



Article

Semi-/Weakly-Supervised Semantic Segmentation Method and Its Application for Coastal Aquaculture Areas Based on Multi-Source Remote Sensing Images—Taking the Fujian Coastal Area (Mainly Sanduo) as an Example

Chenbin Liang ^{1,2,3}, Bo Cheng ^{1,*} , Baihua Xiao ², Chenlinqiu He ^{1,4}, Xunan Liu ⁵, Ning Jia ⁵ and Jinfen Chen ^{1,4}

¹ Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; liangchenbin2019@ia.ac.cn (C.L.); heclq@radi.ac.cn (C.H.); chenjf@aircas.ac.cn (J.C.)

² State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China; baihua.xiao@ia.ac.cn

³ School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100190, China

⁴ College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100190, China

⁵ National Marine Hazard Mitigation Service, Beijing 100194, China; xnliu@nmhms.org.cn (X.L.); njia@nmhms.org.cn (N.J.)

* Correspondence: chengbo@aircas.ac.cn; Tel.: +86-1362-101-2563



Citation: Liang, C.; Cheng, B.; Xiao, B.; He, C.; Liu, X.; Jia, N.; Chen, J. Semi-/Weakly-Supervised Semantic Segmentation Method and Its Application for Coastal Aquaculture Areas Based on Multi-Source Remote Sensing Images—Taking the Fujian Coastal Area (Mainly Sanduo) as an Example. *Remote Sens.* **2021**, *13*, 1083. <https://doi.org/10.3390/rs13061083>

Academic Editor: Thomas Blaschke

Received: 25 February 2021

Accepted: 9 March 2021

Published: 12 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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/>).

Abstract: Coastal aquaculture areas are some of the main areas to obtain marine fishery resources and are vulnerable to storm-tide disasters. Obtaining the information of coastal aquaculture areas quickly and accurately is important for the scientific management and planning of aquaculture resources. Recently, deep neural networks have been widely used in remote sensing to deal with many problems, such as scene classification and object detection, and there are many data sources with different spatial resolutions and different uses with the development of remote sensing technology. Thus, using deep learning networks to extract coastal aquaculture areas often encounters the following problems: (1) the difficulty in labeling; (2) the poor robustness of the model; (3) the spatial resolution of the image to be processed is inconsistent with that of the existing samples. In order to fix these problems, this paper proposes a novel semi-/weakly-supervised method, the semi-/weakly-supervised semantic segmentation network (Semi-SSN), and adopts 3 data sources: GaoFen-2 image, GaoFen-1(PMS)image, and GanFen-1(WFV)image with a 0.8 m, 2 m, and 16 m spatial resolution, respectively, and through experiments, we analyze the extraction effect of the model comprehensively. After comparing with other the-state-of-art methods and verifying on an open remote sensing dataset, we take the Fujian coastal area (mainly Sanduo) as the experimental area and employ our method to detect the effect of storm-tide disasters on coastal aquaculture areas, monitor the production, and make the distribution map of coastal aquaculture areas.

Keywords: coastal aquaculture areas; semantic segmentation; semi-/weakly-supervised learning; GAN; conditional adversarial learning

1. Introduction

Recent successful advances of deep learning make it become an increasingly popular choice in many fields of application. Following this wave of success and due to the increased availability of data and computational resources, the usage of deep learning in remote sensing is finally taking off in remote sensing as well. Coastal aquaculture areas, as a typical area for remote sensing, are vulnerable to storm-tide disasters, and are important for the government's scientific management and planning of aquaculture resources. In order to obtain the information of aquaculture areas, there are more and more researchers paying attention to using remote sensing technology and machine learning, and a series of research works has ensued [1–7]. At present, researchers use expert experience [8–10], characteristic

learning [11–14], threshold segmentation [15,16], and semantic segmentation networks [6] to extract aquaculture areas, and the practice has proven that these methods work well in this field. Reference [6] adopted a semantic segmentation network based on hybrid dilated convolution (HDC) [17] to extract aquaculture areas and summed up its four improvements compared to traditional machine learning: (1) the extraction results have clearer boundaries; (2) attenuation of the impact of sediments in seawater on the extraction results; (3) avoiding the influence of ships and other floatage; (4) avoiding the misidentification of the internal clearance of the cage culture area.

Although CNN-based approaches have achieved astonishing performance, they require an enormous amount of training data, and the robustness of the model is too poor to be applied to more scenarios. Different from image classification and object detection, semantic segmentation requires accurate per-pixel annotations for each of the training data, which can have a considerable expense and time. To ease the effort of acquiring high-quality labeled data, semi-supervised methods and weakly-supervised methods [18–25] have been applied to the task of semantic segmentation, which is significant for the application of deep learning in remote sensing. They are both incomplete supervised learning based on a small amount of labeled training samples, but weakly-supervised methods requires a lower quality of labeled training samples that cannot be exactly the same as the test and validation samples. Meanwhile, the emergence of the generative adversarial network (GAN) [26] makes semi-/weakly-supervised semantic segmentation more possible, and there are many semi-/weakly-supervised semantic segmentation networks based on GANs. The conditional GAN [27] has a further improvement based on GAN, which feeds y into both the discriminator and generator as an additional input layer, in order to make the generator able to generate samples related to y .

In this paper, we construct our network, the semi-/weakly-supervised semantic segmentation network (Semi-SSN) based on conditional generative adversarial nets (CGANs) and, through a self-training method, generate the pseudo-labels of unlabeled data by the generator to achieve semi-/weakly-supervised learning. Then, we employ Semi-SSN to extract aquaculture in a semi-supervised manner in GF-2 images, make comparative experiments with other state-of-the-art methods, and explore the scientific quality and practicability of our method based on an open remote sensing dataset. Besides, in remote sensing, there are many data sources with different spatial resolutions, and different spatial resolutions of remote sensing images serve different purposes. Ten meter level image remote sensing usually can be used to obtain information in a large area, because of the larger breadth, which is good for mapping the distribution map of coastal aquaculture areas. Meter level and sub-meter level remote sensing images are convenient to obtain the spatial distribution and capture more accurate information, so they are suitable for change detection such as disaster emergency response, production monitoring, etc. However, in practice, the resolution of the image to be processed can be inconsistent with that of the existing samples. Therefore, we employed Semi-SSN to extract aquaculture areas in a weakly-supervised manner with different spatial resolution remote sensing images. Taking the Fujian coastal area (mainly Sanduo) as the experimental area, we explore the application effect of Semi-SSN in different scenarios.

In short, we propose a novel method, Semi-SSN, based on conditional adversarial learning to extract aquaculture areas, in order to deal with the following problems: (1) the difficulty in labeling; (2) the poor robustness of the model; (3) the spatial resolution of the image to be processed is inconsistent with that of the existing samples. After comparing with other state-of-the-art methods and verifying on an open remote sensing dataset, we take the Fujian coastal area (mainly Sanduo) as an example and use our method to carry out disaster emergency response, production monitoring, and map making.

2. Materials and Methods

2.1. Related Work

In 2014, Goodfellow et al. [26] first proposed the generative adversarial network (GAN), which has been widely used in object detection [27], semantic segmentation [27,28], etc. The GAN (Figure 1) is composed of two neural networks: the generator G and the discriminator D . The generator trains with the objective to maximize the probability of the discriminator making mistakes, i.e., building a mapping function from a generator distribution p_g to the data space as $G(z; \theta_g)$ (where θ_g are the parameters of the generator). The discriminator $D(x; \theta_d)$ (where θ_d are the parameters of the discriminator) aims to estimate the probability that the sample came from the training data rather than the generator distribution p_g . Both networks are trained simultaneously with value function $V(G, D)$:

$$\min_G \max_D V(G, D) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] , \quad (1)$$

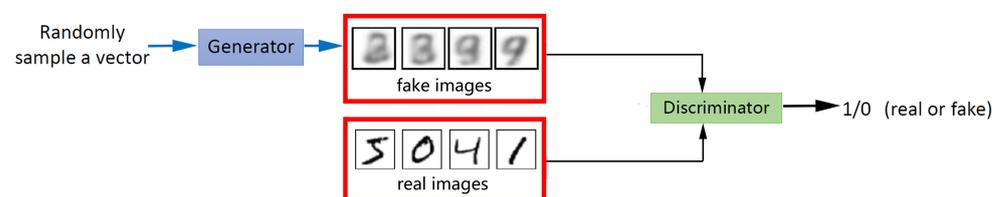


Figure 1. GAN architecture.

Based on its theoretical foundation, M. Mirza et al. [29] proposed conditional generative adversarial nets (CGANs) to make improvements (Figure 2), which makes the generator and discriminator conditioned on some extra information y . Both the prior noise distribution $p_z(z)$ and y are input into the generator, and in the discriminator, x and y are treated as inputs. The value function would be as follows:

$$\min_G \max_D V(G, D) = E_{x \sim p_{data}(x)} [\log D(x | y)] + E_{z \sim p_z(z)} [\log(1 - D(G(z | y)))] , \quad (2)$$

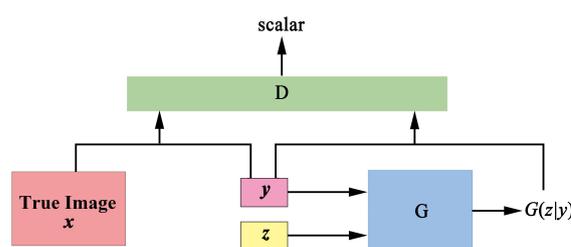


Figure 2. Conditional generative adversarial net (CGAN) architecture.

