Semi-supervised learning for hyperspectral landmine detection

Victor Sineglazov[1,*,†], Volodymyr Lytvynenko[2,†], Kyrylo Lesohorskyi[3,†]

[1] *Institute of Cybernetics of the National Academy of Sciences of Ukraine, Kyiv, Ukraine*
[2] *Kherson National Technical University, Beryslavs'ke Hwy, 24, Kherson, Kherson Oblast, 73008, Ukraine*
[3] *National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine*

Abstract

This paper concerns the problem of hyperspectral imagery multi-class classification. A hybrid three step iterative method is proposed. The method consists of pre-processing, training and proxy-labeling steps. During the pre-processing step the data is normalized and filtered to minimize the noise and smooth the data. The second step involves training a convolutional neural network to train a discriminator. Third step utilizes the discriminator from the second step to label high-confidence samples from the unlabeled pool. An overview of each component is given. The efficiency of dimensionality reduction is evaluated. Results indicate that dimensionality reduction step both speeds up the subsequent learning and improves the efficiency of semi-supervised learning.

Keywords

hyperspectral imaging, semi-supervised learning, multi-class classification, dimensionality reduction, proxy-labeling

1. Introduction

Hyperspectral imaging is an advanced remote sensing method that collects detailed information about the earth's surface. It projects light waves in hundreds of narrow spectral ranges, providing a rich dataset for identifying and characterizing materials, assessing biophysical parameters and monitoring environmental conditions. A number of studies are devoted to the use of supervised learning for hyperspectral classification. The most common task for hyperspectral image processing is multi-class classification. This is a simple, yet effective way to model connections between a discrete target variable and a set of input data. There are a number of problems associated with the use of data for hyperspectral classification:

- high dimensionality of hyperspectral data - hyperspectral data consists of hundreds of light spectrums, which complicates the development of accurate models, which leads to the curse of dimensionality [1] - the accuracy of machine learning models can be reduced by increased data dimensions;

[1*] Corresponding author.
[†] These authors contributed equally.
svm@nau.edu.ua (V. Sineglazov); immun56@gmail.com (V. Lytvynenko); lesogor.kirill@gmail.com (K. Lesohorskyi)
0000-0002-3297-9060 (V. Sineglazov); 0000-0002-1536-5542 (V. Lytvynenko); 0000-0003-2773-7398 (K. Lesohorskyi)

- availability of labeled data - supervised learning algorithms require a large volume of labeled data to be effective. However, collecting labeled data for hyperspectral regression can be difficult and expensive;

In this work, a hybrid approach is used to alleviate these issues. Deep neural networks allow for the effective embedding of high-dimensional features into low-dimensional embeddings, but they require even more labeled samples to learn the low-dimensional representation. To partially solve the curse of dimensionality, the number of hyperspectral bands is reduced during the pre-processing stage to decrease the number of parameters of our model and partially assist in solving the curse of dimensionality.
Semi-supervised learning, on the other hand, allows utilization of a combination of labeled and unlabeled data to train the model, achieving higher accuracy than a normal supervised learning.

2. Problem Statement

When considered under the semi-supervised framework, hyperspectral image classification in it's core is similar to other multi-class classification tasks. More formally, we start by introducing a labeled dataset $L=\{(x_0, y_0),\ldots,(x_n, y_n)\}$, where $x \in X$, $y \in Y$, X is the input space, Y is the label space, n is the size of labeled dataset. It is complemented by unlabeled dataset $U=\{x_{n+1},\ldots,x_{n+m}\}$, $x \in X$, m is the size of the unlabeled dataset. A combined dataset $D=U \cup L$ is used to train model parameters θ of an approximator function $f(x,\theta)\approx y$'. The approximator function is learnt to minimize a loss function and make predicted labels as close to the groundtruth labels as possible.
Training is an iterative process $\theta_i, L_i, U_i = T(\theta_{i-1}, L_{i-1}, U_{i-1})$, where i is the training iteration (or epoch), T is a training step. Depending on the internal implementation of the algorithm, labeled and unlabeled data sets can be updated between the training iterations.
Hyperspectral imagery has a lot in common with classic computer vision tasks – the data is represented as a 3-dimensional cube of data $x_{H\times W\times C} \in X$, where x is the data sample, H is the image height, W is the image width, C is the number of channels. However, unlike conventional image recognition, however, the number of channels C is usually in the range of 100-500, instead of the normal 3-4 channels. This leads to increase in the input size, e.g. for an input with H = 48, W = 48 the normal configuration (C = 3) would result in 6912 features. Hyperspectral images of the same resolution with C = 100 would have 230400 features, an increase of ~ 33 times, which can be even larger in practice.
Semi-supervised framework with deep neural networks is able to address the issues of sample size and high dimensionality, however the problem of noise is still present. As such, hyperspectral images must be pre-processed to reduce noise and remove artifacts. The pre-processing is modeled as a function $P(x_{H\times W\times C}) = x'_{H\times W\times C}, x\,X, x'\,X$, where x is the original input, x' is pre-processed input.

