

This CVPR workshop paper is the Open Access version, provided by the Computer Vision Foundation. Except for this watermark, it is identical to the accepted version; the final published version of the proceedings is available on IEEE Xplore.

GAF-NAU: Gramian Angular Field encoded Neighborhood Attention U-Net for Pixel-Wise Hyperspectral Image Classification

Sidike Paheding, Abel A. Reyes, Anush Kasaragod, Thomas Oommen Michigan Technological University Houghton, MI, USA

{spahedin, areyesan, akasarag, toomme}@mtu.edu

Abstract

Hyperspectral image (HSI) classification is the most vibrant area of research in the hyperspectral community due to the rich spectral information contained in HSI can greatly aid in identifying objects of interest. However, inherent non-linearity between materials and the corresponding spectral profiles brings two major challenges in HSI classification: interclass similarity and intraclass variability. Many advanced deep learning methods have attempted to address these issues from the perspective of a region/patch-based approach, instead of a pixel-based alternate. However, the patch-based approaches hypothesize that neighborhood pixels of a target pixel in a fixed spatial window belong to the same class. And this assumption is not always true. To address this problem, we herein propose a new deep learning architecture, namely Gramian Angular Field encoded Neighborhood Attention U-Net (GAF-NAU), for pixel-based HSI classification. The proposed method does not require regions or patches centered around a raw target pixel to perform 2D-CNN based classification, instead, our approach transforms 1D pixel vector in HSI into 2D angular feature space using Gramian Angular Field (GAF) and then embed it to a new neighborhood attention network to suppress irrelevant angular feature while emphasizing on pertinent features useful for HSI classification task. Evaluation results on three publicly available HSI datasets demonstrate the superior performance of the proposed model. The source code available at https://github.com/MAIN-Lab/GAF-NAU/

1. Introduction

Hyperspectral images (HSIs) contain abundant spectral bands/channels where each band measures the radiated energy from objects in narrow bandwidths. The detailed spectral information has several applications in fields of agriculture, forestry, urban and natural resources management



Figure 1. High-level schematic of the proposed methodology for pixel-wise HSI classification. GAF: Gramian Angular Field. GAF-NAU: Gramian Angular Field encoded Neighbor Attention U-Net.

pertaining to land use classification [19], defense [25], mineral mapping [23], vegetation health analysis [14], etc. The challenging part, however, is to differentiate the variability in the spectral signatures of materials in a scene.

The conventional HSI classification involves the following steps: removing noisy bands (such as water absorption bands), data normalization, feature extraction, spectral and/or spatial feature based classification. Spectral-feature based classification approaches treat each HSI pixel as a 1D spectrum and then perform pixel-wise categorization, while spatial-feature based methods take into account the characteristics of pixels of its neighborhood in the spatial domain. Moreover, the spectral-spatial based classification frameworks utilize both spectral and spatial features during HSI classification. The limitation of the conventional classification approach is that feature engineering and classification are two separate tasks, and the hand-crafted features can hinder a classifier's performance if it is not carefully designed. Recent advances in machine learning, particularly, deep learning have shown great promises in achieving end-to-end HSI classification framework by automating the feature engineering process and has been able to produce models with demonstrated good performance [1, 13].

Among various deep learning approaches for HSI classification, convolutional neural network (CNN) [12] has become the most popular one. CNN and its variants extract informative features from the original data via a series of hierarchical layers such as convolution, max pooling and fully connected layers. CNN has been employed to perform 1-D spectral, 2-D spectral-spatial, and 3-D spectral-spatial based HSI classification. [1, 8, 24]. However, most recent works that involve deep CNN empathize on utilizing both spatial and spectral features to obtain benefits of 2D or 3D convolutional operations, instead of working on 1D pixel-vector itself due to the complex spectral property of HSI such as high-dimensionality and spectral correlation. Although 2D or 3D based CNN based frameworks have shown promises compared to 1D CNN methods, it assumes pixels within a fixed spatial size of neighborhood share similar spectral characteristics, which ignores that the possibility of pixels in the same neighborhood may represent a different class. This problem becomes more prominent if the spatial resolution is coarser. To address this issue, a robust deep learning framework capable of transforming 1D spectral vectors of hyperspectral data into 2D spectral feature matrices is needed to enable utilizing 2D CNN architectures, however, this type of approach has rarely been studied.