Then, the methodology of conditional adversarial learning was applied to a wide range of discrete labels [30,31], tackling prediction from a normal map [32], future frame prediction [33], semantic segmentation [34,35], and image generation from sparse annotations [36,37].

The semantic segmentation network [38–42] is a method of interpreting images at the pixel level, which requires an enormous amount of labeled data and has a considerable expense. Pinheiro and Collobert [18] and Pathak et al. [43] employed MIL to generate labels for supervised training. Hong et al. [44] used image level supervised images and a few fully annotated images to train their semantic segmentation network. In order to ease the effort of acquiring high-quality labeled data, semi-supervised semantic segmentation is imperative, and the emergence of GAN makes it more possible.

Recently, with the rise of the GAN and its improved network, adversarial learning has been widely used in semantic segmentation. Pauline et al. [28] trained a convolutional semantic segmentation network along with an adversarial network that discriminates segmentation maps coming either from the ground truth or from the segmentation network to detect and correct higher-order inconsistencies between ground truth and the map generated by the segmentation net. Huaqing Liu et al. [45] proposed semi-cGAN, based on CGAN, to segment lumbosacral structures on thin-layer computed tomography with a few labeled data. Souly et al. [46] leveraged a massive amount of available unlabeled or weakly labeled data and non-real images created through the GAN to achieve semi-supervised learning, and subsequently, Hung W C [47] made improvements based on it. Konstantinos et al. [48] adapted source domain images to the target domain based on a GAN-based method, which outperformed many unsupervised domain adaptation scenarios, and produced plausible samples. Xue et al. [49] proposed a novel semantic segmentation network, named SegAN, used a fully convolutional neural network as the segmenter, and adopted a multi-scale L1 loss function to train the critic and segmenter. Cherian et al. [50] presented a semantically-consistent GAN framework based on CycleGAN, dubbed Sem-GAN, which improved the quality of the translated images significantly.

Based on the methodology of CGAN and previous research, this paper proposes a novel network, Semi-SSN, that introduces conditional adversarial learning into the semantic segmentation network to realize semi-/weakly-supervised learning. We adopt the confidence maps generated by the discriminator and the predicted maps generated by the generator of the unlabeled data to produce the pseudo-labels for the model's training.

2.2. Network and Algorithm

The self-training method in semi-/weakly-supervised learning means that the classifier can be used to generate pseudo-labels of unlabeled data, after it is sufficiently trained on labeled data. If we take the confident predictions and assume that they are correct, we can add the unlabeled data with pseudo-labels into the training. If the noise in the pseudo-labels is sufficiently low, the model can benefit from the additional training data to obtain improved accuracy.

This paper proposed a self-training semi-supervised semantic segmentation method, which is divided into two processes: (1) using labeled data to train the classifier; (2) obtaining pseudo-labels of unlabeled data based on the classifier and then further training the classifier. At the same time, this paper introduced adversarial loss into the network, which not only improves the accuracy of semantic segmentation, but also reduces the noise of the pseudo-labels, thereby improving the accuracy of the entire model. The specific algorithm and network architecture are as follows.

2.2.1. Network Architecture

Based on the methodology of conditional adversarial learning, we propose a semi-/weakly-supervised semantic segmentation network (Semi-SSN), as shown in Figure 3. In this framework, we made the generator-classifier and discriminator of the GAN as a kind of semantic segmentation network; the generator-classifier, i.e., $S(\cdot)$, generates the prediction map of the labeled image X or unlabeled image \hat{X} , and the latter, i.e., $D(\cdot)$, takes $[X, S(X)]$, or $[X, Y]$, or $[\hat{X}, S(\hat{X})]$ as the input and outputs a confidence map, which infers the regions where the prediction results are close enough to the ground truth distribution.

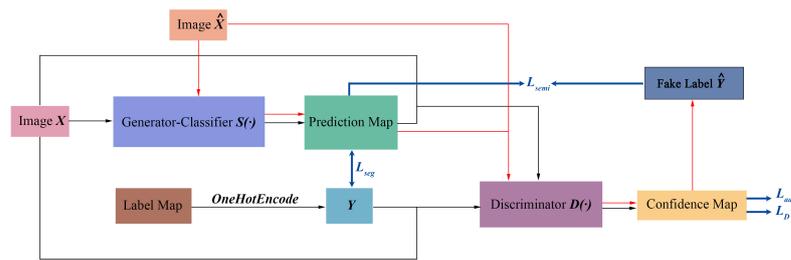


Figure 3. Semi-/weakly-supervised semantic segmentation network (Semi-SSN) architecture. The black workflow is the process of training labeled image X ; The red workflow is the process of training unlabeled image \hat{X} .

1. Generator-classifier:

According to the training tips proposed by DCGAN [51], we made some improvements to SegNet to obtain the generator-classifier, i.e., the baseline model in this paper:

- (1) Use Leaky-ReLU activation for all layers except for the output, which uses Softmax.
- (2) Replace deterministic spatial pooling functions (such as max pooling) with strided convolutions.
- (3) Replace upsampling with deconvolutions (the difference between the two methods is that the latter's parameters can be learned in the training process).
- (4) Use batch normalization on all layers except for the output layer.

2. Discriminator

We chose a simple dilation convolution network, the context network as discriminator (Table 1), and we used Leaky-ReLU activation for all layers.

Table 1. Context network. In this network, C is the category number.

Layer	1	2	3	4	5	6	7	8
Dilated rate	1	1	2	4	8	16	1	1
Kernel size	3×3	1×1						
Filter	$2C$	$2C$	$4C$	$8C$	$16C$	$32C$	$32C$	C

2.2.2. Algorithm

At the start, we trained our model based on labeled datasets. Given an input image X of size $H \times W \times channels$, where $channels$ is the number of bands, with its label map Y of size $H \times W \times C$, where C is the category number, we could obtain the predicted probability map $S(X)$ and confidence map $D(X, S(X))$ or $D(X, Y)$.

Firstly, we fixed the generator-classifier, i.e., $S(\cdot)$, and trained the discriminator, i.e., $D(\cdot)$, and we minimized the spatial cross-entropy loss L_D .

$$L_D = - \sum_{h,w} \log(1 - D(X, S(X)))^{(h,w)} + \log(D(X, Y))^{(h,w)}, \quad (3)$$

After training the discriminator k times, fix the discriminator, and update the generator-classifier by minimizing L_{class} , which is combined by the generator-classifier's cross-entropy loss L_{seg} and the adversarial loss L_{adv} .

$$L_{seg} = - \sum_{h,w} \sum_{c \in C} Y^{(h,w,c)} \log(S(X)^{(h,w,c)}), \quad (4)$$

$$L_{adv} = - \sum_{h,w} \log(D(X, S(X)))^{(h,w)}, \quad (5)$$

$$L_{class} = L_{seg} + \lambda_{adv} L_{adv}, \quad (6)$$

where $\lambda_{adv} \leq 1$ is a constant for balancing the multi-task training.

Iterate the above steps n times. After obtaining the trained generator-classifier and discriminator based on labeled data, we trained in a semi-/weakly-supervised manner the generator-classifier based on unlabeled data. Given an input image \hat{X} of size $H \times W \times C$, where $channels$ is the number of bands, we utilized the prediction map $S(\hat{X})$ and the confidence map $D(\hat{X}, S(\hat{X}))$ generated by the trained discriminator to generate the fake label map \hat{Y} .

$$\hat{Y} = \text{OneHotEncode}(I(D(\hat{X}, S(\hat{X})) > T_{semi}) \cdot \arg \max(S(\hat{X}))), \quad (7)$$

where the threshold, T_{semi} , is equal to the validation accuracy of the generator-classifier that was trained by L_{class} and labeled data and $I(\cdot)$ is the indicator function, $\text{OneHotEncode}(\cdot)$ indicates one-hot encoding of the vector, i.e., when $D(\hat{X}, S(\hat{X}))^{(h,w)}$ is greater than T_{semi} , $S(\hat{X})^{(h,w)}$ can be regarded as a true value of $\hat{X}^{(h,w)}$. The resulting semi-/weakly-supervised loss is defined by:

$$L_{semi} = - \sum_{h,w} \sum_{c \in C} \hat{Y}^{(h,w,c)} \log(S(\hat{X})^{(h,w,c)}), \quad (8)$$

Then, we trained the generator-classifier by minimizing $L_{semi-class}$, which combines semi-/weakly-supervised loss L_{semi} and the adversarial loss L_{adv} .

$$L_{semi-class} = \lambda_{semi} L_{semi} + \lambda_{adv} L_{adv}, \quad (9)$$

where $\lambda_{adv}, \lambda_{semi} \leq 1$ is a constant for balancing the multi-task training. And the specific details of training are in Algorithm 1.

Algorithm 1. Minibatch stochastic gradient descent training of our model. The number of steps to apply to the discriminator, k , is a hyperparameter. We used $k = 1$, the least expensive option, in our experiments.

