Forward-Forward Algorithm for Hyperspectral Image Classification

Abel A. Reyes-Angulo
Michigan Technological University
1400 Townsend Dr, Houghton, MI 49931
areyesan@mtu.edu

Sidike Paheding
Fairfield University
1073 N Benson Rd, Fairfield, CT 06824
spaheding@fairfield.edu

Abstract

The back-propagation algorithm has long been the de-facto standard in optimizing weights and biases in neural networks, particularly in cutting-edge deep-learning models. Its widespread adoption in fields like natural language processing, computer vision, and remote sensing has revolutionized automation in various tasks. The popularity of back-propagation stems from its ability to achieve outstanding performance in tasks such as classification, detection, and segmentation. Nevertheless, back-propagation is not without its limitations, encompassing sensitivity to initial conditions, vanishing gradients, overfitting, and computational complexity. The recent introduction of a forward-forward algorithm (FFA), which computes local goodness functions to optimize network parameters, alleviates the dependence on substantial computational resources and the constant need for architectural scaling. This study investigates the application of FFA for hyperspectral image classification. Experimental results and comparative analysis are provided with the use of the traditional back-propagation algorithm. Preliminary results show the potential behind FFA and its promises.

1. Introduction

Deep Learning (DL) [13] has been revolutionizing many different fields due to its ability to achieve unprecedented performance when applied to real-world problems, including applications in agriculture [9, 16], medicine [3], cybersecurity [15], and many others [1, 2, 11, 17, 23].

Hyperspectral imagery (HSI) contains an extensive array of continuous spectral information across numerous narrow bands. The inherent high-dimensionality of HSI data poses substantial obstacles to accurate classification tasks (i.e. crop or soil classification), owing to intricate spectral variations and a scarcity of labeled samples. Deep learning models, specifically convolutional neural networks (CNNs) [12], have exhibited remarkable accomplishments in numerous computer vision tasks, notably including the
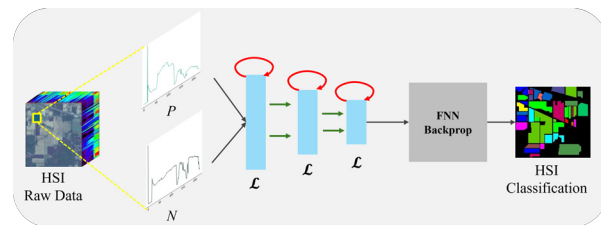

Figure 1. Illustration of the proposed methodology. We aim to combine the use of forward-forward algorithm (FFA) with traditional Backprop for HSI classification task. FNN: Feedforward Neural Network.

classification of hyperspectral images [7, 14]. Nevertheless, the conventional backpropagation algorithm, which is commonly employed for training deep learning models, may confront certain limitations within this domain, such as computational or energy cost, and sensitivity to initial conditions [5, 22].

By propagating errors in a backward manner through the network, the backpropagation algorithm calculates gradients that guide the adjustment of the model's parameters to minimize the objective function [21]. While backpropagation has proven effective in deep learning applications, it encounters difficulties when dealing with hyperspectral data such as limited availability of labeled samples. As a result, there is a need for alternative training approaches that can enhance the performance of deep learning models, specifically in the context of hyperspectral image classification.

In this study, we investigate the performance of the forward-forward algorithm (FFA) [5] for the task of hyperspectral image classification. The FFA explores the relationships among input data samples by feeding forward both the original data (positive data) and an alternative version of this data (negative data), encouraging the model to learn robust features that capture the underlying characteristics of the HSI.

Our initial experiments show that solely using the FFA does not yield better results compared to the backpropagation algorithm. However, considering the advantages of

both methods, we propose to combine the forward-forward pass algorithm with traditional backpropagation for hyperspectral image classification. The idea is to incorporate the forward-forward algorithm as an initial learning stage, allowing the model to learn more discriminative features. Subsequently, the model is fine-tuned using the backpropagation algorithm, which refines the learned representations and optimizes the classification performance. Figure 1 illustrates the proposed methodology for the HSI classification task. To the best of our knowledge, this work is the first attempt to utilize FFA for HSI classification tasks.

The remaining sections of this paper are structured as follows: Section 2 summarizes related works in HSI classification task. Section 3 presents an overview of the FFA explored in this work. Section 4 details the proposed hybrid approach. Section 5 presents the benchmark dataset utilized for the experiments. Section 6 provides the experimental setup and analyzes the obtained experimental results. Section 7 provides insights regarding the limitations in the use of the proposed methodology, with a particular focus on the FFA. Finally, Section 8 provides a summary of the findings from our study.

