

Structured Receptive Field Networks and applications to hyperspectral image classification

Demetrio Labate^a, Kazem Safari^a, Nikolaos Karantzas^a, Saurabh Prasad^b, Farideh Foroozandeh Shahraki^b

^a Department of Mathematics, University of Houston

^b Department of Electrical and Computer Engineering, University of Houston

ABSTRACT

Deep neural networks have achieved impressive performance in problems of object detection and object category classifications. To perform efficiently though, such methods typically require a large number of training samples. Unfortunately, this requirement is highly impractical or impossible in applications such as hyperspectral classification where it is expensive and labor intensive to generate labeled data for training. A few ideas have been proposed in the literature to address this problem such as transfer learning and domain adaptation. In this work, we propose an alternative strategy to reduce the number of network parameters based on Structured Receptive Field Networks (SRFN), a class of convolutional neural networks (CNNs) where each convolutional filter is a linear combination from a predefined dictionary. To better exploit the characteristics of hyperspectral data to be learned, we choose a filter dictionary consisting of directional filters inspired by the theory of shearlets and we train a SRFN by imposing that the convolutional filters form sparse linear combinations in such dictionary. The application of our SRFN to problems of hyperspectral classification shows that this approach achieves very competitive performance as compared to conventional CNNs.

Keywords: Hyperspectral data, convolutional neural networks, deep learning, multiscale analysis.

1. INTRODUCTION

Hyperspectral imagery (HSI) is an imaging modality that collects hundreds of contiguous wavebands from the electromagnetic spectrum with wavelengths that can range from 400 to 2500 nm and spatial resolutions of 1 m or less. By including wavelengths beyond the visible range, the rich information provided by such imagery has the potential to facilitate robust object recognition and characterization. For this reason, HSI is widely used and has gained increasingly more prominence in applications ranging from remote sensing for ground cover analysis and urban mapping through environmental monitoring and mineral resource exploitation.

Traditional approaches to hyperspectral classification such as kernel based methods [1, 2] typically operate by mapping data into a feature space followed by design and optimization of a classifier. Taking advantage of spatial information has been also widely exploited to improve classification accuracy. Neighboring pixels in HSI data are highly correlated since spatial structures such as buildings and roads are often larger than pixel size so that presence of a material at one pixel affects the likelihood of the same material being present in a neighboring pixel. For this reason, methods designed to jointly assess spatial and spectral properties of pixels have also become very successful [3, 4].

In recent years, deep neural networks (DNNs) have emerged in a wide range of data classification problems including HSI classification. Compared with conventional methods from machine learning

where features are manually selected, DNNs are designed to learn hierarchical features from raw input data by feeding low-level features extracted by the bottom layers into a sequence of layers extracting increasingly more abstract and higher-level features [5, 6]. As a result, this approach has the ability to generate very robust and discriminative features for classification. In contrast to linear transforms and kernel methods that generate features by convolution with fixed filters, the filters of a DNN are learned from training data and a large number of labeled samples are typically required to achieve high classification performance. However, this requirement is impractical or impossible in HSI applications where it is very laborious and expensive to collect and label data. For this reason, a number of strategies have been proposed to apply DNNs for HSI classification despite the limited number of labeled training samples. Among such strategies, transfer learning – where the network or a section of the network is pre-trained on a different dataset – is a popular solution to facilitate the training process. Even though the performance of the classifier tends to decrease when moving away from the type of data where it was trained, this approach has shown promising result in HSI classification [7–9]. Another method proposed in the literature to deal with the scarcity of labeled data is active learning – an iterative procedure that involves selecting a small set of the most informative unlabeled samples. Some researchers have recently proposed interesting methods based on active learning to perform highly efficient training of DNNs for hyperspectral classification [10, 11]. Finally, other studies have proposed clever domain adaptation and semi-supervised learning techniques for a similar task [12, 13].

