A new satellite-derived dataset for marine aquaculture areas in the China's coastal region

Yongyong Fu[1], Jinsong Deng[1], Hongquan Wang[1], Alexis Comber[2], Wu Yang[1], Wenqiang Wu[1], Shixue You[1], Yi Lin[3], Ke Wang[1]

[1]College of Environmental and Resource Sciences, Zhejiang University, Hangzhou, 310058, China
[2]School of Geography, University of Leeds, Leeds LS1 9JT, UK
[3]Department of Geography, University of Hong Kong, Hong Kong SAR 999077, China

Correspondence to: Jinsong Deng (jsong_deng@zju.edu.cn)

Abstract. China has witnessed extensive development of the marine aquaculture industry over recent years. However, such rapid and disordered expansion posed risks to coastal environment, economic development, and biodiversity protection. This study aimed to produce an accurate national-scale marine aquaculture map at a spatial resolution of 16 m, using a proposed deep convolution neural networks (CNNs) based model and applied it to satellite data from China's GF-1 sensor in an end-to-end way. The analyses used homogeneous CNNs to extract high-dimensional features from the input imagery and preserve information at full resolution. Then, a hierarchical cascade architecture was followed to capture multi-scale features and contextual information. This hierarchical cascade homogeneous neural network (HCHNet) was found to achieve better classification performance than current state-of-the-art models (FCN-32s, Deeplab V2, U-Net, and HCNet). The resulting marine aquaculture area map has an overall classification accuracy >95% (95.2%-96.4, 95% confidence interval). And marine aquaculture was found to cover a total area of ~1100 km^2 (1096.8 km^2-1110.6 km^2, 95% confidence interval) in China, of which more than 85% are marine plant culture areas, with 87% found in the Fujian, Shandong, Liaoning, and Jiangsu provinces. The results confirm the applicability and effectiveness of HCHNet when applied to GF-1 data, identifying notable spatial distributions of different marine aquaculture areas and supporting the sustainable management and ecological assessments of coastal resources at a national scale. The national-scale marine aquaculture map at 16 m spatial resolution is published in the Google Maps kmz File Format with georeferencing information at https://doi.org/10.5281/zenodo.3881612 (Fu et al., 2020).

1 Introduction

Marine aquaculture, which refers to the breeding, rearing, and harvesting of aquatic plants or animals in marine waters, has significant potential for food production, economic development, and environmental protection in coastal areas (Burbridge et al., 2001; Campbell and Pauly, 2013; Gentry et al., 2017). It has become a fast-growing industry in China due to the significant increase in the demand for seafood, support from policies, and technology innovation (Liang et al., 2018). The marine aquaculture production in China has increased from 10.6 million tons in 2000 (Bureau of Fisheries of the Ministry of

1

Agriculture, 2001) to 20.7 million tons in 2019 (Bureau of Fisheries of the Ministry of Agriculture, 2020). However, such rapid and disordered growth may cause severe economic losses and environmental problems, such as water pollution (Tovar et al., 2000), biodiversity decrease (Galil, 2009; Rigos and Katharios, 2010), and marine sediment pollution (Porrello et al., 2005; Rubio-Portillo et al., 2019). Therefore, accurate mapping and monitoring of marine aquaculture can provide evidence

35   to support the sustainable management of coastal marine resources.

Previous research in this domain can be grouped into visual interpretation, analyses enhanced by including ancillary data such as information about spatial structure, object-based image analysis (OBIA), and deep learning based methods. Visual interpretation is used less frequently as it requires too much time and effort. Enhanced analyses that incorporate features such as texture or average filtering (Fan et al., 2015; Lu et al., 2015; Xiao et al., 2013) are commonly employed for pixel-

40   based approaches. However, these are subject to noise (the salt-and-pepper effect) and decreased accuracy (Zheng et al., 2017). OBIA has been widely used for the detailed interpretation of marine aquaculture from remote sensing images (Fu et al., 2019a; Wang et al., 2017; Zheng et al., 2017). It first partitions the image into segments and then classifies segments based on their internal properties (Blaschke et al., 2014). However, since almost all these methods are proposed based on the handcrafted features, it is inherently difficult for them to achieve balance between high discriminability and good robustness

45   (L. Zhang et al., 2016). To solve such problems, the remote sensing community has started to incorporate deep fully convolutional neural networks (FCN) within marine aquaculture detection tasks using high spatial resolution (HSR) images at local scales (Cui et al., 2019; Fu et al., 2019b; Shi et al., 2018). However, the opportunities associated with analyses of the high volumes of publicly available and free remote sensing data at medium resolution, such as Landsat, Sentinel-2 A/B, and GaoFen-1 wide-field-of-view (GF-1 WFV) imagery, have not been exploited. Therefore, it is necessary to develop a

50   detection system applying deep FCNs to such data to provide more reliable and effective mapping and monitoring over wider areas, supporting evaluations of marine aquaculture areas at a national scale.

However, there are several critical limitations for accurate mapping of marine aquaculture areas using deep FCN-based methods when applied to medium-resolution data. The first is the coexistence of multi-scale objects, such as the large sea areas as well as small aquaculture areas, making it difficult to focus FCN on small marine aquaculture objects. A common

55   approach is to use inputs of different sizes from the original images (Eigen and Fergus, 2015; Liu et al., 2016; Zhao and Du, 2016) to ensure that different object sizes are prominent in different parts of the FCN structure, but such methods take more time due to the repetitive sampling of the input imagery. Some researchers have generated multi-scale features using atrous convolution (Chen et al., 2018) or pooling operations at different scales (He et al., 2015; Zhao et al., 2017). However, such approaches may be limited to a certain range of receptive fields, as operations may be applied to invalid zones when pooling

60   with a larger pooling size or atrous convolution with a higher atrous rate. The second critical limitation is that the final features may have a smaller size than the input imagery due to consecutive pooling operations in FCN, making it hard to identify land cover details. To solve this problem, researchers have used deconvolution operation (Noh et al., 2015) or fused features (Pinheiro et al., 2016; Ronneberger et al., 2015), but FCN may fail to identify relatively small marine aquaculture areas. Finally, some researchers have tried to refine the classification approach by including known boundaries (Bertasius et

2

65   al., 2016; Fu et al., 2019c; Marmanis et al., 2018), but such methods require additional classification steps to perform boundary extraction.

In conclusion, although present methods have been successfully applied for dense classifications, the challenge of using them to accurately extract the marine aquaculture areas from medium resolution images at a national scale remains.To overcome these limitations, we proposed a novel framework for the large-scale marine aquaculture mapping. The main

70   contributions of our study can be summarized as follows:

(1) We present a unified CNN-based framework for national-scale marine aquaculture extraction.

(2) A hierarchical cascade homogeneous neural network (HCHNet) model is proposed to learn discriminative and robust features.

(3) We provide the first detailed national-scale marine aquaculture map with a spatial resolution of 16 m.

75   The rest of the paper proceeds as follows. Section 2 briefly presents a description of the study area and different types of marine aquaculture. Section 3 introduces the input data and method to develop the proposed deep learning architectures, implementation details, and methodological choices. The results are presented in Section 4. Section 5 then discusses the methods and the limitations of using deep learning methods with medium resolution data, and finally, Section 6 concludes the paper.

80   **2 Study area**

The study area included all of the potential marine aquaculture areas in China's coastal regions (Fig. 1). Due to the large amount of coastline and associated resources, many coastal marine aquaculture areas have rapidly developed in coastal regions. After a visual inspection on the HSR images from Google Earth, we empirically set the width of the study area along with the coastal line for detection as 30 km. According to the types of cultivated aquatic products, the marine

85   aquaculture areas in China can be classified into marine animal culture (MAC) areas and marine plant culture (MPC) areas.

MAC areas are cultured with marine animals, such as fish, crustaceans, shellfish, etc., in connected cages (Fig. 2k), or wooden rafts (Fig. 2d). Most of the cages and rafts are small (normally 3×3 or 5×5 m in size) and simple in form (normally square). The materials used to construct these cages are collected locally and include bamboo, wooden boards, plastic foam floats, and polyvinyl chloride or nylon nets. Because of the low investment costs and ease of construction,

90   farmers typically make the cages themselves. As they cannot withstand waves generated by typhoons or sea currents, most of cages must be installed in inshore waters and sheltered sites (Fig. 2c, i).

MPC areas are generally cultivated with seaweed, such as kelp, undaria, gracilaria, etc. Most of the seaweed is twisted around ropes about 2 m in length. The ropes are linked or tied to one (Fig. 2 f) or two floating lines (Fig. 2j), which are about 60 m long and kept at the sea surface by buoys made from foam or plastic and anchored by lines tied to wooden pegs driven

95   into the sea bottom. As most of the MPC are submerged in the sea water, the features of MPC in remotely sensed images are usually influenced by different environments (Fig. 2b, e, g, h , j), making it difficult for classification.

