Article

GASSF-Net: Geometric Algebra Based Spectral-Spatial Hierarchical Fusion Network for Hyperspectral and LiDAR Image Classification

Rui Wang [1,*], Xiaoxi Ye [1], Yao Huang [1], Ming Ju [1] and Wei Xiang [2,3]

1   School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China;
   yexiaoxi@shu.edu.cn (X.Y.); huangyaoem@shu.edu.cn (Y.H.); jm735@shu.edu.cn (M.J.)
2   School of Computing, Engineering and Mathematical Sciences, La Trobe University,
   Melbourne, VIC 3086, Australia; w.xiang@latrobe.edu.au
3   College of Science and Engineering, James Cook University, Cairns, QLD 4878, Australia
*   Correspondence: rwang@shu.edu.cn

Abstract: The field of multi-source remote sensing observation is becoming increasingly dynamic through the integration of various remote sensing data sources. However, existing deep learning methods face challenges in differentiating between internal and external relationships and capturing fine spatial features. These models often struggle to effectively capture comprehensive information across remote sensing data bands, and they have inherent differences in the size, structure, and physical properties of different remote sensing datasets. To address these challenges, this paper proposes a novel geometric-algebra-based spectral–spatial hierarchical fusion network (GASSF-Net), which uses geometric algebra for the first time to process multi-source remote sensing images, enabling a more holistic approach to handling these images by simultaneously leveraging the real and imaginary components of geometric algebra to express structural information. This method captures the internal and external relationships between remote sensing image features and spatial information, effectively fusing the features of different remote sensing data to improve classification accuracy. GASSF-Net uses geometric algebra (GA) to represent pixels from different bands as multivectors, thus capturing the intrinsic relationships between spectral bands while preserving spatial information. The network begins by deeply mining the spectral–spatial features of a hyperspectral image (HSI) using pairwise covariance operators. These features are then extracted through two branches: a geometric-algebra-based branch and a real-valued network branch. Additionally, the geometric-algebra-based network extracts spatial information from light detection and ranging (LiDAR) to complement the elevation data lacking in the HSI. Finally, a genetic-algorithm-based cross-fusion module is introduced to fuse the HSI and LiDAR data for improved classification. Experiments conducted on three well-known datasets, Trento, MUUFL, and Houston, demonstrate that GASSF-Net significantly outperforms traditional methods in terms of classification accuracy and model efficiency.

Keywords: multi-source remote sensing data; geometric algebra; cross-channel fusion; multi-branch hierarchical fusion

1. Introduction

In urban smart construction, spatial information in remote sensing images plays an essential role. The hyperspectral image (HSI) is a pervasive tool in the field of remote sensing due to its ability to provide high spectral resolution. It combines sample images with spectral information and provides a 3D data representation with rich spectral bands and abundant spatial–spectral information [1]. This technology has been extensively utilized in numerous Earth observation tasks, including land cover classification [2], scene classification [3], object detection [4], medical diagnosis [5], and urban planning [6].

Convolutional neural network (CNN)-based classification models primarily focus on the spatial–spectral characteristics of the hyperspectral image (HSI). Li et al. [7] introduced a spectral context-aware transformer (SCAT) algorithm that better captures spatial information in his; consequently, this results in a notable improvement in the accuracy of categorization. Zhang et al. [8] introduced an unsupervised convolutional neural network model based on spectral–spatial features. This model fully utilizes the spectral–spatial features to better extract sample features. Through 2D CNN, the spatial characteristics of the HSI can be extracted, while 3D CNNs are effective for spectral feature extraction. Additionally, Gao et al. [9] introduced a new multi-scale, two-branch feature fusion method based on an attention mechanism that addresses the limitations of previous methods that relied on static convolutional kernels and a step-by-step approach to feature extraction. By utilizing a multi-branch network structure, the depth of the network is reduced, and it better adapts to different scales of features while accounting for both spectral and spatial data.

However, the application of the HSI is limited in complex terrains and multi-object conditions, including the ability to directly capture the spectral properties of ground objects. Therefore, it can be augmented with other remotely sensed data when performing classification tasks. By integrating complementary information extracted from multimodal data, more robust and reliable decisions can be made in feature classification tasks [10].

In recent years, many multimodal remote sensing data fusion methods have been proposed for consideration [11], and feature extraction can be performed efficiently by using deep learning networks, which provide a solid foundation for addressing the above challenges. Li et al. [12] put forth the concept of unsupervised fusion networks with diminished adaptive learning capabilities, which are capable of directly encoding spatial and spectral transformations across a range of resolutions. The CNN-Fus [13] method fuses the HSI and MSI using a subspace representation and a CNN noise reducer, and it is used for grayscale image denoising, which shows better performance. A cross-modal-learning X-shaped interactive self-encoder network (XINet) [14] couples two disconnected U-nets through a parameter sharing strategy to realize information exchange and complementarity between modalities. In contrast, traditional light detection and ranging (LiDAR) incorporates additional data, including 3D coordinate information, reflection intensity data, time stamps, and echo information [15]. This enables more effective compensation for the absence of elevation data in hyperspectral data. Hong et al. [16] introduced a deep encoder–decoder network structure (End-Net) by reconstructing multimodal data with feature fusion to realize cross-modal activation of neurons, which can be effectively fused to multimodal images. Zhang et al. [17] introduced the interleaved perception convolutional neural network (IP-CNN), a two-branch CNN architecture for integrating different input data. Zhang et al. [18] proposed a new three-channel CNN to extract the spectral, spatial, and elevation information of remote sensing images, and a multilevel feature fusion (MLF) module was employed to integrate shallow and deep features. In addition, a mutually guided attention (MGA) module was introduced to achieve a comprehensive fusion of spatial and elevation data. Ding et al. [19] proposed a novel approach to the utilization of both local and global features simultaneously and employed a probabilistic approach for classification estimation through decision fusion. Lu et al. [20] introduced a new classification method based on coupled adversarial learning (CALC). This method trains a coupled adversarial feature learning (CAFL) sub-network, which enables the unsupervised extraction of high-level semantic features from hyperspectral images (HSIs) and LiDAR data. The CAFL sub-network generates multiple category-estimated probabilities by learning the low-, intermediate-, and high-level features, which are then combined in an adaptive manner to produce the final, accurate classification results. Yu et al. [21] proposed a shadow-mask-driven multimodal endowment image decomposition (smMIID) approach to overcome the shortcomings of existing intrinsic image decomposition (IID)-based frameworks in terms of information diversity and modal relevance.

Researchers, such as Nitta [22,23], generalized the neural network on the quaternion domain and designed the BP neural network. Li et al. [24] applied principal component analysis (PCA) to the HSI to obtain and maintain the orthogonal structure of the HSI and encoded the first three principal components (PCs) as the three imaginary parts of the quaternion. Voronin et al. [25] used a quaternion framework to represent remote sensing images, and Rao et al. [26] proposed an innovative quaternion-based network for HSI classification (QHIC-Net). This network captures both the local dependencies among spectral channels for individual pixels and the global structural relationships that define edges or shapes formed by pixel groups. Zhou et al. [27] investigated the mapping of real HSI features into quaternion features and proposed a new separable quaternion convolutional neural network (SQNet) to classify hyperspectral images using quaternion convolutional neural networks. However, quaternions have only four parts, and thus they cannot handle the complex spectral spatial structure information in remote sensing data.

Remote sensing imagery typically includes rich spectral and spatial details, while existing deep learning methods primarily address the external spatial relationships between pixels. These methods often overlook the intrinsic relationships among various attributes within a single pixel, making it challenging to accurately identify and distinguish the internal and external relationships between different features in the image, which can result in the loss of structural features and critical information across different spectral bands. Furthermore, remote sensing data generally comprise multiple spectral bands, and the correlation between these bands is essential for precise classification, but traditional deep learning models struggle to capture the connections between comprehensive information. The interactions between different bands are often ignored or oversimplified, which leads to limitations in the model's ability to extract comprehensive spectral–spatial features, thus affecting classification outcomes.

To address this limitation, we propose geometric algebra (GA) networks for the first time in multi-source remote sensing image classification. Weights within the geometric algebraic neurons can fully capture the structural features according to the algebraic rules, and each pixel in the remote sensing image is represented as a multivector in this method, which allows the whole remote sensing image to be represented by a genetic matrix instead of several independent real-valued matrices. The complex internal and external relationships and spatial information inherent in the remote sensing data are extracted more efficiently through the integrated processing of the internal correlation information and the global overall information of the image. The GA-based convolution and the real-valued convolution are hierarchically fused across different domains to mitigate the instability associated with extracting HSI depth features, thereby enhancing the model's overall performance. Based on this framework, we propose a GA-based spectral–spatial hierarchical fusion network (GASSF-Net) for multi-source remote sensing data. The primary contributions of this paper are as follows:

(1) In response to the complex spectral and spatial information of remote sensing data, as well as the holistic relationships between different bands, this study extends convolutional layers into the geometric algebra domain for the integration and categorization of multi-source remote sensing images. By using geometric algebra matrices to represent the entire remote sensing image, both internal correlations and holistic spatial relationships can be processed simultaneously. This multi-dimensional representation captures the complex interactions between spectral and spatial features more effectively than traditional real-valued matrices.

(2) To enhance the correlation of spectral dimensions while improving model performance, the multi-source feature extraction (MSFE) module uses pairwise ensemble operators (PEOs) to preserve the spectral and spatial information of his, thereby deeply mining spectral features.

(3) A GA and real-valued domain multi-dimensional fusion module (GRMF) is proposed as a means to extract deep features from the HSI. GA convolution effectively captures the relationships between different spectral bands, and its integration with the real-

valued convolution enables more comprehensive information extraction. In addition, the GA network extracts features from LiDAR data, thus improving the correlation between spatial and spectral information. These neurons can fully capture structural features according to algebraic rules, leading to more efficient extraction of complex internal correlations and spatial information, thus improving the overall performance of the model.

(4) A GA-based cross-fusion (GACF) module is employed to achieve comprehensive multi-source feature fusion in the spectral–spatial domain, which enables feature-level fusion while preserving holistic relationships between different attributes.

The following is a detailed account of the article in question. Section 2 provides a detailed analysis of the methodology used for each module of the experiment, while Section 3 presents the findings of the experimental research and the subsequent analysis. Section 4 discusses the proposed model. Lastly, Section 5 provides the conclusion of this paper.

2. Materials and Methods

2.1. Preliminary Foundations of Geometric Algebra

Geometric algebra (GA), first described by William K. Clifford and also referred to as Clifford algebra [28,29], is a branch of mathematics that combines algebra and geometry, combining Hamilton's quaternions and Grassmann's dilation algebra to enable high-dimensional geometric computations, and it is an expansion algebra of algebraic domains, such as real, complex, and quaternion [30–32].