For number of supervised training iterations **do**

For k steps **do**

- Select a minibatch with $\{x_1, x_2, \dots, x_m\}$ with ground truth $\{y_1, y_2, \dots, y_m\}$ from the labeled training set.
- Obtain the probability maps $\{S(x_1), S(x_2), \dots, S(x_m)\}$
- Update the discriminator by ascending its stochastic gradient:

$$\nabla_{\theta_d} \sum_{i=1}^m \sum_{h,w} \log(1 - D(x_i, S(x_i)))^{(h,w)} + \log(D(x_i, y_i))^{(h,w)}$$

End For

- Select a minibatch with $\{x_1, x_2, \dots, x_m\}$ with ground truth $\{y_1, y_2, \dots, y_m\}$ from labeled training set.
- Update the generator by ascending its stochastic gradient:

$$\nabla_{\theta_g} \sum_{i=1}^m \sum_{h,w} \left[\lambda_{adv} \log(D(x_i, S(x_i)))^{(h,w)} + \sum_{c \in C} y_i^{(h,w,c)} \log(S(x_i)^{(h,w,c)}) \right]$$

End For

For number of semi-/weakly-supervised training iterations **do**

- Select a minibatch with $\{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m\}$ from the unlabeled training set.
- Obtain the fake label maps $\{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_m\}$
 $\hat{y}_i = \text{OneHotEncode}(I(D(\hat{x}_i, S(\hat{x}_i)) > T_{semi}) \cdot \arg \max(S(\hat{x}_i)))$
- Update the generator by ascending its stochastic gradient:

$$\nabla_{\theta_g} \sum_{i=1}^m \sum_{h,w} \left[\lambda_{adv} \log(D(\hat{x}_i, S(\hat{x}_i)))^{(h,w)} + \lambda_{semi} \sum_{c \in C} \hat{y}_i^{(h,w,c)} \log(S(\hat{x}_i)^{(h,w,c)}) \right]$$

End For

2.3. Data Source

There are two mainly types of coastal aquaculture areas: cage culture area and raft culture area (Figure 4). Cage culture areas have small grid cages that can be clearly observed in high-resolution remote sensing images. They are mainly made of plastic material and appear as uneven rectangles on the image. The raft culture area is usually arranged together and appears as a dark and more uniform rectangle on the image, and there are small light points on the edge of the raft culture area in optical images.

This paper employed four GF-2 images, three GF-1(PMS) images, and three GF-1(WFV) images (Table 2) to explore the scientific quality and practicality of our method, Semi-SSN. The technical indicators of the GF-2, GF-1(PMS), and GF-1(WFV) images are as shown in Table 3.

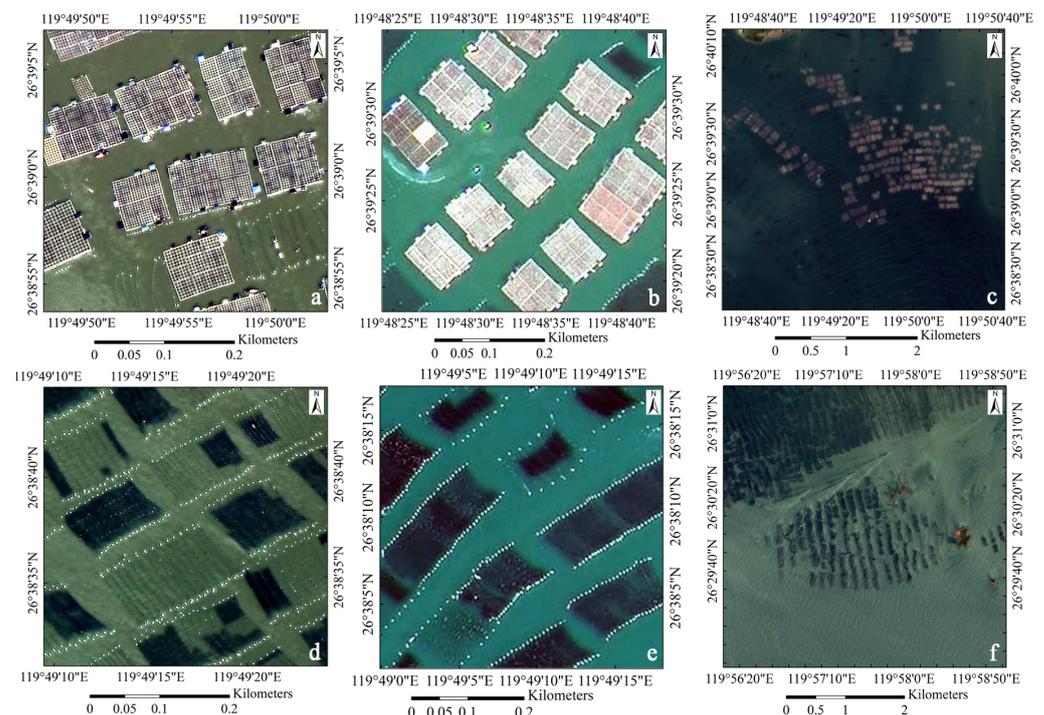


Figure 4. Two main types of aquaculture areas. (a) Cage culture area in a GF-2 image; (b) cage culture area in a GF-1(PMS) image; (c) cage culture area in a GF-1(WFV) image; (d) raft culture area in a GF-2 image; (e) raft culture area in a GF-1(PMS) image; (f) raft culture area in a GF-1(WFV) image.

Table 2. Data sources and their uses.

Data Source	Use	Date	Central Geographical Coordinates
GF-2	Sample Construction	30 March 2019	119°52'E 26°43'N
		30 March 2019	119°49'E 26°32'N
	Production Monitoring	22 June 2016 19 September 2016	119°54'E 26°42'N 119°48'E 26°42'N
GF-1(PMS)	Sample Construction	23 September 2019	119°56'E 26°49'N
	Disaster Emergency Response	8 April 2018	119°42'E 26°36'N
		19 September 2018	119°54'E 26°36'N
GF-1(WFV)	Sample Construction	28 January 2019	119°50'E 26°55'N
	Map making	15 May 2016	119°55'E 26°54'N
		15 July 2016	119°50'E 26°50'N

Table 3. The technical indicators of the GF-2, GF-1(PMS), and GF-1(WFV) images.

Spectral Range		Resolution		
		GF-2	GF-1(PMS)	GF-1(WFV)
Pan	0.45–0.90 μm	1 m	2 m	16 m
	0.45–0.52 μm	4 m	8 m	16 m
Multispectral	0.52–0.59 μm	4 m	8 m	16 m
	0.63–0.69 μm	4 m	8 m	16 m
	0.77–0.89 μm	4 m	8 m	16 m

3. Results

There were two experiments. In one experiment, we used Semi-SSN to extract coastal aquaculture areas in GF-2 images based on semi-supervised method with different labeled GF-2 data amounts. In the other experiment, we employed Semi-SSN to extract coastal aquaculture areas in higher (lower) spatial resolution remote sensing images based on weakly-supervised method with lower (higher) spatial resolution remote sensing images.

3.1. Training Objective

3.1.1. Sample Construction

In deep learning, the dataset mainly includes the training set, validation set, and test set. The training set is used to train the weight parameters of the model. Neither the validation set, nor the test set is involved in the training of the model. The former is used to adjust the hyperparameters of the model, preliminarily to evaluate the prediction ability of the model, and to prevent over-fitting during training. The latter is used to evaluate the robustness and generalization ability of the model.

According to Table 2, we chose two GF-2 images, one GF-1(PMS) image, and one GF-1(WFV) image for sample construction. We selected four areas from GF-2 images with a size of 5000×5000 pixels, an area from GF-1(PMS) images with a size of 2000×2000 pixels, and an area from GF-1(WFV) images with a size of 2000×2000 pixels for category marking. We randomly cut these images and labels for a size of 128×128 pixels. Then, according to Tables 4 and 5, we constructed training set and validation set, respectively.

Table 4. Division of samples in the semi-supervised experiment.

Labeled Data Amount	Training Set		Validation Set
	Labeled	Unlabeled	
Full	8000	-	2000
1/2	4000	4000	2000
1/4	2000	6000	2000
1/8	1000	7000	2000

Table 5. Division of samples in the weakly-supervised experiment.

Labeled Data/Unlabeled Data	Training Set		Validation Set
	Labeled	Unlabeled	
GF-2/GF-1(PMS)	1000	7000	2000
GF-2/GF-1(WFV)	1000	7000	2000
GF-1(PMS)/GF-2	1000	7000	2000
GF-1(PMS)/GF-1(WFV)	1000	7000	2000
GF-1(WFV)/GF-2	1000	7000	2000
GF-1(WFV)/GF-1(PMS)	1000	7000	2000

Then, we chose a test area in the GF-2 image with a size of 2000×2000 pixels, a test area in the GF-1(PMS) image with a size of 1000×1000 pixels, and a test area in the GF-1(WFV) image with a size of 1000×1000 pixels, as shown in Figure 5, to further verify the effect of the model.

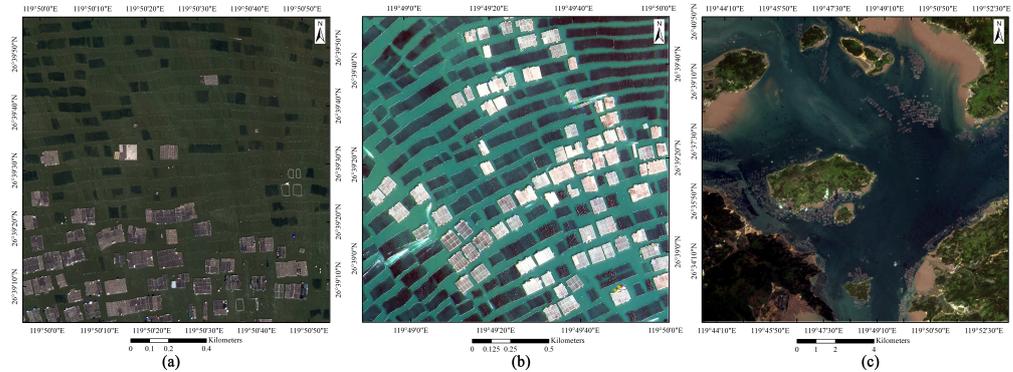


Figure 5. Test image. (a) GF-2 test image; (b) GF-1 (PMS) test image; (c) GF-1 (WFV) test image.