3. Related Works

A general framework for hyperspectral multi-class classification consists of two stages: data pre-processing and discrimination. Pre-processing steps may vary depending on the

discriminator used. Recent research uses the variety of discriminator models, such as SVM, matched filters, and neural networks. Among these methods, neural networks is the most robust as it allows to efficiently learn the low-dimensionality embeddings from the high-dimensionality dataset.

3.1. Approaches to mine detection

Despite widely acknowledged danger and economic damage mines are known to posses [2], landmine discovery and removal operations are still widely performed with the same tools that were developed during WW2, namely a magnetic detector and a probe. This leads to increase risk to the operators performing a survey. This lead to an increased interest in leveraging remote sensing technologies to detect mines and create precise maps of the minefields.

As such, a variety of techniques were considered for remote probing. Hyperspectral imagery[3] is an emergent technology that allows to capture a broader band of light than normal cameras do, allowing for reach features to be extracted from the data. One of the first projects to study mine detection using infrared wavelengths was conducted at Defence Research & Development Canada (DRDC). DRDC began its research in support of the Canadian Army on mine and unexploded ordnance detection in 1978 and in collaboration with Itres Research on hyperspectral imaging for mine detection in 1989. The algorithms developed during this project can be applied to pre-processed images from hyperspectral imagers. An early project proposed a hierarchical image processing algorithm to detect a sparsely distributed bright region a few pixels wide in a monochromatic image [4]. Fusion of visible and SWIR bands can provide better detection results. Basic fusion of two spectral bands provides acceptable segmentation of objects from the background, regardless of the illumination conditions. In other words, choosing a set of two or three spectral bands from an image has been shown to be as effective in differentiating artificial objects from the background as using all spectral bands simultaneously [5]. Such fusion has the potential to detect mine-like objects in an image using an integrated camera with visible and SWIR sensors and more sophisticated and specialized detection algorithms. [6] describes a Defense Advanced Research Projects Agency (DARPA)-sponsored experiment to test the feasibility of detecting buried mines using midwave infrared (MWIR) (3 to 5 μm) and longwave infrared (LWIR) (8 to 12 μm) hyperspectral bands. The project focuses on detecting surface disturbances caused by buried mines. . Previous experiments have shown the ability of VNIR and SWIR imagers to detect surface disturbances [7], [8]. However, the problem was the high rate of false alarms caused by surrounding vegetation and rocks. According to the authors, the main rationale for detecting buried mines using spectral properties is that the surface properties are somewhat different from the properties of the subsurface soil. The impact of soil on the surface changes some of its physical and chemical properties.

More recent approaches leverage artificial neural networks to detect landmines more reliably. In India, researchers proposed a hierarchical algorithm for mine detection using infrared images, which consists of pre-processing (contrast enhancement - filtering - smoothing), segmentation, feature extraction, and ANN-based classification [9]. The authors tested the algorithm on surface mines in two soil types: black cotton and sand. During pre-processing, the image is converted to gray color. The two most important preprocessing steps are contrast enhancement and noise removal. During the tests, the authors used a small NN with 1 hidden layer and 4 neurons. The results obtained on a simple dataset are good, but it is not expected that the algorithm will work well on another field or soil type, since the data used in

the training stage is not complete enough. In 2015, TELOPS, a Canadian research company specializing in infrared and hyperspectral imaging, demonstrated the feasibility of detecting buried objects using an airborne LWIR hyperspectral imager [10]. From the aircraft, they obtained thermal hyperspectral images of areas that contain previously buried artificial objects. They found that the disturbed soil directly above the buried target is warmer than the undisturbed soil area next to it. Comparing the emissivity data obtained by decoupling temperature and emissivity, buried targets are displayed as part of the hottest ground region within the scene, but additional classification or information is needed to distinguish buried objects from other naturally hot regions.

It is worth noting that most of the works reviewed pay much attention to the quality of electro-optical sensors and pre-processing, but the classifier itself is quite simple. The use of modern artificial intelligence technologies, namely convolutional neural networks (CNN), will have a positive effect on both the overall accuracy and specificity of the classifier. Another common problem is the limited training dataset. Semi-supervised learning methods are used to solve this problem.