The most basic algebraic elements in GA are vectors, which extend traditional vector algebra by introducing additional structures and operations, such as points, lines, planes, and polynomials. The core concepts of GA are polynomial rings and algebraic expansion fields. A polynomial ring is an algebraic structure consisting of polynomials and their coefficients that can be used to represent geometric objects and transformations [33,34]. Investigating geometric properties and transformations, and performing accurate and efficient geometric computations and analyses, are of significant utility in the fields of mathematics and physics [35–38].

Suppose that there exists a GA set in an n-dimensional space, denoted by $G_{p,q}$, where $n = p + q$. Then, the orthogonal basis vectors $\{e_1, e_2, \ldots, e_n\}$ that exist in the space $G_{p,q}$ are defined as

$$\begin{cases} e_i^2 = 1, i = 1, \ldots, p \\ e_i^2 = -1, i = p+1, \ldots p+q \end{cases} \tag{1}$$

In the GA framework used in this paper, q = 0. In the following, $G_{p,0}$ is denoted by G_n. The complete orthogonal basis $\{1, \{e_i\}, \{e_i e_j\}, \ldots, \{e_1 e_2 \ldots e_n\}\}$ of G_n can be simplified to the form of $\{1, \{e_i\}, \{e_{ij}\}, \ldots, \{e_{12\ldots n}\}\}$.

The power set $\gamma = \{1, \ldots, n\}$ can be employed to transform the basis into an ordered basis. The index set is derived as follows:

$$B := \{(a_1, \ldots a_r) \in \gamma, 1 \leq a_1 \ldots a_r \leq n\} \tag{2}$$

Then, the number of complete orthogonal bases of the order 2^n exists in space G_n, as follows:

$$\{e_I := e_{a_1} \ldots e_{a_r} | I \in B\} \tag{3}$$

GA unifies the concepts of inner and outer products by means of the geometric product, which no longer relies on coordinate information and makes computation more flexible. The geometric product of two multivectors, $v, w \in G_n$, is denoted by

$$v \otimes_n w = v \cdot w + v \wedge w \tag{4}$$

where $v \cdot w$ represents the inner product of the vectors and $v \wedge w$ signifies the outer product of the same two vectors. In the G_n space, the orthogonal basis vectors are orthogonal, and the inner product is 0, so the geometric product above is equivalent to the outer product. Therefore, the geometric product is non-exchangeable, and then we have

$$e_i e_j = -e_j e_i, i, j = 1, \ldots, n, i \neq j \tag{5}$$

In GA, an arbitrary multivector Z is described as shown in (6):

$$v = E_0 + \sum_{1 \leq i \leq n} E_i(K)e_i + \sum_{1 \leq i \leq j \leq n} E_{ij}(K)e_{ij} + \ldots + E_{1\ldots n}(K)e_{1\ldots n}, E(K) \in \mathbb{R} \tag{6}$$

where $\mathbb{R}$ denotes the real number field. The above equation can be simplified as

$$v = \sum_{t=0}^{n} \langle v \rangle_t = \sum_{I \in B} [v]_I e_I \tag{7}$$

where $[v]_I \in \mathbb{R}$ denotes the value of each component of the multivector.

2.2. Methods

The conventional techniques for processing multi-channel images typically extract features by merely summing the outputs of the individual channels, which inevitably results in the loss of the intricate relationships between space and structure. In contrast, GA offers a robust mathematical model for the representation and manipulation of geometric objects in multi-dimensional space, encompassing vectors and rotations, which effectively preserves key spectral and spatial structure information in remote sensing images. By representing each multi-spectral pixel with GA multivectors, the intrinsic connection between pixels can be captured with greater precision, which facilitates the acquisition of more detailed and accurate image representations. The combination of geometric algebra with other techniques allows for the extraction of useful features from remote sensing images with greater precision, which can then be employed in high-precision classification. Additionally, geometric algebra networks consider additional contextual information and geometric relationships during the feature extraction process, resulting in more accurate and reliable classification outcomes. In this section, we apply geometric algebraic convolution to the fusion classification of multi-source remote sensing images and propose the GASSF-Net model. Figure 1 presents the overall structure of the proposed network model, which comprises a feature extraction network, a multi-dimensional fusion module, and a cross-fusion module for fusion classification.

In our proposed model, multi-source remote sensing image feature extraction is performed initially to capture the spectral and spatial information of remote sensing images while maintaining their interdependence and reducing the number of spectra to be processed in subsequent feature extraction stages. Then, this is followed by the input of GA-based convolutional and real-valued convolutional layer branch fusion networks. These features are then fed into the GA-based cross-fusion module for fusion and classification. The functions of each module are described in detail below.
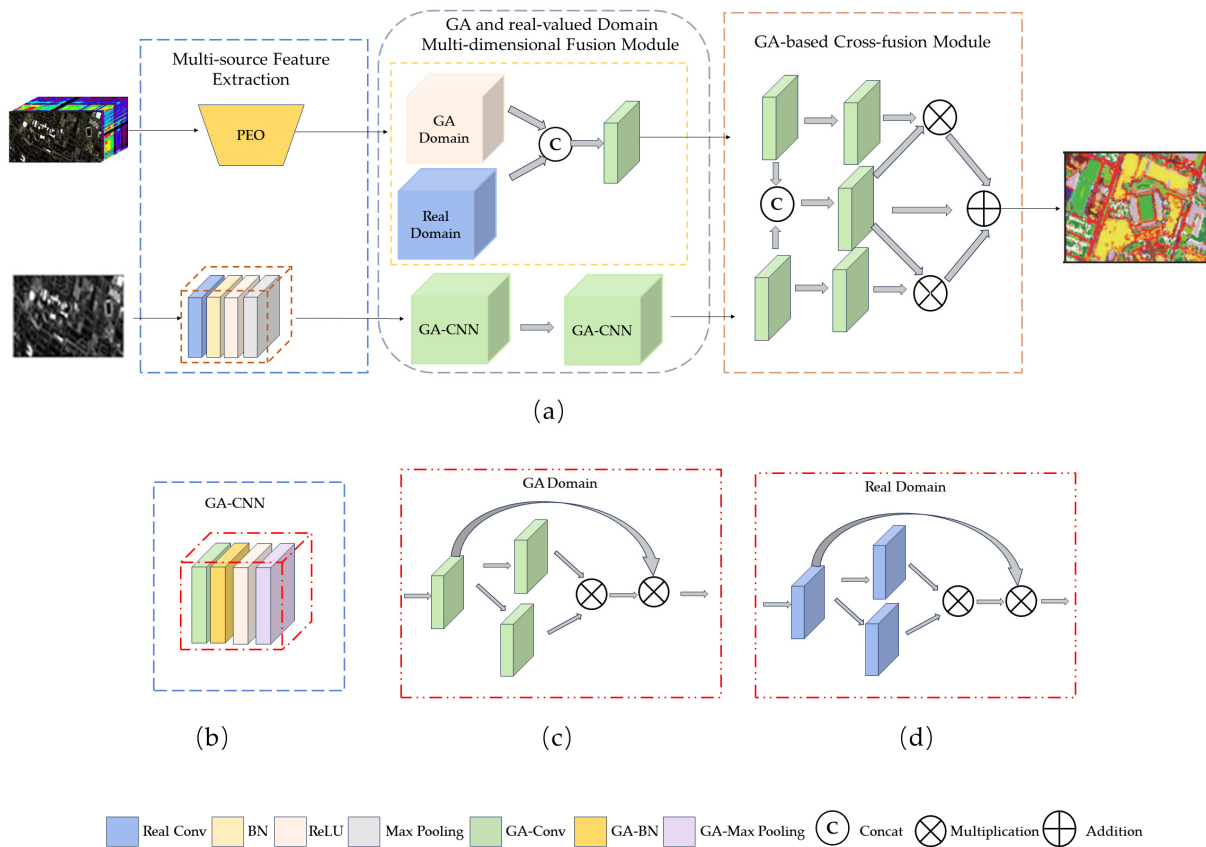
Figure 1. The flow chart structure of the method is given. Figure (**a**) illustrates the overall architecture of the proposed approach. Figure (**b**) shows the GA-CNN. Figure (**c**) shows in detail the branches of the geometric algebraic domain in the GRMF block. Figure (**d**) shows in detail the branches of the real-valued domain in the GRMF block.

2.2.1. Multi-Source Feature Extractor Module

$X_{Hsi} \in H \times W \times D$ denotes the hyperspectral image (HSI) of a given region, while $X_{LiDAR} \in H \times W \times d$ denotes the corresponding light detection and ranging (LiDAR) image, where H and W represent the height and width of the HSI and LiDAR images and D and d represent the number of spectral bands of the hyperspectral image and the lidar image, respectively.

To extract both spectral and spatial information of the stellar meter in an efficient manner, we innovatively introduce the pairwise ensemble operator (PEO) to extract the spectral spatial features of the HSI, as inspired by the literature [39,40]. The specific introduction is shown in Figure 2. Firstly, the PEO acquires the input features through feature mapping and generates kernels corresponding to them through automatic computation. In contrast to fixed kernels, these kernels can be adapted dynamically to the particulars of the input features, thereby enhancing the efficiency of information extraction. The method is highly adaptive, thus ensuring the accurate capture and amplification of key information while effectively suppressing noise and redundant data, thereby enhancing the accuracy and completeness of information extraction. Then, this information is combined with the features extracted from the three-layer filter to obtain F_{hsi}^{spe}. As opposed to conventional convolutional neural networks, we refrain from utilizing batch normalization (BN) and rectified linear unit (ReLU) activation functions throughout the training process, which accelerate training and enhance the generalization capacity of the model to a certain extent. By avoiding these operations, our model is better able to preserve the intrinsic characteristics of hyperspectral data, thus ensuring more precise feature extraction and classification performance. Therefore, by directly multiplying the features obtained from the triple-layer

filter, we retain the original appearance of the HSI as much as possible, thus ensuring the authenticity and reliability of the subsequent analysis. Unlike the conventional convolution, the PEO exhibits symmetric inverse properties, which not only ensure the stability and consistency of information between disparate spectra but also facilitate the discernment and reinforcement of subtle differences in spatial location, i.e., spatial specificity.
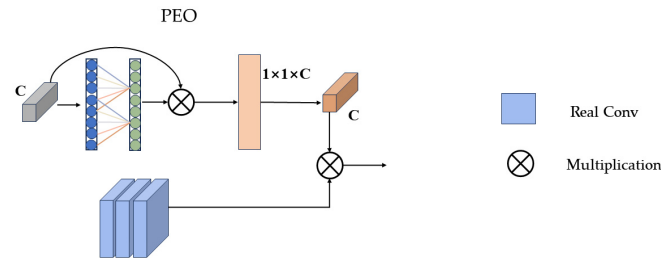

Figure 2. The detailed structure of the PEO within the MSFE block.