2. Related Work

CNN models that focus on spectral information or pixel values use one-dimensional spectral signatures, indicated as $\mathbf{x}_i$ within the real space R^B, as their input, where B is the count of spectral bands present in an HSI. Hu et al. [6] introduced a basic 1D CNN structure, consisting of a convolution layer, a pooling layer, and a fully connected layer for classifying HSI, treating each pixel's spectral signature as a one-dimensional series. This model demonstrated superior accuracy when compared with a two-layer neural network and the support vector machine classifier. To counter the significant correlation among spectral bands in HSI, Gao et al. [4] converted the one-dimensional spectral data into two-dimensional spectral feature matrices and utilized compact convolutional kernels, specifically sized 3×3 or 1×1, to create convolutional layers. More recently, a novel 1D CNN technique, Plastic-Net [8], was developed to detect plastic materials from ATR-FTIR spectra. Plastic-Net processes one-dimensional spectral data from ATR-FTIR, translating it into Gramian Angular Fields (GAF), thereby creating a two-dimensional matrix. This conversion facilitates the use of two-dimensional CNN on the GAF matrices, which has proven to be more accurate in sorting mixed plastic waste than the 1D CNN. To enhance computational efficiency, this framework also incorporates Piecewise Aggregate Approximation (PAA) [10] to reduce the size of the input GAF matrices. Additionally, Wu et al. [25] integrated recurrent layers with convolutional layers to simultaneously capture contextual details and features that are invariant locally from the one-dimensional spectral data. This com-

bined architecture outperformed both one-dimensional and two-dimensional CNN models in terms of accuracy. Lastly, Paheding et al. [18] introduced GAF-NAU leveraging the GAF pixel-wise conversion to effectively perform classification with the use of hyperspectral pixel data solely, alleviating the dependency of the pixel neighborhood reference values. This method, in addition, combined the use of GAF with deep learning architecture models, such as U-Net [19], achieving state-of-the-art performance over several benchmark datasets.

3. Methods

3.1. The backpropagation

Introduced by Rumelhart et al. [20, 21], in the backpropagation algorithm for training artificial neural networks, the process involves computing the gradient of the cost function with respect to the network's parameters and using this information to update the weights via gradient descent. Backpropagation has demonstrated strong generalization capabilities and effectiveness in handling non-linearities, making it applicable to a wide range of neural network types including feedforward networks, recurrent networks, and convolutional networks.

3.2. The Forward-Forward algorithm

Forward-forward algorithm (FFA) [5] involves substituting the conventional forward and backward passes from the backpropagation algorithm, with two forward passes that function in a parallel manner but on distinct data with opposing objectives. Figure 2 illustrates a comparison between the traditional backpropagation algorithm and the newer forward-forward algorithm. The affirmative pass involves "real data" and modifies the weights to improve the goodness in each hidden layer, while the negative pass operates on "negative data" and modifies the weights to diminish the goodness within every hidden layer. In [5], two distinct criteria were investigated for measuring quality: the sum of the squared neural activities and the negative sum of the squared activities, although numerous other criteria can also be utilized. In the original FFA, the sum of the squares of the activities in the layer is expressed as

$$G = \sum_j z_j^2, \qquad (1)$$

where z_j represents the activity of the jth hidden unit.

Furthermore, the positive and negative passes adjust the weights locally, and the probability of the outputs are expressed as follows:

$$prob(\text{positive}) = \sigma(G - \theta) \qquad (2)$$

where σ despite a logistic distribution function, and and θ a given threshold.
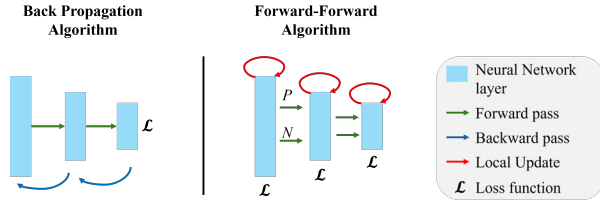
Figure 2. A brief illustration comparing the backpropagation and the forward-forward algorithms.

To facilitate the contrast of positive and negative data during the supervised training process of FFA, we need to develop a method for merging the data with their corresponding labels. In [5], Hinton proposed to overlay the label information onto the data itself or embed the label within the input data. However, in this work, we take a different approach by appending the label at one end of the spectral signature of each sample. Moreover, we explore various methods for encoding the label information. For instance, we experiment with one-hot encoding representation, binary representation, and decimal representation. Through the analysis of experimental results, it is found that using the one-hot encoding representation to append the label information with the hyperspectral signature data yields better outcomes. As a result, all the experiments reported in this work utilizing FFA are trained using this approach. Figure 3 visually depicts the imputation method employed, where the label-encoded information is appended at one end of the pixel's hyperspectral signatures.

4. FFA for hyperspectral image classification

In this preliminary study, we contemplate the implementation of FFA with the use of fully connected layers and 1D convolutional layers for HSI classification. The fully connected layer performs a linear transformation on the input data, usually in a vector shape, through weights matrix multiplication operations, in addition to a bias term. The fully connected layers comprise the contributions of all inputs for a final prediction. In contrast, convolutional layers in CNN can be used to capture local dependencies in sequential data, such as time series or text. Unlike fully connected layers that operate on the entire input, CNN considers the local receptive field of the input at a time. This allows them to extract features that are sensitive to local patterns and variations.