Figure 1: Location of the study area, the spatial distribution of labeled samples, and acquired GF-1 wide-field-of-view (WFV) image swaths in China.
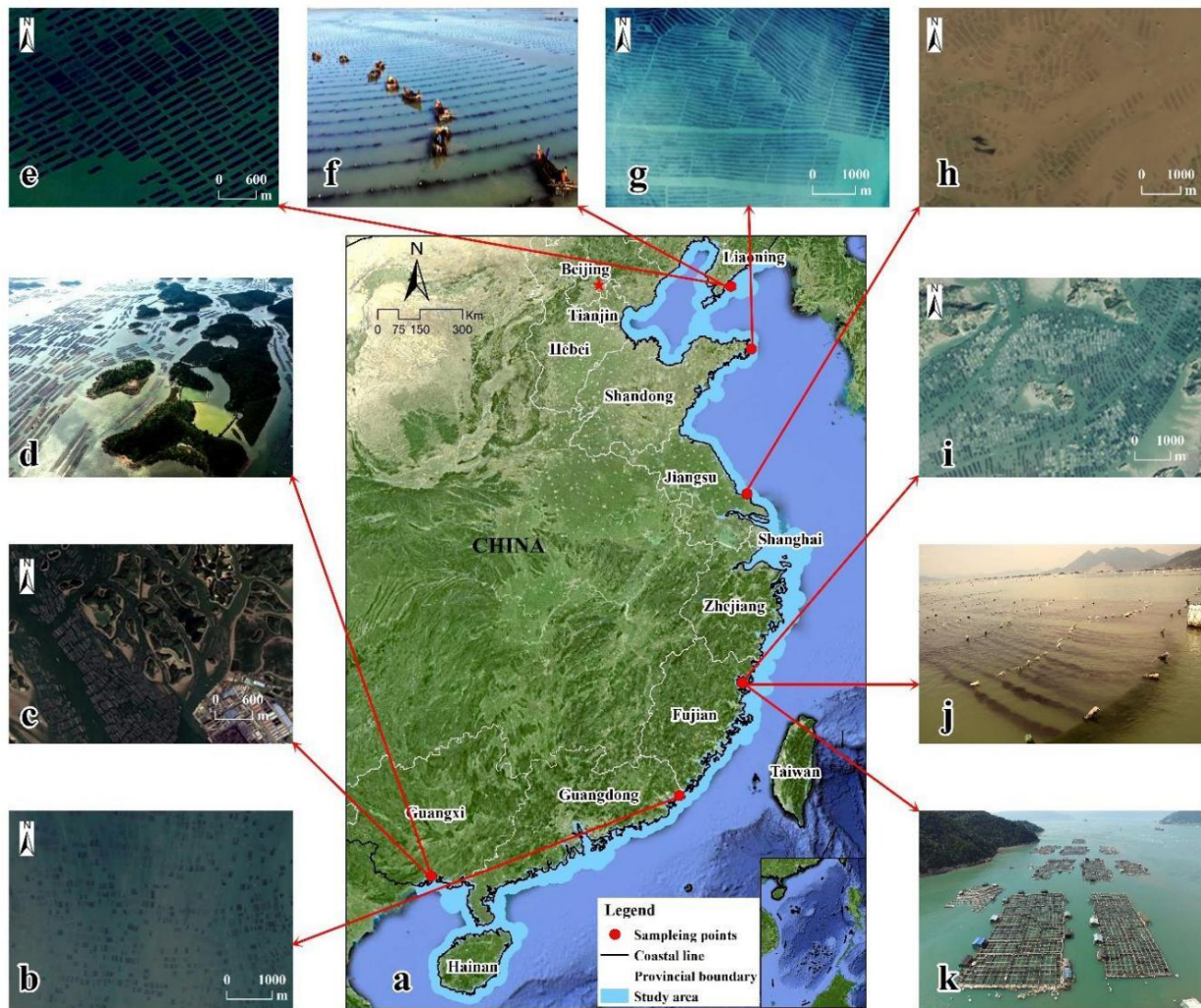
Figure 2: Location of the Sampling points (a). And image examples of typical marine aquaculture areas on ground or from high spatial resolution (HSR) images: (b) (e) (g) (h) (i) Marine plant culture (MPC) areas from HSR images; (c) (i) Marine animal culture (MAC) areas from HSR images; (d) (k) photos of MAC areas on ground; (f) (j) photos of MPC areas on ground.

3 Materials and methods

Due to the large number of factors that could potentially affect classification performance, implementation of the FCN-based method at the national scale is a challenge. To reduce the influence of various land covers, we used the coastal line vector (Chuang et al., 2019) to exclude mainland areas after preprocessing all the input images (Fig. 3); then, we produced the marine aquaculture map by utilizing the HCHNet method, which was trained and tested on dataset validated by field survey or HSR images.
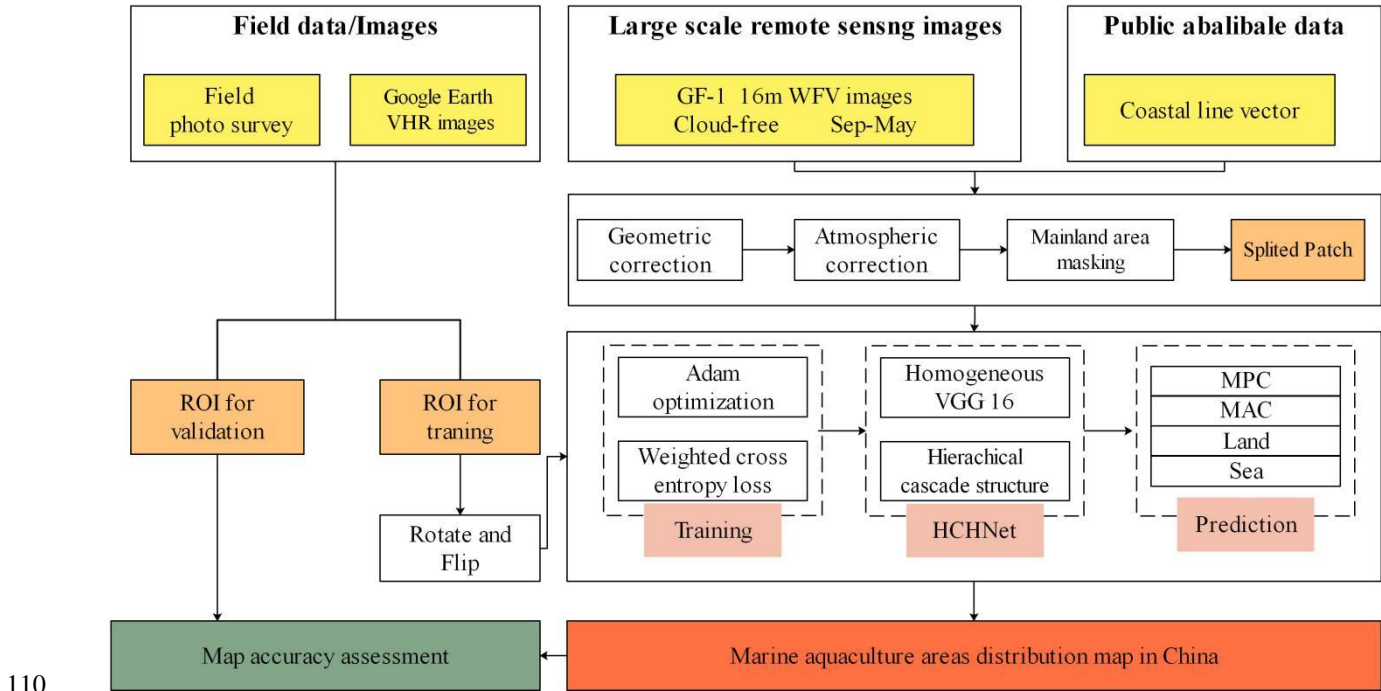
5

Figure 3: Schematic flowchart of the marine aquaculture mapping (ROI: region of interest. HCHNet: hierarchical cascade homogeneous neural network).

3.1 Data and preprocessing

In this study, images from the WFV sensors of GF-1 were selected as the primary data source. The GF-1 satellite, which is the first satellite of the China high-resolution earth observation satellite program, was launched by the China Aerospace Science and Technology Corporation in April 2013. This satellite carries four integrated WFV sensors, providing multi-spectrum data with a two-day revisit cycle and a swath width of 800 km when the four sensors are combined. Each WFV sensor has four multi-spectral bands at 16 m spatial resolution: B1 (450–520 nm, blue), B2 (520–590 nm, green), B3 (630–690 nm, red), and B4 (770–890 nm, near infrared). Compared with other frequently used medium resolution satellite imagery (e.g. Landsat, Sentinel), the wide coverage ability, high-frequency revisit time, and 16-m spatial resolution of the data significantly improves the capabilities for large-scale marine aquaculture areas observation and monitoring. A total of 35 quantified GF-1 WFV images spanning the 2016–2019 period were finally selected from the China Centre for Resources Satellite Data and Application to cover the whole coastal region in China and to filter for cloud coverage (Fig. 1). The product filenames are listed in Table S1 of the Supplementary material.

