remote sensing

MDPI

Article

Hyperspectral and LiDAR Fusion Using Deep Three-Stream Convolutional Neural Networks

Hao Li [1,2,3,*,†], Pedram Ghamisi [4,†], Uwe Soergel [2] and Xiao Xiang Zhu [1,5]

1 Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM), Arcisstr. 21, 80333 Munich, Germany; xiao.zhu@dlr.de
2 Institute for Photogrammetry (ifp), University of Stuttgart, 70174 Stuttgart, Germany; soergel@ifp.uni-stuttgart.de
3 GIScience Research Group, Institute of Geography, Heidelberg University, 69120 Heidelberg, Germany
4 Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Exploration, Chemnitzer Str. 40, D-09599 Freiberg, Germany; p.ghamisi@gmail.com
5 Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, 82234 Wessling, Germany
* Correspondence: leebobgiser316@gmail.com; Tel.: +49-6221-54-5534
† These authors contributed equally to this work.

Received: 5 September 2018 ; Accepted: 15 October 2018; Published: 16 October 2018

Abstract: Recently, convolutional neural networks (CNN) have been intensively investigated for the classification of remote sensing data by extracting invariant and abstract features suitable for classification. In this paper, a novel framework is proposed for the fusion of hyperspectral images and LiDAR-derived elevation data based on CNN and composite kernels. First, extinction profiles are applied to both data sources in order to extract spatial and elevation features from hyperspectral and LiDAR-derived data, respectively. Second, a three-stream CNN is designed to extract informative spectral, spatial, and elevation features individually from both available sources. The combination of extinction profiles and CNN features enables us to jointly benefit from low-level and high-level features to improve classification performance. To fuse the heterogeneous spectral, spatial, and elevation features extracted by CNN, instead of a simple stacking strategy, a multi-sensor composite kernels (MCK) scheme is designed. This scheme helps us to achieve higher spectral, spatial, and elevation separability of the extracted features and effectively perform multi-sensor data fusion in kernel space. In this context, a support vector machine and extreme learning machine with their composite kernels version are employed to produce the final classification result. The proposed framework is carried out on two widely used data sets with different characteristics: an urban data set captured over Houston, USA, and a rural data set captured over Trento, Italy. The proposed framework yields the highest OA of 92.57% and 97.91% for Houston and Trento data sets. Experimental results confirm that the proposed fusion framework can produce competitive results in both urban and rural areas in terms of classification accuracy, and significantly mitigate the salt and pepper noise in classification maps.

Keywords: data fusion; extinction profiles (EPs); composite kernels; convolutional neural networks (CNN); feature extraction (FE)

1. Introduction

With the rapid development of imaging techniques, it is now possible to obtain multi-sensor data captured over the same region. Such multi-sensor information can demonstrate different characteristics, including spectral information obtained by passive sensors (e.g., multispectral and hyperspectral images) [1,2], or elevation and shape information obtained by light detection and ranging (LiDAR)

sensors [3,4]. Therefore, it is important to develop robust and accurate multi-sensor fusion methods, which could integrate the complementary information from individual sources.

Urban and rural areas are inherently complex due to the existence of many classes with similar spectral responses, which makes the classification of the available classes a challenging task. It is generally optimistic to assume that a single sensor can provide enough information for classification of complex areas [5]. To be more specific, the classification of land cover classes only based on spectral signatures derived from hyperspectral images (HSI) may result in very limited discriminant capabilities, especially for different classes that are made of the same material (e.g., building roof and road both made from asphalt). Conversely, individual consideration of the LiDAR data is not enough for the classification of objects (e.g., residential and commercial buildings) with the same elevation but made of different materials, where classes can be easily mixed considering only elevation information.

Due to the above-mentioned facts, multi-sensor data fusion may lead to accurate land cover classification by taking advantage of the promising aspects of each sensor. In other words, it is natural to imagine that, through the joint use of hyperspectral and LiDAR data, one can combine detailed spectral and spatial information from the hyperspectral data with elevation information from the LiDAR data to increase the discriminating power of the subsequent classifier.

Furthermore, much research has confirmed this higher discriminating power obtained by the joint use of HSI and LiDAR data [6–14]. In [6], a generalized graph-based fusion method and morphological profiles (MPs) have been investigated for HSI and LiDAR data fusion, which could simultaneously reduce the dimensionalities in feature space and fuse heterogeneous data for accurate classification. In [7], the fusion of HSI and LiDAR data was taken into account for the classification of cloud-shadow mixed remote sensing scenes, which processed the cloud-shadow and shadow-free areas separately. This generated more specific training samples for the cloud-shadow area in order to achieve higher classification performance. In [8], the joint use of HSI and LiDAR data was explored for accurate land cover classification of both urban and rural areas, which employs extinction profiles (EPs) to extract spatial and contextual features from multisensory data sources. Then, through a graph-based fusion process, a deep learning-based classifier was adopted for further boosting of the framework performance. In [9], a novel HSI and LiDAR fusion method named the sparse and low-rank component analysis was proposed to improve classification performance, where low-rank components helped to handle spectral redundancy and sparsity properties dealt with spatial smoothness. Therefore, this method could efficiently degrade the influence of the Hughes phenomenon and lead to region-wise homogeneous classification map. In addition, the deep fusion of HSI and LiDAR data for accurate land cover classification was developed in [10], in which the feature fusion framework was purely based on deep neural networks architectures, including abstract feature extraction and classification. To sum up, it is well-founded to believe that the fusion of heterogeneous features, such as spectral, spatial and elevation features, could provide more robust and reliable signatures of different land cover classes during the classification task.

Spatial information has been confirmed to be significantly important in HSI data processing, especially for that of high spatial resolution [1]. In this case, the spatial feature extraction recently became a hot topic in the HSI community. Among the existing spatial information extraction methods, mathematical morphology profiles have attracted a lot of attention. Morphological profiles (MPs) and attribute profiles (APs) have been intensively investigated due to the capability of generating discriminating spatial information for classification [15–21]. MPs could be established by stacking a set of opening profiles and closing profiles that are reconstructed with a structuring element (SE) of increased size. In [15], MPs using morphological transformations were carefully designed for extracting informative spatial features from the high-resolution image. Based on the MPs, APs were introduced in [17] as generalized MPs, which considered using attribute filters (AFs) to produce multilevel spatial information profiles of the image. In [18,20], APs were employed to model the contextual information of the ground appearance in order to improve the performance of image classification and building extraction. In [21], APs and their extended multi-attribute profiles (EMAP) were investigated for

the fusion of HSI and LiDAR data, where heterogeneous features were integrated with a subspace multinomial logistic regression approach. The result further proved that modeling of contextual and spatial information is of great importance, especially in very high-resolution image processing. APs have been proven to be a more robust tool, compared to MPs [17], since APs are naturally based on any attributes of images (e.g., pure geometric, spatial resolution or spectral-related characteristics). Otherwise, there are still general limitations of APs, like their performance is higHy dependent on the attributes threshold, which demands the manual initialization of parameters.

Extinction profiles (EPs) were proposed in [22], aimed at addressing the main limit of APs' threshold dependence, since EPs are extrema-oriented profiles and free of threshold setting burdens [23]. EPs have been successfully investigated for their wide use in spatial feature extraction of panchromatic images [22], HSI [23], and LiDAR-derived images [24], respectively. Despite the benefit of using the spatial features for classification, traditional morphological feature extraction methods still appear to be insufficient in invariant feature learning, e.g., the morphological-level features are naturally redundant and the classification result is still affected by image noise problems (such as the salt and pepper noise).

Recently, with great developments of deep learning concept in remote sensing applications [25–29], deep learning architectures have been investigated to progressively extract high-level and abstract features, which are more reliable and invariant due to their independence from most local details of the input data [26]. Specifically, convolutional neural networks (CNN), as fundamental deep learning architectures, have been successfully designed for various feature extraction and classification applications [30–33]. In [30], a deep feature extraction method has been proposed, where the spectral-spatial hierarchical features extracted by CNN were employed for accurate land cover classification. To improve the feature representation by involving temporal information, an end-to-end recurrent CNN architecture for change detection in multispectral images has been proposed in [31]. By iteratively selecting the spectral bands from HSI, a self-improving CNN was proposed in [32], which was proven to be effective in HSI classification. In case of Synthetic Aperture Radar (SAR) images key-point matching, a Pseudo-Siamese CNN has been developed to determine the corresponding patches between SAR images and very high resolution (VHR) optical images in [33]. Moreover, other deep learning architectures (such as, generative adversarial networks and recurrent neural networks [34,35]) have also been proved to be efficient for feature learning and classification in high-dimensional data like HSI. For instance, a successful work on employing generative adversarial networks (GAN) for HSI classification has been reported in [34]. In [35], an unsupervised HSI feature extraction framework based on Conv-Dconv recurrent neural networks (RNN) has been investigated for better HSI feature leaning. Overall, it is believed that, in comparison to other "shallow" feature extraction models, deep learning architectures are able to extract high-level, hierarchical, and abstract features, which are generally more robust to the nonlinear input data [26].

