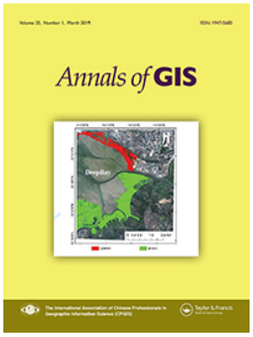

Annals of GIS

ISSN: 1947-5683 (Print) 1947-5691 (Online) Journal homepage: https://www.tandfonline.com/loi/tagi20

A small-patched convolutional neural network for mangrove mapping at species level using high-resolution remote-sensing image

Luoma Wan, Hongsheng Zhang, Guanghui Lin & Hui Lin

To cite this article: Luoma Wan, Hongsheng Zhang, Guanghui Lin & Hui Lin (2019) A small-patched convolutional neural network for mangrove mapping at species level using high-resolution remote-sensing image, Annals of GIS, 25:1, 45-55, DOI: 10.1080/19475683.2018.1564791

To link to this article: https://doi.org/10.1080/19475683.2018.1564791

Published online: 09 Jan 2019.

ANNALS OF GIS
2019, VOL. 25, NO. 1, 45–55
https://doi.org/10.1080/19475683.2018.1564791

Taylor & Francis
Taylor & Francis Group

A small-patched convolutional neural network for mangrove mapping at species level using high-resolution remote-sensing image

Luoma Wan[a], Hongsheng Zhang[a,b], Guanghui Lin[c] and Hui Lin[a,b,d]

[a]Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong; [b]Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China; [c]Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China; [d]Department of Geography and Resource Management, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

ABSTRACT

The understanding of mangrove forest structure and dynamics at species level is essential for mangrove conservation and management. To classify mangrove species, remote-sensing technologies provide a better way with high spatial resolution image. The spatial structure is usually viewed as effective complementary information for classification. However, it is still a challenge to design handcrafted features for mangrove species due to their non-structure texture. To leverage the advantage of convolutional neural networks (CNNs) in abstract feature exploration, a small patch-based CNN is proposed to overcome the requirement of fixed and large input which limits the applicability of CNNs to fringe mangrove forests. The function of down-sampling technology was substantially reduced to make deeper network for small input in our work. Meanwhile, the inception structure is used to exploit the multi-scale features of mangrove forests. Furthermore, the network is optimized with lesser convolution kernels and a single fully connected layer to reduce overfitting via reducing the training parameters. Compared to the features of grey level co-occurrence matrix with support vector machine, our proposed CNN shows better performance in classification accuracy. Moreover, the differences between mangrove species can be perceptive via CNN visualization, which offers better understanding of mangrove forests.

ARTICLE HISTORY
Received 14 July 2018
Accepted 17 December 2018

KEYWORDS
Small patch; mangrove species; CNNs

Introduction

Mangrove forests are the most productive ecosystems in the world with providing essential and unique ecological goods and services to human beings and adjacent systems including coastline and flood protection, shrimp farming and habitat provision for various terrestrial, estuarine and marine species (Giri et al. 2011; Wang, Sousa, and Gong 2004; Polidoro et al. 2010). Dominating the intertidal zone of tropical and subtropical coastlines, different mangrove species adapting to specific environment are distributed parallel to the coast or riverine system. Usually, they form distinct zones to be 'foundation species', which can control population and ecosystem dynamic, including fluxes of energy and nutrients, hydrology, food webs, and biodiversity (Polidoro et al. 2010; Ellison et al. 2005; Claridge and Burnett 1993). However, mangrove forests decline rapidly in both area and species (Polidoro et al. 2010; Giri et al. 2008), resulting from serious human intervention and environment pollution. Thus, mangrove forests monitoring at species level can help understand their structure and dynamics, and then offer guidance for the management, such as conservation, assessment of mangrove reconstruction, and invasive species monitoring, etc (Vaiphasa, Skidmore, and de Boer 2006).

Remote sensing provides a sufficient tool for mangrove monitoring over a large scale (Jia et al. 2016). In terms of monitoring mangrove at species level, the median–low resolution remote-sensing images like Landsat TM are not appropriate due to coarser resolution (Wang, Sousa, and Gong 2004). High-resolution remote-sensing images offer a cost-efficient way with more detailed information. Still, some serious challenges remain, such as the spectral similarity of mangrove and associate species and unclear zonation between species, especially when vegetation is sparse or degraded (Heumann 2011). The exploratory experiment has indicated that individual mangrove species could not be separated only using spectral features (Heumann 2011; Blasco et al. 1998), and more information or new methods are needed (Ji et al. 2008). Textural characteristics (e.g. grey-level co-occurrence matrix or lacunarity) of the canopy and leaves are the main complementary features for mangrove community's distinction (Kuenzer et al. 2011; Wang et al. 2015, 2004; Myint et al. 2008). Compared to pixel-based classification, object-based classification (Myint et al. 2008;

CONTACT Hongsheng Zhang stevenzhang@cuhk.edu.hk

Liu et al. 2007; Wan et al. 2014), integration pixel-based and object-based classification (Wang, Sousa, and Gong 2004), fuzzy classification (Neukermans et al. 2008) and NDVI-based classification (Valderrama-Landeros et al. 2018) were also applied for mangrove species mapping. A principle behind them is that more robust features considering spatial information (the affinity of a pixel and its neighbours) were developed to improve mangrove species mapping, which is also indicated in Vaiphasa, Skidmore, and de Boer (2006; Heumann 2011; Myint et al. 2008). However, these features usually were handcrafted and in low-level, which are suitable for the objects with a clear edge, shape, and texture. For forests with non-structure features, an advanced feature or representation for accurate mangrove species mapping is needed.