The images were projected into the UTM map projection, and atmospheric correction was undertaken using the FLAASH atmospheric correction model embedded in the ENVI software (v5.3.1). The "Maritime" model was set as Aerosol Model. And all the other parameters can be automatically set by using the extension tools (https://github.com/yyong-

6

fu/ENVI_FLAASH_EasyToUse). A 30 km buffer was used to extract the images of the coastal areas, and the final set of clipped images consisted of four image bands at a spatial resolution of 16 m, used as inputs in the following parts.

3.2 Hierarchical cascade homogeneous neural network

As shown in Fig. 4, the proposed HCHNet is an FCN-based neural network, which can be trained and applied to a large area in an end-to-end way. Specifically, a homogeneous CNN was designed to extract high-dimensional features from the input images. A hierarchical cascade structure was followed to extract multi-scale contextual information gradually based on high-dimensional features. The following subsections introduce three important components of the proposed HCHNet method, including (1) an encoder based on a homogeneous CNN; (2) hierarchical cascade structure; and (3) a loss function.
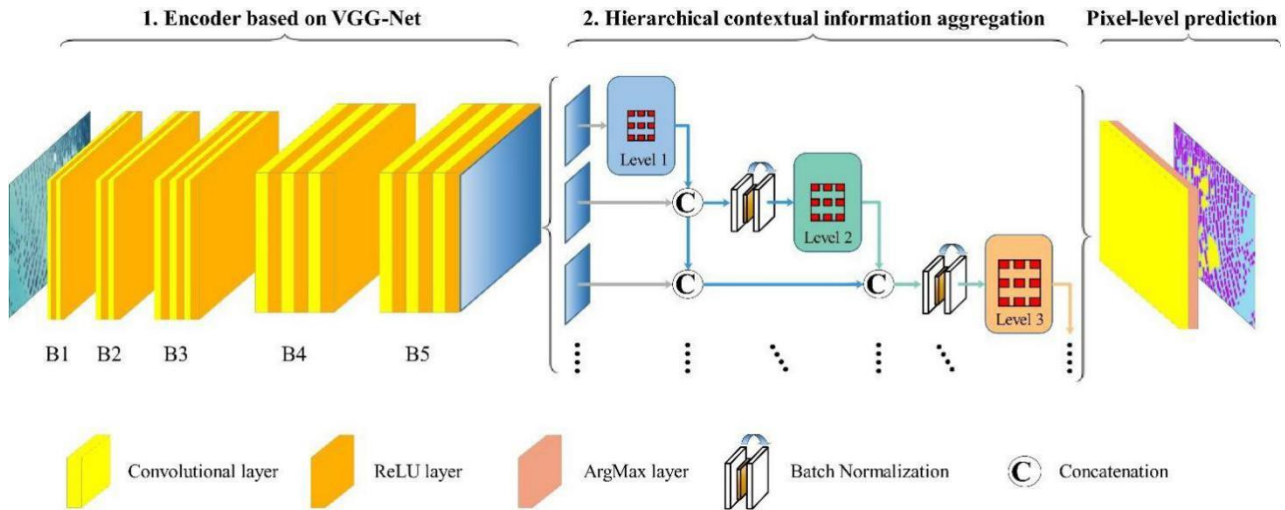

Figure 4: Overview of the proposed HCHNet approach.

3.2.1 Encoder based on a homogeneous CNN

Traditional CNN uses down-sampling process to improve the local invariance, and the prediction results are usually only labels at the patch level. For semantic segmentation, FCN can enlarge the down-sampling feature maps to full-sized outputs by using interpolation (Badrinarayanan et al., 2017) or deconvolution (Noh et al., 2015). However, foreground objects, such as the marine aquaculture areas in our study, occupy a smaller portion of the GF-1 WFV images than in the natural images, making it hard for the FCN to recover the details missing from consecutive pooling operations via learning.

As representations with high-resolution are important for the preservation of detailed information, the homogeneous CNN (Shi et al., 2018) was used as the encoder. One of the advantages of the homogeneous CNN is that it retains the full resolution of features by removing all of the pooling operations. As shown in Fig. 4, we built the encoder based on the widely used VGG-16 model. The VGG-16 is constructed of thirteen convolutional layers and followed by three fully connected layers. To preserve the spatial information and control the model size, the fully connected layer was removed and

convolutional kernels in the corresponding layers was reduced. As a result, the encoder can preserve full resolution features
150 as the input image.

3.2.2 Hierarchical cascade structure

Although removing of pooling operations can preserve more detailed information, it can also decrease the receptive field of
the underlying neural network (Liu et al., 2018). In this case, with fixed and limited receptive fields, it may cause more
misclassifications because of the loss of multi-scale contextual information. To solve this problem, the hierarchical cascade
155 structure proposed in a previous study (Fu et al., 2019b) was used. This structure generally enlarges the receptive field and
increases the sampling rate by creating a hierarchical cascade structure using the atrous convolution layers (as shown in the
central part of Fig. 4). To reduce memory usage, batch normalization operations were used to replace the attention modules,
allowing feature maps from different levels to be concatenated and easing the training process (Ioffe and Szegedy, 2015).
Each atrous convolution layer in the proposed structure is formulated as follow:

$$F_1 = C_{k,D_1}[F_o], \tag{1}$$

$$F_l = C_{k,D_l}[\mathcal{L}(F_o \copyright F_1 \copyright F_2 \copyright F_3 \copyright \ldots \copyright F_{l-1})], l > 1, \tag{2}$$

$$D_1 < D_2 < D_3 < \cdots < D_l, \tag{3}$$

160 where F_o denotes the feature maps from the output of our encoder network. $C_{k,D_l}[\cdot]$ denotes an atrous convolution operation
with the kernel size of $K \times K$ and dilation rate of d at the l-level. F_l ($l = 1,\ldots$, n) denotes the features at the l-level in the
structure. '$\copyright$' denotes the concatenation operation. $'\mathcal{L}(\cdot)'$ denotes the batch normalization. D_l denotes the dilation rate value
at the l-level.

3.2.3 Loss function

165 A significant problem during the training of FCN is the imbalance of classes. Such imbalances can make training inefficient,
with relatively small marine aquaculture areas contributing little to the model training process. In contrast, much of the
coastal area contains negative samples, such as the sea area, which may dominate the training process and decrease marine
aquaculture identification accuracy. To address this class imbalance problem, Eigen and Fergus (2015) proposed reweighting
each class based on a loss calculation. Following this idea, the weighted loss was used to deal with the class imbalance
170 problem and to allow effective training of all examples. The loss was defined as:

$$\text{WL}(p, q) = -\sum_{i=1}^{n} \alpha_i p(x_i) \log(q(x_i)), \tag{4}$$

where p $\in \{0,1\}$ represents the ground truth class, q $\in [0,1]$ is the predictive probability from the model for the classes, and
α_i represents the weight for each class.

8

3.3 Implementation details

As shown in Fig. 4, the encoder was first applied, which is a homogenous CNN with 13 layers (Table 1), to produce high-dimensional and abstract features with full resolution from the input imagery. And then, multi-scale contextual information was captured by using the hierarchical cascade structure with atrous rates of 3, 6, and 9. To regulate the model's memory consumption and to prevent it growth too wide, 1×1 kernels were employed in the hierarchical cascade structure to keep all the channels of concatenated features fixed to 128, which have the same output feature maps of other atrous convolution layers.

Table 1: Detailed configuration of the encoder in the proposed HCHNet method. (l, $k \times k \times n$, s) of configurations means there are l convolution layers with n convolution kernels and their size is $k \times k$ and stride is s. (h, w, c) of the output size means the output image or feature maps have a height of h, a width of w, and a channel of c.

Layer name	Layer type	Configurations	Output size
Input	Image data	-	$256 \times 256 \times 4$
B1	Convolution, ReLU	2, $3 \times 3 \times 32$, 1	$256 \times 256 \times 32$
B2	Convolution, ReLU	2, $3 \times 3 \times 64$, 1	$256 \times 256 \times 64$
B3	Convolution, ReLU	3, $3 \times 3 \times 128$, 1	$256 \times 256 \times 128$
B4	Convolution, ReLU	3, $3 \times 3 \times 256$, 1	$256 \times 256 \times 256$
B5	Convolution, ReLU	3, $3 \times 3 \times 256$, 1	$256 \times 256 \times 256$

As for the training and testing of HCHNet, we selected a total of 705 patches. Each one of them has non-overlapping 256 $\times 256$ pixels from raw images (Fig. 1). The ground truth maps for each patch were obtained by visual interpretation. From them, we randomly selected 80% to construct the training dataset. Considering the relatively small training dataset, data augmentation was applied to make the training process more effective and reduce overfitting: each patch was flipped in the horizontal and vertical direction and was rotated counterclockwise by 90. As a result, there were 4656 patches can be used for the training of HCHNet.

