Article

An Efficient Attention-Based Convolutional Neural Network That Reduces the Effects of Spectral Variability for Hyperspectral Unmixing

Baohua Jin, Yunfei Zhu, Wei Huang *, Qiqiang Chen and Sijia Li

College of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China
* Correspondence: hnhw235@zzuli.edu.cn; Tel.: +86-182-3714-0566

Abstract: The purpose of hyperspectral unmixing (HU) is to obtain the spectral features of materials (endmembers) and their proportion (abundance) in a hyperspectral image (HSI). Due to the existence of spectral variabilities (SVs), it is difficult to obtain accurate spectral features. At the same time, the performance of unmixing is not only affected by SVs but also depends on the effective spectral and spatial information. To solve these problems, this study proposed an efficient attention-based convolutional neural network (EACNN) and an efficient convolution block attention module (ECBAM). The EACNN is a two-stream network, which is learned from nearly pure endmembers through an additional network, and the aggregated spectral and spatial information can be obtained effectively with the help of the ECBAM, which can reduce the influence of SVs and improve the performance. The unmixing network helps the whole network to pay attention to meaningful feature information by using efficient channel attention (ECA) and guides the unmixing process by sharing parameters. Experimental results on three HSI datasets showed that the method proposed in this study outperformed other unmixing methods.

Keywords: hyperspectral unmixing; convolutional neural network; endmember bundle; spectral variability; attention mechanism

1. Introduction

Hyperspectral images (HSIs) have many applications. When they are used in remote sensing, for instance, different ground objects can be distinguished [1]. Often, HSI pixels contain reflections from many different ground objects and are then referred to as mixed pixels. The presence of mixed pixels will reduce the HSI processing performance [2,3]. Therefore, hyperspectral unmixing (HU) is used to obtain the spectral features and abundances of the substances (endmembers) in mixed pixels [4].

Originally, a linear mixing model (LMM) based on the photon interaction mechanism at work in the target object was adopted for HU [5]. An LMM is very interpretable. Vertex component analysis (VCA) [6] and N-FINDR [7] are representative models too, but mixed pixels are common in different scenes; therefore, there are limitations to these methods [8]. A nonlinear mixing model (NLMM) can be considered for mixed pixels. In principle, an NLMM needs to consider more complex factors. Although the traditional unmixing method shows excellent performance [9–11], outliers and high noise distortions can cause it to lose a lot of detailed information from HSIs during dimensionality reduction [12]. In addition, there is a large amount of redundant information in HSIs, which increases processing difficulties.

At present, there are many LMM-based methods for HU following the traditional view that the endmember spectrum has no spectral variability. However, this assumption is usually not valid for real datasets because the radiation or reflectivity of materials may

change significantly with changes in the environment, including changes in the illumination and atmosphere, perhaps resulting in estimation errors propagated throughout the unmixing process. Therefore, spectral variability (SV) has attracted wide attention. Fu et al. [13] proposed a dictionary adjustment method to solve the SV problem, where SV is regarded as an endmember dictionary in the spectral library that does not match the observed spectral features. In fact, some SVs are caused by additive perturbations that destroy the original pure endmember spectrum. An interference matrix can be used to model this kind of spectral variation. Thouvenin et al. [14] considered SVs as additional endmember disturbance information and developed a perturbed LMM model (PLMM) on the basis of minimum volume constrained non-negative matrix factorization (MVCNMF) [15]; however, this model lacks specific physical meaning. To clarify the physical meaning, Drumetz et al. [16] proposed the extended LMM (ELMM), which effectively simulates changes in reflectivity due to changes in lighting by multiplying the diagonal matrix and the endmember. Although the physical meaning of this method is clear, the ELMM model assumes that all wavelengths have a fixed scaling ratio, and thus, this model has some limitations when the endmembers are under the influence of a more complex environment. In a variety of complex hyperspectral scenes, an LMM's data unmixing and reconstruction ability is limited.

An NLMM is constructed from an LMM by considering specific nonlinear factors to improve the unmixing performance. Initially, the Hapke model [17], which is based on radiative transfer theory (RTT), was proposed. It uses a mathematical model to express complex nonlinear mixed phenomena and then solves the model. However, it has serious limitations, including difficulties with complex and vegetation-covered scenes. Later, in keeping with the physical meaning of the model, a simplification in the form of the bilinear mixture model (BMM) [18], which can be applied to two-layer mixed scenarios, was proposed. Fan et al. [19] made further improvements, allowing the model to tackle a variety of mixed-material scenarios. However, these models have some limitations. For example, an endmember must be extracted in advance by another algorithm before its abundance can be estimated. This issue leads to many limitations in the event of complex SV scenarios. Data-driven NLMM has also attracted much attention. Unlike the model-based NLMM, the data-driven method does not require that the nonlinear mixed form be known and only needs data to carry out endmember extraction and abundance inversion. The kernel method is a representative data-driven method. It projects the obtained original nonlinear data onto a high-dimensional space and then performs linear HU in that space. Relevant algorithms mainly include kernel fully constrained least squares (FCLS) [20,21] and non-negative matrix factorization (NMF) unmixing based on the kernel model [22,23].

Recently, deep learning (DL) has developed rapidly, and it has been widely used in computer vision and natural language processing [24,25]. DL has attracted much attention in terms of HU due to its strong feature representation and learning abilities. Initially, some basic network frameworks were applied to HU [26], but these methods required ground truth or endmember training sets with known abundances, which caused an issue given that ground truth availability is very limited. Autoencoder (AE) networks have been widely used in HU due to their characteristics and good performances [27–29]. An AE network can reconstruct the output data to find the low-dimensional representation (abundance score) of an HSI. In addition, a convolution neural network (CNN) is able to extract structural features from an HSI, and thus, it is also very suitable for HU tasks. Su et al. [30] proposed a deep autoencoder network (DAEN) for unmixing hyperspectral data with outliers. In [31], Hong proposed a framework called WU-NET, which was used to deal with SV. In addition, the two-stream autoencoder network (TANET) [32] uses super-pixel segmentation as a preprocessing technique to extract endmember bundles for two-stream autoencoder unmixing. However, during dimensionality reduction, the AE will inevitably lose feature information from an HSI. In [33], an end-to-end pixel-based CNN was proposed to solve the unmixing task, and the multilayer perceptron (MLP) structures were used to obtain the pixel abundance. In [34], Arun et al. used CNNs for HU. Among

them, a long short-term memory network was better at unmixing than a linear hybrid encoder–decoder method.

Attention-based methods originated from natural language processing (NLP). In recent years, attention mechanisms have been used in many fields, such as image classification [35–37] and target detection [38,39]. At present, attention mechanisms are shown to play a good role in capturing HSI features. Sun et al. [40] designed a successive pooling attention network for the semantic segmentation of remote sensing images. Fu et al. [41] designed a recurrent thrifty attention network for remote sensing recognition by using a self-attention mechanism. Zeng et al. [42] designed a residual network based on an attention mechanism to conduct HU for limited training samples. Zhu et al. [43] improved its performance by using a squeeze excitation (SE)-driven attention mechanism to consider the differences between optical detection and ranging LiDAR heights to guide the unmixing process. Attention-based networks have great potential for capturing features, and there is still a lot of room for exploration in this field. Hence, our network optimizes the modeling by considering the HSI feature information in the unmixing process.

This study re-examined the limitations of the nonlinear hybrid model and the existing solution mixing schemes to propose workarounds for these shortcomings. By extracting the information from the HSI with physically meaningful endmembers, the attention module can also learn the hyperspectral feature information. Accordingly, an efficient attention-based CNN for HU was proposed in this study.

The main contributions of this study are as follows:

1. This study proposed an efficient attention-based convolutional neural network called the EACNN, which simulates endmembers in a physically meaningful and self-supervised way and captures hyperspectral information effectively, allowing for the HU of complex scenes.
2. Inspired by the attention mechanism approach, an efficient convolution block attention module (ECBAM) for HU was proposed. It can effectively extract the rich spatial–spectral information of HSI.
3. A joint attention feature extraction strategy was proposed. For the HSI data, the network is only allowed to learn its useful bands for HU. On the other hand, the endmember bundles aggregate spatial information to a certain extent, and the amount of data is less than that of the original HSI; therefore, it is more efficient to extract spatial information from the endmember bundles.

The rest of the paper is structured as follows. Section 2 briefly introduces the relevant category models for advanced HU, while Section 3 provides details for related methods and the proposed EACNN network framework. Next, Section 4 validates the proposed method and evaluates the experimental results from different datasets. Section 5 discusses the above experiments. Finally, conclusions are reached in Section 6.

2. Relevant Research Works

In previous work, researchers proposed many unmixing algorithms, such as fully constrained least-squares unmixing (FCLSU) [44], graph-regularized $l_{1/2}$-NMF (GLNMF) [45], unmixing based on the graph Laplacian (GraphL) [46], the Merriman–Bence–Osher (MBO) scheme for solving a graph's total variation subproblem (gtvMBO) [47], deep autoencoder unmixing (DAEU) [27] and the endmember-guided unmixing network with pixelwise (EGU-pw) [48]. Their advantages and disadvantages are shown in Table 1.

