the source image patches could lead to better classification results [18]. It is worth noting that BoF only relies on the zeroth-order information, *i.e.* frequency of textons, as the texture descriptor. More recently, the Vector of Locally Aggregated Descriptors (VLAD) [19] was proposed to capture the first-order information of clusters of local descriptors to represent a global image feature. At a high-level, this advancement is a natural extension to BoF, providing richer information for material classification.

This paper contributes the following novelties. Firstly, we represent region-wise texture features in hyperspectral imagery using the VLAD encoding. Further, we propose a material segmentation method that combines the segmentation maps resulting from pixel-wise and region-wise classification, followed by a segment boundary refinement step.

2. SPATIO-SPECTRAL MATERIAL SEGMENTATION

Suppose we are given a collection of M training pixels, where pixel $\mathbf{x}_i$ is assigned a material label y_i belonging to a finite set $\{1, \ldots, C\}$. With the training data set $\{(\mathbf{x}_i, y_i) : i = 1, \ldots, M\}$, we aim to partition a novel image into disjoint regions $R_1, \ldots, R_K$, where each region is assigned one of the C material labels above. In other words, $\cup_{k=1}^{K} R_k = \Omega$ and $R_k \cap R_l = \varnothing, \forall k \neq l$, where Ω denotes the spatial do-

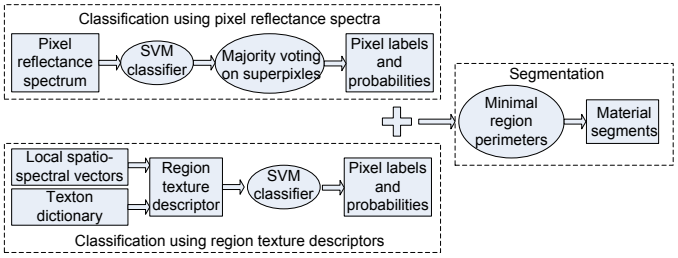

Fig. 1: The proposed material segmentation framework.

main of an image. For the sake of simplicity, we assume that the training and test images are sampled at the same discrete wavelengths λ_j, $j = 1, \ldots, J$. Let $S(\mathbf{x}, \lambda)$ denote the spectral reflectance of an image at the pixel $\mathbf{x}$ and wavelength λ. With this notation, the reflectance spectrum at a pixel $\mathbf{x}$ is denoted by a vector $\mathbf{S}(\mathbf{x}) \triangleq [S(\mathbf{x}, \lambda_1), \ldots, S(\mathbf{x}, \lambda_J)]^T$.

As shown in Fig. 1, the proposed method consists of two components. The top one extracts pixel-level reflectance feature, while the bottom one extracts region-level texture feature, producing two independent material label maps. The two material maps are then fed into a segment boundary refinement step by minimizing the total region perimeters. For the first component, we adopt a pixel-wise classification method using solely pixel reflectance spectra in [20]. Here, we focus on describing the main contribution, which is the region texture descriptor.

2.1. Region Classification with Texture Descriptor

2.1.1. Pixel-wise Spatio-spectral Feature

(a) Per-channel local spatial feature (b) Local spatio-spectral feature across bands

Fig. 2: Formation of the local spatio-spectral vector at each pixel: (a) the per-channel feature vector obtained by reordering image values in the 3×3 neighborhood of each pixel. (b) the local spatio-spectral vector formed by concatenating the per-channel feature vectors.

In our work, the local spatio-spectral vector of single pixel is formulated using each pixel's local neighborhood as follows. Let us consider a 3×3 neighbourhood around a reference pixel $\mathbf{x}$. In each spectral channel λ_j, we formulate the spatial feature $T(\mathbf{x}, \lambda_j)$ by concatenating the reflectance values $S(\mathbf{x}', \lambda_j)$, where $\mathbf{x}'$ is a pixel in the neighbourhood as depicted in Fig. 2a. Subsequently, the feature vectors over all the wavelengths $\lambda_j, j = 1, \ldots, J$, are concatenated into a vector $T(\mathbf{x}) \in \mathbb{R}^{9J}$, representing the local texture feature in pixel $\mathbf{x}$. The process is demonstrated in Fig. 2b.