3.1.2. Implementation Details

We used the mean intersection over union (*mIoU*) as the evaluation function, as Equation (11).

$$IoU = \frac{TP}{TP + TN + FN}, \quad (10)$$

$$mIoU = \frac{1}{N} \sum_{i=1}^N IoU_i, \quad (11)$$

where *TP* is the pixel number that is correctly extracted as aquaculture areas, *FP* is pixel number that is mistakenly extracted as aquaculture areas, and *FN* is the pixel number that is not extracted, but is aquaculture areas in reality. *N* is the number of categories. IoU_i is the intersection over union of the *i*th class.

We implemented our network using the Keras and TensorFlow framework. We trained our network on eight GPUs with 12 GB memory. We used Adam [52] as the optimizer, where the weight decay was 10^{-4} and the initial learning rate was set as 10^{-7} . In our experiment, T_{semi} is a dynamic value that changes with the validation accuracy (i.e., *mIoU* on the validation set) of each iteration, as Equation (12). λ_{semi} is determined according to the labeled data amount, as Equation (13). We set λ_{adv} as 0.04 (the detail is given in Section 4.1).

$$T_{semi} = 0.6 \times validation_accuracy, \quad (12)$$

$$\lambda_{semi} = labeled_data_amount, \quad (13)$$

3.2. Semi-Supervised Experiment

In the experiment, we randomly divided the GF-2 sample as Table 4, calculated the *mIoU* on the validation set and test set (Table 6), and obtained the extracted result of the test area (Figure 6).

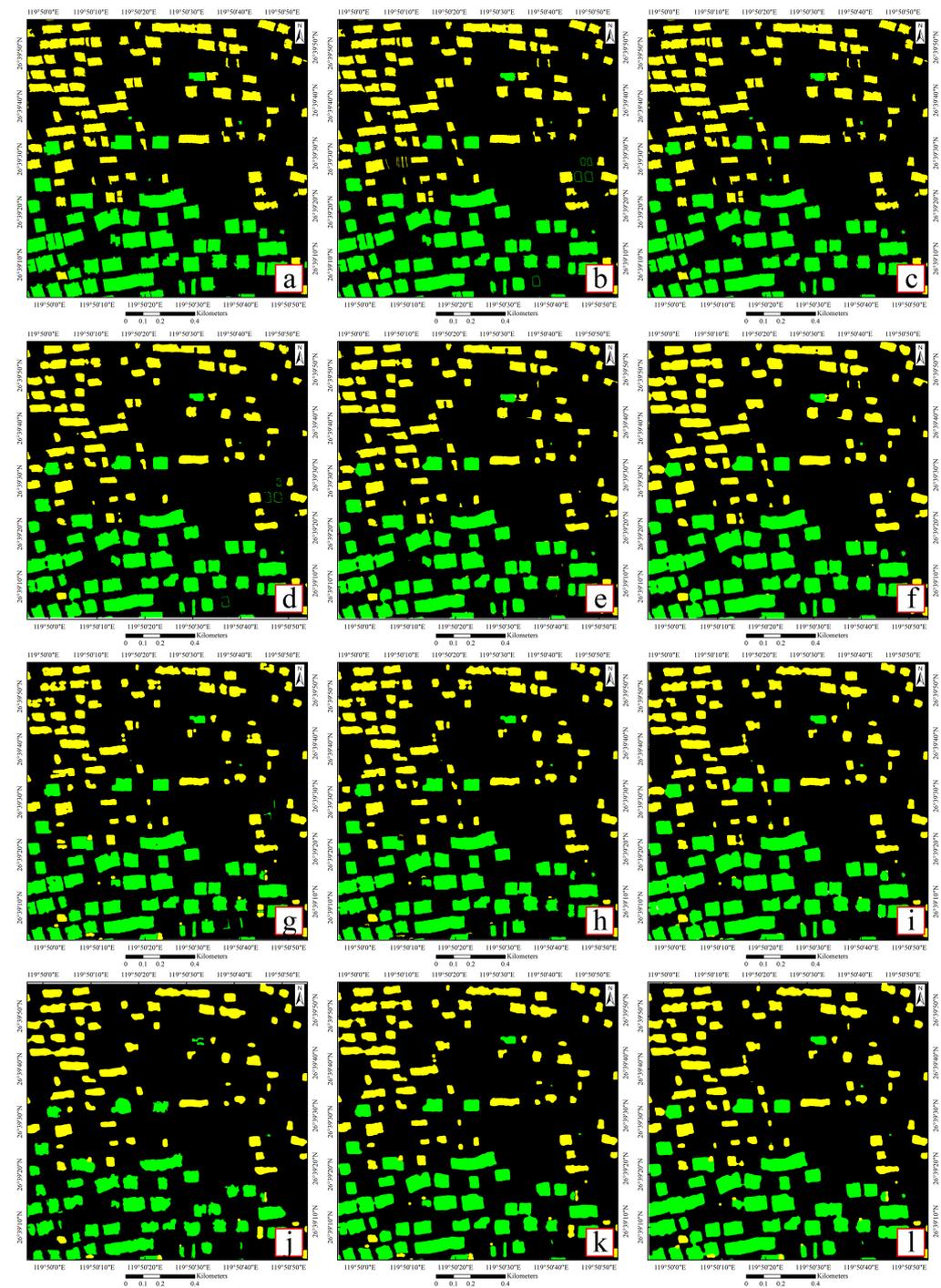


Figure 6. Extraction result, where green stands for cage culture areas, yellow stands for raft culture areas, and black for background. (a) Ground truth; (b) labeled data amount: full, model: baseline; (c) labeled data amount: full, model: baseline + L_{adv} ; (d) labeled data amount: 1/2, model: baseline; (e) labeled data amount: 1/2, model: baseline + L_{adv} ; (f) labeled data amount: 1/2, model: baseline + L_{adv} + L_{semi} ; (g) labeled data amount: 1/4, model: baseline; (h) labeled data amount: 1/4, model: baseline + L_{adv} ; (i) labeled data amount: 1/4, model: baseline + L_{adv} + L_{semi} ; (j) labeled data amount: 1/8, model: baseline; (k) Labeled data amount: 1/8, model: baseline + L_{adv} ; (l) labeled data amount: 1/8, model: baseline + L_{adv} + L_{semi} .

Table 6. *mIoU* on the validation set and test set.

	Methods	Labeled Data Amount			
		1/8	1/4	1/2	Full
Validation Accuracy	baseline	0.7047	0.7434	0.7937	0.8530
	baseline + L_{adv}	0.7512	0.7842	0.8253	0.8725
	baseline + L_{adv} + L_{semi}	0.7961	0.8103	0.8417	N/A
Test Accuracy	baseline	0.6842	0.7618	0.8036	0.8847
	baseline + L_{adv}	0.7635	0.7876	0.8309	0.9058
	baseline + L_{adv} + L_{semi}	0.8053	0.8244	0.8610	N/A

3.3. Weakly-Supervised Experiment

Another currently hot issue is how to use a small amount of existing labeled samples to extract aquaculture areas in different spatial resolution remote sensing images and comprehensively analyze practical problems according to the data from various sources based on the weakly-supervised semantic segmentation method.

In this experiment, we used Semi-SSN to extract coastal aquaculture areas in lower (higher) spatial resolution images based on weakly-supervised method with a small amount of labeled higher (lower) spatial resolution data with unlabeled lower (higher) data (Table 5), obtained the extraction results of unlabeled data source (Figure 7), and calculated *mIoU* on the validation set and test set (Table 7).

