Naval Research Laboratory

Washington, DC 20375-5320

NRL/5510/MR--2021/1

Semi-supervised Learning on Hyperspectral Imagery

Leslie N. Smith

Kristen Nock

Krish Ganotra (*summer intern*)

Navy Center for Applied Research in AI
Information Technology Division

Colin Olson

Advanced Processing Section
Optical Sciences Division

Mark Snyder

Jacobs
Hanover, MD

March 10, 2021

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
10-03-2021	NRL Memorandum Report	

4. TITLE AND SUBTITLE

Semi-supervised Learning on Hyperspectral Imagery

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Leslie N. Smith, Kristen Nock, Krish Ganotra*, Colin Olson, and Mark Snyder**

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
1J06

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375-5320

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/5510/MR--2021/1

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
One Liberty Center
875 N. Randolph Street, Suite 1425
Arlington, VA 22203-1995

10. SPONSOR / MONITOR'S ACRONYM(S)

ONR

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

*Summer Intern
**Jacobs, 7740 Milestone Parkway, Hanover, MD 21076

14. ABSTRACT

Classification of hyperspectral imagery (HSI) is an important research topic for both military and commercial applications. Significant research into deep learning techniques have been concentrated on this task. However, classifying highdimensional HSI data with a limited number of training samples remains an open issue. Semi-supervised learning offers a solution as it requires labeling only a small subset of the pixels and leverages a large number of unlabeled data during its training. In this paper we investigate a variety of both supervised and semi-supervised methods on a new, large HSI dataset called AeroRIT. We determine that the results previously reported in the semi-supervised literature for smaller, older HSI datasets did not generalize to the AeroRIT dataset. Overall, the semi-supervised methods did not provide a significant improvement in performance relative to supervised learning in the limited labeled pixel regime.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Leslie N. Smith
UU	UU	UU	UU	25	**19b. TELEPHONE NUMBER** *(include area code)* (202) 767-9532

This page intentionally left blank.

CONTENTS

FIGURES

TABLES

This page
intentionally
left blank

EXECUTIVE SUMMARY

Automatic identification of land cover and man-made structures from hyperspectral satellite imagery is important in both military and commercial applications. Deep learning techniques have demonstrated high performance in the task of classifying every pixel of hyperspectral imagery (HSI) but these methods rely on training with large quantities of manually labeled samples. Since semi-supervised learning has demonstratively reduced the labeling burden with numerous computer vision applications, we investigated semi-supervised learning methods in the context of a new, larger HSI dataset called AeroRIT, which has not been the subject of any previous study of semi-supervised learning.

In this study we used the following four methods:

1. Supervised learning in the limited labeled pixel regime
2. Superpixel method: a SSL method not based on neural networks
3. Small shallow network (both CNN and dense layers) in the limited labeled pixel regime
4. Cross-Consistency Training (CCT) Algorithm that demonstrated state-or-the-art results for RGB data

Our conclusions from this study are:

- Evaluation metrics for HSI performance are not standardized and can significantly impact the performance and conclusions reported in the literature
- The results previously reported in the semi-supervised literature for smaller, older HSI datasets did not generalize to the AeroRIT dataset
- Semi-supervised learning did not provide a significant improvement in performance relative to supervised learning in the limited labeled pixel regime

We conclude that improving HSI performance in the limited labeled pixel regime remains an open problem.

This page
intentionally
left blank

SEMI-SUPERVISED LEARNING ON HYPERSPECTRAL IMAGERY

1. INTRODUCTION

Since the first satellite imagery, land cover classification has been widely studied. Over the years, the technological advancement of satellite imagery have evolved to capture high resolution, multispectral images causing the automatic identification of land cover and man-made structures to become valuable in military and commercial applications. In remote sensing, multispectral imagery captures object radiance at many wavelength bands. This increases the discriminative ability of objects relative to conventional color images (i.e., RGB). Hyperspectral imagery (HSI) is a subset of multispectral imagery where the wavelength resolution is fine (generally resulting in hundreds of bands) and their range is high.

Due to advances in deep learning for computer vision tasks, convolutional neural networks (CNNs) are now widely used for the analysis of remote sensing imagery. In particular, CNNs are used for semantic segmentation, which is the task of classifying every pixel of imagery (RGB and HSI). Although CNN based classifiers have demonstrated good performance on HSI data, their training relies heavily on having a large quantity of labeled data. Unfortunately, per pixel labeling of HSI data is a heavy investment that erects a barrier to using deep learning techniques ubiquitously on new HSI data collections.

Semi-supervised learning offers a solution as it requires labeling only a small subset of the pixels and leverages a large number of unlabeled data during its training. In this report, we investigate this scenario of a new, large HSI dataset where only a few pixels are labeled and most of the pixels remain unlabeled. As we describe below, the new AeroRIT dataset [1] is a large, recently collected HSI dataset that is available in the public domain. It has not been used in any previous study of semi-supervised learning for HSI. We evaluate the efficacy of several semi-supervised learning methods on AeroRIT with a consistent framework and discovered significant divergence from previously published results.

2. HYPERSPECTRAL IMAGERY (HSI)

Hyperspectral imagery (HSI) collects and processes information from across the electromagnetic spectrum instead of just assigning primary colors (red, green, blue) to each pixel. Whereas the human eye sees color of visible light in mostly three bands (long wavelengths - perceived as red, medium wavelengths - perceived as green, and short wavelengths - perceived as blue), spectral imaging divides the spectrum into many more bands. This technique of dividing images into bands can be extended beyond the visible. The light striking each pixel is broken down into many different spectral bands across the electromagnetic spectrum in order to provide more information on what is imaged. This additional information from each pixel can be used to perform advanced tasks like finding objects, identifying materials, vehicle tracking and occlusion handling. The goal of hyperspectral imaging is to obtain the spectrum for each pixel in the image of a scene, with the purpose of finding objects, identifying materials, or detecting processes. Not only does HSI have rich spectral information, the spatial relationship among different spectra in neighboring regions enable elaborate

Manuscript approved February 26, 2021.

spectral-spatial models in segmentation approaches. Taking the spatial aspect into account during the analysis improves the model robustness and efficiency.

Signals coming from the Earth surface are changed by atmospheric perturbations such as clouds and atmospheric aerosols. Hence, the reflectance is preferably used, which is defined as the ratio between the emitted flux of the surface and the incidental flux. This ratio gives the reflecting effectiveness of a given object for each light wavelength band and is an intrinsic property of the materials, which makes it discriminative for classification purposes.

To prepare for classification, preprocessing, atmospheric correction and normalization are commonly employed. Atmospheric correction methods take the intensity images as input and produce reflectance images, by getting rid of the light diffusion effects and radiative phenomena of the atmosphere. Band selection is commonly employed because the hundreds of measured wavelengths in hyperspectral imagery is unnecessary for classification. Dropping uninformative bands improves the efficiency of classifiers. In addition, it is a common practice in the machine learning community to normalize the data beforehand to zero-mean and unit-variance for which classifiers are known to behave well.

In practice, hyperspectral images are three-dimensional cubes with two spatial dimensions (width and height) and a spectral one (bands). The term hypercube is commonly used to refer to HSI data cubes. Fortunately, all values in the hypercube are expressed in the same unit, either light intensities or reflectances, which makes operations on subsets of the cube mathematically valid. This property allows 3D convolutions operations on the hypercube.

Often a pixel of HSI data corresponds to a surface made of several various materials which produce a spectra mixture. This might be due to the fine spatial resolution of the sensor or pixels that are at the boundaries of different materials. Reference spectra of pure materials are called end-members and these spectra can be used as labeled data in the limited labeled data scenario.

2.1 AeroRIT dataset

In past years, the top three baseline hyperspectral remote sensing datasets have been Indian Pines, University of Pavia, and Salinas. These three datasets are shown in Figure 1. The largest of these is the University of Pavia dataset.