In this research, we propose an effective pixel-wise HSI classification framework that represents a 1D spectral signature of a pixel vector as a 2D feature map in Gramian Angular Fields (GAF) and then embeds it to a deep network consisting of neighborhood attention gate, progressive expansion layer, and U-Net framework. Figure 1 shows a highlevel schematic of our method. The proposed approach not only alleviates spectral correlation challenges as it considers CNN-based operations on angular fields, but also allows for integrating advanced CNN architectures into the HSI classification framework to account for interclass and intraclass spectral variations. The key contributions are summarized as follows:

- To the best of our knowledge, this study is the first attempt to introduce the notion of 2D attention learning and 2D U-Net for a pixel-wise HSI classification task.
- The concept of attention gate with progressive expansion layer that connects encoder and decoder in U-Net framework is introduced, which utilizes attention from both higher and lower feature maps to highlight relevant features for better classification accuracy.
- We demonstrate the state-of-the-art classification performance on three benchmark HSI datasets by comparing existing CNN-based pixel-wise HSI classification frameworks.

In the rest of the paper, Section 2 reviews related pixelwise HSI classification methods. Section 3 introduces the proposed approach, while the experimental analysis is presented in Section 4. Finally, we conclude the work in Section 5.

2. Related Work

Spectral-based or pixel-based CNN models consider 1D spectral signature, denoted as $\mathbf{x}_i \in R^B$, as an input, where B represents the number of spectral bands in a given HSI data. In [9], a simple 1D CNN architecture (i.e., one convolution, one pooling and one fully connected layer) is introduced for HSI classification by considering the spectral signature of each pixel as a 1D array, and yielded better accuracy than two-layer neural network and supper vector machine classifier. To alleviate the influence of strong correlation among HSI spectral bands, Gao et al. [5] reshaped the 1D spectral vectors of hyperspectral data into 2D spectral feature matrices and then adopted small convolution kernels with size of 3×3 or 1×1 to form convolutional layers. Recently, a new 1D CNN approach named Plastic-Net [10] is introduced to identify plastic components from ATR-FTIR (attenuated total reflection-Fourier transform infrared spectroscopy) spectra. The PlasticNet receives 1D spectral signal collected from ATR-FTIR and represents it as Gramian angular fields (GAF) to form 2D matrix. This transformation allows for applying 2D CNN on 2D GAF, which produced higher accuracy in classifying mixed plastic waste comparing to 1D CNN. Their framework also employed a Piecewise Aggregate Approximation (PAA) [11] method to reduce the dimension of the input GAF matrices with the aim of ameliorating computational burden of 2D CNN. In [22], recurrent layers are combined with convolutional layers to extract both contextual information and locally-invariant features from 1D spectral vector, the architecture achieved better accuracy than 1D and 2D CNN methods. Charmisha et al. [4] performed a comparative analysis on the effect of a dimensionality reduction (DR) method, namely dynamic mode decomposition (DMD), for the performance of 1D CNN. The results indicate a comparable classification accuracy can be obtained with the reduced feature dimension by applying DMD. Zhu et al. [26] improved the generalization capability of a CNN classifier by introducing generative adversarial network (GAN) in the HSI classification framework which tends to reduce overfitting issues associated with 1D-CNN. Sidike et al. [20] proposed a pixel-wise deeper CNN classifier, namely deep Progressively Expanded Network (dPEN), which can extract pertinent features from raw data using very limited training data, and outperformed 1D-CNN and other popular machine learning classifiers.

Recently, U-Net and its variants, as well as attention gates, have been employed and have shown promising results for HSI classification task, where a 2D patch-based semantic segmentation approach is typically implemented [6, 7, 15, 16, 18]. To the best of our knowledge, there is no approach thus far that performs attention U-Net based pixel-wise HSI classification. In this work, we introduce a neighborhood attention U-Net scheme where attention

gates encapsulate neighboring feature maps and progressive expansion layers. The proposed architecture encodes 1D pixel vector using Gramian Angular Summation Fields (GASF) and Gramian Angular Difference Fields (GADF) which are then fed to a neighborhood attention U-Net to perform pixel-wise classification. It is worth mentioning that U-Net based deep learning architectures are designed to output a 2D segmentation map in the last stage of the network, however, in the HSI classification task, it is expected to output a single value that indicates a class label. To address this issue, we also employ a majority voting scheme onto the segmentation map to calculate a class label, More details will be provided in Section 3.

3. Proposed Methodology



Figure 2. Conversion of 1D spectra to the polar coordinate system and to GASF and GADF.