Deep convolutional neural networks (DCNNs) have been prevalent in image processing because of its strong ability of learning representative and discriminative features in a hierarchical way. Stacking convolutional layers and pooling layers, high-level features can be further learnt. The first and last convolutional layers were usually visualized to explain what kind of features were learnt and why they are so effective for classification (Zeiler and Fergus 2014). Also, it has been widely used for remote-sensing applications. Four main categories were found according to 190 records, including object detection (vehicle detection Chen et al. 2014; Li et al. 2017b; Qu, Zhang, and Sun 2017, aircraft detection Pan et al. 2017; Wang et al. 2017; Zhang et al. 2018, 2016, ship detection Zou and Shi 2016; Yao et al. 2017; Lin, Shi, and Zou 2017, road detection Cheng et al. 2017, building detection Guo et al. 2016; Tian et al. 2017, airport Xiao et al. 2017; Zhang et al. 2017; Cai et al. 2017), object recognition (ocean front Lima et al. 2017, building shape Tian et al. 2017 and others Liu and Zheng 2017), scene classification (Hu et al. 2015; Zhong, Fe, and Zhang 2016; Li et al. 2017a) and semantic segmentation. These studies mainly focus on urban with appreciable features, like regular shape, clear edge and structure texture. However, few studies about mangrove forests taking advantage of their ability in feature extraction were found.

Mangrove species classification is a task of semantic segmentation (pixelwise classification in Maggiori et al. 2017), which is to assign a label to each pixel in an image. Usually, CNN models, such as AlexNet (Krizhevsky, Sutskever, and Hinton 2012), GoogleNet (Szegedy et al. 2015), VGG, and ResNet (He et al. 2016), consist of stacked convolutional layers, followed by fully connected layer to output a single label for the input. An image patch centred at a pixel is fed on CNN models, and the output is the for the pixel, then a strategy of 'sliding window' is used to get the labels for all pixels in an image (Marmanis et al. 2016; Liu et al. 2017). In addition, the Fully Convolutional Network (FCN) is another option for semantic segmentation (Lin, Shi, and Zou 2017; Long, Shelhamer, and Darrell 2015). Instead of fully connected layers, FCN employs deconvolutional layers (transposed convolutional layers in Liu et al. 2017) to decode a full-resolution label map from a reduced convolutional layer, providing an end-to-end way for classification. Generally, FCN rather than prior CNNs is viewed as a better choice for semantic segmentation in computation and performance. However, FCNs do not leverage the spatial information. More importantly, the training samples of ground truth for FCN are tough to obtain for mangrove because of the difficulties in outlining the boundary of different mangrove species. Hence, this paper adopts CNNs rather than FCNs. For fringe mangrove, the fixed and large input size of CNNs, such as the size of 256 for AlexNet, becomes a serious problem, such a large homogenous patch is difficult to obtain and the severe overlap between adjacent image patches will seriously reduce the model's performance.

In this paper, a small-patched CNN is proposed for mangrove mapping at species level. The main objective of this paper is twofold:

(1) A novel CNNs is designed for the need of smaller input patch which classifies the mangrove forests using CNNs. Tailored for strip-like zone of mangrove forests, the stride of 1 for convolution layers and pooling layers is adapted to keep the size of feature maps constant and make deeper network possible. Also, the inception structure from GoogLeNet is used to exploit multi-scale feature;
(2) The ability of CNNs in feature representation of mangrove forests was investigated.

Study sites and preprocessing

Study sites

DeepBay is a special place where there are two famous mangrove reservation zones, Shenzhen Mangrove Forest Nature Reserve (SZMFNR) and Mai Po Marshes Nature Reserve (MPMNR). SZMRNR (22°30′–22°32′ N and 113°59′–114°03′ E) is in the northeast region of the DeepBay, while the MPMNR (22°28′–22°31′ N and 113°59′–114°04′ E) lies on the opposite shores of Deep Bay in New Territories, Hong Kong. Both are well protected and play an important role in ecosystem balance.

Mangrove forests in SZMFNR are about 100 ha. The climate here is tropical zone maritime climate with annual

precipitation of 1962 mm and average temperature of 22.5°C (Congjiao et al. 2016). Governed by the continent of mainland China, in MPMNR, the summer is hot and humid with winds from the south and southeast, while winter is cool and dry because of the continental air streams from the north and northeast (Jia et al. 2014). Because of well protection by Hong Kong government since 1975, mangrove forests in MPMNR are the largest one with area of 170 ha in Hong Kong (Wang et al. 2015).

The dominant species in SZMFNR include *Avicennia marina* (AM), *Kandelia obovata* (KO), *Sonneratia apetala* (SA) and *Sonneratia caseolaris* (SC). In 1993, SC and SA from Hainan and Bangladesh are brought in to reconstruct the mangrove forests. Floating through the DeepBay, SA and SC can be firstly found in MPMNR in 2000, but they are viewed as invasive species due to their high adaptability and threat of potential colonization and have been consistently removed by the Agriculture, Fisheries and Conservation Department (AFCD) to protect native species (Peng 2003; Ren et al. 2009; Wong and Fung 2014). In addition, four dominant species including *Avicennia marina, Acanthus ilicifolius* (AI), *Kandelia obovata*, and *Aegiceras corniculatum* (AC) were found during our fieldwork (Wang et al. 2015), which is different from the report of eight species of mangroves in this area (AFCD 2015).

Image preprocessing

A WorldView 2 standard image (level 2) covering SZMFNR and MPMNR was acquired on 14 November 2010. It was delivered in a geo-registered UTM/WGS84 projection with a 16-bit depth standard ENVI format with basic preprocessing including radiation correction, geometrical reference and ortho rectification, resulting in eight resampled multispectral bands and a panchromatic band with spatial resolution of 2 m and 0.5 m. For generalization in the input image requirement, general bands including blue, green, red, and near-infrared (NIR) band were used in this paper. Since we focus on mangrove species differentiation, the main regions growing mangrove forests in both reserves (Figure 1) are depicted manually based on coarse classification using NDVI with a threshold of 0.2.