The PEO is capable of maintaining the integrity of the data when extracting spectral spatial features while simultaneously delving more profoundly into the hidden spatial structural information, thus guaranteeing the accuracy and completeness of the process.

The typical absence of elevation data in hyperspectral datasets is addressed by integrating the LiDAR data. To this end, a real-valued convolutional network is utilized in the feature extraction phase with the specific objective of processing the LiDAR data in order to extract their key features. To improve the network's training efficiency and ensure model stability, a batch normalization layer has been incorporated into the network architecture. The incorporation of this network facilitates the acceleration of the training process and the reduction of internal covariate bias, thereby markedly enhancing the model's generalization capacity, which is of particular significance for supplementing the elevation information in the HSI.

2.2.2. GA and Real-Valued Domain Multi-Dimensional Fusion Module

Real-valued CNNs transform an RGB image into single-channel feature maps in an input layer by summing outputs from the different channels. When extracting features of remote sensing images, this method ignores the relationship between channels, which may lead to loss of information. To solve the previous problems, we propose extending the CNNs from the real-number domain to the GA domain [41]. In this domain, the convolution does not decrease the order of the input layers, and it preserves information regarding the interrelationships between the channels [42]. GA neurons are a generalized description of real neurons when extended from the real domain to the GA domain. They encapsulate raw pixel inputs as multivectors, and they can be regarded as a mapping function from $(G_{p,q})^n$ to $G_{p,q}$.

$$y = \sum_{i=1}^{n} w_i \otimes_{p,q} x_i + \theta,$$ (8)

where $x_i \in G_i$, $w_i \in G_i$, and $\theta \in G_i$ represent the input, weights, and bias multivector, respectively, uses the geometric product instead of the multiplication of real neurons.

Unlike traditional CNNs, the convolutional layer based on GA algorithms utilizes a multi-dimensional convolutional kernel, thus providing multi-channel characteristics for multi-dimensional signals, as shown in Figure 3. This allows for the maintenance of internal relationships between input features. In the l-th convolutional layer, the output of the j-th neuron is represented as $y_j^{l_{conv}}$.

$$y_j^{l_{conv}} = \sum_{I \in B} g^{(l)} \left(\sum_i c_{kj}^{(l)} \cdot x_k^{(l-1)} + \theta_j^{(l)} e_I \right)$$ (9)
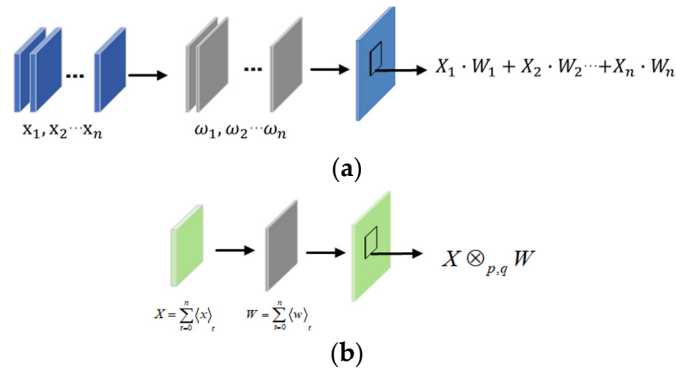
Figure 3. Convolutional performance in different domains. (**a**) Real-valued convolution; in contrast, the GA convolution (**b**) has multiple cores and can perform multi-channel extraction of multi-dimensional signals.

The *k*-th input from the preceding layer is designated as $x_k^{(l-1)}$, while the convolution kernels are represented by $c_{kj}^{(l)}$.

Applying geometric algebra convolution to remote sensing data with multiple bands and strong correlations between spectral and spatial features allows for the extraction of more valuable features. It captures the internal and external relationships between pixels and spatial features while preserving the global contextual information between spectral and spatial data and the inherent geometric structure of the HSI. Furthermore, real-valued convolution can extract inter-relationships between channels from remote sensing data, which uses local connection and weight sharing properties of convolution to avoid the instability caused by hierarchical extraction and improves the extraction efficiency. The features of F_{hsi}^{spe} are extracted through the simultaneous utilization of two branches of geometric algebraic: the field and the real range field.

$F_{hsi}^{G_{1\_1}}$ and $F_{hsi}^{C_{1\_2}}$ are defined as the two branches; the following is the first grouping operation.

$$F_{hsi}^{G_{1\_1}} = G_{s1}(F_{hsi}^{spe}) \tag{10}$$

$$F_{hsi}^{C_{1\_2}} = C_{s1}(F_{hsi}^{spe}) \tag{11}$$

where G_{si} denotes the geometric algebraic convolution by; for real-valued convolutions, it is denoted by C_{si}. Subsequently, $F_{hsi}^{G_{1\_1}}$ is split into two branches through geometric algebraic convolution, designated as $F_{hsi}^{G_{2\_1}}$ and $F_{hsi}^{G_{2\_2}}$, respectively. $F_{hsi}^{C_{1\_2}}$ is also extracted using a real-valued convolutional layer. The extracted branches are denoted by $F_{hsi}^{C_{2\_3}}$ and $F_{hsi}^{C_{2\_4}}$. This enhances the ability of the network to handle different datasets with greater resilience.

$$F_{hsi}^{G_{2\_1}}, F_{hsi}^{G_{2\_2}} = G_{s2}(F_{hsi}^{G_{1\_1}}) \tag{12}$$

$$F_{hsi}^{C_{2\_3}}, F_{hsi}^{C_{2\_4}} = C_{s2}(F_{hsi}^{C_{1\_2}}) \tag{13}$$

In order to more effectively preserve the spectral–spatial correlation state of the hyperspectral image, a multi-dimensional fusion strategy is employed to generate a more comprehensive feature representation. The fusion of $F_{hsi}^{G_{2\_1}}$ and $F_{hsi}^{G_{2\_2}}$ is conducted in the geometric algebraic domain and normalized by the Softmax function, resulting in $F_{hsi}^{G_{3\_1}}$. The same operation is performed in the real domain, yielding $F_{hsi}^{C_{3\_2}}$.

$$F_{hsi}^{G_{3\_1}} = \text{Softmax}(F_{hsi}^{G_{2\_1}} \times F_{hsi}^{G_{2\_2}}) \tag{14}$$

$$F_{hsi}^{C_{3\_2}} = \text{Softmax}(F_{hsi}^{C_{2\_3}} \times F_{hsi}^{C_{2\_4}}) \tag{15}$$

The integration of multi-scale information facilitates the capture of a more comprehensive range of features, thereby enhancing their representation. The generation of $F_{hsi}^{G_{4\_1}}$ and $F_{hsi}^{C_{4\_1}}$ is achieved through the fusion of these features in the geometric algebraic and real-number domains.

$$F_{hsi}^{G_{4\_1}} = F_{hsi}^{G_{1\_1}} \times F_{hsi}^{G_{3\_1}} \tag{16}$$

$$F_{hsi}^{C_{4\_2}} = F_{hsi}^{C_{1\_2}} \times F_{hsi}^{C_{3\_2}} \tag{17}$$

Real-valued convolution is an appropriate method for capturing local features with both local connectivity and weight-sharing properties. In contrast, geometric algebraic convolution is well-suited to handling high-dimensional data and complex geometric relationships. The combination of these two approaches can enhance feature complementarity and improve the model's generalization ability.

$$F_{hsi}^{5\_1} = \text{Concat}(F_{hsi}^{G_{4\_1}}, F_{hsi}^{C_{4\_2}}) \tag{18}$$

The GA-based convolutional layer is then employed to compute the fused feature maps, thereby maintaining the interrelationships between the extracted spectral–spatial information.

$$F_{hsi}^{G_{6\_1}} = G_{s3}(F_{hsi}^{5\_1}) \tag{19}$$

The comprehensiveness of the feature extraction procedure is improved by using GA-based convolutional layers and real-valued convolutional layers for hyperspectral feature extraction and fusion, respectively. The geometric algebra convolutional layers can capture the internal and external relationships of hyperspectral image features as well as complex spatial features. When fused with channel information extracted through real-valued convolution, this ensures that higher-level features remain stable and invariant during the deep structure and hierarchical extraction process. This network architecture takes full advantage of convolutional layers based on genetic algorithms to more effectively extract global features from hyperspectral data.

The GA-based convolutional layer allows for more efficient extraction of radar features from LiDAR data. The objective of this network architecture is to fully leverage the benefits of GA-based convolutional layers while addressing the limitations of deep networks in elevation feature extraction. The employed GA-based convolutional network incorporates not only geometric algebraic convolution but also pooling and activation functions, which are essential for executing nonlinear operations. In the geometric algebraic domain, maximum pooling is applied to the imaginary component, and the activation function for multiple vectors represents an operation that applies an activation function in the real domain to each quantity. In GA-based convolutional networks, the same sigmoid function $g^{(l)}$ is employed for each channel, which is analogous to real-valued convolutional neural networks. The proposed GA-based convolutional layer retains a greater degree of interrelated information and extracts more optimal features than the real-valued convolutional layer. By integrating these techniques, it is possible to extract the desired features from LiDAR data with greater efficiency, thus facilitating more accurate characterization for subsequent feature fusion and other tasks. We then obtain the feature representation F_{LiDAR} of LiDAR.

2.2.3. Cross-Fusion Module Based on GA

After extracting features from the HSI and LiDAR, a GA-based cross-fusion module (GACF) was designed to achieve comprehensive fusion of spectral–spatial information from HSI data and elevation–spatial information from LiDAR data. Because different remote sensing images exhibit significant differences in size, structure, and physical properties, it is necessary to consider the heterogeneity of various data sources during the fusion process. Through the incorporation a geometric algebra network, different features can be effectively fused between the real and imaginary parts of the geometric algebra, thereby

maintaining the internal and external relationships between features as well as the integrity and consistency of the overall information.

The GACF module enables efficient information exchange and fusion at the feature level. The GA-based approach overcomes the limitations of single-feature representation methods in multimodal data fusion, thus preserving the inherent structural features of data sources while extracting deep-level features. The module not only enhances the representational capacity of fused features but also ensures the stability of features.