**Table 1.** Advantages and disadvantages of different unmixing algorithms.

Method	Advantages	Disadvantages
FCLSU	The addition of the ℓ_2-norm as a regularization term effectively improves the computational efficiency	The obtained solution tends toward a local optimum solution
GLNMF	Constructs the nearest-neighbor graph of the local window to represent the structural characteristics of data	Global feature information is lost
GraphL	Introduces a normalized graph Laplacian to improve the efficiency	Improved efficiency but insufficient use of HSI feature information
gtvMBO	Graph total variation (gtv) regularization is introduced to capture the similarity between spectra and improve scalability	Regularization schemes are designed only for endmembers, leading to biased estimates of the abundance
DAEU	Spectral features are obtained via a stacked non-negative sparse autoencoder	During dimensionality reduction, feature information is lost
EGU-pw	Shared network weights improve the unmixing performance and make it more interpretable	Using the AE network, feature information is lost in the dimensionality reduction process and the captured features are limited

Although the FCLSU algorithm performs unmixing well, its final solution tends toward a local optimum solution, which is unfavorable for HSIs containing large amounts of data. The GLNMF approach hinges on converting a non-negative matrix with higher dimensions into two non-negative matrices with lower dimensions. However, when it is directly applied to the estimation of abundance, it often falls into the local minimum problem. The GraphL and gtvMBO methods improve the efficiency using the Laplacian graph operation, but their optimizations are aimed at the endmembers only, with no consideration of the abundance.

With the development of technology, researchers put forward the DAEU, which uses a neural network, with its strong ability to fit nonlinear problems and process a large amount of data. It extracts the hidden input features through its encoder and reconstructs the input through a decoder, which can achieve good results. However, during dimensionality reduction, it will lose rich HSI feature information, which greatly reduces its unmixing performance. The EGU-pw is an end-to-end, two-stream deep unmixing network that simulates the physical properties of endmembers in the real world through self-supervision technology. Although it produces excellent results, it also ignores the rich HSI feature information.

3. Proposed Method

Figure 1 shows the basic EACNN framework, including its endmember network (EN) and unmixing network (UN). First, through an effective clustering method, the required pseudo-pure endmember bundles are obtained by aggregating the HSI feature information. Next, the EN maps the pseudo-pure endmember bundles to the network layer, which obtains the global spatial and spectral information in the HSI through an efficient convolutional block attention module (ECBAM). The UN obtains the spectral information in the HSI that is useful for learning via effective channel attention (ECA) [49]. Finally, the EACNN uses a parameter-sharing strategy to make the two networks communicate closely, allowing the EN to embed the inherent physical properties of the endmembers into the UN, the UN to feed back its information to guide the EN, and the two sides to promote each other in such a manner as to make the whole network learn in a more accurate and reliable way.
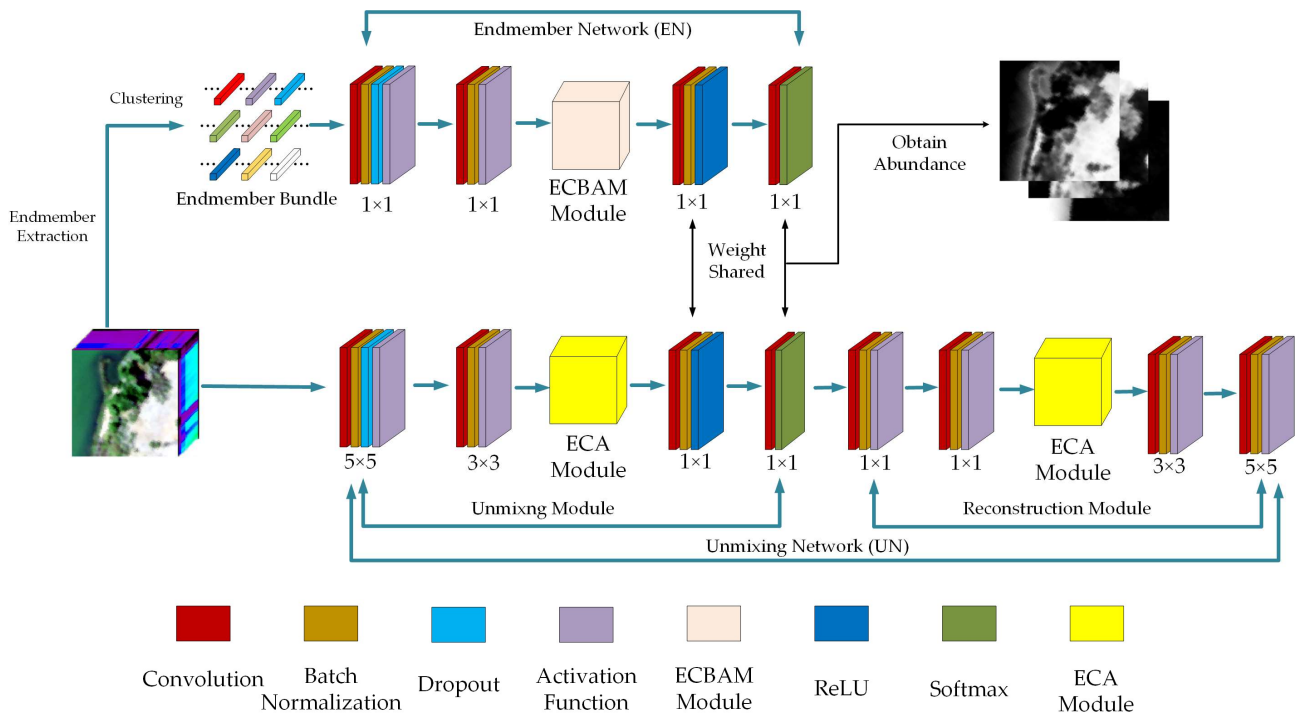
Figure 1. Network architecture of the proposed efficient attention-based convolutional neural network (EACNN).

Next, the proposed EACNN framework is described in detail.

3.1. Endmember Network

In the traditional blind unmixing process, an unsupervised unmixing task can be accomplished by adding an abundance non-negativity constraint (ANC) and abundance sum-to-one constraint (ASC) to the network; however, the accuracy and robustness will be limited and no clear physical meaning can be assigned to the endmembers. Therefore, the EN learns the physical properties of endmembers by using pure or relatively pure pseudo-endmember bundles as the input. Inspired by [50], the endmember bundles required for the EN can be obtained via the following steps.

Based on [51–53], the spectral characteristics of adjacent pixels are highly correlated, which indicates that pure spectral pixels are more likely to appear in areas with uniform spatial distributions. First, the HSI is randomly divided into partially overlapping blocks, with the number of partitions set according to [52]. Then, the number of endmembers is automatically estimated using the HySime algorithm, and the endmembers are extracted from each block via VCA. Finally, the repeated endmembers are removed using the K-means clustering algorithm, and the extracted endmembers are aggregated into K clusters. According to the experiment, the K value should be set to about 20% of the pixels in the HSI. The participation of the extracted endmember bundles in unmixing can both clarify the physical meaning of unmixing and effectively reduce the influence of SV on the whole-network unmixing process, which is more conducive to the accurate estimation of the abundances in the HSI.

The endmember bundle input extracted by this method is defined as $\{x_i\}_{i=1}^{P_e} \in \mathbb{R}^B$, with B bands consisting of P_e pixels, the pure abundance $\left\{y_{i=1}^{N_e}\right\} \in \mathbb{R}^C$ has C categories and the ith endmember of the lth EN is defined as $t_i^{(l)}$, which is expressed as

$$t_i^{(l)} = \begin{cases} g\left(W_{EN}^{(l)}, B_{EN}^{(l)}, x_i\right), & l = 1 \\ g\left(W_{EN}^{(l)}, B_{EN}^{(l)}, t_i^{(l-1)}\right), & l = 2, \ldots, k \end{cases} \tag{1}$$

where $g(\cdot)$ is the nonlinear activation function. W_{EN} and B_{EN} represent the weights and biases of each layer.

As shown in Figure 1, the EN convolution layer uses the 1×1 convolution core. After each output of the convolution layer, a batch normalization (BN) layer is used. The output of the BN layer is

$$t_{BN_i}^{(1)} = \alpha \hat{t}_i^{(1)} + \beta \tag{2}$$

where $\hat{E}_i^{(1)}$ is the z-score of $E_i^{(1)}$ and α and β represent the parameters for network learning.