In this paper, a novel framework is proposed for the fusion of HSI and LiDAR data based on CNN and composite kernels. The proposed framework designs a three-stream CNN to extract high-level and invariant spectral features (HSI), spatial features (obtained by performing EPs on HSI), and elevation features (obtained by applying EPs on LiDAR-derived data). To effectively classify the heterogeneous spectral, spatial, and elevation features obtained by the three-stream CNN, instead of feature stacking, a multi-sensor composite kernels (MCK) scheme is carefully designed based on either a support vector machine (SVM) or extreme learning machine [36] (ELM). This MCK scheme considers three individual kernels suitable for the joint use of spectral, spatial, and elevation features, as their superior performance is shown in [37]. The main contributions of this paper are described below:

1. A three-stream CNN is designed in the proposed framework, which can effectively extract high-level features from spectral as well as spatial and elevation features produced by EPs. This baseline allows us to simultaneously take advantage of heterogeneous complementary features (from HSI and LiDAR) to achieve higher discriminating power during classification tasks.

2. The proposed framework progressively combines low-level features (obtained by extinction profiles) with high-level features (obtained by CNN) for invariant feature learning. This consideration could significantly reduce the salt and pepper noise, as well as further promote the classification performance.
3. A novel fusion scheme is proposed based on multi-sensor composite kernels, where three different base kernels for spectral, spatial and elevation features are taken into account. To be more specific, the MCK scheme provides us with an efficient framework for multi-sensor data fusion, where complementary information from heterogeneous features can be joint used for accurate classification by establishing and optimizing the corresponding MCK.

The proposed fusion framework is tested in both urban and rural scenes. As a result, the proposed method achieves significant noise reduction in classification maps and competitive classification accuracy compared to state-of-the-art approaches.

The rest of this paper is organized as follows: Section 2 explains the design of the proposed fusion framework in detail. Section 3 is devoted to the description of experiment details on two commonly used multi-sensor data sets. Section 4 reports evaluation results and detail comparisons between the proposed framework and state-of-the-art methods. At last, Section 5 gives the main concluding remarks and wraps up this paper.

2. Methods

2.1. Workflow of the Proposed Fusion Framework

In this section, the general workflow of the proposed fusion scheme is shown in Figure 1. First, two EPs were generated from HSI (EP_{HSI}) and LiDAR-derived data (EP_{LiDAR}), individually. These profiles can be regarded as spatial features and elevation features of the co-registered area. Next, a three-stream CNN feature extraction approach was designed to extract hierarchical high-level abstract features from spectral, spatial and elevation features, respectively. Then, two machine learning classifiers, SVM and ELM, which was embedded with a multi-sensor composite kernels scheme, were adopted to achieve the final data fusion and classification step. The detailed design with emphasis on the CNN feature extraction layers is shown in Figure 2.

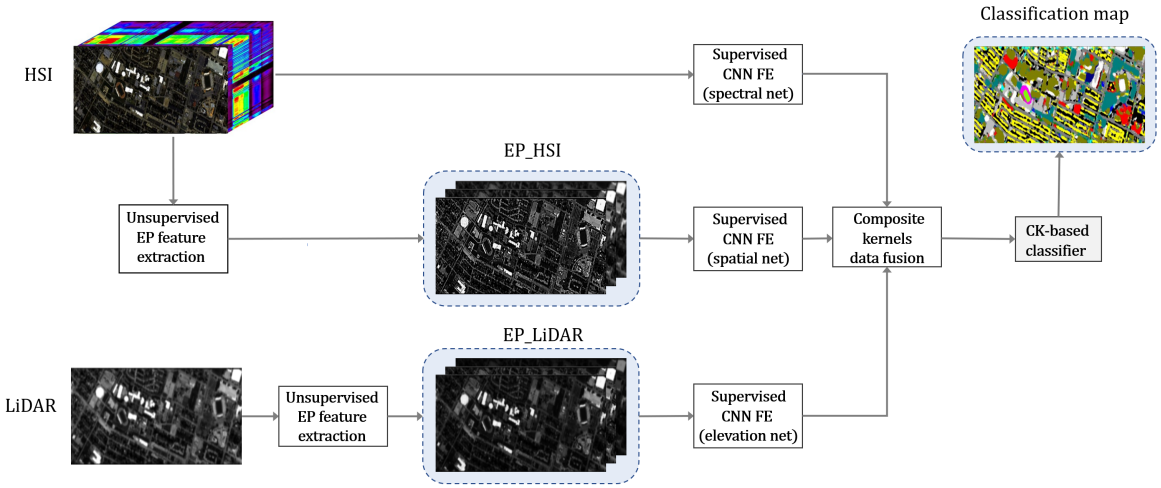

Figure 1. Workflow of the proposed multi-sensor fusion framework of hyperspectral images (HSI) and light detection and ranging (LiDAR) data for land cover classification, which consists of three main parts: unsupervised EPs feature extraction, supervised CNN feature extraction and multi-sensor composite kernels data fusion.

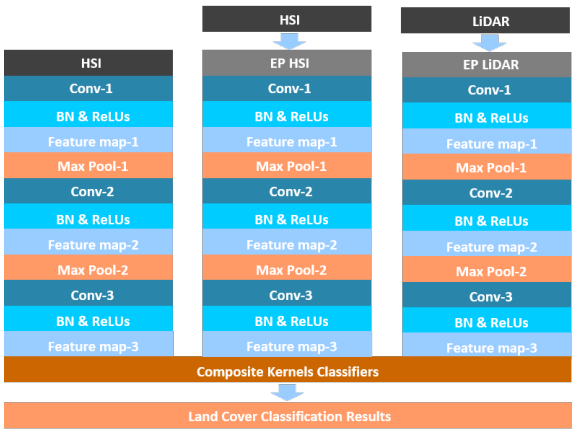

Figure 2. Detailed design of the proposed multi-sensor data fusion framework based on convolutional neural networks and composite kernels. SVM and ELM are used as classifiers for evaluating the classification performance.

2.2. Extinction Profiles

Actually, EPs [22] are based on extinction filters (EFs) [38], which are extrema-oriented connected independent filters. Different from the attribute profiles (APs), whose performance are higHy dependent on the manual initialization of the attributes threshold, EPs are naturally based on the number of extrema, which can be set automatically. To this end, EPs could get rid of the time-demanding burden of manually determining attributes-threshold values [22].

Extinction filters (EFs), which are the main building blocks of EPs, can be defined as follows: Let $Max\left(\mathbf{F}\right) = \{\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_N\}$ be a set of regional maxima of the grayscale image $\mathbf{F}$. $\mathbf{M}_i$ is an image with the same size as $\mathbf{F}$, whose entities are zero except in the positions of the pixels that include the regional maximum $\mathbf{M}_i$ (the gray-value is the value of the maximum). Each regional maximum $\mathbf{M}_i$ corresponds to a certain extinction value that is defined by Vachier [38]. The EFs of $\mathbf{F}$, which are set to keep the n maxima with the highest extinction values, is then given as:

$$\mathrm{EF}^n\left(\mathbf{F}\right) = R_{\mathbf{F}}^{\delta}\left(\mathbf{G}\right), \tag{1}$$

where $R_{\mathbf{F}}^{\delta}\left(\mathbf{G}\right)$ is the reconstruction by dilation [39] of the mask image $\mathbf{F}$ from the marker image $\mathbf{G}$. The marker image $\mathbf{G}$ is then given as:

$$\mathbf{G} = \max_{i=1}^{n}\left\{\mathbf{M}_i^{'}\right\}, \tag{2}$$

where $\mathbf{M}_1^{'}$ represents the regional maximum with the highest extinction value, $\mathbf{M}_2^{'}$ as maximum with the second highest extinction value, etc. [22].

By applying a series of thinning and thickening EFs to a grayscale image, extinction profiles (EPs) can be derived. The threshold values progressively decreases or increases to shape the profile to simultaneously extract spatial and contextual information of the input image. The mathematical definition of the EPs is given as below:

$$\mathrm{EP}(\mathbf{F}) = \left\{ \begin{array}{ll} \Pi_{\phi^{\lambda_m}}, & m = (s - i + 1), \quad \forall i \in [1, s]; \\ \Pi_{\gamma^{\lambda_m}}, & m = (i - s), \quad \forall i \in [s + 1, 2s]. \end{array} \right\}, \tag{3}$$

where $\Pi_{\phi^{\lambda}}$ is the thickening extinction profile, $\Pi_{\gamma^{\lambda}}$ is the thinning extinction profile, and $\lambda : \{\lambda_m\}\ (m = 1, \ldots, s)$ is a set of ordered threshold values (i.e., $\lambda_i \subset \lambda_j, i \leqslant j$). Here, s is the amount of thresholds [22].

To generalize the concept of extinction profiles from grayscale images to hyperspectral data [23], the extended extinction profiles (EEP) were proposed by applying EPs on the most informative

features [23] that were generated by dimension reduction approaches, such as the principal component analysis (PCA) or independent component analysis (ICA). In addition, to extract complementary features and boost the performance of the EPs, one can produce multi-EP (MEP) by stacking different EPs (e.g., area, height, volume, diagonal of the bounding box, and standard deviation) and the raw image together. Moreover, an extended multi-EP (EMEP) can be further derived, Figure 3 shows the details of EMEP formulation. The definition of EMEP is given as follows:

$$EMEP = \{MEP(\mathbf{C}_1), MEP(\mathbf{C}_2), \ldots, MEP(\mathbf{C}_m)\} \quad (4)$$

where $\mathbf{C}_k = \{\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_m\}$ stands for different independent components (ICs) provided by the ICA. In this way, the EMEP could yield more spatial features than a single EPs.