In the experiment, we trained the HCHNet for 30 epochs using a batch size of 4 and the Adam optimizer. The Adam parameters were set as: $\beta 1 = 0.9$, $\beta 2 = 0.999$, and a learning rate of 0.0001. The HCHNet was built and implemented using the Keras (v2.2.4) on top of Tensorflow (v1.8.0). All of the experiments were undertaken on a computer with an graphics processing unit (GPU) of NVIDIA GeForce GTX 1070.

3.4 Comparing methods and accuracy assessment

3.4.1 Comparing methods

To assess the effectiveness and advantage of our proposed method, we provided a comparison with four state-of-the-art FCN-based methods. We summarized the main information as follows:

FCN-32s: the first FCN-based method proposed by Zhang et al. (2015) for semantic segmentation. It was constructed based on the VGG-16, in which the original fully connected layers are convolutionized. The model predicts classification results by upsampling the final feature maps 32 times directly. It does not use any structure to extract multi-scale feature

200 maps or get more detailed information from the shallow layers. Thus, it can be used as a baseline model for our proposed HCHNet methods.

U-Net: a typical FCN-based model with encoder–decoder structure,which was proposed by Ronneberger et al. (2015) for semantic segmentation of medical images. The encoder has a similar structure to VGG-16. Different from the FCN 32s, U-Net combined the feature maps in the decoder and mirrored feature maps in the encoder by using long-span connections to
205 provide precise localization and high classification accuracy.

Deeplab V2: Chen et al. (2018) proposed the Deeplab V2 (VGG-16 as the backbone) model for semantic segmentation, which used the atrous spatial pyramid pooling (ASPP) structure to capture multi-scale contextual information, and then used the fully connected conditional random fields (CRFs) as a post-processing tool to refine the prediction results.

HCNet: HCNet was proposed by Fu et al. (2019b) to map the detailed spatial distribution of marine aquaculture from HSR
210 images. This model has a variant of VGG-16 as an encoder, in which the stride and padding of the last two pooling layers are set as one for high- resolution feature maps. It combines the long-span connections, while also combining a hierarchical cascade structure.

The above models are suitable for classification and comparison purposes, because nearly all of these methods are VGG-16 based neural networks and employed typical structures for multi-scale information extraction. In the training phase, all of
215 the above models, including the proposed HCHNet, were trained from scratch using the same patches and experimental settings as in the HCHNet method.

3.4.2 Accuracy assessment and comparison

To ensure representativeness of each class in the whole sample population for accuracy assessment, we followed the widely used random stratified sampling method (Padilla et al., 2014; Ramezan et al., 2019) to generated 4000 randomly selected
220 points in the coastal zone. Based on the visual interpretation results from HSR images, several commonly-used accuracy statistics were calculated from the error matrix, such as producer accuracy (PA), user accuracy (UA), overall accuracy (OA), and the kappa coefficient. Meanwhile, we also conducted the area accuracy assessment (the percentage of overlapping areas in the ground truth) based on more than 120 randomly selected 256×256 patches, which accounts for nearly 20% of the total samples.
225 After that, we compared the performances of our proposed method with four state-of-the-art FCN-based models. The accuracy comparison was undertaken using the test dataset. To provide a quantitative assessment between our proposed method and other methods, we calculated the widely-used F1 score (F1), precision, and recall (POWERS, 2011) as follows:

$$\text{Precision} = \frac{TP}{TP + FP}. \tag{5}$$

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{6}$$

10

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{7}$$

where TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives.

To compare different models' discriminate ability for the marine aquaculture areas, these accuracy values were calculated for each class, and the mean F1 values of the MAC and MPC areas were used to assess the performance of the different methods.

4 Results

4.1 Spatial distribution of marine aquaculture areas in China
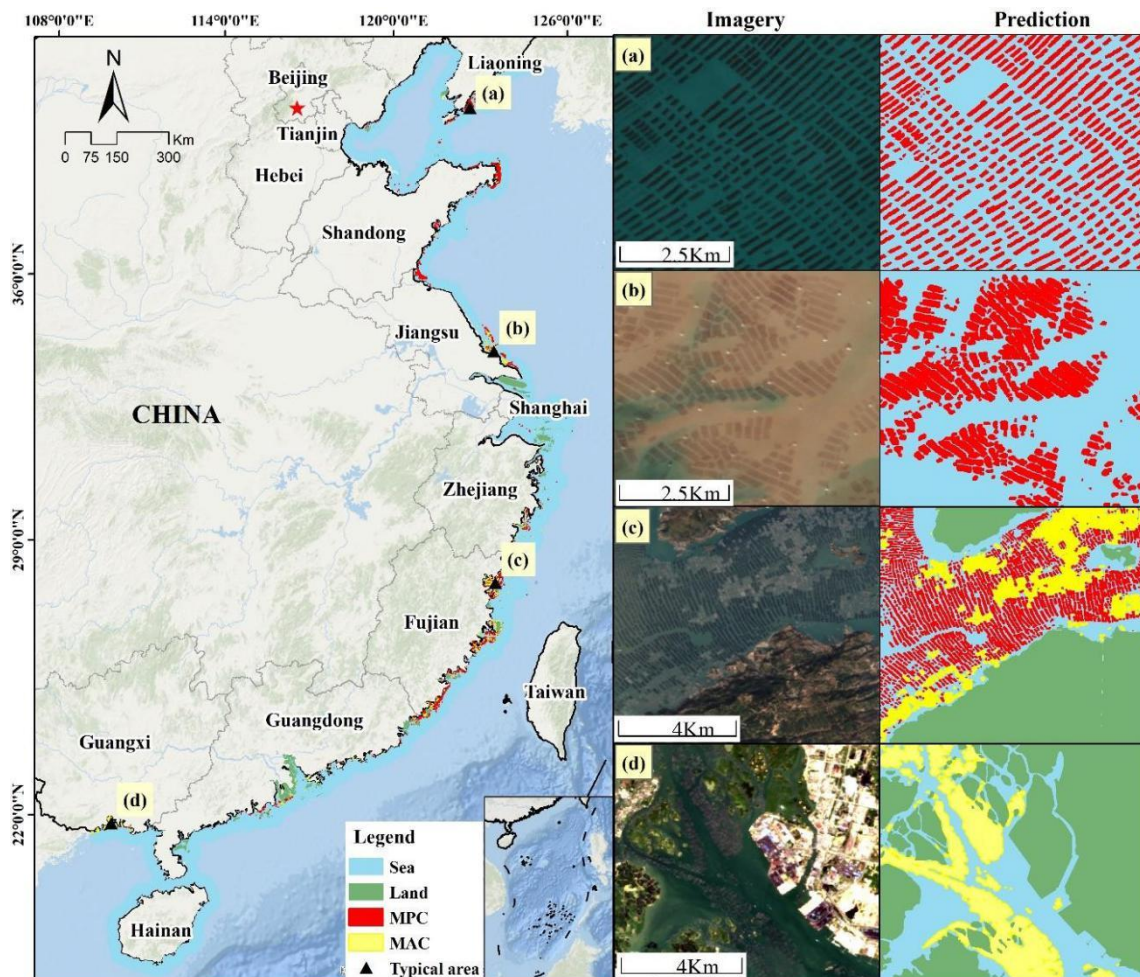

Figure 5: Spatial distribution of China's marine aquaculture and zoom views of imagery and prediction results for typical areas of (a), (b), (c), and (d).

11

The final classification results are shown in Fig. 5, with the corresponding Google Maps kmz file are published at https://doi.org/10.5281/zenodo.3881612 (Fu et al., 2020). The prediction results of typical areas (Fig. 5a–d) demonstrate the applicability and robustness of the HCHNet method to different marine aquaculture areas (i.e., MPC and MAC) over different study sites (i.e., from Liaoning to Guangxi provinces).

