SIMPLE LINEAR LAYERS LEAD A STRONG BASELINE FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Ringo S.W. Chu

University College London

Wayne Luk

Imperial College London

ABSTRACT

Hyperspectral images (HSIs) classification has been widely employed in remote sensing applications, due to their proficiency in using hundreds of spectral bands to distinguish material precisely. While existing learning based approaches such as Convolutions, RNN or Transformer have achieved vast success, we demonstrate that a simple multi-layer perceptron (MLP) baseline can excel in this task. This is attributed to the fact that HSI classification primarily involves detecting unique spectral peak patterns, thus a simple model can already be well-suited. Our experimental results validate that our MLP outperforms well-known models on four datasets. Furthermore, we conduct analysis to understand the effect by modifying various elements in the model.

Index Terms— Hyperspectral Image, Classification

1. INTRODUCTION

We study Hyperspectral Image (HSI) Classification, a technique that unveils the unseen beyond our human visible spectrum. Unlike traditional images comprise of three spectral bands (Red, Green, Blue), HSI can extend up to few hundreds bands. Each pixel in an HSI contains a spectral signals, with multiple signal peaks correspond to specific wavelengths at which materials exhibit significant reflectance or absorption. HSIs are commonly used in satellite remote sensing, to oil spill detection, early cancer diagnosis and environmental monitoring.

Many of the most notable models for HSI classification leverage machine learning techniques, as documented in [1, 2]. Including but not limited to Support Vector Machine, Markov random fields, Convolutions, Recurrent Networks, Transformers. In addition, current research also explores varying learning paradigms for this task, such as Contrastive Learning [3] and Few shot Learning [4]. These models marked a significant milestones in this field. Nonetheless, this research aims to challenge the necessity of employing such complex models for achieving effective HSI classification. It prompts the question: Is it feasible to design a simpler model for this task?

Therefore, we simplify our approach to the most basic settings. We showcase that effective HSI classification can be accomplished using only Multilayer Perceptron (MLP) layers. Drawing inspiration from [5, 6], HSI classification is not more complicating than identifying spectral similarity and peak patterns within each classification class. All we need is to perform regression over the spectral channel of HSI pixels with MLP layers, thereby allowing it to learn intricate spectral patterns.

By employing a simple Multilayer Perceptron (MLP) baseline, our model not only achieves state-of-the-art quantitative performance, but also demonstrates efficiency. To validate our claims, we conducted comprehensive evaluations on four datasets: Indian Pines, Salinas, Kennedy Space Centre, and Botswana. The results demonstrate that our model surpasses convolutional, RNN, and Transformer models in terms of both accuracy and efficiency. Additionally, We performed analysis on model component to gain an intuitive understanding of our method.

Our research has substantial potential for practical applications in the field of HSIs. The inherent simplicity of our proposed model has the potential to streamline hardware deployment processes, similar to the approach presented in [7]. We provide discussions on related work, proposed methodology, and experimentation in the following sections.

2. RELATED WORK

2.1. Hyperspectral Image Classification

In this section, we summarize some of the works conducted in HSI classification. Chen *et al.* [8] were among the pioneers to explore MLP for HSI classification. Their approach involved applying PCA to reduce the spectral dimension of the HSI cube, followed by MLP layers for classification. Subsequently, a large body of research focused on utilizing 3D Convolutions (3D CNN), due to their ability to jointly process spatial and spectral dimensions. For instance, Chen *et al.* [9] employed 3D CNN on HSIs with virtual samples to address the issue of limited training data. Li *et al.* [10] utilized a new 3D CNN architecture conceptually similar to [9], but with smaller kernels to reduce model size and enhance accuracy. Zhong *et al.* [11] incorporate fully residual connection to reach performance gain. Ahmad *et al.* [12] extended the work of [8] by applying 3D CNN to PCA-reduced HSIs. Roy *et al.* [13] proposed a hybrid architecture that combines 3D and 2D CNNs to facilitate dual spectral-spatial feature learning. He *et al.* [14] introduced multi-scale CNN kernel as an efficient method to mitigate the need for spectral reduction algorithms. Liu et al. [7] optimized the approach in [15] by employing fixed kernel sizes throughout the CNN model, enabling efficient hardware deployment.

