remote sensing

MDPI

Article

Multi-Scale and Context-Aware Framework for Flood Segmentation in Post-Disaster High Resolution Aerial Images

Sultan Daud Khan [1,*] and Saleh Basalamah [2]

[1] Department of Computer Science, National University of Technology, Main IJP Road, Sector I-12, Islamabad 44000, Pakistan
[2] Department of Computer Engineering, Umm Al-Qura University, Mecca 24382, Saudi Arabia
* Correspondence: sultandaud@nutech.edu.pk

Abstract: Floods are the most frequent natural disasters, occurring almost every year around the globe. To mitigate the damage caused by a flood, it is important to timely assess the magnitude of the damage and efficiently conduct rescue operations, deploy security personnel and allocate resources to the affected areas. To efficiently respond to the natural disaster, it is very crucial to swiftly obtain accurate information, which is hard to obtain during a post-flood crisis. Generally, high resolution satellite images are predominantly used to obtain post-disaster information. Recently, deep learning models have achieved superior performance in extracting high-level semantic information from satellite images. However, due to the loss of multi-scale and global contextual features, existing deep learning models still face challenges in extracting complete and uninterrupted results. In this work, we proposed a novel deep learning semantic segmentation model that reduces the loss of multi-scale features and enhances global context awareness. Generally, the proposed framework consists of three modules, encoder, decoder and bridge, combined in a popular U-shaped scheme. The encoder and decoder modules of the framework introduce Res-inception units to obtain reliable multi-scale features and employ a bridge module (between the encoder and decoder) to capture global context. To demonstrate the effectiveness of the proposed framework, we perform an evaluation using a publicly available challenging dataset, FloodNet. Furthermore, we compare the performance of the proposed framework with other reference methods. We compare the proposed framework with recent reference models. Quantitative and qualitative results show that the proposed framework outperforms other reference models by an obvious margin.

Keywords: flood segmentation; remote sensing; deep learning; disaster assessment; scene understanding

Citation: Khan, S.D.; Basalamah, S. Multi-Scale and Context-Aware Framework for Flood Segmentation in Post-Disaster High Resolution Aerial Images. *Remote Sens.* **2023**, *15*, 2208. https://doi.org/10.3390/rs15082208

Academic Editor: Fumio Yamazaki

Received: 2 March 2023
Revised: 9 April 2023
Accepted: 20 April 2023
Published: 21 April 2023

1. Introduction

The rapid growth in urban population and severe atmospheric conditions lead to floods. Floods cause major societal and economic disruption, lead to the loss of life of humans and animals and cause severe damage to property. Due to the frequent occurrence of floods and the severity of damage cause by floods, several researchers and government agencies devised different methods and techniques for flood monitoring. However, most of the current flood monitoring techniques are based on manual analysis and require an expert to manually analyze huge amounts of data acquired through different sensors. This manual analysis of data is a tedious job and is always prone to errors due to limited human capabilities. Another traditional way of flood monitoring is to exploit optical imagery (acquired through optical sensors) to compute different water indices. These techniques adopt different threshold methods to identify water bodies in the image [1]. However, these methods suffer from the following limitations: (1) These methods provide information only about the existence of water bodies. (2) These methods do not provide real-time and automated flood monitoring analysis.

Due to rapid advancement in sensing technologies, significant amounts of data (in the form of satellite or aerial images) are readily available. These high resolution satellite

images contain detailed information which facilitates response teams to timely analyze the whole scene. They use this information to generate impact maps which summarize the magnitude of the damage in flooded area [2,3]. Currently, much analysis of satellite images is performed manually by an expert, which is a tedious and time consuming job. Due to the availability of large amounts of satellite images and increased demand for extracting crucial information from images, researchers have employed computer vision techniques to automate the process. To automatically extract detailed information from satellite images, researchers have adopted different image segmentation techniques. Image segmentation is a field of computer vision which predicts the confidence score of each pixel and transforms the input image into high-level semantic information.

Image semantic segmentation is a high-level computer vision task that provides an aid to scene understanding. Due to the high demand for scene understanding, image semantic segmentation is the center of interest for many researchers. Image semantic segmentation has numerous applications, including urban planning, smart agriculture, building mapping, etc. A comprehensive review of different segmentation models aimed at solving different computer vision tasks can be found in [4]. Despite the success of segmentation models in various computer vision tasks, few efforts have been made towards flood segmentation in satellite images.

Flood segmentation in aerial images is a challenging task compared to generic semantic segmentation in ground-level images due to following reasons: (1) Satellite images contain complex textures, since the images are acquired from a distant camera at an oblique angle. (2) Due to the complex background, patterns from the same class appear different (intra-class heterogeneity), while different patterns share similar features (inter-class homogeneity) [5]. (3) The size of objects in satellite images is very small and covers only small portion of the whole image. (4) In satellite images, there are significant variations in shape and scale of the same/different objects. Despite the success of deep learning models in various semantic segmentation from satellite images, few efforts have been made towards flood segmentation from satellite images. Recently, Rahnemoonfar et al. [6] proposed a dataset, FloodNet, and evaluated and compared the performance of different deep learning models, including PSPNet [7], ENet [8] and DeepLabv3 [9] on a proposed dataset. From experimental results, we have observed that due to aforementioned problems, existing segmentation models, including U-Net [10], U-Net++ [11], DeepLabv3 [9], FCN [12], ENet [8], SegNet [13], PSPNet [7] and Tiramisu [14] face challenges in characterizing patterns and identifying their boundaries.

To mitigate the aforementioned challenges, a novel framework is proposed that exploits both multi-scale and contextual information. To extract multi-scale information, the framework introduces Res-inception units and employs a pyramid scene pooling network to capture global context. Generally, the framework consists of three modules, encoder, decoder and bridge, combined in a popular U-shaped scheme as adopted in U-Net [10]. This scheme enhances the receptive field of the network, which enables the network to combine the local multi-scale feature and global contextual information to produce a precise segmentation mask.

The contribution of this work is summarized as follows:

1. For flood segmentation in satellite images, it is crucial to extract multi-scale features. For this purpose, the proposed framework introduces Res-inception units that extract multi-scale features from multiple layers of the network.
2. The framework integrates a pyramid scene parsing network as bridge between the encoder and decoder modules to capture global contextual information.
3. The effectiveness of proposed framework is gauged on the challenging publicly available FloodNet dataset. From quantitative and qualitative comparisons, it is demonstrated that the framework achieves the best performance and is better than the other reference methods.

We organize the rest of paper as follows: Section 2 discusses the related work. The technical details and architecture of proposed framework are provided in Section 3. Sec-

tion 4 discusses the performance of proposed framework along with comparison with reference methods. Finally, conclusions are given in Section 5.

2. Related Work

Due to superior performance of deep learning models, researchers have also made several strides towards employing deep learning models for scene understanding in satellite images. In this section, we discuss different segmentation models in satellite images. For convenience, we categorize the semantic segmentation methods into two categories. In the first category, we discuss the handcrafted feature-based methods developed for semantic segmentation in satellite images, while in the second category, we discuss the deep learning models.