In 2019, Rangnekar et al. [1] collected the AeroRIT dataset, which is nearly 8 times larger than the University of Pavia dataset. The AeroRIT data, shown in Figure 2, collected by drones flying over Rochester Institute of Technology, contains 5 relevant classes: buildings, roads, vegetation, cars, and water. A significant portion of pixels are also classified as unspecified, generally due to lack of HSI data for that pixel (shows up as 0 value for all bands). Additionally, there is a significant class imbalance skewed towards more pixels of vegetation and buildings and less of cars and water. We leverage this dataset in our research for its robust size and because it contains a complex distribution.

3. BACKGROUND

3.1 HSI in Deep Learning

Deep learning classification methods for HSI can be divided into two categories: methods that use just spectral information and those that use spatial-spectral information. Typically, fully supervised spatial-spectral

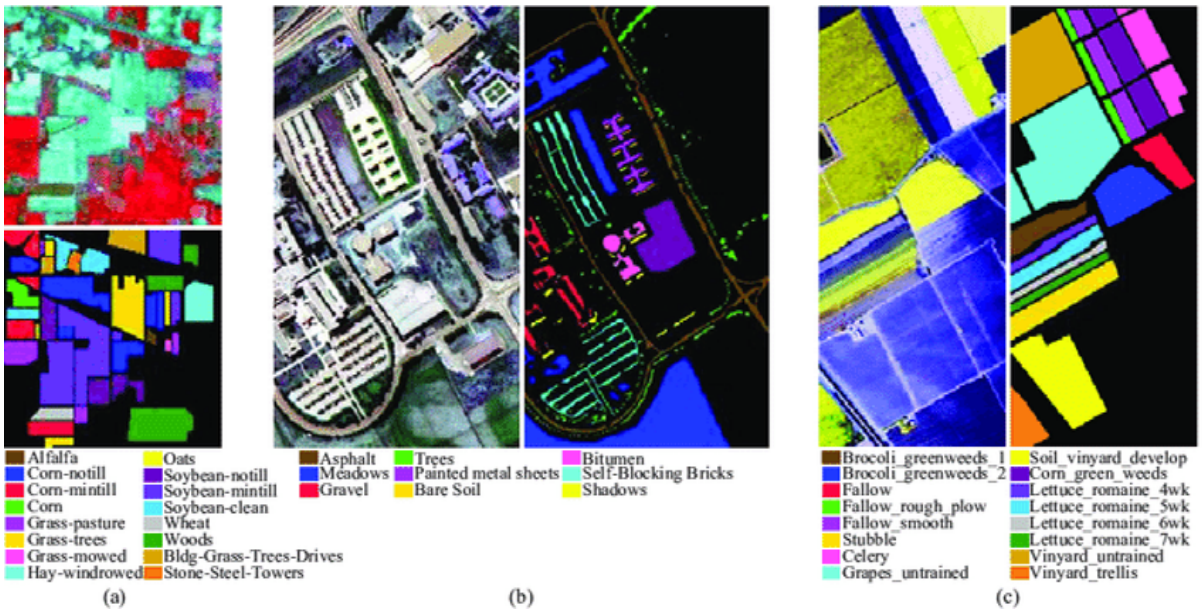


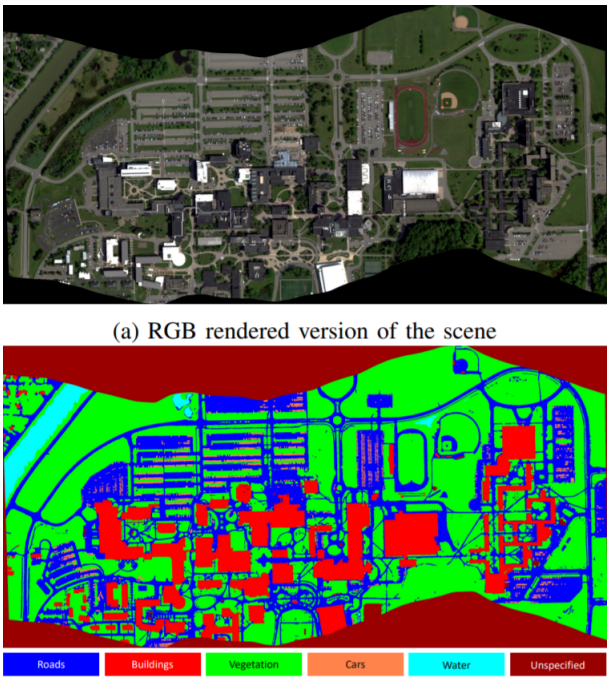
Fig. 1—Pseudo-color and ground truth references for the following hyperspectral remote sensing datasets: (a) Indian Pines; (b) University of Pavia; and (c) Salinas

methods with lots of labeled training data exceed the performance of methods that use spectral information alone. In Section 5 we investigate how both types of methods work in semi-supervised scenario where the labeled examples are severely limited.

In the arena of spectral information classifiers, autoencoders (AE) are unsupervised models with encoders and decoders that are trained to reconstruct the input spectral data [2]. This encoder-decoder architecture is common in HSI classifiers, even in supervised models such as Unets [3]. The AE learns to compress the original data with minimal information loss. The most straightforward evolution from SVMs or Random Forests to deep learning is using a deep fully-connected network on single pixel data (i.e., spectral band information) for classifier. This has been updated with the use of 1D convolutional neural networks (CNNs) which learn filters to be applied on individual spectra [4]. More recently, semi-supervised methods using pseudo-labeling has been introduced [5].

The use of CNNs on color (i.e., RGB) imagery has proven very successful. Typically a small kernel size (i.e., 3x3) is defined in each layer that learn the spatial features to perform a task, such as classification. The kernel size defines the size of the input image that influences the layer's output and this influence size is known as the receptive field. The obvious way to extend CNNs to HSI data is to use 3D CNNs where the two of the dimensions correspond to the input's height and width, and the third dimension covers the bands. This is a promising approach [6] that handles directly the hyperspectral cube with 3D CNN which work simultaneously on the three dimensions using 3D convolutions. In supervised learning, 3D CNNs typically provide the best classification performance. However, 3D convolutions introduce substantially more weights into the networks that must be learned and increases the computational burden of the training.

Using a combination of 2D+1D convolutional layers improves the efficiency of the training with a small decrease in performance for the fully supervised cases. For example, in [7] the first layers reduce the spectral



(a) RGB rendered version of the scene

(b) Semantic labeling for the scene

[h]

Fig. 2—The AeroRIT scene overlooking Rochester Institute of Technology's university campus. The spatial resolution is 1973 x 3975 pixels and covers the spectral range of 397 nm - 1003 nm in 1 nm steps.

dimension using a 1 x 1 x n kernel (n is the number of bands/channels), and then the spatial ones with a k x k x 1 kernel. There is an assumption of separability that generally holds in practice.

While supervised deep learning requires large labeled datasets for training, annotating a hyper-spectral image often demands professional field exploration, which is unimaginably costly in terms of manpower and time. Therefore there has been interest in applying semi-supervised learning to HSI to reduce the need for labeling large numbers of pixels. A majority of semi-supervised methods for HSI are not neural network based, such as superpixel method [8] and Support Vector Machine methods [9]. Early efforts to apply semi-supervised deep learning methods to HSI include fine-tuning [10] and the use of pseudo labeling [5], while more recent methods have been based on graph neural networks [11, 12].

3.2 Performance Metrics

In literature, we noticed a lack of standardization of measuring performance for semantic segmentation models and the choice of performance metric is especially vital when considering class imbalance (more pixels of one class than another in the dataset). A few of the most common performance metrics are described below. We use a combination of the following metrics for all our results.