The high dimensionality, *i.e.* $9J$, of the local spatio-spectral vector $T(\mathbf{x})$ above often poses a number of disadvantages such as expensive and inefficient computational

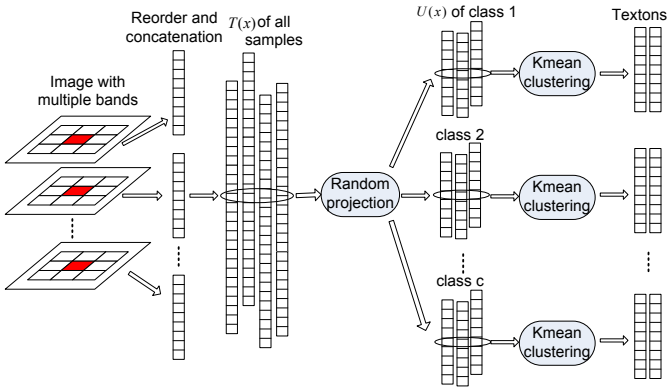

Fig. 3: Flowchart of generating spatio-spectral textons.

load, irrelevant and noisy features. To avoid these disadvantages, random projection method proposed in [21] is applied to $T(\mathbf{x})$ by a projection matrix $\Phi \in \mathbb{R}^{d \times 9J}$ whose elements are independent, zero-mean, unit-variance Gaussian random variables. Subsequently, the texture feature $T(\mathbf{x})$ is projected into a d-dimensional space through the linear projection Φ

$$U(\mathbf{x}) = \Phi T(\mathbf{x}), \tag{1}$$

where $U(\mathbf{x})$ is the random projection of the spatio-spectral vector $T(\mathbf{x})$ into $\mathbb{R}^d$. In addition, we apply the following normalization step as it was reported in [18] to lead to improved empirical classification results.

2.1.2. Formulation of the Region Descriptor

Having obtained the randomly projected feature $U(\mathbf{x})$ at each pixel $\mathbf{x}$, we proceed to obtain the representative texture feature for each region (super-pixel). To this end, we aggregate the local spatio-spectral vectors into a VLAD descriptor per region according to the following procedure.

In the first step, we obtain a code book of "visual words", which we also term as the spatio-spectral "textons", from the training data by k-means clustering of the above local spatio-spectral vectors. Let us denote the N textons, *i.e.* the centroids of the resulting clusters, $V_1, \ldots, V_N$, where $V_n \in \mathbb{R}^d, \forall n = 1, \ldots, N$. In our experiments, the total number of textons is $N = C \times N_t$, where C is the number of classes and N_t is the texton number per class. Fig. 3 summarizes the process of generating the codebook of textons.

The VLAD descriptor for each region measures the total deviation of its local features $U(\mathbf{x})$ from each centroid. By considering the deviations from all the textons, VLAD could effectively signify the textons representative of the local features in each region. This first-order information is an extension from the zeroth-order information in the bag-of-visual-words descriptor, which simply counts the occurrences of the nearest textons in each region.

Specifically, we assign each pixel $\mathbf{x}$ in a given region $\mathcal{X}$ to its nearest texton, in the random texture feature space, and

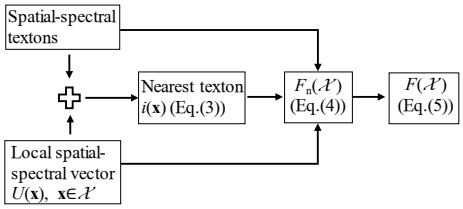

Fig. 4: Formulation of the region texture descriptor.

denote the texton index it is associated with as

$$i(\mathbf{x}) \triangleq \underset{n}{\operatorname{argmin}} \|U(\mathbf{x}) - V_n\|. \quad (2)$$

Next, we compute the total deviation vector for texton V_n as a sum of feature differences over its associated pixels as

$$F_n(\mathcal{X}) \triangleq \sum_{\mathbf{x}\in\mathcal{X}:i(\mathbf{x})=n} (U(\mathbf{x}) - V_n), \forall n = 1, \ldots, N. \quad (3)$$

Lastly, we form the VLAD descriptor for each region $\mathcal{X}$ by concatenating the above d-dimensional deviation vectors over all the textons such that

$$F(\mathcal{X}) \triangleq [F_1(\mathcal{X})^T, \ldots, F_N(\mathcal{X})^T]^T, \quad (4)$$

which is a dN-dimensional vector.