2.1. Handcrafted Feature-Based Models

Before the advent of deep learning networks, most of best performance algorithms relied on handcrafted features. These models extract handcrafted features, for example, Histograms of Oriented Gradients (HOG) [15], Scale Invariant Feature Transform (SIFT) [16], Local Binary Pattern (LBP) [17] and Gray Level Cooccurrence Matrix (GLCM) [18] use features to train a statistical classifier that obtains a semantic segmentation map by classifying the pixels of the input image. Low-level features, namely, semantic textons are proposed in [19], which combines decision trees to classify image pixels. The authors of [20] combine appearance and motion features and employ a probabilistic model based on conditional random field for semantic segmentation in road scenes. Markov Random Field (MRF) is employed in [21] to segment objects in street scene images. In [22], color and texture descriptors are computed for superpixels and train two separate classifiers based on KNN classifiers to classify superpixels to generate the segmentation map. Similarly, in [23], color and texture features are extracted from different regions of the image and train an SVM model to classify the pixels. LBP features are extracted from each region from the image, which are combined with spectral features in [24] for segmentation of high resolution satellite images. An entropy-based technique is proposed in [25] for automatic segmentation of color aerial images. The authors also evaluated the performance of the model on grey aerial images and conclude that the model performed better on color images than grey images. A non-supervised multicomponent aerial image segmentation model is proposed in [26] that employs a self-organizing map (SOM) and hybrid genetic algorithm (HGA). The self-organizing map is used to extract discriminating features from the image. Based on extracted features, different regions of the image are clustered into homogeneous regions by employing the hybrid genetic algorithm (HGA). A land cover segmentation model is proposed in [27] that employs the Structured Support Vector Machines (SSVM) model to learn appearance features and local class interactions. An adaptive mean-shift clustering algorithm is employed in [28] for semantic segmentation in satellite images. The model first extracts color and texture features from different areas of the image and then employs a mean-shift clustering algorithm to combine the homogeneous region of the image. A semantic segmentation model is proposed in [29] for urban aerial images. The model embeds geographic context in a pairwise CRF model and trains the random forest model on multiple descriptors to obtain class likelihood of superpixels.

Although these handcrafted feature-based models perform well in simple semantic segmentation tasks, these models exhibit poor performance in complex scenes. This may be attributed to the following reasons: (1) These models reply on manual computation of complex features which increases the computational cost. (2) Handcrafted features are not robust and are prone to noise and illumination changes. (3) These models lack global context and multi-scale features, because of which these models generally confuse different patterns, leading to misclassification.

2.2. Deep Learning Models

Deep learning models achieved tremendous success in various visual tasks, including object detection [30], image recognition [31] and semantic segmentation [12]. With the success of deep learning models in natural images, researchers have explored and applied various deep learning models in aerial image analysis to extract meaningful information for scene understanding.

Generally, semantic segmentation from aerial images can be categorized in the following categories: (1) road extraction, (2) building extraction and (3) land-cover segmentation.

Road extraction from satellite images offers crucial information for intelligent traffic monitoring. This information can be utilized to detect newly constructed roads and automatically update maps accordingly. Because of this reason, a significant amount of work [32–37] is reported in the literature regarding road extraction from satellite images. A detailed survey of road extraction from satellite images is reported in [38].

Building extraction from satellite images has wide range of applications in urban planning [39], disaster management [40,41] and population estimation [42]. Although several models [43–47] have been proposed in recent years for automatic building footprints' extraction from satellite images, these models suffer from a scale problem. Due to the different sizes of buildings, it becomes challenging for the models to precisely extract building footprints from satellite images. For example, the MFBI model is proposed in [48] to address the problem of multiple scales. For multiple region extraction, an attention module with multi-scale guidance framework is proposed in [49]. A multi-scale encoder–decoder framework is reported in [50] to extract local and global features to model the complex and diverse shapes of buildings from satellite images.

Land cover segmentation provides high-level semantic information about the land classified into forests, vegetation, grasslands and barren lands. Such information is useful for land use management [51] and precision agriculture [52]. Due to immense advantages of land cover segmentation, several researchers have developed various deep learning models [53–57] for automatic segmentation of land cover types from high resolution satellite images.

In addition to the above-mentioned methods, several methods have been reported to extract high-level semantic information for other tasks, including slum segmentation [58], farmland segmentation [59,60] and segmentation of residential solar panels [61,62]. A fully convolutional network (FCN) is proposed in [63] to identify slums in satellite images. Similarly, a deep fully convolutional network is proposed in [64] for sea–land segmentation in satellite images. The network follows a similar pipeline as that of the popular U-Net [10] (initially introduced for bio-medical image segmentation); however, instead of using convolutional layers in the encoder and decoder parts, DeepUNet introduced DownBlocks in the encoder part and UpBlocks in the decoder part. These two blocks are connected via U-connection and Plus connections to obtain more precise segmentation results. TreeUNet [65] extended DeepUNet by introducing skip connections to discriminate the pixels of apparently similar classes for land cover segmentation in satellite images. Similarly, a deep learning framework, ResUNet-a, is proposed in [66] that integrates atrous convolution layers, pyramid scene parsing and residual connection with UNet to identify the boundaries of different patterns. Recently, an attention mechanism has been introduced in deep learning networks to model long range dependencies and further refine the feature maps. In this strategy, the network focuses more on the object of interest and pays little attention to the background. A channel attention mechanism that is integrated with FCN is proposed in [67] for semantic segmentation of aerial images. Similarly, a hybrid attention mechanism is introduced in [68] to capture global relationships for a better representation of features.

3. Methodology

In this section, we discuss the details of the proposed framework for flood segmentation in satellite images. The detailed architecture of the framework is illustrated in Figure 1.

The goal of the framework is to enrich each pixel of the input image with a suitable class label. Generally, the structure of proposed framework is similar to U-Net [10]. However, the framework integrates Res-inception units along with a pyramid pooling module to effectively extract multi-scale features and global contextual information. Generally, the framework consists of three modules: (1) encoder, (2) decoder and (3) bridge. The encoder module consists of four Res-inception units, where each unit is followed by max-pooling layer. The decoder module comprises three Res-inception units. The encoder and decoder modules are combined with a bridge module, resulting in a U-shaped structure. The encoder modules takes an image of arbitrary size and extracts multi-scale hierarchical features. The feature map generated from the fourth Res-inception unit is then passed through a bridge module, which extracts global contextual information by employing the pyramid pooling scheme. The decoder module up-samples the feature maps obtained from the bridge module and concatenates the features maps with respective features maps from the encoder to recover spatial information. We then employ a convolutional layer followed by a Softmax layer to obtain a dense semantic map, where each pixel in the map represents the appropriate class. We discuss each module of the framework as follows:

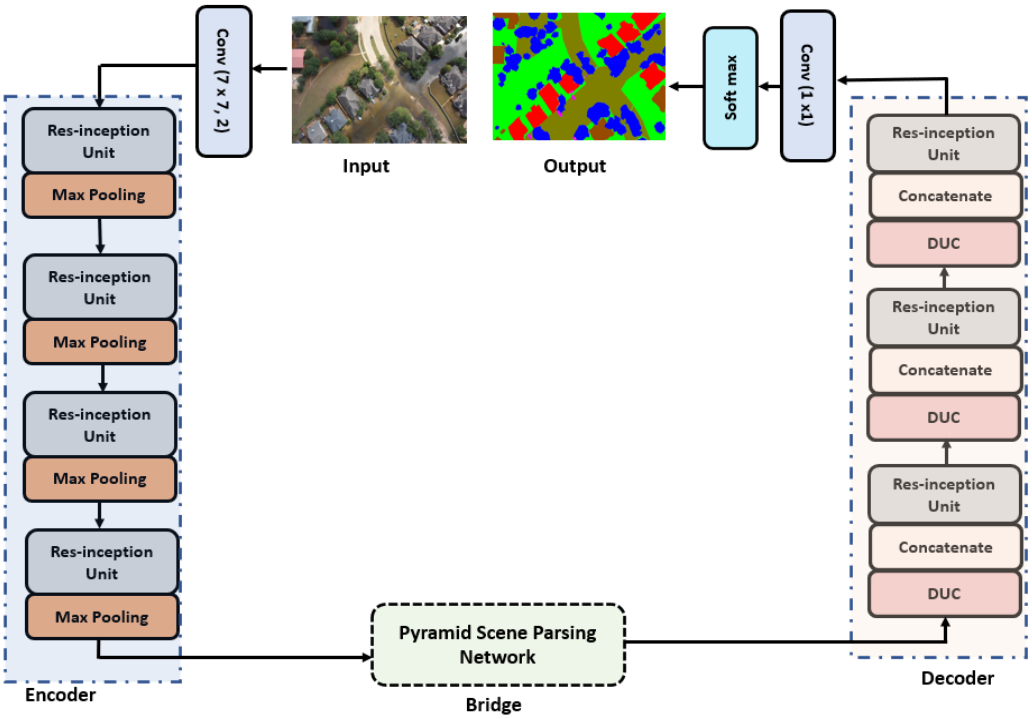

Figure 1. Detailed architecture of proposed framework for flood segmentation in satellite images.

Generally, the encoder part is considered the feature extractor or reduction part, which extracts hierarchical features and reduces the feature map after passing through subsequent pooling layers. Typically, the encoder part of classical U-Net is shallow and cannot extract multi-scale features, which is crucial for dense segmentation in satellite images [55]. To remedy this problem, instead of using simple convolutional and pooling filters, we use Res-inception units to aggregate rich features from multiple branches with different kernel sizes. We argue that this setting increases the width of the network and makes it capable of learning multi-scale features.

The structural diagram of the Res-inception unit is shown in Figure 2. As illustrated from Figure 2, the input feature maps from the preceding layer are provided as input to three convolutional blocks in parallel. Ω_{t-1} is the feature map from the previous layer. $\alpha_{n\times n}$ is the convolutional operation of size $n \times n$, and β represents the batch normalization operation. C_1 represents the feature maps obtained from the first convolutional block, C_2 is the output of the second convolutional block and C_3 represents the feature maps obtained

from the third convolutional block. We obtain the current feature map Ω_t after passing the Ω_{t-1} through the Res-inception module using the following process.

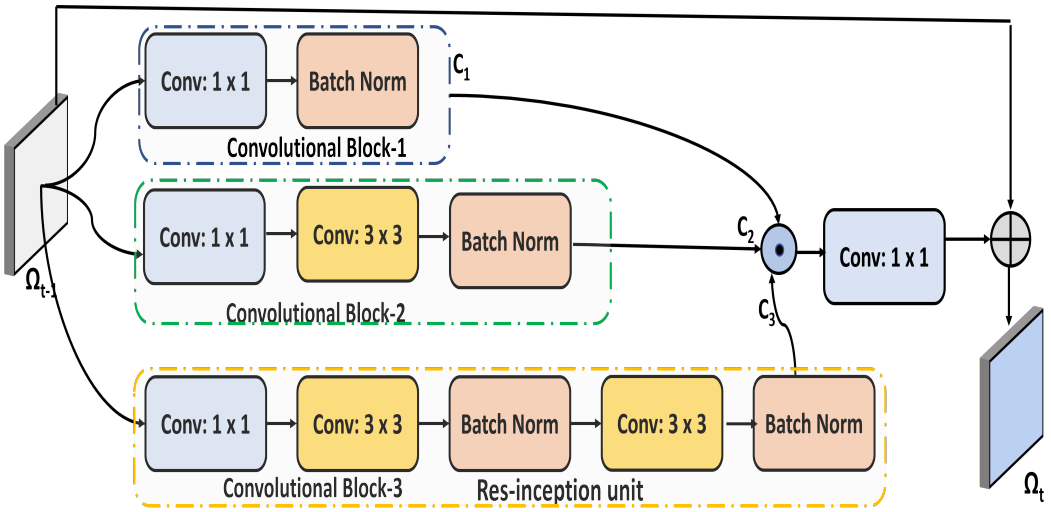

Figure 2. Structural diagram of Res-inception unit.

The feature map of first convolutional block C_1 is obtained by employing a convolutional filter of size 1×1, followed by a batch normalization layer as formulated in Equation (1). The purpose of using the batch normalization layer is to eliminate the gradient vanishing problem. Similarly, the input from the previous layer Ω_{t-1} is applied to the second convolutional block that employs a 1×1 convolutional layer, followed by a 3×3 convolutional layer and batch normalization layer as formulated in Equation (2). The third convolutional block applies a convolutional kernel of size 1×1, followed by two convolutional layers of size 3×3 and two batch normalization layers as formulated in Equation (3). Then the resulting feature map Ω_t is obtained by first concatenating the feature maps of C_1, C_2, C_3 using the operation $\odot$ and then applying a convolutional operation of size 1×1. The resulting feature map is then summed with the input Ω_{t-1} as formulated in Equation (4). Due to this unique structure, the network becomes wider instead of deeper, which makes the network capable of learning spatial patterns as well as multi-scale features. We argue that Res-inception learns both depth-wise and spatial patterns by employing different convolutional operations with different sizes. Res-inception learns depth-wise patterns by employing a convolutional operation of size 1×1 and learns spatial patterns by employing a convolutional layer of size 3×3. The unit further increases the representation power of the learned feature by concatenating feature maps learned from convolutional blocks. The Res-inception unit also reduces the computation complexity by factorizing the large convolutional operations into small ones without compromising the performance.

$$C_1 = \beta(\alpha_{1\times1}(\Omega_{t-1})) \tag{1}$$

$$C_2 = \beta(\alpha_{3\times3}(\alpha_{1\times1}(\Omega_{t-1}))) \tag{2}$$

$$C_3 = \beta(\alpha_{3\times3}(\beta(\alpha_{3\times3}(\alpha_{1\times1)}(\Omega_{t-1}))))) \tag{3}$$

$$\Omega_t = \alpha_{1\times1}(C_1 \odot C_2 \odot C_3) \oplus \Omega_{t-1} \tag{4}$$

Although the encoded feature map Ω_t obtained from the encoder module captures multi-scale features by employing Res-inception units, it cannot capture contextual information. As mentioned above, flood segmentation in satellite images is a challenging task due to inter-class homogeneity and intra-class heterogeneity. This is due to the following reasons: (1) images are acquired from a distant camera, (2) images contain complex textures (3) and there is uneven distribution of samples of different classes. To precisely character-

ize different patterns with similar appearances and alleviate semantic ambiguity among different patterns, it is imperative to consider contextual information [69].