4.1. Fully Connected FFA network for pixel-wise HSI classification

The classification of hyperspectral images poses a substantial challenge attributed to the data's high dimensionality and spectral complexity. Deep learning architectures have shown promise in extracting discriminative features from hyperspectral images. Nonetheless, the efficacy of these architectures is heavily contingent upon the quality of the learned representations. In this study, we explore the use of the FFA with Fully Connected layers to enhance for HSI classification task. FFA comprises the use of a few hidden layers to extract features from the HSI data, scale it to a latent space, and produce the final pixel-wise classification. A total of 3 hidden layers were used in our FFA, with the following number of units: 784, 500, and 500, respectively.

4.2. Convolutional FFA network for pixel-wise HSI classification

In this work, we explore the implementation of the aforementioned types of neural network layers, limited to 1D data. Convolutional layers for 1D data are usually implemented as neural networks that convolve the input of the hyperspectral image with a set of learnable filters. These layers detect local relationships and patterns within the interaction of the hyperspectral bands per pixel. The capture of this spectral signature enables the architecture to learn discriminative representations of the data. By using the FFA technique, the network aims to update the weights layer by layer in a forward pass only, relieving the computational load of computing the gradient during the backpropagation of the error to update the learned parameters.

The implementation of the 1D CNN layer allows us to implement an FFA architecture similar to the type used for HSI classification [7]. The proposed FFA comprises the use of 1D CNN layers in the early stages to capture feature representations within a different latent dimensional space, while the fully connected layers are used at the end to learn how to properly discriminate among the classes. The network is configured as follows: an initial 1D convolutional layer with 64 kernels of size 64, followed by a set of two hidden layers with 128 and 256 feature maps, respectively, both with a kernel size of 36. Then, a max-pooling operation is applied to downscale the dimensions of the tensor by a factor of two. Another 1D CNN layer is applied with 256 kernels of size 36. This last 1D CNN layer is followed by another max-pooling operation with similar characteristics as the previous one, and a flattening operation. Finally, two fully connected layers are added with 100 and N units, where N represents the number of classes in the HSI dataset.

4.3. Combination of FFA with backpropagation

Given the similarities between the nature of FFA and the training procedure in contrasting learning, we propose the utilization of FFA during the initial stage of training. In this stage, each sample is contrasted with different output choices, enabling the model to learn how to effectively discriminate between the correct prediction and other alternatives. Subsequently, the model proceeds to refine its learning through traditional backpropagation using the
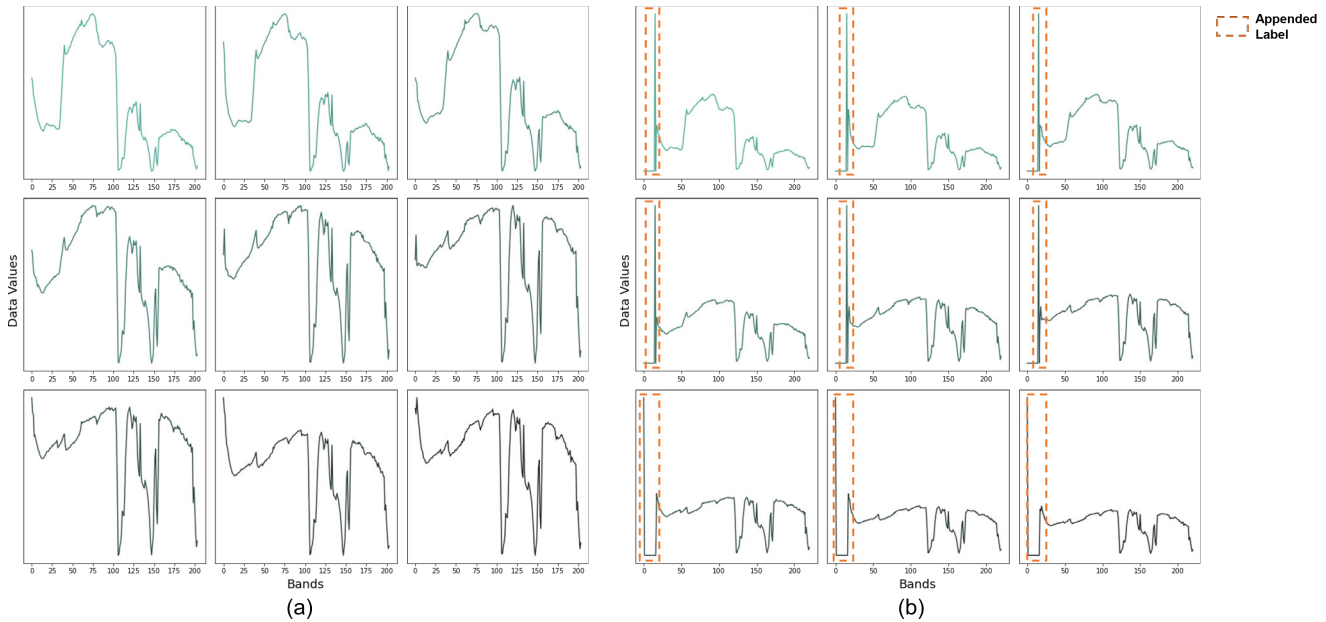
Figure 3. Illustration of label embedding method. (a) shows spectral signatures from the Salinas hyperspectral dataset. (b) shows the same samples with the appended label information, which is appended at the beginning of spectral signatures.

