ABSTRACT
We propose an active learning algorithm for labeling hyperspectral images (HSI). Pixels with ambiguous class affinity are iteratively estimated using geometric and statistical properties of the data. These pixels are then labeled with ground truth data, yielding a small but potent set of labeled pixels, which are consequently used to label the remaining data. The proposed method enjoys quasilinear complexity in the number of sample pixels, as well as competitive empirical performance on real HSI. Substantial improvement in labeling accuracy compared to unsupervised learning and existing active learning methods is observed with just a few well-selected label queries.

Index Terms—Hyperspectral images, active learning, statistical learning, unsupervised learning, diffusion geometry

1. INTRODUCTION
Hyperspectral imagery (HSI) is an important data source in the geosciences [1]. Hyperspectral sensors record reflectance of a scene at localized electromagnetic ranges, allowing for precise discrimination of materials on the basis of their spectral properties. The effectiveness of HSI for material differentiation has led to its proliferation, and the consequent deluge of HSI data is problematic for manual analysis. Indeed, the large number of pixels and high ambient dimensionality of HSI demand efficient statistical and machine learning algorithms [2] to automatically process and glean insight from the glut of hyperspectral data now available.

While HSI abounds, it is time-consuming and inefficient for humans to manually label substantial proportions of the pixels in a large HSI. Given the huge size of many HSI scenes, even traditional supervised learning may require onerous human resources to label a nontrivial percentage of pixels in the scene for classification. Indeed, these methods often require up to 10% of all pixels be used for training in order to achieve robust statistical performance, which may be a prohibitive human burden. The large sample sizes for training are required when the training set is generated as a uniform random sample over all pixels in the image, which leads to regions in the HSI that are challenging to classify to be sampled on average as much as regions that are easy to classify, leading to inefficient use of labeling resources. It may not be necessary to have as many training samples from a simple region in the HSI as are needed from a complicated region; moreover samples near boundaries between regions (in the high-dimensional spectral space and/or in the spatial domain, depending on the classification techniques used) should be expected to be more informative than labels near the “core” the classes.

To move beyond this classical training regime, innovative methods for unsupervised [3, 4, 5], zero and low-shot [6, 7, 8], and active learning [3, 9, 10] of HSI are necessary. These methods eliminate (unsupervised, zero-shot) or significantly reduce (low-shot, active learning) the need for labeled training pixels in segmenting or classifying HSI, and are suitable for the big data regime in which it is costly to label pixels for training. In active learning, a scientific user—or algorithm—determines a small subset of the data to query for labels. In this setting, generating labels for training is costly or restrictive, so it must be done parsimoniously rather than randomly.

We propose an active learning algorithm that automatically determines which pixels in an HSI to query for ground truth labels, based on geometric and statistical properties learned from the data without supervision. The algorithm iteratively estimates points to query using diffusion distance, and is called iterative active diffusion learning (IADL). The algorithm significantly improves on unsupervised clustering of HSI data with only very few labeled pixels.

2. PROPOSED ALGORITHM
The proposed IADL algorithm proceeds in three major stages. First, diffusion maps [11, 12] and empirical density estimations are used to compute modes of an HSI dataset; by mode,
where we mean high density point in a cluster. Second, pixels to query for labels are computed, by measuring a pixel’s diffusion distance to its nearest labeled pixel. If the two nearest labeled pixels to a given unlabeled pixel are roughly at the same diffusion distance, then this pixel’s class label should be considered ambiguous, as it lies near the cluster boundaries. The pixels to query are computed iteratively, that is, after each new label is acquired, the most ambiguous pixel is re-computed. Third, the remaining pixels are labeled according to their nearest spectral diffusion distance neighbor of higher density and spatial regularization is applied. The re-computation of ambiguous pixels in the active learning stage produces a spectrally diverse set of training pixels, which is expected to enhance labeling accuracy compared to fully unsupervised learning. The first and third steps of the algorithm form the basis for the diffusion learning framework, which enjoys state-of-the-art performance for HSI clustering [3].

The IADL algorithm starts by computing class modes using diffusion distances and kernel density estimation [3]. For HSI data $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^D$, interpreting each pixel as a point in a $D$-dimensional Euclidean space with $D$ the number of spectral bands, a Markov transition matrix $P$ with state space $X$ is built as follows. Let $W$ be the $N \times N$ weight matrix $W(x, y) = e^{-\|x-y\|^2/\sigma^2}, x \in N_N(y)$ and $W(x, x) = 0$ otherwise, where $N_N(x)$ is the set of $k$-nearest neighbors of $y$ in $X$ with respect to Euclidean distance and $\sigma$ is a scale parameter. The degree of $x$ is $\deg(x) := \sum_{y \in X} W(x, y)$. $P$ is the $N \times N$ transition matrix $P(x, y) = W(x, y)/\deg(x)$. The matrix $P$ encodes the transition probabilities of a random walk on $X$. By construction, $P$ has a spectral decomposition $\{\lambda_n, \Phi_n\}_{n=1}^N$, and we define the diffusion distance between $x, y \in X$ as $d_t^2(x, y) = \sum_{n=1}^N \lambda_n^t (\Phi_n(x) - \Phi_n(y))^2$. Note that the diffusion distance depends on a parameter $t$, which corresponds to how long the diffusion process runs. We set $t$ to be 30 in this paper; diffusion learning is relatively robust to choice of $t$ [3].