To aggregate rich contextual information, we integrate a pyramid scene parsing network (PSPNet) [7] as a bridge between the encoder and decoder modules. PSPNet is a fully convolutional network (FCN) [12] that employs sub-region pooling operations of different scales to capture global context. The structural diagram of bridge network is illustrated in Figure 3. The bridge module takes feature map Ω_t as an input and employs pyramid pooling operations to obtain pooled maps of different sizes. In this work, we choose the number of levels of the pyramid to be 4, and consisting of different scales, 1×1, 2×2, 4×4 and 8×8. These sub-region pooling operations of different scales divide the feature map into different sub-regions and obtain a pooled representation for each sub-region. With this pyramid of pooled operations, the network captures both local and global context. After each pooling operation, the network employs a convolutional operation of size 1×1 to reduce the dimension. The pooled feature maps are up-sampled to the size of feature map Ω_t, and are then concatenated together to generate the final global feature map.

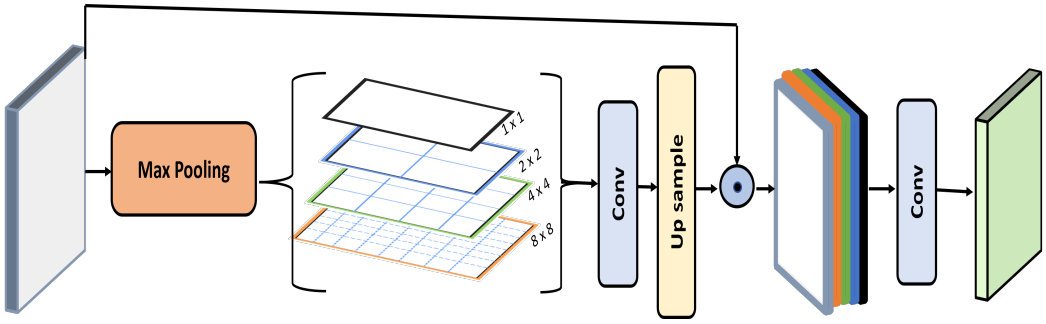

Figure 3. Structural diagram of bridge network.

The final global feature map is then provided as an input to the decoder module. The decoder module projects the global feature map onto a dense segmentation map, where each pixel of the map represents a class. The decoder module consists of a hierarchy of three decoder units, where each decoder unit contains a Res-inception unit and dense upsampling convolution (DUC) [70]. To recover spatial information and boost the feature representation, the decoder employs the DUC layer (instead of bilinear interpolation) to up-sample the feature map and then concatenate the up-sampled feature map with the pooled version of the corresponding encoder. The concatenation step combines high- and low-quality feature maps to boost the efficiency of the framework and precisely characterize the boundaries of different patterns [5]. In the end, the decoder employs a 1×1 convolutional layer followed by a softmax layer to generate the dense segmentation map.

For training the framework, we initialize the weights using the strategy adopted in [71] and employ stochastic gradient descent with a fixed learning rate. Due to the high resolution of the satellite images and to minimize the computational load on the GPU, we use a set of 10 images per batch. We train the network for 100 epochs, and after each epoch we reshuffle the training set to ensure that each image in the training set is seen only once by the network during entire training process.

Cross-entropy loss is a commonly used objective function to minimize the pixel loss for segmentation problems; however, it alone is not suitable in our case. This because there is considerable variation in the number of pixels of different classes in the flood dataset. Furthermore, the number of samples per class is also different. This creates a class imbalance problem which may change the shape of loss function, and the network may over-represent the bigger class compared to the smaller one. Therefore, we use the hybrid loss function [55]. This loss function is the linear combination of both cross-entropy loss and dice-loss as formulated in Equation (5).

$$L_{hybrid} = L_{cross} + L_{dice} \tag{5}$$

where L_{cross} and L_{dice} are given by Equations (6) and (7), respectively

$$L_{cross}(k, \hat{k}) = -\frac{1}{T_c} \sum_{n}^{T_c} k \ln(\hat{k}) + (1-k) \ln(1-\hat{k}) \tag{6}$$

where k and $\hat{k}$ are the ground truth label and the predicted label of the pixel, respectively, and T_c is the total number of classes.

$$L_{dice} = \frac{2|\Psi_g \cap \Psi_p|}{|\Psi_g| + |\Psi_p|} \tag{7}$$

where Ψ_g is the annotated mask with each pixel representing the class label and Ψ_p is the predicted segmentation mask, where each pixel represents the predicted label.

4. Experiment Results

In this section, we performed extensive evaluation of the proposed framework and also compared its performance with other reference methods. We evaluated and compared the performance of different frameworks on the publicly available challenging dataset, FloodNet. The FloodNet dataset was first proposed by Rahnemoonfar [6], and is the only comprehensive dataset available currently . The dataset was collected on August 2017 from the hurricane landfall that took place in Texas and Louisiana. DJI Mavic Pro quadcopters were used to collect images from the height of 200 feet. The dataset consists of a total of 2343 images. Each image of the dataset covers a spatial resolution of 1.5 cm with pixel resolution of 4000 × 3000 pixels.

The dataset contains of 9 different classes and contains complex structures and textures. These classes are labeled as follows: 1 → Building—flooded, 2 → Building—non-flooded, 3 → Road—flooded, 4 → Road—non-flooded, 5 → Vehicle, 6 → Pool, 7 → Tree, 8 → Water, 9 → Grass. Sample images and their corresponding ground truth masks are shown in Figure 4.

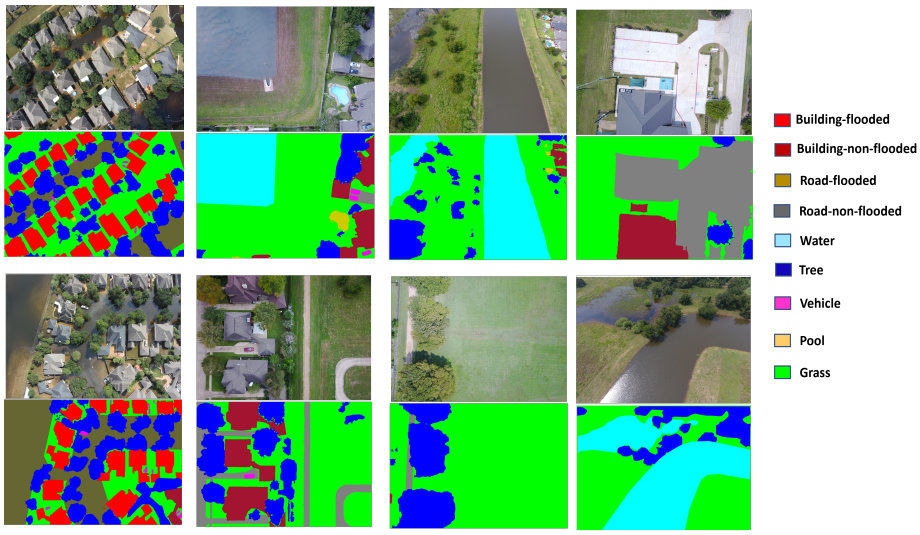

Figure 4. Sample images of the dataset. First and third row show the sample images, while second and fourth row represent the ground truth masks.

