



Editorial

Special Issue Review: Artificial Intelligence and Machine Learning Applications in Remote Sensing

Ying-Nong Chen ^{1,2,*}, Kuo-Chin Fan ², Yang-Lang Chang ³ and Toshifumi Moriyama ⁴

¹ Center for Space and Remote Sensing Research, National Central University, No. 300, Jhongda Rd., Jhongli Dist., Taoyuan City 32001, Taiwan

² Department of Computer Science and Information Engineering, National Central University, No. 300, Jhongda Rd., Jhongli Dist., Taoyuan City 32001, Taiwan

³ Department of Electrical Engineering, National Taipei University of Technology, Taipei 10608, Taiwan

⁴ Graduate School of Engineering, Nagasaki University, 1-14 Bunkyo-Machi, Nagasaki 852-8521, Japan

* Correspondence: yingnong@csrsr.ncu.edu.tw

Abstract: Remote sensing is used in an increasingly wide range of applications. Models and methodologies based on artificial intelligence (AI) are commonly used to increase the performance of remote sensing technologies. Deep learning (DL) models are the most widely researched AI-based models because of their effectiveness and high performance. Therefore, we organized a Special Issue on remote sensing titled “Artificial Intelligence and Machine Learning Applications in Remote Sensing.” In this paper, we review nine articles included in this Special Issue, most of which report studies based on satellite data and DL, reflecting the most prevalent trends in remote sensing research, as well as how DL architecture and the functioning of DL models can be analyzed and explained is a hot topic in AI research. DL methods can outperform conventional machine learning methods in remote sensing; however, DL remains a black box and understanding the details of the mechanisms through which DL models make decisions is difficult. Therefore, researchers must continue to investigate how explainable DL methods for use in the field of remote sensing can be developed.

Keywords: artificial intelligence; machine learning; deep learning; remote sensing



Citation: Chen, Y.-N.; Fan, K.-C.; Chang, Y.-L.; Moriyama, T. Special Issue Review: Artificial Intelligence and Machine Learning Applications in Remote Sensing. *Remote Sens.* **2023**, *15*, 569. <https://doi.org/10.3390/rs15030569>

Received: 5 January 2023

Accepted: 17 January 2023

Published: 18 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Remote sensing technologies are used for various applications [1], such as the classification of hyperspectral images (HSIs); the detection of mangroves, land cover and usage, environmental pollution, and other target objects; the estimation of rainfall; and the prevention of disasters. Deep learning (DL) methods [2] have been applied in numerous studies on remote sensing technologies because of their exceptional performance in feature extraction and data classification and interpretation. Developing a high-performance DL model for specific applications requires considerable relevant labeled data. Because most applications of remote sensing technologies involve big data analysis, DL methods are particularly useful in remote sensing.

Researchers have made considerable progress in applying conventional machine learning (ML) methods to remote sensing [3,4]; however, these methods have several drawbacks. First, many conventional ML models are based on heuristics and parameter-tuning, which rely on the domain knowledge and personal experience of the person developing the model. Although domain knowledge and personal experience may be helpful in certain task-specific cases, this dependence limits the generalizability of ML models and makes them unsuitable for use in combination with remote sensing systems. Second, conventional ML models extract only shallow features. These shallow features are normally extracted from statistical data or simple nonlinear correlations. Models that rely on such features have only limited performance for recognition tasks, such as HSI classification

and precipitation–nonprecipitation classification, or can be used in preprocessing procedures such as dimension reduction; they can seldom be used to make great breakthroughs or extract salient deep features. The use of only shallow features makes achieving high performance in HSI classification or rainfall estimation almost impossible [5]. Third, conventional ML models cannot be trained through unsupervised learning; therefore, they exhibit limited performance in the processing of extensive labeled data. By contrast, adversarial networks [6], a class of DL networks, can be trained on unlabeled samples. In addition, with the rapid advancement of sensors, satellites, and other relevant technologies, additional requirements, including those related to online and incremental learning, must be considered. DL methods can often account for such requirements, whereas conventional ML methods cannot.