same deep learning architecture. This process facilitates the model in adjusting the extraction of meaningful characteristics from the high-level spectral information and fine-tuning the latent representation for accurate final predictions. After this initial phase, the model transitions to the standard back-propagation technique, which leverages feedforward neural network for further refinement. During this stage, the model fine-tunes its representation, optimizing its capacity to capture significant features from the high-dimensional spectral data.

5. Datasets

For simplicity of this proof-of-concept, we perform experiments over three publicly available dataset[1]: The Salinas valley, the Indian Pines, and the Pavia University.

5.1. The Salinas Valley

The Salinas dataset is a popular hyperspectral image dataset that is commonly used in the realms of remote sensing and image processing. It is named after the Salinas Valley in California, USA, where the data was collected. The dataset consists of a hyperspectral image of size 512×217 pixels, with 224 spectral bands covering the range from 0.2 to 2.4 micrometers. Each pixel in the image represents a small area on the ground, and the spectral bands capture the reflectance of the surface at different wavelengths. The Sali-

nas dataset was collected using an Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor, which was flown over an agricultural area in the Salinas Valley. The image contains 16 different crop types, including lettuce, broccoli, and bare soil, among others.

5.2. The Indian Pines

Collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over an agricultural area in Indiana, USA, the Indian Pines dataset consists of a hyperspectral image of size 145×145 pixels, with 224 spectral bands covering the range from 0.4 to 2.5 micrometers. Each pixel in the image represents a small area on the ground, and the spectral bands capture the reflectance of the surface at different wavelengths. The dataset contains 16 different land cover classes, including crops, trees, roads, and buildings, among others.

5.3. Pavia University

The University of Pavia HSI dataset, a component of the Pavia scenes, was captured using the Reflective Optics System Imaging Spectrometer (ROSIS) sensor over Pavia, a city in northern Italy. This dataset comprises an image of dimensions 610×610, encompassing a total of 103 spectral bands. It features nine different land cover types, each with a spatial resolution of 1.3 meters.

---

[1]www.ehu.eus/ccwintco/index.php/Hyperspectral Remote Sensing Scenes

Table 1. Summary of the qualitative results using the different techniques, mentioned in this work, over Salinas, Indian Pines, and Pavia University datasets. Best performance in marked in **bold** font.

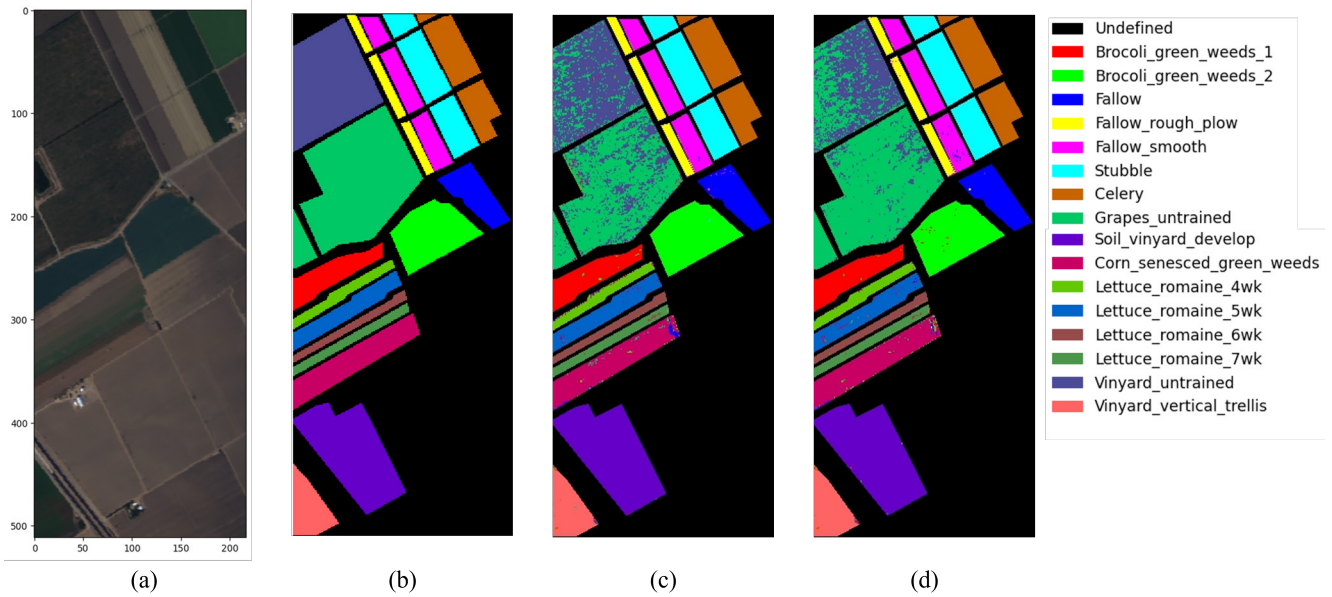| Method	Salinas			Indian Pines			Pavia University		
	OA	AA	κ	OA	AA	κ	OA	AA	κ
BP	0.9190	**0.9605**	0.9099	**0.8109**	**0.7759**	**0.7842**	0.9148	0.8885	0.8866
FFA (Dense layers)	0.8392	0.8470	0.8206	0.5921	0.5841	0.5370	0.7936	0.7635	0.7325
FFA (Dense & Conv layers)	0.9101	0.8518	0.8666	0.6660	0.5822	0.6194	0.8554	0.8458	0.8101
FFA + BP	**0.9221**	0.9564	**0.9130**	0.7365	0.6978	0.6978	**0.9251**	**0.9122**	**0.9011**