3.1. Gramian Angular Fields (GAF)

A HSI can be represented as a 3D cube as $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{H \times W} \in R^{H \times W \times B}$ where H and W are the height and width of the spatial image, respectively. B denotes the total number of spectral bands. $\mathbf{x}_i \in R^B$ is the i^{th} sample in the HSI cube with the B-dimension, which belongs to one of the available class $y_i \in \{1, 2, \dots, C\}$. For 1D-CNN classification, each pixel is given as input as a 1D vector to a CNN architecture where 1D convolution and 1D pooling are applied. In contrast, for 2D-CNN classification, either the 1D vector of spectra needs to be transformed into the 2D matrix, or a neighborhood window/patch of a center (target) pixel is treated as a sample. In this study, we employ GAF to transform a 1D pixel vector to the 2D matrix and then send it to a deeper network for prediction. Originally, GAF is used to encode time-series as images to capture correlation structures [21] and use that output to process 2D-CNN. Similarly, GAF can be used to encode spectra as a 2D feature map to use as input to the 2D-CNN. In GAF, a

1D signal is represented in a polar coordinated system and the angles of each data point are converted into matrices using various operations as described in the following. In the proposed framework, pixel vector is first normalized into [0, 1], expressed as

$$\tilde{\mathbf{x}}_{i} = \sum_{i=1}^{N} \frac{\mathbf{x}_{i} - \mathbf{X}_{min}}{\mathbf{X}_{max} - \mathbf{X}_{min}}$$
(1)

where N (i.e., $H \times W$) is the total number of pixel in an HSI. Next, the $\tilde{\mathbf{x}}_i$ is represented in polar coordinate system by converting the normalized vector to angular cosine and radius with the equation below:

$$\phi_j = \arccos(\tilde{\mathbf{x}}_i^j), r_j = \frac{j}{B}, j = 1, 2, \dots, B$$
 (2)

where $\tilde{\mathbf{x}}_i^j$ represents the value of j^{th} band in the normalized pixel vector $\tilde{\mathbf{x}}_i$. $\phi \in R^B$ is the angle vector and $r \in R^B$ is the radius vector.



Figure 3. A typical architecture of 1D-CNN.



Figure 4. The architecture of GAF Encoded 2D-CNN.

There are two types of GAF: Gramian Angular Summation Fields (GASF) and Gramian Angular Difference Fields (GADF), which can respectively be calculated as

$$GASF = \cos(\phi_i + \phi_j) = \tilde{\mathbf{x}}^T \tilde{\mathbf{x}} - \sqrt{I - \tilde{\mathbf{x}} \mathbf{x}^2}^T \sqrt{I - \tilde{\mathbf{x}}^2}$$
(3)

and

$$GADF = \sin(\phi_i - \phi_j) = \sqrt{I - \tilde{\mathbf{x}}^2}^T \tilde{\mathbf{x}} - \tilde{\mathbf{x}}^T \sqrt{I - \tilde{\mathbf{x}}^2}$$
(4)

where *i* and *j* denotes i^{th} and j^{th} spectral band, respectively. *I* is a unit row vector of size $1 \times B$ (i.e., the total number of spectral bands).

The resulting GASF and GADF have a size $B \times B$, where B is the number of the spectral band of HSI. The size of



Figure 5. Schematic of the proposed GAF-NAU architecture, presented on the Indian Pines dataset. The first step consists of sending a HSI pixel to the GAF encoder module and then feed the output to the proposed AG-PE (Attention Gate with Progressive Expansion) embedded U-Net architecture.

the GAF can be varied, and thus the corresponding prediction output is subject to change [10]. The conversion of 1D spectra to polar coordinates and to GASF and GADF is illustrated in Figure 2. The 1D-CNN architecture is shown in Figure 3 and the GAF encoded 2D-CNN architecture is shown in Figure 4.

3.2. Proposed GAF-NAU Method

We herein introduce a new HSI classification approach, namely Gramian Angular Field encoded Neighborhood Attention U-Net (GAF-NAU), which consists of neighbor attention mechanism, progressively expanded attention unit, and U-Net framework. The objective is to perform pixelwise classification by using the encoded GAF feature map as an input to the proposed deep learning architecture to perform pixel-wise categorization. The proposed GAF-NAU architecture is depicted in Figure 5 and illustrated in Section 3.2.2.