DL methods can be used to extract features and construct models simultaneously. DL networks, in contrast to conventional networks based on heuristics and parameter-tuning, can extract salient and deep features automatically. In addition, deep layers in the architecture of such networks can often extract high-level features that suitably represent complex data corresponding to various spectral bands in multispectral, hyperspectral, and ultraspectral images. Furthermore, DL methods can be used to overcome problems related to unlabeled data, and a well-trained DL model can usually be used for several task-specific applications in which few or no relevant labeled data may be available. Therefore, we organized a Special Issue on remote sensing titled “Artificial Intelligence and Machine Learning Applications in Remote Sensing” to present state-of-the-art research on the applications of artificial intelligence, ML, and DL in the processing of remote sensing data. In this paper, we review nine of the papers published in this Special Issue.

The remainder of this paper is organized as follows. We summarize the fundamental DL approaches in Section 2 and review the papers published in this Special Issue in Section 3. An overall discussion of this Special Issue is presented in Section 4, and our conclusions are presented in Section 5.

2. Deep Learning

DL is a major subfield of ML that developed out of research on conventional multilayer perceptrons (MLPs). Although conventional MLPs contain only one hidden layer, DL architectures often have numerous hidden layers, thus enabling the use of unlabeled raw data or labeled data sets in model training. As the raw or labeled data are input into the DL model and are processed in each hidden layer, the model learns increasingly deep high-level features. Eventually, the raw or labeled data are encoded as high-dimensional discriminant feature vectors.

Although numerous DL architectures have been proposed [7,8], most share basic components such as convolutional, pooling, fully connected, and softmax layers. DL models expand upon the basic structure of the MLP by incorporating numerous hidden layers for learning complicated regressions and mapping raw data to specific outputs. Because of their powerful regression capabilities, DL models contain numerous (even millions of) parameters that must be tuned. Therefore, constructing DL models requires considerable effort and resources, including time and training data sets, especially because most DL models are trained through supervised learning. Developing DL models in the early days of DL was difficult. However, two solutions have helped researchers overcome the aforementioned obstacles. First, because of recent advancements in computational hardware and software, researchers can conveniently and efficiently implement DL models. According to [9], DL platforms—including tensor processing units, graphics processing units (GPUs), and central processing units—each have distinct advantages for certain types of DL models. Second, as discussed in the previous section, most remote sensing applications involve big data analysis, and large-scale annotated remote sensing data sets can easily be used to train DL models.

Convolutional neural network (CNN) architecture has become the most typical and representative DL architecture because it has been applied in image processing and com-

puter vision. Generally, CNNs are designed and are especially suitable for processing data in multiple arrays. Figure 1 illustrates typical CNN architecture, which contains convolutional, pooling, and fully connected layers and a softmax layer. Convolutional layers are used to extract or enhance spatial features in the input data. Pooling layers reduce the spatial dimensions of the input data, thus reducing the number of trainable parameters in the subsequent layers and enabling them to concentrate on a greater range of input patterns. As the input is processed in deeper layers, the number of suitable high-level features that are extracted increases. Finally, the fully connected and softmax layers are used for data classification.

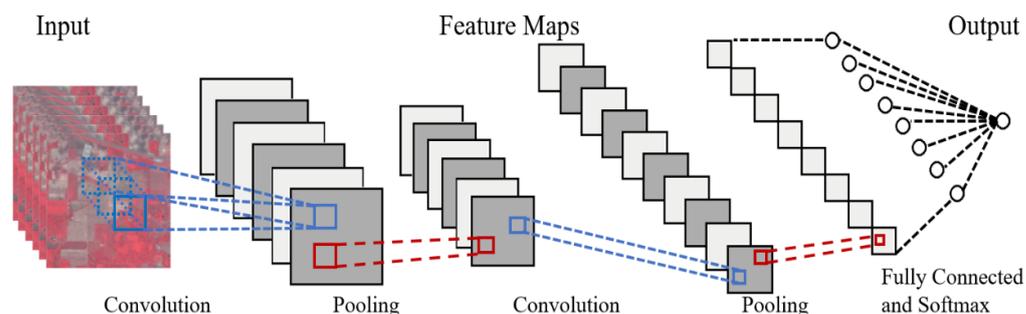


Figure 1. Typical convolutional neural network architecture, including convolutional, pooling, and fully connected layers and a softmax layer.