(a)   (b)

(c)   (d)
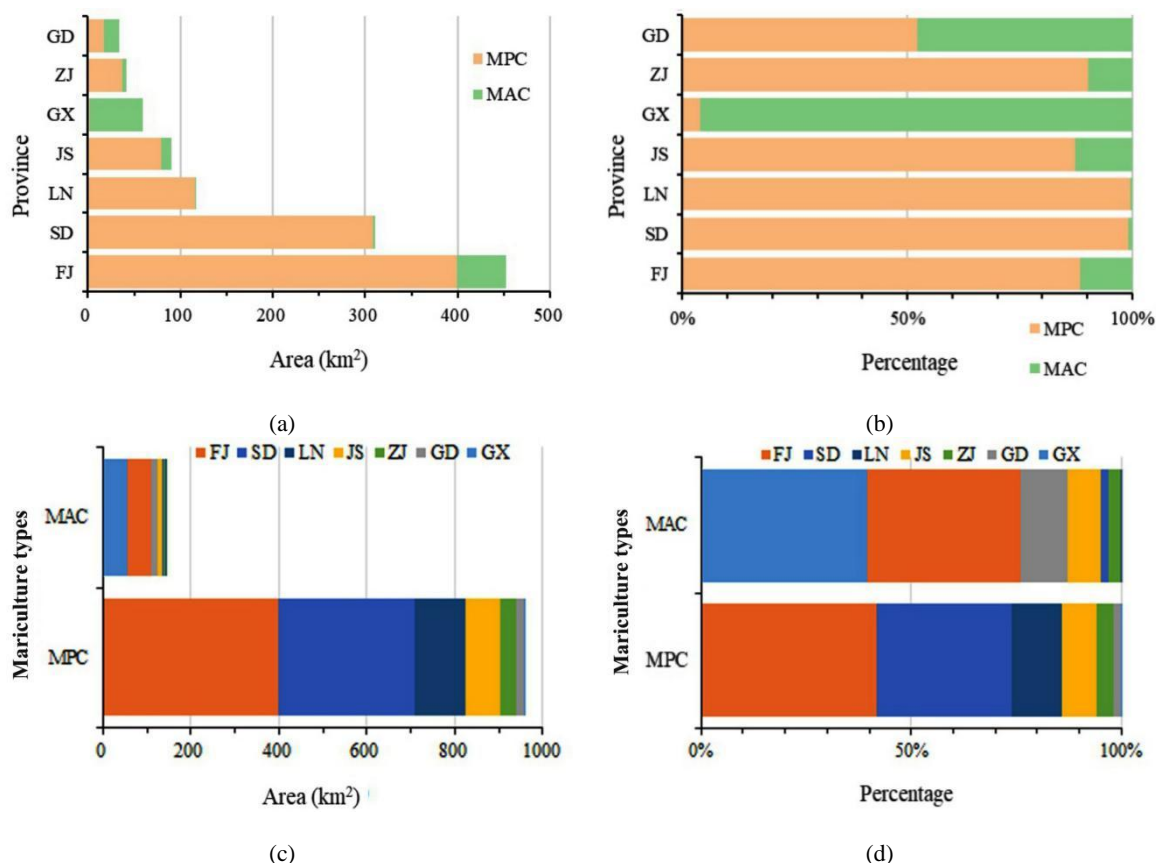
Figure 6: The amount (area) and percentage of different types of marine aquaculture distribution in the coastal provinces of China. LN, SD, JS, ZJ, FJ, GX, and GD indicate Liaoning, Shandong, Jiangsu, Zhejiang, Fujian, Guangxi, and Guangdong provinces, respectively.

According to the classification results, the total area of marine aquaculture in China is approximately 1103.67 km^2. As can be seen from Fig. 6a, marine aquaculture is mainly distributed in the coastal areas of Fujian, Shandong, Liaoning, and Jiangsu provinces. Fujian and Shandong provinces have the largest areas of over 300 and 450 km^2, respectively. Furthermore, nearly 100 km^2 of marine aquaculture areas are found in Liaoning and Jiangsu provinces, respectively. Figure 6b shows that over 85% of the marine aquaculture areas in these four coastal areas are MPC. Fig. 6b also shows that the provinces in North China, such as Liaoning and Shandong, tend to have more MPC areas, with the provinces in South China having more MAC areas.

12

250 Figure 6c shows that most of the marine aquaculture areas in China are MPC areas, with an area of over 950 km^2, six times larger than the MAC area. Guangxi and Fujian provinces have the largest areas of MAC, which account for more than 70% of the total MAC areas in China (Fig. 6d). The largest areas of MPC are found in Fujian and Shandong provinces, accounting for more than 70% of the total MPC areas in China.

4.2 Accuracy assessment of the marine aquaculture area map in China

255 To quantitatively evaluate the classification performance of our proposed HCHNet method, we used the random stratified sampling method to perform the evaluation. A total of 4000 reference pixels were randomly selected, with 1000 pixels from the classification results for each classes. We then obtained the ground truth of each point by visual interpretation based on HSR images from Google Earth point-by-point.

As shown in Table 2, the error matrix shows that the overall accuracy was 95.83%, and the kappa coefficient was 0.94.
260 The land and sea areas get the highest classification accuracy with both of the PA and UA values greater than 91%. The MPC areas have relatively high PA and UA values of approximately 95%. Most of the misclassifications of MPC areas are related to the sea area. This is because the MPC areas are submerged in a complex sea environment, which can easily be affected by waves, seafloor topography, shadows of clouds, etc. The MAC areas have a relatively lower UA value of 89.1%, which may be caused by the relatively high complexity and small numbers of training patches of MAC areas.

265 **Table 2: Accuracy assessment of the classification results in China based on visual interpretation (pixels). (PA: producer accuracy. UA: user accuracy)**

Predicted class	Ground truth					
	Sea	Land	MPC	MAC	Sum	UA
Sea	981	12	5	2	1000	98.10%
Land	5	987	3	5	1000	98.70%
MPC	20	3	974	3	1000	97.40%
MAC	31	76	2	891	1000	89.10%
Sum	1037	1078	984	901		
PA	94.60%	91.56%	98.98%	98.89%		
Overall accuracy	95.83%					
Kappa coefficient	0.94					

After that, we employed the bootstrapping (Efron and Tibshirani, 1997), which is suitable for estimating classification accuracy (Duan et al., 2020; Lyons et al., 2018), to estimate the uncertainty level. We bootstrapped the overall accuracy from 4000 independent reference points. The bootstrapping was performed for 1000 iterations, and the mean of the distribution
270 used for the evaluation and the confidence intervals was set as 95% quantile. Eventually, we obtained the overall accuracy of 95.8% (95.2%-96.4, 95% confidence interval). Meanwhile, we derived that the marine aquaculture area in China is 1103.67 km^2 (1096.8 km^2-1110.6 km^2, 95% confidence interval).

To further assess the validity of our proposed HCHNet method, we also evaluated the area accuracy (percentage of the over lapping area) based on the test dataset, including 120 randomly selected patches with a size of 256 × 256, which

275   accounts for 20 % of the labeled samples. As shown in Table 3, the land and sea areas have the best classification accuracy with the area accuracy values greater than 93%. Meanwhile, the MPC and MAC also have a relatively high area accuracy values of 81.8% and 72.5%, respectively.

Table 3. Area accuracy assessment of different classes based on randomly selected patches.

Class	Sea area	Land area	RCA	CCA
Area accuracy:	93.6%	98.4%	81.8%	72.5%

4.3 Comparison with the state-of-the-art methods

280   To assess the quality of the proposed HCHNet method, the performance was compared with the results from other state-of-the-art FCN-based methods, using the same test dataset. As can be seen from the prediction results in Fig. 7, FCN-32s and Deeplab V2 are unable to reliably identify marine aquaculture, especially the small and isolated marine aquaculture areas (Fig. 7b,d,e) with much coarser predicted results than other approaches. HCHNet identified more MPC and MAC areas than U-Net and HCNet (Fig. 7a,d,e), and also identified detailed information, even with the narrow channels among neighboring

285   MPC areas (Fig. 7b,e).

To provide a quantitative comparison, several commonly-used accuracy metrics were calculated from the test dataset for MAC and MPC areas. Table 4 shows that the FCN-32s and Deeplab V2 achieved similar accuracy values, with mean F1 values less than 40%. The U-Net and HCNet achieved a similar classification performance, with the mean F1 values of approximately 70%. Compared with these state-of-the-art methods, our proposed HCHNet approach obtained the best

290   classification performance, with the mean F1 value of 76.3%. Besides, the HCHNet also achieved a good balance between precision and recall values of MAC, identifying more accurate and existing MAC areas. The difference between them is less than 4% for the HCHNet, while the difference values of other methods are more than 28%.

Table 4. Quantitative comparison of MPC and MAC areas between our method and other methods, where the best accuracy values are in bold (%).