The distribution of number of images and distribution of number of instances per class are shown in Figure 5 and Table 1, respectively. From Figure 5 and Table 1, it is obvious that the distribution of images per class and instances per class is uneven, which causes a class imbalance problem. The class imbalance problem will lead the network to become more biased towards the class that contains a larger number of instances. To avoid this problem, we randomly selected 200 samples from each class and adopted a data augmentation technique to generate flipped, scaled and shuffled samples of the original images. It is worth mentioning that the resolution of satellite images is large, which

increases computational complexity during the training process. To avoid this problem, we cropped patches of the size 512 × 512 from each sample of the training set and used this set of those patches for training the network. For training the network, we adopted the strategy in [6] and used 70% of the data for training, and the remaining 30% was used for validation.

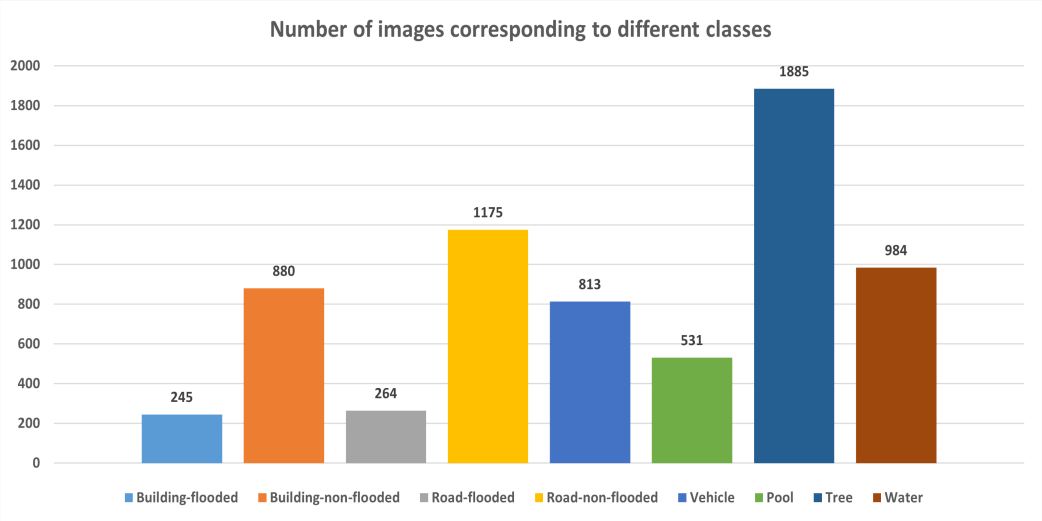

Figure 5. Distribution of images corresponding to different classes.

Table 1. Number of instances per each class.

Number of Instances	Class
3248	Building—flooded
3427	Building—non-flooded
495	Road—flooded
2155	Road—non-flooded
4535	Vehicle
1141	Pool
19,682	Tree
1374	Water

To evaluate the pixel-wise performance of proposed framework, we computed the precision, recall and F1-score for each class. Precision is formulated as: $\frac{TP}{TP+FP}$, recall is formulated as $\frac{TP}{TP+FN}$, and the F1-score is computed as $2 * \frac{Precision*Recall}{Precision+Recall}$. These evaluation metrics measure how well the model precisely classifies pixels into different classes. For computing these evaluation metrics, we computed True Positive (*TP*), False Positive (*FP*), True Negative (*TN*) and False Negative (*FN*). TP measures the number of pixels correctly classified as the given class, FP measures number of incorrectly classified pixels, TN measures the number of pixels correctly classified as not the given class, while FN measures the number of pixels incorrectly classified as not the given class. We report the results in Table 2. It is evident from Table 2 that the proposed framework achieves high precision and recall values for large and medium patterns, for example, Building—non-flooded, Road—non-flooded, Water, Tree and Grass. However, the proposed method achieves low precision and recall values for the Vehicle and Pool classes. This may be attributed to the fact that the sizes of vehicles and pools are small. The framework utilizes the feature maps of the last Res-inception unit in the encoder module, which usually leads to the loss of information about small objects due to subsequent pooling layers.

Table 2. Pixel-wise performance of proposed method using precision, recall and F1-score.

	Precision	Recall	F1-Score
Building—flooded	0.75	0.71	0.73
Building—non-flooded	0.90	0.94	0.92
Road—flooded	0.82	0.84	0.83
Road—non-flooded	0.92	0.93	0.92
Water	0.94	0.93	0.94
Tree	0.92	0.94	0.93
Vehicle	0.67	0.70	0.69
Pool	0.76	0.69	0.72
Grass	0.94	0.95	0.95

To quantify the performance of the proposed framework and compare its performance with other reference methods, we used the Jaccard similarity index, also known as Intersection Over Union (IoU). The Jaccard similarity index is an evaluation metric widely used for assessing the performance of segmentation models and is formulated as $\frac{\Psi_p \cap Psi_g}{\Psi_p \cup \Psi_g}$, where Ψ_g is the annotated mask and Ψ_p is the predicted mask. We computed the Jaccard similarity index for each class and then computed mean Intersection-over-Union (mIoU) to summarize the performance of the methods.

For performance comparisons, we chose popular segmentation models, including U-Net [10], U-Net++ [11], DeepLabv3 [9], FCN [12], ENet [8], SegNet [13], PSPNet [7] and Tiramisu [14]. These segmentation models achieve superior performance in various segmentation tasks. We employed pre-trained models of these methods and fine-tuned the models on the FloodNet dataset. For each method, we computed the Jaccard similarity index for all classes and also computed the mIoU. The results of the methods, along with proposed framework, are reported in Table 3. From Table 3, we observe that the proposed framework outperforms other reference methods by a considerable margin. From the Table 3, it is observed that FCN achieved a comparatively low performance. This may be due to the fact that FCN usually lost spatial information during the decoding stage because of which FCN faces difficulties in detecting small objects such as vehicles and pools. It is also unable to differentiate between the boundaries of two different patterns. Furthermore, due to lack of contextual information, FCN is unable to capture inter-class differences between two different classes such as building—flooded and road—flooded classes. SegNet, on the other hand, achieved slightly better results than FCN. This is due to how SegNet, in contrast to FCN, up-samples the feature maps by utilizing the indices of max-pooling layers saved during the encoding stage. This strategy is useful to alleviate the problem of insufficient spatial information and also minimizes memory requirements. However, excessive down-sampling during the encoding stage and lack of contextual information hurt the performance of SegNet. ENet achieved a good performance compared to FCN and SegNet. In contrast to SegNet, which overly utilizes down-sampling, ENet utilizes dilation convolutions to obtain rich contextual information and retain spatial information. It is worth mentioning that excessive down-sampling leads to the loss of spatial information such as boundaries and edge information, as well as the details of small objects. However, down-sampling filters have large receptive fields that may be useful in capturing the global contextual information and are always useful in differentiating two different patterns. The proposed framework follows a similar pipeline to that of U-Net and U-Net++; however, the proposed framework outperforms these models. One of the reasons for the lower performance of U-Net is that the encoder module of the network is of limited depth and cannot capture the fine-grained multi-scale information that is required for semantic segmentation in high resolution satellite images. U-Net++ achieves a performance gain over U-Net by redesigning the skip-connection to obtain fine-

grained details of foreground patterns by enriching the feature maps of the encoder module before fusing the feature maps with corresponding feature maps of the decoder module. PSPNet achieves comparable performance. PSPNet aggregates contextual information from different regions of the image by employing sub-region pooling operations of different scales. This model is well-suited for semantic segmentation task in satellite images; however, the model has limited capability to extract finite features and sometimes ignores semantic boundary information of different patterns. Compared to reference models, the proposed model employs Res-inception units in the encoder module to extract multi-scale features, and integrates a PSPNet as a bridge network to aggregate global contextual information, which is crucial for differentiating boundaries of different patterns.