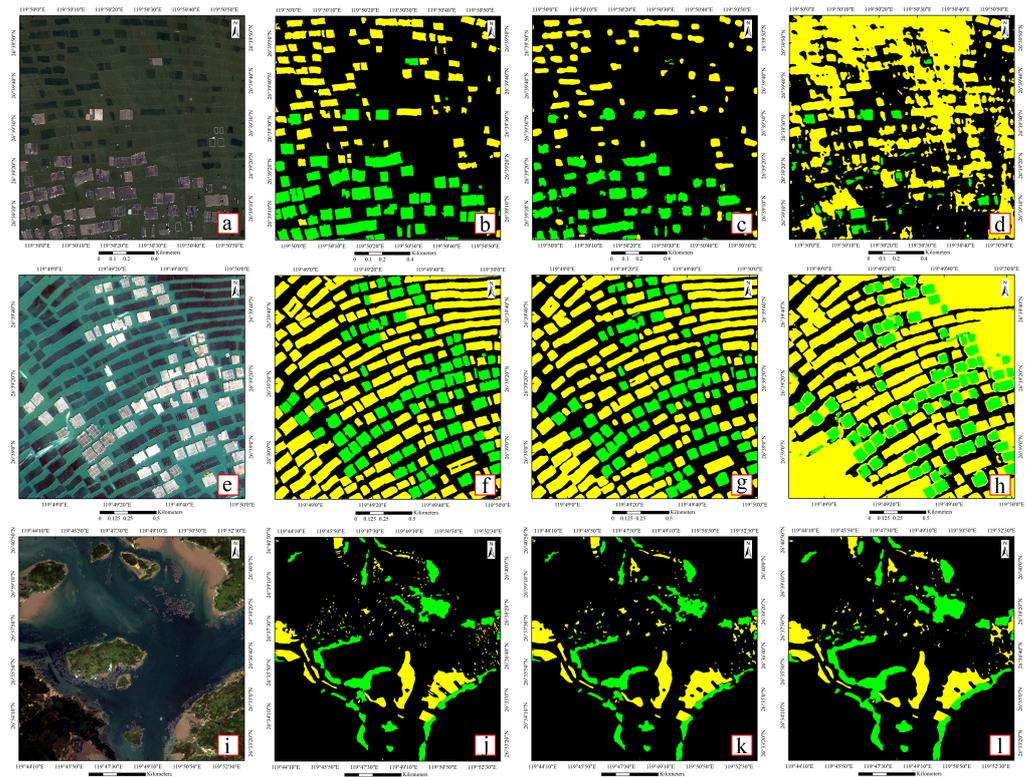


Figure 7. Extraction result based on the weakly-supervised method. (a) GF-2 Image; (b) ground truth of GF-2 image; (c) extraction result of GF-2 image based on labeled GF-1 (PMS) data; (d) extraction result of GF-2 image based on labeled GF-1 (WFV) data; (e) GF-1 (PMS) image; (f) ground truth of GF-1 (PMS) image; (g) extraction result of GF-1 (PMS) image based on labeled GF-2 data; (h) extraction result of GF-1 (PMS) image based on labeled GF-1 (WFV) data; (i) GF-1 (WFV) image; (j) ground truth of GF-1 (WFV) image; (k) extraction result of GF-1 (WFV) image based on labeled GF-2 data; (l) extraction result of GF-1 (WFV) image based on labeled GF-1 (PMS) data.

Table 7. *mIoU* on the validation set and test set.

Unlabeled Data Source	Labeled Data Source	Validation Accuracy	Test Accuracy
GF-2	GF-1(WFV)	0.4418	0.3259
GF-2	GF-1(PMS)	0.6976	0.6368
GF-1(PMS)	GF-1(WFV)	0.6754	0.6309
GF-1(PMS)	GF-2	0.7977	0.8198
GF-1(WFV)	GF-1(PMS)	0.8297	0.8329
GF-1(WFV)	GF-2	0.8426	0.8646

4. Discussion

4.1. Hyperparameter Selection

The hyperparameters are mainly determined according to the performance of the model in the validation set. There are three hyperparameters: λ_{adv} and λ_{semi} are used to balance the multi-task learning, and T_{semi} is used to control the sensitivity in the semi-/weakly-supervised learning described in Equation (7). In our experiment, T_{semi} is a dynamic value that changes with the validation accuracy of each iteration, as Equation (12). λ_{semi} is determined according to the labeled data amount, as Equation (13). We used GF-2 labeled data to evaluate the effect on λ_{adv} , as shown in Table 8, and chose final values as 0.04.

Table 8. The effect on λ_{adv} .

λ_{adv}	0	0.02	0.035	0.04	0.045	0.05	0.06
Validation Accuracy	0.8530	0.8613	0.8683	0.8725	0.8652	0.8612	0.8566

4.2. Analysis of the Results

4.2.1. Semi-Supervised Results

It can be seen (Table 6) that the adversarial loss and unlabeled data could improve the *mIoU* to different degrees; the former can help improve the validation accuracy by 1.9–4.7% and improve the test accuracy by 2.1–8.0%, and the latter can help improve the validation accuracy by 4.8–9.2% and improve the test accuracy by 5.7–12.2%, compared to the baselines. The extraction results have a clear boundary with few fragments (Figure 6), especially the addition of adversarial loss, which makes the model more sensitive to structure information, which greatly avoids the misidentification of floatage on the water surface. In order to facilitate observation, we zoom in on local details (Figure 8). It can be seen that Semi-SSN is convenient to filter out some impurities such as floatage on the water surface after adding adversarial loss.

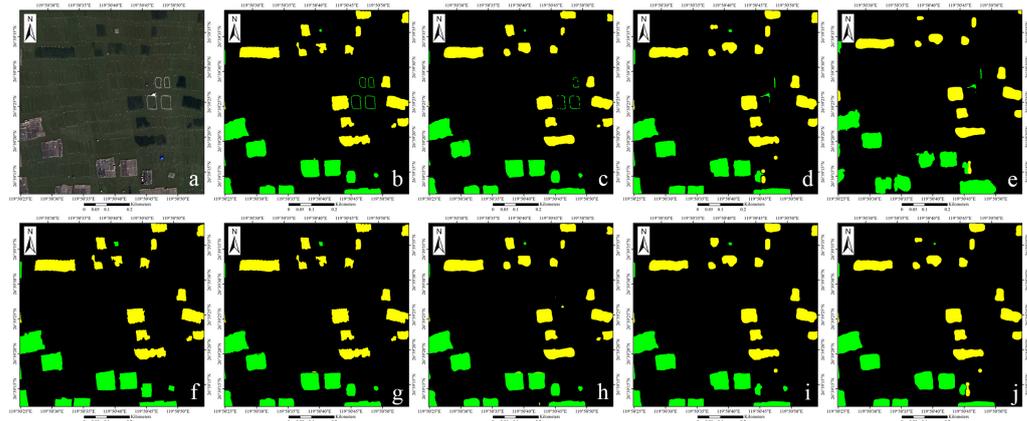


Figure 8. The details of the extraction result. (a) image; (b) labeled data amount: full, model: baseline; (c) labeled data amount: 1/2, model: baseline; (d) labeled data amount: 1/4, model: baseline; (e) labeled data amount: 1/8, model: baseline; (f) ground truth; (g) labeled data amount: full, model: baseline + λ_{adv} ; (h) labeled data amount: 1/2, model: baseline + λ_{adv} ; (i) labeled data amount: 1/4, model: baseline + λ_{adv} ; (j) labeled data amount: 1/8, model: baseline + λ_{adv} .

4.2.2. Weakly-Supervised Results

From the overall point of view, using the higher spatial resolution labeled data to extract the coastal aquaculture areas in the lower spatial resolution image, both validation accuracy and test accuracy can reach about 80%. The extraction result (Figure 7g,k,l) has a clear boundary, which can be directly put into practical applications.

While using lower spatial resolution labeled data to extract aquaculture areas in higher spatial resolution images, the situation is relatively complicated, and the common point is that the accuracy is not enough to meet the needs of practical applications. Using labeled GF-1(PMS) data to achieve extraction in a GF-2 image cannot extract the aquaculture areas precisely (Figure 7c), but the extraction result can simply reflect the spatial distribution of the aquaculture areas, which means it can be used for rough analysis and extraction. However, using labeled GF-1(WFV) data to extract the aquaculture area in a GF-1(PMS) or GF-2 image (Figure 7d,h), the raft culture area is easily confused with seawater, and the cage area culture cannot be extracted completely.

In short, Semi-SSN is conducive to using the existing labeled data to extract aquaculture areas in different spatial resolution remote sensing images and can help reduce the time cost of labeling. According to our method, labeled higher spatial resolution samples can be used to extract aquaculture areas in lower spatial resolution images, and partially labeled lower spatial resolution data (GF-1(PMS)) can be conducive to roughly extracting the distribution of aquaculture areas in higher spatial resolution images (GF-2). However, in general, using labeled lower spatial resolution samples to extract aquaculture areas in higher spatial resolution images cannot obtain ideal results.

4.3. Algorithm Validation

4.3.1. Comparison with Other Methods

We made comparative experiments with FCN8s, UNet, SegNet, and HDCUNet [6] based on 8000 labeled GF-2 samples (Table 9). Note that the validation accuracy and test accuracy of our baseline model are slightly inferior to HDCUNet, but they are effectively improved with the addition of the adversarial loss. Besides, it can be seen (Tables 6 and 9) that our method can reach the approximate effect of FCN8s, UNet, and SegNet based on a smaller amount (1/2, 1/4, and 1/2) of labeled samples, after adding L_{adv} and L_{semi} to our baseline.

Table 9. *mIoU* of five methods.

Method	FCN8s	UNet	SegNet	HDCUNet	Baseline	Baseline + L_{adv}
Validation Accuracy	0.8286	0.7910	0.8347	0.8693	0.8530	0.8725
Test Accuracy	0.8451	0.8249	0.8607	0.8926	0.8847	0.9058

4.3.2. Algorithm Validation on an Open Dataset

Tong et al. [53] constructed a large-scale land cover dataset with GaoFen-2 (GF-2) satellite images, named GID, and it contained two sub-datasets: a large-scale classification set (LCS) and a fine land cover classification set (FLCS), which is provided online at <http://captain.whu.edu.cn/GID/> (accessed on 1 November 2020). The LCS contains 150 pixel level annotated GF-2 images (the training set contains 120 images; the validation set contains 30 images), and the FLCS is composed of 30,000 multi-scale image patches (training set) coupled with 10 pixel level annotated GF-2 images (validation set).

In this paper, we pretrained our baseline based on LCS and generated coarse pixel level labels based on the patch level labels of the FLCS training set. Then, we used the FLCS to explore the effectiveness of our method, trained Semi-SSN based on labeled samples and unlabeled samples with different ratios, and made a comparative experiment with Tong et al. [53] Given that Tang et al. [53] used overall accuracy (OA) as the evaluation criterion (as Equation (14)), we made a comparison based on this (Table 10).

$$OA = \frac{TP + TN}{TP + FP + TN + FN'} \quad (14)$$

Table 10. Result on the validation set of the fine land cover classification set (FLCS).