- **Overall Accuracy** – Percent of pixels in your image that are classified correctly. Overall Accuracy falls victim to class imbalance and the following metrics are often more appropriate.
- **Average Accuracy** – Average of accuracy for each class. The accuracy for each class is the percent of pixels of that class that are classified correctly. Although average accuracy is more robust against class imbalance, it is not as robust as the following metrics.

- **Mean Intersection-Over-Union (MIoU)** – Average of IoU for each class. Intersection-Over-Union is the area of overlap between the predicted segmentation and the ground truth divided by the area of union between the predicted segmentation and the ground truth. MIoU is very robust against class imbalance, is straightforward to calculate, and extremely effective. This metric is illustrated in Figure 3.

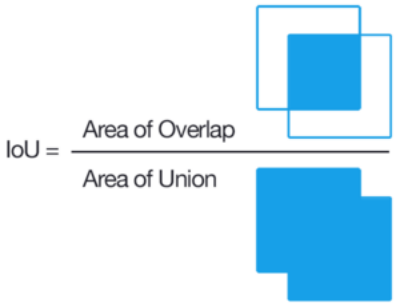


Fig. 3—Mean IoU

- **Mean Dice Score (mDICE)** – Average of Dice scores for each class. The Dice score is 2 * the Area of Overlap divided by the total number of pixels in both images. mDICE is also very robust against class imbalance, but penalizes single instances of bad classification slightly less than MIoU. mDICE can be considered to be between MIoU and average accuracy. This metric is illustrated in Figure 4.

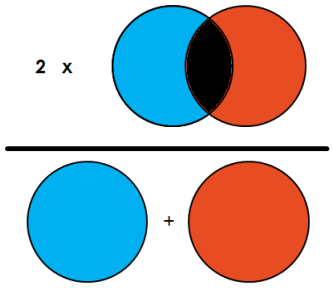


Fig. 4—Mean DICE

3.3 Data Imbalance

Almost every dataset will have an unequal representation of classes based on available data. When one or more of the classes are significantly under-sampled compared to the others, the imbalance can have adverse effects on model performance. An example of this is during training, the model might learn to under-perform in identifying minority classes. Some methods of overcoming this challenge include data-level methods such as undersampling and oversampling, as well as algorithm-level methods like kernel modifications and weighted approaches.

However, we are the first to investigate the possibility that data imbalance can become beneficial to training neural networks. Data imbalance is problematic when it is based on class availability/scarcity. It becomes beneficial when it is based on how difficult it is to learn a class (i.e., as of now it has not been

discussed in the recognition application literature that each classes does not train equally well). Hence, when a subset of classes are similar and hard to differentiate, those classes benefit from additional labeled samples.

When the data is imbalanced, as with the AeroRIT dataset, overall accuracy can be misleading. With the overall accuracy metric the classification errors in the minority classes are treated as less important than that of the majority classes. Thus different evaluation metrics are often required when working with imbalanced classification.

4. METHODS

4.1 Semantic Segmentation

The goal of semantic image segmentation is to label each pixel of an image with a corresponding class of what object the pixel represents While it is possible to analyze the HSI data on a per-pixel basis, semantic segmentation can lead to a more structure-aware approach. By working with patches, spatial and spectral resolution is considered. In this paper, four semantic segmentation approaches are investigated.

1. Rangenekar et al. [1] compare semantic segmentation models to establish a baseline, and the U-Net-m architecture proved to be the best for the AeroRIT dataset.

2. Sellars et al. [8] demonstrate a novel semi-supervised graph-based approach that uses superpixel segmentation for the classification of hyperspectral images.

3. Ouali et al. [13] demonstrate a novel cross-consistency training (CCT) method for semi-supervised semantic segmentation on RGB images.

4. We implemented dense and 1-D convolutional neural networks to provide simple baselines in the limited data regime. In addition, we tested classical ML methods such as Random Forests, Extra Trees, and SVMs for comparison.

Our investigation was built upon recent research advancements in semi-supervised, semantic segmentation and hyperspectral imaging. Supervised learning uses labeled data to define a task (i.e., which classes for recognition). Unsupervised learning attempts to use unlabeled data to learn useful representations for a downstream task. Semi-supervised learning is a combination of the two that utilizes the strengths of each method. Semi-supervised learning uses a small amount of labeled data to define the desired task and like unsupervised learning, semi-supervised learning uses large amounts of unlabeled data to avoid overfitting.

Unlike the fully supervised used by Rangnekar et al. [1], semi-supervised learning combines a small amount of labeled data with a large amount of unlabeled data during training. Semi-supervised learning falls between unsupervised learning and supervised learning. There is a vast amount of literature on semi-supervised, unsupervised, and self-supervised learning, with a number of surveys and books [14–17] available for interested readers.

4.2 U-Net-m Algorithm

Rangenekar et al. [1] found a variation of a U-Net (called U-Net-m), a dense semantic segmentation framework, to perform the best on the AeroRIT dataset so we started with this model as our first approach. After replicating the results in the AeroNet paper [1] using U-net-m, we reduced the number of labeled training samples. We run the U-Net-m framework with a limited number of labeled examples. To do this, we adjust the network from originally working on 64x64 patches, to train at the single center-pixel level. In this way, the network can train on single instances of the 5 class labels (Road, Building, Vegetation, Car, Water). See Section 5.1 for more details.

Although the U-Net-m trains on 64x64 "chips" of the whole AeroRIT image, one chip is NOT a sample. Rather, each pixel is considered a sample. By choosing the same number of training samples per class, this helps alleviate the challenges of class imbalance while training.

We varied the number of training samples per class from 1 to 128 by powers of 2. From a coding approach, this meant changing the custom DataLoader (named AeroCLoader) in helpers/utils.py so that there is an option to limit the number of samples per class. This option was used for the training DataLoader. The validation and testing sets were left untouched in this process.

4.3 Superpixel algorithm

The superpixel algorithm [8] is a state-of-the-art method on hyperspectral imagery that is not based on deep neural networks and we chose this method as one of our baselines for comparison. Sellars et al. proposed a novel method of formulating superpixels from HSI data. A superpixel can be defined as a group of pixels that share common characteristics. According to Sellars et al. [8] superpixel maps such as the one shown in Figure 5 have many desirable properties: they are computationally and representationally efficient, the individual superpixels are perceptually meaningful and as superpixels are the result of an over-segmentation they are very good at conserving image structures.

Sellars et al. [8] developed a graph-based model for semi-supervised learning on HSI images. They leverage spatial continuity for improving classification performance; that is based on the observation that a pixel is likely to be of the same class as its nearest neighbors. They had been using common HSI datasets such as Indian Pines, but our goal was to get the model working on the AeroRIT dataset. Thus, we altered it to run on the AeroRIT dataset.