The effectiveness of CNNs relies on three factors: local connectivity, weight sharing, and pooling. Local connectivity enables hidden neurons in CNNs to extract spatially local input patterns. Because of local connectivity, the shallow layers of a CNN can extract low-level features, such as textures and edges, within a relatively small receptive field, whereas the deeper layers of the CNN can extract high-level semantic features within a relatively large receptive field. Weight sharing effectively reduces the number of trainable parameters by using the same filtering mask to convolve all the visual fields. In addition to reducing the number of trainable parameters, weight sharing helps achieve translation invariance; that is, the same filter can extract an important feature regardless of the feature's location.

Pooling operations are commonly used in CNNs to reduce the dimensions of input feature maps. They not only reduce the number of trainable parameters in a CNN but also enable the subsequent layers to concentrate on a greater range of input patterns. Most DL models use max pooling, which involves applying a maximum filter to each feature map to obtain the maximum value in each subfield.

3. Artificial Intelligence and ML Applications in Remote Sensing

In this section, we review nine articles published in the Special Issue on remote sensing titled "Artificial Intelligence and Machine Learning Applications in Remote Sensing." Most of the applications presented in the articles are based on satellite data.

3.1. Deep Learning Method Pre-Trained on Natural Images for Seismic Random Noise Attenuation [10]

Seismic signals collected using onshore or offshore sensors are often contaminated by random or complex noise, resulting in seismic data with a low signal-to-noise ratio (SNR). Therefore, improving the quality of seismic data by suppressing random noise is the main goal of prestack and poststack seismic data processing.

To achieve the aforementioned goal, researchers have proposed various conventional methods on the basis of their personal experience or domain knowledge. Most such methods involve discriminating signals from noise by projecting them into more-discriminating feature spaces. However, even in such feature spaces, seismic signals and noise overlap considerably, and distinguishing between the two remains difficult.

As the number and variety of successful applications of DL in remote sensing have continued to increase, researchers have developed some DL-based denoising methods. However, DL denoising models must be trained on numerous samples to prevent overfitting, and the training samples must be sufficiently representative of the entire solution space. Researchers must focus on these two factors to continue improving the generalizability of DL denoising models.

To improve the generalizability of prior DL denoising models, Zhao et al. [10] proposed a DL model in which transfer learning is used to suppress random seismic noise. Their model comprises one pretrained and one post-trained network. The pretrained network is a denoising CNN (DnCNN) with dilated convolution trained on natural images, and the post-trained network is a U-Net trained through semisupervised learning on fewer seismic images. The connection between the pretrained and post-trained networks is based on transfer learning.

The pretrained network is a DnCNN with 13 layers, 32 convolution kernels, and a kernel size of 3 that includes a rectified linear unit (ReLU) activation function after each hidden layer. To ensure high training efficiency, dilated convolution is used to increase the size of the receptive field in the first few layers without increasing the depth of the network, thus enabling the network to extract global features at the start of training.

After the pretrained network is established, transfer learning is used to transfer the parameters of the pretrained network to a new network for training. The objective of transfer learning is to improve the efficiency of the training and performance of the new network in processing closely related data or performing relevant tasks. The pretrained network is trained exclusively on natural images and does not learn detailed seismic signal information; instead, the parameters of the network pretrained on natural images are transferred to a new network, which is trained on seismic data, thus rendering repetition of the entire training process unnecessary.

The post-trained network trained on seismic data is a U-Net containing residual units and dropout layers. The input of the post-trained network is seismic data processed by the pretrained network. The post-trained network contains 13 convolutional layers and four pooling layers.

According to the experimental results, the pretrained network successfully denoised input data, as reflected in various performance metrics (peak SNR, mean standard error, and structural similarity). The post-trained network was discovered to successfully extract detailed seismic data structural features to further denoise the input data. Overall, the proposed DL model was found to achieve higher performance in suppressing seismic random noise in terms of both performance metrics and training efficiency than other state-of-the-art methods.