In this paper, we propose a new strategy to generate deep features for hyperspectral classification without the need of extensive training data. Our method is based on Structured Receptive Field Networks (SRFN) [14], a class of convolutional neural networks (CNNs) where convolutional filters are expressed as linear combinations from a predefined dictionary and only the coefficients of the linear combination are learned during training. We will apply this framework to reduce significantly the number of training parameters of a network by imposing that the convolutional filters of our SRFN are sparse in an appropriate dictionary $\mathcal{D}$. A major novelty of our approach is that our filter dictionary $\mathcal{D}$ consists of a recently proposed collection of discrete directional filters inspired by the theory of shearlets. The main advantage of our design strategy is to impose a geometric bias on the convolutional filters with the result of facilitating the training process while learning more efficiently the fundamental structures of the input data. To illustrate and validate our approach, we run numerical experiments of hyperspectral classification using the University of Houston dataset. Our results show that our method provides very competitive classification performance and outperforms conventional CNNs while training with a significant lower number of parameters.

The rest of the paper is organized as follows. In Section 2, we recall the construction of SRFNs and present our strategy that uses a new class of discrete filters to build a SRFN with special geometric bias. In section 3, we illustrate the application of our approach to problems of hyperspectral classification and compare it to a conventional CNN. We have final remarks in Section 4.

2. GEOMETRIC-BIASED CONVOLUTIONAL NEURAL NETWORKS

We present a new deep learning approach for hyperspectral classification that exploits the strategy of SRFNs to reduce the number of parameters of the network. In order to ensure that the network can efficiently learn the spatial characteristics of the data, we impose a geometric bias on the filters learned by the network. We start by recalling the properties of a SRFN below.

2.1. Structured Receptive Field Networks

The idea of applying structured receptive fields in CNNs was first proposed by Jacobsen et al [14] with the aim of reducing the number of parameters of a CNN while maintaining its “capacity to learn general variances and invariances in arbitrary images”. This idea was inspired by the Scattering Transform [15, 16], a method devised to provide efficient and stable data representations using a convolutional network

implemented through a cascade of wavelet filters and modulus operators over multiple layers. Unlike a CNN, the convolutional filters of the Scattering Transform are fixed from the beginning and chosen according to design principles from wavelet theory. A SRFN by contrast is a CNN that adopts the Scattering Transform idea of selecting fixed filter bases as a prior in the network, but maintains the CNN design of learning the filter parameters from the data. In practice, in a SRFN it is assumed that any $L \times L$ convolutional filter W is expressed as linear combinations of basis filters taken from an appropriate dictionary $\mathcal{D} = \{H_i \in \mathbb{R}^{L \times L}, i = 1, \dots, L^2\}$, that is,

$$W = \sum_{i=1}^{L^2} \alpha_i H_i, \quad (1)$$

where the expansion coefficients α_i are learned during training. Jacobsen et al [14] choose the dictionary $\mathcal{D}$ to consist of a family of Gaussian filters and its smooth derivatives, similar to the filters used in the Scattering Transform, for which it is known that 3-rd or 4-th order is sufficient to capture most local variation found in natural images.

In this work, we adopt the general strategy of SRFNs where only the coefficients α_i in (1) are learned during training with the following critical modifications.

1. We select the filter dictionary with particular criteria that guarantee completeness and high directional sensitivity. To build such dictionary, we use a new construction of *2-dimensional discrete directional Parseval frames with compact support* recently introduced by one of the authors [17]. These filters are designed to provide efficient representations for image patches of small support (e.g., 5×5 pixels) as required for processing HSI.
2. We require that any (matricial) filter of our network is *sparse in our dictionary* by imposing that the linear combination (1) contains very ‘few’ elements. For instance, we impose below that any 5×5 filter is a linear combination of at most 5 basis elements. In our applications below, the 5 basis elements are randomly selected out of the 25 elements in the dictionary and we repeat the random selection of 5 basis elements at each node of the network independently. With this choice, we reduce the total number of network parameters by a factor of 5 with respect to a conventional neural network with the same architecture. In addition, since our filters are highly directional, this constraint effectively imposes that any 5×5 filter has a geometric bias. Note that, since there are several convolutional node in each layer (32 or 64 in our scheme below), potentially all filters can be used by the network.