Field survey and sample collection

The field survey for MPMNR was conducted on 10 November 2015. Due to the strict protection of mangrove forest and limited access policy, collecting reference data in MPMNR is extremely difficult. Hence, the samples collected in MPMNR are mainly via visual interpretation through panchromatic band of WorldView 2 data and GoogleEarth; the samples from other studies are also used for reference. In SZMFNR,

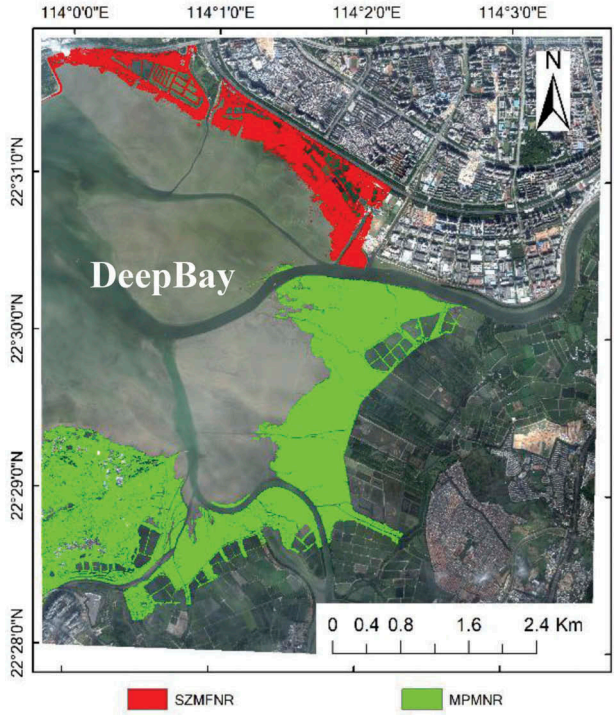

Figure 1. WorldView 2 image (true color) of study area and mangrove regions.

the *in-situ* samples were collected on 11 April 2017. Thanks to the guidance under the local manager, the data collected are relatively easy and trusted. Finally, reference samples of seven classes (six mangrove species and the other vegetation, Figure 2) with around 1100 pixels for each were randomly collected. To balance the training samples, the dominant species with much more samples are cut down and the final number for training and testing can be found in Table 1. Although the time of data collection is far behind the day of image acquisition, it is still acceptable to assign the data labels to the image since both reserves are well protected, and no strong human disturbance and no destroyed disaster happen during these years. The effect of the SA clearance from AFCD on the cover change can be easily avoided because the trace is very clear in image with high spatial resolution.

Methods

Generally, a typical CNN contains five parts: input, convolutional layers, pooling layers, fully connected layers and output. The change of input usually leads to the change of the whole CNN architecture; however, no any evidence shows that a rule can be followed to design a new CNN in a new domain, but follows the classic ones. Firstly, the proposed architecture of small patch-based CNNs is

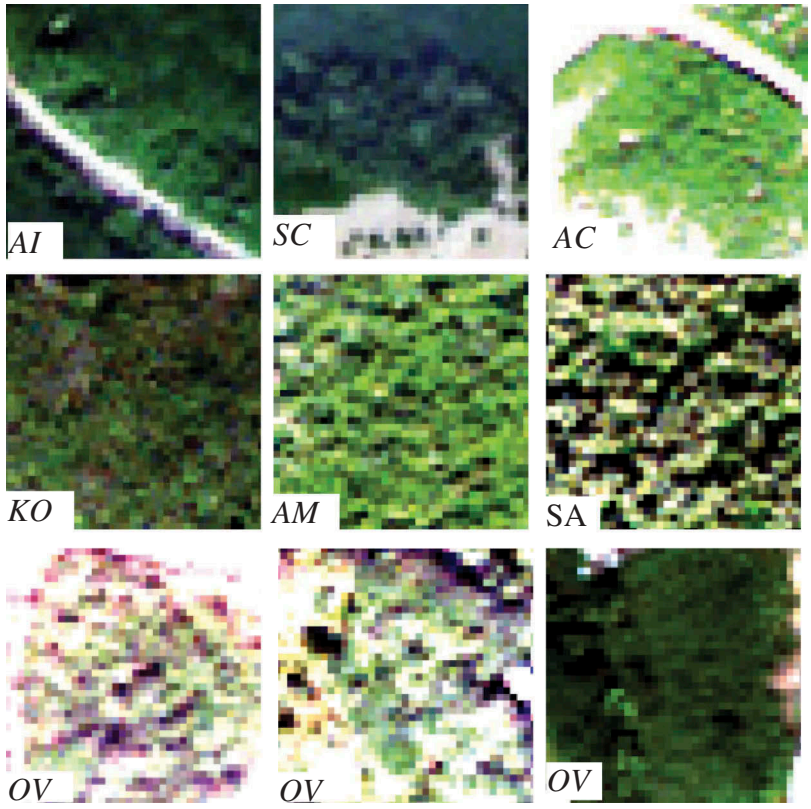

Figure 2. Samples of seven classes in WV2. AI, Acanthus ilicifolius; SC, Sonneratia caseolaris; AC, Aegiceras corniculatum; KO, Kandelia obovate; AM, Avicennia marina; SA, Sonneratia apetala; OV*, other vegetation (non-mangrove).

Table 1. Samples of six mangrove species and other vegetation.

Class	Training number	Testing number
AI	824	354
SC	823	354
AC	844	353
KO	845	353
AM	845	354
SA	844	354
OV*	736	316

AI, Acanthus ilicifolius; SC, Sonneratia caseolaris; AC, Aegiceras corniculatum; KO, Kandelia obovate; AM, Avicennia marina; SA, Sonneratia apetala; OV*, other vegetation (non-mangrove)

presented (Figure 3), and then some of the novel or unusual parts of our CNNs are described below.

Pooling layers

Pooling layers inserted in-between successive convolution layers in CNNs mainly function to reduce the dimension of feature maps to reduce parameter number and computation, and hence to control overfitting (CS231n). However, it becomes a problem and limits the network depth for the input with small patches, because the reduction in feature maps' dimension possibly causes no sufficient size for pooling operator in latter part of CNNs. Precisely, pooling operator summarizing a neighbourhood centred at the location of the pooling unit does not make any change to the output unless the moving stride of pooling operator is greater than 1. In Basaeed, Bhaskar, and Al-Mualla (2016), a CNN architecture with no pooling layers was proposed and convolution layers with a stride achieves the same function of sub-sampling. In our work, we adapt pooling layers with a stride of 1 to keep the feature maps' dimension and then make deeper network possible. In addition, it may improve performance and help reduce overfitting (Krizhevsky, Sutskever, and Hinton 2012). The reasons why keep pooling layers rather than remove them are as follows: (1) max or average pooling can help with anti-noise, especially for high spatial resolution image with high intraclass variance; (2) pooling layers will not add extra trainable parameters for CNNs; (3) less modification makes classic CNN architecture adaptation easy and further transfer learning possible.