The fusion process of the GACF module is represented as follows:

$$F_{hsi}^1 = G_{s4}(F_{hsi}^{6\_1}) \tag{20}$$

$$F_{LiDAR}^1 = G_{s5}(F_{LiDAR}) \tag{21}$$

$$F^1 = \text{Concat}(F_{hsi}^1, F_{LiDAR}^1) \tag{22}$$

In order to achieve further fusion of the features, F_{hsi}^1, F_{LiDAR}^1, and F^1 are subjected to additional convolutional processing, resulting in the generation of F_{hsi}^2, F_{LiDAR}^2, and F^2, which are cross-fused separately to obtain F^3 and F^4. The fused features of F^3 and F^4 are summed with F^2 for classification purposes.

$$F_{hsi}^2 = G_{s6}(F_{hsi}^1) \tag{23}$$

$$F_{LiADR}^2 = G_{s7}(F_{LiADR}^1) \tag{24}$$

$$F^2 = G_{s8}(F^1) \tag{25}$$

$$F^3 = F_{LiADR}^2 \times F^2 \tag{26}$$

$$F^4 = F_{hsi}^2 \times F^2 \tag{27}$$

$$F_{out} = F^2 + F^3 + F^4 \tag{28}$$

During the cross-fusion process, the completeness of information between individual features is maintained, along with the interconnections and complementarities between HSI and LiDAR features.

Our network parameters can be updated by jointly optimizing the cross-entropy loss function, denoted as

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij}\log(out_{ij}) \tag{29}$$

where N denotes the number of samples while C indicates the number of categories to be distinguished. The value of variable y_{ij} is equal to 1 when the true category of the i sample is equal to j, and it is 0 otherwise. The variable out_{ij} represents the probability that the model classifies sample i as belonging to category j.

3. Experiments and Results

All experiments were carried out using a Linux operating system using a GeForce RTX 3090 GPU and trained on the PyTorch framework. The optimizer used is Adam, and a single GPU was used for training over 200 epochs. The initial learning rate was set to 0.01, and it decreased by 1/2 every 80 epochs. To assess the reliability of the proposed model, we tested it on three remote sensing datasets.

3.1. Presentation of the Datasets

(1) Trento data capture rural areas surrounding the city of Trento, Italy, with dimensions of 600×166 pixels. The hyperspectral imagery obtained from the AISA Eagle sensor comprises 63 spectral bands, covering spectral wavelengths from 420.89 to 989.09 nm with a spectral resolution of 9.2 nm and a spatial resolution of 1 m. The LiDAR imagery captured by the Optech ALTM 3100EA sensor is a single-channel image containing elevation heights corresponding to ground locations. The dataset

comprises 30,214 ground truth pixels, which have been categorized into six classes. Detailed information regarding the sample quantities for each class is provided in Table 1, and the visualization is demonstrated in Figure 4.

Table 1. The number of samples in each class in the Trento dataset.

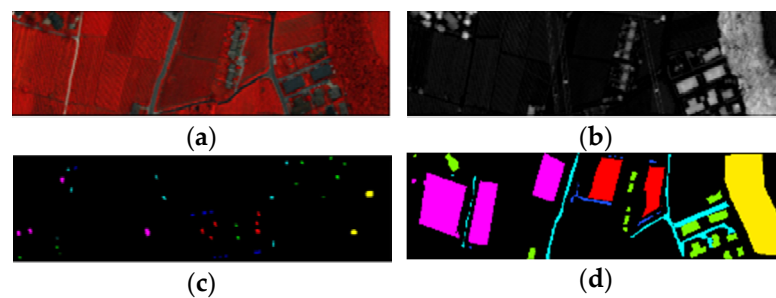| NO.	Class	Train	Test	Total
1	Apple trees	129	3905	4034
2	Buildings	125	2778	2903
3	Ground	105	374	479
4	Woods	154	8969	9123
5	Vineyard	184	10,317	10,501
6	Roads	122	3052	3174
	Total	819	29,395	30,214

Figure 4. Visualisation of the Trento dataset. (**a**) False-color image for HSI. (**b**) Grayscale image for LiDAR. (**c**) Distribution of training samples. (**d**) Distribution of testing samples.

(2) MUUFL data is a registered aerial hyperspectral–LiDAR dataset. It was acquired through a single aerial flight over the Southern Mississippi University Gulf Park Campus on 8 November 2010, capturing two modal images. The images have dimensions of 325 × 220 pixels. The hyperspectral data acquired using the CASI-15000 imaging sensor include 64 spectral bands ranging from 375 to 1050 nm, with a spatial resolution of 0.54 m. The LiDAR data consist of two elevation rasters, providing elevation heights corresponding to ground positions. The dataset comprises 53,687 ground truth pixels, which have been categorized into 11 classes. Detailed information regarding the sample quantities for each class is provided in Table 2, and the visualization is demonstrated in Figure 5.

Table 2. The number of samples in each class in the MUUFL dataset.

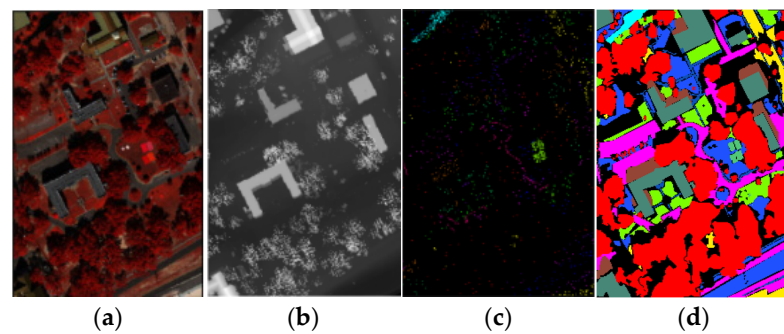| NO.	Class	Train	Test	Total
1	Trees	2325	20,921	23,246
2	Mostly grass	427	3843	4270
3	Mixed ground surface	689	6193	6882
4	Dirt and sand	183	1643	1826
5	Road	669	6018	6687
6	Water	47	419	466
7	Building shadow	224	2009	2233
8	Building	624	5616	6240
9	Sidewalk	139	1246	1385
10	Yellow curb	19	164	183
11	Cloth panels	27	242	269
	Total	5373	48,314	53,687

Figure 5. Visualisation of the MUUFL dataset. (**a**) False-color image for HSI. (**b**) Grayscale image for LiDAR. (**c**) Distribution of training samples. (**d**) Distribution of testing samples.

(3) Houston 2013 data were obtained by the University of Houston through the National Science Foundation (NSF)-funded Center for Airborne Laser Ranging (NCALM). The corresponding scene information was collected using the ITRES CASI-1500 imaging sensor over the University of Houston's campus and its surrounding neighborhoods. The images have dimensions of 349×1905 pixels. The hyperspectral dataset includes 144 spectral bands spanning from 364 to 1046 nm, with a spectral resolution of 10 nm and a spatial resolution of 2.5 m. The LiDAR dataset is a single-channel image providing elevation heights corresponding to ground positions. The dataset comprises 15,029 ground truth pixels, which have been categorized into 15 classes. Detailed information regarding the sample quantities for each class is provided in Table 3; the visualization is demonstrated in Figure 6.

Table 3. The number of samples in each class in the Houston dataset.

| NO.	Class	Train	Test	Total
1	Healthy grass	198	1053	1251
2	Stressed grass	190	1064	1254
3	Synthetic grass	192	505	697
4	Trees	188	1056	1244
5	Soil	186	1056	1242
6	Water	182	143	325
7	Residential	196	1072	1268
8	Commercial	191	1053	1244
9	Road	193	1059	1252
10	Highway	191	1036	1227
11	Railway	181	1054	1235
12	Parking lot 1	192	1042	1233
13	Parking lot 2	184	285	469
14	Tennis court	181	247	428
15	Running track	187	473	660
	Total	2832	12,197	15,029
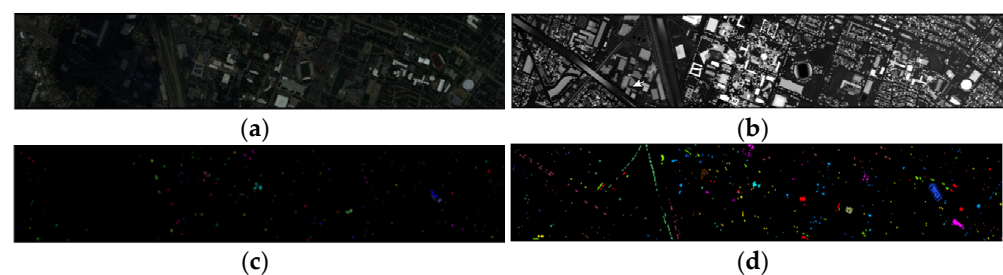
Figure 6. Visualisation of the Houston dataset. (**a**) False-color image for HSI. (**b**) Grayscale image for LiDAR. (**c**) Distribution of training samples. (**d**) Distribution of testing samples.

3.2. Multi-Source Data Fusion Analysis

The effectiveness of the comparative classification methods used in our experiments was assessed using three standard metrics: overall accuracy (OA), average accuracy (AA), and Kappa coefficient ($Kappa$). OA was previously employed to indicate the ratio of correctly classified pixels to the total number of pixels. AA is the average accuracy across all categories. $Kappa$ is an indicator used for consistency testing., adjusted for the level of agreement that could occur by chance, expressed as a percentage. Higher values of the three metrics (OA, AA, and $Kappa$) indicate better classification performance in remote sensing image classification tasks. The definitions of these indices are as follows:

$$OA = \frac{N_c}{N_a} \tag{30}$$

$$AA = \frac{1}{C} \sum_{i=1}^{C} \frac{N_c^i}{N_a^i} \tag{31}$$

$$Kappa = \frac{OA - p_e}{1 - p_e} \tag{32}$$

where N_c represents the number of correctly identified samples and N_a represents the total number of samples. N_c^i and N_a^i represent the number of samples in each category in N_c and N_a, respectively. The hypothesized probability of chance congruence, p_e, is calculated using Equation (33).

$$p_e = \frac{N_r^1 \times N_p^1 + \ldots N_r^i \times N_p^i + \ldots + N_r^C \times N_p^C}{N_a \times N_a} \tag{33}$$

where the counts of actual samples and predicted samples for each class are denoted by N_r^i and N_p^i, respectively.