4.4 Cross-Consistency Training (CCT) Algorithm

Cross-Consistency Training (CCT) Algorithm is a recent semi-supervised method in semantic segmentation for achieving state-of-the-art performance for RGB imagery. Ouali et al. [13] builds upon semi-supervised methods in deep learning, such as pseudo labeling, entropy minimization, and consistency regularization. Pseudo-labeling [18] defines a base model that is trained on labeled data and the model is used to predict labels for unlabeled data. Entropy minimization encourages the model to output confident predictions on unlabeled data. A common underlying assumption in many semi-supervised learning methods is that the classifier's decision boundary should not pass through high-density regions of the marginal data distribution and one way to implicitly achieve entropy minimization is through the use of a "sharpening" function. Consistency regularization utilizes unlabeled data by relying on the assumption that the model should output the same final predictions when fed perturbed versions as on the original image. Ouali et al. implement a

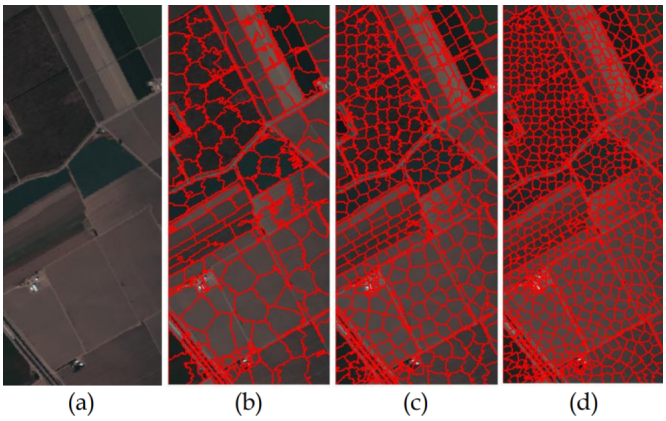


Fig. 5—The Salinas HSI segmented using the HMS algorithm. Figure (a) shows a RGB version of the image and Figures (b)-(d) show the image segmented using 280, 569 and 1034 superpixels respectively. Note that due to the content sensitive nature of the HMS extension, there are a larger number of smaller superpixels in content dense regions.

SSL method in their framework where the invariance of the predictions is enforced over a variety of three different perturbations injected into the encoder's outputs. This is shown in the Cross-Consistency Training framework in Figure 6.

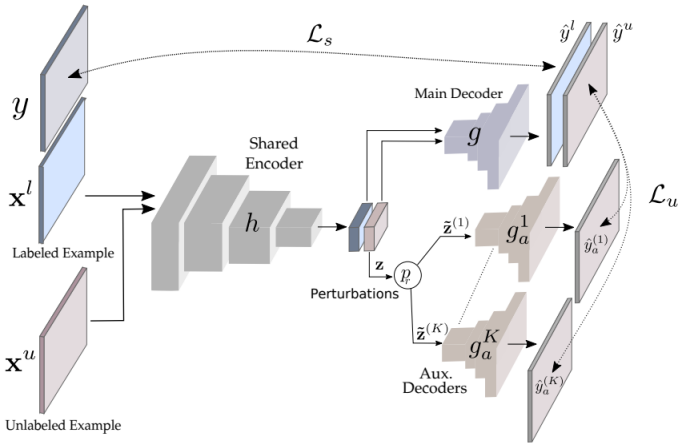


Fig. 6—Illustration of the Cross-Consistency Training approach. Both labeled and unlabeled images are passed through the encoder and main decoder to obtain two main predictions. Various perturbations are applied to z, the output of the encoder. The unsupervised loss is then computed between the outputs of the auxiliary decoders and that of the main decoder.

The CCT method for semi-supervised semantic segmentation was shown to achieve state-of-the-art results for RGB images. We wanted to examine the effectiveness of using CCT for semi-supervised on HSI domain. This method operates under the cluster assumption, in that unlabeled data can be leveraged to enhance decision boundaries in low density regions. This ties into semantic segmentation because low density regions are more apparent within the hidden representations, i.e. encoder outputs, than within the original input data. In the original CCT training process, perturbations are applied to the encoder's output. The three types of perturbations used include feature based, prediction based, and random. Due to time constraints, and because

the original perturbations were designed to work on 2D images, we had to turn off these perturbations in order to run our experiments.

We altered a state-of-the-art semi-supervised learning for semantic segmentation on RGB images to work on HSI samples; that is the Cross-Consistency Training (CCT) technique. This involved editing the model layers to ingest the AeroRIT 51 channel bands instead of the standard three channels that are inherent to RGB images.

4.5 Dense and Convolutional Neural Network with Semi-Supervised Learning

For the sake of including simple baselines, we implemented a dense network and a one-dimensional CNN. In dense networks, each of the neurons in a network layer receives input from all the neurons present in the previous layer. Dense networks are connected in such a way that it trains on all the combinations of the features from the previous layer. On the other hand, the CNN works as a set of filters and relies on consistent features within relatively small fields the size of the kernel. These filters or convolutions have shared weights, and can perform more efficiently with a smaller number of coefficients than fully connected dense networks. Diagrams for our two networks can be seen in Figure 7.

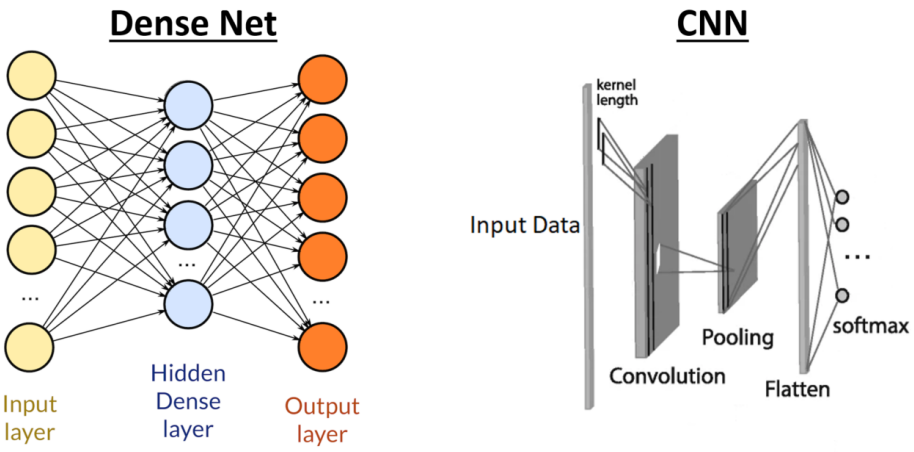


Fig. 7—Illustration of a simple dense network (left) and convolutional neural network (right).

Further, we perform testing of the most appropriate kernel size for our CNN. In our CNN the stride of the convolution is the length of the kernel size. This kernel size can be exploited to capture relationships between adjacent values in the input. This tuning of the kernel size is to determine if there are correlations among neighboring bands in the input data. In addition, other classical machine learning approaches, such as Random Forest, Extra Trees, and support vector machines (SVMs) were tested on the same AeroRIT data as additional baselines.

Deep learning has demonstrated its superior performance in many applications. One consideration, however, is that deep learning networks work best when handling large datasets. In this case we are purposely trying to train on a small amount of training data. As another test to our CNN, we exploit semi-supervised approaches. The approaches we tried were derived from the approaches used in the CCT algorithm. These

perturbations include feature-based (injecting random noise) and random-based (applying spatial dropout) perturbations to the hidden encoder's outputs.

5. RESULTS

5.1 U-Net-m Algorithm on AeroRIT Dataset

The original U-Net-M model input data was based on spatial location – left side of image as train, middle as validation, right as test. The benefit of splitting the data with this "spatial" technique is that it generates a training, validation and test set that can be used to benchmark performance without the need for cross-validation splits. However, this raises some problems since the class balance is not equal across the 3 regions. For example, there are no pixels labeled as "water" in the middle (validation) or right (test) regions. So, we adjusted the U-Net-m to instead split train/validation/test randomly across the whole image. We call this our "random" sampling approach. Since "chips" have a 50% overlap, we ensured that test and validation "chips" did not overlap with each other or with train "chips". From a coding approach, this meant altering sampling_data.py (new file named sampling_data_random.py) to sample randomly across the whole image rather than spatially before running train.py.

We compare "random" and "spatial" sampling approaches with the performance of the U-Net-m configuration as illustrated in Figure 8. There are three main pheneomena observed. The first observation is that when there are less than 32 samples per class, the overall accuracy is irregular. We believe this is due to the fact that HSI samples are information dense and complex. As expected, we observe that overall there is a trend of positive correlation between the samples per class and the network performance.