In addition, it is necessary to mention that the computational cost of EPs and EMEP are almost the same since the main time demanding part is constructing of the max-tree and min-tree, which only established once for each image [22,23].

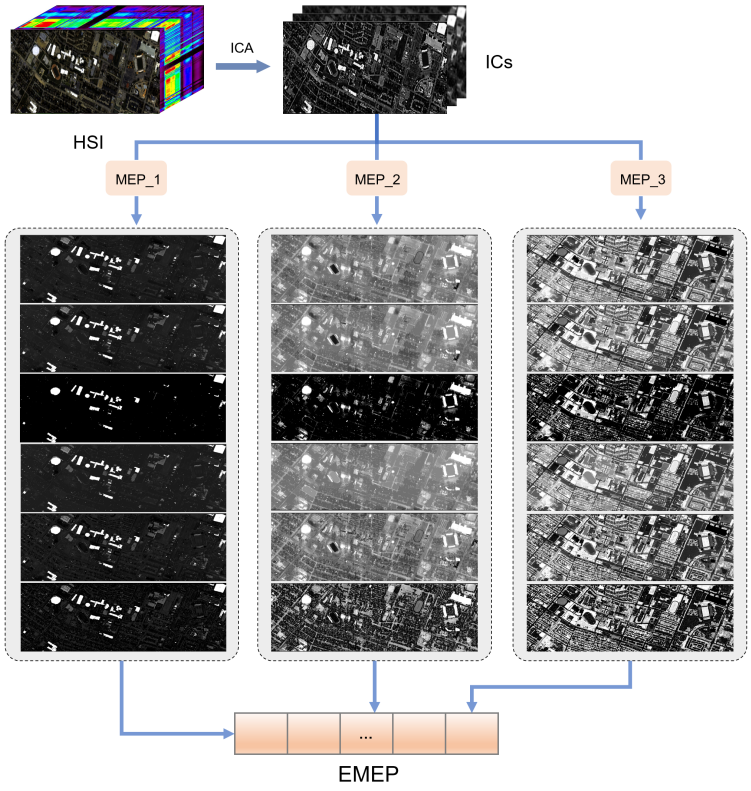

Figure 3. The construction workflow of the EMEP with respect to the Houston data set. In the first steps, ICA analysis is applied to HSI data, and three ICs are extracted. Then, each IC is used to produce MEP including five extinction value types (i.e., area, volume, standard deviation, diagonal of the bounding box and height) as well as the IC itself. At last, the EMEP is obtained by stacking these MEP together.

2.3. Convolutional Neural Networks Feature Extraction

Among all of the deep learning models, convolutional neural networks [40] have gained great research interests due to their advantages of using local connections to handle spatial dependencies and sharing weights to reduce the number of training parameters. Moreover, due to the layer-wise nature of the CNN, it makes hierarchical feature extraction doable.

In general, CNN architectures contain three parts: convolutional layers, pooling layers,

and nonlinear transformations [41,42], which are shown in Figure 4. The convolutional layer acts as the most important part of the CNN architecture, which is defined as follows:

$$\mathbf{x}_i^l = f\left(\sum_{j=1}^{P} \mathbf{x}_j^{l-1} * k_{ij}^l + b_i^l\right), \tag{5}$$

where P indexes the feature map numbers, $\mathbf{x}_j^{l-1}$ is the jth feature map of the $(l-1)$th layer, and $\mathbf{x}_i^l$ is the ith feature map of the current (l)th layer. k_{ij}^l refers to the weight of (l)th layer, which connects the ith and jth feature maps. b_i^l stands for the bias of the jth feature map in the ith layer. The function f is a nonlinear activation function, and $*$ is the convolution operation.

A pooling layer can be added after the convolution layers in order to cluster spatially discriminating signatures of the singal [43]. By pooling over a small window into a single value, CNN could additionally extract invariant features as well as reduce the size of the feature maps. The neuron in the pooling layer combines a small $n \times n$ patch of the corresponding convolution layer. In this paper, the following max pooling [44] is employed:

$$\mathbf{x}_{max} = \max_{n \times n}\left(\mathbf{x}_i^{n \times n} u\left(n, n\right)\right), \tag{6}$$

where $u\left(n, n\right)$ is a window function to the patch of the convolution layer, and $\mathbf{x}_{max}$ is the maximum in the neighborhood.

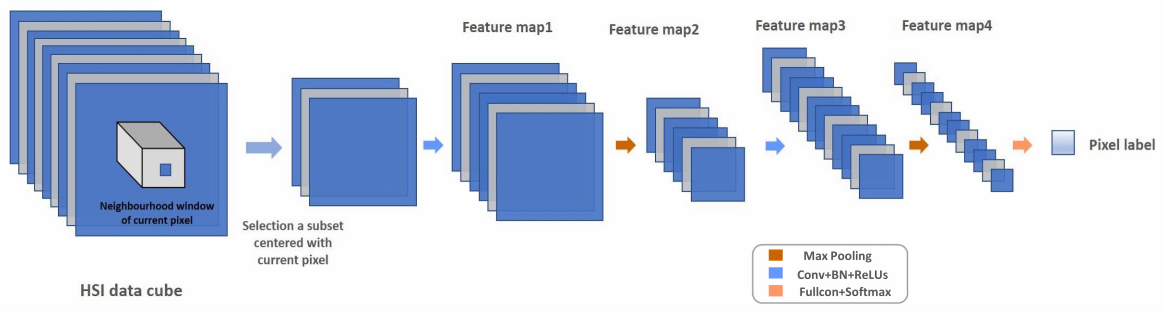

Figure 4. An example of CNN feature extraction architecture for HSI data, which has three convolution layers and two pooling layers. At last, the fully connected layer is shown as a part of the softmax layer.

Commonly, there would be a fully connected layer before the last softmax layer, where the input feature maps would be further reshaped into a vector feature. Let us assume the input training patch to be a $c \times c \times b$ dimensional subset notated by $\mathbf{t}_n$ $(n = 1, 2, 3 \ldots N)$ and the corresponding training label to be $\mathbf{y}_n$, where N refers to the total amount of training samples. Based on the above theory, the input subset $\mathbf{t}_n$ would be reduced to a vector feature $\mathbf{v}_n$ and fed into a softmax classifier as an input. As an output, the softmax layer leads to multi-classes label possibilities of each input training patch, which is shown as follows:

$$q_n^m = \frac{e^{v_n^m}}{\sum_{m=1}^{M} e^{v_n^m}}, \tag{7}$$

where M is the number of classes, v_n^m is the mth value of vector $\mathbf{v}_n$ and q_n^m refers to the possibility of $\mathbf{v}_n$ belonging to mth classes. Moreover, a predicted label would be determined by the maximal possibilities as well as the loss function L as follows:

$$L = (-1) \sum_{m=1}^{M} \sum_{n=1}^{N} y_n^m \log q_n^m, \tag{8}$$

During the back-propagation of CNN, the weights k_{ij}^{l} and biases b_{i}^{l} would be optimized with a stochastic gradient descent algorithm in order to achieve a minimized loss. With the convergence of the loss function, these parameters would eventually be fixed.

Once the CNN training is achieved, the trained networks could further perform as feature extractors by cutting off the coefficients of the last softmax layer [30]. Firstly, all spectral, spatial and elevation features are feedforwad into the corresponding feature extraction networks with a pixel-wise sliding window method, where input patches are sampled with a fixed window size centering at every pixel. Then, the feature maps produced by the third convolutional layers are extracted as deep feature outputs, which are shown as *Feature map-3* in Figure 5. To be more specific, the softmax layer plays different roles during the CNN training and classification. In the feature extraction, so to say, training step, the softmax is taken into account to adjust the parameters in the back-propagation algorithm. After the training, once all the parameters are fixed, the softmax is used as a multi-class classifier, which can be replaced by any other classifiers. Here, MCK classifiers (SVM and ELM) are implemented to replace the softmax layer for the final classification step, which simultaneously achieve multi-sensor data fusion and classification.

It is believed that the HSI imaging process is inherently nonlinear [26], so the success of CNN feature extraction mostly relies on the fact that networks progressively learn invariant information, and allows us to extract high-level abstract features, which are more reliable due to their independence from the most local details of input data. Furthermore, the consideration of spatial neighborhood information during CNN feature extraction could lead to higher capability in handling image noise problems such as the salt and pepper noise.

2.4. Data Fusion Using Multisensor Composite Kernels

The concept of the kernel method is using a nonlinear mapping function $\Phi(\cdot)$ to transfer the input data $\mathbf{x}_i$ from the original feature space $\mathcal{H}$ into a higher Hilbert kernel feature space $\mathcal{H}^{'}$, while the nonlinear problem of feature space $\mathcal{H}$ could be transferred into a linear problem of $\mathcal{H}^{'}$. This theoretical elegance of the kernel trick makes it an effective tool for HSI analysis due to its insensitivity to the Hughes phenomenon [45].