Methods	Labeled Data Amount (OA)			
	1/8	1/4	1/2	Full
Tong et al. [53]	N/A	N/A	N/A	0.7004
baseline	0.4609	0.5233	0.6298	0.6962
baseline+ L_{adv}	0.5051	0.5709	0.6523	0.7191
baseline+ L_{adv} + L_{semi}	0.5652	0.6619	0.6913	N/A

Table 10 shows the evaluation results on the FLCS dataset; the *OA* of our baseline model trained fully labeled data can reach above 69%, and *OA* can exceed that in Tong et al. [53] after adding L_{adv} . The adversarial loss brings consistent performance improvement (2.2–4.8%) over different amounts of training data, and incorporating the proposed semi-supervised learning scheme brings overall a 6.1–13.8% improvement.

4.4. Application and Analysis

4.4.1. Disaster Emergency Response

On 11 July 2018, typhoon “Maria”, the eighth super typhoon, landed on Huangqi Peninsula, Lianjiang county, Fujian Province, with a local maximum wind level of 14, accompanied by heavy rain, and Sanduo suffered the most serious damage. Based on two GF-1(PMS) images, the disaster situation of the area is discussed as shown in Figure 9a,b.

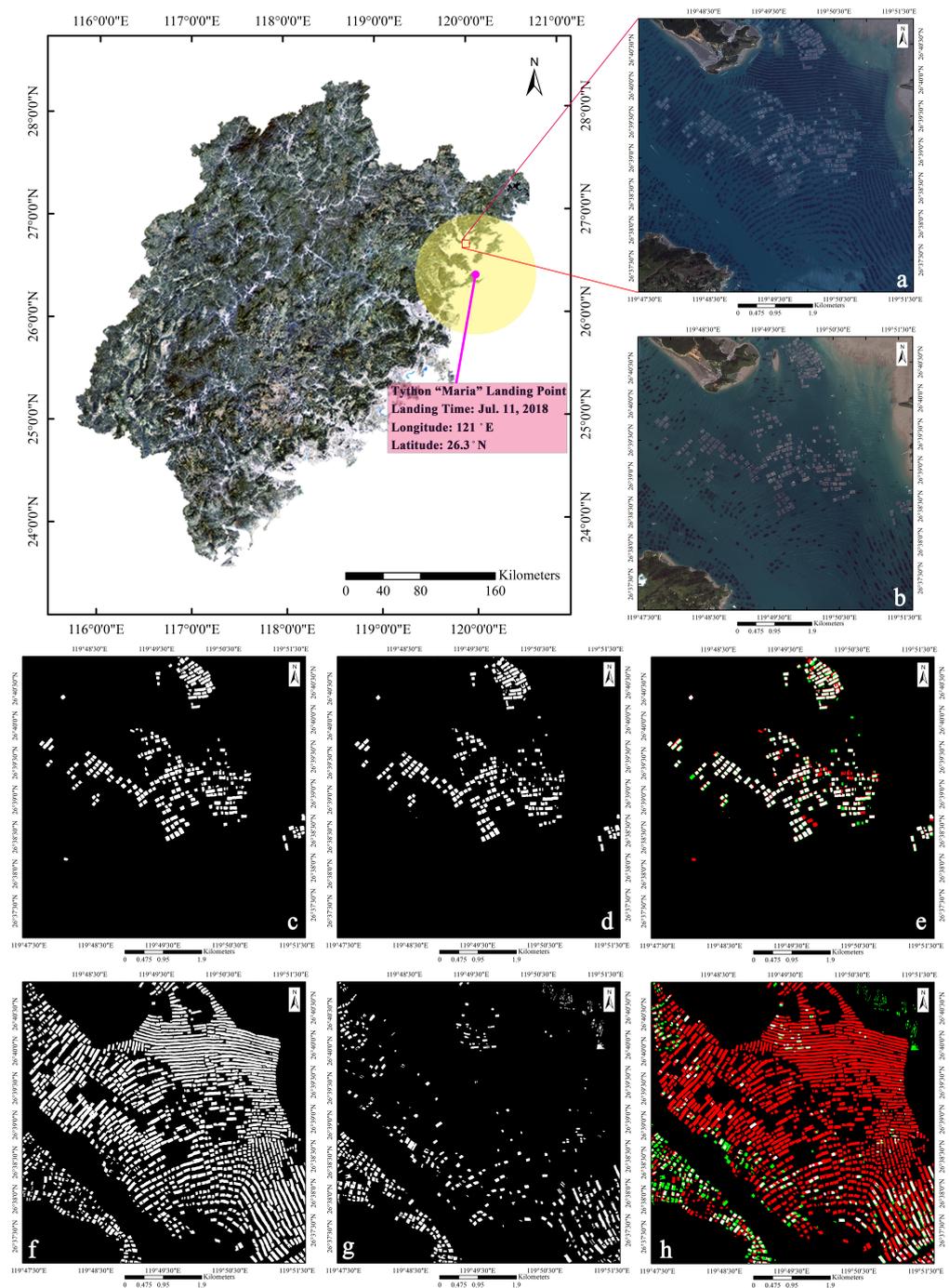


Figure 9. Change detection before and after typhoon “Maria”, where white stands for coastal aquaculture areas, red stands for the reduction of aquaculture areas, and green stands for the increase of aquaculture areas. (a) GF-1 (PMS) image imaged on 8 April 2018; (b) GF-1 (PMS) image imaged on 19 September 2018; (c) cage culture area before typhoon “Maria”; (d) cage culture area after typhoon “Maria”; (e) change area of the cage culture area; (f) raft culture area before typhoon “Maria”; (g) raft culture area after typhoon “Maria”; (h) change area of the raft culture area.

We used Semi-SSN to extract coastal aquaculture areas based on the existing labeled GF-2 samples and unlabeled samples collected from these two images (Figure 9a,b) and detected the changes of the intersecting parts (3500×3500 pixels).

We found that typhoon “Maria” had a relatively small impact on the cage culture area (Figure 9e), while the raft culture area (Figure 9h) had a large area loss. Besides, for the incremental part after the typhoon, the patch-like increase was likely caused by the broken

aquaculture areas being blown away by the typhoon, and the more complete increase may be the newly added aquaculture areas for subsequent reconstruction after the disaster. According to the statistics, the raft culture area after the typhoon was reduced by about 174,664 m², and about 16,388 m² were added; and the cage culture area was reduced by about 5992 m² and about 4108 m² added.

What is more, from 11 to 13 July 2018, we followed the field investigation team from the National Marine Hazard Mitigation Service (NMHMS) who went to the Fujian coastal areas to investigate the impact and destruction of the disaster. As shown in Figure 10, the raft culture area in Sanduo was seriously damaged, and the aquaculture facilities were destroyed and scattered by storm surges and coastal waves, which further supports the results of this experiment. Through field investigation, the main reason for the loss of the raft culture area was that it was easily pushed away by storm surge and nearshore waves, which usually causes a large area of damage. However, for the cage culture area, the construction was relatively stable, which was not easily destroyed; therefore, the specific damage to the species cultured in the cage culture area caused by the typhoon is not easy to directly measure through image interpretation.



Figure 10. Field investigation of disaster situation. (a) Cage culture area; (b) raft culture area.

4.4.2. Production Monitoring

Aquaculture production activities are mainly carried out from May to September every year. The number and density of aquaculture areas are very important for production quality and water environment, so it is necessary to monitor the production activities. After an investigation, there was no super typhoon passing through the area from 22 June to 19 September in 2016, so we used the intersection region (10,000 × 10,000 pixels) of these two GF-2 images (Figure 11a,b) to discuss the area change of the two kinds of aquaculture areas in the peak season.

We used Semi-SSN to extract coastal aquaculture areas based on the existing labeled GF-2 samples and unlabeled samples collected from these two images (Figure 11a,b) and compared the changes of the intersecting parts.

It can be seen intuitively that the increase and decrease of the raft culture area (Figure 11h) is obvious, but the change of the cage culture area (Figure 11e) is not obvious. Through the investigation, the cage culture area can be used for two to three years, in general, and does not need to be redeployed every year, so it has a lower probability of a large number of changes. From the extraction results, it is found that the cage culture area did not change significantly in the peak season of 2016, which shows that this year is very likely not the year when the cage culture area needed to be replaced and redeployed, while the raft culture area needs to be redeployed every year, and its significant increase or decrease indicates that this kind of aquaculture had been actively carried out during this period. However, at the same time, the density of the raft culture area increased significantly, which is very important to the yield and biological environment. Therefore, it is necessary to monitor and control it in time. According to the statistics, in the peak

season, the raft culture area was reduced by about 50,806 m² and 72,485 m² added, and the cage culture area was reduced by about 5045 m² and about 4733 m² added.