To demonstrate the superiority of the GACF module and the advantages of multi-source data fusion in classification, we assessed the performance of classification using HSI and LiDAR data alone versus multi-source data. We trained HSI and LiDAR data using the MSFE module and the GRMF module and compared their classification abilities with the results obtained using both data simultaneously. Figure 7 presents the outcomes of the distinct classification methodologies.
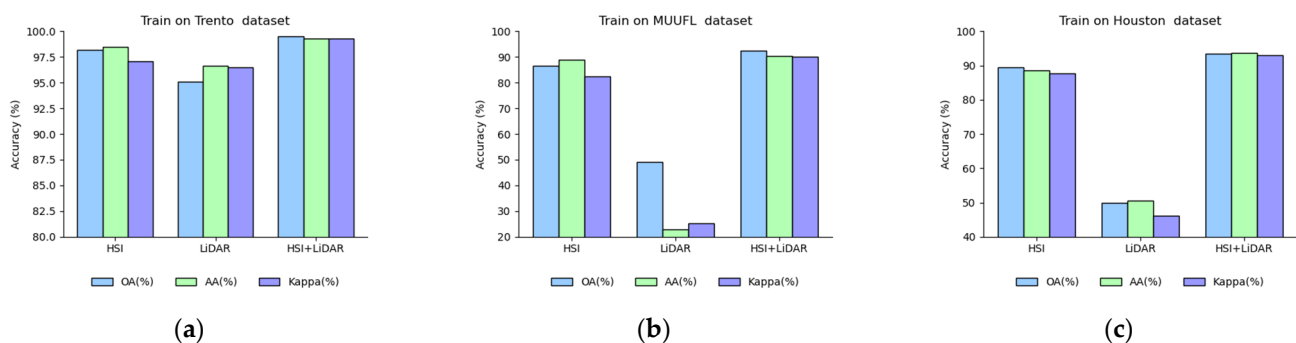

Figure 7. Classification effects of different modes. (**a**) Trento dataset. (**b**) MUUFL dataset. (**c**) Houston dataset.

Figure 7 shows that using multiple sources of data for classification resulted in better performance in all three datasets compared to using species data alone. In the Houston dataset, using the HSI alone improved the OA, AA and $Kappa$ by almost 40% compared to using LiDAR alone. In the MUUFL dataset, the OA improved by about 40%, while the AA and $Kappa$ improved by almost 57%. After fusing the two data sources, both datasets achieved an OA of over 90%, and the AA and $Kappa$ were also improved.

This demonstrates that LiDAR data can effectively compensate for the missing elevation information in HSI data. Our proposed fusion network can then better combine the two sets of data features to achieve high-precision classification.

3.3. Classification Performance

We investigated the performance impact of data blocks at different sizes (7×7, 9×9, 11×11, 13×13, 15×15, and 17×17) and different batches (16, 32, 64, and 128), and the outcomes are presented in Figures 8–10. In the Trento dataset, the highest *OA* value was achieved for a batch size of 16 when the dataset size was the same, where the highest *OA* value of 99.5% was achieved for a scale size of 17×17. In the MUUFL dataset, when the dataset batches are the same, the *OA* value peaks for a size of 13×13, where the *OA* reaches a maximum of 92.88% for a batch size of 128. In the Houston dataset, the *OA* peaked at 94.46% at a batch size of 32 and a scale size of 11×11.
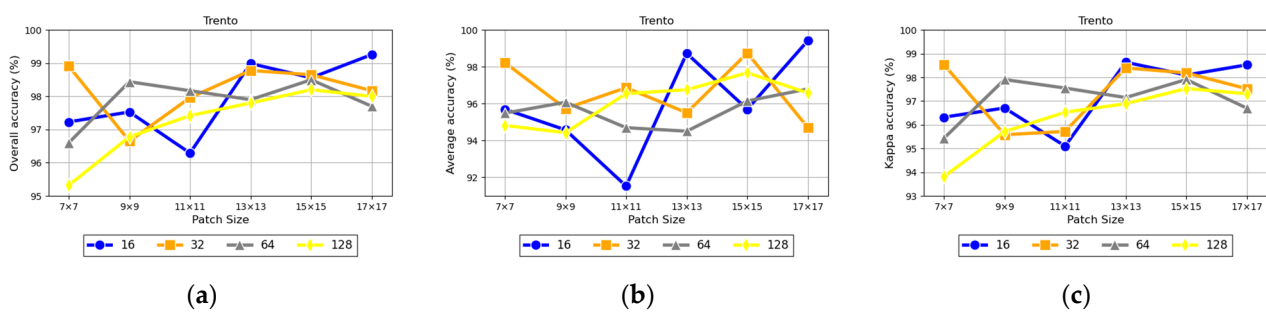

Figure 8. Trento dataset. (**a**) *OA*. (**b**) *AA*. (**c**) *Kappa*.
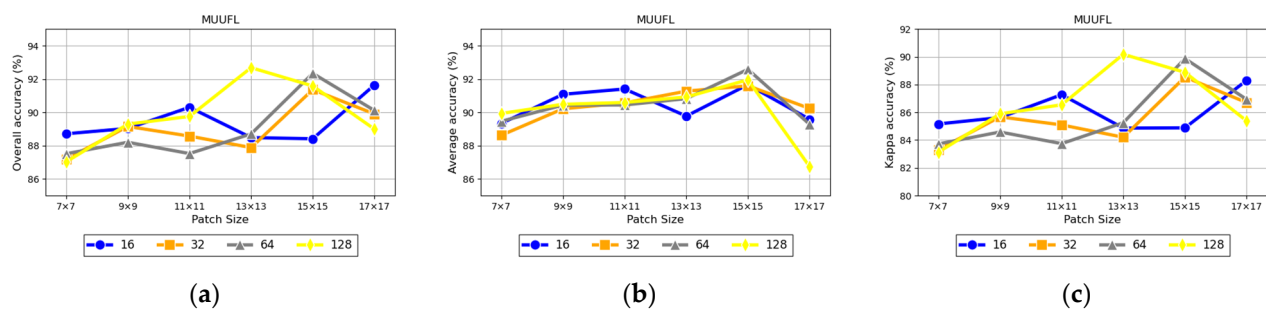
Figure 9. MUUFL dataset. (**a**) *OA*. (**b**) *AA*. (**c**) *Kappa*.
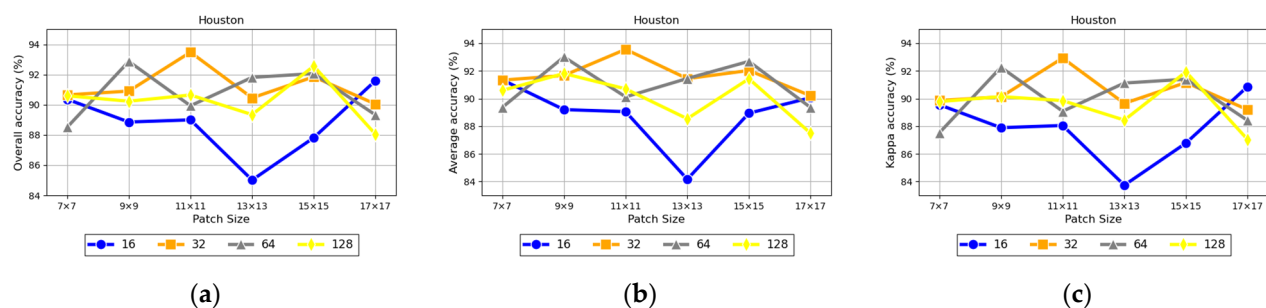
Figure 10. Houston dataset. (**a**) *OA*. (**b**) *AA*. (**c**) *Kappa*.

3.4. Ablation Study

The proposed network includes the MSFE module, which preserves the integrity of the information and the corresponding spatial structure of the HSI. In addition, the GRMF module mines the internal and external relationships and the overall nature of the HSI spectral–spatial features. Finally, the GACF module is the module in the network that

performs the fusion of different data features. These three modules work in concert to enhance the classification performance of remotely sensed data. To assess the effectiveness of MSFE, GRMF, and GACF, we conducted ablation experiments. These experiments helped determine the key modules and guided future research.

Both the MSFE and the GRMF use a simple real-valued conv2d for transformation. For the fusion module, the sole fusion technique employed is summation fusion. We utilized identical experimental configurations and evaluated the final classification outcomes for all three datasets. Table 4 shows the experimental results and the best results are in bold.

Table 4. Results for different combinations of modules in the Trento, MUUFL, and Houston datasets.

Module		Various Combinations						
GAFE **GRMF** **GACF**	**Metrics**	$\checkmark$ $\times$ $\times$	$\times$ $\checkmark$ $\times$	$\times$ $\times$ $\checkmark$	$\checkmark$ $\checkmark$ $\times$	$\checkmark$ $\times$ $\checkmark$	$\times$ $\checkmark$ $\checkmark$	$\checkmark$ $\checkmark$ $\checkmark$
Trento	*OA* (%)	98.12	97.97	97.31	99.12	98.21	97.81	**99.26**
	AA (%)	96.59	96.57	95.29	99.22	96.72	95.83	**99.43**
	Kappa × 100 (%)	97.48	97.27	96.40	99.34	97.60	97.06	**98.53**
MUUFL	*OA* (%)	84.19	84.71	84.84	90.30	86.61	85.51	**92.68**
	AA (%)	85.04	89.65	86.47	**91.61**	87.85	88.24	90.92
	Kappa × 100 (%)	79.35	80.16	80.24	87.28	82.48	81.21	**90.18**
Houston	*OA* (%)	88.32	87.43	88.14	89.74	88.51	89.28	**93.49**
	AA (%)	88.74	87.23	87.40	90.70	88.75	89.32	**93.56**
	Kappa × 100 (%)	87.29	86.35	87.07	88.86	87.52	88.35	**92.93**

Table 4 shows that using MSFE + GRMF and MSFE + GACF in all three datasets improves the *OA* compared to using MSFE alone. In the Trento dataset, the *OA*, *AA* and *Kappa* have all improved, and, in the MUUFL dataset, the *OA* and *AA* increased by approximately 6% and 2%, respectively, while Kappa increased by 8% and 2%, respectively. When using MSFE + GRMF, the *OA* and *AA* in the Houston dataset improved by 1.42% and 1.96%, respectively, while Kappa increased by 1.57%. This confirms the ability of GRMF to explore fine spatial information in remote sensing images and preserve the overall relationships between internal and external features, as well as the fusion of multi-source information through GACF.