With respect to the multi-sensor data fusion, heterogeneous features obtained by different sensors would mostly have different scales, attributes, dimension channels and statistical significances [46], while this fact also leads to the use of the multi-sensor composite kernels scheme that could treat heterogeneous data sets separately. Based on the kernel concept, composite kernels (CK) can be regarded as a multiple kernel learning (MKL) method, where the multi-sensor data could be implicitly fused in a high-dimensional feature space. It is believed that learning from multiple kernels could provide better similarity generating performance. For instance, by involving multi-scale RBF kernels with different scale parameters σ, the best kernel with an optimal discriminating capability would be derived [47]. In addition, it is also possible to embed the MCK method with machine learning classifiers (such as SVM) in order to simultaneously optimize both different kernels parameter like: σ and SVM's hyperplane parameter C during the training step [19,24].

In terms of heterogeneous features, like spectral, spatial and elevation features, they may have different contributions in the classification task. So, coupling different multi-sensor features to construct multi-scale composite kernels can help to refine these contributions and promote the adoption of complementary information from these heterogeneous features. It is necessary to clarify that CK mainly contains three types: simple summation kernels, weighted summation kernels, and cross-information kernels. Among three different types, simple summation kernels combining heterogeneous features naturally come from the concatenation of individual transformations of multi-sensor information. Although weighted summation kernels introduce a trade-off between heterogeneous features aiming at better discriminating capability, an additional *prior* knowledge might be required, which remains unknown in most case [48]. In this context, to balance the effect of different features, simple summation composite kernels are implemented in the proposed framework.

Let $\mathbf{x}_i^w$, $\mathbf{x}_i^s$ and $\mathbf{x}_i^e$ be the output of CNN spectral, spatial, and elevation features in their original feature spaces $\mathcal{H}$, respectively, which correspond with three nonlinear feature mapping functions $\Phi_1(\cdot)$, $\Phi_2(\cdot)$, and $\Phi_3(\cdot)$ into Hilbert space $\mathcal{H}_1'$, $\mathcal{H}_2'$, and $\mathcal{H}_3'$. Then, the following transformation is generated:

$$\Phi(\mathbf{x}_i) = \{\Phi_1(\mathbf{x}_i^w), \Phi_2(\mathbf{x}_i^s), \Phi_3(\mathbf{x}_i^e)\}, \tag{9}$$

A three-stream composite kernels could be then calculated with the dot product of $\Phi(\mathbf{x}_i)$:

$$\begin{aligned} K(\mathbf{x}_i, \mathbf{x}_j) &= \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle \\ &= K_w\left(\mathbf{x}_i^w, \mathbf{x}_j^w\right) + K_s\left(\mathbf{x}_i^s, \mathbf{x}_j^s\right) \\ &+ K_e\left(\mathbf{x}_i^e, \mathbf{x}_j^e\right), \end{aligned} \tag{10}$$

Here, the CK is formulated as the sum of positive definite matrices [49] independent from the spectral, spatial and elevation components. As for the dimensionality, $dim\left(\mathbf{x}_i^w\right) = N_w$, $dim\left(\mathbf{x}_i^s\right) = N_s$, and $dim\left(\mathbf{x}_i^e\right) = N_e$, $dim(K) = dim(K_w) = dim(K_s) = dim(K_e)$. Importantly, solving optimization problems in composite kernels requires the same number of constraints as in conventional SVM and ELM algorithms. Therefore, no additional computational burden is introduced during the classification process. Moreover, the nonlinear feature mapping functions in SVM and ELM could have different constructions. For more details, please refer to [48,50].

For convenience reason, the following denotations are used in the experiment section: HSI shows the classification accuracy of the hyperspectral data. $\mathrm{EP_{HSI}}$ and $\mathrm{EP_{LiDAR}}$ show the classification accuracy of EMEP and EPs applied to hyperspectral images and LiDAR-derived images, respectively. $\mathrm{EP_{HSI}} + \mathrm{EP_{LiDAR}}$ denotes the classification accuracy of the stack of $\mathrm{EP_{HSI}}$ and $\mathrm{EP_{LiDAR}}$.

3. Experiment

3.1. Data Descriptions

To evaluate the performance of our proposed fusion framework, two data sets containing both hyperspectral and LiDAR data were carefully investigated in this paper.

The first data set is from an urban area of Houston, USA, which was originally distributed for the 2013 GRSS Data Fusion Contest [51]. The image size of the HSI and LiDAR-derived data is 349 × 1905 with a 2.5 m spatial resolution. The HSI data has in total 144 spectral bands, which ranges from 0.38 to 1.05 µm. Here, the HSI data is cloud-shadow removed (The enhanced data set was provided by Prof. N. Yokoya from Technical University of Munich). Figure 5 shows a false color HSI together with the corresponding training and test samples. The number of training and test samples are shown in Table 1.

The second data set is from a rural area of Trento, Italy. The area of interest is in the southern area of the city. Trento data set is composed of an HSI and LiDAR-derived DSM as well, which is the same as in the Houston data set. The image size is 166 × 600 pixels. The HSI data were obtained by the AISA Eagle sensor, and the LiDAR DSM data were captured by the Optech ALTM 3100EA sensor, with a spatial resolution of 1 m. The HSI consists of 63 bands ranging from 0.40 to 0.98 µm and a spectral resolution of 0.09 µm. Figure 6 demonstrates a false color HSI together with the corresponding training and test samples, and the number of training and test samples is shown in Table 2.

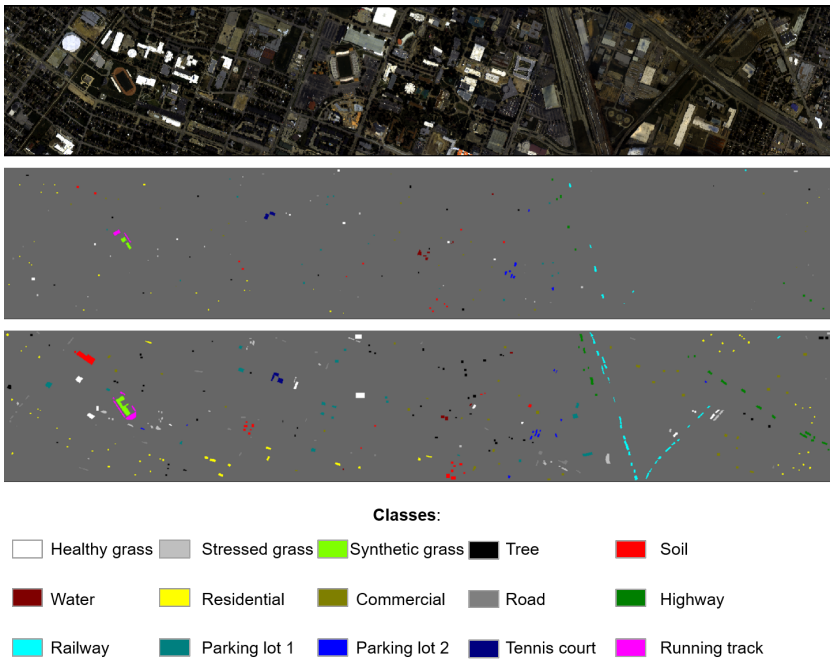

Figure 5. Houston data set, from top to bottom: false color image, training samples image, and test samples image.

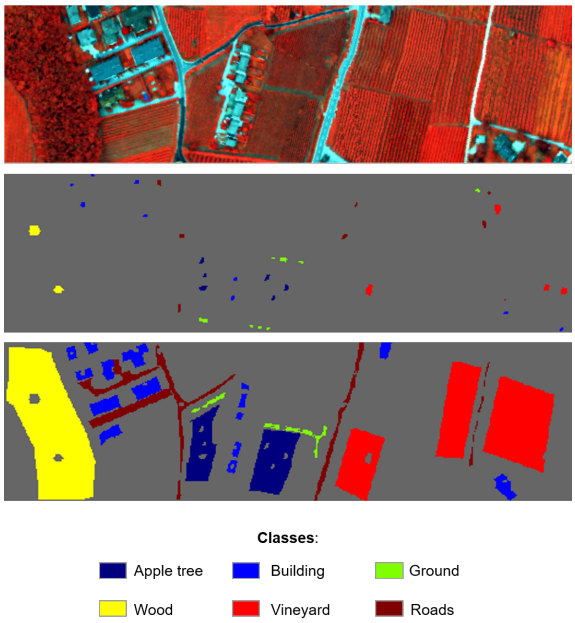

Figure 6. Trento data set, from top to bottom: false color image, training samples image, and test samples image.

3.2. Algorithm Setup

With respect to the EPs of hyperspectral and LiDAR-derived data used in this work, it is only necessary to set up the number of desired extrema, then the generation of profiles will be fully automatic. Since EPs are naturally data independent, this allows us to use a uniform parameter setting for co-registered multi-sensor data such as HSI and LiDAR-derived data. Here, the setup followed the same as suggested in [23]. For EMEP of HSI, we considered first three informative components of ICA with EPs for area, volume, height, standard deviation and diagonal of the bounding box while the values of n are automatically denoted by $\lceil 3^j \rceil$, where $j = 0, 1, \ldots, s-1$, with s equals to seven in our case. All the EPs were produced using the four-connected connectivity rule.