After the first convolution layer, a dropout layer is used to effectively alleviate overfitting, remove noise and diminish the SV to a certain extent to enhance the generalization ability of the model. The output of the dropout layer is denoted as $E_{D_i}^{(1)}$. An effective convolutional attention module is used for unmixing. As shown in Figure 2, the ECBAM consists of two attention modules, namely, the efficient channel attention and spatial attention modules. First, efficient channel attention produces the input. To effectively aggregate the spatial feature information, it is implemented using a fast 1D convolution of size k, where the kernel size k represents the coverage of local cross-channel interactions and is confirmed using an adaptive method.
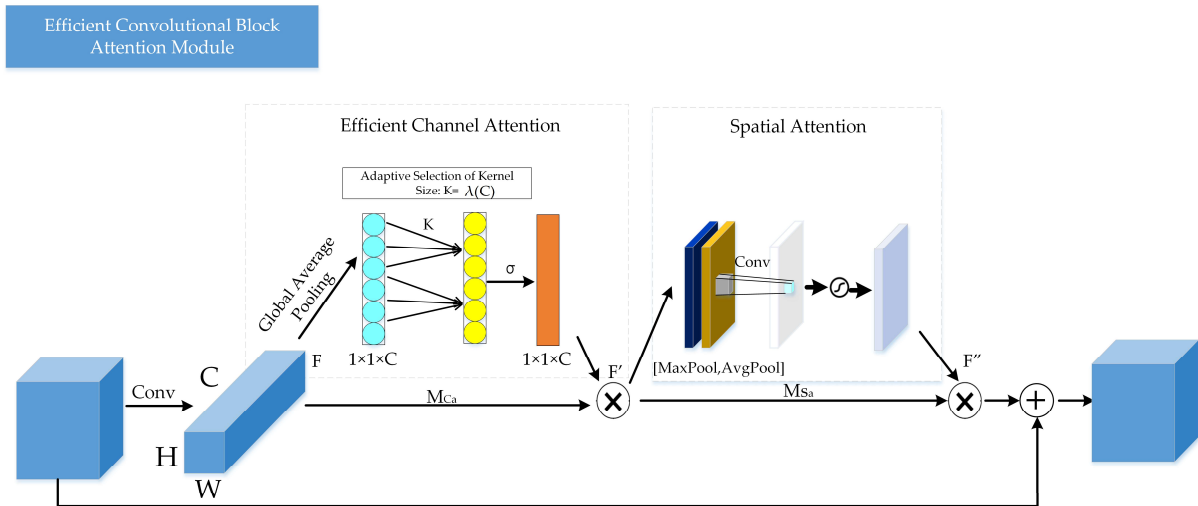

Figure 2. Schematic diagram of the efficient convolutional block attention module (ECBAM).

Given an intermediate feature map $F \in \mathbb{R}^{W \times H \times C}$ as the input, where W, H and C represent the width, height and channel dimension, because our goal was to obtain useful spectral information from HSIs by capturing local cross-channel interaction, we only focused on the interaction between each channel and its k neighbors. Therefore, the weight of y_i is calculated using

$$w_i = \sigma \left(\sum_{j=1}^{k} \alpha_i^j y_i^j \right), y_i^j \in \Omega_i^k \tag{3}$$

where Ω_i^k represents the set of k channels adjacent to y_i. By capturing the local feature information across channels, only the part of the network between all channels needs to be learned so that the overall efficiency is very high. For this operation, the attention of each channel includes the parameters of $k \times C$. At the same time, in order to further reduce the complexity of the unmixing module, all channels have the same parameters, which are expressed as follows:

$$w_i = \sigma \left(\sum_{j=1}^{k} \alpha^j y_i^j \right), y_i^j \in \Omega_i^k \tag{4}$$

In general, ECA is accomplished using a 1D convolution with a kernel size of k, which is expressed as follows:

$$M_{Ca}(F) = \sigma(C1D(GAV(F))) \tag{5}$$

where C1D indicates a 1D convolution and GAV denotes the global average pooling.

The kernel size k determines the interaction coverage captured, which is adaptive to the channel dimension C. The mapping φ between the kernel size k and channel C is expressed as follows:

$$C = \varphi(k) \tag{6}$$

The mapping φ is known from [49], and k is nonlinearly proportional to C. The approximate mapping φ is mapped using the following exponential function:

$$C = \varphi(k) = 2^{(\varepsilon * k - b)} \tag{7}$$

Finally, the value of k can be obtained as follows:

$$k = \lambda(C) = \left| \frac{\log_2(C)}{\varepsilon} + \frac{b}{\varepsilon} \right|_{odd} \tag{8}$$

where $|v|_{odd}$ represents the odd number closest to v.

The spatial attention map is generated with the spatial attention module. First, average pooling and max pooling operations are applied along the channel axis, and then average and max operations are performed on the input features in the channel dimension. Two 2D features are obtained then a hidden layer containing a single convolution kernel is used to convolute them. Finally, these features are concatenated together according to the channel dimension to obtain a 2D spatial attention map. The specific calculation used in the spatial attention module is as follows:

$$\begin{aligned} M_{Sa}(F) &= \sigma\Big(f^{7\times7}([\text{AvgPool}(M_{Sa}(F)); \text{MaxPool}(M_{Ca}(F))])\Big) \\ &= \sigma\Big(f^{7\times7}\Big(\Big[F_{avg}^{Sa}; F_{max}^{Sa}\Big]\Big)\Big) \end{aligned} \tag{9}$$

where σ denotes the sigmoid function and $f^{7\times7}$ denotes the convolution operation with a filter size of 7×7.

Overall, for a given intermediate feature $F \in \mathbb{R}^{H \times W \times C}$, the ECBAM can be generalized as follows:

$$\begin{aligned} F' &= M_{Ca}(F) \otimes F \\ F'' &= M_{Sa}(F') \otimes F' \end{aligned} \tag{10}$$

The last nonlinear activation function of the first two convolution blocks is defined as $a_i^{(1)}$, which can be expressed as follows:

$$a_i^{(1)} = g\left(t_{Drop_i}^{(1)}\right) \tag{11}$$

Next, we imposed ASC constraints on the last two convolution blocks through the ReLU layer as follows:

$$a_{ReLU_i}{}^{(1)} = g_{ReLU}(t_{BN_i}^{(1)}) = \max(0, t_{BN_i}^{(1)}) \tag{12}$$

The ANC constraint was imposed through the softmax layer:

$$a_{Soft_i^{(1)}} = g_{Soft}\left(t_i^{(1)}\right) = \frac{e^{t_i^{(1)}}}{\sum\limits_{j=1}^{C} e^{t_j^{(1)}}} \tag{13}$$

The cross-entropy is used to measure the EN loss, which can be expressed as follows:

$$L_{EN} = \frac{-1}{N_e} \sum_{i=1}^{N_e} \left[y_i \log a_{Soft_i} + (1 - y_i) \log(1 - a_{Soft_i}) \right] \tag{14}$$

The results obtained when applying just the ANC and ASC constraints in the blind unmixing method were not very satisfactory because the blind unmixing network was prone to producing no physical noise when dealing with conditions such as noise, unknown materials and meaningful spectral features. According to previous experiments, the endmember bundles can effectively guide the unmixing process to derive physical meanings for the endmembers. Embedding the network to guide the UN unmixing process should help it to obtain more accurate representations of the abundance.

3.2. Unmixing Network

The UN structure is roughly similar to that of the EN because the UN and EN can share weights in the unmixing process, allowing the attributes of the endmembers to be fully taken into account in the unmixing process. The UN is made up of two similar parts: the unmixing and reconstruction modules.

In order to effectively share the information obtained by the EN with the UN, the sharing strategy involves learning in a partially shared fashion following the extraction of spectral feature information.

Due to the existence of many different SVs in hyperspectral data, linear activation functions cannot be used in unmixing. At the same time, a linearization operation cannot fully reproduce the original spectral details of an HSI with complex SVs. Therefore, across the whole network, a large number of nonlinear activation functions are used. The UN learns by sharing some parameters. Note that the ECBAM is used in the EN because the endmember bundle input processes the spatial information. With a certain degree of aggregation, the extraction of spatial feature information for the EN is efficient, and to a certain extent, the extracted information can be guaranteed to promote network learning. At this time, the UN's extraction of spatial information from the original HSI will add factors that are irrelevant to the unmixing and may even affect the unmixing performance. Thus, the proposed method favors a balance between accuracy and efficiency to further improve the estimation of abundances. The UN and EN have similar settings; for specific settings, please refer to the EN outlined above.

Through its extraction and reconstruction of the detailed features of the HSI, the network can obtain satisfactory unmixing results. It should be noted that HSIs contain a wider range of information and a larger amount of data than natural images of the same size; therefore, lightweight modules can be used in multi-scale feature extraction. Only a few parameters can effectively improve the network performance.

The abundances of an unmixing module can be derived directly through a_{soft}. The optimization of the UN reconstruction module L_{UN} is obtained by minimizing the reconstruction error:

$$L_{UN} = \min_{w_{UN}, B_{UN}} \| X - r_m(g_u(w_{UN}, B_{UN}, X)) \|_F^2$$

$$\text{s.t.} w_{UN} = w_{EN}, \ B_{UN} = B_{EN} \tag{15}$$

where r_m and g_u correspond to the mapping functions of the unmixing and reconstruction modules with respect to the weights W_{UN} and biases B_{UN} of the unmixing module, respectively. The overall EACNN loss can be expressed as follows:

$$\text{Loss} = L_{EN} + L_{UN} \tag{16}$$

As shown in [46], the endmembers can be obtained indirectly. Further, the endmembers ($E = \{e_i\}_{i=1}^{C}$) can be estimated via a simple linear model when abundances are known:

$$\min_E \|X - Ey\|_F^2 \quad \text{s.t. } E \geq 0 \tag{17}$$

It should be noted that our model uses a large number of nonlinear activation functions, while the linear model is simply convenient for visualizing endmembers and comparing the endmembers to reference endmembers. The specific network parameters are shown in Table 2.

Table 2. Specific network details of the EACNN.