Upon comparing GRMF alone with both MSFE + GRMF and GRMF + GACF, in the Trento dataset, there is a minimal difference in classification performance between the use of GRMF and the use of GRMF + GACF. However, the *OA* improves by at least 1% when using MSFE + GRMF, suggesting that the two modules collaborating during feature extraction can capture spatial and spectral features more efficiently. In the MUUFL dataset, compared to GRMF alone, both MSFE + GRMF and GRMF + GACF improved the *OA*, *AA* and *Kappa*; MSFE + GRMF improved the *OA*, *AA* and *Kappa* by 5.59%, 1.96%, and 7.21%, respectively. In the Houston dataset, the other two methods achieved improvements in the *OA* by about 2.32% and 1.85%, *AA* by about 3.47% and 2.09%, and *Kappa* by 2.51% and 2% over GRMF alone. These findings indicate that MSFE can be used to compensate for spatial information in remote sensing data and to thoroughly explore the spectral–spatial details.

Simultaneous use of MSFE + GACF and GRMF + GACF resulted in a slight improvement in the *OA*, *AA* and *Kappa* compared to using GACF alone in both the Trento and Houston datasets. Meanwhile, in the MUUFL dataset, the *OA* improved by 1.77% and 0.67%, *AA* by 1.38% and 1.77%, and *Kappa* by 2.24% and 0.97% compared to using GACF alone. The highlights the significance of the feature extraction module in extracting spatial and spectral features from remote sensing data.

The most favorable classification outcomes were achieved when all three modules were utilized together, resulting in significant improvements in the *OA*, *AA* and *Kappa*. This demonstrates that the three designed modules not only achieve good classification results separately but also when combined.

Concurrently using MSFE, GRMF, and GACF can fully utilize the role of each module in maximizing the extraction of spatial and spectral feature details in remote sensing data while retaining the interrelated structural domain information. This leads to a fully integrated set of features for classification.

3.5. Comparative Experimental Analysis

To substantiate the distinction of our proposed model, we conducted comparisons with various network models designed for the classification of HSI and LiDAR data. The methods for comparison include the Interleaved Perception CNN (IP-CNN) [16], the Adaptive Mutual Learning Multimodal Data Fusion Network (AM³-Net) [43],the Cross-Channel Reconstruction Module (CCR-Net) [44], the Deep Codec Network (End-Net) [18], the Hierarchical Random Wandering Network (HRWN) [45], Coupled Adversarial Learning Classification (CALC) [20], and the Multiscale Spatial Spectral Network (AMSSE) [46]. To ensure fairness in the experiments, the same training and test samples were utilized for comparison purposes. Tables 5–7 show the detailed results of the various methods applied to the three experimental datasets and the best results are in bold. These include the classification accuracy for each category as well as the *OA*, *AA* and *Kappa* values. Specifically, CCR-Net, IP-CNN, and End-Net use feature-level fusion, while and HRWN use decision fusion strategies. CALC and AMSSE, on the other hand, use both characteristic-level and decision-level fusion methods. Most of the architectures employed in this study are based on deep CNNs. Based on Tables 5–7, the following conclusions can be drawn.

Table 5. *OA*, *AA* and *Kappa* in the Trento dataset under different algorithms.

NO.	Class Name	Performance							
		AM³-Net	End-Net	CCR-Net	IP-CNN	AMSSE	HRWN	CALC	Proposed
1	Apple trees	96.92	89.52	79.76	98.04	99.95	98.85	98.49	**100.00**
2	Buildings	99.59	95.73	95.49	97.38	**99.33**	95.36	96.47	98.85
3	Ground	97.96	92.44	80.86	**99.29**	82.81	96.25	98.12	89.57
4	Woods	99.03	96.53	**100.00**	98.84	**100.00**	98.88	99.98	**100.00**
5	Vineyard	95.55	86.47	**100.00**	97.39	**100.00**	87.49	99.88	**100.00**
6	Roads	92.40	88.59	99.42	72.78	**99.71**	94.48	92.10	99.24
	OA (%)	96.88	91.12	96.50	95.21	98.95	94.05	98.57	**99.26**
	AA (%)	96.91	91.55	92.59	93.95	94.30	95.21	97.51	**99.43**
	Kappa × 100 (%)	95.83	88.29	98.61	93.85	99.16	92.10	98.01	**98.53**

Table 6. *OA*, *AA* and *Kappa* in the MUUFL dataset under different algorithms.

NO.	Class Name	Performance							
		AM³-Net	End-Net	CCR-Net	IP-CNN	AMSSE	HRWN	CALC	Proposed
1	Trees	70.85	82.93	89.05	91.23	91.03	90.08	92.88	**97.46**
2	Mostly grass	76.00	78.50	90.27	75.05	**89.85**	71.98	90.58	86.41
3	Mixed ground surface	54.01	69.25	51.28	77.69	76.63	79.39	**83.21**	82.79
4	Dirt and sand	76.81	83.21	88.37	84.20	95.41	74.60	95.05	**95.50**
5	Road	63.85	87.85	87.90	**92.22**	89.93	78.25	85.70	89.73
6	Water	**100.00**	97.65	99.68	72.48	99.68	97.45	100.00	100.00
7	Building shadow	75.91	87.56	**96.78**	83.07	94.96	84.22	96.64	95.02
8	Building	86.99	91.66	94.79	92.98	94.22	96.64	**97.67**	94.44

Table 6. *Cont.*

NO.	Class Name	Performance							
		AM³-Net	End-Net	CCR-Net	IP-CNN	AMSSE	HRWN	CALC	Proposed
9	Sidewalk	59.01	71.15	**84.86**	69.18	81.38	60.51	83.48	73.93
10	Yellow curb	**100.00**	92.56	90.91	64.76	90.91	68.33	93.94	84.85
11	Cloth panels	97.98	97.27	99.16	79.50	98.32	89.89	98.32	**100.00**
	OA (%)	70.42	82.53	85.07	85.40	89.45	89.45	91.15	**92.68**
	AA (%)	78.31	85.46	88.46	80.30	91.12	91.12	92.49	**90.92**
	Kappa × 100 (%)	62.91	77.66	84.47	81.04	86.20	86.20	88.38	**90.18**

Table 7. *OA*, *AA* and *Kappa* in the Houston dataset under different algorithms.

NO.	Class Name	Performance							
		AM³-Net	End-Net	CCR-Net	IP-CNN	AMSSE	HRWN	CALC	Proposed
1	Healthy grass	91.10	82.15	83.10	88.30	83.10	**92.10**	88.03	82.91
2	Stressed grass	98.13	83.65	84.87	85.62	**100.00**	98.37	95.08	96.15
3	Synthetic grass	**100.00**	**100.00**	99.80	85.10	99.01	**100.00**	99.75	99.01
4	Trees	92.41	93.09	93.28	93.92	93.84	93.97	97.78	**98.58**
5	Soil	96.34	99.91	99.53	95.90	99.62	99.91	97.83	**99.72**
6	Water	92.57	95.10	95.80	94.69	95.10	93.71	**97.38**	89.51
7	Residential	82.20	81.90	88.53	92.09	**93.47**	84.79	93.34	89.09
8	Commercial	79.43	76.35	81.11	78.04	77.78	77.07	89.14	**92.12**
9	Road	75.32	84.89	88.10	84.89	88.20	85.57	86.83	**95.09**
10	Highway	93.41	81.47	61.49	62.69	80.00	92.01	79.71	**96.24**
11	Railway	82.86	83.58	82.83	81.19	97.15	80.28	92.08	**98.01**
12	Parking lot 1	84.67	91.35	**94.24**	85.40	88.38	86.80	82.12	83.00
13	Parking lot 2	37.30	83.51	86.32	83.65	81.40	51.72	**92.27**	84.21
14	Tennis court	**100.00**	**100.00**	93.12	89.87	**100.00**	99.28	98.71	**100.00**
15	Running track	**100.00**	98.31	99.79	97.76	92.60	**100.00**	97.76	99.79
	OA (%)	87.75	87.24	87.15	85.39	90.85	89.42	91.15	**93.49**
	AA (%)	87.10	89.02	88.79	86.62	91.43	89.10	92.52	**93.56**
	Kappa × 100 (%)	86.71	85.35	86.28	84.20	90.07	88.52	90.44	**92.93**

First, feature-level fusion methods, such as IP-CNN, that fuse features achieve preservation of the complementary structure of the HSI and LiDAR, as well as the integrity of the fusion information. Nevertheless, the utilization of the Gram matrix for the purpose of preserving multi-source complementary information results in the accumulation of a considerable amount of superfluous data, rendering the extraction of pivotal information a challenging endeavor. This allows for more effective exchange of feature information, resulting in clearer features on the edges and structures of the information, ultimately leading to a clearer appearance of the fused features.

Second, the classification accuracy can be improved by implementing a decision-level fusion strategy. AM³-Net represents both overall and local information through shallow and deep appearance features. This is due to the strong complementary relationship between shallow and deep information. Different weights are utilized when fusing the three levels of features. Similarly, HRWN achieves classification based on the random wandering layer of the LiDAR weight map, and both methods achieve better classification accuracy.

Third, AMSSE-Net utilizes MMHF to capture spatial information and feature maps of different sensory fields and fuses the data using a combined strategy involving characteristic levels and weighted merging. CALC improves the classification performance of the model by utilizing both characteristic-level and decision-level fusion methods. In the fusion stage, advanced semantic and complementary information is mined and utilized, which increases adversarial training and efficiently maintains intricate details in both HSI and LiDAR data. The adaptive probabilistic fusion strategy further enhances classification performance.

To improve the feature extraction and fusion of various remote sensing data, we pro-pose the GASSF-Net model. Considering the above methods and challenges faced, we first extracted the features of HSI data through the PEO before performing feature fusion. Subsequently, we utilize the GRMF module to simultaneously mine the spectral–spatial information of the HSI through GA networks and real-valued networks, which ensures the correlation of internal and external relationships between high-dimensional signals and maintains the comprehensive information while mining the advanced semantic details of the HSI.To complement the HSI features, we utilize a GA-based network to extract elevation and spatial details from LiDAR data. Finally, we introduce a cross-fusion module that effectively keeps detailed information in both HSI and LiDAR data to achieve interactive complementation of rich information, thus better fusing multi-source data using GA-based methods.

The performance of the different comparison methods was similarly evaluated for the Trento and MUUFL datasets with fewer feature classes and higher spatial resolution. The visualized classification figures are shown in Figures 11 and 12. For the Trento dataset, our method achieved the best results for the *OA* and *AA*, with higher results for *Kappa*. The highest level was achieved in five categories. Moreover, our method achieves the highest *OA* and *Kappa* values for the challenging MUUFL dataset, and it is capable of achieving high classification accuracies for classes with exceptionally large sample sizes (e.g., class 1 trees).

For the Houston dataset, the best results were obtained for all three metrics. The visualized classification graph is shown in Figure 13, from which we can see that we achieve smoother classification results. This shows that our method can mine more features and achieve higher classification accuracy, and our method aligns more closely with the ground truth map.