In the following, we will denote a modified CNN where we adopt the type of filter assignment strategy described above as a *Geometric-Biased CNN* (GBCNN).

2.2. Filter dictionaries for image patches

The dictionary selection is critical for our GBCNN. For that, we adopt a new construction of multidimensional discrete frames with compact support and directional characteristics that was recently introduced by one of the authors [17]. This construction is inspired by the theory of continuous directional transforms such as curvelet and shearlet transforms and motivated by their applications to edge detection [18–20]. It is known that discrete versions of such transforms are very efficient to approximate natural images, due to their directional sensitivity and ability to represent edges [21–23]. However, such discrete constructions are Fourier-based and not suitable to design filters with short support in space domain. The new approach by Atreas et al [17] is designed precisely for this task. It can generate discrete multidimensional Parseval frames that not only have compact support of any prescribed length, but also directional vanishing moments to ensure pronounced directional sensitivity. This construction also allows for additional

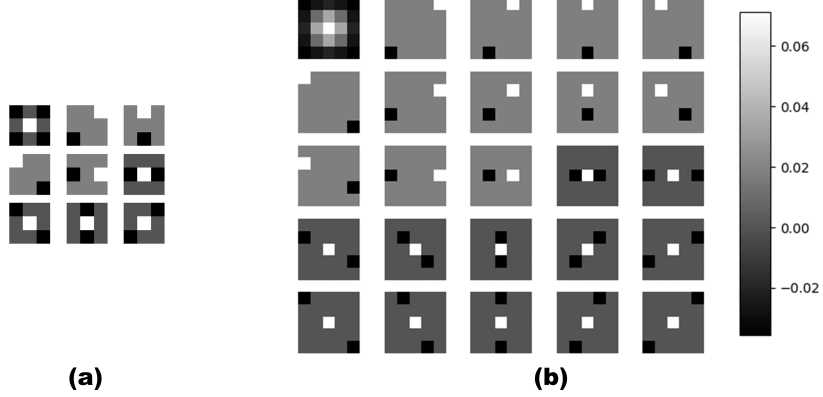


Figure 1. Two-dimensional discrete dictionaries (frames) with compact support and directional characteristics. (a) 3×3 filter dictionary includes from top left: low-pass filter; 8 first-order finite difference filters. (b) 5×5 filter dictionary includes from top left: low-pass filter; 12 first-order finite difference filters; 12 second-order finite difference filters.

flexibility such as sparsity, in the sense that most filter entries are zeros. This property is helpful to speed up the training process of our GBCNN.

To construct such a filter dictionary, we can start with any low-pass filter H_0 whose Fourier transform is given by $H_0(\cdot) = \sum_{k \in J} \alpha_k e^{2\pi i k \cdot} \in L^2(\mathbb{T}^2)$, where $|J| = N < \infty$,

(a) $\alpha_k > 0$ for all $k \in J$ and

(b) $\overline{H_0(q)} = \delta_{0,q}$ for all $q \in \{0, 1/2\}^2$ where $\delta_{0,q}$ denotes the Kronecker delta.

One can prove that, for $v \geq N - 1$ and $c := (\sqrt{\alpha_k})_{k \in J} \in \mathbb{R}^{1 \times N}$, the rows of

$$B = D \operatorname{diag}(c) \in \mathbb{R}^{v \times N}$$

hold the high-pass filter coefficients inducing a Parseval frame for $L^2(\mathbb{R}^2)$ if the rows of $(\frac{c}{D})$ form a Parseval frame for $\mathbb{R}^N$ and if $D^T c = 0$ [17, Thm. 2.5].