It is noteworthy from a comparison of results with these four metrics, that overall accuracy is a misleading metric. Even with few samples per class, one obtains overall accuracies similar to having 128 samples per class. We investigated this and found that the most common class is vegetation and its somewhat unique spectra makes it easy to distinguish from the other classes. In addition, if the model learned to mostly guess vegetation, it will obtain a relatively high performance. For the rest of this paper, we report the overall accuracy but discount its value.

Lastly, we note that the "random" sampling approach results in higher average accuracy, mean IOU, and mean DICE. We believe that when the samples are selected randomly across the whole image you can achieve a larger variety of samples and improves the generalization performance.

5.2 Superpixel algorithm on AeroRIT Dataset

Unlike the U-Net-m, the superpixel algorithm requires one full image rather than many smaller "chips" since the algorithm requires spatial information to be conserved. The ground truth labels for AeroRIT also had to be converted from RGB to 1-dimensional numeric labels. We preprocessed the AeroRIT HSI data to match the scale of the Indian Pines dataset. This meant dividing all HSI values by 10 and adding 950. (spectral_data = (spectral_data/10)+950)

Similar to how we varied the number of labeled training samples per class for the U-Net-m, we varied the number of labeled training samples for the superpixel algorithm. However, since the superpixel algorithm is semi-supervised, the unlabeled samples still aid in the training process. Considering scenarios where we

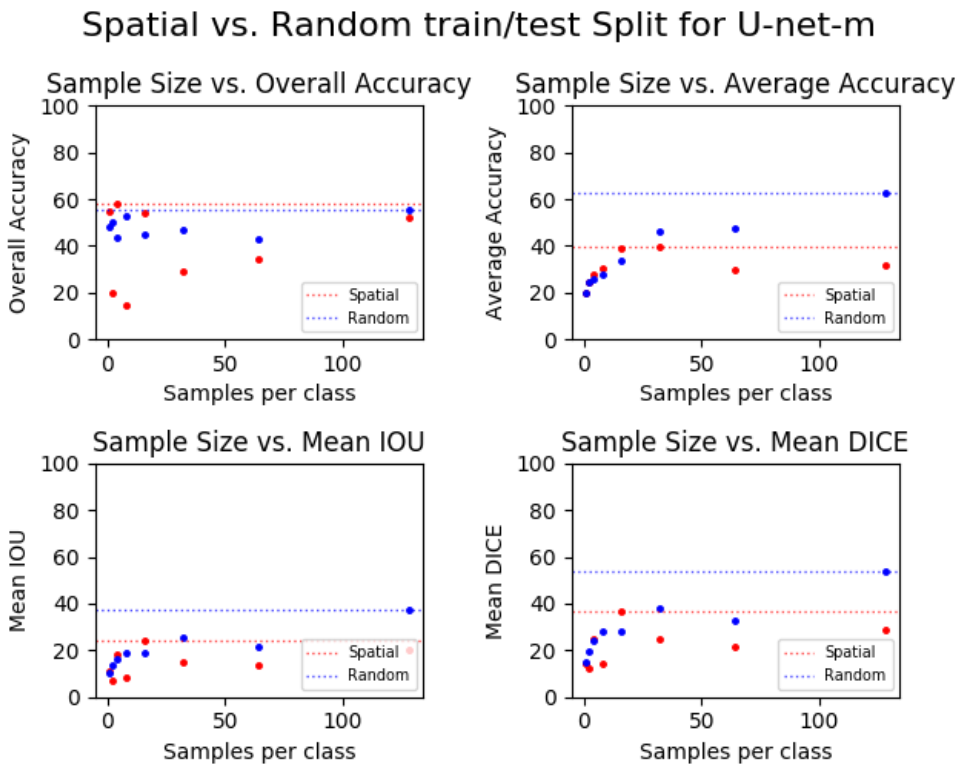


Fig. 8—Performance comparison for U-net-m with spatial vs. random test/train split with varying number of labeled training samples

only have 1 labeled sample per class (and the rest unlabeled), this should give the superpixel algorithm an advantage over the U-Net-m.

The U-Net-m correctly ignores pixels labeled as ”unspecified” since they often do not contain valid data. For the superpixel algorithm to correctly work on the AeroRIT dataset, we made sure that it also ignored the ”unspecified” pixels.

Our results, shown in Figure 9, depict that the Superpixel algorithm performs slightly worse than the U-Net-M algorithm except in terms of average accuracy. The average accuracy is obtained from both the minority and majority classes and is less likely to take advantage of the distribution of the majority classes of vegetation, roads, and buildings. Accordingly, the Superpixel algorithm performs slightly better on the minority classes of water and cars. However, it is noteworthy that the Superpixel algorithm's performance on AeroRIT is significantly worse than the results reported in the paper [8].

5.3 Cross-Consistency Training algorithm on AeroRIT Dataset

The CCT algorithm was originally presented as a novel cross-consistency based semi-supervised approach for semantic segmentation with RGB images. We wanted to examine the effectiveness of using CCT for semi-supervised on HSI domain. The original algorithm operates under the cluster assumption, in that unlabeled data can be leveraged to enhance decision boundaries in low density regions. Another important aspect in

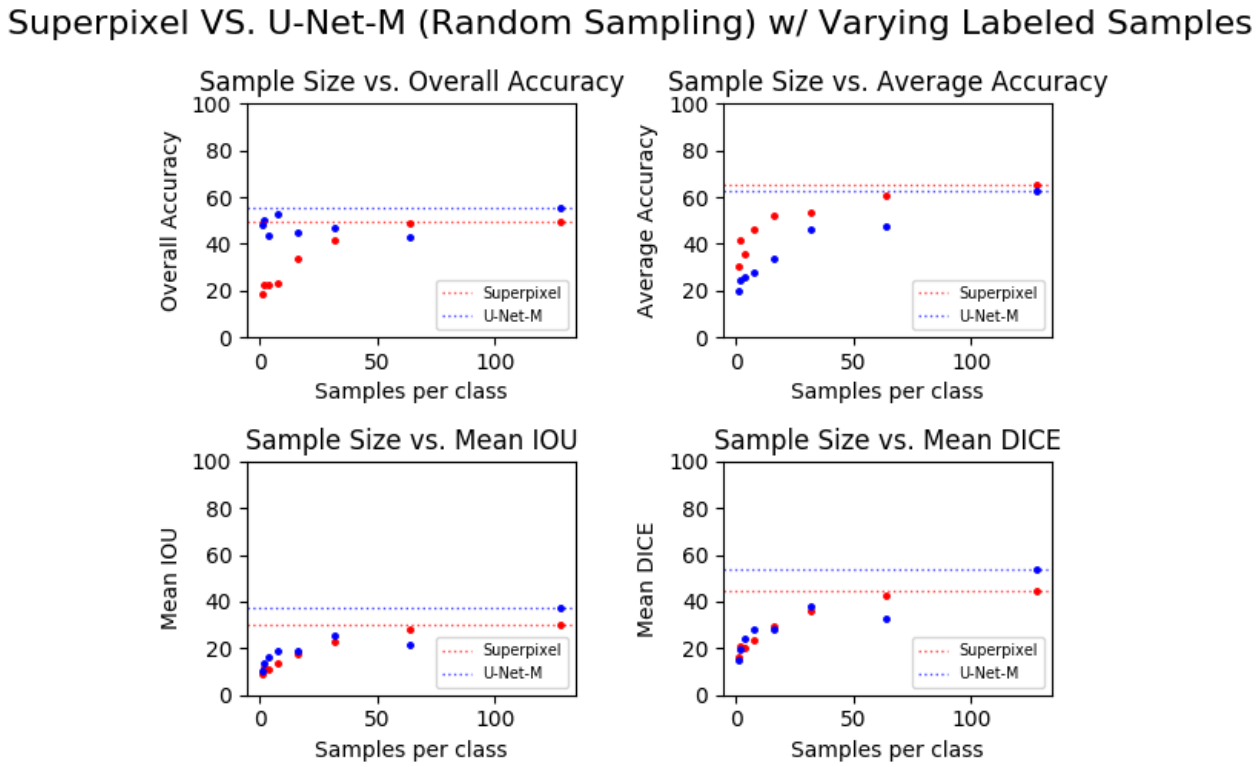


Fig. 9—Performance comparison for Superpixel and U-net-m (w/ random test/train split) with varying number of labeled training samples

the CCT training is the perturbations applied to the encoder's output. The three types of perturbations used include feature based, prediction based, and random.