3.2.1 Model Overview

The main purpose of this research is to explore an alternate approach that enables deep 2D CNN-based methods to perform 1D pixel-wise classification in HSI, which has been rarely studied in the literature. In addition, we hypothesize that adding progressive expansion layer to the selfattention gating module can facilitate U-Net based deep net-



Figure 6. Illustration of the progressively expanded layer for a case of the Maclaurin series with three terms. ξ represents an intermediate feature map in the network, c_1 , c_2 , and c_3 represent first three terms' coefficients of the Maclaurin series of a nonlinear function, respectively. And p_1 , p_2 , and p_3 represent the corresponding terms' powers. $S_k(k = 1, 2, 3)$ are called progressively expanded feature maps.

work to learn more discriminate features than gating-based features propagated through the skip connections. The proposed method consists of two major steps: first, each pixel in HSI is passed through a GAF transformation which combines GASF and GADF to better capture spectral correlation. Second, the encoded GAF feature representation is fed to neighborhood attention U-Net to highlight salient features while suppressing irrelevant regions in GAF to better describe pixel category. Detailed architecture is described



Figure 7. Indiana Pines hyperspectral dataset with classification maps. (a) True color composite of HSI, (b) Ground truth. Classification maps from (c) 1D-CNN, (d) dPEN, (e) Plastic-Net (f) GAF U-Net, (g) GAF-NAU.

in the following section.

3.2.2 Model Architecture

As shown in Figure 5, GAF-NAU contains three major components: the initial GAF transformation, Attention Gate with Progressive Expansion (AG-PE) block, and U-Net framework. In the AG-PE block, the PE layer is applied to 1×1 convolution (i.e., ω_X). PE layer was originally developed for 1D pixel-wise multispectral image classification, and found to be effective in learning complex data structures by generating a nonlinear boundary that may better fit the nonlinear features [20]. In this study, we embed it in the self-attention gate and further extend it to perform 2D PE on each feature map. Figure 6 illustrates an example of a 2D PE layer. For a better explanation, let ξ represents an input 2D feature map, then the k^{th} progressively expanded feature map, denoted as S_k , can be expressed as

$$S_k = \sum_{k=1}^K c_k \xi^{p_k} \tag{5}$$

where c_k and p_k are the coefficient and power of the k^{th} term in the Maclaurin series of a nonlinear function (e.g., arctan). K denotes the total number of terms in the Maclaurin series is used. We set K = 2 in this study. The coefficient and power of expanded terms are applied to input feature maps in an element-wise fashion (every node in the hidden layer). The purpose of introducing the PE layer in the self-attention gate is to contract robust feature maps using a nonlinear function with its corresponding Maclaurin series to better fit complex data structures. Since the

PE layer does not contain trainable parameters, less model memory is consumed.

The AG-PE block is used to extract features not only from the same level of the encoder and decoder path in U-Net, but also from the neighbor's levels (upper and lower) as shown in Figure 5, and thereby we named our method as neighborhood attention. The output of the AG-PE block is element-wise multiplication of input feature-maps and attention coefficients (i.e., $\alpha \in [0, 1]$): $\tilde{X}^l = X^{l-1} \cdot X^l \cdot X^{l+1} \cdot \alpha^l$ where X^l corresponds to the feature map in layer l. In Figure 5, the gating signal gis used to determine focus regions, and it contains contextual information to trim lower-level feature responses [17]. To obtain the gating coefficient, we employ additive attention [2, 17], computed by

$$q_{att}^{l} = \psi^{T}(\sigma_{1}(PE(\omega_{X^{l-1}}^{T}X^{l-1}) + PE(\omega_{X^{l}}^{T}X^{l}) + PE(\omega_{X^{l+1}}^{T}X^{l+1}) + \omega_{q}^{T}g + b_{g})) + b_{\psi}$$
(6)

$$\alpha^{l} = PE(\sigma_{2}(q_{att}^{l}(X^{l-1}, X^{l}, X^{l+1}, g; \Theta_{att})))$$
(7)

where a ReLU and Sigmoid activation functions are represented as σ_1 and σ_2 , respectively. *PE* denotes progressive expansion operation. The set of parameters such as convolutions (ω_X and ω_g) and bias terms (b_g and b_{ψ}) are represented by Θ_{att} . A more detailed explanation of similar terms can be found in [17]. The proposed AG-PE block is incorporated into the modified U-Net architecture to highlight important features taken from the skip connections from the encoder to the decoder path as shown in Figure 5. The number of filter is set to 128 in the beginning and double it through the encoder path, and downsampling by 2 in the decoder path.