Hence, to generate a set of high pass filters of our filter dictionary $H_1, \dots, H_L \in L^2(\mathbb{T}^s)$ where $H_i(\cdot) = \sum_{k \in J} b_k^i e^{2\pi i k \cdot}$, $i = 1, \dots, L$, we proceed as follows.

1. We set $c := (\sqrt{a_k})_{k \in J}$ and observe that, for any $\lambda \in \mathbb{R}^L$ and $\hat{B} := (b_{i,k})_{k \in J, i=1, \dots, L}$, the matrix

$$D_1(\lambda) = \operatorname{diag}(\lambda) \hat{B} \operatorname{diag}(c)^{-1} \in \mathbb{R}^{L \times N}$$

satisfies $D_1(\lambda) c^T = 0$ since the high-pass filters satisfy $H_i(0) = 0$, $i = 1, \dots, L$.

2. We solve the optimization problem

$$\begin{cases} \max \operatorname{trace}(c^T c + D_1(\lambda)^T D_1(\lambda)) \\ \text{subject to } \|c^T c + D_1(\lambda)^T D_1(\lambda)\| \leq 1 \end{cases}$$

to obtain $\lambda^* \in \mathbb{R}^L$ such that the singular values s_ν of $(\frac{c}{D_1(\lambda^*)}) \in \mathbb{R}^{(L+1) \times N}$ satisfy $s_\nu \leq 1$.

3. One can prove [17, Lemma 3.1] that there exists a completion matrix D_2 for which the rows of

$$\begin{pmatrix} c \\ D_1(\lambda^*) \\ D_2 \end{pmatrix} \in \mathbb{R}^{(v+1) \times N}, \quad v \geq N-1$$

form a Parseval frame for $\mathbb{R}^N$ provided that the singular values s_ν of $(D_1(\lambda^*)) \in \mathbb{R}^{(L+1) \times N}$ satisfy $s_\nu \leq 1$. Hence, we obtain a high-pass filter matrix by

$$B = \begin{pmatrix} D_1(\lambda^*) \\ D_2 \end{pmatrix} \text{diag}(c).$$

4. Finally, we can choose to either keep the full high-pass Parseval frame filter bank B or to omit the filters obtained by the completion matrix D_2 . The latter can be done with the aid of [17, Thm. 3.2] when our pre-designed filter bank $D_1(\lambda^*)$ is such that $\text{rank}(D_1(\lambda^*)) = N$. Then the collection of filters $H_0, H_1, \dots, H_L$ induces a frame for $\mathbb{R}^N$.

As an application of this algorithm, we show in Fig. 1 a Parseval frame of 3×3 filters and a Parseval frame of 5×5 filters. Each Parseval frame includes a b-spline low pass filter and a set of directional finite difference filters. Note that the high-frequency filters exhibit pronounced responses along 8 orientations and are highly sparse.

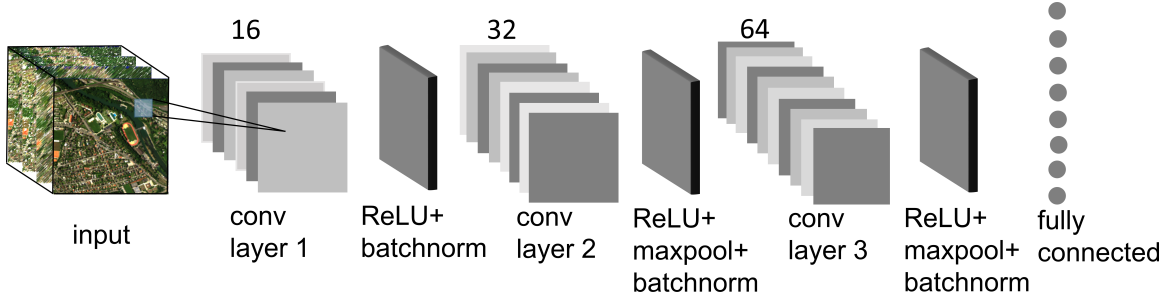


Figure 2. 2D CNN architecture used for our experiments, consisting of 3 convolutional layers followed by a fully connected layer. With respect to a conventional CNN, our GBCNN uses a different training strategy to determine the convolutional filters.