To quantitatively analyze the computational cost of different models, Table 8 shows the total parameters of the neural network models (in millions), the training time (in seconds), and the testing time (in seconds) for different models on the Houston dataset. It can be seen that End-Net, as a lightweight network, has a smaller number of parameters and shorter training and testing times, but it is less efficient in learning. In contrast, IP-CNN has a significant increase in the number of parameters due to two-stage training, while AMSSE-Net and AM^3-Net deeply mine the spectral features through pairwise ensembles, which means that the training and testing time is mainly consumed in this part. Our proposed model maintains a moderate level of training and testing time, but it still achieves the best classification results, indicating its high training and testing efficiency while guaranteeing high classification accuracy.
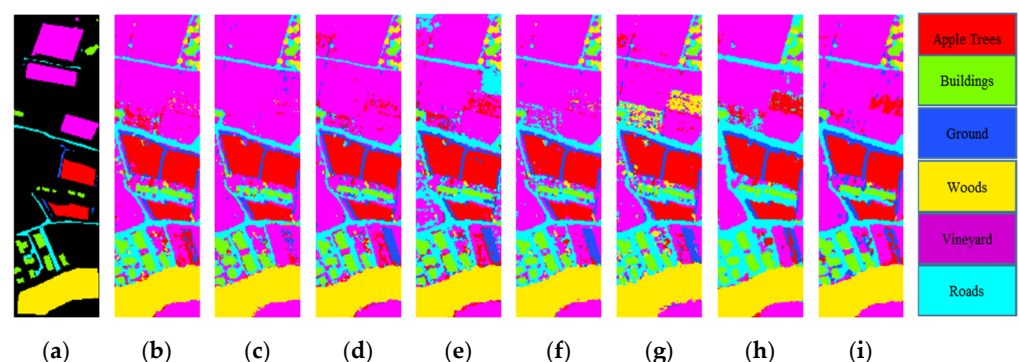

Figure 11. Visualization and classification maps for the Trento dataset. (**a**) Ground truth map. (**b**) AM^3-Net (**c**) End-Net. (**d**) CCR-Net. (**e**) IP-CNN. (**f**) AMSSE-Net. (**g**) HRWN. (**h**) CALC. (**i**) Proposed.
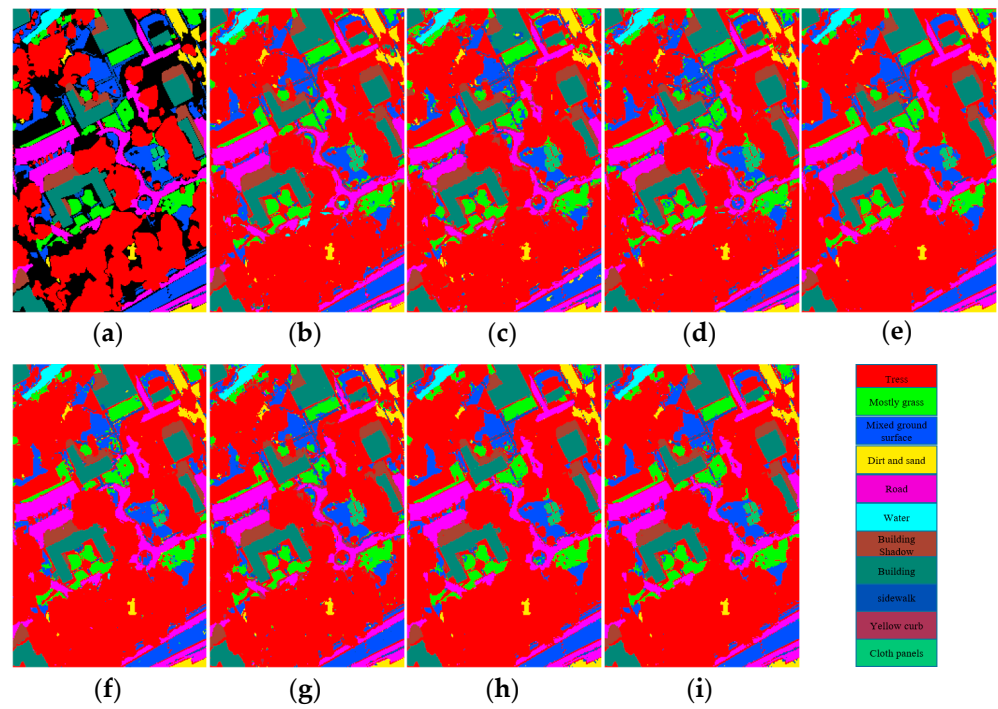
Figure 12. Visualization and classification maps for the MUUFL dataset. (**a**) Ground truth map. (**b**) AM^3-Net. (**c**) End-Net. (**d**) CCR-Net. (**e**) IP-CNN. (**f**) AMSSE-Net. (**g**) HRWN. (**h**) CALC. (**i**) Proposed.
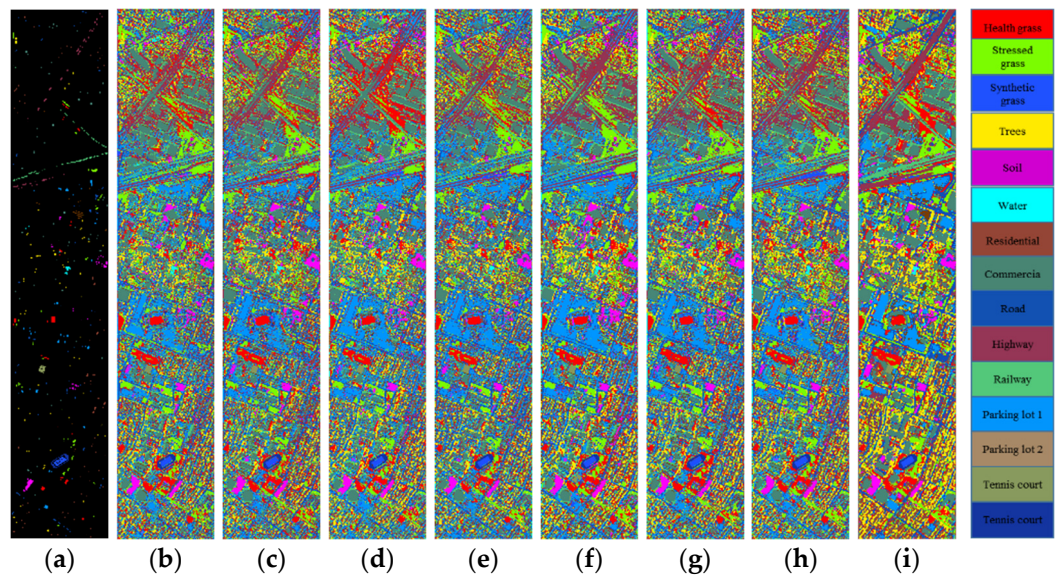
Figure 13. Visualization and classification maps for the Houston dataset. (**a**) Ground truth map. (**b**) AM^3-Net. (**c**) End-Net. (**d**) CCR-Net. (**e**) IP-CNN. (**f**) AMSSE-Net. (**g**) HRWN. (**h**) CALC. (**i**) Proposed.

Table 8. Analysis of the model complexity of various models on the Houston 2013 dataset.

Complexity	AM^3-Net	End-Net	CCR-Net	IP-CNN	AMSSE	HRWN	CALC	Proposed
Parameters (M)	2.78	0.088	0.071	3.45	2.15	1.37	0.327	2.32
Training Time (s)	81.97	91.41	327.9	970	580	45.74	430	415
Testing Time (s)	3.64	1.79	1.53	4.22	2.96	1.35	2.75	2.14

4. Discussion

The GASSF-Net method, as proposed in this paper, has made notable advancements in the field of remote sensing spectral–spatial information extraction. A comprehensive evaluation of GASSF-Net using several standard datasets, including Trento, MUUFL, and Houston, demonstrates that the method excels in terms of feature extraction accuracy, classification accuracy, and generalization ability. The ablation experiments on single-source and multi-source data demonstrate that the LiDAR data features extracted through geometric algebra can effectively address the limitations of HSI data regarding elevation. Additionally, the integration of diverse data features allows for the extraction of a more comprehensive range of information. The results of the ablation experiments on different modules also demonstrate the crucial role of geometric algebra in this method. The fundamental innovation of GASSF-Net lies in its incorporation of geometric algebraic techniques for the extraction of spectral–spatial information. The geometric algebra network can comprehensively capture the complex internal and external relationships and comprehensive information within the data, and when combined with the channel information extracted through the real-valued network, it ensures efficient encoding and interpretation of features. The branch fusion method significantly enhances the extraction of spectral spatial features, thereby enabling GASSF-Net to more accurately reflect the information present in remote sensing images. Concurrently, the GA-based cross-fusion module achieves feature-level fusion and complementarity while maintaining the integrity of the unique spectral spatial information of each data source. This module is capable of dealing with the complexity of different data sources and effectively fusing multimodal data. In comparison to other methods, GASSF-Net exhibits superior performance in feature extraction accuracy, classification accuracy, and generalization ability.

While GASSF-Net demonstrates satisfactory performance in the current experiments, there are still some avenues that warrant further investigation. Firstly, although this paper focuses on the fusion of hyperspectral images and LiDAR data, the framework of GASSF-Net can be extended to other types of remote sensing data, including synthetic aperture radar (SAR) images or multispectral images. Secondly, the incorporation of geometric algebra with its non-commutative multiplication results in an increase in the computational complexity of the algorithm. We think this issue will be addressed with the advent of geometric-algebra-parallel computing technology. Furthermore, there is scope for further optimization of the genetic algorithm's design to enhance the efficiency and effectiveness of the cross-fusion module and to fully use the extracted feature information.

GASSF-Net offers novel insights into the processing of remote sensing data and provides robust support for classification tasks in real-world applications. To illustrate, in certain domains, such as land cover classification, environmental monitoring, and urban planning, GASSF-Net can enhance the precision and dependability of classification, thereby furnishing more precise guidance for pertinent decisions. This design not only enhances the comprehensiveness and robustness of feature representation but also effectively addresses the pivotal challenges inherent to multi-source data fusion. The incorporation of a genetic algorithm in feature fusion represents a significant advancement over existing methodologies, as it is better equipped to navigate the intricate relationships between data sources, thereby enhancing classification performance.