The final layer of GAF-NAU produces a 2D output map where the values will be ranging from [1 C] and C denotes the total number of classes contained in the dataset. To determine the predicted label to the input pixel vector, Majority Voting (MV) strategy is adopted. following the MV rule, the majority class label from the output is considered as the final predicted label.

4. Experimental Results



Figure 8. University of Pavia scene hyperspectral dataset with classification maps. (a) True color composite of HSI, (b) Ground truth. Classification maps from (c) 1D-CNN, (d) dPEN, (e) Plastic-Net (f) GAF U-Net, (g) GAF-NAU.

4.1. Datasets

The datasets used in our experiments are Indian Pines (IP), University of Pavia (UP), and Salinas Valley (SV),



Figure 9. Salinas Valley hyperspectral dataset with classification maps. (a) True color composite of HSI, (b) Ground truth. Classification maps from (c) 1D-CNN, (d) dPEN, (e) Plastic-Net (f) GAF U-Net, (g) GAF-NAU.

which are publicly available ¹ and widely used for evaluation of HSI classification methods. The details of the datasets are described as follows.

Indian Pines (IP): Indian Pines dataset [3] was acquired by an AVIRIS sensor covering the Indian Pines test area in the region of Northwestern Indiana. This dataset consists of 145×145 pixels and 224 spectral reflectance bands covering the range of 400–2500 nm. This scene is distributed as follows: two-thirds of agriculture presence and one-third of forest and/or natural perennial vegetation. There are a total of 16 classes, which are not all mutually exclusive, identified in the ground truth. By removing the water absorption

¹http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote _Sensing_Scenes

Metric	1D-CNN [9]	dPEN [20]	Plastic-Net [8]	GAF U-Net	GAF-NAU
OA	74.39	77.62	73.00	75.67	81.07
AA	76.36	81.28	63.29	71.69	74.67
κ	70.57	74.54	69.23	71.89	78.31

Table 1. Accuracy (in %) comparison on the Indian Pines dataset.

Note:	The highest	accuracy i	is highl	lighted	in	bold	font.
	0	2	0	0			

Table 2. Accuracy (in %) comparison on the University of Pavia dataset.

Metric	1D-CNN [9]	dPEN [20]	Plastic-Net [8]	GAF U-Net	GAF-NAU
OA	90.17	91.21	89.70	89.70	91.12
AA	89.37	90.54	87.76	87.70	90.49
κ	86.85	88.29	86.29	86.25	88.09

Table 3. Accuracy (in %) comparison on the Salinas Valley dataset.

Metric	1D-CNN [9]	dPEN [20]	Plastic-Net [8]	GAF U-Net	GAF-NAU
OA	91.43	92.53	90.69	93.82	94.59
AA	95.06	96.48	95.06	96.57	97.00
κ	90.45	91.68	89.63	93.11	93.97

region, 200 spectral bands are remained. Figure 7 (a) and (b) shows the pseudocolor image and the ground truth map of the IP dataset, respectively.

University of Pavia (UP): As part of the Pavia scenes, the University of Pavia HSI dataset was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor over the city of Pavia, located in northern Italy. The size of the image is 610×610 with a total of 103 bands. There are nine types of land cover with a spatial resolution of 1.3 meters. Figure 8 (a) and (b) shows the pseudocolor image and the ground truth map of the UP dataset, respectively.

Salinas Valley (SV): The Salinas Valley HSI dataset was collected by the AVIRIS sensor over the Salinas Valley in California. The dataset comprises 512 by 217 pixels with a spatial resolution of 3.7-meter pixels, and contains a total of 220 spectral reflectance bands. In this dataset, the water absorption bands (20 in total) were also discarded, which left for 204 bands. The provided ground truth contains a total of 16 classes that cover regions of vegetables, bare soils, and vineyard fields. Figure 9 (a) and (b) shows the pseudocolor image and the corresponding ground truth map, respectively.