Figure 4. Classification maps over the Salinas HSI dataset. (a) True color composite of HSI, and (b) ground truth. Classification maps from (c) traditional BP, and (d) FFA+BP.
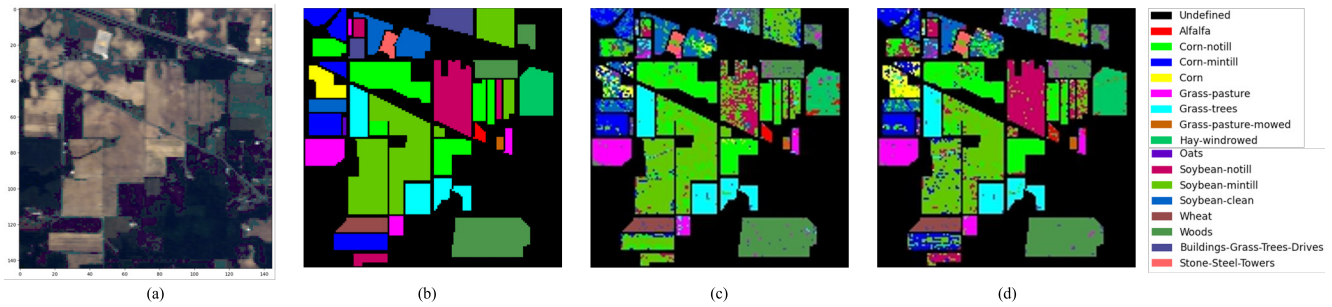

Figure 5. Classification maps over the Indian Pines HSI dataset. (a) True color composite of HSI, and (b) ground truth. Classification maps from (c) traditional BP, and (d) FFA+BP.

6. Results

6.1. Experimental setup

To ensure a fair comparison, we evaluate the different techniques reported in this work using the same training and test datasets. The Salinas, Indian Pines, and Pavia University HSI datasets are randomly split into training, validation, and testing sets with a ratio of 8:1:1, respectively. We repeat the entire experimental process three times to obtain robust performance estimates and average the results.