5. Conclusions

This paper innovatively proposes a multi-source remote sensing image fusion classification method that combines geometric algebra, GASSF-Net. This method is the first to adopt a geometric algebra network to comprehensively capture the complex spectral–spatial information and the comprehensive context within remote sensing images. Through the multi-dimensional representation provided by the geometric algebra network, we successfully address the internal and external correlations between spectral and spatial features, which are often difficult to capture with traditional real-valued CNNs, along with more refined spatial details. Additionally, we designed a GA-based cross-fusion module

(GACF), which deeply explores the inherent structural information between spectral bands while preserving the differences between different data sources. This approach achieves a comprehensive fusion of HSI and LiDAR data at the feature level. The module not only ensures efficient information exchange during multi-source data fusion but also addresses the consistency of heterogeneous data feature representation during the fusion process, thereby enhancing the comprehensiveness and robustness of feature representation. Experiments conducted on three well-known remote sensing datasets demonstrate that GASSF-Net significantly improves land cover classification accuracy and model generalization ability.

References

1. Li, J.; Zheng, K.; Liu, W.; Li, Z.; Yu, H.; Ni, L. Model-guided coarse-to-fine fusion network for unsupervised hyperspectral image super-resolution. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 5508605. [CrossRef]
2. Jahan, F.; Zhou, J.; Awrangjeb, M.; Gao, Y. Integration of heterogeneous features from co-registered hyperspectral and LiDAR data for land cover classification. In Proceedings of the 2018 10th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS), Beijing, China, 19–20 August 2018; pp. 1–6.
3. Cheng, G.; Yang, C.; Yao, X.; Guo, L.; Han, J. When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2811–2821. [CrossRef]
4. Zhang, Y.; Zhou, Y.; Xu, S.; Hong, D.; Gao, H.; Li, C.; Zhong, Q.; Zhang, B. Depthwise separable convolutional autoencoders for hyperspectral image change detection. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 5505105.
5. Li, J.; Zheng, K.; Gao, L.; Ni, L.; Huang, M.; Chanussot, J. Model-Informed Multistage Unsupervised Network for Hyperspectral Image Super-Resolution. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5516117. [CrossRef]
6. Kuras, A.; Brell, M.; Teien, S.; Burud, I.; Thiis, T. Feature-level based hyperspectral and LiDAR data fusion for urban analysis. In Proceedings of the 2023 Joint Urban Remote Sensing Event (JURSE), Heraklion, Greece, 17–19 May 2023; pp. 1–4.
7. Li, N.; Xue, J.; Jia, S. Spectral context-aware transformer for cholangiocarcinoma hyperspectral image segmentation. In Proceedings of the 2022 5th International Conference on Image and Graphics Processing, Beijing China, 7–9 January 2022; pp. 209–213.
8. Zhang, S.; Xu, M.; Zhou, J.; Jia, S. Unsupervised spatial-spectral CNN-based feature learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5524617. [CrossRef]
9. Gao, H.; Zhang, Y.; Chen, Z.; Li, C. A multiscale dual-branch feature fusion and attention network for hyperspectral images classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 8180–8192. [CrossRef]
10. Li, J.; Hong, D.; Gao, L.; Yao, J.; Zheng, K.; Zhang, B.; Chanussot, J. Deep learning in multimodal remote sensing data fusion: A comprehensive review. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102926. [CrossRef]
11. Li, Z.; Sui, H.; Luo, C.; Guo, F. Morphological convolution and attention calibration network for hyperspectral and LiDAR data classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 5728–5740. [CrossRef]
12. Li, J.; Zheng, K.; Yao, J.; Gao, L.; Hong, D. Deep unsupervised blind hyperspectral and multispectral data fusion. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6007305. [CrossRef]
13. Dian, R.; Li, S.; Kang, X. Regularizing Hyperspectral and Multispectral Image Fusion by CNN Denoiser. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 1124–1135. [CrossRef]
14. Li, J.; Zheng, K.; Li, Z.; Gao, L.; Jia, X. X-shaped interactive autoencoders with cross-modality mutual learning for unsupervised hyperspectral image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5518317. [CrossRef]
15. Wang, S.; Shu, Q.; Ma, X.; Xiao, J.; Zhou, W. Research Progress in Data Fusion of LiDAR and Hyperspectral Imaging Technology. *Remote Sens. Technol. Appl.* **2024**, *39*, 11–23.
16. Hong, D.; Gao, L.; Hang, R.; Zhang, B.; Chanussot, J. Deep encoder–decoder networks for classification of hyperspectral and LiDAR data. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 5500205. [CrossRef]
17. Zhang, M.; Li, W.; Tao, R.; Li, H.; Du, Q. Information fusion for classification of hyperspectral and LiDAR data using IP-CNN. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5506812. [CrossRef]

18. Zhang, T.; Xiao, S.; Dong, W.; Qu, J.; Yang, Y. A mutual guidance attention-based multi-level fusion network for hyperspectral and LiDAR classification. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 5509105. [CrossRef]

19. Ding, K.; Lu, T.; Fu, W.; Li, S.; Ma, F. Global–local transformer network for HSI and LiDAR data joint classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5541213. [CrossRef]

20. Lu, T.; Ding, K.; Fu, W.; Li, S.; Guo, A. Coupled adversarial learning for fusion classification of hyperspectral and LiDAR data. *Inf. Fusion* **2023**, *93*, 118–131. [CrossRef]

21. Yu, W.; Huang, H.; Zhang, M.; Shen, Y.; Shen, G. Shadow Mask-Driven Multimodal Intrinsic Image Decomposition for Hyperspectral and LiDAR Data Fusion. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5525915. [CrossRef]

22. Nitta, T. A quaternary version of the back-propagation algorithm. In Proceedings of the ICNN'95—International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995; Volume 5, pp. 2753–2756.

23. Nitta, T. An extension of the back-propagation algorithm to three dimensions by vector product. In Proceedings of the 1993 IEEE Conference on Tools with AI (TAI-93), Boston, MA, USA, 8–11 November 1993; pp. 460–461.

24. Li, H.; Li, H.; Zhang, L. Quaternion-based multiscale analysis for feature extraction of hyperspectral images. *IEEE Trans. Signal Process.* **2019**, *67*, 1418–1430. [CrossRef]

25. Voronin, V.; Semenishchev, E.; Cen, Y.; Zelensky, A.; Agaian, S. Image enhancement in a quaternion framework for remote sensing applications. In *Proc. SPIE, Artificial Intelligence and Machine Learning in Defense Applications II*; SPIE: Bellingham, WA, USA, 2020; Volume 11543, pp. 135–140.

26. Rao, S.P.; Panetta, K.; Agaian, S. Quaternion based neural network for hyperspectral image classification. In *Proc. SPIE, Mobile Multimedia/Image Processing, Security, and Applications 2020*; SPIE: Bellingham, WA, USA, 2020; Volume 11399, pp. 179–190.

27. Zhou, H.; Zhang, X.; Zhang, C.; Ma, Q. Quaternion convolutional neural networks for hyperspectral image classification. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106234. [CrossRef]

28. Hestenes, D.; Sobczyk, G. *Clifford Algebra to Geometric Calculus: A Unified Language for Mathematics and Physics*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 5.

29. Sommer, G. *Geometric Computing with Clifford Algebras: Theoretical Foundations and Applications in Computer Vision and Robotics*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.

30. Mao, X.; Shen, C.; Yang, Y.B. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *Advances in Neural Information Processing Systems 29 (NIPS 2016)*; Curran Associates: Red Hook, NY, USA, 2016.

31. Shang, F.; Hirose, A. Quaternion neural-network-based PolSAR land classification in Poincare-sphere-parameter space. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 5693–5703. [CrossRef]

32. Wang, R.; Cao, Z.; Wang, X.; Xue, W.; Cao, W. GA-STIP: Action recognition in multi-channel videos with geometric algebra based spatio-temporal interest points. *IEEE access* **2018**, *6*, 56575–56586. [CrossRef]

33. Shen, M.; Wang, R.; Cao, W. Joint sparse representation model for multi-channel image based on reduced geometric algebra. *IEEE Access* **2018**, *6*, 24213–24223. [CrossRef]

34. Li, Y. A digital image watermarking algorithm based on Clifford algebra. *Acta Electron. Sin.* **2008**, *36*, 852–855.

35. Cao, W.; Lyu, F.; He, Z.; Cao, G.; He, Z. Multimodal medical image registration based on feature spheres in geometric algebra. *IEEE Access* **2018**, *6*, 21164–21172. [CrossRef]

36. Wang, R.; He, Y.; Huang, C.; Wang, X.; Cao, W. A novel least-mean kurtosis adaptive filtering algorithm based on geometric algebra. *IEEE Access* **2019**, *7*, 78298–78310. [CrossRef]

37. Wang, R.; Shen, M.; Wang, T.; Cao, W. L1-norm minimization for multi-dimensional signals based on geometric algebra. *Adv. Appl. Clifford Algebras* **2019**, *29*, 33. [CrossRef]

38. Wang, R.; Shen, M.; Cao, W. Multivector sparse representation for multispectral images using geometric algebra. *IEEE Access* **2019**, *7*, 12755–12767. [CrossRef]

39. 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. 11534–11542.

40. Li, D.; Hu, J.; Wang, C.; Li, X.; She, Q.; Zhu, L.; Zhang, T.; Chen, Q. Involution: Inverting the inherence of convolution for visual recognition. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 12321–12330.

41. Li, Y.; Wang, Y.; Wang, R.; Wang, Y.; Wang, K.; Wang, X.; Cao, W.; Xiang, W. GA-CNN: Convolutional neural network based on geometric algebra for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5540314. [CrossRef]

42. Bhatti, U.A.; Yu, Z.; Yuan, L.; Zeeshan, Z.; Nawaz, S.A.; Bhatti, M.; Mehmood, A.; Ain, Q.U.; Wen, L. Geometric algebra applications in geospatial artificial intelligence and remote sensing image processing. *IEEE Access* **2020**, *8*, 155783–155796. [CrossRef]

43. Wang, J.; Li, J.; Shi, Y.; Lai, J.; Tan, X. AM^{3}Net: Adaptive mutual-learning-based multimodal data fusion network. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 5411–5426. [CrossRef]

44. Wu, X.; Hong, D.; Chanussot, J. Convolutional neural networks for multimodal remote sensing data classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5517010. [CrossRef]

45. Zhao, X.; Tao, R.; Li, W.; Li, H.C.; Du, Q.; Liao, W.; Philips, W. Joint classification of hyperspectral and LiDAR data using hierarchical random walk and deep CNN architecture. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7355–7370. [CrossRef]
46. Gao, H.; Feng, H.; Zhang, Y.; Xu, S.; Zhang, B. AMSSE-Net: Adaptive multiscale spatial–spectral enhancement network for classification of hyperspectral and LiDAR data. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5531317. [CrossRef]