	Endmember Network	Unmixing Network
Convolution block 1	1×1 Conv Batch normalization Dropout Tanh	5×5 Conv Batch normalization Dropout Tanh
Convolution block 2	1×1 Conv Batch normalization Tanh	3×3 Conv Batch normalization Tanh
Attention block 1	Efficient convolutional block attention module	Efficient channel attention
Convolution block 3	1×1 Conv Batch normalization ReLU	1×1 Conv Batch normalization ReLU
Convolution block 4	1×1 Conv Softmax	1×1 Conv Softmax
Convolution block 5		1×1 Deconv Batch normalization Sigmoid
Convolution block 6		1×1 Deconv Batch normalization Sigmoid
Attention block 2		Efficient channel attention
Convolution block 7		3×3 Deconv Batch normalization Sigmoid
Convolution block 8		5×5 Deconv Batch normalization Sigmoid

4. Experimental Results

One synthetic and two real datasets that are commonly used in HU tasks were selected and the datasets were obtained from [54]. Through quantitative evaluation and testing, the unmixing performances of several advanced unmixing algorithms were compared with the proposed EACNN. The network used the above datasets as training and the ground truth for verification; the details were as follows:

(1) Synthetic dataset: the first dataset contained 200×200 pixels and 224 effective spectral bands, while the ground truth contained five randomly selected endmembers.

(2) Jasper Ridge dataset: this contained 100×100 pixels and 198 effective spectral bands, while the ground truth contained four endmembers: "Road", "Soil", "Water" and "Tree".

(3) Samson dataset: the last dataset contained 95×95 pixels and 156 effective spectral bands, while the ground truth contained three endmembers: "Rock", "Tree" and "Water".

4.1. Experimental Details and Evaluation Indicators

Several of the most advanced algorithms in blind unmixing, including the FCLSU, GLNMF, GraphL, gtvMBO, DAEU and EGU-pw, were compared with the EACNN.

For the existing advanced algorithm models, the optimal parameter settings given in the literature were adopted. For the proposed EACNN, the power was set to 0.99 and the dropout rate was set to 0.9. The learning rate was iteratively updated by multiplying the initial rate by $(1 - (\text{iter})/(\text{maxIter}))^{\text{power}}$. The network model ended after 300 batches.

To evaluate the unmixing performance, the abundance and endmembers were evaluated using the root-mean-square error (RMSE) and spectral angle distance (SAD), respectively. These values are defined as follows:

$$\text{RMSE}(\hat{y}_j, y_j) = \frac{1}{N} \sum_{j=1}^{N} \sqrt{\|\hat{y}_j - y_j\|_2^2} \tag{18}$$

where $\hat{y}_j$ and y_j are the estimated abundance and actual abundance, respectively;

$$\text{SAD}(\hat{e}_i, e_i) = \arccos\left(\frac{\hat{e}_i^t e_i}{\|\hat{e}_i\|_2 \|e_i\|_2}\right) \tag{19}$$

where $\hat{e}_i$ and e_i represent the extracted endmembers and reference endmembers, respectively.

4.2. Results for the Synthetic Dataset

The synthetic dataset was simulated by randomly selecting five reference endmembers from the United States Geological Survey (USGS) spectral library. The complete hyperspectral dataset contained a total of 200×200 pixels, where each pixel was recorded on 224 spectral bands from 0.4 µm to 2.5 µm. The simulated dataset contained non-Gaussian SV noise and other complex SVs caused by causes. Please refer to [55] for detailed information on the synthetic dataset.

The resulting abundance maps, endmembers and quantitative measurements obtained during unmixing are given in Figures 3 and 4 and Table 3, respectively. Thanks to the different SVs in the composite dataset, it can be clearly seen that the FCLSU algorithm performed well on the composite dataset without taking the endmember variation into consideration and that the results for the GLNMF algorithm in terms of abundance estimation and endmember extraction were poor, which may have been due to the fact that when SVs are present, the GLNMF algorithm could not be well constructed, and its abundance map has serious noise. The gtvMBO algorithm's unmixing was biased due to its non-negative endmember constraint. Compared with the traditional unmixing method, the DL-based model estimated the endmembers well, further indicating its potential. Although the results of the two endmembers extracted by the EACNN were a little different from those extracted by the EGU-pw, its performance indicators revealed that EACNN performed very well. Its stability and effectiveness were demonstrated, and it produced more accurate unmixing results in blind HU tasks.
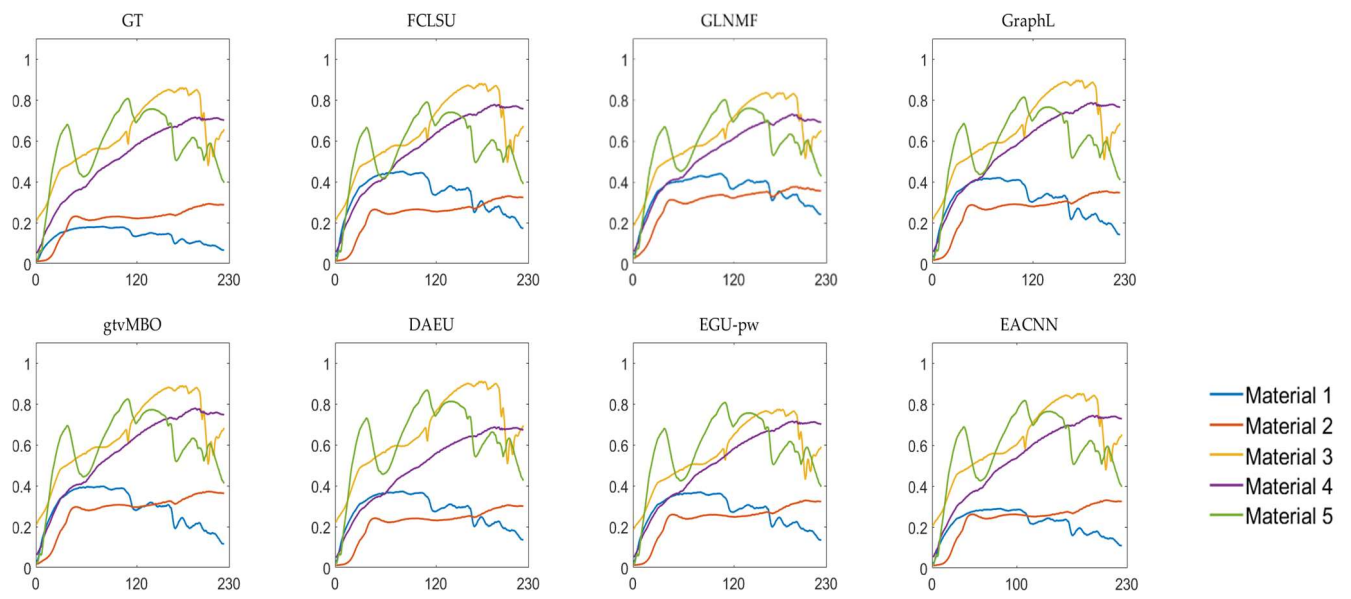
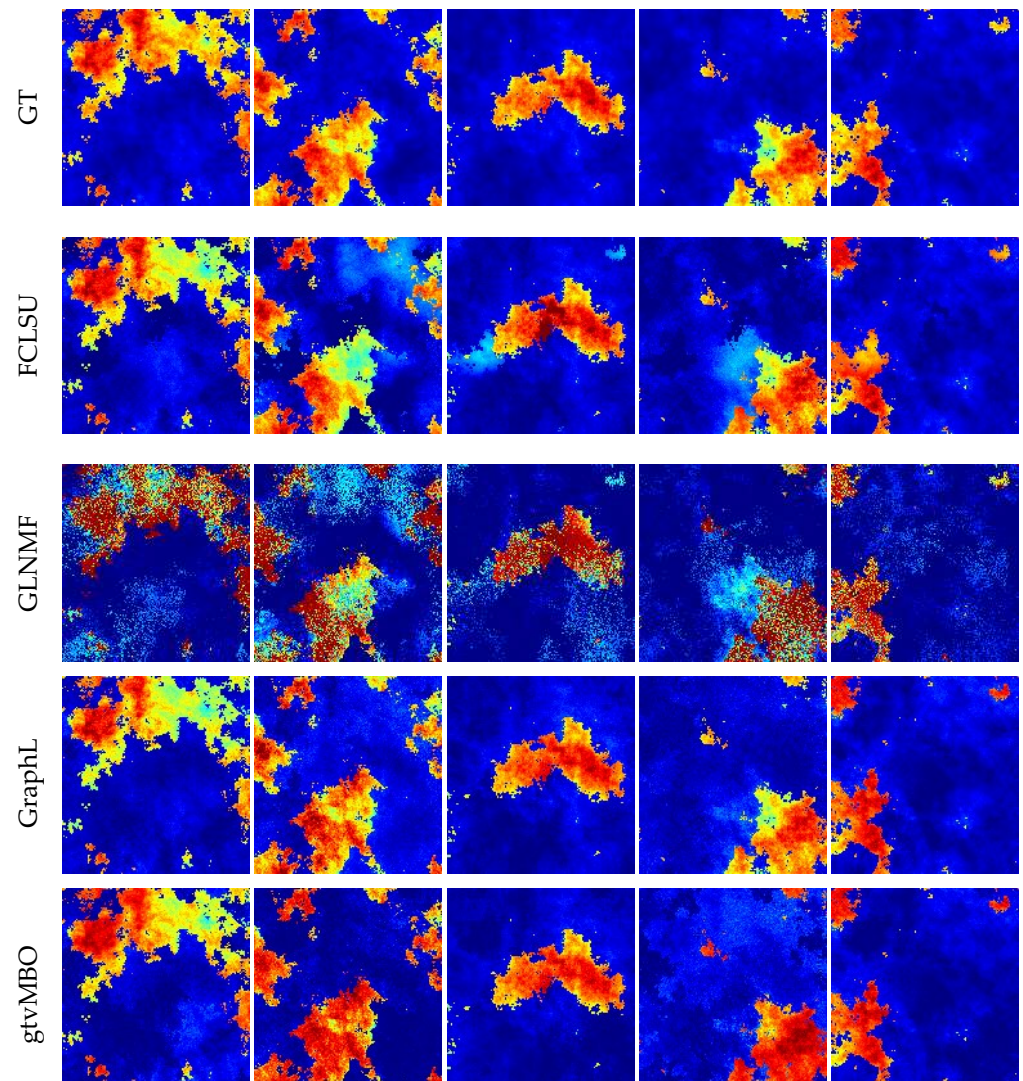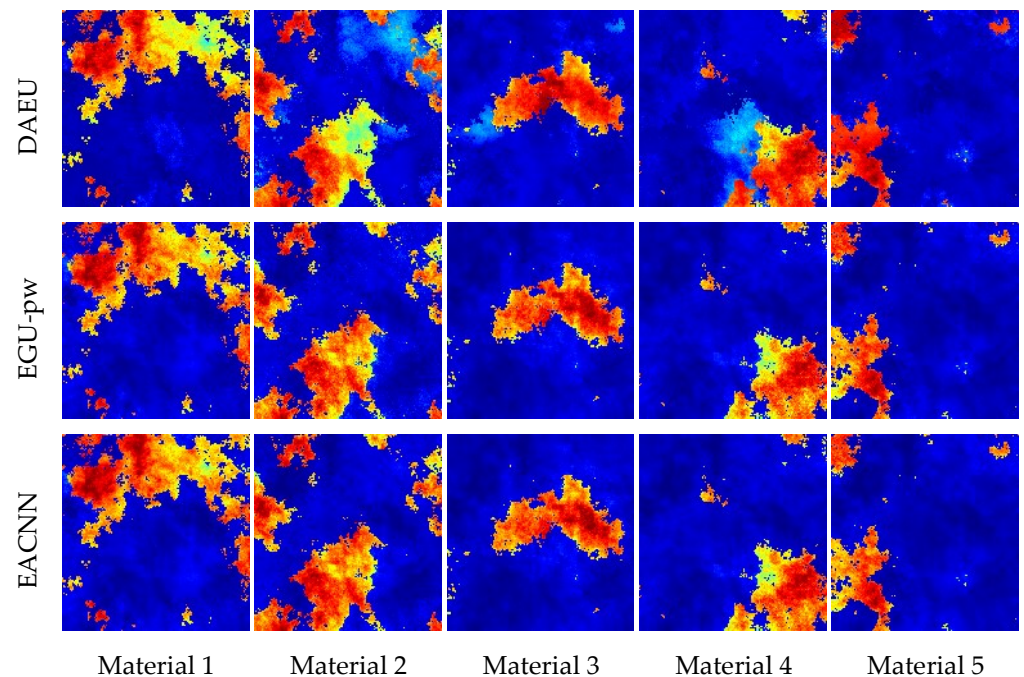**Figure 3.** The endmember extraction results for the synthetic data.