4.2. Training Protocols

Our proposed architecture is compared against other pixel-based deep learning architectures, including 1D-CNN

[9], dPEN [20] and Plastic-Net [8]. All the analyses have been placed under the same dataset split proportion for training (10%), validation (10%) and testing (80%). To assess the performance of the different methods, Overall Accuracy (OA), Average Accuracy (AA) and Kappa coefficient (κ) have been used as evaluation metrics. The proposed GAF-NAU and GAF U-Net use a fixed window size of 32×32 from the GAF matrix as an input for all the datasets. In the encoder path of GAF-NAU architecture, the number of filters in the convolutional layers is doubled after each down-sampling operation starting from 128 filters in the first convolutional layer. During the network training, the number of epochs is set to 150 epochs with a batch size of 64. And a learning rate scheduler was set, starting at 10^{-3} with constantly decreasing by a factor of $e^{-0.01}$ after each epoch.

4.3. Results

Our proposed GAF-NAU architecture is verified with the IP, UP, and SV datasets, the corresponding results shown in Tables 1, 2 and 3, respectively. It is noticeable that the proposed method outperforms the other state-of-the-art methods in the majority of experiments. For instance, GAF-NAU outperforms all competing methods in the IP dataset in terms of OA and κ . For the UP dataset, our method produces competitive performance compared to the dPEN al-

gorithm. In the case of the SV dataset, our method again yields the best classification accuracy over OA, AA and κ . It can be also observed that GAF-NAU consistently outperforms GAF U-Net for all three datasets, indicating the contribution of AG-PE block to the network performance. In terms of training speed, our proposed model takes longer time to train compared to other competing methods due to computational complexity of our algorithm. For instance, our method requires approximately 11.9 mins, 50.2 mins and 26.7 mins training time for IP, UP and SV datasets, respectively. In future work, we will explore strategies to improve its computational efficiency while maintaining good classification accuracy.

The classification maps for IP, UP and SV datasets are demonstrated in Figures 7, 8 and 9, respectively. It can be visually observed that the classification maps produced from our GAF-NAU method tend to be less noisy and smoother in the majority of cases, which indicates better performance in handling interclass and/or intraclass variations. For instance, in Figure 7, it can be seen that other competing methods incorrectly classify Soybean-notil as other classes, while our approach alleviates this problem at a certain level.

4.4. Ablation Study

In this section, we conduct a set of ablation studies to understand the contribution of various aspects of our proposed model. In Table 4, we evaluate the performance of three different models: 1) GAF-NAU, 2) GAF-NAU without the use of PE block in the AG, 3) GAF-NAU without the use of neighborhood AG-PE. As demonstrated in this table, the performance of the GAF-NAU model is decreased in the absence of any of those components, although the number of trainable parameters are reduced.

Table 5 presents OA on each dataset with different GAF matrix sizes, and it is observed that the GAF matrix size of 32×32 seems a reasonable choice considering OA obtained from all three datasets. There is a slight decrement of OA as the GAF matrix size increases for the IP dataset, while it is a no certain rule applied for both UP and SV datasets. However, it can be observed that the training time increases as the GAF matrix size increases.

Furthermore, we verify the effect of increasing the training size and then evaluate the performance of our proposed model as shown in Table 6. As expected, the accuracy gradually increases as the number of training samples increases, which echos with typical deep learning models on the number of training samples requirement for achieving better performance.

5. Conclusion

We proposed a new pixel-wise hyperspectral image classification framework, which we refer to as Gramian An-

Table 4. Model performance (overall accuracy in %) and trainable parameters by varying components of the proposed GAF-NAU architecture.

Model	IP	UP	SV	#Params
GAF-NAU	81.07	91.12	94.59	158.1 M
GAF-NAU w/o PE	79.91	90.63	93.98	154.0 M
GAF-NAU w/o AG-PE	75.67	89.70	93.82	23.7 M

Table 5. Model performance (overall accuracy in %) and training time by varying GAF matrix size for three different datasets.

GAF				Training
matrix size	IP	UP	SV	Time (s)
GAF-NAU (16 x 16)	82.01	89.10	93.44	2130.69
GAF-NAU (32 x 32)	81.07	91.12	94.59	3012.11
GAF-NAU (48 x 48)	79.73	90.99	94.79	4872.69

Table 6. Model performance (overall accuracy in %) by varying the fraction of training samples.