3. RESULTS

We illustrate the application of our deep learning approach to hyperspectral classification and compare the performance of our GBCNN with respect to a conventional CNN with the same architecture. For this experiment, we adopt a simple 3-layer 2-dimensional CNN architecture as illustrated in Fig. 2. This network consists of

1. Layer 1: convolution (filter size 5×5) + relu + batchnorm, number kernels = 16;
2. Layer 2: convolution (filter size 5×5) + relu + maxpool + batchnorm, number kernels = 32;
3. Layer 3: convolution (filter size 3×3) + relu + maxpool + batchnorm, number kernels = 64;
4. Output: fully connected layer with no non-linearity.

Our benchmark data is the University of Houston hyperspectral dataset that was originally released as part of the 2013 GRSS data fusion contest (hyperspectral.ee.uh.edu). Data was acquired using a ITRES-CASI sensor - Vis-VNIR and it contains 144 spectral bands between 380-1050 nm; spatial domain extends over 349×1905 pixels with a spatial resolution of 2.5m. This data includes both landcover and urban regions that are labeled into 15 classes as shown in Fig. 3.

3.1. Experimental setup

The input data to the network are hyperspectral image patches of size $7 \times 7 \times 144$ pixels. We have a total of 9394 such image patches that we split into 3773 of them for training and the remaining 5621 for testing. Note that we did not adopt the same split originally assigned in the 2013 GRSS data fusion contest: we combined the originally assigned training and testing data and randomly selected 40% for training and 60% for testing.

The number of training and testing images belonging to each class is listed in Table 1.

	Class name	Training size	testing size
1	Grass_healthy	417	609
2	Grass_stressed	171	575
3	Grass_synthetic	176	489
4	Tree	155	311
5	Soil	565	298
6	Water	111	90
7	Residential	209	266
8	Commercial	239	379
9	Road	234	292
10	Highway	355	517
11	Railway	393	472
12	Parking_lot1	317	548
13	Parking_lot2	113	142
14	Tennis_court	164	222
15	Running_track	154	411

Table 1. Number of 7×7 image patches per class used for training and testing

The output of the network is a prediction table consisting of integer values between 1 to 15, corresponding to the 15 classes of interest.

3.2. Preprocessing and training

To generate a classification model, data was first normalized by multiplying by 10^{-4} and subtracting the mean and standard deviation channel-wise. The experimental setup is as follows: we trained the network for 300 epochs, with batch size 256, choosing cross entropy softmax with logits as cross function and using the standard Adam optimizer learning rate 10^{-3} .

As discussed above, during the training phase of our GBCNN, each 5×5 or 3×3 convolutional filter is expressed as a sparse linear combination from our dictionary of discrete Parseval frames with 5×5 or 3×3 support and directional vanishing moments described above. We impose that each 5×5 filter is a linear combination of 5 basis filters and that each 3×3 filter is a linear combination of 3 basis filters. By so doing, the overall number of parameters of the network is reduced from 93391 (conventional CNN) to 24783 (our GBCNN), with a 73.5% reduction in the number of parameters.