Figure 4. *Cont.*

Figure 4. The abundance result for the synthetic data estimated by the proposed methods.
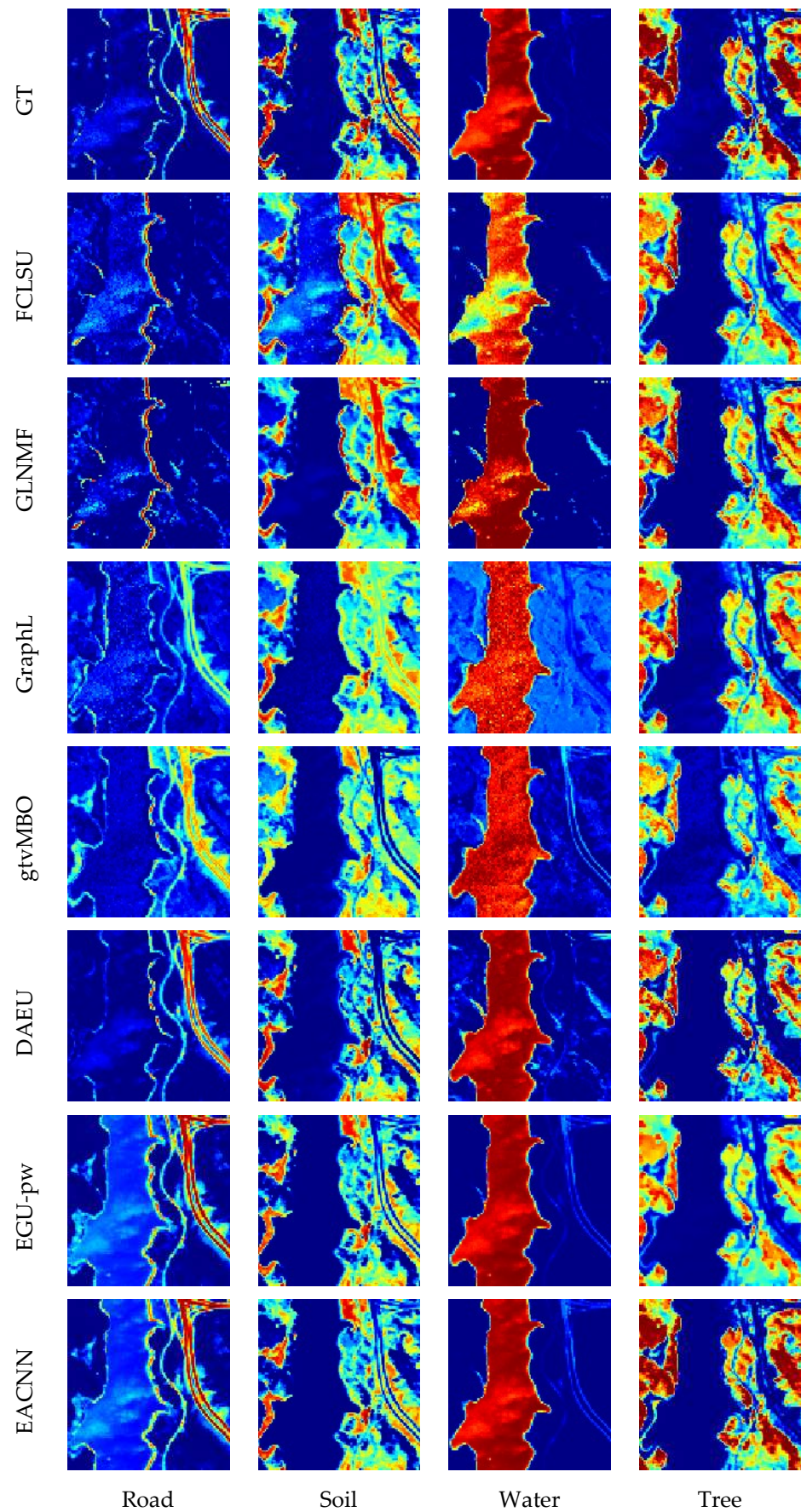
Table 3. Experimental results for the synthetic data. The best one is shown in bold.

	Methods	FCLSU	GLNMF	GraphL	gtvMBO	DAEU	EGU-pw	EACNN
	Material 1	0.0275	0.1133	0.0172	0.0438	0.0072	**0.0061**	0.0062
	Material 2	0.0270	0.0738	0.0213	0.0308	0.0069	0.0056	**0.0021**
SAD	Material 3	0.0066	0.0227	0.0039	0.0055	0.0032	0.0017	**0.0016**
	Material 4	0.0132	0.0436	0.0115	0.0191	0.0037	**0.0021**	0.0022
	Material 5	0.0072	0.0213	0.0056	0.0061	0.0033	0.0040	**0.0029**
Mean SAD		0.0163	0.0549	0.0119	0.0210	0.0049	0.0039	**0.0030**
RMSE		0.0461	0.1322	0.0355	0.0535	0.0329	0.0206	**0.0197**

4.3. Results for the Jasper Ridge Dataset

The spectral resolution of the Jasper Ridge dataset is 9.46 nm and it has a total of 512×614 pixels. Each pixel was recorded on 224 spectral bands in the wavelength range of 0.38 μm to 2.5 μm. Although the HSI is too complex, it is impossible to obtain its basic facts. Therefore, a widely used region was selected for the experiment. This region had a sub-region size of 100×100 pixels and encompassed 198 spectral bands. The research scenario consisted of four endmembers: "Road", "Soil", "Water" and "Tree".