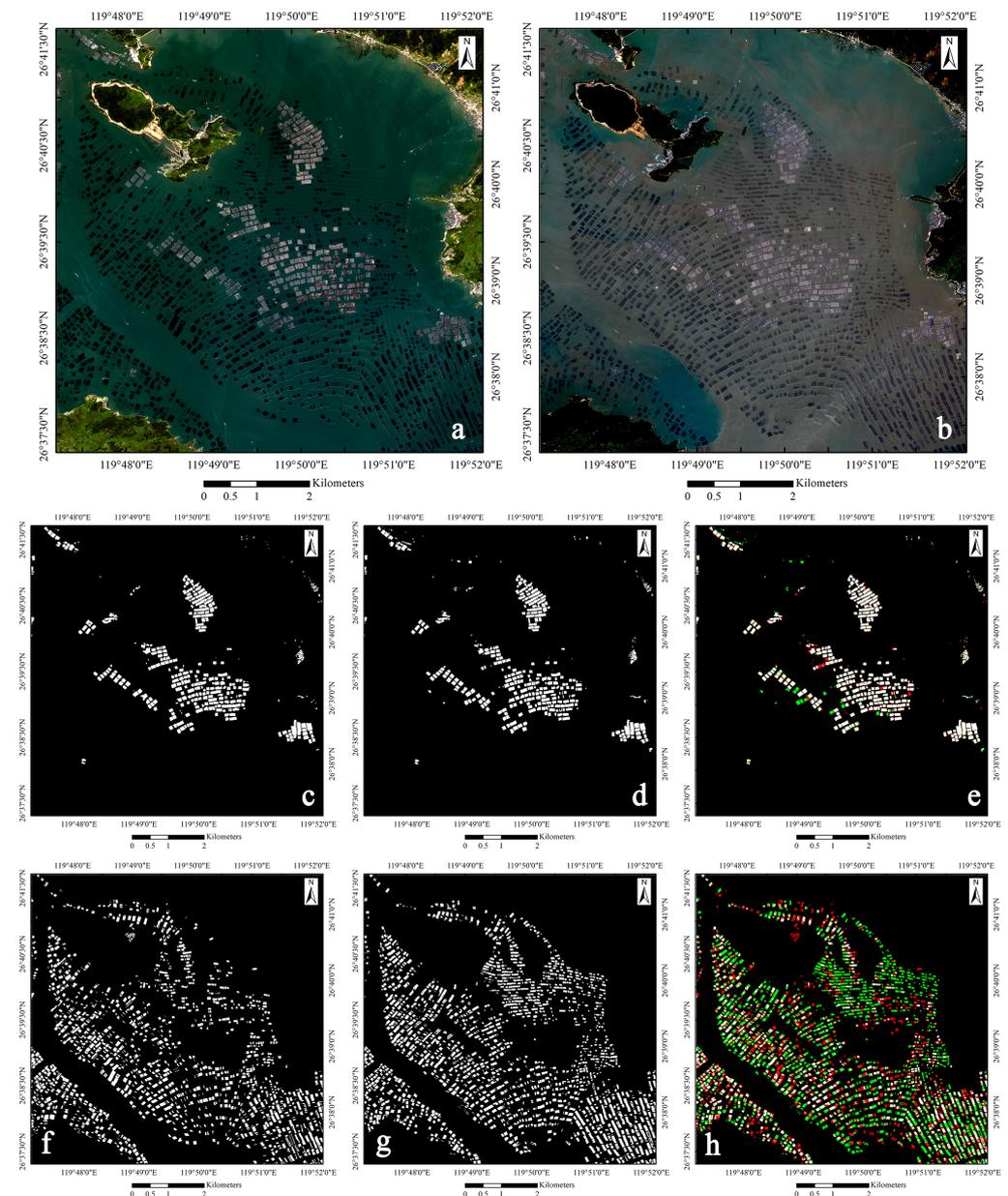


Figure 11. Production monitoring, where white stands for coastal aquaculture areas, red stands for the reduction of aquaculture areas, and green stands for the increase of aquaculture areas. (a) GF-2 image imaged on 22 June 2016; (b) GF-2 image imaged on 19 September 2016; (c) cage culture area on 22 June 2016; (d) cage culture area on 19 September 2016; (e) change area of the cage culture area; (f) Raft culture area on June 22, 2016; (g) raft culture area on 19 September 2016; (h) change area of the raft culture area.

4.4.3. Map Making

We chose two GF-1(WFV) images covering Sanduo and carried out image fusion to avoid the interference of clouds; the detail information is shown in Table 2. Then, we used Semi-SSN to extract coastal aquaculture areas based on the existing labeled GF-2 samples and unlabeled samples collected from these two images and made the distribution map of coastal aquaculture areas in Sanduo, as shown in Figure 12. According to the statistics, in 2016, the raft culture area was 911,872 m², and the cage culture area was 810,496 m².

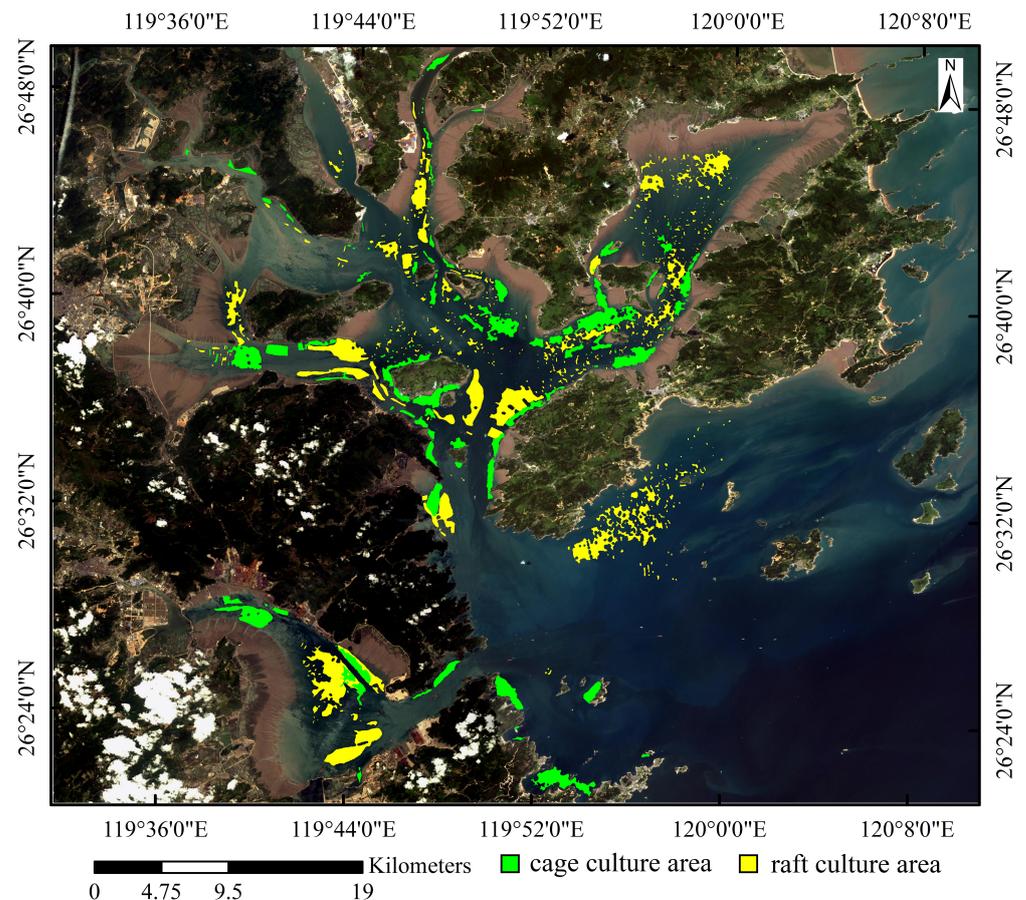


Figure 12. The distribution map of the coastal aquaculture areas.

5. Conclusions

In this work, we proposed a semi-/weakly-supervised semantic segmentation network (Semi-SSN) based on conditional adversarial learning for extracting aquaculture areas, and after experiments and analysis, we drew the following conclusions:

1. For semi-supervised extraction in GF-2 images, both adversarial loss and unlabeled samples are conducive to improving the validation accuracy and test accuracy, and especially the former can make the model more sensitive to structure information;
2. For multi-scale spatial resolution remote sensing images, labeled higher spatial resolution samples are conducive to extracting aquaculture areas in lower spatial resolution images, but not vice versa;
3. Through experiments, Semi-SSN can reach the approximate effect of other state-of-the-art methods based on relatively fewer labeled samples and adversarial loss and perform better in an open remote sensing dataset (FLCS) than Tong et al.'s method [53];
4. Applying Semi-SSN to detect the change before and after typhoon "Maria", the raft culture area was more vulnerable than the cage culture area in this disaster.
5. Employing Semi-SSN to monitor the production, in the peak season (2016), the distribution density of the raft culture area increased significantly during this period, while the cage culture area did not change significantly.
6. Making the distribution map of coastal aquaculture areas in 2016, the raft culture area was 911,872 m², and the cage culture area was 810,496 m².

In short, Semi-SSN is convenient for practical application and provides a new paradigm for solving the following problems: (1) the difficulty in labeling; (2) the poor robustness of the model; (3) the spatial resolution of the image to be processed is inconsistent with that of the existing samples. In the future, we will be devoted to improving the transfer capacity and robustness of our model, which is the general dilemma of deep learning models in

dealing with remote sensing problems, and further explore the performance of our model on different spectral resolutions, radiometric resolutions, radar data, and different remote sensing tasks, to make it more suitable for the actual needs.

Author Contributions: Conceptualization, B.C. and C.L.; methodology, C.L. and B.X.; validation, C.L., B.X., and C.H.; formal analysis, B.C., B.X., and C.L.; investigation, C.L., X.L., and C.H.; resources, B.C.; data curation, C.L., X.L., N.J., C.H., and J.C.; writing—original draft preparation, C.L.; writing—review and editing, B.C. and B.X.; visualization, C.L. and C.H.; supervision, B.C. and B.X.; project administration, B.C.; funding acquisition, B.C. All authors read and agreed to the published version of the manuscript.

Funding: National Natural Science Foundation of China (61731022, 61531019) project.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This work was partially funded by the National Natural Science Foundation of China (61731022, 61531019) and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant Numbers XDA19010401). The authors would also like to thank the National Marine Hazard Mitigation Service (NMHMS) for supporting our research work.