Dataset	10% data	15% data	20% data
IP	81.07	84.91	85.98
UP	91.12	91.54	91.77
SV	94.59	95.31	95.46

gular Field encoded Neighborhood Attention U-Net (GAF-NAU). It combines multiple unique components, including GAF encoding, neighborhood attention gate, and progressive expansion layer, to ensure good feature representation. The experimental evaluation on three HSI datasets confirms the efficacy of the proposed network. Moreover, an ablation study was performed to investigate characteristics of the proposed GAF-NAU, which we found that 1) GAF matrix size affects classification output, and 2) Attention Gate with Progressive Expansion (AG-PE) has a noticeable contribution to the classification performance. It is anticipated that further gains in accuracy of GAF-NAU may be obtained by more detailed tuning of hyperparameters and proper combination of different network components, we plan to conduct such experiments in our future work.

Acknowledgement

Research reported in this publication was supported in part by funding provided by the National Aeronautics and Space Administration (NASA), under award number 80NSSC20M0124, Michigan Space Grant Consortium (MSGC).

References

- Nicolas Audebert, Bertrand Le Saux, and Sébastien Lefèvre. Deep learning for classification of hyperspectral data: A comparative review. *IEEE geoscience and remote sensing magazine*, 7(2):159–173, 2019. 1, 2
- [2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014. 5
- [3] Marion F. Baumgardner, Larry L. Biehl, and David A. Landgrebe. 220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3, Sep 2015. 6
- [4] KS Charmisha, V Sowmya, and KP Soman. Dimensionally reduced features for hyperspectral image classification using deep learning. In *International Conference on Communications and Cyber Physical Engineering 2018*, pages 171–179. Springer, 2018. 2
- [5] 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
- [6] Renlong Hang, Zhu Li, Qingshan Liu, Pedram Ghamisi, and Shuvra S Bhattacharyya. Hyperspectral image classification with attention-aided cnns. *IEEE Transactions on Geoscience* and Remote Sensing, 59(3):2281–2293, 2020. 2
- [7] Siyuan Hao, Wei Wang, and Mathieu Salzmann. Geometryaware deep recurrent neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 59(3):2448–2460, 2020. 2
- [8] 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, 7
- [9] 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, 7
- [10] 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, 4
- [11] 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
- [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] 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
- [14] Liang Liang, Liping Di, Lianpeng Zhang, Meixia Deng, Zhihao Qin, Shuhe Zhao, and Hui Lin. Estimation of crop lai using hyperspectral vegetation indices and a hybrid inver-

sion method. *Remote Sensing of Environment*, 165:123–134, 2015. 1

- [15] Moule Lin, Weipeng Jing, Donglin Di, Guangsheng Chen, and Houbing Song. Context-aware attentional graph u-net for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2021. 2
- [16] Xiaoguang Mei, Erting Pan, Yong Ma, Xiaobing Dai, Jun Huang, Fan Fan, Qinglei Du, Hong Zheng, and Jiayi Ma. Spectral-spatial attention networks for hyperspectral image classification. *Remote Sensing*, 11(8):963, 2019. 2
- [17] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999, 2018. 5
- [18] Arati Paul and Sanghamita Bhoumik. Classification of hyperspectral imagery using spectrally partitioned hyperunet. *Neural Computing and Applications*, 34(3):2073– 2082, 2022. 2
- [19] Paheding Sidike, Vijayan K Asari, and Vasit Sagan. Progressively expanded neural network (pen net) for hyperspectral image classification: A new neural network paradigm for remote sensing image analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146:161–181, 2018. 1
- [20] 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. 2, 5, 7
- [21] Zhiguang Wang and Tim Oates. Encoding time series as images for visual inspection and classification using tiled convolutional neural networks. In Workshops at the twenty-ninth AAAI conference on artificial intelligence, 2015. 3
- [22] Hao Wu and Saurabh Prasad. Convolutional recurrent neural networks forhyperspectral data classification. *Remote Sensing*, 9(3):298, 2017. 2
- [23] Naoto Yokoya, Jonathan Cheung-Wai Chan, and Karl Segl. Potential of resolution-enhanced hyperspectral data for mineral mapping using simulated enmap and sentinel-2 images. *Remote Sensing*, 8(3):172, 2016.
- [24] Shiqi Yu, Sen Jia, and Chunyan Xu. Convolutional neural networks for hyperspectral image classification. *Neurocomputing*, 219:88–98, 2017. 2
- [25] P WT Yuen and Mark Richardson. An introduction to hyperspectral imaging and its application for security, surveillance and target acquisition. *The Imaging Science Journal*, 58(5):241–253, 2010.
- [26] Lin Zhu, Yushi Chen, Pedram Ghamisi, and Jón Atli Benediktsson. Generative adversarial networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 56(9):5046–5063, 2018. 2