Convolution layers

Inspired from the cells in cat's visual cortex, which are sensitive to small sub-regions of the visual field (also

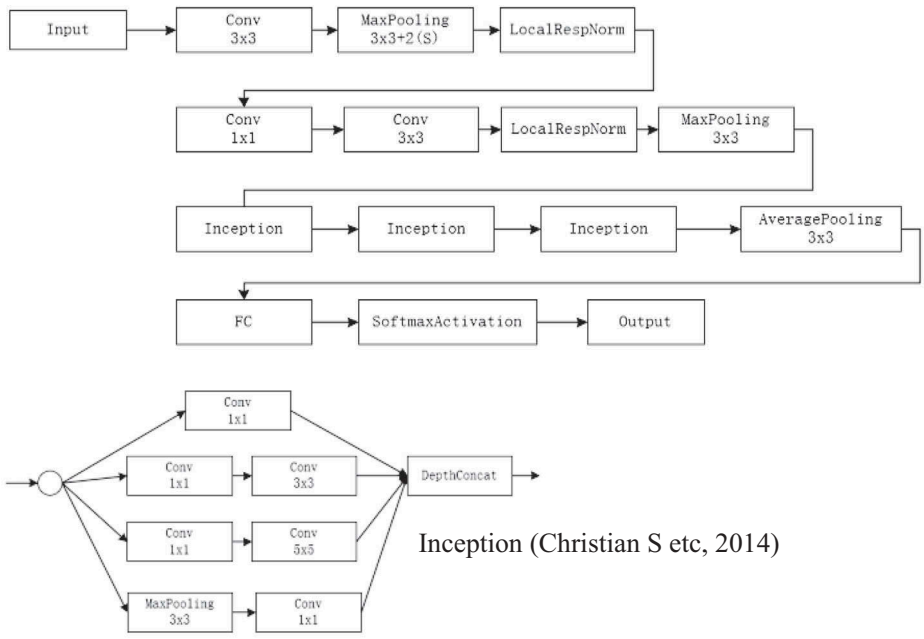

Figure 3. The architecture of proposed CNN.

called a receptive field), the convolution layers acting as local filters over the entire visual field were introduced to exploit the strong spatially local correlation in natural images (deep learning tutorial, LISA lab, University of Montreal). To capture multi-scale features, the inception structure with multi-convolution kernels with different sizes are adopted. Due to the limitation of input size, the kernels with size of 3 by 3 and 5 by 5 were applied in convolutional layers.

Reduce overfitting

Transfer learning and pre-trained CNNs are valid ways to reduce overfitting for applications in new domains. In our work, we train the network from scratch because of the change in CNN architecture and the lag between different model domains. For so many trainable parameters (e.g. 60 million parameters for AlexNet), data augmentation (Krizhevsky, Sutskever, and Hinton 2012) and dropout (Krizhevsky, Sutskever, and Hinton 2012; Szegedy et al. 2015) are common tricks. A light CNN with less parameters here is needed in consideration of the following factors: (1) the target classes of mangrove species mapping is much less than 1000 classes of applications in computer vision; (2) the dataset for mangrove forests is not enormous even with the help of data augmentation like random clip, vertical/horizontal reflection transformation.

The CNNs go deeper and deeper after the comparable result to human beings got by AlexNet, especially after the residual network was developed. However, many effective features for mangrove species separation rather than an optimal CNN are the result we aim for. Hence, a CNN with less convolution layers is acceptable only if there is a minor cost of performance. Therefore, the stacked convolution layers with inception structures and fully connected layers from AlexNet and GoogLeNet mainly construct our CNNs. Furthermore, the fully connected layers were cut down but kept only one to reduce the training parameters based on the evidence that fully connected layers contain larger number of parameters than convolution layers (Szegedy et al. 2015; He et al. 2016). Finally, the number of convolution kernels and neurons in fully connected layers is also cut down in view of the species of mangrove is far less than 1000 for ImageNet.

Results

Through trials of layer construction, the network of two convolution layers with 64 and 192 kernels, six inception layers with 64, 120, 128, 132, 208, and 256 kernels, and a single fully connected layer with 100 neuros was adapted, resulting in 4.64 million parameters to be trained. ReLU was used as the activation function. The largest input size was reduced to 15 pixels by 15 pixels, because the width of narrow fringe mangrove forests with a homogenous species is from 18 m to 30 m and the inputs with larger size will cause severe 'margin-effect' (Wan et al. 2018). The random selection of pixels to generate small patches as the training samples functions random crop. Consequently, only vertical/horizontal reflection transformation was used to enlarge

training samples. To reduce overfitting, dropout with value of 0.4, the tech of early stop, and the batch size of 20 were also adapted in our work. The whole experiment was conducted in Python based on Keras with the backend of tensorflow. It performs over the platform of windows 10 OS with a core processor of i7-7700K, and 32 GB memory. A GTX 1070 display card with 8 GB display memory was assisted in training acceleration.

Mangrove mapping at species level

When train the small-patched CNN, the epoch was set to 60 from which it starts to converge for different settings. The input patches with size of 9 and 13 were also fed for training to support the selection of 15 pixels by pixels for our CNN (Figure 4). Clearly, training with input size of 15 can converge soon with minor loss, and the procedure stopped at epoch of around 27.

To assess the performance of small-patched CNN, the grey level co-occurrent matrix (GLCM) features as well as the spectral features with support vector machine (SVM) were used for comparison. For GLCM, the processing window was set to 15 × 15 and the co-occurrence shift was set to 1 × 1 to be consistent with the kernel used in the small-patched CNN. The grey-scale quantization levels of 64 were adapted to compute the texture measures of contrast, correlation, homogeneity and entropy (Wang et al. 2015; Wan et al. 2018). For SVM, the kernel of RBF with gamma of 0.25 was used as well as the default values for the rest parameters in Envi. In addition, GoogLeNet from which the small-patched CNNs are derived was also applied for comparison.