Unlike Convolution, sequential models such as RNNs or Transformer offer a global perspective for processing HSI spectral information. For examples, Mou *et al.*[16] adopt a sequential approach to HSI classification by treating each spectral signal as a sequence vector, using Gated Recurrent Unit with customised activation functions. The remarkable success of Transformers in vision domain has led to their adaptation for HSI as well. For example, Hong *et al.* [17] employ Transformer layers to compute self-attention across the whole spectral dimension. In addition, they introduce group-wise spectral embedding and cross-layer adaptive fusion as extensions to the Transformer backbone, enabling the model to learn more local spectral representations.

2.2. Discussion

The current body of work in HSI classification has primarily focused on the use of 3D CNNs. While these models have demonstrated their ability to model spatial-spectral information and extract discriminative features from HSIs, they often overlook global dependencies, particularly along the spectral dimension. Additionally, the use of 3D convolutions adds complexity to the model pipeline. Attempts to alleviate this complexity through PCA dimension reduction techniques may incur information loss and degrade performance.

On the other hand, both RNNs and Transformers have the advantage of processing the entire spectral signals globally. However, RNNs cannot be easily parallelized due to their sequential nature. Transformers, while capable of capturing global dependencies, can exhibit high intra-block complexity due to self-attention computations. Furthermore, Transformers often require large amounts of labeled data due to their lack of inductive bias, raising questions about their suitability for HSI classification, especially with limited labeled samples.

In this work, we revisit the approach proposed in [8]. The key difference lies in our use of independent regression on spectral channels before classification layers. Our method can also be considered a simplified version of the Transformer, where we replace self-attention with MLP as a drop-in replacement.

3. METHOD

We present our simple MLP baseline model below. We first extract the nearest neighbor patch from the HSI. This patch is then fed into an MLP layer, which performs regression over the spectral dimension, thereby producting a latent factor. Subsequently, we flatten this latent vector and fit forward to a second MLP for classification. A detailed illustration of this model can be found in Figure 1.

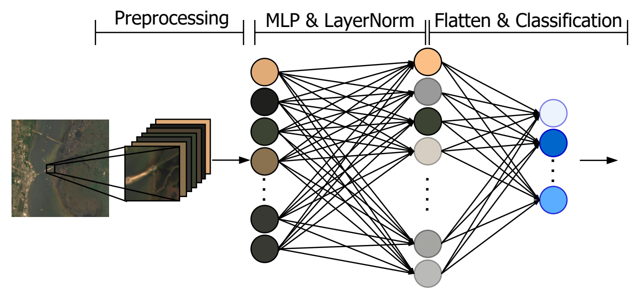

Fig. 1: Illustration of our Simple MLP Model architecture.

3.1. Pre-Processing

HSIs can be described visually as 3D cubes, where each spatial pixel represents a 1-dimensional vector containing wavelength information. Considering the high correlation and shared characteristics among neighboring pixels in HSIs, we extract 3D HSI patches instead of processing a 1D spectral vector. Given a hyperspectral cube $\mathcal{H} \in \mathbb{R}^{H\times W\times B}$, where H, W denotes height and width, B represents the number of spectral bands. We extract local patches using nearest-neighbor strategy. These patches are represented as $(x_1, x_2, ..., x_n) \in \mathbb{R}^{p\times p\times B}$, where p denotes the patch size. The truth labels of x_i are determined by the label of the central pixel within the patch.

3.2. The Simple MLP Baseline

Our simple MLP baseline, consists of only two linear layers. The primary purpose of this model is to directly regress over the spectral dimension to extract features and perform classification. The formal expressions for the model are as follows:

$$\mathbf{h}_i = \sigma(\mathsf{LayerNorm}(\mathbf{W}_1\mathbf{x}_i))$$
$$\mathbf{y}_i = \mathbf{W}_2\mathbf{h}_i$$

The first layer is shown as $\mathbf{W_1}\mathbb{R}^{B\times B} \in$, and is responsible for regressing over the spectral axis to capture signal patterns. We then incorporate LayerNorm, which promotes more stable training and reduces internal covariate shifts. σ denotes the activation function, for which we opt Gaussian Error Linear Unit as default. The second layer $\mathbf{W_1} \in \mathbb{R}^{(p\times p\times B)\times c_{cls}}$, takes the flattened latent vector from h and computes a probability logit. The output logits are then used to predict the class labels.

4. EXPERIMENT

4.1. Setup

Dataset Description. We use four benchmark datasets are used to evaluate our proposed model. These include Indian Pines, Salinas, Kennedy Space Centre (KSC) and Botswana. The spectral bands for these datasets are 200, 204, 176 and 145 respectively. For Indian Pines and Salinas, we select 10%, 5% and 85% as training, validation and testing split. For KSC and Botswana, we select 15%, 5% and 80% as training, validation and testing split.