3.2. BO-DRNet: An Improved Deep Learning Model for Oil Spill Detection by Polarimetric Features from SAR Images [11]

International trade today is large in scale and has only increased in scale since the outbreak of the COVID-19 pandemic. Numerous large ships are frequently traveling to and from ports worldwide, and many are required to wait outside ports because of various factors, which may increase the risk of marine oil spills. Such spills can cause severe global economic and ecological problems, such as harming sea creatures, coating birds' wings and leaving them unable to fly, inducing toxicity in animals, damaging habitats, contaminating seafood and drinking water, and devastating natural resources, which can cause chain reactions. Accurately and rapidly detecting marine oil spills to prevent such disasters and chain reactions is crucial.

Drawing on recent advancements in synthetic-aperture radar (SAR) technology and DL methods, researchers have developed DL marine oil spill detection models that use polarimetric SAR (Pol-SAR) data, such as Mueller matrices, copolarized complex correlation coefficients, degrees of polarization, the total power of the SAR scattering target, geometric

intensity, conformity coefficients, mean scattering angles, and entropy values. Many such models outperform conventional ML models in detecting oil spills.

Despite this high performance, some problems—such as insufficient extracted features due to insufficient model depth, loss of target information due to small receptive field, and fixed model hyperparameters—continue to limit the performance of such DL models. To address these problems, Wang et al. [11] proposed a DL model called BO-DRNet that is based on Bayesian optimization (BO), DeepLabv3+, and ResNet-18.

BO-DRNet is a CNN based on encoder–decoder architecture. Its encoder integrates ResNet-18 and DeepLabv3+, extracting features from the input through the downsampling layers. Because ResNet-18 learns the residual function on the basis of the layer input, it can help overcome the problems of complex network convergence and gradient disappearance, and using a pretrained ResNet-18 can help extract more high-level features. Moreover, in the encoder, after ResNet-18 is used for feature extraction, the key component of DeepLabv3+, namely atrous spatial pyramid pooling (ASPP), is used to expand the receptive field without modifying the size of the feature map, which facilitates the extraction of multiscale information.

The BO-DRNet decoder is used for convolution and deconvolution. It uses a skip connection to combine the ResNet-18 and ASPP extraction features for input data recovery.

After the encoding and decoding, the hyperparameters of the model are subjected to BO. Compared with conventional hyperparameter optimization methods, which are impractical when a model has an excessive number of hyperparameters, BO is more efficient because it involves fewer steps and does not require a derivative.

Wang et al. used three data sets to construct and evaluate the effectiveness of BO-DRNet and compared the performance of BO-DRNet with that of FCN-8s, DeepLabv3 + Xception, and DeepLabv3 + ResNet-18. Although BO-DRNet was discovered to outperform the other DL models in oil spill detection in the experiments, it still failed to detect numerous oil spill pixels, which may have been because the number of samples used to train the model was insufficient. The performance of BO-DRNet can be further improved through further training using a greater number of training sets. In addition, Wang et al. used quadpolarimetric oil spill SAR images collected by RADARSAT-2 to assess the performance of BO-DRNet, and BO-DRNet achieved the highest accuracy among the tested networks, indicating that it is robust and has outstanding general adaptability.

3.3. Tree Recognition on the Plantation Using UAV Images with Ultrahigh Spatial Resolution in a Complex Environment [12]

Increasing greenhouse gas emissions pose a major threat to our world, and the construction of forest monitoring systems is crucial to addressing this threat. Forests not only serve as the largest terrestrial ecosystems but also benefit humans by providing raw materials for the manufacturing of products that play major roles in the economy and society. However, manually checking trees is laborious and costly. Therefore, unmanned aerial vehicle (UAV) images and remote sensing technologies can be used to reduce the cost associated with detecting vegetation.

Most studies on the application of UAV images have focused on plant detection and counting. Experimental results have indicated that the use of UAV images can reduce the cost of forest monitoring. Color is the most commonly used feature in the detection of individual tree crowns (ITCs). For example, the normalized excessive green index is often used to measure green vegetation in the UAV images that are employed in automatic forest monitoring, indicating that spectral information from UAV images can be used for preliminary classification of vegetation.

To facilitate the detection and counting of ITCs, Guo et al. [12] proposed a random forest (RF)-based method including two texture features. The proposed method consists of three main stages: spectral index selection, potential tree crown region (region of interest (ROI)) extraction, and ROI identification.