For each experiment, the respective models are trained for 250 epochs using the Adam optimizer with a fixed learning rate of 1×10^{-3}. When employing backpropagation, categorical cross-entropy is utilized as the loss function.
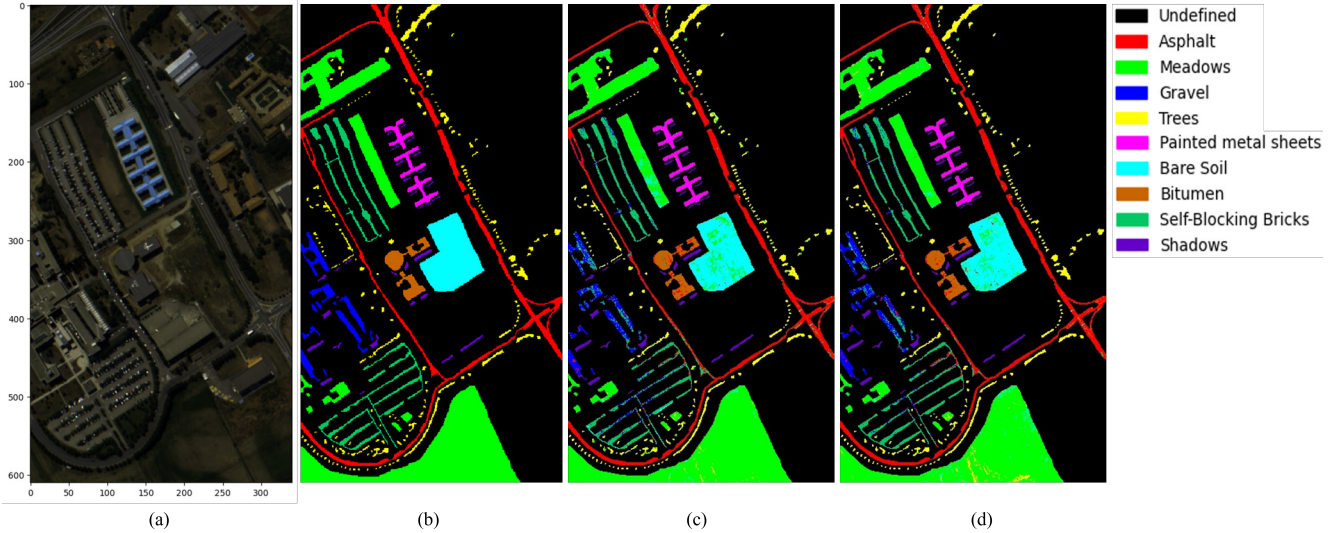
Figure 6. Classification maps over the Pavia University HSI dataset. (a) True color composite of HSI, and (b) ground truth. Classification maps from (c) traditional BP, and (d) FFA+BP.

Table 2. Comparison with other machine learning models in HSI classification task. Key: [**Best**, **Second Best**]

Dataset	Metric	Method				
		1D-CNN [6]	dPEN [24]	Plastic-Net [7]	GAF-NAU [18]	FFA+BP
Salinas	OA	0.9143	**0.9253**	0.9069	**0.9459**	0.9221
	AA	0.9506	**0.9648**	0.9506	**0.9700**	0.9564
	κ	0.9046	**0.9168**	0.8963	**0.9397**	0.913
Indian Pines	OA	0.7439	**0.7762**	0.7300	**0.8107**	0.7365
	AA	0.7636	**0.8128**	0.6329	0.7467	0.6978
	κ	0.7057	**0.7454**	0.6923	**0.7831**	0.6978
Pavia University	OA	0.9017	**0.9121**	0.8970	0.8970	**0.9251**
	AA	0.8937	**0.9054**	0.8776	0.8770	**0.9122**
	κ	0.8685	**0.8829**	0.8629	0.8625	**0.9011**

However, when FFA is used, a custom loss function is implemented to measure the distance between the *goodness* of each positive and negative sample with respect to the provided threshold.

All the experiments are run on an NVIDIA RTX 3070 graphic card with 8GB of dedicated GPU.

6.2. Performance comparison

To assess and contrast the performance of the different techniques presented in this work, we employed the following evaluation metrics:

- **Overall Accuracy** (OA)**:** This metric provides the percentage of correctly classified pixels from the respective HSI dataset.
- **Average Accuracy** (AA)**:** This metric is computed by averaging each class accuracy score, thus providing a class-specific evaluation of the technique's performance.
- **Kappa Coefficient** (κ)**:** This metric measures the agree-

ment between the predicted and true class labels, in which the accuracy that could be achieved by chance is taken into account.

The use of these evaluation metrics collectively enables a comprehensive assessment of the various techniques presented in this work for HSI classification.

Table 1 summarizes the experimental results obtained by evaluating the discussed techniques using the aforementioned performance metrics on the Salinas, Indian Pines, and Pavia University HSI datasets. As shown in Table 1, when considering the Salinas HSI dataset, the combination of FFA and backpropagation (BP) achieved the best performance in terms of OA (0.9221) and κ (0.9130). However, BP alone achieved the highest AA (0.9605). On the other hand, for the Indian Pines HSI dataset, BP exhibits the best performance across all evaluation metrics, with OA (0.8109), AA (0.7759), and κ (0.7842). Nevertheless, the utilization of FFA in combination with BP demonstrated a
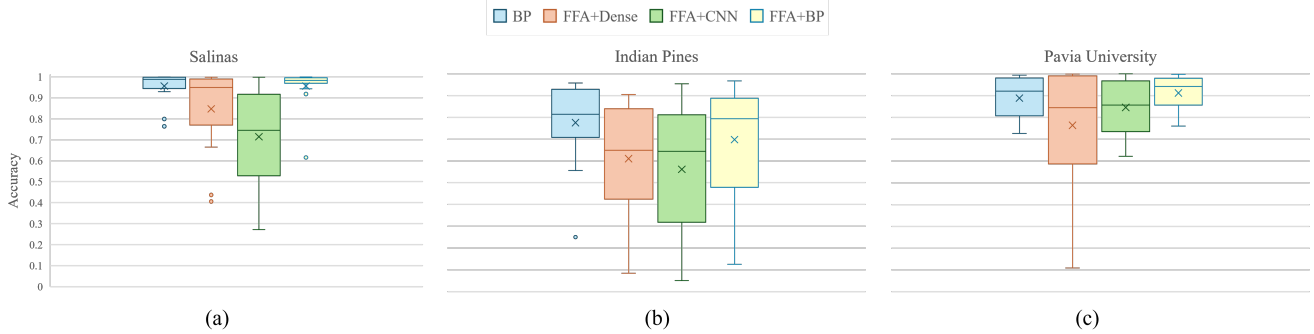
Figure 7. Comparison of the different approaches of the accuracy performance across individual classes over the (a) Salinas, (b) Indian Pines, and (c) Pavia University datasets.