Table 4 shows the estimates of abundance and endmembers for the Jasper Ridge dataset, and Figure 5 shows the abundance map for this dataset. Figure 6 shows the endmember extraction results. The FCLSU and gtvMBO strictly adhered to the ASC constraints, resulting in poor estimation in terms of the endmembers and abundances. While it retained the details, the GLNMF did not capture the global feature information well, presumably due to the complexity of the ground feature distribution. It can be seen from Figure 5 that GraphL and gtvMBO performed well at distinguishing narrow roads due to their use of a graph structure and that non-local similarity played an important role in the narrow pixel information provided in the abundance map. However, it can be seen that there was a lot of noise in the abundances obtained by GraphL. In addition, the DAEU performed well. EGU-pw, which has a two-stream architecture, achieved excellent results by refactoring the HSI. Compared with the former, the EACNN pays more attention to the capture of feature information and network guidance, and thus, achieved excellent Mean SAD and RMSE values.

Figure 5. The abundance results of the Jasper Ridge dataset estimated using the proposed method.

Table 4. Experimental results for the Jasper Ridge dataset.

	Methods	FCLSU	GLNMF	GraphL	gtvMBO	DAEU	EGU-pw	EACNN
SAD	Road	0.6121	0.7670	0.3344	0.1152	0.1235	**0.1082**	0.1135
	Soil	0.1672	0.1141	0.0788	0.1460	0.1187	**0.0786**	0.0816
	Water	0.2286	**0.0629**	0.3918	0.3915	0.1565	0.1832	0.1273
	Tree	0.1642	**0.0983**	0.2170	0.2434	0.1583	0.1089	0.0988
Mean SAD		0.2931	0.2606	0.2555	0.2240	0.1392	0.1197	**0.1053**
RMSE		0.1481	0.1115	0.1453	0.1360	0.1215	0.0896	**0.0700**
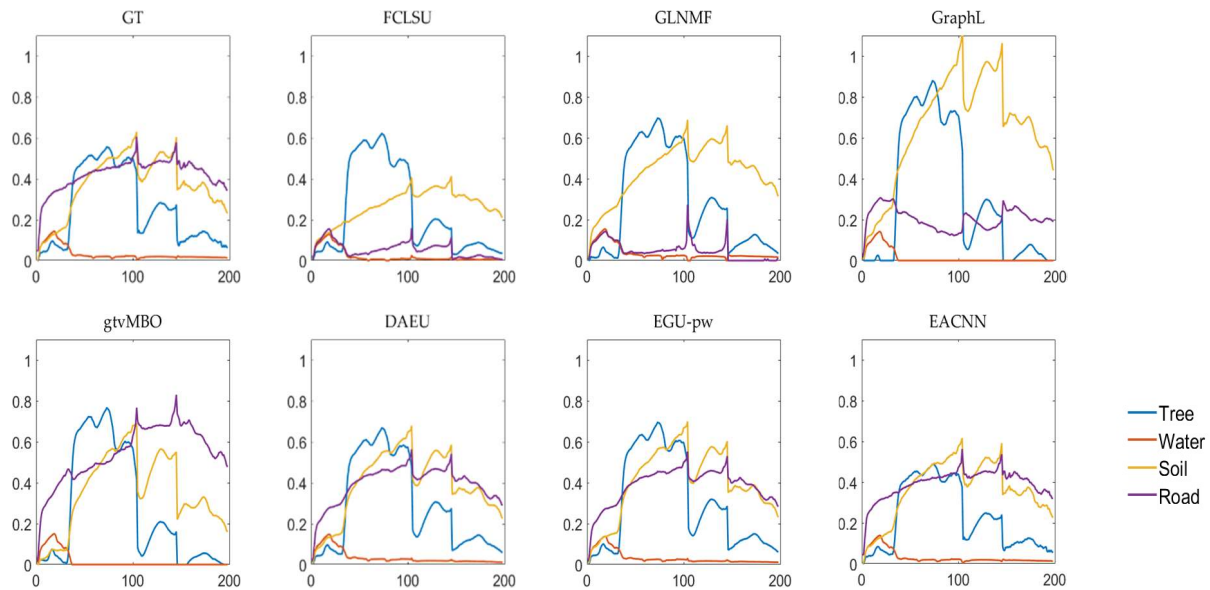
Figure 6. The endmember extraction results of the Jasper Ridge dataset.

4.4. Results for the Samson Dataset

The spectral resolution of the Samson dataset is 3.13 nm. Each pixel was recorded on 156 spectral bands ranging from 0.401 μm to 0.889 μm in terms of wavelength. A region encompassing 95 × 95 pixels was selected for the experiment. In this region, the hyperspectral data were not degraded by blank bands or serious noise bands. It contained three endmembers: "Soil", "Tree" and "Water".

Finally, experiments were conducted on the Samson dataset. Because this dataset was not degraded by blank bands or serious noise bands, it can more directly reflect the performances of the different unmixing methods. Table 5 shows the abundances and endmembers estimated for the Samson dataset.

Table 5. Experimental results for the Samson dataset.

	Methods	FCLSU	GLNMF	GraphL	gtvMBO	DAEU	EGU-pw	EACNN
SAD	Rock	0.0313	**0.0239**	0.0254	0.0453	0.0693	0.0317	0.0328
	Tree	0.0468	**0.0284**	0.0613	0.0964	0.0531	0.0572	0.0519
	Water	0.1126	0.1823	0.3249	0.3734	0.1105	0.1096	**0.1026**
Mean SAD		0.0636	0.0782	0.1372	0.1717	0.0776	0.0661	**0.0624**
RMSE		0.1801	0.1870	0.1392	0.0958	0.0332	0.0232	**0.0171**

As can be seen in Figure 7, all the methods captured the approximate shape of the endmembers and their change, but the reflectance values were different. The performances of the FCLSU and GLNMF algorithms on this dataset were relatively general. There was some obvious noise in the GLNMF abundance map, and the abundance estimation for

GraphL was 0.04 higher than those of the FCLSU and GLNMF algorithms, but the estimated endmember results were not very ideal. In addition, because the gtvMBO method imposes non-negative constraints on the endmembers via its hard threshold operators, it had many endmember estimation results close to zero. In contrast, because the distribution of the dataset was simple and there were no complex SVs, it was very obvious that compared with the traditional unmixing methods, the DL-based model obtained very excellent unmixing results. Compared with DAEU and EGU-pw, the EACNN further improved the abundance estimation for the Samson dataset by nearly 0.016 and 0.006, respectively. In other words, compared with the unmixing method based on an AE, the EACNN puts more emphasis on the spatial and spectral information contained in the HSI in the unmixing process and guides the network in obtaining the best results through its convolution operation and attention mechanisms.
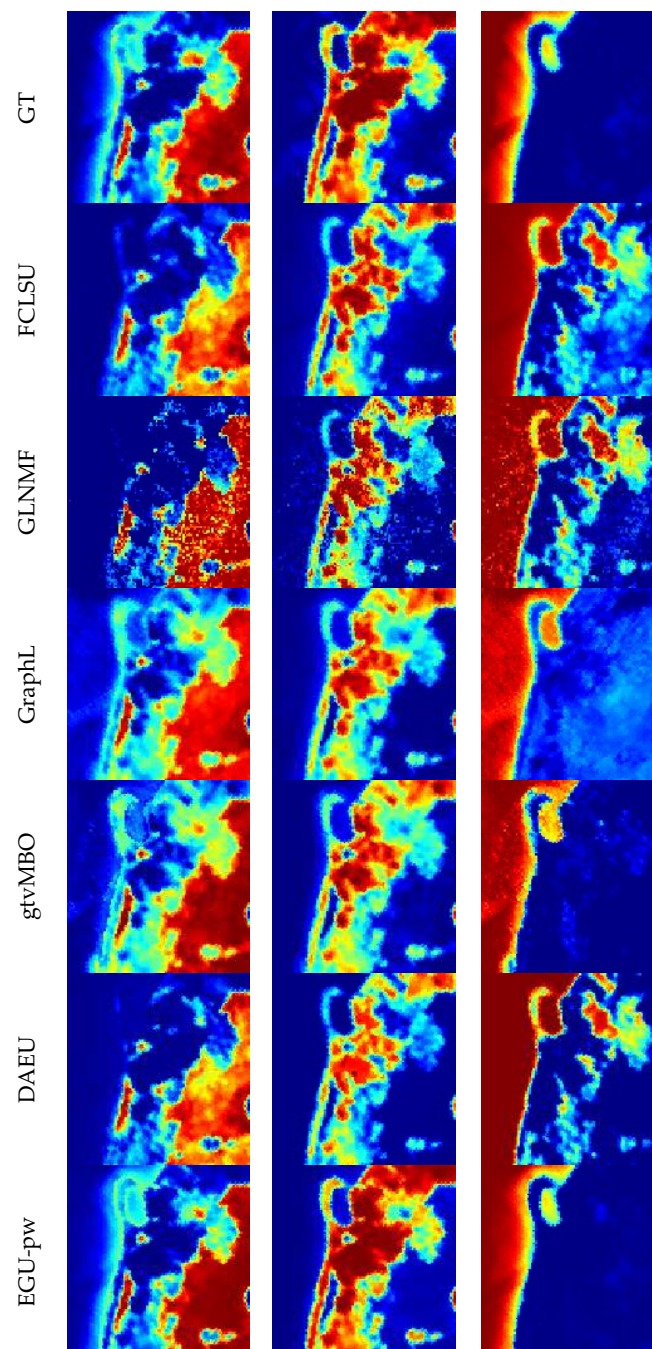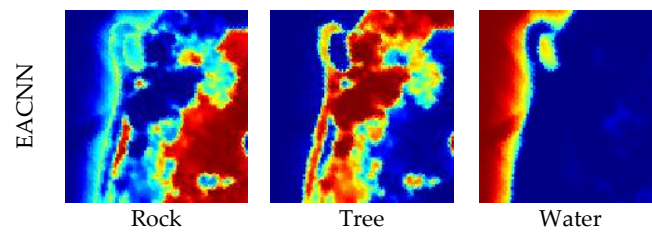

Figure 7. *Cont.*

Figure 7. The abundance results of the Samson Dataset estimated using the proposed method.