3.2. Pre-processing

Data preprocessing is an important step in the data analysis process to ensure its accuracy, reliability, and readiness for further analysis. In this work we use a 4-stage pre-processing framework to minimize the impact of the noise and decrease the dimensionality of the input.

Normalization is the basic technique [11] aimed at adjusting the flow of changes in lightning, atmospheric conditions and other factors on the spectral data. The first step is to align the values of different bands as their reflectance values can be highly different. This is achieved by applying a min-max[12] normalization. Min-max normalization is a common technique to normalize multi-band images, be it hyperspectral or simple RGB and is defined as:

$$x'_{i,j} = \frac{x_{i,j} - \min\limits_{x_{H,W}} x}{\max\limits_{x_{H,W}} x - \min\limits_{x_{H,W}} x}, \tag{1}$$

where $x'_{i,j}$ is the normalized pixel value, $x_{i,j}$ is the original pixel value, $\max\limits_{x_{H,W}} x$ is the maximum intensity value and $\min\limits_{x_{H,W}} x$ is the minimal intensity value within the channel. Another common normalization technique to stabilize the training is spectral normalization [13]. Although it is a powerful tool, it's primary use case is the stabilization of generative models[14], and as such it is ill-suited for discriminator training. The removal of noise may result in the influx of unwanted signals and artifacts that may interfere with the results of the analysis[15]. A variety of filters, such as a median filter [16] for removing impulse noise:

$$x'_i = median\left[x_{i-n}, x_{i-n+1}, \ldots, x_i, \ldots, x_{i+n-1}, x_{i+n}\right], \tag{2}$$

where x' is filtered signal, x is the original signal, median is the median operation. Median filters are linear, and often are defined via kernels in case of multidimensional data. A Gaussian filter [8] is used for smoothing data to decrease the interference (in this work we utilize a two-dimensional version of the filter):

$$G(i,j) = \frac{e^{-(i^2+j^2)/2\sigma^2}}{2\pi\sigma^2}, \tag{3}$$

where *i,j* are dimensions along which the filtering is performed, σ is the standard deviation of the distribution. This filter is also used only in one band to decrease the inference.

Depending on the approach, the next step is often skipped, however it can have a great impact on the discriminator's performance, as it is shown that higher dimensionality leads to rapid decrease in discriminator's accuracy and exponential increase in model's parameters [17]. As such, dimensionality techniques are often applied, especially for higher-dimensional data. Some of the bands may not be important for the analysis, and as such can be discarded to decrease the dimensionality. Principal component analysis (PCA) is a common technique [18], used to decrease the dimensionality of the data and identify the main components, which allows for the largest portion of information to be reduced and the influx of noise.

In this work we explore a class of problems that can be classified based on narrow, independent bands of data, as such linear methods are preferable due to their specificity. In these scenarios, linear dimensionality reduction methods are preferable as they are simpler and have comparable effect as nonlinear methods. However, for certain types of problems, it is beneficial to use non-linear dimensionality reduction methods, such as, Isomap[19], that perform better when noise forms non-linear cross-band structures.

3.3. Discriminator

Multiple types of discriminators are considered applied to the hyperspectral data classification. Several methods were used historically to process hyperspectral images, however neural networks are considered the state-of-the art approach. They are able to efficiently learn low-dimensional embeddings for hyperspectral data. Convolutional neural networks (CNN) are the most common approach to processing high-dimensional data, as convolutional layers are the most effective approach to reducing dimensions while minimizing the number of parameters. There are two types of convolutional layers - 3d and 2d convolutions which operate on the same principle:

A discrete convolution on a finite set of values is determined using the formula

$$(f*g)[n]=\sum_{m=-M}^{M} f[n-m]g[m], \tag{4}$$

where *f* and *g* are some discrete functions, *M* is a matrix size, *n*, *m* are matrice's values. If *I* (image) and *K* (kernel) are matrices, the two-dimensional convolution is determined using the formula

$$(I*K)[m,n]=\sum_{i=0}^{k_1}\sum_{j=0}^{k_2} I[m+i,n+j]K[i,j], \tag{5}$$

where *m* and *n* are the current position of the upper left corner of the kernel relative to the upper left corner of the image; k_1 is the kernel's height; k_2 is the kernel's width.