From the confusion tables (Table 2), the overall accuracy and kappa coefficient from our method is much higher

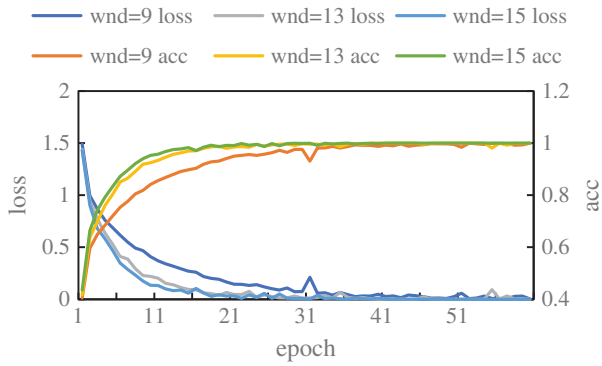

Figure 4. Training with different sizes for input patches; 'wnd' is abbreviated for 'window' and the value is the side length of a square window, the 'acc' is for 'accuracy'.

than that from GoogLeNet with over accuracy of 17.17%. Nothing but chaos can be observed from the classification result (Figure 5). It is mainly because a large window with size of 224 by 224 pixels in GoogLeNet introduces abundant irrelated information, weakening the contribution from the valid information of narrow and strip-like mangrove forests. In addition, little information of mangrove forests within the larger windows gradually vanished when it passed through the pooling layers in GoogLeNet. A worse situation is that invalid information flows through the latter layers in GoogLeNet, making results out of control. Inspired from the limitation in the requirement of a large window and information loss in pooling layers, our small-patched CNNs were proposed.

Compared to the conventional features (GLCM), our model gets higher overall accuracy of 98.8105% and kappa coefficient of 0.986064, about 30% improvement. Using our proposed CNN (Figure 5(a)), the strip feature of mangrove can be clearly perceived, while that using

Table 2. Confusion matrix of mangrove classification using different methods.

	GoogLeNet							
	AI	SC	AC	KO	AM	SA	OV*	Total
AI	63	43	48	38	60	46	23	321
SC	15	35	40	17	16	13	8	144
AC	25	32	45	23	38	35	19	217
KO	103	68	79	93	86	91	63	583
AM	47	54	41	67	37	40	53	339
SA	67	50	64	78	91	93	59	502
OV*	34	30	21	34	26	25	30	200
Total	354	312	338	350	354	343	255	2306

	SVM+GLCM							
	AI	SC	AC	KO	AM	SA	OV*	Total
AI	274	0	144	12	7	13	9	459
SC	0	309	7	34	8	4	42	404
AC	49	8	119	11	5	52	30	274
KO	3	11	20	9	310	15	10	378
AM	19	1	31	46	13	207	11	328
SA	7	4	19	219	10	59	59	377
OV*	2	21	13	23	0	4	155	218
Total	354	354	353	354	353	354	316	2438

Ours

	AI	SC	AC	KO	AM	SA	OV*	Total
AI	352	0	0	2	0	0	0	354
SC	0	342	0	0	0	12	0	354
AC	2	0	353	0	0	1	0	356
KO	0	0	0	351	0	0	0	351
AM	0	0	0	0	354	0	0	354
SA	0	12	0	0	0	341	0	353
OV*	0	0	0	0	0	0	316	316
Total	354	354	353	353	354	354	316	2438

	Overall accuracy	Kappa coefficient
GoogLeNet	17.17%	0.0295
SVM+GLCM	65.3404%	0.5952
Ours	98.8105%	0.986064

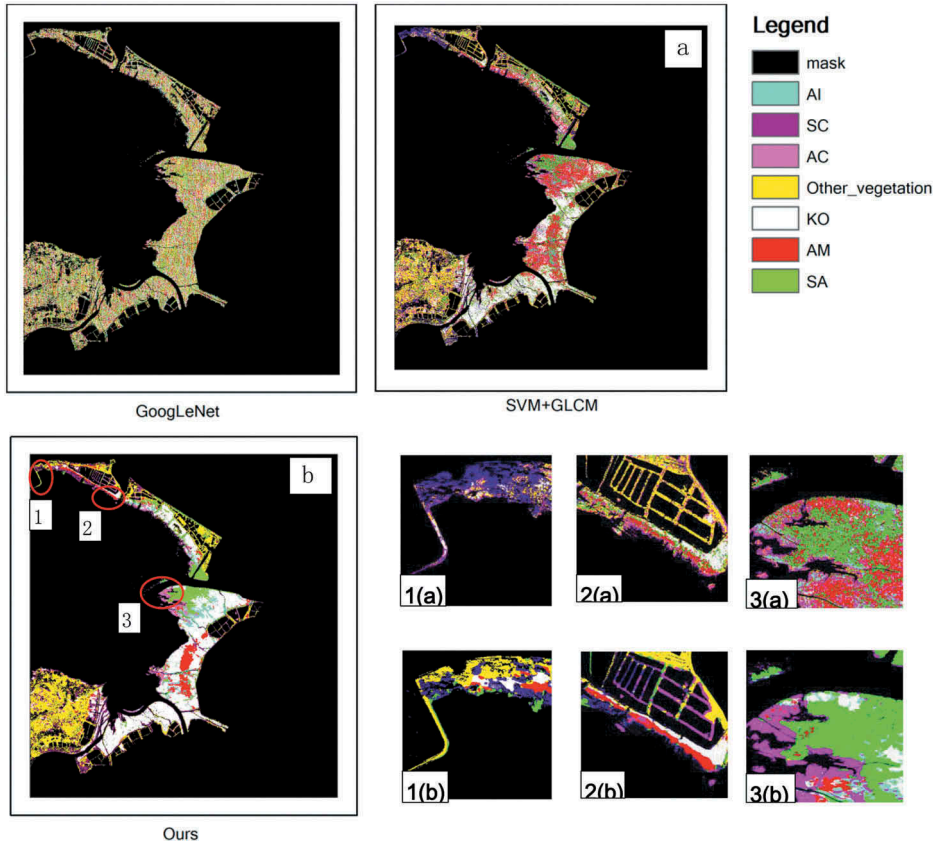

Figure 5. Mangrove classification at species level using different methods, and the main differences between SVM with GLCM features and ours in three regions: landward mixed region of non-vegetation and different mangrove species (1); narrow striped region with a clear feature of 'zonation' (2); seaward region with many creeks (3); the details can be found from the right parts.