As for CNN feature extraction, the CNN architecture shown in Figure 7 was taken into account. three similar architecture networks have been trained for each individual data set, namely spectral net, spatial net, and elevation net. The size of the sampling neighborhood window was set to 28×28, the pooling window to 2×2, and all data were linearly mapped into $[-0.5, 0.5]$. In consideration of getting sampling image patches, all images input to CNN were padded with a mirror edges with 14 pixels. By classifying every pixel with a sliding window method, all images kept constant sizes during CNN feature extraction. Having the limited number of training samples and the small input patches size, the CNN was restrained to only three convolution layers, and pooling layers were followed as well. For training CNN from scratch, the mini-batch size was set to 100, and the training epoch number was 120. To build a robust network, sparse-based regularization techniques, like *batch normalization* (BN) and *rectified linear unit* (ReLU), were considered as well [52,53].

INPUT	[28*28*P]
CONV-1	Kernel size:5*5*P kernel:32 weights:(5*5*P)*32+32
Feature map-1	24*24*32
POOL-1	12*12*32
CONV-2	Kernel size:5*5*32 kernel:64 weights:(5*5*32)*64+64
Feature map-2	8*8*64
POOL-2	4*4*64
CONV-3	Kernel size:4*4*64 kernel:128 weights:(4*4*64)*128+128
Feature map-3	1*1*128
Full connections	Kernel size:1*1*128 weights:(1*1*128)*M+M
Feature map-3	1*1*M
Softmaxloss	
Output label	Probability vector:1*M

Figure 7. Detailed information on CNN designs, which lists the feature map size of each layer. Kernels size, number, and weights are also explained in detail. P is the input feature map number and M is the number of classes.

For SVM and ELM classifiers, SVM with an RBF kernel was used in this work, and the optimal hyperplane parameters C and σ were tuned in the range of $C = -4, -3, -2, \ldots, 16$ and $\sigma = -12, -11, \ldots, 3$ using the fivefold cross-validation algorithm. Then, ELM with the sigmoid activation function was adopted, and the hidden layer parameters a_i and b_i were randomly generated based on uniform distribution from the range $[-1, 1]$. Meanwhile, the number of hidden nodes L was set to 1000, as suggested in [36].

4. Discussion

4.1. Classification Results

Tables 1 and 2 list the overall accuracy (OA), average accuracy (AA), and Kappa coefficient (K) for the Houston and Trento data sets, which indicated the performance of our proposed framework.

As shown in Table 1, morphological EP_{LiDAR} can significantly outperform in classification accuracy compared with the results which the classifiers SVM and ELM are applied directly to LiDAR, which improves OA by 28.96% and 36.26% for SVM and ELM, respectively. Otherwise, HSI achieves less improvement compared to EP_{HSI}. The possible reasons could be that the EMEP used for EP_{HSI} only consider the first three independent components, which cannot fully consider the rich spectral information consisted in HSI. Due to this fact, HSI could better classify several classes (e.g., Grass Healthy, Grass Stressed, and Tree) than EP_{HSI}. Moreover, it can be seen that the joint use of spectral, spatial, and elevation information derived from HSI and LiDAR data outperforms the individual use of each single data source. For instance, $EP_{HSI} + EP_{LiDAR}$ improves EP_{LiDAR} by 10.99% and 14.52% using SVM and ELM, respectively. Similarly, $EP_{HSI} + EP_{LiDAR}$ slightly outperforms EP_{HSI} in OA, AA, and K, which further confirms the superior classification performance by considering the complementary information from HSI and LiDAR. The proposed framework obtains the best classification performances in Houston data set, which achieves to the best OA over 92%. In case of

ELM, our proposed framework significant improves $EP_{HSI} + EP_{LiDAR}$ by 8.34% in OA and 5.95% in AA. Due to the fact that most morphological-level features (such as, EMEP and EPs) are naturally redundant and require further invariant feature extraction, so this improvement could be attributed to the advantages of simultaneously making use of low-level morphological features and high-level deep features. In addition, experiments were also conducted based on the deep fusion method proposed in [10]. In this case, instead of applying the implicit MCK fusion scheme in kernel space, deep features were explicitly concatenated and then fed into a Softmax classifier. The proposed framework improves the deep fusion method by 1.97% in terms of OA, which confirms that the consideration of MCK fusion scheme could outperform the feature stacking strategy. By including all three streams information into one MCK fusion framework, the proposed framework reports the best classification result in 8 classes for Grass Stressed, Soil, Residential, Commercial, Road, Parking Lot 1, Parking Lot 2 and Tennis Court.

As shown in Table 2, the effectiveness of EPs features is further confirmed that EPs can significantly improves HSI and LiDAR in terms of OA, AA, and K for both SVM and ELM results. This fact confirms that the consideration of contextual information extracted by EPs is beneficial in both urban and rural areas. In addition, $EP_{HSI} + EP_{LiDAR}$ shows the same superior performance with respect to the individual EP_{HSI} and EP_{LiDAR}. Due to the fact that rural areas as Trento data set are of less complex land cover classes structures than Houston data set, the fused profiles $EP_{HSI} + EP_{LiDAR}$ could be already satisfying enough to derive classification results with OA over 97%, while the proposed framework could further promote the classification performance in both OA, AA, and K. Moreover, compared to the deep fusion method [10], the proposed framework reports better classification accuracy in 5 classes for Apple trees, Buildings, Ground, Wood, and Vineyard. In this context, the effectiveness of our proposed framework has been fully demonstrated in both urban and rural scenes.

To sum up, the classification results from Houston and Trento data sets show that the proposed framework outperforms either the individual data source or the fused EPs profiles in terms of classification accuracy, and also confirm the robustness of the proposed three-stream CNN feature extraction based on EPs features in both urban and rural areas. Moreover, the consideration of MCK fusion scheme is proven to be more effective than common feature stacking method, which achieves multi-sensor data fusion by establishing and training the corresponding composite kernels.

Figures 8–11 give a demonstration of the selected classification maps that are reported in Tables 1 and 2 in the following order: Flase color image together with ground truth mask, (a) is the SVM output on HSI data, (b) is the SVM output on the stack of $EP_{HSI} + EP_{LiDAR}$ data, (c) is the proposed fusion framework output using an SVM classifier, (d) is the ELM output on HSI data, (e) is the ELM output on the stack of $EP_{HSI} + EP_{LiDAR}$ data, and (f) is the proposed fusion framework output using an ELM classifier. In case of Figure 8, for the Houston data set, it is obvious to see that the joint use of hyperspectral and LiDAR data can extract more land cover details about inter-classes areas. For instance, the left part of Houston has been misclassified to Residential by the individual use of HSI, while the fused $EP_{HSI} + EP_{LiDAR}$ can better distinguish between Residential and Commercial. In addition, the adoption of EPs can reduce the salt and pepper noise and homogenize the classification map compared to HSI with a certain degree. As can be seen, the proposed framework can further mitigate this salt and pepper phenomena and provide the most homogeneous classification maps. This improvement can be attributed to the design of the three-stream CNN feature extractors, which considers the spatial neighboring information by involving a series of convolution and pooling operations during networks training steps. In the meantime, it is evident that the proposed framework can extract more precise shapes of different class objects, which is of great importance in accurate classification, especially in complex urban scenes.

Table 1. Classification results of Houston data set using SVM and ELM. The best result is shown in bold.

		HSI		LiDAR		EP_{HSI}		EP_{LiDAR}		$EP_{HSI} + EP_{LiDAR}$		Proposed Framework		Deep Fusion [10]
Classes	**Train/Test**	**SVM**	**ELM**	**SVM**	**ELM**	**SVM**	**ELM**	**SVM**	**ELM**	**SVM**	**ELM**	**SVM**	**ELM**	**Softmax**
Grass Healthy	198/1053	84.14	**90.98**	38.89	6.36	76.45	72.27	58.88	62.30	78.16	80.82	76.92	76.92	81.58
Grass Stressed	190/1064	95.68	97.65	42.27	34.59	83.65	79.42	57.71	42.58	80.73	84.77	96.33	**98.40**	92.11
Grass Synthetic	192/505	100.00	**100.00**	98.61	71.09	100.00	100.00	99.01	85.54	100.00	100.00	89.50	90.50	93.07
Tree	188/1056	**98.96**	94.70	61.84	29.64	81.82	81.91	68.94	62.97	96.02	88.35	93.75	96.12	94.22
Soil	186/1056	97.82	99.62	61.84	10.32	96.02	95.36	80.87	76.42	96.40	98.67	95.64	**99.62**	98.77
Water	182/143	95.10	88.81	74.83	66.43	95.80	92.31	76.92	76.92	95.80	95.80	95.10	95.80	**97.90**
Residential	196/1072	78.82	82.92	44.31	48.51	75.75	65.11	80.41	64.83	72.48	74.44	92.26	**95.15**	89.27
Commercial	191/1053	44.54	77.02	44.54	60.87	85.66	46.44	72.65	70.56	91.45	78.16	93.35	**94.59**	91.17
Road	193/1059	74.88	65.16	55.34	40.60	65.82	68.18	57.41	57.79	68.78	69.08	87.25	**93.20**	87.63
Highway	191/1036	79.83	77.61	9.07	36.39	75.77	66.22	67.95	64.00	73.65	74.13	85.04	88.42	**90.06**
Railway	181/1054	89.28	72.87	32.26	0	87.86	95.16	**99.72**	99.53	87.86	92.12	95.35	93.83	96.49
Parking Lot 1	192/1041	63.11	41.02	33.62	0	79.54	75.31	74.26	77.23	84.05	82.90	92.03	**92.41**	84.63
Parking Lot 2	184/285	75.79	68.42	34.04	39.30	81.40	75.09	55.09	65.61	77.89	78.95	79.65	**84.56**	81.75
Tennis Court	181/247	100.00	99.19	26.32	95.95	100.00	100.00	100.00	100.00	100.00	100.00	100.00	**100.00**	100.00
Running Track	187/473	98.10	98.73	32.56	95.56	**99.37**	98.73	73.78	86.89	99.15	98.73	77.59	87.74	86.26
OA(%)		82.66	81.80	44.80	33.45	82.77	77.27	73.76	69.71	84.75	84.23	90.22	**92.57**	90.60
AA(%)		79.75	77.06	43.52	39.73	80.19	80.77	70.56	72.68	81.47	86.53	84.36	**92.48**	85.31
K		0.8123	0.7939	0.4019	0.3826	0.8133	0.7939	0.7156	0.7073	0.8347	0.8557	0.8938	**0.9193**	0.8981

Table 2. Classification results of Trento data set using SVM and ELM. The best result is shown in bold.