	Indian Pines			Salinas			KSC			Botswana		
	OA	AA	κ	OA	AA	κ	OA	AA	κ	OA	AA	κ
Conv3D	75.81	77.84	71.92	93.34	96.75	92.57	80.29	77.65	77.95	95.04	95.24	94.62
SSRN	83.12	84.02	80.41	94.80	97.67	94.22	75.01	68.11	71.81	97.73	97.64	97.54
HybridSN	87.89	87.09	86.00	96.61	98.26	96.21	83.45	73.55	81.57	97.29	97.68	97.50
RNN	84.52	83.80	82.17	91.90	94.50	90.96	54.30	41.60	47.87	66.90	68.88	64.15
ViT	75.13	73.98	71.52	90.68	92.87	89.62	37.78	31.68	35.46	51.52	52.72	47.43
SpectralFormer	85.06	86.30	84.35	91.41	94.45	90.41	31.78	31.63	34.80	82.12	82.88	80.63
Ours	88.06	89.26	87.24	96.90	98.54	96.54	80.57	73.33	78.31	97.65	97.71	97.54

Table 1: Classification results for the Indian Pines, Salinas, Kennedy Space Center, and Botswana Datasets. Values highlighted in blue represent the best results, while those in red denote the second-best results. All metrics are reported as percentages (%).

Implementation. To train our MLP baseline model, we utilize cross-entropy as the loss function to capture the distribution of our dataset. We select Adam optimizer to update the weights and biases of our model, with an initial learning rate of of 0.0001, β_1=0.9 and β_2=0.999. We trained our model on a single 2080Ti, where batch size are set to 256. We set the pre-processing patch size to $p = 3$. This patch size allows us to best capture local spatial and spectral information.

Evaluation Metrics. The classification performance of each model is quantitatively evaluated using three widely used indices, all reported in percentage (%). Overall Accuracy (OA) measures the proportion of correctly classified examples over the total test examples, while Average Accuracy (AA) represents the average classification performance across all classes. Cohen's Kappa (κ) is a metric to evaluate classification results where the class distribution is imbalanced. All reported results are the average over 5 trials.

Comparison Models. To ensure a comprehensive comparison, we selected several widely-used and highly-cited baseline models from different categories. In the convolution-based model category, we have chosen *3DCNN* [10], *SSRN* [11], *HybridSN* [13]. For recurrent network-based models, we selected *RNN* [13]; In the Transformer-based model category, our choices include *ViT* [17] and *SpectralFormer* [17].

4.2. Quantitative Evaluation.

We present the comparison of our model with other baseline models in Table 1. Despite its simplicity, our model achieves competitive results and outperforms other models based on the designed metrics, except for the KSC dataset, where it is ranked second after *HybridSN*. A visualization of these results can be found in Figure 2. Botswana is omitted in the figure due to limited space.

4.3. Computational Analysis

We access the practical performance and reported ours results in Table 2. Our minimalist design has led to large reductions in MACs, parameter counts, and inference time. In theory, 3D CNNs can have a computational complexity of

Model	MACs	Inference Time	Parameter
Conv3D	0.82G	55.7s	2331K
SSRN	0.18G	151.2s	575K
HybridSN	3.04G	79.5s	6699K
RNN	0.03G	163.1s	4297K
ViT	0.02G	201.1s	105k
SpectralFormer	0.04G	84.2s	429K
Ours	0.41M	17.7s	60K

Table 2: Computational analysis across all evaluated models.

$\mathcal{O}(N^3K^3C_{in}C_{out})$, where N, K, C represent the spatial-temporal, kernel and channel dimensions. RNN and Transformer have complexities of $\mathcal{O}(Bd^2)$ and $\mathcal{O}(B^2d)$, with d denotes the hidden dimension. In contrast, our model regress directly on spectral dimension only, resulting in $\mathcal{O}(B^2)$, which makes our model more efficient to others.

4.4. More Analysis

Importance of LayerNorm. We observe the crucial role of Layer Normalization (LayerNorm). Table 3 presents the effect of removing the LayerNorm layer from the MLP, showing in a significant performance drop upon its removal. We hypothesize that LayerNorm plays a vital role in mitigating optimization weakness, particularly when the simple model has limited expressiveness and train on small sample set.

	OA	AA	κ
None	83.62	84.21	81.67
LayerNorm	88.06	89.26	87.24

Table 3: Effect of Layer Normalisation on our Model. Results are evaluated on the Indian Pines Dataset.