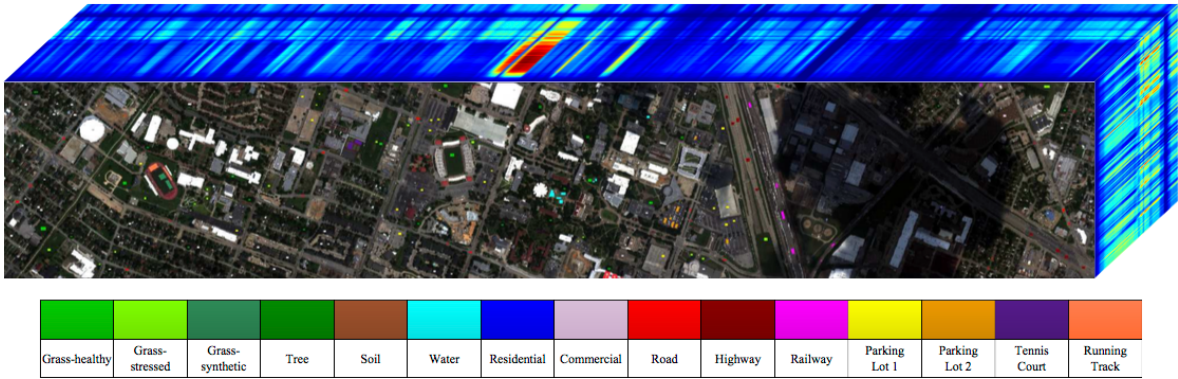


Figure 3. University of Houston hyperspectral dataset. This data contains 144 spectral bands and covers a spatial region of 349×1905 pixels with a spatial resolution of 2.5m per pixel. Data includes 15 labeled classes.

3.3. Testing results

Classification results on the University of Houston dataset are reported in Table 2. We show side-by-side the average accuracy (over 15 classes) and per-class accuracy of our GBCNN and of a CNN with the same architecture. Each accuracy number in the table is the average over 50 runs of the network and we include the standard deviation as a measure of the stability of the test.

The classification results show that our GBCNN outperforms a conventional CNN with the same architecture in terms of average accuracy and yields a significantly lower standard deviation suggesting that the model is more stable. Examination of the per-class accuracy shows that the improvement of our approach is especially remarkable in classes associated with strong directional characteristics such as Road, Highway and Railway. Improvement is negligible or less remarkable for classes such as Grass that are well separated using spectral characteristics alone.

4. CONCLUSION

In this work, we have introduced a new deep learning network called GBCNN for hyperspectral classification that reduces the number of training parameters of the network while imposing a geometrical bias on the convolutional filters. A major novelty of our approach is that the convolutional filters learned by our network are sparse linear combination from a predefined dictionary of highly directional filters. As compared with a conventional CNN, our approach yields a superior classification performance that can be explained with its improved ability to efficiently learn spatial features from the input data. The method presented in this paper is expected to have a wider applicability and to be especially promising for problems of classification with small data size.

Acknowledgments

D.L. acknowledges support from NSF grants DMS 1720487 and 1720452.

REFERENCES

1. U. B. Gewali, S. T. Monteiro, and E. Saber, “Machine learning based hyperspectral image analysis: a survey,” *arXiv preprint arXiv:1802.08701*, 2018.
2. 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.

	GBCNN		CNN	
	mean	std	mean	std
Average accuracy	97.18	1.80	91.85	9.98
Grass healthy	99.51	0.99	98.35	3.47
Grass stressed	99.58	1.22	98.72	2.90
Grass synthetic	99.52	0.51	96.10	14.31
Tree	99.24	0.88	95.87	11.92
Soil	99.62	0.87	95.06	14.73
Water	97.78	1.64	97.97	2.60
Residential	97.22	2	93.54	11.32
Commercial	96.46	3.23	96.44	4.45
Road	93.90	6.87	85.05	21.74
Highway	96.08	4.89	78.96	25.05
Railway	96.29	7.54	85.12	26.40
Parking lot1	91.65	9.33	83.48	21.83
Parking lot2	94.29	3.51	91.99	11.31
Tennis court	98.25	2.75	97.59	8.71
Running track	99.71	0.57	93.73	16.86

Table 2. Classification accuracy on the University of Houston dataset. The table compares our GBCNN against a CNN with the same architecture. Accuracy results are average taken over 50 runs of the network (with different random initializations). The “std” column reports the standard deviation of the accuracy taken over the 50 measures. We denote in bold font the highest accuracy value (average and per-class).