4.5. Ablation Experiments and Analyses

It can be seen from Table 6 that the ablation experiments on the network modules verified the importance of all the modules in the proposed EACNN network model, including the ECBAM and ECA attention modules and their combinations with different networks. Finally, the reliability and fairness of the ablation experiment were ensured by setting the hyperparameters consistently.

Table 6. Analysis of the ablation experiment for the Samson dataset.

Structure	RMSE
EN + UN (no spectral bundles)	0.0351
EN(ECBAM) + UN(ECA) (no spectral bundles)	0.0286
EN + UN (no feature extraction module)	0.0232
EN(ECA) + UN(ECBAM)	0.0203
EN(ECBAM) + UN(ECBAM)	0.0197
EN(ECA) + UN(ECA)	0.0195
EN(CBAM) + UN(ECA)	0.0175
EN(ECBAM) + UN(ECA)	**0.0171**

In fact, the main limitation faced by the EACNN algorithm is the accuracy and robustness of the endmember bundle extraction algorithm. After consideration of the inherent attributes of the endmembers in the EACNN method, the abundance estimation was significantly improved.

Using the same endmember bundles after removing the attention modules in the ECBAM and ECA, the EACNN obtained the worst unmixing result, which shows that the single two-stream network had some limitations with regard to HU. Thus, adding the ECBAM and ECA attention modules to the two-stream network improved the abundance and endmember estimation, especially the estimation of abundance. For the ablation experiment, it should be noted that the ECBAM attention module comprehensively obtained the spatial and spectral information from the HSI and had a better effect on the endmember bundles aggregating the HSI spatial information than the HSI itself. Hence, the ECBAM attention module was combined with the EN to add the endmember bundles and share its parameters with the UN, which can reasonably embed more detailed endmember information. In addition, especially for targets with complex ground feature distributions and serious SVs, when the HSI dataset is large, its role in obtaining spatial information is limited, and capturing the dependencies between all bands is not efficient and unnecessary. The introduction of the ECA to the UN obtained more effective spectral information for unmixing from the HSI and increased the complexity of the model while posting obvious improvements. It brought about further improvements to the network in terms of the abundances and endmembers. The results in the above experiments explained the proposed point of view well and also showed that the EACNN's multi-attention joint learning obtained the information that the network needed to pay attention to and help guide the network in improving unmixing results.

5. Discussion

This section discusses the results of the synthetic dataset, the real datasets and the ablation experiment in Section 4. First, the quantitative analysis of six different algorithms on the synthetic dataset, Jasper Ridge dataset and Samson dataset found that compared with other algorithms, the abundance results obtained using FCLSU on the three datasets were not satisfactory, and it could be seen that the effect fluctuated according to the complex distribution of real ground objects. On the other hand, according to the visual effect of the abundance map, it can be seen that the visualization quality of the three methods—GLNMF, GraphL and gtvMBO—fluctuated greatly, especially for GLNMF. Due to the SV caused by various factors in the synthetic dataset, the visualization effect of GLNMF was the most affected, which also showed that GLNMF could not handle the influence of SV well. The method based on deep learning achieved good results on the above three datasets, which showed its potential in the task of unmixing. Due to the dimensionality reduction factor, DAEU lost the rich spatial–spectral information in HSI, and the abundance map obtained by DAEU displayed the phenomenon of identification error, especially in the Samson dataset, which was less affected by noise. The result of EGU-pw was excellent, but its use of the feature information of HSI was still insufficient.

On the other hand, by making full use of spatial–spectral information, the EACNN's unmixing performance was improved. By sharing the parameters of two networks, the overall network obtained more comprehensive feature information and reduced the influence of SVs. Experiments on synthetic, real datasets and the final ablation experiments showed that our model displayed good performance.

Of course, the proposed method also had some shortcomings, and it also had some dependence on the extraction effect of the endmember bundle. In the future, we will seek a more simple and efficient way, such as a multi-modal method, to improve the precision of HU while balancing performance and efficiency.

6. Conclusions

In this study, we proposed an efficient attention-based convolutional neural network that reduced the effects of spectral variability for hyperspectral unmixing called the EACNN, which mainly includes two parts: the EN and UN. The EN learns from pure or nearly pure endmember bundles and then shares its parameters with the UN. Compared with other advanced algorithms, sharing parameters with the same weights takes into account the potential inherent properties of endmembers, which can better guide the network in learning the endmembers, thus making the unmixing results more accurate and acceptable. In the past, many unmixing networks lost the spatial and spectral information of hyperspectral images in varying degrees under dimensionality reduction learning, thus losing the significance of these images. Thanks to the ECA and ECBAM attention mechanism, the EACNN can learn feature information from differently scaled HSIs to help the network to develop in a better direction. Endmember bundles are well-equipped to deal with SVs. We will search for more specific physical constraints in future work to stabilize the physical meaning of DL-based unmixing and improve the framework to make it more efficient.

Abbreviations

The abbreviations used in this paper are as follows:

AE	Autoencoder
CNN	Convolutional neural network
HU	Hyperspectral unmixing
SV	Spectral variability
HSI	Hyperspectral image
EACNN	Efficient attention-based convolutional neural network
ECBAM	Efficient convolutional block attention module
LMM	Linear mixing model
NLMM	Nonlinear mixing model
VCA	Vertex component analysis
DL	Deep learning
NLP	Natural language processing
SE	Squeeze excitation
NMF	Non-negative matrix factorization
ELMM	Extended linear mixing model
RTT	Radiative transfer theory
BMM	Bilinear mixture model
EN	Endmember network
UN	Unmixing network
ANC	Abundance non-negativity constraint
ASC	Abundance sum-to-one constraint
ECA	Effective channel attention
FCLSU	Fully constrained least-squares unmixing
MBO	Merriman–Bence–Osher
DAEU	Deep autoencoder unmixing
EGU	Endmember-guided unmixing
RMSE	Root-mean-square error
SAD	Spectral angle distance

References

1. Hong, D.; He, W.; Yokoya, N.; Yao, J.; Gao, L.; Zhang, L.; Chanussot, J.; Zhu, X. Interpretable hyperspectral artificial intelligence: When nonconvex modeling meets hyperspectral remote sensing. *IEEE Geosci. Remote Sens. Mag.* **2021**, *9*, 52–87. [CrossRef]
2. Jin, Q.; Ma, Y.; Mei, X.; Li, H.; Ma, J. Gaussian mixture model for hyperspectral unmixing with low-rank representation. In Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2019), Yokohama, Japan, 28 July–2 August 2019; pp. 294–297.
3. Jin, Q.; Ma, Y.; Mei, X.; Li, H.; Ma, J. UTDN: An Unsupervised Two-Stream Dirichlet-Net for Hyperspectral Unmixing. In Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2021), Toronto, ON, Canada, 6–11 June 2021; pp. 1885–1889.
4. Keshava, N.; Mustard, J.F. Spectral unmixing. *IEEE Signal Process. Mag.* **2002**, *19*, 44–57. [CrossRef]
5. Bioucas-Dias, J.M.; Plaza, A.; Dobigeon, N.; Parente, M.; Du, Q.; Gader, P.; Chanussot, J. Hyperspectral Unmixing Overview: Geometrical, Statistical, and Sparse Regression-Based Approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 354–379. [CrossRef]
6. Nascimento, J.M.P.; Dias, J.M.B. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 898–910. [CrossRef]
7. Winter, M.E. N-FINDR: An algorithm for fast autonomous spectral endmember determination in hyperspectral data. Spies International Symposium on Optical Science(SPIE). In *Imaging Spectrom V*; International Society for Optics and Photonics: Bellingham, WA, USA, 1999; Volume 3753, pp. 266–275.
8. Gader, P.; Parente, M.; Heylen, R. A Review of Nonlinear Hyperspectral Unmixing Methods. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 1844–1868.
9. Halimi, A.; Altmann, Y.; Dobigeon, N.; Tourneret, J.Y. Nonlinear Unmixing of Hyperspectral Images Using a Generalized Bilinear Model. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4153–4162. [CrossRef]
10. Altmann, Y.; Halimi, A.; Dobigeon, N.; Tourneret, J.Y. Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery. *IEEE Trans. Image Process.* **2012**, *21*, 3017–3025. [CrossRef]
11. Heylen, R.; Scheunders, P. A Multilinear Mixing Model for Nonlinear Spectral Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 240–251. [CrossRef]