Normalizing Peak Signals. To validate our claim that effective HSI classification requires only detecting spectral peaks patterns, we conducted an experiment where we intentionally removed signal peaks to increase the difficulty of spectral

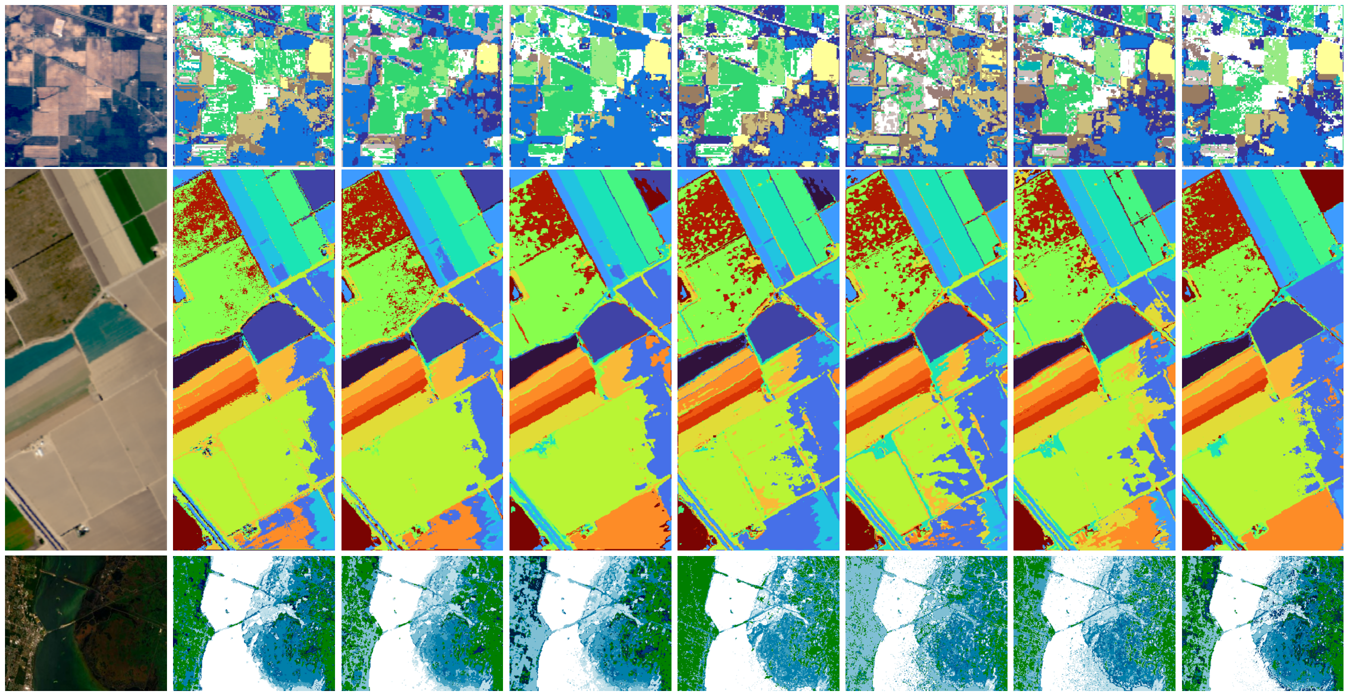

Fig. 2: Visualisation maps obtailed from all compared Models. From top to bottom: Indian Pines, Salinas, and KSC.

classification. We use two signal normalization algorithms: ① Gaussian smoothing, for suppressing high-frequency components with kernels, and ② Median filtering, which replaces outlier data point with the median value within a window.

Our assumption was that if peak signals are reduced in the HSI data, our model would struggle to classify HSI pixels accurately, leading us to resort to more complex models. We validated this assumption and reported the results in Table 4. After signal normalization, our model's performance deteriorates, while the Transformer model was still able to achieve reasonable performance, as they're designed to capture complex pattern. This experiment further supports our claim that our simple baseline model is already effective for HSI classification, as HSI more concerns detecting signal peaks pattern, where a simple model can achieve the same.

	Guassian Filtering		Median-Filtering	
	OA (%)	AA (%)	OA (%)	AA (%)
SpectralFormer	84.76	82.26	84.13	85.74
Ours	84.31	84.14	83.98	83.81

Table 4: Effects of normalizing HSI spectral signals on classification performance. Results are evaluated on Indian Pines Dataset.