Before the VLAD descriptor is fed into a region classifier, it is l^2-normalized and then projected to a low-dimensional subspace by PCA. We employ a VLAD subspace of 80 dimensions in all our experiments. We summarize the above process of generating the region texture descriptor in Fig. 4.

Using the above texture descriptor, we predict the material class of each region in the hyperspectral image. We adopt the majority voting approach to obtain the ground truth material label of each region from its constituent pixels. With the region texture descriptors and region labels as inputs, we train a region SVM classifier with the RBF kernel. As a result, we employ the resulting classifier to predict the probability $p(c|\mathcal{X})$ that a region $\mathcal{X}$ in the test set is made of the c-th material. Subsequently, we populate the region material labels and probabilities to its pixels $\mathbf{x}$, *i.e.* $l^{region}(\mathbf{x}) = l(\mathcal{X})$ and $p^{region}(c|\mathbf{x}) = p(c|\mathcal{X}), \forall c \in \{1, \ldots, C\}$.

2.2. Region Boundary Refinement

Now we further refine the segmentation maps by enforcing a minimal length constraint on the segment boundaries. Let us $p^{pixel}(c|\mathbf{x})$ denote the material label map obtained from pixel-wise spectrum classification (top part of Figure 1). In addition, this map is combined with the region-wise material map to form the label map $p_c(\mathbf{x}) \triangleq p^{pixel}(c|\mathbf{x}) + p^{region}(c|\mathbf{x})$. Next, the boundary of the combined label map is further refined by the optimization procedure presented in [20] which aims to minimize the total perimeters of the resulting region.

3. EXPERIMENTS AND DISCUSSIONS

In this section, we evaluate the proposed approach on a hyperspectral image dataset introduced in [20] and compare it to state-of-the-art methods.

3.1. Dataset

The hyperspectral images dataset are collected from five source databases and covers diverse materials, such as cloth, soil, vegetation, *etc*. and the images are grouped into four categories, including portrait, landscape, office, and fruit & vegetable. Every image in this dataset contains 28 bands spanning the wavelength from 430 nm to 700 nm with 10 nm increments. Moreover, the ground truth material label is provided for each image pixel.

3.2. Experimental settings

In the experiments, we augment the background as an additional class to the materials of interest, *i.e*. foreground materials. We determine the training sample size based on the number of images and number of classes within each category. For the pixel reflectance feature, we employed 100, 200, 100 and 400 training pixels per class per image for the portrait, landscape, office and fruit & vegetables categories, respectively. Furthermore, for the region classifier, we sample 100, 300, 300 and 400 training regions per class per image from these four categories and each region contains 200 to 400 pixels.

For each of the four categories, we randomly sample 75% of the images as the training data and the other 25% as test data. With the extracted training features, we train an SVM classifier [22] equipped with the radial basis function kernel to generate a material map. In addition, the set of SVM hyperparameters is fine-tuned for each image category via a one-time cross-validation procedure. Furthermore, the weight α of minimal region perimeters constraint [20] is set to 0.125 for portraits and 0.15 for the other categories. We use the SLIC algorithm [23] to generate the regions and the implementation of the region VLAD descriptor in our experiments is inherited from the VLFeat library [24].

3.3. Comparison with prior works

We compare our proposed method with five alternatives, including Multiple Feature Learning (MFL) [25], pixel-wise classification by SVM with Markov Random Field (SVM-MRF) and pixel-wise classification by SVM with minimal region perimeters (P-MRP) [20]. We also compare the proposed method with two deep learning methods, including the fully convolutional network with a VGG16 backbone [26] (VGG16-FCN8s) and I-CNN+CRF [27], which is a shallower version of SegNet [28].

The algorithms under study are evaluated using the following metrics: overall accuracy (OA), average accuracy (AA), and mean of class-wise intersection over union (mIoU).

Table 1: Performance of the variants of the proposed method as compared to three baseline methods on all categories.