		HSI		LiDAR		EP_{HSI}		EP_{LiDAR}		$EP_{HSI} + EP_{LiDAR}$		Proposed Framework		Deep Fusion [10]
Classes	**Train/Test**	**SVM**	**ELM**	**SVM**	**ELM**	**SVM**	**ELM**	**SVM**	**ELM**	**SVM**	**ELM**	**SVM**	**ELM**	**Softmax**
Apple trees	129/3905	91.52	89.55	41.25	0	**100.00**	100.00	99.05	95.29	98.80	99.80	99.59	94.11	99.28
Buildings	125/2778	84.67	78.04	84.88	65.08	98.78	99.42	96.22	73.72	99.32	99.60	99.89	**100.00**	90.03
Ground	105/374	96.52	94.12	37.70	0	**96.52**	92.25	58.29	61.50	32.62	69.79	71.93	73.80	54.01
Wood	154/8969	96.43	86.41	93.77	87.85	100.00	100.00	97.38	95.55	99.87	100.00	99.97	**100.00**	99.72
Vineyard	184/10317	78.32	56.50	74.07	97.96	99.64	99.12	61.11	67.12	99.88	95.95	99.62	**99.93**	99.34
Roads	122/3252	67.92	74.02	66.48	0	70.97	74.51	70.02	69.79	84.93	88.96	90.83	85.29	**97.80**
OA(%)		85.35	74.35	75.49	67.34	96.70	96.89	81.43	80.37	97.27	97.18	**97.91**	97.33	97.83
AA(%)		73.63	79.77	56.88	35.84	80.84	**94.22**	68.87	67.16	73.63	92.66	79.47	92.19	77.17
K		0.8067	0.7573	0.6740	0.3018	0.9558	0.9306	0.7646	0.7259	0.9634	0.9119	**0.9729**	0.9063	0.9710

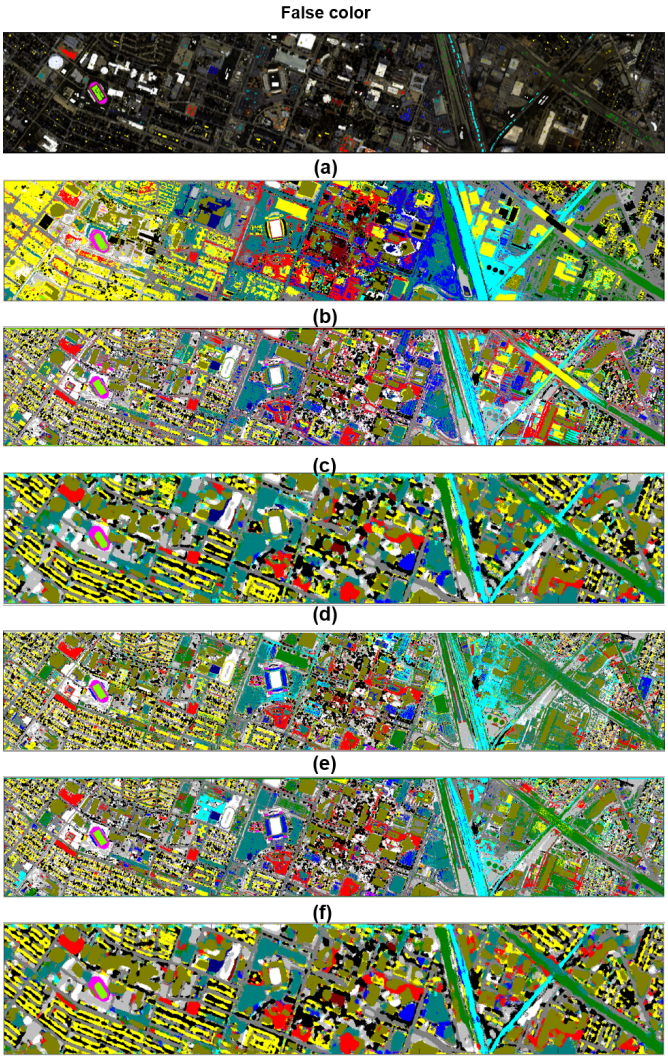

Figure 8. Houston: Classification maps, (**a**–**c**) SVM result on: HSI, $EP_{HSI} + EP_{LiDAR}$, and Proposed fusion framework; (**d**–**f**) ELM result on: HSI, $EP_{HSI} + EP_{LiDAR}$, and Proposed fusion framework.

To have a more detailed visual comparison in the Houston area, the classification maps of two sub-areas are shown in Figure 9. The undesirable misclassification can be seen when adopting HSI only, where two sub-areas are misclassified to either Residential or Railway, which results in poor classification accuracy. By taking advantage of the complementary information of HSI and LiDAR, as well as the discriminant capability of morphological EPs features, $EP_{HSI} + EP_{LiDAR}$ can better differentiate different objects and yield more detailed spatial pattern. The reduction of salt and pepper noise is till not satisfying enough, especially in the transitional region of different classes. In this context, our proposed fusion framework can further outperform the fused EPs features due to the consideration of CNN feature extraction based on morphological EPs, which makes use of the spatial neighboring information by adopting convolution and pooling layers. To this end, such consideration could further lead to superior performances in both the reduction of noise and the preservation of spatial patterns. The above-mentioned fact confirms that the attempt of combining low-level features and high-level features in the proposed framework is more effective when considering better noise reduction and higher classification accuracies.

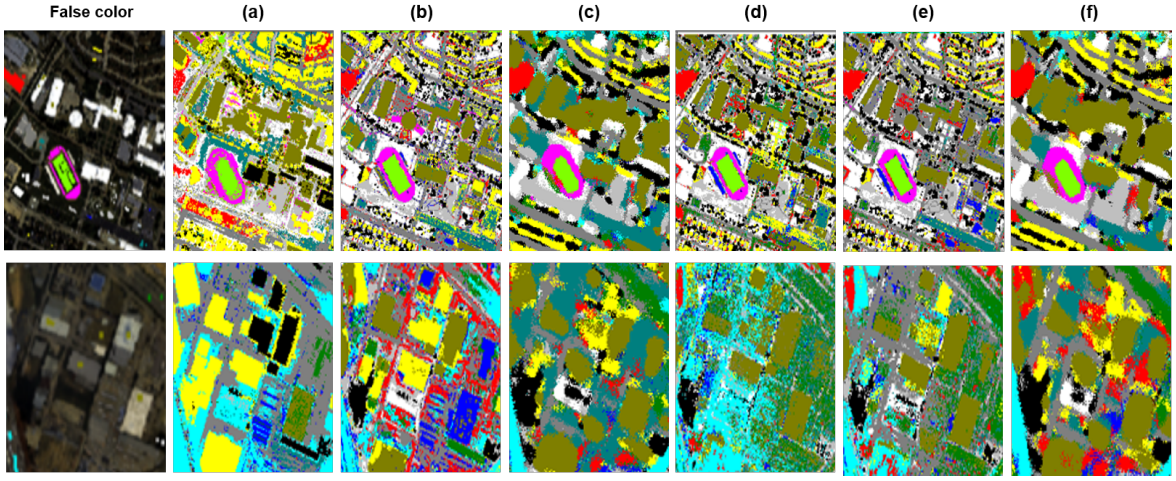

Figure 9. Houston: Close classification maps of two sub-areas, (**a**–**c**) SVM result on: HSI, $EP_{HSI} + EP_{LiDAR}$, and Proposed fusion framework; (**d**–**f**) ELM result on: HSI, $EP_{HSI} + EP_{LiDAR}$, and Proposed fusion framework.