handcrafted features shows a serious phenomenon of 'pepper-noise'. It is caused by the vast variance of inter-class, which is a side effect coupled with detailed structure information brought by high spatial resolution images (Figure 5 (b)). The main differences between them can be found in three **parts**: first, the landward region of SZMFNR (Figure 5 (1)) where non-mangrove and SC are mixed; they are easily confused (Figure 5(1a)) when using GLCM and spectral features, but ours can resolve this problem (Figure 5(1b)). The second part is the narrow-striped region where different mangrove species grow with a clear feature of 'zonation'. Three different mangrove species can be identified clearly by our CNN, but they are mixed using GLCM (Figure 5(2)). The last one is the seaward region where ACs grow along the creeks. Using GLCM with spectral features, ACs are submerged by surroundingswhile ours can recognise them all (Figure 5(3)).

CNN features visualization

To explore the features extracted by our model, the filters were visualized. Considering the small filters in our model, instead of visualizing the filters directly, we here visualize the inputs that maximize the activation of the filters; the details can be found in Chollet (2016). The filters in first convolution layer were visualized to expect to get low-level features of different mangrove species. Unlike the previous works on CNN visualization in which there are clear low-level features such as the edges and direction extracted by the first convolution layers, there are more features exploited by our CNNs, and they can be observed from visualization of the first convolution layer (Figure 6). The first 32 filters mainly focus on the colour with little difference in texture of 'pepper noise', while some of the rest filters (patches = {(5,2), (5,5), (5,6),(6,2),(6,8)}) exhibit the 'dot matrix' with directionality, showing the ability of individual tree detection, especially for the seventh filter in row 5, the outline of trees can be observed. It indicates that for mangrove forests discrimination, the low features of minor difference in colour still make contributions. Instead of clear edge, the low features of 'dots' presented by trees also play an important role. In addition, the results from the last filter in row 5 and first filter in row 6 possibly give us a clue that the external environment can offer extra information for mangrove

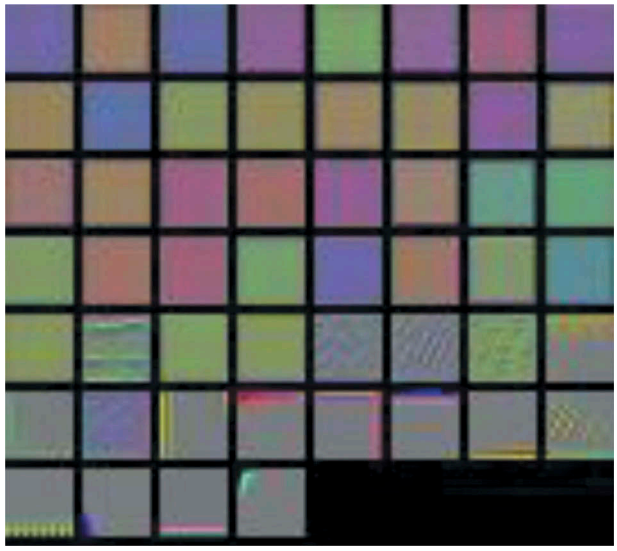

Figure 6. Visualization of first convolution layer (48 of 64 filters were listed).

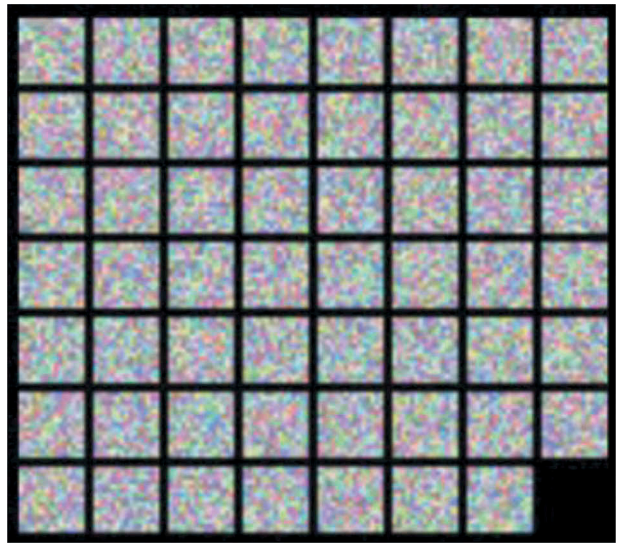

Figure 7. Visualization of the output of last inception (56 of 1024 filters were listed).

species mapping. These features may be the reasons why our model outperforms better than GLCM and spectral features with SVM.

To get insight into the high-level features exploited by our CNNs, the output of the final inception was also visualized (Figure 7). Based on the result, no distinguished features but chaos were observed, meaning that the high-level features are statistical features rather than structure feature, which is unexpected. The statistical features resulting from the distinguished low-level features indicate that the micro-feature for mangrove species is statistical, but this conclusion should be supported by more extra experiments with different visualization technologies, because different visualization technologies have different visual effects and whether the capacity of our model can be completely exposed via visualization is still unknown. However, the statistical features from ours are better than those from GLCM and spectral features.

Conclusion and discussion

This paper explores the ability of CNNs in mangrove species mapping and assesses the performance in feature extraction. Because of the requirement of large input size in conventional CNNs which is invalid in the classification of mangrove forests with a stripe zonation, a small-patched CNN with multi-scale features extractor was proposed. According to the results, the small-patched CNNs can give an acceptable classification than the result with handcrafted features. Because of the characteristic of small patch for input, the stripe region can be well detected. The multi-scale feature extractor helps with the fragmentation reduce. In addition, the low- and high-level features extracted by small-patched CNNs were also exploited. The experiments show that our model can capture many distinguished low-level features including colours, dot patterns and directionalities, and high-level statistical features. Compared to GLCM and spectral features, both are statistical, but ours are better.

Nevertheless, aimed for the applicability of conventional CNNs to mangrove species classification, some work, like the hyperparameter settings and more methods of reducing overfitting, was not perfected. Although some tricks were adapted to reduce overfitting, this problem did not resolve in our work possibly because of insufficiency in data or more parameters to be trained, which will be done in the future.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was jointly supported by National Natural Science Foundation of China [41401370], the Research Grants Council (RGC) General Research Fund [CUHK 14601515, CUHK 14635916 and CUHK 14605917].