12. Ye, Q.; Li, Z.; Fu, L.; Zhang, Z.; Yang, W.; Yang, G. Nonpeaked Discriminant Analysis for Data Representation. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 3818–3832. [CrossRef] [PubMed]
13. Fu, X.; Ma, W.; Bioucas-Dias, J.M.; Chan, T. Semiblind Hyperspectral Unmixing in the Presence of Spectral Library Mismatches. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 5171–5184. [CrossRef]
14. Thouvenin, P.; Dobigeon, N.; Tourneret, J. Hyperspectral Unmixing With Spectral Variability Using a Perturbed Linear Mixing Model. *IEEE Trans. Signal Process.* **2016**, *64*, 525–538. [CrossRef]
15. Miao, L.; Qi, H. Endmember Extraction From Highly Mixed Data Using Minimum Volume Constrained Nonnegative Matrix Factorization. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 765–777. [CrossRef]
16. Imbiriba, T.; Borsoi, R.A.; Bermudez, J.C.M. Generalized Linear Mixing Model Accounting for Endmember Variability. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 1862–1866.
17. Hapke, B. Bidirectional reflectance spectroscopy 1: Theory. *J. Geophys. Res. Earth Surf.* **1981**, *86*, 3039–3054. [CrossRef]
18. Somers, B.; Cools, K.; Delalieux, S.; Stuckens, J.; Zande, D.; Verstraeten, W.W.; Coppin, P. Nonlinear Hyperspectral Mixture Analysis for tree cover estimates in orchards. *Remote Sens. Environ.* **2009**, *113*, 1183–1193. [CrossRef]
19. Fan, W.; Baoxin, H.U.; Miller, J.; Mingze, L.I. Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data. *Int. J. Remote Sens.* **2009**, *30*, 2951–2962. [CrossRef]
20. Zhang, L.; Wu, B.; Huang, B.; Li, P. Nonlinear estimation of subpixel proportion via kernel least square regression. *Int. J. Remote Sens.* **2007**, *28*, 4157–4172. [CrossRef]
21. Broadwater, J.; Chellappa, R.; Banerjee, A.; Burlina, P. Kernel fully constrained least squares abundance estimates. In Proceedings of the 2007 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2007), Barcelona, Spain, 23–28 July 2007.
22. Dobigeon, N.; Févotte, C. Robust nonnegative matrix factorization for nonlinear unmixing of hyperspectral images. In Proceedings of the 2013 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Gainesville, FL, USA, 26–28 June 2013; pp. 1–4.
23. Fang, B.; Li, Y.; Zhang, P.; Bai, B. Kernel sparse NMF for hyperspectral unmixing. In Proceedings of the 2014 International Conference on Orange Technologies, Xi'an, China, 20–23 September 2014; pp. 41–44.
24. Mario, S.; Fausto, S.; Vincenzo, B. A new board for CNN stereo vision algorithm. In Proceedings of the 2000 IEEE International Symposium on Circuits and Systems (ISCAS), Geneva, Switzerland, 28–31 May 2000; pp. 702–705.
25. Fahad, S.K.A.; Yahya, A.E. Inflectional Review of Deep Learning on Natural Language Processing. In Proceedings of the 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE), Shah Alam, Malaysia, 11–12 July 2018; pp. 1–4.
26. Licciardi, G.A.; Frate, F.D. Pixel Unmixing in Hyperspectral Data by Means of Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4163–4172. [CrossRef]
27. Palsson, B.; Sigurdsson, J.; Sveinsson, J.R.; Ulfarsson, M.O. Hyperspectral Unmixing Using a Neural Network Autoencoder. *IEEE Access* **2018**, *6*, 25646–25656. [CrossRef]
28. Qu, Y.; Guo, R.; Qi, H. Spectral unmixing through part-based non-negative constraint denoising autoencoder. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 209–212.
29. Qu, Y.; Qi, H. uDAS: An Untied Denoising Autoencoder With Sparsity for Spectral Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 1698–1712. [CrossRef]
30. Su, Y.; Li, J.; Plaza, A.; Marinoni, A.; Gamba, P.; Chakravortty, S. DAEN: Deep Autoencoder Networks for Hyperspectral Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4309–4321. [CrossRef]
31. Hong, D.; Chanussot, J.; Yokoya, N.; Heiden, U.; Heldens, W.; Zhu, X.X. WU-Net: A Weakly-Supervised Unmixing Network for Remotely Sensed Hyperspectral Imagery. In Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 373–376.
32. Jin, Q.; Ma, Y.; Mei, X.; Ma, J. TANet: An Unsupervised Two-Stream Autoencoder Network for Hyperspectral Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–15. [CrossRef]
33. Zhang, X.; Sun, Y.; Zhang, J.; Wu, P.; Jiao, L. Hyperspectral Unmixing via Deep Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1755–1759. [CrossRef]
34. Arun, P.V.; Buddhiraju, K.; Porwal, A. CNN based sub-pixel mapping for hyperspectral images. *Neurocomputing* **2018**, *311*, 51–64. [CrossRef]
35. Xiao, T.; Xu, Y.; Yang, K.; Zhang, J.; Peng, Y.; Zhang, Z. The application of two-level attention models in deep convolutional neural network for fine-grained image classification. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 842–850.
36. Haut, J.M.; Paoletti, M.E.; Plaza, J.; Plaza, A.; Li, J. Visual Attention-Driven Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8065–8080. [CrossRef]
37. Sun, L.; Fang, Y.; Chen, Y.; Huang, W.; Wu, Z.; Jeon, B. Multi-Structure KELM with Attention Fusion Strategy for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–17. [CrossRef]
38. Liu, N.; Han, J.; Yang, M. PiCANet: Learning Pixel-Wise Contextual Attention for Saliency Detection. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3089–3098.

39. Wu, X.; Li, W.; Hong, D.; Tian, J.; Tao, R.; Du, Q. Vehicle Detection of Multi-source Remote Sensing Data Using Active Fine-tuning Network. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 39–53. [CrossRef]

40. Sun, L.; Cheng, S.; Zheng, Y.; Wu, Z.; Zhang, J. SPANet: Successive Pooling Attention Network for Semantic Segmentation of Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4045–4057. [CrossRef]

41. Fu, L.; Zhang, D.; Ye, Q. Recurrent Thrifty Attention Network for Remote Sensing Scene Recognition. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 8257–8268. [CrossRef]

42. Zeng, Y.; Ritz, C.; Zhao, J.; Lan, J. remote sensing attention-based residual network with scattering transform features for hyperspectral unmixing with limited training samples. *Remote Sens.* **2020**, *12*, 400. [CrossRef]

43. Han, Z.; Hong, D.; Gao, L.; Yao, J.; Zhang, B.; Chanussot, J. Multimodal Hyperspectral Unmixing: Insights From Attention Networks. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [CrossRef]

44. Heinz, D.C.; Chein, I.C. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 529–545. [CrossRef]

45. Lu, X.; Wu, H.; Yuan, Y.; Yan, P.; Li, X. Manifold Regularized Sparse NMF for Hyperspectral Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2815–2826. [CrossRef]

46. Qin, J.; Lee, H.; Chi, J.T.; Lou, Y.; Chanussot, J.; Bertozzi, A.L. Fast Blind Hyperspectral Unmixing Based On Graph Laplacian. In Proceedings of the 2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHPERS), Amsterdam, The Netherlands, 24–26 September 2019; pp. 1–5.

47. Qin, J.; Lee, H.; Chi, J.T.; Drumetz, L.; Chanussot, J.; Lou, Y.; Bertozzi, A.L. Blind Hyperspectral Unmixing Based on Graph Total Variation Regularization. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 3338–3351. [CrossRef]

48. Hong, D.; Gao, L.; Yao, J.; Yokoya, N.; Chanussot, J.; Heiden, U.; Zhang, B. Endmember-Guided Unmixing Network (EGU-Net): A General Deep Learning Framework for Self-Supervised Hyperspectral Unmixing. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 6518–6531. [CrossRef] [PubMed]

49. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 11531–11539.

50. Somers, B.; Zortea, M.; Plaza, A.; Asner, G.P. Automated Extraction of Image-Based Endmember Bundles for Improved Spectral Unmixing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 396–408. [CrossRef]

51. Jin, Q.; Ma, Y.; Pan, E.; Fan, F.; Huang, J.; Li, H.; Sui, C.; Mei, X. Hyperspectral Unmixing with Gaussian Mixture Model and Spatial Group Sparsity. *Remote Sens.* **2019**, *11*, 2434. [CrossRef]

52. Eches, O.; Dobigeon, N.; Tourneret, J. Enhancing Hyperspectral Image Unmixing With Spatial Correlations. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4239–4247. [CrossRef]

53. Giampouras, P.V.; Themelis, K.E.; Rontogiannis, A.A.; Koutroumbas, K.D. Simultaneously Sparse and Low-Rank Abundance Matrix Estimation for Hyperspectral Image Unmixing. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4775–4789. [CrossRef]

54. Remote Sensing Laboratory School of Surveying and Geospatial Engineering. Available online: https://rslab.ut.ac.ir/data (accessed on 10 July 2022).

55. Drumetz, L.; Veganzones, M.A.; Henrot, S.; Phlypo, R.; Chanussot, J.; Jutten, C. Blind Hyperspectral Unmixing Using an Extended Linear Mixing Model to Address Spectral Variability. *IEEE Trans. Image Process.* **2016**, *25*, 3890–3905. [CrossRef]