significant improvement compared to using any of the FFA variants individually. Figures 4, 5, and 6 demonstrate classification maps of different methods using the Salinas, Indian Pines, and Pavia University datasets, respectively. Additionally, experimental results, from the Pavia University dataset, indicate the best performance achieved by the combined FFA+BP approach consistently across the three evaluation metrics (OA: 0.9251, AA: 0.9122, and κ: 0.9011).

Figure 7 shows a comparative analysis of classification accuracy for individual classes using different approaches across the three hyperspectral datasets: (a) Salinas, (b) Indian Pines, and (c) Pavia University. The approaches include Backpropagation (BP), FFA with Dense layers (FFA+Dense), FFA with Convolutional Neural Network layers (FFA+CNN), and a hybrid of FFA and Backpropagation (FFA+BP). The boxplots illustrate the mean accuracy (marked by an 'X'), interquartile range (boxes), max/min values (without outliers), and outliers are plotted as dots.

Generally, FFA+BP seems to be a strong contender, often performing well across all three datasets, with particularly consistent high performance in the Salinas and Pavia University datasets. Analysing BP, by itself, shows the most consistency in Salinas in contrast to the performance over Pavia University and Indian Pines datasets. The FFA+Dense approach does exhibit a broader range of accuracy, FFA+CNN displays a wider interquartile range in both Salinas and Indian Pines datasets.

6.3. Comparison with other models

In order to provide a better understanding of this work's contribution, we compare the performance of the proposed approach with several machine learning (ML) models for HSI classification tasks. Table 2 summarizes the performance of a traditional 1D-CNN [6], dPEN model [24], Plastic-Net [7], and GAF-NAU [23] in contrast with our proposed FFA+BP. Reported results indicate that FFA+BA yields promising results. Particularly, in the Pavia Univer-

sity dataset, the proposed method outperforms all the other competing methods across all metrics. We hypothesize the variation in the proposed method's performance across different datasets may indicate that its effectiveness could be dependent on the specific characteristics of the dataset, such as the number of classes and samples per class, the inherent variability within the data, or the distribution of the samples.

7. Discussion

In the investigation of the Forward-Forward Algorithm (FFA) for hyperspectral image classification, our analysis reveals significant insights into its performance, and areas for future exploration such as the use of the algorithm in combination with fully connected neural networks or convolutional neural networks. Comparatively, FFA's integration with the backpropagation (BP) technique has demonstrated an enhanced ability to process high-dimensional hyperspectral data. Through the experiments conducted on the Salinas, Indian Pines, and Pavia University datasets, the FFA+BP method notably outperformed standalone FFA and feedforward network with BP methods in terms of Overall Accuracy (OA) and Kappa coefficient (κ), underscoring its potential in extracting relevant features for classification tasks.

However, the FFA's efficacy is intrinsically linked to the generation and selection of negative samples. This process is critical, as it directly impacts the algorithm's learning dynamics and its subsequent ability to distinguish between classes. The requirement for two forward passes, incorporating both positive and negative data, introduces an additional computational burden when compared to traditional BP. Optimizing this aspect is crucial for enhancing the FFA's practicality and scalability. Future research directions should focus on developing more sophisticated methods for generating negative samples, exploring the integration of FFA with other neural network architectures, and extending its applications beyond classification to other

remote sensing challenges. Further, we plan to evaluate and optimize FFA in terms of computational efficiency, and also investigate its performance across diverse hyperspectral datasets will be pivotal in broadening its applicability and effectiveness in real-world scenarios.

8. Conclusion

We proposed the integration of the Forward-Forward algorithm and traditional backpropagation during the early stages of training. The empirical outcomes demonstrated that this method significantly improves feature representation and boosts the accuracy of classifying hyperspectral images in several scenarios. By incorporating the FFA algorithm, the network was able to capture useful feature representation by adjusting network parameters in every hidden layer. Subsequent fine-tuning through backpropagation facilitated the extraction of discriminative task-specific features. This combined approach exemplifies the potential for leveraging the strengths of different learning algorithms to achieve superior results in hyperspectral image classification tasks.