We ran into challenges in altering the CCT codes to run on our AeroRIT dataset. We had to alter the convolutional network layers in the model to handle the additional dimensionality of the HSI data. We also had to remove any instances of data augmentation within the model because the standard methods of augmentation on RGB images aren't suitable for HSI data.

Another hurdle we encountered is that the original CCT code is not written to accept single pixel data. When we tried to set input patch data surrounding pixels of interest as 'unspecified' the model would consider the data as significantly unbalanced. This led to the model learning to only classify pixels as 'unspecified'.

We settled on inputting a small number of input labeled 64x64 patches, N. We varied the number of input patches from 10, and then 100 to 500 in increments of 100. We used the same patch sizes as in the AeroRIT approach with size 64x64. The input train and test distribution of pixels, for each number of patches (N) are shown in Tables 1 and 2. In each semi-supervised case the number of unlabeled data equals the same number of input training patches.

The results of applying CCT to the AeroRIT dataset are shown in Figure 10 and in Tables 3 and 4. In summary, the addition of unlabeled data through the semi-supervised approach for the CCT method did not offer reliable improvement in performance.

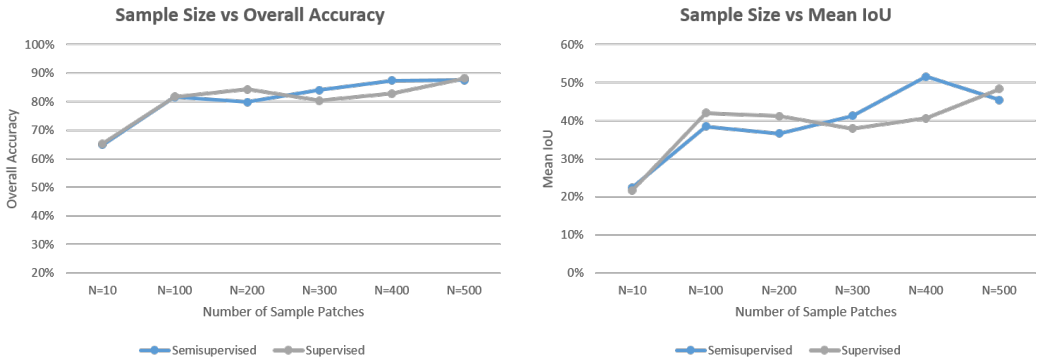


Fig. 10—Performance comparison for Cross-Consistency Training method for both the supervised and semi-supervised case for various numbers of input sample patches.

	Buildings	Vegetation	Roads	Water	Cars	Unspecified
N=10	3383	17524	13602	3994	526	1931
N=100	71206	139566	108989	21477	5171	63191
N=200	139893	302382	209865	33740	14680	118640
N=300	198055	456458	319879	57357	21849	175202
N=400	230891	659308	419341	68965	37493	222402
N=500	290521	831978	522850	88088	37146	277417

Table 1—Distribution of input sample pixels with varying number of labeled training sample patches, N. Each patch is of size 64x64 pixels.

	Buildings	Vegetation	Roads	Water	Cars	Unspecified
N=3127	1392980	6058966	3079434	22872	168472	2085468

Table 2—Distribution of test sample pixels in the 3127 test patches. Each patch is of size 64x64 pixels.

5.4 Dense and Convolutional Network with Semi-Supervised Learning

In Figure 11 we compare the results of using a 1-D CNN and a dense network trained on only supervised training data, as well as a CNN with one auxiliary decoder trained by semi-supervised learning. The results in Figure 11 show that there is an improvement in both overall accuracy and MIOU by using semi-supervised approaches.

Results of the role for kernel size and stride with the CNN are shown in Figure 12. In summary, the larger kernels and strides are universally better than smaller for the AeroRIT dataset. On the other hand, Figure 13 shows that increasing the number of auxiliary decoders (k) does not increase the network performance. We found both of these results to be surprising and are still attempting to understand the underlying cause.

Further, we investigated classical machine learners such as Random Forests, Extra Trees, and SVMs as alternatives to our small dense network. The overall accuracy and MIOU of these various approaches are shown in Figure 14. These results show that for our limited amount of input data, the classical machine

	10	100	200	300	400	500
Semi-supervised	64.9%	81.6%	79.9%	84.1%	87.4%	87.6%
Supervised	65.3%	81.7%	84.3%	80.4%	82.9%	88.1%

Table 3—Comparison of overall accuracy performance for Cross-Consistency Training method for both the supervised and semi-supervised case.

	10	100	200	300	400	500
Semi-supervised	18.7%	32.1%	30.6%	34.4%	43.0%	37.9%
Supervised	18.1%	35.1%	34.3%	31.6%	33.9%	40.3%

Table 4—Comparison of MIoU performance for Cross-Consistency Training method for both the supervised and semi-supervised case for various numbers of input sample patches.

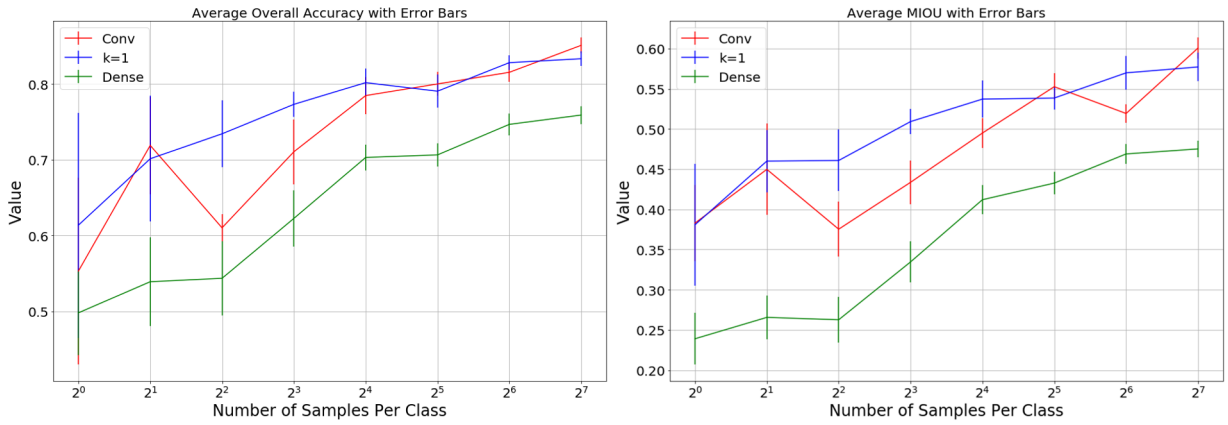


Fig. 11—Comparison of network performance for a fully supervised CNN, a fully supervised dense network, and a semi-supervised CNN with one auxiliary decoders (k).

learners outperform our dense network. On the other hand, the performance of the CNN as shown in Figure 11 exceeds that of the classical ML methods.

5.5 Increasing training data imbalance for hard classes

Figure 15 contains a plot of the means and standard deviations of the spectra for each of the five classes in AeroRIT: water, vegetation, roads, buildings, and cars. This shows the statistics for each of these classes. The large standard deviations reflect that there are huge differences between spectra within its own class. It is also apparent that the water and vegetation classes are unique, whereas the other three classes are fairly similar to each other. The implication of this is that the class accuracies for water and vegetation are likely to be higher than the class accuracies for roads, buildings and cars. This in fact was observed across all of our experiments.