We report qualitative comparisons of different methods in Figure 6. The first row of Figure 6 shows sample input satellite images (randomly sampled from test set). The second column represents the ground truth annotation, while the remaining columns display the output of different methods. From Figure 6, it is obvious that the proposed framework obtains results close to the ground truth. The performance of SegNet is relatively lower than the other competing methods. PSPNet, on the other hand, produces comparable results; however, PSPNet faces challenges in identifying small objects, for example vehicles.

Table 3. Comparison of different methods using Jaccard index and mIoU. BF: Building—flooded, BnF: Building—non-flooded, RF: Road—flooded, RnF: Road—non-flooded, W: Water, T: Tree, V: Vehicle, P: Pool, G: Grass.

Method	BF	BnF	RF	RnF	W	T	V	P	G	mIoU
U-Net	21.83	69.49	23.84	72.45	63.62	70.19	21.48	37.52	77.16	50.84
DeepLabv3+	28.1	78.1	32	81.1	73	74.5	33.6	40	87.1	58.61
FCN	18.75	28.42	20.64	37.84	42.01	40.95	19.24	22.94	52.49	31.48
U-Net++	24.34	71.46	27.61	74.96	69.78	72.6	25.16	38.64	80.75	53.92
ENet	21.82	41.41	14.76	52.53	47.14	62.56	26.21	16.57	75.57	39.84
SegNet	17.82	38.54	10.81	48.76	34.97	55.23	19.32	21.53	68.74	35.08
PSPNet	65.61	90.92	78.69	90.9	91.25	89.17	54.83	66.37	95.45	80.35
Tiramisu	31.53	75.68	24.72	78.61	72.85	74.97	23.74	40.18	84.62	56.32
Proposed	72.71	91.27	84.62	92.34	93.64	93.72	67.82	71.49	94.86	84.72

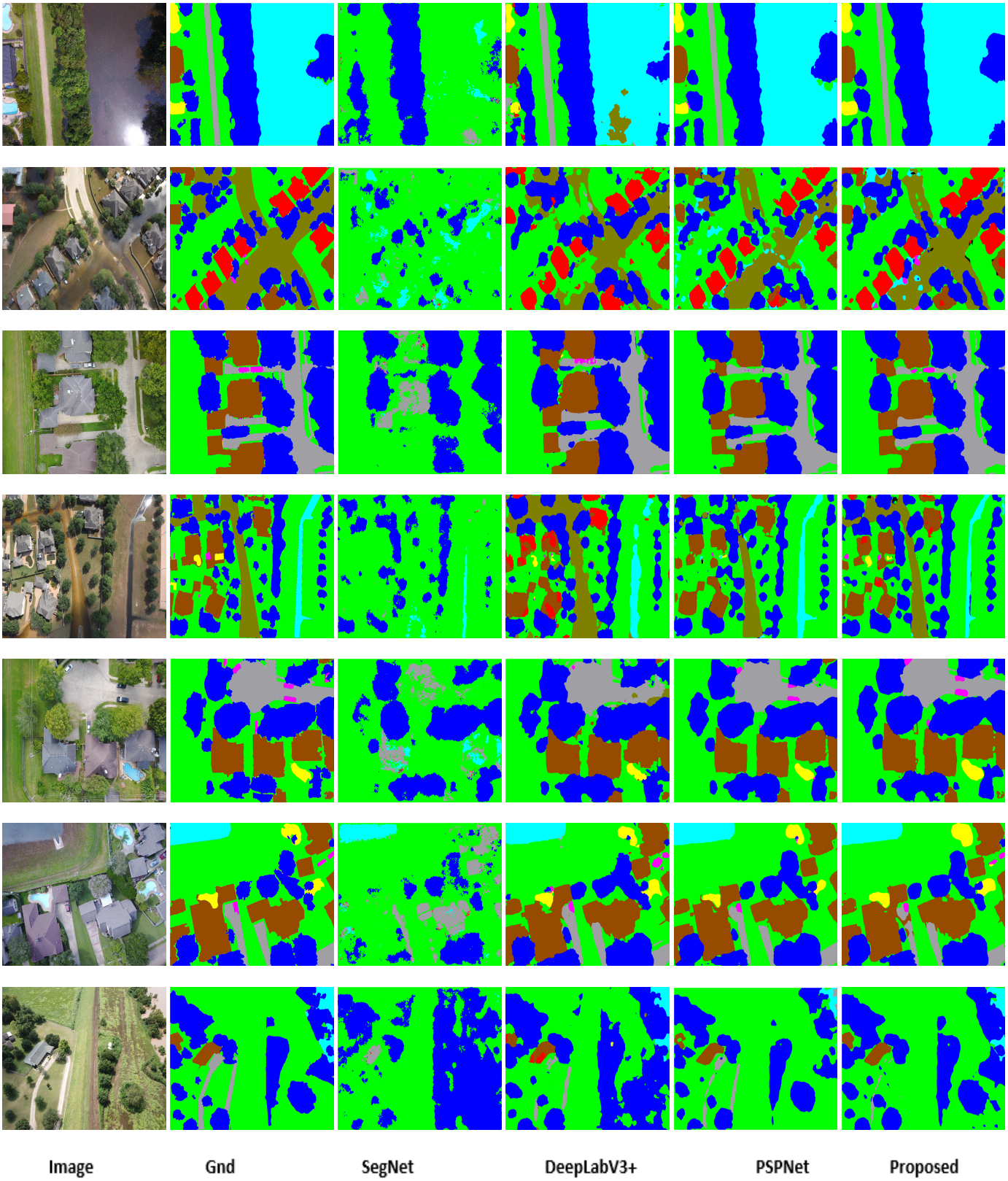

Figure 6. Qualitative comparison of different methods for flood segmentation.

5. Conclusions

In this work, we proposed a novel framework for flood segmentation in high resolution satellite images. We observed that multi-scale features and learning global context are

crucial for flood segmentation in satellite images. The proposed framework addressed these problems by introducing a Res-inception unit in the encoder and decoder modules and utilizes PSPNet as a bridge module. Although the framework follows a similar pipeline as that of U-Net, it achieves better performance, enhances global context awareness and reduces the loss of multi-scale features. The encoder module extracts multi-scale features by incorporating Res-inception units, while the decoder recovers the spatial information lost during the down-sampling operations in the encoder module. The framework effectively integrates a bridge module between the encoder and decoder modules which enhances the global contextual intelligence. The framework was evaluated on the publicly available benchmark dataset, FloodNet. From the experiment results, we show the effectiveness of the proposed framework, and demonstrate that the proposed framework achieves superior results compared to the reference methods.

Despite achieving good performance, the framework also suffers from some limitations. Inclusion of Res-inception units in encoder and decoder modules, and integration of a bridge module increase the complexity of the framework. Furthermore, too many parameters increase the training time. To remedy these issues, the proposed framework needs to be optimized, which is an iterative process and involves extensive experimentation of trying different combinations of optimization algorithms, learning rates, weight initialization techniques, regularization methods, hyperparameters and careful tuning of these parameters to achieve optimal results without compromising the performance of the proposed framework.