In the spectral index selection stage, to facilitate accurate feature extraction, UAV images are preprocessed using a suitable spectral index, such as the normalized green–blue difference index, normalized green–red difference index, green–red difference index, normalized blue–green vegetation index, normalized excessive green index, excess green index, or excess red index.

In the stage of potential tree crown region (ROI) extraction, some foreground and background features that are not identified in the first step may remain. To address this problem and more effectively separate the trees from the background, morphological image processing methods are applied.

Finally, in the ROI identification stage, some effective color and texture features—namely the mean and variance of color components, energy, contrast, correlation, entropy, homogeneity, and local binary patterns—are used to classify objects in the extracted regions.

When the three aforementioned stages are complete, an RF is used to determine whether the extracted regions are tree crown regions. To evaluate the effectiveness of their model, Guo et al. performed five-fold cross-validation with a sample of 1400 training set images.

According to the experimental results, the proposed tree recognition method, which uses a suitable spectral index to mitigate the effect of high-coverage weeds in UAV images, is effective. The method uses multiple closed morphological filters to connect scattered canopy branches and decrease the interference of canopy debris.

The proposed method uses an RF of 500 trees to classify ROIs. The average accuracy of the method in detecting canopies in the test set increased from 75.76% or 77.15% when a single texture feature was used to 92.94% when the two texture features were used, indicating that using multiple texture features is an effective approach to improving the effectiveness of methods for detecting vegetation in high-resolution UAV images. The proposed method expands the applications of UAV images in forest monitoring and may serve as an effective tool for improving the precision and efficiency of forest monitoring.

3.4. Consolidated Convolutional Neural Network for Hyperspectral Image Classification [13]

HSIs have been used in various applications in the field of remote sensing, including vegetation monitoring, area change detection, disaster prevention, land cover classification, weather forecasts, and atmospheric research. Hyperspectral sensors generate hundreds of wavelength bands that range from visible to near-infrared. These wavelength bands typically provide abundant spectral and spatial information that can be used to analyze a target area. Each pixel in an HSI typically corresponds to hundreds of reflected electromagnetic radiation bands. However, these bands usually correspond to similar nonlinear and high-dimensional features, which can make the analysis of HSIs challenging. The performance of HSI classification models is highly dependent on spatial and spectral information and is strongly affected by factors such as data redundancy and insufficient spatial resolution.

In [13], Chang et al. developed a consolidated CNN (C-CNN) for hyperspectral image classification. They compared the performance of their network in HSI classification with that of other state-of-the-art DL networks.

The proposed C-CNN comprises a three-dimensional CNN (3D-CNN) combined with a two-dimensional CNN (2D-CNN). The 3D-CNN is used to represent spatial–spectral features from the spectral bands, whereas the 2D-CNN is used to learn abstract spatial features. Principal component analysis is used to preprocess the original HSIs before they are input to the C-CNN to reduce spectral band redundancy. Chang et al. used image-augmentation techniques, including rotation and flipping, to increase the number of training samples and prevent overfitting. In addition, they used both dropout and L2 regularization to further reduce the model's complexity and prevent overfitting.

The results of experiments performed using the Indian Pines, Pavia University, and Salinas Scene benchmark hyperspectral data sets indicated that the proposed model provides the optimal trade-off between accuracy and computational time relative to related methods. Finally, by using multiple GPUs (2xNVIDIA GeForce RTX 2080 Ti) and NVIDIA

NVLink for GPU interconnection, the authors reduced the processing time for each of the three aforementioned data sets by approximately 44%.

3.5. MDPrePost-Net: A Spatial-Spectral-Temporal Fully Convolutional Network for Mapping of Mangrove Degradation Affected by Hurricane Irma 2017 Using Sentinel-2 Data [14]

Jamaluddin et al. [14] developed a new fully convolutional network (FCN) called MDPrePost-Net for detecting mangrove degradation in the southwest Florida coastal zone induced by Hurricane Irma in 2017. The network extracted spatial–spectral–temporal features from images of the study area before and after Hurricane Irma (from Sentinel-2) to identify intact and degraded mangrove areas. A total of 10 input bands were used (Blue, Green, Red, NIR, SWIR-1, SWIR-2, NDVI, CMRI, NDMI, and MMRI). The target data consisted of three classes (nonmangrove, intact mangrove, and degraded mangrove). The authors combined their manually delineated images with images classified using an RF to obtain the final target data set.