Figure 16 show five randomly chosen individual spectra for each of the two classes buildings and cars. It is clear in this Figure that the differences between a inter-class spectra can be less than the intra-class spectra. This demonstrates the difficulty of image semantic segmentation for HSI data.

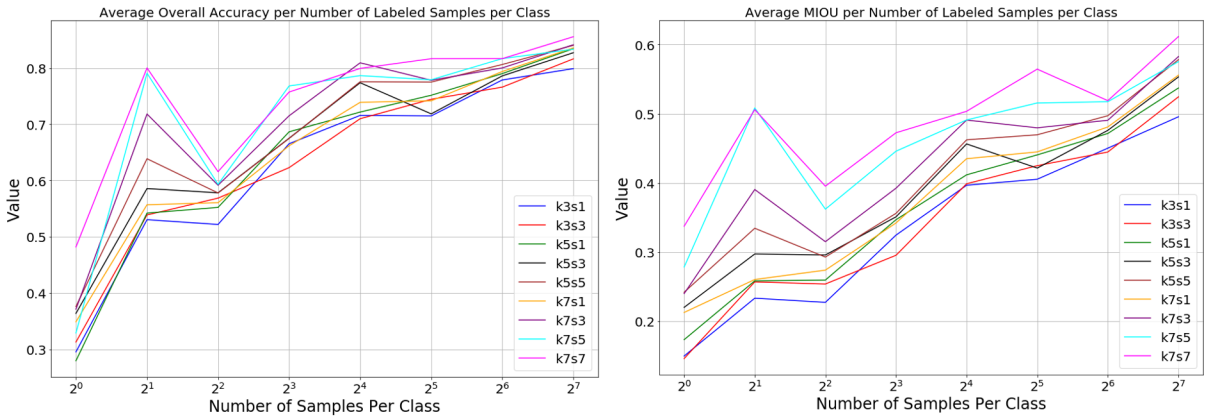


Fig. 12—Performance comparison of over kernel size and stride.

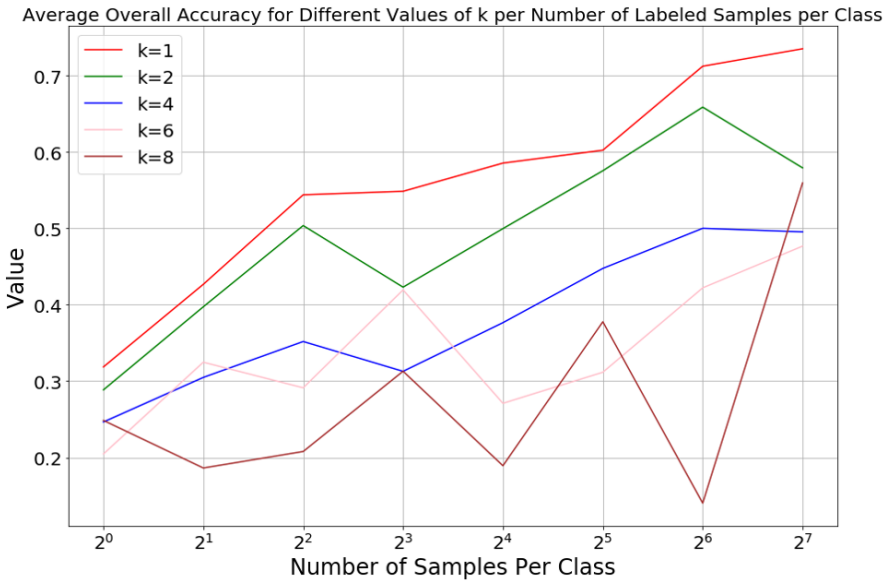


Fig. 13—Comparison of network performance for different number of auxiliary decoders (k).

This lead to a novel idea that we tested; what if we utilize class imbalance to improve the performance of the hard to distinguish classes. In other words, once we identify the hard classes, what if the labeled training data included significantly more of those classes than of the easier classes. The intuitive concept behind this idea is to flip class imbalance, which is commonly considered a problem, from being based on availability of classes in images to being based on difficulty of the class for the classification task.

Table 5 shows a comparison of the class accuracies for a balanced number of labeled samples per class (ratio=1) versus having a greater number of labeled samples for roads, buildings, and cars. Specifically, for a ratio of 1, every class had 128 labeled training samples, while for a ratio of 4, the water and vegetation classes had 128 labeled training samples but the roads, buildings, and car classes each had 512 labeled training samples. As is apparent from these results, the class accuracies for the hard classes improve while

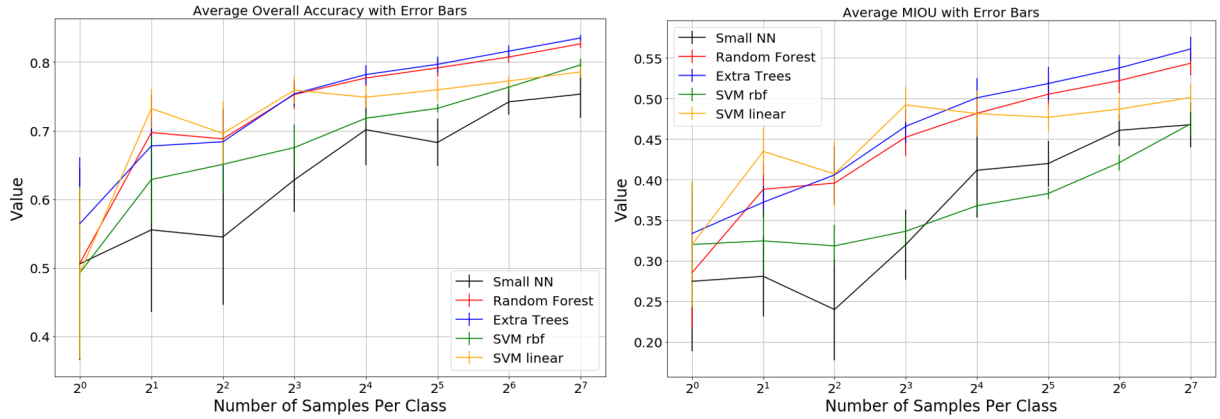


Fig. 14—Comparison of dense network, and other classical machine learning approaches.

Ratio	Water	Vegetation	roads	buildings	cars
1	96.0±1	94.9±1	78.1±6	62.7±7	60.5±7
4	93.3±2	92.1±2	81.7±4	70.0±7	68.1±5

Table 5—Comparison of class accuracies for balanced number of labeled samples per class (ratio=1) versus having a greater number of labeled samples for roads, buildings, and cars. The class accuracies for the hard classes improve while the accuracies for easy classes decrease slightly.

the accuracies for easy classes decrease slightly. Hence, the results do indicate that there is a some benefit to this method but with a performance tradeoff for the easier classes.

6. CONCLUSIONS

In addition to limiting the number of labeled samples, working with hyperspectral data has numerous challenges, including per pixel classification, imbalanced data, and inconsistent performance metrics. Even with all these challenges, our team made significant progress in being able to train the hyperspectral dataset AeroRIT with few labeled examples. Our work compared four algorithms: fully supervised learning with limited labeled examples, the state-of-the-art in HSI semi-supervised learning (i.e., CCT), shallow networks, and conversion of the state-of-the-art in semi-supervised learning for semantic segmentation to work on the AeroRIT dataset.