Author Contributions: Conceptualization, S.D.K. and S.B.; methodology, S.D.K.; software, S.D.K.; validation, S.D.K., S.B; formal analysis, S.D.K.; writing—original draft preparation, S.D.K.; writing—review and editing, S.D.K and S.B.; visualization, S.D.K.; supervision, S.B.; project administration, S.B.; funding acquisition, S.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References

1. Memon, A.A.; Muhammad, S.; Rahman, S.; Haq, M. Flood monitoring and damage assessment using water indices: A case study of Pakistan flood-2012. *Egypt. J. Remote Sens. Space Sci.* **2015**, *18*, 99–106. [CrossRef]
2. Schumann, G.J.; Brakenridge, G.R.; Kettner, A.J.; Kashif, R.; Niebuhr, E. Assisting flood disaster response with earth observation data and products: A critical assessment. *Remote Sens.* **2018**, *10*, 1230. [CrossRef]
3. Abid, S.K.; Sulaiman, N.; Chan, S.W.; Nazir, U.; Abid, M.; Han, H.; Ariza-Montes, A.; Vega-Muñoz, A. Toward an integrated disaster management approach: How artificial intelligence can boost disaster management. *Sustainability* **2021**, *13*, 12560. [CrossRef]
4. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A review on deep learning techniques applied to semantic segmentation. *arXiv* **2017**, arXiv:1704.06857.
5. Lal, S.; Nalini, J.; Reddy, C.S.; Dell'Acqua, F. DIResUNet: Architecture for multiclass semantic segmentation of high resolution remote sensing imagery data. *Appl. Intell.* **2022**, *52*, 15462–15482.
6. Rahnemoonfar, M.; Chowdhury, T.; Sarkar, A.; Varshney, D.; Yari, M.; Murphy, R.R. Floodnet: A high resolution aerial imagery dataset for post flood scene understanding. *IEEE Access* **2021**, *9*, 89644–89654. [CrossRef]
7. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017 ; pp. 2881–2890.
8. Paszke, A.; Chaurasia, A.; Kim, S.; Culurciello, E. Enet: A deep neural network architecture for real-time semantic segmentation. *arXiv* **2016**, arXiv:1606.02147.
9. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
10. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
11. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–11.

12. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
13. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef]
14. Jégou, S.; Drozdzal, M.; Vazquez, D.; Romero, A.; Bengio, Y. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 11–19.
15. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–26 June 2005; Volume 1, pp. 886–893.
16. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
17. Li, M.; Staunton, R.C. Optimum Gabor filter design and local binary patterns for texture segmentation. *Pattern Recognit. Lett.* **2008**, *29*, 664–672. [CrossRef]
18. Suresh, A.; Shunmuganathan, K. Image texture classification using gray level co-occurrence matrix based statistical features. *Eur. J. Sci. Res.* **2012**, *75*, 591–597.
19. Shotton, J.; Johnson, M.; Cipolla, R. Semantic texton forests for image categorization and segmentation. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska, 23–28 June 2008; pp. 1–8.
20. Sturgess, P.; Alahari, K.; Ladicky, L.; Torr, P.H. Combining appearance and structure from motion features for road scene understanding. In Proceedings of the BMVC-British Machine Vision Conference, BMVA, London, UK, 7–10 September 2009.
21. Zhang, C.; Wang, L.; Yang, R. Semantic segmentation of urban scenes using dense depth maps. In Proceedings of the European Conference on Computer Vision, Heraklion, Greece, 5–11 September 2011; Springer: Berlin/Heidelberg, Germany, 2010; pp. 708–721.
22. Ghiasi, M.; Amirfattahi, R. Fast semantic segmentation of aerial images based on color and texture. In Proceedings of the 2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP), Zanjan, Iran, 10–12 September 2013; pp. 324–327.
23. Wang, X.Y.; Wang, T.; Bu, J. Color image segmentation using pixel wise support vector machine classification. *Pattern Recognit.* **2011**, *44*, 777–787. [CrossRef]
24. Wang, A.; Wang, S.; Lucieer, A. Segmentation of multispectral high-resolution satellite imagery based on integrated feature distributions. *Int. J. Remote Sens.* **2010**, *31*, 1471–1483. [CrossRef]
25. Barbieri, A.L.; De Arruda, G.; Rodrigues, F.A.; Bruno, O.M.; da Fontoura Costa, L. An entropy-based approach to automatic image segmentation of satellite images. *Phys. A Stat. Mech. Its Appl.* **2011**, *390*, 512–518. [CrossRef]
26. Awad, M.; Chehdi, K.; Nasri, A. Multicomponent image segmentation using a genetic algorithm and artificial neural network. *IEEE Geosci. Remote Sens. Lett.* **2007**, *4*, 571–575. [CrossRef]
27. Volpi, M.; Ferrari, V. Semantic segmentation of urban scenes by learning local class interactions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
28. Banerjee, B.; Varma, S.; Buddhiraju, K.M. Satellite image segmentation: A novel adaptive mean-shift clustering based approach. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 4319–4322.
29. Volpi, M.; Ferrari, V. Structured prediction for urban scene semantic segmentation with geographic context. In Proceedings of the 2015 Joint Urban Remote Sensing Event (JURSE), Lausanne, Switzerland, 30 March–1 April 2015; pp. 1–4.
30. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
31. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
32. Ren, Y.; Yu, Y.; Guan, H. DA-CapsUNet: A dual-attention capsule U-Net for road extraction from remote sensing imagery. *Remote Sens.* **2020**, *12*, 2866. [CrossRef]
33. Khan, S.D.; Alarabi, L.; Basalamah, S. DSMSA-Net: Deep Spatial and Multi-scale Attention Network for Road Extraction in High Spatial Resolution Satellite Images. *Arab. J. Sci. Eng.* **2023**, *48*, 1907–1920. [CrossRef]
34. Wulamu, A.; Shi, Z.; Zhang, D.; He, Z. Multiscale road extraction in remote sensing images. *Comput. Intell. Neurosci.* **2019**, *2019*, 2373798. [CrossRef]
35. Lian, R.; Huang, L. DeepWindow: Sliding window based on deep learning for road extraction from remote sensing images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 1905–1916. [CrossRef]
36. Xu, Y.; Xie, Z.; Feng, Y.; Chen, Z. Road extraction from high-resolution remote sensing imagery using deep learning. *Remote Sens.* **2018**, *10*, 1461. [CrossRef]
37. Li, P.; He, X.; Qiao, M.; Cheng, X.; Li, Z.; Luo, H.; Song, D.; Li, D.; Hu, S.; Li, R.; et al. Robust deep neural networks for road extraction from remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 6182–6197. [CrossRef]
38. Chen, Z.; Deng, L.; Luo, Y.; Li, D.; Junior, J.M.; Gonçalves, W.N.; Nurunnabi, A.A.M.; Li, J.; Wang, C.; Li, D. Road extraction in remote sensing data: A survey. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102833. [CrossRef]
39. Hu, Q.; Zhen, L.; Mao, Y.; Zhou, X.; Zhou, G. Automated building extraction using satellite remote sensing imagery. *Autom. Constr.* **2021**, *123*, 103509. [CrossRef]