References

[1] Md Zahangir Alom, Tarek M Taha, Chris Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, Mahmudul Hasan, Brian C Van Essen, Abdul AS Awwal, and Vijayan K Asari. A state-of-the-art survey on deep learning theory and architectures. *electronics*, 8(3):292, 2019. 1

[2] Dimitri Bourilkov. Machine and deep learning applications in particle physics. *International Journal of Modern Physics A*, 34(35):1930019, 2019. 1

[3] Travers Ching, Daniel S Himmelstein, Brett K Beaulieu-Jones, Alexandr A Kalinin, Brian T Do, Gregory P Way, Enrico Ferrero, Paul-Michael Agapow, Michael Zietz, Michael M Hoffman, et al. Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface*, 15(141):20170387, 2018. 1

[4] Hongmin Gao, Yao Yang, Chenming Li, Hui Zhou, and Xiaoyu Qu. Joint alternate small convolution and feature reuse for hyperspectral image classification. *ISPRS International Journal of Geo-Information*, 7(9):349, 2018. 2

[5] Geoffrey Hinton. The forward-forward algorithm: Some preliminary investigations. *arXiv preprint arXiv:2212.13345*, 2022. 1, 2, 3

[6] Wei Hu, Yangyu Huang, Li Wei, Fan Zhang, and Hengchao Li. Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors*, 2015, 2015. 2, 6, 7

[7] Wei Hu, Yangyu Huang, Li Wei, Fan Zhang, and Hengchao Li. Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors*, 2015:1–12, 2015. 1, 3, 6, 7

[8] Shengli Jiang, Zhuo Xu, Medhavi Kamran, Stas Zinchik, Sidike Paheding, Armando G McDonald, Ezra Bar-Ziv, and Victor M Zavala. Using atr-ftir spectra and convolutional neural networks for characterizing mixed plastic waste. *Computers & Chemical Engineering*, 155:107547, 2021. 2

[9] Andreas Kamilaris and Francesc X Prenafeta-Boldú. Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, 147:70–90, 2018. 1

[10] Eamonn Keogh, Kaushik Chakrabarti, Michael Pazzani, and Sharad Mehrotra. Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and information Systems*, 3(3):263–286, 2001. 2

[11] Uzair Khan, Sidike Paheding, Colin P Elkin, and Vijaya Kumar Devabhaktuni. Trends in deep learning for medical hyperspectral image analysis. *IEEE Access*, 9:79534–79548, 2021. 1

[12] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 1

[13] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015. 1

[14] Shutao Li, Weiwei Song, Leyuan Fang, Yushi Chen, Pedram Ghamisi, and Jon Atli Benediktsson. Deep learning for hyperspectral image classification: An overview. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6690–6709, 2019. 1

[15] Samaneh Mahdavifar and Ali A Ghorbani. Application of deep learning to cybersecurity: A survey. *Neurocomputing*, 347:149–176, 2019. 1

[16] Maitiniyazi Maimaitijiang, Vasit Sagan, Paheding Sidike, Sean Hartling, Flavio Esposito, and Felix B Fritschi. Soybean yield prediction from uav using multimodal data fusion and deep learning. *Remote sensing of environment*, 237:111599, 2020. 1

[17] Maryam M Najafabadi, Flavio Villanustre, Taghi M Khoshgoftaar, Naeem Seliya, Randall Wald, and Edin Muharemagic. Deep learning applications and challenges in big data analytics. *Journal of big data*, 2(1):1–21, 2015. 1

[18] Sidike Paheding, Abel A Reyes, Anush Kasaragod, and Thomas Oommen. Gaf-nau: Gramian angular field encoded neighborhood attention u-net for pixel-wise hyperspectral image classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 409–417, 2022. 2, 6

[19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015. 2

[20] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985. 2

[21] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986. 1, 2

[22] Randall S Sexton and Jatinder ND Gupta. Comparative evaluation of genetic algorithm and backpropagation for train-

ing neural networks. *Information sciences*, 129(1-4):45–59, 2000. 1

[23] Nahian Siddique, Sidike Paheding, Colin P Elkin, and Vijay Devabhaktuni. U-net and its variants for medical image segmentation: A review of theory and applications. *Ieee Access*, 9:82031–82057, 2021. 1, 7

[24] Paheding Sidike, Vasit Sagan, Maitiniyazi Maimaitijiang, Matthew Maimaitiyiming, Nadia Shakoor, Joel Burken, Todd Mockler, and Felix B Fritschi. dpen: Deep progressively expanded network for mapping heterogeneous agricultural landscape using worldview-3 satellite imagery. *Remote sensing of environment*, 221:756–772, 2019. 6, 7

[25] Hao Wu and Saurabh Prasad. Convolutional recurrent neural networks forhyperspectral data classification. *Remote Sensing*, 9(3):298, 2017. 2