Our conclusions from this study are:

- Evaluation metrics for HSI performance are not standardized and can significantly impact the performance and conclusions reported in the literature
- The results previously reported in the semi-supervised literature for smaller, older HSI datasets did not generalize to the AeroRIT dataset.
- Semi-supervised learning did not provide a significant improvement in performance relative to supervised learning in the limited labeled pixel regime

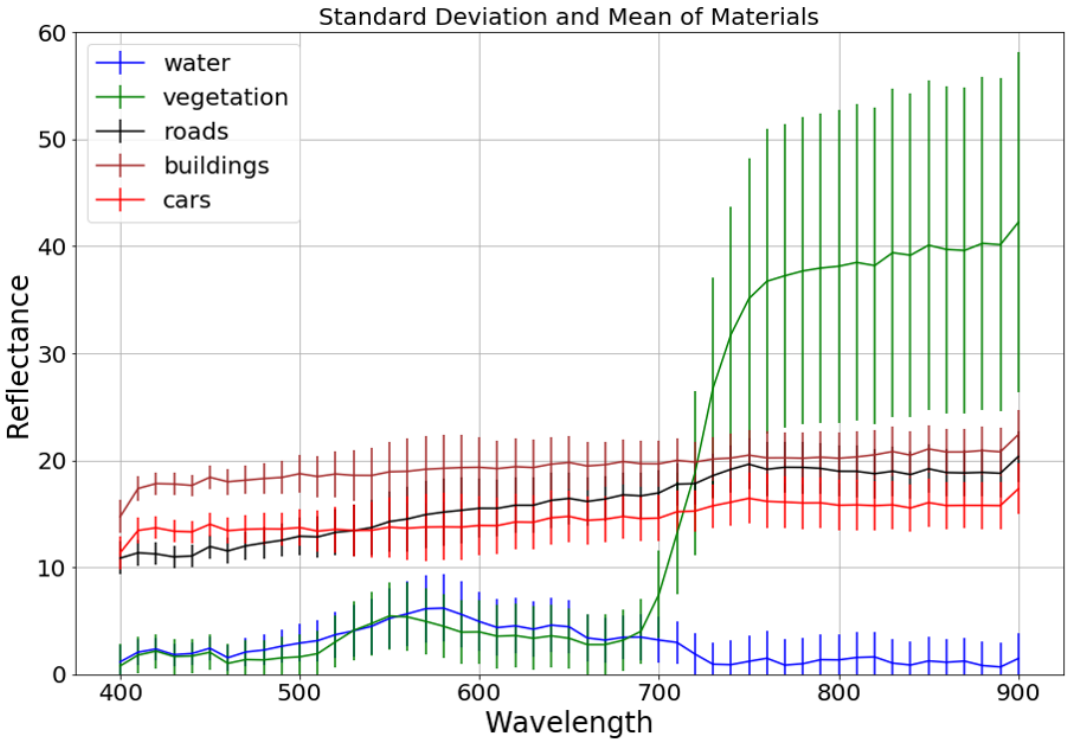


Fig. 15—Mean and standard deviation of each class spectra over the entire AeroRIT dataset.

We conclude that improving HSI performance in the limited labeled pixel regime remains an open problem.

ACKNOWLEDGMENTS

The authors wish to acknowledge that this work was supported by the US Naval Research Laboratory's NISE program.

REFERENCES

1. A. Rangnekar, N. Mokashi, E. J. Ientilucci, C. Kanan, and M. J. Hoffman, "Aerorit: A new scene for hyperspectral image analysis," *IEEE Transactions on Geoscience and Remote Sensing* (2020).

2. G. E. Hinton and R. S. Zemel, "Autoencoders, minimum description length and Helmholtz free energy," Proceedings of the Advances in neural information processing systems, 1994, pp. 3–10.

3. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," Proceedings of the International Conference on Medical image computing and computer-assisted intervention (Springer), 2015, pp. 234–241.

4. W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *Journal of Sensors* **2015** (2015).

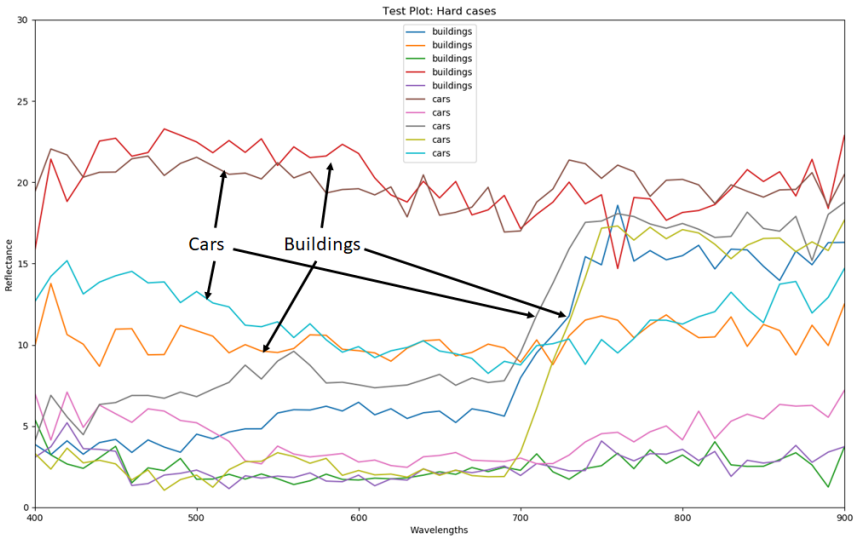


Fig. 16—A few random selections of individual spectra for the "hard" buildings versus cars classes. Illustrates that some spectra for these two classes are very similar.

5. H. Wu and S. Prasad, "Semi-supervised deep learning using pseudo labels for hyperspectral image classification," *IEEE Transactions on Image Processing* **27**(3), 1259–1270 (2017).

6. Y. Li, H. Zhang, and Q. Shen, "Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sensing* **9**(1), 67 (2017).

7. A. B. Hamida, A. Benoit, P. Lambert, and C. Ben-Amar, "Deep learning approach for remote sensing image analysis, 2016.

8. P. Sellars, A. I. Aviles-Rivero, and C. B. Schönlieb, "Superpixel contracted graph-based learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing* **58**(6), 4180–4193 (2020).

9. L. Bruzzone, M. Chi, and M. Marconcini, "A novel transductive SVM for semisupervised classification of remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing* **44**(11), 3363–3373 (2006).

10. E. Protopapadakis, A. Doulamis, N. Doulamis, and E. Maltezos, "Semi-Supervised Fine-Tuning for Deep Learning Models in Remote Sensing Applications," *arXiv preprint arXiv:2006.00345* (2020).

11. Z. Zhang, "Semi-supervised Hyperspectral Image Classification Algorithm based on Graph Embedding and Discriminative Spatial Information," *Microprocessors and Microsystems* p. 103070 (2020).

12. H. Zeng, Q. Liu, M. Zhang, X. Han, and Y. Wang, "Semi-supervised Hyperspectral Image Classification with Graph Clustering Convolutional Networks," *arXiv preprint arXiv:2012.10932* (2020).

13. Y. Ouali, C. Hudelot, and M. Tami, "Semi-Supervised Semantic Segmentation with Cross-Consistency Training," Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 12674–12684.

14. X. J. Zhu, "Semi-supervised learning literature survey," University of Wisconsin-Madison Department of Computer Sciences, 2005.

15. J. E. Van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Machine Learning* **109**(2), 373–440 (2020).

16. O. Chapelle, B. Scholkopf, and A. Zien, "Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]," *IEEE Transactions on Neural Networks* **20**(3), 542–542 (2009).

17. X. Zhu and A. B. Goldberg, "Introduction to semi-supervised learning," *Synthesis lectures on artificial intelligence and machine learning* **3**(1), 1–130 (2009).

18. D. H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," Proceedings of the Workshop on challenges in representation learning, ICML, volume 3, 2013.