40. Rudner, T.G.; Rußwurm, M.; Fil, J.; Pelich, R.; Bischke, B.; Kopačková, V.; Biliński, P. Multi3Net: Segmenting flooded buildings via fusion of multiresolution, multisensor, and multitemporal satellite imagery. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 702–709.
41. Li, S.; Tang, H.; Huang, X.; Mao, T.; Niu, X. Automated detection of buildings from heterogeneous VHR satellite images for rapid response to natural disasters. *Remote Sens.* **2017**, *9*, 1177. [CrossRef]
42. Wu, S.S.; Qiu, X.; Wang, L. Population estimation methods in GIS and remote sensing: A review. *GISci. Remote Sens.* **2005**, *42*, 80–96. [CrossRef]
43. Na, Y.; Kim, J.H.; Lee, K.; Park, J.; Hwang, J.Y.; Choi, J.P. Domain adaptive transfer attack-based segmentation networks for building extraction from aerial images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5171–5182. [CrossRef]
44. Zhang, L.; Wu, J.; Fan, Y.; Gao, H.; Shao, Y. An efficient building extraction method from high spatial resolution remote sensing images based on improved mask R-CNN. *Sensors* **2020**, *20*, 1465. [CrossRef]
45. Liu, H.; Luo, J.; Huang, B.; Hu, X.; Sun, Y.; Yang, Y.; Xu, N.; Zhou, N. DE-Net: Deep encoding network for building extraction from high-resolution remote sensing imagery. *Remote Sens.* **2019**, *11*, 2380. [CrossRef]
46. Zhang, Z.; Guo, W.; Li, M.; Yu, W. GIS-supervised building extraction with label noise-adaptive fully convolutional neural network. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 2135–2139. [CrossRef]
47. Protopapadakis, E.; Doulamis, A.; Doulamis, N.; Maltezos, E. Stacked autoencoders driven by semi-supervised learning for building extraction from near infrared remote sensing imagery. *Remote Sens.* **2021**, *13*, 371. [CrossRef]
48. Bi, Q.; Qin, K.; Zhang, H.; Zhang, Y.; Li, Z.; Xu, K. A multi-scale filtering building index for building extraction in very high-resolution satellite imagery. *Remote Sens.* **2019**, *11*, 482. [CrossRef]
49. Li, K.; Hu, X.; Jiang, H.; Shu, Z.; Zhang, M. Attention-guided multi-scale segmentation neural network for interactive extraction of region objects from high-resolution satellite imagery. *Remote Sens.* **2020**, *12*, 789. [CrossRef]
50. Ma, J.; Wu, L.; Tang, X.; Liu, F.; Zhang, X.; Jiao, L. Building extraction of aerial images by a global and multi-scale encoder-decoder network. *Remote Sens.* **2020**, *12*, 2350. [CrossRef]
51. Stow, D.A.; Hope, A.; McGuire, D.; Verbyla, D.; Gamon, J.; Huemmrich, F.; Houston, S.; Racine, C.; Sturm, M.; Tape, K.; et al. Remote sensing of vegetation and land-cover change in Arctic Tundra Ecosystems. *Remote Sens. Environ.* **2004**, *89*, 281–308. [CrossRef]
52. Anand, T.; Sinha, S.; Mandal, M.; Chamola, V.; Yu, F.R. AgriSegNet: Deep aerial semantic segmentation framework for IoT-assisted precision agriculture. *IEEE Sens. J.* **2021**, *21*, 17581–17590. [CrossRef]
53. Perumal, B.; Kalaiyarasi, M.; Deny, J.; Muneeswaran, V. Forestry land cover segmentation of SAR image using unsupervised ILKFCM. *Mater. Today Proc.* **2021**. [CrossRef]
54. Bengana, N.; Heikkilä, J. Improving land cover segmentation across satellites using domain adaptation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *14*, 1399–1410. [CrossRef]
55. Khan, S.D.; Alarabi, L.; Basalamah, S. Deep Hybrid Network for Land Cover Semantic Segmentation in High-Spatial Resolution Satellite Images. *Information* **2021**, *12*, 230. [CrossRef]
56. Atik, S.O.; Ipbuker, C. Integrating convolutional neural network and multiresolution segmentation for land cover and land use mapping using satellite imagery. *Appl. Sci.* **2021**, *11*, 5551. [CrossRef]
57. Sravya, N.; Lal, S.; Nalini, J.; Reddy, C.S.; Dell'Acqua, F. DPPNet: An Efficient and Robust Deep Learning Network for Land Cover Segmentation From High-Resolution Satellite Images. *IEEE Trans. Emerg. Top. Comput. Intell.* **2022**, *7*, 128–139.
58. Rehman, M.F.U.; Aftab, I.; Sultani, W.; Ali, M. Mapping Temporary Slums From Satellite Imagery Using a Semi-Supervised Approach. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
59. Xu, L.; Ming, D.; Zhou, W.; Bao, H.; Chen, Y.; Ling, X. Farmland extraction from high spatial resolution remote sensing images based on stratified scale pre-estimation. *Remote Sens.* **2019**, *11*, 108. [CrossRef]
60. Gao, X.; Liu, L.; Gong, H. MMUU-Net: A Robust and Effective Network for Farmland Segmentation of Satellite Imagery. *J. Phys. Conf. Ser.* **2020**; *1651*, 012189. [CrossRef]
61. Zhuang, L.; Zhang, Z.; Wang, L. The automatic segmentation of residential solar panels based on satellite images: A cross learning driven U-Net method. *Appl. Soft Comput.* **2020**, *92*, 106283. [CrossRef]
62. Li, P.; Zhang, H.; Guo, Z.; Lyu, S.; Chen, J.; Li, W.; Song, X.; Shibasaki, R.; Yan, J. Understanding rooftop PV panel semantic segmentation of satellite and aerial images for better using machine learning. *Adv. Appl. Energy* **2021**, *4*, 100057. [CrossRef]
63. Wurm, M.; Stark, T.; Zhu, X.X.; Weigand, M.; Taubenböck, H. Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 59–69. [CrossRef]
64. Li, R.; Liu, W.; Yang, L.; Sun, S.; Hu, W.; Zhang, F.; Li, W. DeepUNet: A deep fully convolutional network for pixel-level sea-land segmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3954–3962. [CrossRef]
65. Yue, K.; Yang, L.; Li, R.; Hu, W.; Zhang, F.; Li, W. TreeUNet: Adaptive tree convolutional neural networks for subdecimeter aerial image segmentation. *ISPRS J. Photogramm. Remote Sens.* **2019**, *156*, 1–13. [CrossRef]
66. Diakogiannis, F.I.; Waldner, F.; Caccetta, P.; Wu, C. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 94–114. [CrossRef]
67. Luo, H.; Chen, C.; Fang, L.; Zhu, X.; Lu, L. High-resolution aerial images semantic segmentation using deep fully convolutional network with channel attention mechanism. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3492–3507. [CrossRef]

68. Niu, R.; Sun, X.; Tian, Y.; Diao, W.; Chen, K.; Fu, K. Hybrid multiple attention network for semantic segmentation in aerial images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5603018. [CrossRef]
69. Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 325–341.
70. Mehta, S.; Rastegari, M.; Caspi, A.; Shapiro, L.; Hajishirzi, H. Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 552–568.
71. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1026–1034.