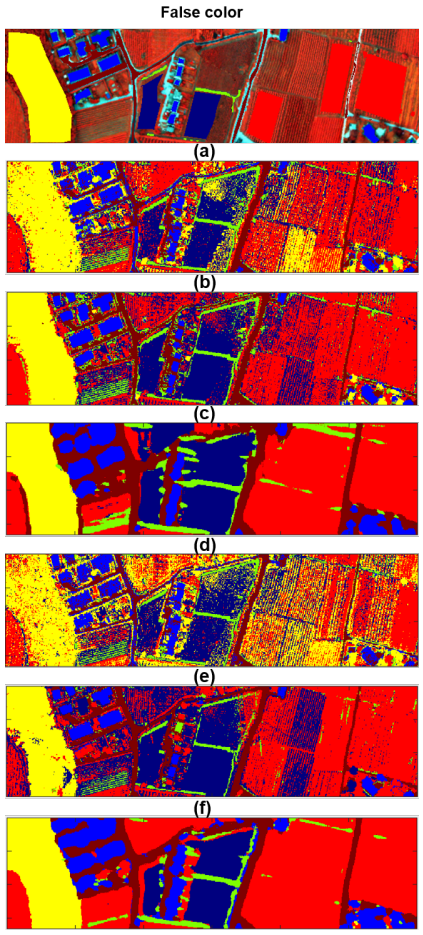

Figure 10. Trento: Classification maps, (**a**–**c**) SVM result on: HSI, $EP_{HSI} + EP_{LiDAR}$, and Proposed fusion framework; (**d**–**f**) ELM result on: HSI, $EP_{HSI} + EP_{LiDAR}$, and Proposed fusion framework.

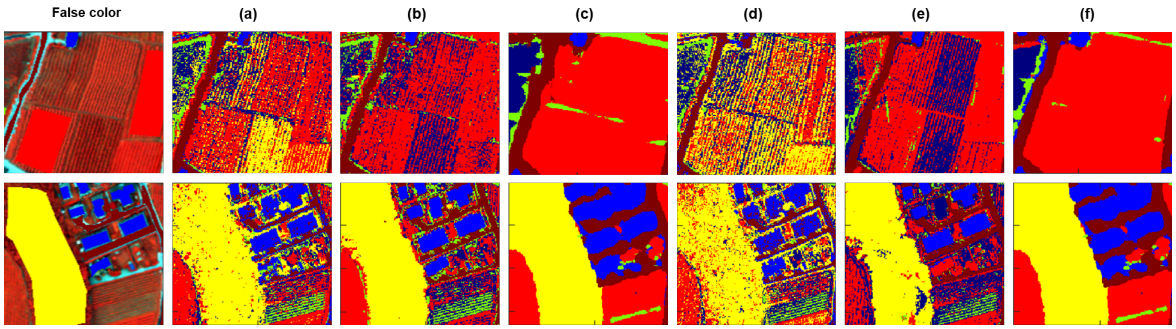

Figure 11. Trento: Close classification maps of two sub-areas, (**a**–**c**) SVM result on: HSI, $EP_{HSI} + EP_{LiDAR}$, and Proposed fusion framework; (**d**–**f**) ELM result on: HSI, $EP_{HSI} + EP_{LiDAR}$, and Proposed fusion framework.

Figure 10 illustrates the classification maps obtained by different methods on the Trento data set. From the visual comparison, one can obtain the same conclusion as for the Houston data set that the proposed framework significantly reduces the salt and pepper noise, and leads to better classification accuracy at the meantime. As can be seen in the first sub-area of Figure 11, either HSI or the fused $EP_{HSI} + EP_{LiDAR}$ classifies the large Vineyard area into Building by mistake, while the proposed framework shows a significant improvement in obtaining more homogeneous and accurate classification results. However, the proposed framework slightly outperforms the fused $EP_{HSI} + EP_{LiDAR}$ in terms of the classification accuracy (see Table 2). This outperformance in noise reduction confirms the advantage of our proposed framework in producing more homogeneous classification maps without degrading the classification accuracy. By taking this advantage, the proposed framework can also help to save the post-processing effort.

4.2. Comparison to State-of-the-Art

With respect to the widely used Houston data set, which focuses on more challenging complex urban areas, Table 3 compares the classification accuracy of the proposed fusion framework with the state-of-art methods introduced in [8–12]. Since in all those works, exactly the same sets of training and test samples were used, therefore the obtained results are fully comparable. From Table 3, our proposed fusion framework improves the graph-based feature fusion (GBFF [6]) method based on EPs and CNN in [8], the EPs fusion method introduced in [12], the deep two-stream fusion method introduced in [10], the sparse and low-rank method introduced in [9], and the low-rank and total variation method introduced in [11] in terms of OA by 1.55%, 1.25%, 2.64%, 1.27%, and 0.12%, respectively. Our proposed fusion framework also achieves the highest Kappa coefficient and the second highest AA among these state-of-art methods. There are two main reasons for this superior performance: First, this improvement is due to the benefits of invariant features learned by combining high-level CNN feature extraction and low-level morphological EPs, which are of great importance for the following fusion and classification steps. Next, compared to different feature fusion strategies introduced in the aforementioned papers [8–12], the proposed MCK fusion scheme takes advantage of multiple kernel learning methods and allows us to integrate multi-sensor data in a more robust and effective way, which eventually leads to further accuracy improvement.

Table 3. Classification results of Houston data set using standard training and test samples. The best result is shown in bold.

	Method in [8]		Method in [12]	Method in [9]	Method in [10]	Method in [11]	Proposed
	HSI + LiDAR	GBFF	EPs Fusion	SLRCA	Deep Fusion	OTVCA	
OA(%)	83.33	91.02	89.93	91.30	91.32	92.45	**92.57**
AA(%)	82.21	91.82	91.02	91.95	91.96	**92.69**	92.48
K	0.8188	0.9033	0.8910	0.9056	0.9057	0.9181	**0.9193**

5. Conclusions

In this paper, a novel framework is proposed for the fusion of multi-sensor HSI and LiDAR-derived data based on convolution neural networks and composite kernels. Using the extinction profiles of HSI and LiDAR data, as well as the original HSI data, a three-stream CNN feature extraction approach is carefully designed to extract abstract and invariant features from different spectral, spatial, and elevation data, respectively. Then, the MCK fusion scheme is implemented to fuse the outputs of CNN feature extraction to produce the final classification results. Results from two data sets confirm the effectiveness and robustness of our fusion framework in both urban and rural areas. The proposed framework reports the highest OA of 92.57% and 97.91% and Kappa coefficient of 0.9193% and 0.9729% for Houston and Trento data sets, individually. Especially for Houston data set, the proposed fusion framework reports competitive classification performance compared to state-of-the-art, which leads to general OA improvements of around 2%. Moreover, the proposed fusion framework achieves a significant improvement in terms of noise reduction for both data sets by producing homogeneous classification maps. Although simple summation of the composite kernels are implemented for the purposed of classifying multi-sensor data, the proposed method achieves superior classification performances on two widely used data sets. This encourages researchers to consider more sophisticated approaches such as the generalized composite kernels in [19] to further improve the classification accuracy in future works. Moreover, the proposed fusion framework can be regarded as a general data fusion framework, which can be easily applied to other data sets containing both hyperspectral and LiDAR data. Last but not least, the attempt of combining low-level hand-crafted features (obtained by EPs) with high-level deep features (obtained by the proposed CNN FE) shows great potentials for HSI feature extraction and noise reduction.

Author Contributions: H.L. and P.G. conceived and designed the experiments. H.L. performed the experiments. H.L. and P.G. wrote the paper manuscript. X.X.Z. and U.S. supervised the work and provided advice during the work. All the authors revised the paper manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank L. Bruzzone of the University of Trento for providing the Trento data set. The same appreciation goes to the National Center for Airborne Laser Mapping (NCALM) at the University of Houston for providing the Houston data set and the IEEE GRSS Image Analysis and Data Fusion Technical Committee for distributing the Houston data set. Moreover, the shadow-free hyperspectral data was provided by N. Yokoya.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Benediktsson, J.; Ghamisi, P. *Spectral-Spatial Classification of Hyperspectral Remote Sensing Images*; Artech House: Norwood, MA, USA, 2015.
2. Ghamisi, P.; Yokoya, N.; Li, J.; Liao, W.; Liu, S.; Plaza, J.; Rasti, B.; Plaza, A. Advances in Hyperspectral Image and Signal Processing: A Comprehensive Overview of the State of the Art. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 37–78. [CrossRef]
3. Eitel, J.U.; Höfle, B.; Vierling, L.A.; Abellán, A.; Asner, G.P.; Deems, J.S.; Glennie, C.L.; Joerg, P.C.; LeWinter, A.L.; Magney, T.S.; et al. Beyond 3-D: The new spectrum of lidar applications for earth and ecological sciences. *Remote Sens. Environ.* **2016**, *186*, 372–392, doi:10.1016/j.rse.2016.08.018. [CrossRef]