MDPrePost-Net consisted of two main submodels: a pre–post deep feature extractor and an FCN classifier. The pre–post deep feature extractor consisted of two key components: a ConvLSTM extractor, which extracted spatial–spectral–temporal correlations from images from before and after Hurricane Irma, and a post feature extractor, which extracted spatial–spectral features from images obtained after Hurricane Irma to provide detailed information about the conditions of the affected mangrove areas. The FCN classifier was constructed using U-Net architecture with a VGG-16 encoder. The intersections of union (IoUs) for the nonmangrove, intact mangrove, and degraded mangrove classes in the testing data set were 99.82%, 96.47%, and 96.82%, respectively. The authors conducted ablation studies to evaluate the effectiveness of each component of the extractor and determined that both components contributed considerably to the performance of the entire network. MDPrePost-Net outperformed other FCNs (FC-DenseNet, U-Net, Link-Net, and FPN) in terms of accuracy metrics. Ultimately, the authors determined that 26.64% (41,008.66 Ha) of the mangroves in the study area had been degraded by Hurricane Irma. The overall accuracy and kappa of the model were 0.973 and 0.960, respectively.

3.6. Change Detection from SAR Images Based on Convolutional Neural Networks Guided by Saliency Enhancement [15]

Change detection is a key task in many remote sensing applications. Unsupervised SAR image change detection involves three main steps: speckle noise reduction, difference map generation, and classification. Regarding speckle noise reduction, the use of conventional denoising methods for removing inherent speckle noise may also result in the loss of useful information. Regarding difference map generation, changed information may be lost in the process of obtaining difference images, which may affect the detection results. Overall, although supervised models can achieve higher performance than unsupervised models, the use of supervised models requires labor-intensive manual annotation, which may affect the generalizability of the model. To improve upon the performance of previous SAR image change detection methods, Li et al. [15] proposed a new automatic SAR image change detection method based on saliency detection and a convolutional wavelet neural network (CWNN).

The proposed SAR image change detection method involves three main steps: difference image generation, salient region extraction, and classification. In the first step, denoising is performed on two SAR images, and the log-ratio operator is used to obtain the initial difference image. In the second step, saliency detection theory is used to obtain visually salient regions, and the automatic threshold Otsu model is used to generate a binarized saliency map for preclassification. Finally, in the third step, a CWNN is used for classification. A wavelet transform is applied in the pooling layer to facilitate the extraction of salient features. The experimental results demonstrated that the proposed SAR image change detection method is effective.

3.7. “Combining Object-Oriented and Deep Learning Methods to Estimate Photosynthetic and Non-Photosynthetic Vegetation Cover in the Desert from Unmanned Aerial Vehicle Images with Consideration of Shadows” [16]

Soil erosion is an important issue for global environment. Recently, three state-of-the-art deep learning models such as DeepLabV3+, PSPNet, and U-Net have been often applied to extract discriminant features from unmanned aerial vehicle (UAV) images. However, the performance of these deep learning models is highly reliant on the acquisition of massive labeled data, which greatly restrict these models’ applicability. In addition, another critical factor is the numerous shadows present in the UAV images, which limit the performance of these models. It is still not clear how much accuracy can be achieved under shaded features.

Therefore, in order to explore the feasibility and efficiency of these three models for shadow-affected UAV images, He et al. [16] proposed a semantic segmentation scheme, which took the MU US Desert as a study area to classify shadow-sensitive photosynthetic vegetation (PV) and non-photosynthetic vegetation (NPV).

To evaluate these three models’ classification performance for shadow-sensitive PV and NPV, numerous label data must be acquired; therefore, an object-oriented label data creation method is proposed in this study. The proposed object-oriented label data creation method consists of three steps: (1) image segmentation, (2) segmented image classification, and (3) the manual correction of the classification result. The first step, image segmentation, is to segment UVA images into segments. Then, in the second step, some of the six classes of segments are used to train the random forest (RS) model, and the obtained RS model is used to classify the segments. In the final step, since the RS model can suffer from misclassifications and omissions, these errors can be fixed by manual visual interpretation. In other words, the proposed object-oriented label data creation is to obtain the ground truth of UVA image.