The procedure for learning the modes of the classes is summarized in Algorithm 1. It first computes a kernel density estimate for each pixel $x_n$ as $p(x_n) = p_0(x_n)/\sum_{m=1}^N p_0(x_m)$, where $p_0(x_n) = \sum_{x_n \in NN(x_n)} e^{-\|x_n-x_m\|^2/\sigma_1^2}$. In our experiments we set $k = 20$, though very large HSI may benefit from larger $k$. The parameter $\sigma_1$ in the exponential kernel is fixed to be one twentieth the mean distance between all pixels.

After computing $p$, the modes of the HSI classes are computed using the time-dependent quantity $\tilde{\rho}_t$ that assigns, to each pixel, the minimum diffusion distance between the pixel and a pixel of higher empirical density:

$$\tilde{\rho}_t(x_n) = \min_{\{p(x_n) \geq \rho_t(x_n)\}} \{d_t(x_n, x_m), x_n \neq \arg \max_{x_m} p(x)\},$$

where $d_t(x_m, x_n)$ is the diffusion distance between $x_m, x_n$, at time $t$. We normalize as $\rho_t(x_n) = \tilde{\rho}_t(x_n)/\max_{x_m} \tilde{\rho}_t(x_m)$. The modes of the HSI are computed as the $K$ maximizers $x'_1, \ldots, x'_K$ of the quantity $D_t(x_n) = p(x_n)\rho_t(x_n)$; here we assume the number of classes $K$ is known. The modes are given unique, arbitrary labels. Under mild assumptions, these modes correspond to different material clusters in the HSI [3, 13]. The use of diffusion distance allows to detect modes in nonlinear, anisotropic, and high dimensional clusters.

**Algorithm 1:** Mode Detection Algorithm

<table>
<thead>
<tr>
<th>Input: $X, K, t$.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Compute ${p(x_n)}_{n=1}^N$.</td>
</tr>
<tr>
<td>2. Compute ${\rho_t(x_n)}_{n=1}^N$.</td>
</tr>
<tr>
<td>3. Compute ${x'<em>i}</em>{i=1}^K$, the $K$ maximizers of $D_t(x_n) = p(x_n)\rho_t(x_n)$.</td>
</tr>
<tr>
<td>Output: ${x'<em>i}</em>{i=1}^K, {p(x_n)}<em>{n=1}^N, {\rho_t(x_n)}</em>{n=1}^N$.</td>
</tr>
</tbody>
</table>

After the modes are learned and arbitrarily labeled, they are used to determine pixels that are most ambiguous with regard to which mode they should be assigned. More precisely, for each pixel $x_n$, let $x_{n_1}, x_{n_2}$ be the labeled pixels $d_t$-nearest to $x_n$ (in the first iteration, only the modes are labeled). If the quantity $F_t(x_n) = |d_t(x_n, x'_{n_1}) - d_t(x_n, x'_{n_2})|$ is close to 0, then there is substantial ambiguity as to the labeled point nearest to $x_n$. An intuitive level, the distances $d_t(x_n, x'_{n_i}), i = 1, 2$ measure affinity to the most likely classes of $x_n$; when these affinities are very close, the pixel $x_n$ is ambiguous and near the class boundaries. By iterating after each query, a spectrally diverse set of labeled pixels is produced, which is expected to be helpful in subsequent labeling of remaining pixels. The active learning procedure is detailed in lines 3-8 of Algorithm 2. Suppose the user is afforded the labels of exactly $L$ pixels. The $L$ labels queried in our active learning regime are the $L$ minimizers of $F_t$, where we re-compute $F_t$ after each query. The active learning setting is most interesting when $\alpha = L/N$ is very small, where $N$ is the total number of pixels in the image. This corresponds to the case when few training labels are available, due to human constraints or difficulty in computing labels with specialized scientific tools.

After the active learning stage of the algorithm, there are labeled according to their nearest label in the spectral domain, unless this label strongly violates spatial regularity. The use of spatial information in the diffusion learning framework has been shown to improve clustering performance [3], and we incorporate it in the proposed active learning algorithm.

More precisely, pixels are labeled in a two-stage process. In the first stage, running in order of decreasing empirical density, the spatial consensus label of each pixel is computed by finding all labeled pixels within distance $r_s \geq 0$ in the spatial domain of the pixel in question; call this set $NN_{r_s}(x_n)$. If one label among $NN_{r_s}(x_n)$ occurs with relative frequency $> 5\%$, that label is the spatial consensus label. Otherwise, no spatial consensus label is given. In detail, let $L_n^{\text{initial}} = \{y_m \mid x_m \in $
The spatial resolution is 1.3 m/pixel. There are 102 spectral bands. The HSI and ground truth are in Figure 1. First, we consider a subset of the Indian Pines dataset, which was captured in Northwest IN, USA by the AVIRIS sensor in 1992. The spatial resolution is 20m/pixel and there are 200 spectral bands. The HSI and ground truth are in Figure 1.