4. Höfle, B.; Hollaus, M.; Hagenauer, J. Urban vegetation detection using radiometrically calibrated small-footprint full-waveform airborne LiDAR data. *ISPRS J. Photogramm. Remote Sens.* **2012**, *67*, 134–147. [CrossRef]
5. Gamba, P.; Acqua, F.D.; Dasarathy, B.V. Urban remote sensing using multiple data sets: Past, present, and future. *Inf. Fusion* **2005**, *6*, 319–326. [CrossRef]
6. Liao, W.; Pizurica, A.; Bellens, R.; Gautama, S.; Philips, W. Generalized graph-based fusion of hyperspectral and LiDAR data using morphological features. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 552–556. [CrossRef]
7. Luo, R.; Liao, W.; Zhang, H.; Zhang, L.; Scheunders, P.; Philips, W. Fusion of Hyperspectral and LiDAR data for Classification of Cloud-shadow Mixed Remote Sensing Scene. *IEEE J-STARS* **2017**, *10*, 53768–3781.
8. Ghamisi, P.; Höfle, B.; Zhu, X.X. Hyperspectral and LiDAR Data Fusion Using Extinction Profiles and Deep Convolutional Neural Network. *IEEE J-STARS* **2017**, *10*, 3011–3024. [CrossRef]
9. Rasti, B.; Ghamisi, P.; Plaza, J.; Plaza, A. Fusion of Hyperspectral and LiDAR Data Using Sparse and Low-Rank Component Analysis. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6354–6365. [CrossRef]
10. Chen, Y.; Li, C.; Ghamisi, P.; Jia, X.; Gu, Y. Deep Fusion of Remote Sensing Data for Accurate Classification. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 53–1257. [CrossRef]
11. Rasti, B.; Ghamisi, P.; Gloaguen, R. Hyperspectral and LiDAR Fusion Using Extinction Profiles and Total Variation Component Analysis. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3997–4007. [CrossRef]
12. Zhang, M.; Ghamisi, P.; Li, W. Classification of hyperspectral and LIDAR data using extinction profiles with feature fusion. *Remote Sens. Lett.* **2017**, *8*, 957–966. [CrossRef]
13. Ghamisi, P.; Benediktsson, J.A.; Phinn, S. Land-cover classification using both hyperspectral and LiDAR data. *Int. J. Image Data Fusion* **2015**, *6*, 189–215. [CrossRef]
14. Ghamisi, P.; Cavallaro, G.; Wu, D.; Benediktsson, J.A.; Plaza, A. Integration of LiDAR and Hyperspectral Data for Land-cover Classification: A Case Study. *arXiv* **2017**, arXiv:1707.02642.
15. Pesaresi, M.; Benediktsoon, J.A. A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Trans. Geosci. Remote Sens.* **2005**, *39*, 309–320. [CrossRef]
16. Ghamisi, P.; Mura, M.D.; Benediktsson, J.A. A Survey on Spectral Spatial Classification Techniques Based on Attribute Profiles. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2335–2353. [CrossRef]
17. Mura, M.D.; Benediktsson, J.A.; Waske, B.; Bruzzone, L. Morphological Attribute Profiles for the Analysis of Very High Resolution Images. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 3747–3762. [CrossRef]
18. Li, J.; Zhang, H.; Zhang, L. Supervised Segmentation of Very High Resolution Images by the Use of Extended Morphological Attribute Profiles and a Sparse Transform. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1409–1413.
19. Li, J.; Marpu, P.R.; Plaza, A.; Bioucas-Dias, J.M.; Benediktsson, J.A. Genralized Composite Kernel Framework For Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4816–4828. [CrossRef]
20. Mura, M.D.; Benediktsson, J.A.; Bruzzone, L. Modeling structural information for building extraction with morphological attribute filters. In *Image and Signal Processing for Remote Sensing XV*; International Society for Optics and Photonics: Bellingham, WA, USA, 2009; Volume 7477.
21. Khodadadzadeh, M.; Li, J.; Prasad, S.; Plaza, A. Fusion of Hyperspectral and LiDAR Remote Sensing Data Using Multiple Feature Learning. *IEEE J-STARS* **2015**, *8*, 2971–2983. [CrossRef]
22. Ghamisi, P.; Souza, R.; Benediktsson, J.A.; Zhu, X.X.; Rittner, L.; Lotufo, R.A. Extinction Profiles for the Classification of Remote Sensing Data. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 5631–5645. [CrossRef]
23. Ghamisi, P.; Souza, R.; Benediktsson, J.A.; Rittner, L.; Lotufo, R.; Zhu, X.X. Hyperspectral Data Classification Using Extended Extinction Profiles. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1641–1645. [CrossRef]
24. Ghamisi, P.; Höfle, B. LiDAR Data Classification Using Extinction Profiles and a Composite Kernel Support Vector Machine. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 659–663. [CrossRef]
25. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
26. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]
27. Mou, L.; Zhu, X.X.; Vakalopoulou, M.; Karantzalos, K.; Paragios, N.; Saux, B.L.; Moser, G.; Tuia, D. Multitemporal Very High Resolution from Space: Outcome of the 2016 IEEE GRSS Data Fusion Contes. *IEEE J-STARS* **2017**, *10*, 3435–3447. [CrossRef]

28. Mou, L.; Ghamisi, P.; Zhu, X.X. Deep Recurrent Neural Networks for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3639–3655. [CrossRef]
29. Lyu, H.; Lu, H.; Mou, L.; Li, W.; Wright, J.; Li, X.; Li, X.; Zhu, X.X.; Wang, J.; Yu, L.; et al. Long-Term Annual Mapping of Four Cities on Different Continents by Applying a Deep Information Learning Method to Landsat Data. *IEEE J-STARS* **2018**, *10*, 471. [CrossRef]
30. Chen, Y.; Jiang, H.; Li, C.; Jia, X.; Ghamisi, P. Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6232–6250. [CrossRef]
31. Mou, L.; Bruzzone, L.; Zhu, X.X. Learning Spectral-Spatial-Temporal Features via a Recurrent Convolutional Neural Network for Change Detection in Multispectral Imagery. *arXiv* **2018**, arXiv:1803.02642.
32. Ghamisi, P.; Chen, Y.; Zhu, X.X. A Self-Improving Convolution Neural Network for the Classification of Hyperspectral Data. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1537–1541. [CrossRef]
33. Hughes, L.H.; Schmitt, M.; Mou, L.; Wang, Y.; Zhu, X.X. Identifying Corresponding Patches in SAR and Optical Images With a Pseudo-Siamese CNN. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 784–788. [CrossRef]
34. Zhu, L.; Chen, Y.; Ghamisi, P.; Benediktsson, J.A. Generative Adversarial Networks for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5046–5063. [CrossRef]
35. Mou, L.; Ghamisi, P.; Zhu, X.X. Unsupervised Spectral-Spatial Feature Learning via Deep Residual Conv-Deconv Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 391–406. [CrossRef]
36. Zhou, Y.; Peng, J.; Chen, C.L.P. Extreme Learning Machine With Composite Kernels for Hyperspectral Image Classification. *IEEE J-STARS* **2015**, *8*, 2351–2360. [CrossRef]
37. Ghamisi, P.; Plaza, J.; Chen, Y.; Li, J.; Plaza, A.J. Advanced Spectral Classifiers for Hyperspectral Images: A review. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–32. [CrossRef]
38. Vachier, C. Extinction values: A new measurement of persistence. In Proceedings of the 1995 IEEE Workshop on Nonlinear Signal and Image Processing, Halkidiki, Greece, 20–22 June 1995; pp. 245–257.
39. Soille, P. *Morphological Image Analysis: Principles and Applications*, 2nd ed.; Springer: New York, NY, USA, 2003.
40. LeCun, Y.; Bengio, Y. *The Handbook of Brain Theory and Neural Networks*; Chapter Convolutional Networks for Images, Speech, and Time Series; MIT Press: Cambridge, MA, USA, 1998; pp. 255–258.
41. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
42. Ye, C.; Zhao, C.; Yang, Y.; Fermüller, C.; Aloimonos, Y. LightNet: A Versatile, Standalone Matlab-based Environment for Deep Learning. *arXiv* **2016**, arXiv:1605.02766.
43. Volpi, M.; Tuia, D. Dense Semantic Labeling of Subdecimeter Resolution Images With Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 881–893. [CrossRef]
44. Scherer, D.; Müller, A.; Behnke, S. Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition. In *Artificial Neural Networks–ICANN 2010*; Diamantaras, K., Duch, W., Iliadis, L.S., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 92–101.
45. Hughes, G.F. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory* **1968**, *14*, 55–63. [CrossRef]
46. Gu, Y.; Chanussot, J.; Jia, X.; Benediktsson, J.A. Multiple Kernel Learning for Hyperspectral Image Classification: A Review. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6547–6565. [CrossRef]
47. Gu, Y.; Wang, C.; You, D.; Zhang, Y.; Wang, S.; Zhang, Y. Representative Multiple Kernel Learning for Classification in Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 2852–2865. [CrossRef]
48. Camps-Valls, G.; Gomez-Chova, L.; Munoz, J.; Vila-Frances, J.; Calpe-Maravilla, J. Composite Kernels for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2005**, *3*, 93–97. [CrossRef]
49. Camps-Valls, G.; Bruzzone, L. Kernel-based methods for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 1351–1362. [CrossRef]
50. Liu, X.; Wang, L.; Huang, G.B.; Zhang, J.; Yin, J. Multiple kernel extreme learning machine. *Neurocomputing* **2015**, *149*, 253–264. [CrossRef]
51. Debes, C.; Merentitis, A.; Heremans, R.; Hahn, J.; Frangiadakis, N.; van Kasteren, T.; Liao, W.; Bellens, R.; Pizurica, A.; Gautama, S.; et al. Hyperspectral and LiDAR Data Fusion: Outcome of the 2013 GRSS Data Fusion Contest. *IEEE J-STARS* **2014**, *7*, 2405–2418. [CrossRef]

52. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML'10), Haifa, Israel, 21–24 June 2010; pp. 807–814.
53. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* **2015**, arXiv:1502.03167.