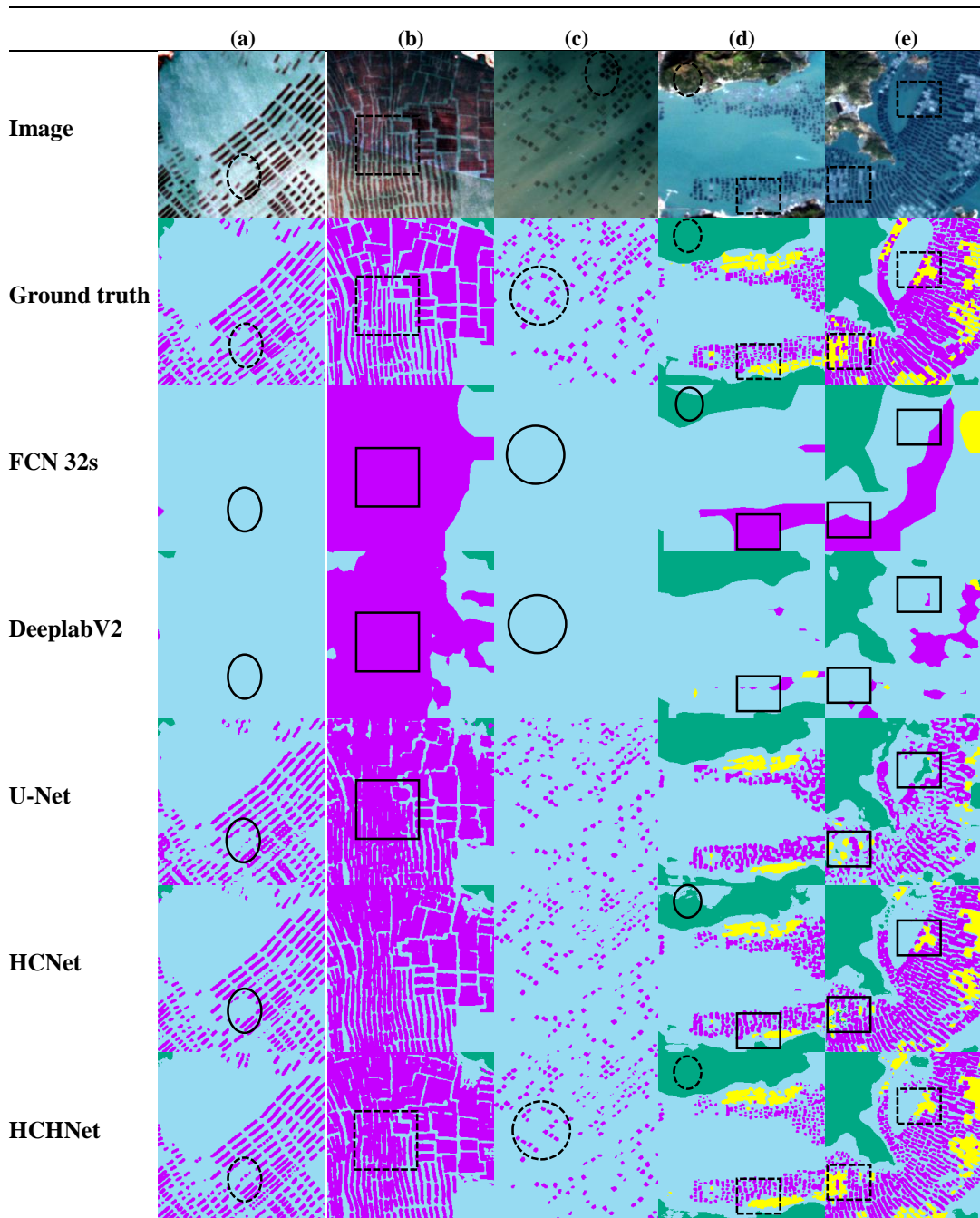| Methods	MPC			MAC			Avg.
	Precision (%)	Recall (%)	F1 (%)	Precision (%)	Recall (%)	F1 (%)	mean F1 (%)
FCN-32s	66.1	54.2	59.6	63.9	8.60	15.2	37.4
DeeplabV2	70.4	40.9	51.7	53.7	5.71	10.3	31.0
U-Net	83.8	77.6	<u>80.6</u>	81.8	53.1	<u>64.4</u>	<u>72.5</u>
HCNet	77.3	79.0	78.1	74.0	40.0	51.9	65.0
Ours-HCHNet	82.2	81.8	**82.0**	68.7	72.5	**70.6**	**76.3**

295

Figure 7: The classification results of MPC and MAC areas comparing the proposed HCHNet method with other approaches. The black solid outlined areas indicate where HCHNet obtains better results. The dotted line shows same locations in other images. The purple, yellow, blue, and green areas in the classification maps represent the MPC, MAC, sea, and land areas, respectively.

5 Discussion

5.1 Date and algorithms for the mapping of marine aquaculture areas in China

This study developed a new algorithm to separate two typical marine aquaculture types based on the most advanced FCN-based models. The HCHNet was applied to medium spatial resolution images from China's GF-1 WFV sensors to map the marine aquaculture areas in China. The input data and the algorithm used in our study were different from current state-of-the-art methods in many ways.

First, China's GF-1 WFV sensors provide a larger number of valid image scenes that are suitable for a wide range of analyses of marine aquaculture areas with high temporal resolution. MPC areas are only visible in several specific months due to phenological development stages. However, it is difficult to capture appropriate images that clearly represent the marine culture areas in these months from other similar satellites, such as Landsat, which are frequently influenced by clouds or waves. The high temporal resolution of the GF-1 WFV data (repeats each 2 days) means that it is possible to observe marine aquaculture areas with much greater frequencies than data from other sources. Additionally, the relatively wide swath of the data makes them highly suitable for such large-scale mapping in China. In addition, it is possible to directly obtain images with 16 m spatial resolution without any additional computations, such as pan-sharpening operations, making the GF-1 WFV data a reliable data source for large-scale marine aquaculture area observation and monitoring.

Second, the proposed FCN-based HCHNet method improves the classification accuracy and efficiency. Much previous research has used OBIA approaches (Fu et al., 2019a; Wang et al., 2017) and other FCN-based methods (Fu et al., 2019b; Shi et al., 2018). The accuracy of the OBIA method depends on segmentation, which does not have universal methods for evaluating the selection of appropriate segmentation parameters (Blaschke, 2010). It also takes a large amount of time to undertake the segmentation process and to design effective features or rules for hard classifications (Zheng et al., 2017), making such approaches more difficult to be implemented operationally for national-scale studies. The proposed HCHNet achieved the best classification performance for three reasons: (1) all of the pooling operations were removed to avoid the shrinking of features, which helps improve the identification of smaller foreground objects; (2) the hierarchical structure was used to enlarge the receptive field to capture more contextual information, which is helpful for reducing the influence of local variance; (3) the weighted loss function was employed to solve the classes imbalance problem.

Third, masking out coastal land areas that do not intersect with marine aquaculture areas was undertaken using publicly available data and provided a simple and straightforward methodological refinement to constrain the marine aquaculture mapping. This was important because of the scale of the classification over large coastal areas in China, which contain various land covers outside of the aims of this study. Previous studies have used a threshold value (Zheng et al., 2017) to mask out these land areas, but in this study, this was done directly.

5.2 Uncertainty and limitations of the marine aquaculture map in China

Accurate mapping of marine aquaculture areas at a regional scale is challenging. There are several potential uncertainties in our methods for mapping marine aquaculture areas. First, because of the medium spatial resolution imagery and the relatively small size of MAC area (Fig. 8a), it is difficult to accurately identify the boundaries of small and isolated MAC areas (generally less than 10 pixels). Overestimation of MAC may occur, where the sea waters among several MAC areas are misclassified as MAC. The HCHNet failed to detect the small MPC areas (Fig. 8b) and harvested MPC areas (Fig. 8c), causing an issue of underestimation. As shown in Fig. 8d, some vegetation that is submerged or close to the sea waters may be misclassified as MPC areas, since these pixels share similar spectrum and shape features.
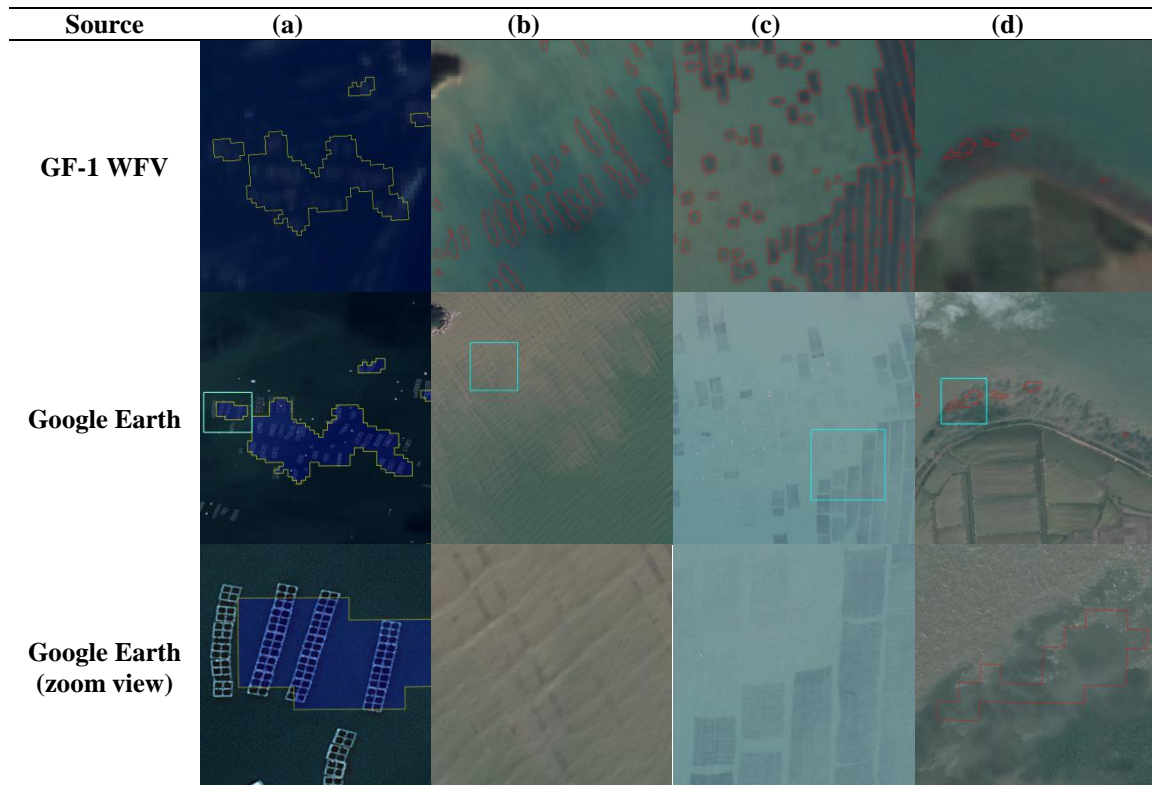
Figure 8: Illustration of potential sources of error in the HCHNet algorithm: the boundaries of relatively small and separate MAC (a) or MPA (b) areas are difficult to accurately identify; (c) harvested MPC areas are also difficult to detect due to shallow waters and the disappeared dark tone; (d) vegetation located close to the water bodies may be misclassified as MPC areas. The first row highlights typical misclassified areas from GF-1 WFV data. The second row shows high spatial resolution images from Google Earth (© Google Maps). The third row is a zoom view of the Google Earth images in the second row (© Google Maps). The red and yellow areas indicate the classification results of MPC and MAC areas, respectively.