The three-dimensional convolution is defined similarly to the two-dimensional convolution with the addition of an additional dimension for the image, kernel, and sum in formula (4). Conv2D and Conv3D convolutional layers are defined using the formula:

$$out(N_i,C_{outj})=bias(C_{outj})+\sum_{k=0}^{C_{in}-1} weight(C_{outj},k)*input(N_i,k), \tag{6}$$

where $input(N_i, k)$ is the k-th channel of the i-th training sample of size H, W for Conv2D and H, W, D for Conv3D; W – image width, H – image height, D – is the image depth (number of channels); $C_{¿}$– number of input channels; N is the number of training samples in the batch; $weight(C_{outj}, k)$ - convolution weights (kernel) between the k-th channel of the input and C_{out}th channel at the output of the layer; $bias(C_{outj})$ is convolution shift for the C_{outj}th channel's at the output of the layer, C_{out} is the number of channels at the output of the layer.

However, apart from using the correct layers, it is also important to pick the appropriate architecture. ResNet [20] is state-of-the art architecture that utilizes classical CNN architecture with residual connections, which assist in gradient propagation along the network. ResNet is arranged as a series of blocks called bottlenecks. For hyperspectral data processing, we utilize a 3D version of ResNet [21], and introduce a more detailed architecture in section 4.

3.4. Loss function

Loss functions play a key role in training deep neural networks because they determine exactly how the model evaluates its performance and adjusts its parameters.The loss function is a key component in the process of developing machine learning models, as it measures how well the model matches the given data. In this work, we use a hybrid method based on semi-supervised learning, however we use a classical proxy labeling approach, and as such our loss does not have a semi-supervised term. Categorical cross-entropy loss is commonly used in training multi-class classifiers [22]:

$$CE = -\sum_{i=0}^{C} t_i \log_2 s_i, \tag{7}$$

where C is the number of classes, $t_i$is the predicted probability, s_i is the groundtruth probability.

3.5. Semi-supervised learning

Semi-supervised learning allows the use of a combination of labeled and unlabeled data. In our case, proxy labeling [23] is the most straightforward approach that is easy to utilize as it defines all the parameters implicitly, with the minimal confidence threshold being the only hyperparameter of the algorithm.

The proxy labeling is an iterative process, where during each iteration a labeled dataset is extended by using the trained model to label unlabeled samples, extend the labeled dataset, and repeat the same process until either all of the labels are labeled or iteration limit is reached. More formally, proxy labeling consists of two stages - train (8) and update (9) that are repeated in the loop:

$$\theta_i, b_i = train(f, \theta_{i-1}, b_{i-1}, L_{i-1}), \tag{8}$$

where θ are model weights, b are model biases, f is the approximator function, L is the labeled dataset, *train* is the training function. During the semi-supervised learning, datasets are updated between the training iterations as well as follows:

$$L_i = L_{i-1} \cup S_i, U = U_{i-1} \cap S, S = \left\{ u \middle| f(u, \theta_i, b_i) \geq \epsilon, u \in U_{i-1} \right\}, \tag{10}$$

where L is the labeled dataset, U is the unlabeled dataset, f is the discriminator model, ϵ is the minimum confidence threshold, W_i, b_i are discriminator's parameters learnt on i-th iteration.

This approach is based on the smoothness assumption (for two points x and x' that are close to each other in the input space, their labels y and y' should match), but does not require an explicit definition of distance metric, as it is instead replaced by the embeddings of the discriminator and confidence of the discriminator.

4. Method

In this work we propose an iterative three stage approach to hyperspectral image classification. The first stage consists of image pre-processing, followed by the iterative proxy-labeling approach to learn the features. The overall training flow is shown in Fig. 1.

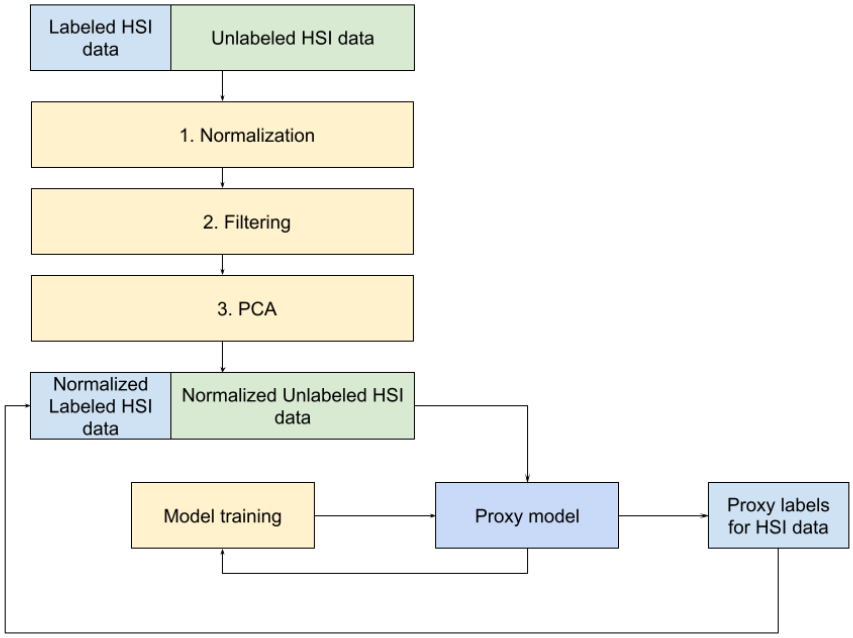

Figure 1: Simplified training flow diagram.