Effect of Patch Sizes. We aim to explore the significance of spatial information in the context of classification. Incorporating neighboring pixels to form patches can help compensate for the absence of spatial context. We varied the patch sizes, reporting the results in Table 5. Our model demonstrated optimal performance when the patch size p was set to 3. Further increasing patch size does not improve performance, instead, it led to a deterioration in accuracy. We hypothesize that an overly large patch size introduces too much contextual variable that prevent model to coverage.

	$p = 1$	$p = 3$	$p = 5$	$p = 8$
OA (%)	82.84	89.15	84.73	84.84
AA (%)	83.78	88.64	85.84	82.94
κ (%)	80.29	87.42	82.34	82.63

Table 5: Effect of Patch Sizes. Results are evaluated on Indian Pines Dataset.

5. LIMITATION & CONCLUSION

Limitations. A limitation of our study is that we did not account for class imbalances, which occurs frequently in HSI datasets. For instance, natural elements often outnumber man-made structures in satellite imagery, which can affect the model's performance.

Conslusion. In summary, we have demonstrated that a simple Multilayer Perceptron (MLP) can surpass the performance of established models for HSI classification. Our findings prompt further inquiries into what constitutes an effective HSI classification model. We hope that our research will contribute to enhancing the practical applicability of these models in real-world applications.

6. REFERENCES

[1] Antonio Plaza, Jon Atli Benediktsson, Joseph W. Boardman, Jason Brazile, Lorenzo Bruzzone, Gustavo Camps-Valls, Jocelyn Chanussot, Mathieu Fauvel, Paolo Gamba, Anthony Gualtieri, Mattia Marconcini, James C. Tilton, and Giovanna Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sensing of Environment*, vol. 113, 2009, Imaging Spectroscopy Special Issue.

[2] 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*, vol. 57, no. 9, pp. 6690–6709, 2019.

[3] Xin Huang, Mengjie Dong, Jiayi Li, and Xian Guo, "A 3-d-swin transformer-based hierarchical contrastive learning method for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[4] Bing Liu, Xuchu Yu, Anzhu Yu, Pengqiang Zhang, Gang Wan, and Ruirui Wang, "Deep few-shot learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 2290–2304, 2019.

[5] O Abilio De Carvalho and Paulo Roberto Meneses, "Spectral correlation mapper (scm): an improvement on the spectral angle mapper (sam)," in *Summaries of the 9th JPL Airborne Earth Science Workshop, JPL Publication 00-18*. JPL publication Pasadena, CA, USA, 2000, vol. 9, p. 2.

[6] Xiaofang Liu and Chun Yang, "A kernel spectral angle mapper algorithm for remote sensing image classification," in *2013 6th International Congress on Image and Signal Processing (CISP)*, 2013, vol. 2, pp. 814–818.

[7] Shuanglong Liu et al., "Optimizing CNN-Based Hyperspectral Image Classification on FPGAs," in *International Symposium on Applied Reconfigurable Computing*, 2019.

[8] Yushi Chen et al., "Deep Learning-Based Classification of Hyperspectral Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.

[9] Yushi Chen et al., "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232–6251, 2016.

[10] Ying Li, Haokui Zhang, and Qiang Shen, "Spectral–spatial classification of hyperspectral imagery with 3d convolutional neural network," *Remote Sensing*, vol. 9, no. 1, 2017.

[11] Zilong Zhong et al., "Spectral–Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 847–858, 2018.

[12] Muhammad Ahmad, Adil Mehmood Khan, Manuel Mazzara, Salvatore Distefano, Mohsin Ali, and Muhammad Shahzad Sarfraz, "A fast and compact 3-d cnn for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[13] Swalpa Kumar Roy, Gopal Krishna, Shiv Ram Dubey, and Bidyut B Chaudhuri, "Hybridsn: Exploring 3d-2d cnn feature hierarchy for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, 2020.

[14] Mingyi He, Bo Li, and Huahui Chen, "Multi-scale 3d deep convolutional neural network for hyperspectral image classification," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3904–3908.

[15] Anirban Santara et al., "BASS Net: Band-Adaptive Spectral-Spatial Feature Learning Neural Network for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 9, pp. 5293–5301, 2017.

[16] Lichao Mou, Pedram Ghamisi, and Xiao Xiang Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3639–3655, 2017.

[17] Danfeng Hong, Zhu Han, Jing Yao, Lianru Gao, Bing Zhang, Antonio Plaza, and Jocelyn Chanussot, "Spectralformer: Rethinking hyperspectral image classification with transformers," 2022, vol. 60, pp. 1–15.