References

AFCD. 2015. *Local Mangrove Species*. Hong Kong: Agriculture, Fisheries and Conservation Department (AFCD).

Basaeed, E., H. Bhaskar, and M. Al-Mualla. 2016. "Supervised Remote Sensing Image Segmentation Using Boosted Convolutional Neural Networks." *Knowledge-Based Systems* 99: 19–27. doi:10.1016/j.knosys.2016.01.028.

Blasco, F., T. Gauquelin, M. Rasolofoharinoro, J. Denis, M. Aizpuru, and V. Caldairou. 1998. "Recent Advances in Mangrove Studies Using Remote Sensing Data." *Marine and Freshwater Research* 49: 287–296. doi:10.1071/MF97153.

Cai, B. W., Z. G. Jiang, H. P. Zhang, D. P. Zhao, and Y. Yao. 2017. "Airport Detection Using End-to-End Convolutional Neural Network with Hard Example Mining." *Remote Sensing* 9 (Nov.): 20. doi:10.3390/rs9111198.

Chen, X. Y., S. M. Xiang, C. L. Liu, and C. H. Pan. 2014. "Vehicle Detection in Satellite Images by Hybrid Deep Convolutional Neural Networks." *IEEE Geoscience and Remote Sensing Letters* 11 (Oct.): 1797–1801. doi:10.1109/LGRS.2014.2309695.

Cheng, G. L., Y. Wang, S. B. Xu, H. Z. Wang, S. M. Xiang, and C. H. Pan. 2017. "Automatic Road Detection and Centerline Extraction via Cascaded End-to-End Convolutional Neural Network." *IEEE Transactions on Geoscience and Remote Sensing* 55 (Jun.): 3322–3337. doi:10.1109/TGRS.2017.2669341.

Chollet, F. 2016. "How Convolutional Neural Networks See the World." Accessed https://blog.keras.io/how-convolutional-neural-networks-see-the-world.html

Claridge, D., and J. Burnett. 1993. *Mangroves in Focus*. Ashmore, QLD: Wet Paper Publications.

Congjiao, P., Q. Jiawei, G. Xudong, Z. Hewei, H. Naxu, Y. Qiong, C. Changpin, and C. Luzhen. 2016. "Vegetation Carbon Stocks and Net Primary Productivity of the Mangrove Forests in Shenzhen, China." *Chinese Journal of Applied Ecology* 27 (Jul.): 2059–2065. doi:10.13287/j.1001-9332.201607.029.

Ellison, A. M., M. S. Bank, B. D. Clinton, E. A. Colburn, K. Elliott, C. R. Ford, D. R. Foster, et al. 2005. "Loss of Foundation Species: Consequences for the Structure and Dynamics of Forested Ecosystems." *Frontiers in Ecology and the Environment* 3: 479–486. doi:10.1890/1540-9295(2005)003[0479:LOFSCF]2.0.CO;2.

Giri, C., E. Ochieng, L. L. Tieszen, Z. Zhu, A. Singh, T. Loveland, J. Masek, and N. Duke. 2011. "Status and Distribution of Mangrove Forests of the World Using Earth Observation Satellite Data." *Global Ecology and Biogeography* 20 (Jan.): 154–159. doi:10.1111/j.1466-8238.2010.00584.x.

Giri, C., Z. Zhu, L. L. Tieszen, A. Singh, S. Gillette, and J. A. Kelmelis. Mar. 2008. "Mangrove Forest Distributions and Dynamics (1975–2005) of the Tsunami-Affected Region of Asia." *Journal of Biogeography* 35: 519–528. doi:10.1111/jbi.2008.35.issue-3.

Guo, Z. L., X. W. Shao, Y. W. Xu, H. Miyazaki, W. Ohira, and R. Shibasaki. 2016. "Identification of Village Building via Google Earth Images and Supervised Machine Learning Methods." *Remote Sensing* 8 (Apr.): 15. doi:10.3390/rs8040271.

He, K., X. Zhang, S. Ren, and J. Sun. 2016. "Deep Residual Learning for Image Recognition." In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 27–30, 770–778. Las Vegas, CA: IEEE.

Heumann, B. W. 2011. "An Object-Based Classification of Mangroves Using a Hybrid Decision Tree-Support Vector Machine Approach." *Remote Sensing* 3 (Nov.): 2440–2460. doi:10.3390/rs3112440.

Hu, F., G. S. Xia, J. W. Hu, and L. P. Zhang. 2015. "Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery." *Remote Sensing* 7 (Nov.): 14680–14707. doi:10.3390/rs71114680.

Ji, M. H., Y. M. Wu, Z. W. Deng, H. Q. Fan, and Z. Zhang. 2008. "Mapping Mangroves from High-Resolution Multispectral Imagery: Using Beilun Estuary, Guangxi, China as a Case Study." In *Remote Sensing and Modeling of Ecosystems for Sustainability V*, vol. 7083, edited by W. Gao and H. Wang. Bellingham: Spie-Int Soc Optical Engineering.

Jia, M. M., M. Y. Liu, Z. M. Wang, D. H. Mao, C. Y. Ren, and H. S. Cui. 2016. "Evaluating the Effectiveness of Conservation on Mangroves: A Remote Sensing-Based Comparison for Two Adjacent Protected Areas in Shenzhen and Hong Kong, China." *Remote Sensing* 8 (Aug.): 20. doi:10.3390/rs8080627.

Jia, M. M., Y. Z. Zhang, Z. M. Wang, K. S. Song, and C. Y. Ren. 2014. "Mapping the Distribution of Mangrove Species in the Core Zone of Mai Po Marshes Nature Reserve, Hong Kong, Using Hyperspectral Data and High-Resolution Data." *International Journal of Applied Earth Observation and Geoinformation* 33 (Dec.): 226–231. doi:10.1016/j.jag.2014.06.006.

Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. "Imagenet Classification with Deep Convolutional Neural Networks." *Advances in Neural Information Processing Systems* 1: 1097–1105.