In the first stage data is pre-processed by normalizing and filtering the original dataset. This stage sets up the dataset for future processing. The next step is dimensionality reduction through PCA. This step analyzes normalized data, identifies important bands and decreases dimensionality. Then, the training loop is engaged with continuous proxy-labeling training.

In this work we use a custom CNN architecture that is optimized[24, 25] to handle a large number of channels by squashing them in the first few convolutional layers with 3D layers. The network architecture is pruned using methods outlined in [26, 27] to ensure the best performance during both inference and training. The exact model architecture is shown in Fig. 2.

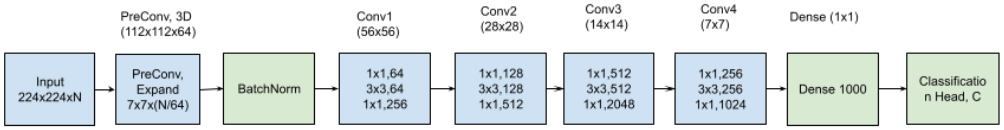

Figure 2: CNN architecture

Semi-supervised learning utilizes a common proxy-label framework - we use 90% as the confidence for unlabeled samples to be transferred to the labeled pool. This update is performed once in 5 epochs of training. Once added to the labeled pool, the samples are removed from the unlabeled samples and treated as ground truth for all of the following training iterations.

It should be noted, that the proposed discriminator architecture is somewhat simplistic compared to the state-of-the-art networks, such as ResNet. The reason for that is twofold. On the one hand, is that the proposed approach only has to classify a fairly modest number of classes (either binary for mine - no mine, or multi-class with up to 20 classes if mines should be classified by their type). As such, the architecture can be simplified to decrease the computational capacity needed during the inference stage. The second reason is that decreased computational cost allows to run the system in a real-time capacity, which is useful for real-time surveying.

As such, the system can be operated in two modes – real-time and batch processing. In the batch mode, the agent first surveys a full area of interest with hyperspectral camera. The collected hypercube is then sliced with a sliding algorithm into a subareas that can be processed by the CNN. Each subarea is pre-processed and passed through the discriminator. Label data is then combined with the telemetry collected by the agent to create the geoinformational database that contains the information about the mines in the area. Unmanned aerial vehicle (UAV) is proposed as the agent, as it offers superior mobility and scanning throughtput [24]

In the realtime mode the data is available in the live feed from the agent. Inference is performed on the agent itself and can be stored on board and broadcasted to ground control station during the survey. In this case, the geoinformational database can be recovered once the mission is compelte to obtain a full map of the minefield. The general flow of both approaches is outlined in Fig. 3