After obtaining the ground truth of UAV image from the proposed object-oriented label data creation method, the UAV image with ground truth is used to evaluate the performance of three state-of-the-art deep learning models: DeepLabV3+, PSPNet, and U-Net. The experimental results show that the DeepLabV3+ model can achieve higher accuracy than that of the U-Net and PSPNet models. Since deep learning semantic segmentation is a cost-effective scheme for land cover classification, this study provides a useful reference for land-use planning and management.

3.8. Surround-Net: A Multi-Branch Arbitrary-Oriented Detector for Remote Sensing [17]

Oriented object detection technology has facilitated major advancements in the field of remote sensing. Oriented object detectors can be categorized into five-parameter systems and eight-parameter systems, which encounter the periodic problem of angle regression and the discontinuous problem of vertex regression during training, respectively. To mitigate these problems, Luo et al. [17] developed a model called Surround-Net. The three main innovative aspects of the proposed model are as follows. First, Surround-Net uses a multibranch strategy to adaptively select the most suitable regression method; this addresses the aforementioned discontinuous problem. Second, it uses a modified Focal Loss and a new Surround IoU Loss function to resolve the inconsistency between classifications and quality estimations. Third, it uses a center vertex attention mechanism to suppress environmental noise in remote sensing images. Luo et al. performed experiments using the DOTA data set to assess the effectiveness of Surround-Net. The experimental results indicated that Surround-Net outperforms comparable detectors.

3.9. “Remote Sensing Image Target Detection: Improvement of the YOLOv3 Model with Auxiliary Networks” [18]

Target detection based on remote sensing images has both civil and military applications. However, two factors related to target detection technology must be addressed: the need for real-time detection and the accuracy of detection of very small targets. To address these two problems, Qu et al. [18] proposed a modified YOLOv3 model with an auxiliary network. The proposed model has four components: (1) an image blocking mod-

ule, which feeds fixed-size images to the network; (2) a Distance IoU loss function, which accelerates the training of the proposed model; (3) a convolutional block attention module, which connects the auxiliary network to the backbone network to facilitate retention of key information during model training; and (4) an adaptive feature fusion method, which reduces the inference overhead to increase the detection speed. The authors performed experiments using the DOTA dataset to evaluate the effectiveness of the proposed model; the results indicate that the proposed model performs highly and significantly outperforms the unmodified YOLOv3 model.

4. Discussion

Scholarly interest in the application of DL-based computer vision technologies in the field of remote sensing has been growing. However, various factors affect the performance of DL models for remote sensing, among which the need for large training data sets is the most crucial, especially for supervised DL models. The quality of remote sensing data, including the removal of noise and compensation for missing data, also strongly affects the performance of DL models. Finally, because no single DL architecture is applicable in all scenarios, researchers must further investigate how effective task-specific DL architecture can be designed.

5. Conclusions

This paper provides an overview of nine papers published in the Special Issue on remote sensing titled “Artificial Intelligence and Machine Learning Applications in Remote Sensing.” The advancements presented in these papers align with current trends in research on DL. Researchers have started to investigate how to best explain DL. Although DL methods perform highly in various remote sensing applications, the detailed mechanisms through which they make decisions remain difficult to understand. In the future, researchers must continue to investigate how explainable DL methods for use in the field of remote sensing can be developed.

Author Contributions: Conceptualization, Y.-N.C.; Methodology, Y.-N.C.; Project administration, Y.-N.C.; Resources, Y.-N.C., K.-C.F., Y.-L.C. and T.M.; Supervision, Y.-N.C.; Writing—original draft, Y.-N.C.; Writing—review and editing, Y.-N.C., K.-C.F., Y.-L.C. and T.M. All authors have read and agreed to the published version of the manuscript.