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

References

1. Xinguo, L.; Nan, J.; Yingbao, Y. Remote Sensing Investigation and Survey of Lake Reclamation and Enclosure Aquaculture in Lake Taihu. *Trans. Oceanol. Limnol.* **2006**, *1*, 93–99.
2. Wang, J.; Gao, J. Extraction of Enclosure Culture in Gehu Lake Based on Correspondence Analysis. *J. Remote Sens.* **2008**, *12*, 716–723.
3. Zhou, X.C.; Wang, X.Q.; Xiang, T.L.; Jiang, H. Method of Automatic Extraction Seaside Aquaculture Land Based on ASTER Remote Sensing Image. *Wetl. Sci.* **2006**, *4*, 64–68.
4. Xie, Y.; Wang, M.; Zhang, X. An Object-oriented Approach for Extracting Farm Waters within Coastal Belts. *Remote Sens. Technol. Appl.* **2009**, *24*, 68–72.
5. Jialan, C.; Dongzhi, Z.; Fengshou, Z. Wakame Raft Interpretation Method of Remote Sensing based on Association Rules. *Remote Sens. Technol. Appl.* **2012**, *27*, 941–946.
6. Cheng, B.; Liang, C.; Liu, X.; Liu, Y.; Ma, X.; Wang, G. Research on a novel extraction method using Deep Learning based on GF-2 images for aquaculture areas. *Int. J. Remote Sens.* **2020**, *41*, 3575–3591. [[CrossRef](#)]
7. Chhetri, M.; Kumar, S.; Pratim, R. P.; Kim, B.G. Deep BLSTM-GRU Model for Monthly Rainfall Prediction: A Case Study of Simtokha, Bhutan. *Remote Sens.* **2020**, *12*, 3174. [[CrossRef](#)]
8. Wu, Y.; Zhang, J.; Tian, G.; Cai, D.; Liu, S. A Survey to Aquaculture with Remote Sensing Technology in Hainan Province. *Chin. J. Trop. Crops* **2006**, *27*, 108–111.
9. Lin, Q.; Lin, G.; Chen, Z.; Chen, Y. The Analysis on Spatial-temporal Evolution of Beach Cultivation and Its Policy Driving in Xiamen in Recent Two Decades. *Geo-Inf. Sci.* **2007**, *9*, 9–13.
10. Lin, G.L.; Zheng, C.Z.; Cai, F.; Huang, F.M. Application of Remote Sensing Technique in Marine Delimiting. *J. Ocean. Taiwan Strait* **2004**, *23*, 219–224.
11. Zhu, C.; Luo, J.; Shen, Z.; Li, J.; Hu, X. Extracted Enclosure Culture in Coastal Waters based on High Spatial Resolution Remote Sensing Image. *J. Dalian Marit. Univ.* **2011**, *37*, 66–69.
12. Li, J.; He, L.; Dai, J.; Li, J. Extract Enclosure Culture in Lakes based on Remote Sensing Image Texture Information. *J. Lake Sci.* **2006**, *18*, 337–342.
13. Peng, L.; Yunyan, D. A CBR Approach for Extracting Coastal Aquaculture areas. *Remote Sens. Technol. Appl.* **2012**, *27*, 857–8642.
14. Ma, Y.; Zhao, D.; Wang, R.; Su, W. Offshore aquatic farming areas extraction method based on ASTER data. *Trans. CASE* **2010**, *26*, 120–124.
15. Lu, Y.; Li, Q.; Wang, H.; Du, X.; Liu, J. A Method of Coastal aquaculture areas Automatic Extraction with High Spatial Resolution Images. *Remote Sens. Technol. Appl.* **2015**, *30*, 486–494.
16. Jian, F.; Hai, H.; Hui, F. Extracting aquaculture areas with RADARSAT-1. *Mar. Sci.* **2005**, *10*, 46–49.
17. Wang, P.; Chen, P.; Yuan, Y.; Liu, D.; Huang, Z.; Hou, X.; Cottrell, G. Understanding convolution for semantic segmentation. In Proceedings of the 2018 IEEE winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 1451–1460
18. Pinheiro, P. O.; Collobert, R. Weakly supervised semantic segmentation with convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; Volume 2, p. 6.

19. George, P.; Liang-Chieh, C.; Kevin, M.; Alan, L.Y. Weakly-and semisupervised learning of a dcnn for semantic image segmentation. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
20. Seunghoon, H.; Hyeonwoo, N.; Bohyung, H. Decoupled deep neural network for semisupervised semantic segmentation. In Proceedings of the Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 7–10 December 2015.
21. Qi, X.; Liu, Z.; Shi, J.; Zhao, H.; Jia, J. Augmented feedback in semantic segmentation under image level supervision. In Proceedings of the 14th European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016.
22. Deepak, P.; Philipp, K.; Trevor, D. Constrained convolutional neural networks for weakly supervised segmentation. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
23. Dai, J.; He, K.; Sun, J. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
24. Seunghoon H.; Donghun, Y.; Suha, K.; Honglak, L.; Bohyung, H. Weakly supervised semantic segmentation using web-crawled videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
25. Amy, B.; Olga, R.; Vittorio, F.; Li, F. What’s the point: Semantic segmentation with point supervision. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016.
26. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; David W.; Sherjil, O.; Aaron, C.; Yoshua, B. Generative adversarial nets. In Proceedings of the 28th Annual Conference on Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 8–13 December 2014.
27. Xiaolong, W.; Abhinav, S.; Abhinav, G. A-fast-rcnn: Hard positive generation via adversary for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
28. Luc, P.; Couprie, C.; Chintala, S.; Verbeek, J. Semantic segmentation using adversarial networks. *arXiv* **2016**, arXiv:1611.08408.
29. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
30. Denton, E.; Chintala, S.; Szlam, A.; Fergus, R. Deep generative image models using a laplacian pyramid of adversarial networks. In Proceedings of the Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 7–10 December 2015.
31. Gauthier J. Conditional Generative Adversarial Nets for Convolutional Face Generation. Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter Semester. 2014. Available online: http://cs231n.stanford.edu/reports/2015/pdfs/jgauthie_final_report.pdf (accessed on 24 February 2021).
32. Wang, X.; Gupta, A. Generative image modeling using style and structure adversarial networks. In *Lecture Notes in Computer Science, Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016*; Springer: Cham, Switzerland, 2016; pp. 318–335.
33. Mathieu, M.; Couprie, C.; LeCun, Y. Deep multi-scale video prediction beyond mean square error. *arXiv* **2015**, arXiv:1511.05440.
34. Yoo, D.; Kim, N.; Park, S.; Paek, A.S.; Kweon, I.S. Pixel level domain transfer. In *Lecture Notes in Computer Science, Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016*; Springer: Cham, Switzerland, 2016; pp. 517–532.
35. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
36. Reed, S. E.; Akata, Z.; Mohan, S.; Tenka, S.; Schiele, B.; Lee, H. Learning what and where to draw. *arXiv* **2016**, arXiv:1610.02454.
37. Karacan, L.; Akata, Z.; Erdem, A.; Erdem, E. Learning to generate images of outdoor scenes from attributes and semantic layouts. *arXiv* **2016**, arXiv:1612.00215.
38. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556
39. 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](#)]
40. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing & Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
41. Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv* **2015** arXiv:1511.07122.
42. Peng, C.; Zhang, X.; Yu, G.; Luo, G.; Sun, J. Large kernel matters—improve semantic segmentation by global convolutional network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4353–4361.
43. Pathak, D.; Shelhamer, E.; Long, J.; Darrell, T. Fully convolutional multi-class multiple instance learning. *arXiv* **2014**, arXiv:1412.7144.
44. Hong, S.; Noh, H.; Han, B. Decoupled deep neural network for semi-supervised semantic segmentation. *arXiv*, **2015**, arXiv:1506.04924.
45. Liu, H.; Xiao, H.; Luo, L.; Feng, C.; Yin, B.; Wang, D.; Li, Y.; He, S.; Fan, G. Semi-supervised Semantic Segmentation of Multiple Lumbosacral Structures on CT. In *Lecture Notes in Computer Science, Proceedings of the International Workshop and Challenge on Computational Methods and Clinical Applications for Spine Imaging, Shenzhen, China, 17 October 2019*; Springer: Cham, Switzerland, 2019; pp. 47–59.
46. Souly, N.; Spampinato, C.; Shah, M. Semi Supervised Semantic Segmentation Using Generative Adversarial Network. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.

47. Hung, W.C.; Tsai, Y.H.; Liou, Y.T.; Lin, Y.Y.; Yang, M.H. Adversarial learning for semi-supervised semantic segmentation. *arXiv* **2018**, arXiv:1802.07934.
48. Bousmalis, K.; Silberman, N.; Dohan, D.; Erhan, D.; Krishnan, D. Unsupervised pixel level domain adaptation with generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3722–3731.
49. Xue, Y.; Xu, T.; Zhang, H.; Long, L.R.; Huang, X. Segan: Adversarial network with multi-scale L1 loss for medical image segmentation. *Neuroin-Form.* **2018**, *16*, 383–392. [[CrossRef](#)]
50. Cherian, A.; Sullivan, A. Sem-GAN: Semantically-consistent image-to-image translation. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 7–11 January 2019; pp. 1797–1806.
51. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv* **2015**, arXiv:1511.06434.
52. Diederik, K.; Jimmy, B. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
53. Tong, X.Y.; Xia, G.S.; Lu, Q.; Shen, H.; Li, S.; You, S.; Zhang, L. Learning Transferable Deep Models for Land-Use Classification with High-Resolution Remote Sensing Images. *arXiv* **2018**, arXiv:1807.05713.