Fig. 1: 50 × 25 subset of the Indian Pines dataset used for analysis, consisting of 3 classes. Left: sum of all spectral bands. Right: the corresponding ground truth labels.

Experimental results are shown in Figure 4. As α increases, the proposed iterative algorithm begins to significantly outperform the one-shot algorithm, which is itself an improvement over fully unsupervised learning.

3.2. Pavia Dataset

Next, we consider a subset of the Pavia dataset, which was captured by the ROSIS sensor during a flight over Pavia, Italy. The spatial resolution is 1.3 m/pixel. There are 102 spectral bands.
bands. A visual representation of the high dimensional data, along with the ground truth labels, are in Figure 2.

![Figure 2: 270 × 50 subset of the Pavia dataset used for analysis, consisting of 6 classes. Top: the sum of all spectral bands. Bottom: the corresponding ground truth labels.](image)

Experimental results are in Figure 4: in the very low \(\alpha\) regime, the proposed iterative method substantially outperforms the one-shot method. While the unsupervised method is quite effective, incorporating a small number of labeled pixels though active learning improves relative performance notably. Moreover, the iterative method produces more rapid gains in accuracy than the one-shot method.

### 3.3. Salinas A Dataset

Finally, we consider the Salinas A dataset, which was captured over Salinas Valley, CA, by the AVRIS sensor. The spatial resolution is 3.7 m/pixel. The image contains 224 spectral bands. A visual representation of the high dimensional data, along with the ground truth labels, appear in Figure 3. Experimental results are shown in Figure 4. While both methods dramatically improve over the unsupervised results with relatively small \(\alpha\), the proposed iterative algorithm is able to improve more rapidly, illustrating its utility. Indeed, the rapid convergence of the accuracy towards 1 illustrates that there is likely to be a region of pixels that are very ambiguous, and once these pixels have been investigated, remaining pixels are easy to label.

![Figure 3: The 86 × 83 Salinas A HSI data consists of 6 classes. Left: the sum of all spectral bands. Right: the labeled ground truth.](image)

### 3.4. Discussion of Experimental Results

The proposed IADL algorithm substantially improves over the one-shot variant, which itself is an improvement over fully unsupervised learning. The proposed method efficiently queries regions of an HSI that are challenging to label without supervision, and makes more efficient use of training labels than classical supervised learning algorithms.

Indeed, a state-of-the-art supervised algorithm, the edge preserving filtering (EPF) algorithm [14], was also applied to each of these datasets. In the very low training regime, with training set of size .1% of total number of pixels, the proposed active learning algorithms substantially outperform the EPF algorithm. Even when the training size is 1% of the total number of pixels, the active learning method outperforms EPF on the Indian Pines and Salinas A dataset. This indicates that the proposed method can perform competitively with state-of-the-art supervised algorithms in the low-sample regime, and moreover, with orders of magnitude fewer labeled pixels. For brevity, these results are not shown.

### 3.5. Computational Complexity

Using cover trees [15, 3], the complexity of computing the class modes and labeling pixels in the proposed scheme is \(O(C^dDN \log(N))\), where \(d\) is the intrinsic dimension of the data. The active learning stage requires \(L\) computations estimating the most ambiguous pixel. Each of these computations is \(O(N)\), so the active learning stage is \(O(LN)\). Thus, the overall complexity of Algorithm 2 is \(O(C^dDN \log(N) + LN)\). In the low-sample active learning regime, \(L \ll N\) and may be assumed \(O(1)\) independent of \(N\). Thus, the overall active learning algorithm is \(O(C^dDN \log(N))\), which is essentially linear in the case \(d\) small. Many active learning algorithms scale as \(O(N^2)\) or worse [16, 17], so the proposed
method enjoys very competitive computational complexity.

4. CONCLUSIONS AND FUTURE DIRECTIONS

The proposed IADL algorithm greatly improves over unsupervised learning of HSI with very few training labels. The experiments show that there is improvement in the iterative re-computation of the most ambiguous pixels after each active learning query.

The proposed IADL algorithm essentially labels pixels that are near class boundaries, in the sense that the labeled pixels corresponding to the class modes (learned without supervision) and the queried pixels (labeled with ground truth through active learning) are insufficient to confidently label these pixels. It is of interest to develop models of data for which this method enjoys performance guarantees. Indeed, the IADL algorithm explicitly incorporates the geometry of data learned via diffusion processes on graphs, and may be viewed as a kind of semisupervised learning on graphs [18, 19]. Rigorous analysis of what features of a data-dependent graph make our method effective is a topic of future work.

5. REFERENCES