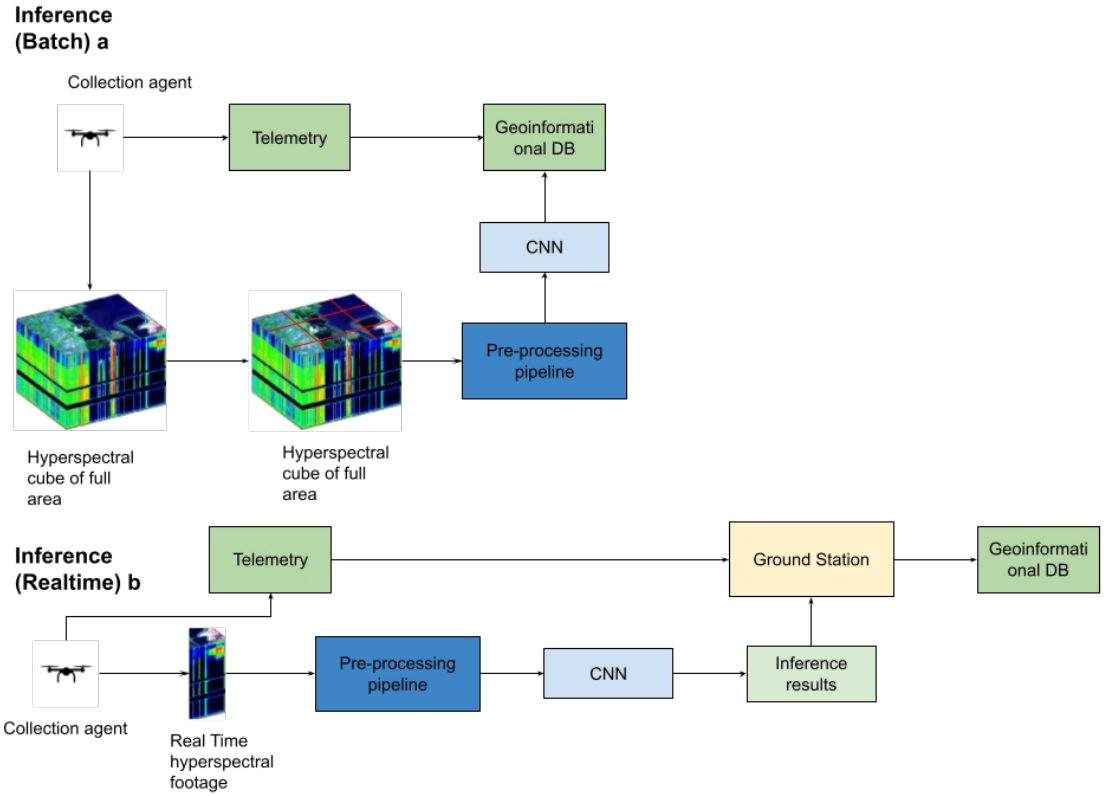

Figure 3: Batch (a) and realtime (b) inference schemas for proposed method.

5. Results

The model was trained with the help of a stochastic gradient descent (SGD) optimizer with an input speed of 0.001, a batch size of 32, and a categorical cross-entropy loss function.

The main evaluation metric used to evaluate the classification efficiency of the model was accuracy.

The Indian Pines dataset is a hyperspectral dataset collected by the AVIRIS sensor over the Indian Pines test site in the southern state of Indiana. The University of Pavia dataset is a hyperspectral dataset collected by the ROSIS sensor over the University of Pavia, Italy.

Results from running the convolutional neural network (CNN) model on the Indian Pines and Pavia University datasets demonstrate the high efficiency of the model. For the Indian Pines data set, accuracy reached 85%, indicating that the model is good at recognizing different classes in folded hyperspectral images. For the University of Pavia data set, the accuracy was 87%, which is also a significant achievement due to the complexity of the data set. Results visualizations are present in the Fig. 4 and 5, tabulated data is presented in Table 1.

It is worth noting, that removing the dimensionality reduction step has a limited effect on the average accuracy, however, extra training epochs were required as not all of the unlabeled data was used and the classifier has not reached its peak discrimination capacity.

The model was tested in the batch mode, and the results were patched together to provide a view of the minefield. In batch mode we were able to achieve an average of 63.26 ms per inference,approximately 17 FPS, which makes the method applicable to real-time problems.

Figure 4: Classification results (Indian Pines dataset). Source data (left) and classification results (right).