Methods	Portrait			Landscape			Office			Fruit & Vegetable		
	OA(%)	AA(%)	mIoU(%)	OA(%)	AA(%)	mIoU(%)	OA(%)	AA(%)	mIoU(%)	OA(%)	AA(%)	mIoU(%)
MFL	79.77	71.19	40.51	73.81	56.63	33.10	54.13	45.66	26.74	91.36	85.91	75.60
SVM-MRF	88.05	79.29	53.27	79.17	62.76	40.45	63.11	54.49	33.67	94.06	92.29	81.73
P-MRP	88.15	79.64	53.78	79.42	64.21	41.00	63.11	54.94	33.66	93.41	92.76	80.65
VGG16-FCN8s	86.98	**81.54**	50.89	78.00	61.72	38.58	59.14	48.29	29.32	92.68	87.24	77.38
I-CNN-CRF	89.43	80.89	55.90	80.01	62.67	40.77	61.67	51.69	31.13	94.79	89.30	83.07
Ours	**89.51**	81.39	**55.92**	**81.72**	**67.91**	**43.90**	**67.43**	**57.62**	**38.00**	**95.14**	**94.32**	**85.34**

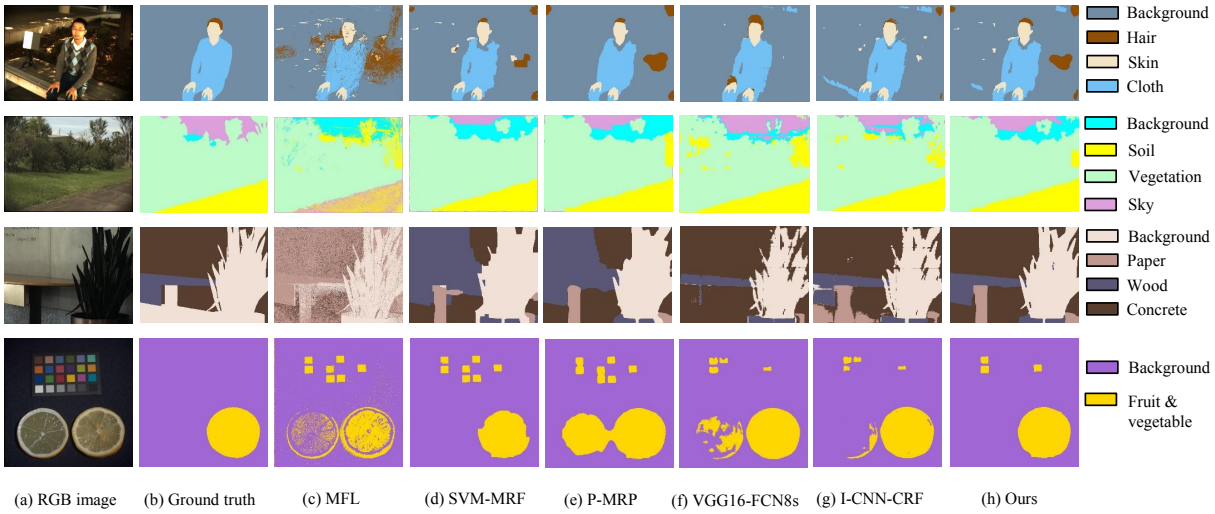

Fig. 5: From top to bottom: segmentation maps for sample portrait, landscape, office, and fruit & vegetable images.

The OA is the overall percentage of correctly labeled pixels. The AA is the average percentage of correctly labeled pixels per class. The mIoU is defined as the ratio of the intersection over the union of the detected pixels and the ground-truth pixels per class. All the reported results are taken as the average of 10 trials of the same experiment with randomly selected training and test sets.

In Table 1, we report the performance of our method and the baseline ones on the four categories. Overall, the proposed method outperforms the baseline ones across all four categories and three evaluation metrics. Notably, it exhibits an improvement approximated to 2%–4% over the baseline methods in terms of all the evaluation metrics. Furthermore, the proposed method can produce better performance than P-MRP, which confirms the complementary roles of the spatial and spectral information in representing material semantics.

Since the presented hyperspectral dataset contains limited training data (tens of images), a deep network can easily overfit the training, leading to sub-optimal test accuracy. This is evidenced by the fact that the I-CNN-CRF (shallow SegNet) performs better than FCN-8s (deeper network) for all the categories. In addition, our method outperforms the other two due to the exploitation of spectral information, which is not a built-in feature of semantic segmentation methods designed for color images.