There are also some limitations of the proposed HCHNet approach. First, the training process requires a large number of high-quality ground truth labels, which may require much manual work and professional interpretation experience. Therefore,

17

further research on accelerating the training or inference process through weak supervision (Lin et al., 2016; Pathak et al., 2015) or a series of model compression methods (Li et al., 2017; Yim et al., 2017; X. Zhang et al., 2016) will be undertaken to enhance the applicability of the approach. Second, our proposed methods can only be used for the monitoring of marine aquaculture areas in surface water; it is unable to detect the submerged cages in some places (such as coastal area of Shandong province in northeastern China).

6 Data availability

The map of marine aquaculture in China's coastal zone at 16 m spatial resolution have been published in the Google Maps kmz File Format with georeferencing information at https://doi.org/10.5281/zenodo.3881612 (Fu et al., 2020).

7 Conclusions

Marine aquaculture areas and the coastal environment they rely on are of significant ecological and socioeconomic value. Accurate and effective mapping approaches are imperative for the monitoring, planning, and sustainable development of marine and coastal resources across local, regional, and global scales. The increasing public availability of remote sensing data, ancillary data, and advanced computer vision algorithms, together provided an effective route for identifying marine aquaculture areas at a national scale. By using the powerful and inherent self-learning mechanism of deep learning, a new algorithm was carefully designed based on the FCN structure and applied to the GF-1 WFV data. The application of this algorithm produced a marine aquaculture area map of China with an overall classification accuracy >95% (95.2%-96.4, 95% confidence interval). The total area of China's marine aquaculture areas was estimated to be approximately 1100 km^2 (1096.8 km^2-1110.6 km^2, 95% confidence interval), of which more than 85% is MPC areas. Most of the marine aquaculture areas are distributed along the coastal areas of Fujian, Shandong, Liaoning, and Jiangsu provinces. Guangxi and Fujian provinces have the largest areas of MAC, and Fujian and Shandong have the largest areas of MPC. The algorithm could be implemented at other regional and global scales with the collection of sufficient samples and the careful investigation of marine aquaculture phenology in these areas.

Author Contributions

Funding acquisition, J.D., W.Y and A.C.; methodology, Y.F.; supervision, J.D. and K.W.; writing—original draft, Y.F.; writing—review & editing, H.W., A.C., S.Y., Y.L., and W.W.

18

Competing interests

The authors declare that they have no conflict of interest.

380 **References**

Badrinarayanan, V., Kendall, A. and Cipolla, R.: SegNet: A Deep Convolutional Encoder-Decoder Architecture for
Image Segmentation., IEEE Trans. Pattern Anal. Mach. Intell., 39(12), 2481–2495, doi:10.1109/TPAMI.2016.2644615,
2017.

Bertasius, G., Shi, J. and Torresani, L.: Semantic Segmentation with Boundary Neural Fields, in 2016 IEEE Conference
385 on Computer Vision and Pattern Recognition (CVPR), pp. 3602–3610, IEEE, Las Vegas, NV, USA., 2016.

Blaschke, T.: Object based image analysis for remote sensing, ISPRS J. Photogramm. Remote Sens., 65(1), 2–16,
doi:10.1016/j.isprsjprs.2009.06.004, 2010.

Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der
Werff, H., van Coillie, F. and Tiede, D.: Geographic Object-Based Image Analysis - Towards a new paradigm, ISPRS J.
390 Photogramm. Remote Sens., 87, 180–191, doi:10.1016/j.isprsjprs.2013.09.014, 2014.

Burbridge, Hendrick, Roth and Rosenthal: Social and economic policy issues relevant to marine aquaculture, J. Appl.
Ichthyol., 17, 194–206, doi:10.1046/j.1439-0426.2001.00316.x, 2001.

Bureau of Fisheries of the Ministry of Agriculture: China Fishery Statistical Yearbook 2001, China Agriculture Press,
Beijing, China., 2001.

395 Bureau of Fisheries of the Ministry of Agriculture: China Fishery Statistical Yearbook 2020, China Agriculture Press,
Beijing, China., 2020.

Campbell, B. and Pauly, D.: Mariculture: A global analysis of production trends since 1950, Mar. Policy, 39, 94–100,
doi:10.1016/j.marpol.2012.10.009, 2013.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A. L.: DeepLab: Semantic Image Segmentation
400 with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs., IEEE Trans. Pattern Anal. Mach.
Intell., 40(4), 834–848, doi:10.1109/TPAMI.2017.2699184, 2018.

Chuang, L., Ruixiang, S., Yinghua, Z., Yan, S., Junhua, M., Lizong, W., Wenbo, C., Doko, T., Lijun, C., Tingting, L.,

Zui, T. and Yunqiang, Z.: Global Multiple Scale Shorelines Dataset Based on Google Earth Images (2015), , doi:10.3974/geodb.2019.04.13.V1, 2019.

405   Cui, B., Fei, D., Shao, G., Lu, Y. and Chu, J.: Extracting Raft Aquaculture Areas from Remote Sensing Images via an Improved U-Net with a PSE Structure, Remote Sens., 11(17), 2053, doi:10.3390/rs11172053, 2019.

Duan, Y., Li, X., Zhang, L., Chen, D., Liu, S. and Ji, H.: Mapping national-scale aquaculture ponds based on the Google Earth Engine in the Chinese coastal zone, Aquaculture, 520, 734666, doi:10.1016/j.aquaculture.2019.734666, 2020.

410   Efron, B. and Tibshirani, R.: Improvements on cross-validation: The .632+ bootstrap method, J. Am. Stat. Assoc., 92(438), 548–560, doi:10.1080/01621459.1997.10474007, 1997.

Eigen, D. and Fergus, R.: Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture, in 2015 IEEE International Conference on Computer Vision (ICCV), pp. 2650–2658, IEEE, Santiago, Chile., 2015.

415   Fan, J., Chu, J., Geng, J. and Zhang, F.: Floating raft aquaculture information automatic extraction based on high resolution SAR images, in 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 3898–3901, IEEE, Milan, Italy., 2015.

Fu, Y., Deng, J., Ye, Z., Gan, M., Wang, K., Wu, J., Yang, W. and Xiao, G.: Coastal aquaculture mapping from very high spatial resolution imagery by combining object-based neighbor features, Sustain., doi:10.3390/su11030637, 2019a.

420   Fu, Y., Ye, Z., Deng, J., Zheng, X., Huang, Y., Yang, W., Wang, Y. and Wang, K.: Finer Resolution Mapping of Marine Aquaculture Areas Using WorldView-2 Imagery and a Hierarchical Cascade Convolutional Neural Network, Remote Sens., 11(14), 1678, doi:10.3390/rs11141678, 2019b.