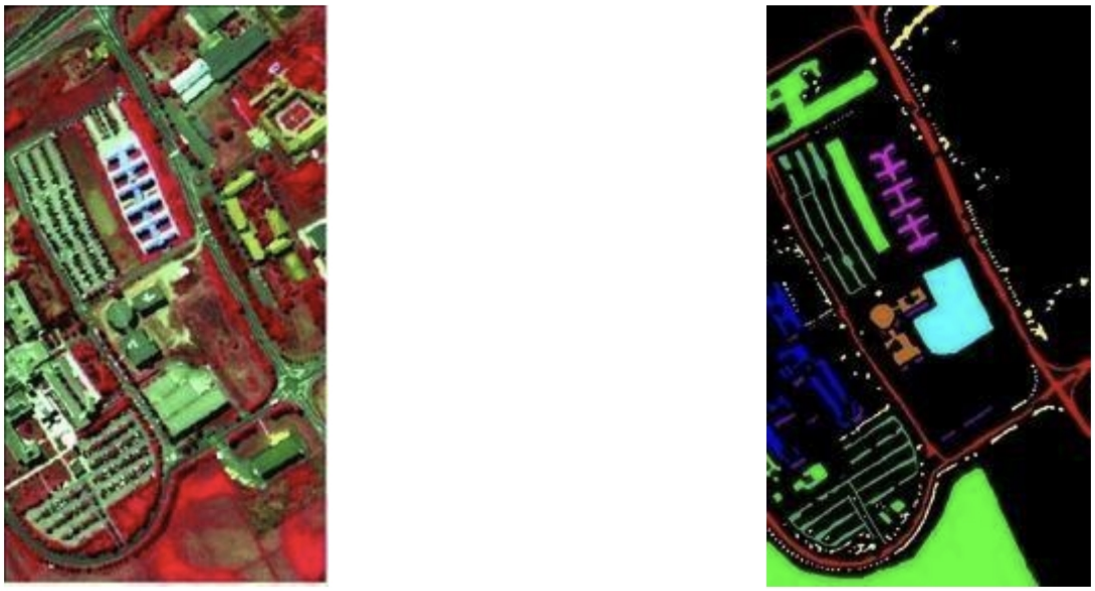

Figure 5: Classification results (Pavia University). Source data (left) and classification results (right).

Table 1
Accuracy results

Dataset	Average Accuracy	Loss	PCA	Epochs
Pavia University	87.32%	0.0019	Yes	200
Indian Pines	85.35%	0.0028	Yes	200
Pavia University	87.25%	0.002	No	300

Indian Pines	85.17%	0.0031	No	300

6. Conclusion

The proposed multi-step hybrid semi-supervised neural network approach is efficient in processing of hyperspectral imagery. The approach is able to efficiently reduce the dimensions of the input data by discarding low-information bands in the pre-processing step via a PCA or other dimensionality reduction tools and then applying a neural network-based discriminator to the processed data to solve the problem of hyperspectral multi-class classification. A semi-supervised learning loop based on proxy-labeling is used to augment the learning capacity, decrease the amount of labeled data needed, and minimize the number of the model's hyperparameters.

The research has several limitations. Firstly, the PCA is a linear dimensionality reduction algorithm, which might remove information, especially in hyperspectral data, as non-linear structures are likely to be lost in the dimensionality reduction procedure. Additionally, used CNN architecture is fairly simplistic and might not have enough learning capacity. Lastly, proxy-labeling threshold is constant during the training process, which might lead to suboptimal labeling as the number of classes grows.

The approach is comparable to the state-of-the-art methods, however it is faster to train and adapt as it has fewer parameters to tune. The drawback of this approach, however, is vulnerability to poorly labeled data. It is also sensitive to non-linear noise structures, as PCA is more likely to remove informative bands in its presence. Further research includes two primary directions: improving pre-processing and optimizing semi-supervised learning. Pre-processing can be improved by sacrificing learning speed for non-linear dimensionality reduction methods, such as Isomap or Laplacian eigenmaps. Semi-supervised learning aspect would benefit greatly from defining an explicit distance metric. As such, autoencoders could be used to learn a distance metric between two samples from the input space, unlocking the application of semi-supervised algorithms with stricter prerequisites.

References

[1] Kuo, Frances Y., and Ian H. Sloan. "Lifting the curse of dimensionality." Notices of the AMS 52.11 (2005): 1320-1328.

[2] Ilegbune, Theresa Oby. "Wartime Environmental Pollution and Endangerment: The Landmine Scourge and the Global Effort to Eliminate It." Ann. Surv. Int'l & Comp. L. 21 (2016): 177.

[3] Lu, Bing, et al. "Recent advances of hyperspectral imaging technology and applications in agriculture." Remote Sensing 12.16 (2020): 2659.

[4] J. E. McFee ; K. L. Russell and M. R. Ito, Detection of surface-laid minefields using a hierarchical image processing algorithm, Proc. SPIE 1567, Applications of Digital Image Processing XIV, 42 (December 1, 1991); doi:10.1117/12.50802