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

References

1. Li, S.; Song, W.; Fang, L.; Chen, Y.; Ghamisi, P. Deep Learning for Hyperspectral Image Classification: An Overview. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6690–6709. [CrossRef]
2. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [CrossRef] [PubMed]
3. Gu, Y.; Chanussot, J.; Jia, X.; Benediktsson, J.A.P. Multiple Kernel Learning for Hyperspectral Image Classification: A Review. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6547–6565. [CrossRef]
4. Ghamisi, P.; Mura, M.D.; Benediktsson, J.A.P. A Survey on Spectral–Spatial Classification Techniques Based on Attribute Profiles. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2335–2353. [CrossRef]
5. Min, M.; Bai, C.; Sun, F.; Liu, C.; Wang, F.; Xu, H.; Tang, S.; Li, B.; Di, D.; Dong, L.; et al. Estimating Summertime Precipitation from Himawari-8 and Global Forecast System Based on Machine Learning. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 2557–2570. [CrossRef]
6. Goodfellow, I.J.; Abadie, J.P.; Mirza, M.; Xu, B.; Farley, D.W.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Proc. NIPS* **2014**, *2*, 2672–2680.
7. Maxwell, A.E.; Warner, T.A.; Guillen, L.A. Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 1: Literature Review. *Remote Sens.* **2021**, *13*, 2450. [CrossRef]
8. Maxwell, A.E.; Warner, T.A.; Guillen, L.A. Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 2: Recommendations and Best Practices. *Remote Sens.* **2021**, *13*, 2591. [CrossRef]
9. Wang, Y.; Wei, G.Y.; Brooks, D. Benchmarking TPU, GPU, and CPU Platforms for Deep Learning. *arXiv* **2019**, arXiv:1907.10701. Available online: <https://arxiv.org/pdf/1907.10701.pdf> (accessed on 24 July 2019).
10. Zhao, H.; Bsi, T.T.; Wang, Z. A Natural Images Pre-Trained Deep Learning Method for Seismic Random Noise Attenuation. *Remote Sens.* **2022**, *14*, 263. [CrossRef]

11. Wang, D.; Wan, J.; Liu, S.; Chen, Y.; Yasir, M.; Xu, M.; Ren, P. BO-DRNet: An Improved Deep Learning Model for Oil Spill Detection by Polarimetric Features from SAR Images. *Remote Sens.* **2022**, *14*, 264. [[CrossRef](#)]
12. Guo, X.; Liu, Q.; Sharma, R.P.; Chen, Q.; Ye, Q.; Tang, S.; Fu, L. Tree Recognition on the Plantation Using UAV Images with Ultrahigh Spatial Resolution in a Complex Environment. *Remote Sens.* **2022**, *13*, 4122. [[CrossRef](#)]
13. Chang, Y.L.; Tan, T.H.; Lee, W.H.; Chang, L.; Chen, Y.N.; Fan, K.C.; Alkhaleefah, M. Consolidated Convolutional Neural Network for Hyperspectral Image Classification. *Remote Sens.* **2022**, *14*, 1571. [[CrossRef](#)]
14. Jamaluddin, I.; Thaipisutikul, T.; Chen, Y.N.; Chuang, C.H.; Hu, C.L. Alkhaleefah, M. MDPrePost-Net: A Spatial-Spectral-Temporal Fully Convolutional Network for Mapping of Mangrove Degradation Affected by Hurricane Irma 2017 Using Sentinel-2 Data. *Remote Sens.* **2021**, *13*, 5042. [[CrossRef](#)]
15. Li, L.; Ma, H.; Jia, Z. Change Detection from SAR Images Based on Convolutional Neural Networks Guided by Saliency Enhancement. *Remote Sens.* **2021**, *13*, 3697. [[CrossRef](#)]
16. He, J.; Lyu, D.; He, L.; Zhang, Y.; Xu, X.; Yi, H.; Tian, Q.; Liu, B.; Zhang, X. Combining Object-Oriented and Deep Learning Methods to Estimate Photosynthetic and Non-Photosynthetic Vegetation Cover in the Desert from Unmanned Aerial Vehicle Images with Consideration of Shadows. *Remote Sens.* **2023**, *15*, 105. [[CrossRef](#)]
17. Luo, J.; Hu, Y.; Li, J. Surround-Net: A Multi-Branch Arbitrary-Oriented Detector for Remote Sensing. *Remote Sens.* **2022**, *14*, 1751. [[CrossRef](#)]
18. Qu, Z.; Zhu, F.; Qi, C. Remote Sensing Image Target Detection: Improvement of the YOLOv3 Model with Auxiliary Networks. *Remote Sens.* **2021**, *13*, 3908. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.