Fu, Y., Liu, K., Shen, Z., Deng, J., Gan, M., Liu, X., Lu, D. and Wang, K.: Mapping Impervious Surfaces in Town–Rural Transition Belts Using China's GF-2 Imagery and Object-Based Deep CNNs, Remote Sens., 11(3), 280, 425   doi:10.3390/rs11030280, 2019c.

Fu, Y., Deng, J., Wang H., Comber, A., Yang, W., Wu, W., You X., Lin Y. and Wang, K.: A new satellite-derived dataset for marine aquaculture in the China's coastal region, Data set, Zenodo, https://doi.org/10.5281/zenodo.3833225, 2020.

Galil, B. S.: Taking stock: Inventory of alien species in the Mediterranean sea, Biol. Invasions, 11(2), 359–372, 430   doi:10.1007/s10530-008-9253-y, 2009.

Gentry, R. R., Froehlich, H. E., Grimm, D., Kareiva, P., Parke, M., Rust, M., Gaines, S. D. and Halpern, B. S.: Mapping the global potential for marine aquaculture, Nat. Ecol. Evol., 1, 1317–1324, doi:10.1038/s41559-017-0257-9, 2017.

He, K., Zhang, X., Ren, S. and Sun, J.: Spatial Pyramid Pooling in Deep Convolutional Networks for Visual 435   Recognition, IEEE Trans. Pattern Anal. Mach. Intell., 37(9), 1904–1916, doi:10.1109/TPAMI.2015.2389824, 2015.

Ioffe, S. and Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift,

in Proceedings of the International Conference on Machine Learning, pp. 448–456, Lille, France., 2015.

Li, H., Kadav, A., Durdanovic, I., Samet, H. and Graf, H. P.: Pruning Filters for Efficient ConvNets, in International Conference on Learning Representations, pp. 1–13, Toulon, France., 2017.

Liang, Y., Cheng, X., Zhu, H., Shutes, B., Yan, B., Zhou, Q. and Yu, X.: Historical Evolution of Mariculture in China During Past 40 Years and Its Impacts on Eco-environment, Chinese Geogr. Sci., 28(3), 363–373, doi:10.1007/s11769-018- 0940-z, 2018.

Lin, D., Dai, J., Jia, J., He, K. and Sun, J.: ScribbleSup: Scribble-Supervised Convolutional Networks for Semantic Segmentation, in 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3159–3167, IEEE, Las Vegas, NV, USA., 2016.

Liu, Y., Zhong, Y., Fei, F. and Zhang, L.: Scene semantic classification based on random-scale stretched convolutional neural network for high-spatial resolution remote sensing imagery, in 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 763–766, IEEE, Beijing, China., 2016.

Liu, Y., Yu, J. and Han, Y.: Understanding the effective receptive field in semantic image segmentation, Multimed. Tools Appl., 77(17), 22159–22171, doi:10.1007/s11042-018-5704-3, 2018.

Lu, Y., Li, Q., Du, X., Wang, H. and Liu, J.: A Method of Coastal Aquaculture Area Automatic Extraction with High Spatial Resolution Images, Remote Sens. Technol. Appl., 30(3), 486–494, doi:10.11873/j.issn.1004-0323.2015.3.0486, 2015.

Lyons, M. B., Keith, D. A., Phinn, S. R., Mason, T. J. and Elith, J.: A comparison of resampling methods for remote sensing classification and accuracy assessment, Remote Sens. Environ., 208, 145–153, doi:10.1016/j.rse.2018.02.026, 2018.

Marmanis, D., Schindler, K., Wegner, J. D., Galliani, S., Datcu, M. and Stilla, U.: Classification with an edge: Improving semantic image segmentation with boundary detection, ISPRS J. Photogramm. Remote Sens., 135, 158–172, doi:10.1016/j.isprsjprs.2017.11.009, 2018.

Noh, H., Hong, S. and Han, B.: Learning deconvolution network for semantic segmentation, in 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1520–1528, IEEE, Santiago, Chile., 2015.

Padilla, M., Stehman, S. V. and Chuvieco, E.: Validation of the 2008 MODIS-MCD45 global burned area product using stratified random sampling, Remote Sens. Environ., 144, 187–196, doi:10.1016/j.rse.2014.01.008, 2014.

Pathak, D., Krahenbuhl, P. and Darrell, T.: Constrained convolutional neural networks for weakly supervised segmentation, in 2015 IEEE International Conference on Computer Vision, pp. 1796–1804, IEEE, Santiago, Chile., 2015.

Pinheiro, P. O., Lin, T. Y., Collobert, R. and Dollár, P.: Learning to refine object segments, in european conference on computer vision, pp. 75–91, Springer Nature, Amsterdam, The Netherlands., 2016.

Porrello, S., Tomassetti, P., Manzueto, L., Finoia, M. G., Persia, E., Mercatali, I. and Stipa, P.: The influence of marine cages on the sediment chemistry in the Western Mediterranean Sea, Aquaculture, 249(1–4), 145–158,

doi:10.1016/j.aquaculture.2005.02.042, 2005.

POWERS, D. M. W.: Evaluation: From precision, recall and f-measure to roc, informedness, markedness and correlation, J. Mach. Learn. Technol., 2(1), 37–63, doi:10.1.1.214.9232, 2011.

475   Ramezan, C. A., Warner, T. A. and Maxwell, A. E.: Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification, Remote Sens., 11(2), 185, doi:10.3390/rs11020185, 2019.

Rigos, G. and Katharios, P.: Pathological obstacles of newly-introduced fish species in Mediterranean mariculture: A review, Rev. Fish Biol. Fish., 20(1), 47–70, doi:10.1007/s11160-009-9120-7, 2010.

Ronneberger, O., Fischer, P. and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, in Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015), pp. 234–241,

480   Munich, Germany., 2015.

Rubio-Portillo, E., Villamor, A., Fernandez-Gonzalez, V., Antón, J. and Sanchez -Jerez, P.: Exploring changes in bacterial communities to assess the influence of fish farming on marine sediments, Aquaculture, 506, 459–464, doi:10.1016/j.aquaculture.2019.03.051, 2019.

Shi, T., Xu, Q., Zou, Z. and Shi, Z.: Automatic Raft Labeling for Remote Sensing Images via Dual-Scale Homogeneous

485   Convolutional Neural Network, Remote Sens., 10, 1130, doi:10.3390/rs10071130, 2018.

Tovar, A., Moreno, C., Mánuel -Vez, M. P. and García -Vargas, M.: Environmental impacts of intensive aquaculture in marine waters, Water Res., 34(1), 334–342, doi:10.1016/S0043-1354(99)00102-5, 2000.

Wang, M., Cui, Q., Wang, J., Ming, D. and Lv, G.: Raft cultivation area extraction from high resolution remote sensing imagery by fusing multi-scale region-line primitive association features, ISPRS J. Photogramm. Remote Sens., 123(1),

490   104– 113, doi:10.1016/j.isprsjprs.2016.10.008, 2017.

Xiao, L., Haijun, H., Xiguang, Y. and Liwen, Y.: Method to extract raft-cultivation area based on SPOT image, Sci. Surv. Mapp., 38(2), 41–43, doi:10.16251/j.cnki.1009-2307.2013.02.033, 2013.

Yim, J., Joo, D., Bae, J. and Kim, J.: A gift from knowledge distillation: Fast optimization, network minimization and transfer learning, in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1063–6919,

495   IEEE, Honolulu, HI, USA., 2017.

Zhang, L., Zhang, L. and Du, B.: Deep learning for remote sensing data: A technical tutorial on the state of the art, IEEE Geosci. Remote Sens. Mag., 4(2), 22–40, doi:10.1109/MGRS.2016.2540798, 2016a.

Zhang, X., Zou, J., He, K. and Sun, J.: Accelerating Very Deep Convolutional Networks for Classification and Detection, IEEE Trans. Pattern Anal. Mach. Intell., 38(10), 1943–1955, doi:10.1109/TPAMI.2015.2502579, 2016b.

500   Zhang, Y., Qiu, Z., Yao, T., Liu, D. and Mei, T.: Fully Convolutional Adaptation Networks for Semantic Segmentation, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 3431–3440, Boston, MA, USA., 2015.

Zhao, H., Shi, J., Qi, X., Wang, X. and Jia, J.: Pyramid scene parsing network, in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6230–6239, IEEE, Honolulu, HI, USA., 2017.

505   Zhao, W. and Du, S.: Learning multiscale and deep representations for classifying remotely sensed imagery, ISPRS J. Photogramm. Remote Sens., 113, 155–165, doi:10.1016/j.isprsjprs.2016.01.004, 2016.

Zheng, Y., Wu, J., Wang, A. and Chen, J.: Object-and pixel-based classifications of macroalgae farming area with high spatial resolution imagery, Geocarto Int., 1–16, doi:10.1080/10106049.2017.1333531, 2017.