[5] L. B. Wolff ; D. A. Socolinsky ; C. K. Eveland ; J. I. Yalcin and J. H. Holloway, Jr. Image fusion of shortwave infrared (SWIR) and visible for detection of mines, obstacles, and camouflage, Proc. SPIE 5089, Detection and Remediation Technologies for Mines and Minelike Targets VIII, 1298 (September 15, 2003).
[6] J. E. McFee, H. T. Ripley, R. Buxton, A. M. Thriscutt, Preliminary study of detection of buried landmines using a programmable hyperspectral imager, Proc. SPIE 2765, Detection and Remediation Technologies for Mines and Minelike Targets, 476 (May 31, 1996); doi:10.1117/12.241250
[7] J. E. McFee and H. T. Ripley, Detection of buried land mines using a CASI hyperspectral imager, Proc. SPIE 3079, Detection and Remediation Technologies for Mines and Minelike Targets II, 738 (July 22, 1997)
[8] G. Suganthi, R. Korah, Discrimination of Mine-Like Objects in Infrared Images Using Artificial Neural Network, Indian Journal Of Applied Research, Volume : 4 Issue : 12 , India, p 206-208, Dec 2014
[9] M.-A. Gagnon, P. Lagueux, J.-P. Gagnon, S. Savary, P. Tremblay, V. Farley, É. Guyot and M. Chamberland. Airborne thermal infrared hyperspectral imaging of buried objects. In SPIE Security+ Defence, pp. 96490T-96490T. International Society for Optics and Photonics, 2015.
[10] Ba, Jimmy Lei, Jamie Ryan Kiros, and Geoffrey E. Hinton. "Layer normalization." arXiv preprint arXiv:1607.06450 (2016).
[11] Patro, S. G. O. P. A. L., and Kishore Kumar Sahu. "Normalization: A preprocessing stage." arXiv preprint arXiv:1503.06462 (2015).
[12] Miyato, Takeru, et al. "Spectral normalization for generative adversarial networks." arXiv preprint arXiv:1802.05957 (2018).
[13] Farnia, Farzan, Jesse M. Zhang, and David Tse. "Generalizable adversarial training via spectral normalization." arXiv preprint arXiv:1811.07457 (2018).
[14] Nettleton, David F., Albert Orriols-Puig, and Albert Fornells. "A study of the effect of different types of noise on the precision of supervised learning techniques." Artificial intelligence review 33 (2010): 275-306.
[15] Gupta, Gajanand. "Algorithm for image processing using improved median filter and comparison of mean, median and improved median filter." International Journal of Soft Computing and Engineering (IJSCE) 1.5 (2011): 304-311.
[16] Wuethrich, Manuel, et al. "A new perspective and extension of the Gaussian Filter." The International Journal of Robotics Research 35.14 (2016): 1731-1749.
[17] Cascianelli, Silvia, et al. "Dimensionality reduction strategies for cnn-based classification of histopathological images." Intelligent Interactive Multimedia Systems and Services 2017 10. Springer International Publishing, 2018.
[18] Kwak, Taehong, Ahram Song, and Yongil Kim. "The impact of the PCA dimensionality reduction for CNN based hyperspectral image classification." Korean Journal of Remote Sensing 35.6_1 (2019): 959-971.
[19] Koonce, Brett, and Brett Koonce. "ResNet 50." Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization (2021): 63-72.
[20] Qiu, Zhaofan, Ting Yao, and Tao Mei. "Learning spatio-temporal representation with pseudo-3d residual networks." proceedings of the IEEE International Conference on Computer Vision. 2017.

[21] Grandini, Margherita, Enrico Bagli, and Giorgio Visani. "Metrics for multi-class classification: an overview." arXiv preprint arXiv:2008.05756 (2020).
[22] Zhang, Zhao, Tommy WS Chow, and Mingbo Zhao. "M-Isomap: Orthogonal constrained marginal isomap for nonlinear dimensionality reduction." IEEE transactions on cybernetics 43.1 (2012): 180-191.
[23] Nath, Vishwesh, et al. "Warm start active learning with proxy labels and selection via semi-supervised fine-tuning." International conference on medical image computing and computer-assisted intervention. Cham: Springer Nature Switzerland, 2022.
[24] Zgurovsky, M., Sineglazov, V., Chumachenko, E. (2021). Classification and Analysis of Multicriteria Optimization Methods. In: Artificial Intelligence Systems Based on Hybrid Neural Networks. Studies in Computational Intelligence, vol 904. Springer, Cham. https://doi.org/10.1007/978-3-030-48453-8_
[25] V.M. Sineglazov, K.D. Riazanovskiy, O.I. Chumachenko, (2020). Multicriteria conditional optimization based on genetic algorithms. In: System research and information technologies, No. 3 (2020) . DOI: https://doi.org/10.20535/SRIT.2308-8893.2020.3.07
[26] Sineglazov , V., & Kot, A. (2021). Design of hybrid neural networks of the ensemble structure. Eastern-European Journal of Enterprise Technologies, 1(4 (109), 31–45. https://doi.org/10.15587/1729-4061.2021.225301
[27] Zgurovsky, M., Sineglazov, V., Chumachenko, E. (2021). Formation of Hybrid Artificial Neural Networks Topologies. In: Artificial Intelligence Systems Based on Hybrid Neural Networks. Studies in Computational Intelligence, vol 904. Springer, Cham. https://doi.org/10.1007/978-3-030-48453-8_3
[28] V. Sineglazov, Main features of terrain-aided navigation systems, 2014 IEEE 3rd International Conference on Methods and Systems of Navigation and Motion Control (MSNMC), Kiev, Ukraine, 2014, pp. 49-52, doi: 10.1109/MSNMC.2014.6979728.