3. M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, “Advances in spectral-spatial classification of hyperspectral images,” *Proceedings of the IEEE*, vol. 101, no. 3, pp. 652–675, 2012.
4. L. He, J. Li, C. Liu, and S. Li, “Recent advances on spectral-spatial hyperspectral image classification: An overview and new guidelines,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 3, pp. 1579–1597, 2017.
5. G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science*, vol. 313, no. 5786, pp. 504–507, 2006. [Online]. Available: <https://science.sciencemag.org/content/313/5786/504>
6. Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
7. J. Lin, R. Ward, and Z. J. Wang, “Deep transfer learning for hyperspectral image classification,” in *2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)*, Aug 2018, pp. 1–5.
8. J. Yang, Y. Zhao, and J. C. Chan, “Learning and transferring deep joint spectralspatial features for hyperspectral classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4729–4742, Aug 2017.
9. L. Windrim, A. Melkumyan, R. J. Murphy, A. Chlingaryan, and R. Ramakrishnan, “Pretraining for hyperspectral convolutional neural network classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 5, pp. 2798–2810, 2018.
10. P. Liu, H. Zhang, and K. B. Eom, “Active deep learning for classification of hyperspectral images,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 2, pp. 712–724, Feb 2017.

11. C. Deng, Y. Xue, X. Liu, C. Li, and D. Tao, "Active transfer learning network: A unified deep joint spectralspatial feature learning model for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 3, pp. 1741–1754, March 2019.
12. X. Zhou and S. Prasad, "Deep feature alignment neural networks for domain adaptation of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 10, pp. 5863–5872, Oct 2018.
13. B. Pan, Z. Shi, and X. Xu, "Mugnet: Deep learning for hyperspectral image classification using limited samples," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, pp. 108 – 119, 2018, deep Learning RS Data. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0924271617303416>
14. J.-H. Jacobsen, J. C. van Gemert, Z. Lou, and A. W. M. Smeulders, "Structured receptive fields in cnns," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. [Online]. Available: <https://ivi.fnwi.uva.nl/isis/publications/2016/JacobsenCVPR2016>
15. S. Mallat, "Group invariant scattering," *Communications on Pure and Applied Mathematics*, vol. 65, no. 10, pp. 1331–1398, 2012. [Online]. Available: <http://dx.doi.org/10.1002/cpa.21413>
16. J. Bruna and S. Mallat, "Invariant scattering convolution networks," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 8, pp. 1872–1886, Aug 2013.
17. N. Karantzas, N. Atreas, M. Papadakis, and T. Stavropoulos, "On the design of multi-dimensional compactly supported parseval framelets with directional characteristics," *Linear Algebra and its Applications*, vol. 582, pp. 1 – 36, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0024379519303155>
18. E. J. Candès and D. L. Donoho, "Continuous curvelet transform: I. resolution of the wavefront set," *Appl. Comp. Harm. Analysis*, vol. 19, pp. 162–197, 2005.
19. G. Kutyniok and D. Labate, "Resolution of the wavefront set using continuous shearlets." *Trans. Am. Math. Soc.*, vol. 361, no. 5, pp. 2719–2754, 2009.
20. K. Guo and D. Labate, "Characterization and analysis of edges using the continuous shearlet transform," *SIAM J. Imaging Sciences*, vol. 2, no. 3, pp. 959–986, 2009.
21. D. Labate, W. Lim, G. Kutyniok, and G. Weiss, "Sparse multidimensional representation using shearlets," *SPIE Proc. 5914, SPIE, Bellingham*, pp. 254–262, 2005. [Online]. Available: [http://en.wikipedia.org/wiki/Bandelet_\(computer_science\)](http://en.wikipedia.org/wiki/Bandelet_(computer_science))
22. K. Guo and D. Labate, "Optimally sparse multidimensional representation using shearlets," *SIAM J. Math. Analysis*, vol. 39, no. 1, pp. 298–318, 2007.
23. —, "Detection of singularities by discrete multiscale directional representations," *The Journal of Geometric Analysis*, vol. 28, pp. 2102–2128, 2018.