In Fig. 5, we provide qualitative segmentation maps to complement the above numerical results. It is obvious that the segmentation maps generated from the proposed method are closer to the ground truth than others. In the last image, the left lemon is labeled as background since its material is plastic. These figures show typical examples of the trend that MFL usually produce small isolated segments (that appears as salt-and-pepper noise). However, the proposed method effectively removes small noisy segments and yields smooth segmentation boundaries.

4. CONCLUSION

This paper presents a framework for hyperspectral material segmentation with a spatio-spectral region texture descriptor represented by the VLAD encoding. The method combines results from both pixel-wise and region classification components to produce a rough probability map at each pixel. Subsequently, the material boundaries in this map are refined by enforcing a minimal boundary length constraint. The experimental results on a hyperspectral image dataset show that the proposed method significantly outperforms prior works in several metrics, over all four image categories.

5. REFERENCES

[1] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 4, pp. 1990–2000, April 2016.

[2] R. N. Clark, *Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy*, vol. 3 of *Manual of Remote Sensing, Remote Sensing for the Earth Sciences*, John Wiley and Sons, 1999.

[3] C. P. Huynh and A. Robles-Kelly, "Hyperspectral imaging for skin recognition and biometrics," in *IEEE International Conference on Image Processing*, Sept 2010, pp. 2325–2328.

[4] D. Slater and G. Healey, "Material classification for 3D objects in aerial hyperspectral images," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1999, vol. 2, p. 273.

[5] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351–1362, 2005.

[6] A. Erturk and S. Erturk, "Unsupervised segmentation of hyperspectral images using modified phase correlation," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 4, pp. 527–531, Oct 2006.

[7] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, Aug 2004.

[8] N. Gat, "Imaging spectroscopy using tunable filters: a review," *Proc. SPIE*, vol. 4056, pp. 50–64, 2000.

[9] A. Bodkin, A. Sheinis, A. Norton, J. Daly, S. Beaven, and J. Weinheimer, "Snapshot hyperspectral imaging: the hyperpixel array camera," *Proc. SPIE*, vol. 7334, pp. 73340H–73340H–11, 2009.

[10] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 8, pp. 2973–2987, 2009.

[11] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognition*, vol. 43, no. 7, pp. 2367–2379, 2010.

[12] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Multiple spectral–spatial classification approach for hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4122–4132, 2010.

[13] K. Bernard, Y. Tarabalka, J. Angulo, J. Chanussot, and J. A. Benediktsson, "Spectral–spatial classification of hyperspectral data based on a stochastic minimum spanning forest approach," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2008–2021, 2012.

[14] X. Kang, S. Li, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification with edge-preserving filtering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2666–2677, 2014.

[15] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM-and MRF-based method for accurate classification of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 4, pp. 736–740, 2010.

[16] D. B. Gillis and J. H. Bowles, "Hyperspectral image segmentation using spatial-spectral graphs," in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2012, pp. 83901Q–83901Q.

[17] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *International Journal of Computer Vision*, vol. 43, no. 1, pp. 29–44, 2001.

[18] M. Varma and A. Zisserman, "A statistical approach to material classification using image patch exemplars," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 2032–2047, 2009.

[19] H. Jégou, M. Douze, C. Schmid, and P. Pérez, "Aggregating local descriptors into a compact image representation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2010, pp. 3304–3311.

[20] Y. Zhang, C. P. Huynh, N. Habili, and K. N. Ngan, "Material segmentation in hyperspectral images with minimal region perimeters," in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 834–838.

[21] L. Liu and P. Fieguth, "Texture classification from random features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 3, pp. 574–586, 2012.

[22] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[23] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC Superpixels Compared to State-of-the-Art Superpixel Methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[24] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," `http://www.vlfeat.org/`, 2008.

[25] J. Li, X. Huang, P. Gamba, J. M. Bioucas-Dias, L. Zhang, J. Atli Benediktsson, and A. Plaza, "Multiple feature learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 3, pp. 1592–1606, 2015.

[26] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[27] Y. Zhang, K. N. Ngan, C. P. Huynh, and N. Habili, "Learning deep spatial-spectral features for material segmentation in hyperspectral images," in *Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2017, pp. 1–7.

[28] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *arXiv preprint arXiv:1511.00561*, 2015.