Kuenzer, C., A. Bluemel, S. Gebhardt, T. V. Quoc, and S. Dech. 2011. "Remote Sensing of Mangrove Ecosystems: A Review." *Remote Sensing* 3 (May.): 878–928. doi:10.3390/rs3050878.

Li, E. Z., J. S. Xia, P. J. Du, C. Lin, and A. Samat. 2017a. "Integrating Multilayer Features of Convolutional Neural Networks for Remote Sensing Scene Classification." *IEEE Transactions on Geoscience and Remote Sensing* 55 (Oct.): 5653–5665. doi:10.1109/TGRS.2017.2711275.

Li, H., K. Fu, M. L. Yan, X. Sun, H. Sun, and W. H. Diao. 2017b. "Vehicle Detection in Remote Sensing Images Using Denoizing-Based Convolutional Neural Networks." *Remote Sensing Letters* 8: 262–270. doi:10.1080/2150704X.2016.1258127.

Lima, E., X. Sun, J. Y. Dong, H. Wang, Y. T. Yang, and L. P. Liu. 2017. "Learning and Transferring Convolutional Neural Network Knowledge to Ocean Front Recognition." *IEEE Geoscience and Remote Sensing Letters* 14 (Mar.): 354–358. doi:10.1109/LGRS.2016.2643000.

Lin, H. N., Z. W. Shi, and Z. X. Zou. 2017. "Fully Convolutional Network with Task Partitioning for Inshore Ship Detection in Optical Remote Sensing Images." *IEEE Geoscience and Remote Sensing Letters* 14 (Oct.): 1665–1669. doi:10.1109/LGRS.2017.2727515.

Liu, Y., D. Nguyen, N. Deligiannis, W. R. Ding, and A. Munteanu. 2017. "Hourglass-ShapeNetwork Based Semantic Segmentation for High Resolution Aerial Imagery." *Remote Sensing* 9 (Jun.): 24.

Liu, Y., and F. B. Zheng. 2017. "Object-Oriented and Multi-Scale Target Classification and Recognition Based on Hierarchical Ensemble Learning." *Computers & Electrical*

Engineering 62 (Aug.): 538–554. doi:10.1016/j.compeleceng.2016.12.026.

Liu, Z. G., J. Li, B. L. Lim, C. Y. Seng, and S. Inbaraj. 2007. "Object-Based Classification for Mangrove with VHR Remotely Sensed Image, Pts 1 and 2." In *Geoinformatics 2007: Remotely Sensed Data and Information*, vol. 67523C, edited by W. Ju and S. Zhao. Bellingham: Spie-Int Soc Optical Engineering.

Long, J., E. Shelhamer, and T. Darrell. 2015. "Fully Convolutional Networks for Semantic Segmentation." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 7–12, 3431–3440. Boston, MA: IEEE.

Maggiori, E., Y. Tarabalka, G. Charpiat, and P. Alliez. 2017. "Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification." *IEEE Transactions on Geoscience and Remote Sensing* 55 (Feb.): 645–657. doi:10.1109/TGRS.2016.2612821.

Marmanis, D., J. D. Wegner, S. Galliani, K. Schindler, M. Datcu, and U. Stilla. 2016. "Semantic Segmentation of Aerial Images with an Ensemble of CNNs." *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 3: 473–480. doi:10.5194/isprsannals-III-3-473-2016.

Myint, S. W., C. P. Giri, W. Le, Z. L. Zhu, and S. C. Gillette. 2008. "Identifying Mangrove Species and Their Surrounding Land Use and Land Cover Classes Using an Object-Oriented Approach with a Lacunarity Spatial Measure." *Giscience & Remote Sensing* 45 (Apr.–Jun.): 188–208. doi:10.2747/1548-1603.45.2.188.

Neukermans, G., F. Dahdouh-Guebas, J. G. Kairo, and N. Koedam. 2008. "Mangrove Species and Stand Mapping in Gazi Bay (Kenya) Using Quickbird Satellite Imagery." *Journal of Spatial Science* 53: 75–86. doi:10.1080/14498596.2008.9635137.

Pan, B., J. H. Tai, Q. Zheng, and S. S. Zhao. 2017. "Cascade Convolutional Neural Network Based on Transfer-Learning for Aircraft Detection on High-Resolution Remote Sensing Images." *Journal of Sensors* 2017: 14.

Peng, L. 2003. "The Characteristics of Mangrove Wetlands and Some Ecological Engineering Questions in China." *Engineering Science* 6: 004.

Polidoro, B. A., K. E. Carpenter, L. Collins, N. C. Duke, A. M. Ellison, J. C. Ellison, E. J. Farnsworth, et al. 2010. "The Loss of Species: Mangrove Extinction Risk and Geographic Areas of Global Concern." *PLoS One* 5: E10095. doi:10.1371/journal.pone.0010095.

Qu, T., Q. Y. Zhang, and S. L. Sun. 2017. "Vehicle Detection from High-Resolution Aerial Images Using Spatial Pyramid Pooling-Based Deep Convolutional Neural Networks." *Multimedia Tools and Applications* 76 (Oct.): 21651–21663. doi:10.1007/s11042-016-4043-5.

Ren, H., H. Lu, W. Shen, C. Huang, Q. Guo, Z. Li, and S. Jian. 2009. "Sonneratia Apetala Buch. Ham in the Mangrove Ecosystems of China: An Invasive Species or Restoration Species?„ *Ecological Engineering* 35: 1243–1248.

Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, et al. 2015. "Going Deeper with Convolutions." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 7–12, 1–9. Boston, MA: IEEE.

Tian, S., Y. Zhang, J. P. Zhang, and N. Su. 2017. "A Novel Deep Embedding Network for Building Shape Recognition." *IEEE Geoscience and Remote Sensing Letters* 14 (Nov.): 2127–2131. doi:10.1109/LGRS.2017.2753821.

Vaiphasa, C., A. K. Skidmore, and W. F. de Boer. 2006. "A Post-Classifier for Mangrove Mapping Using Ecological Data." *ISPRS Journal of Photogrammetry and Remote Sensing* 61 (Oct.): 1–10. doi:10.1016/j.isprsjprs.2006.05.005.

Valderrama-Landeros, L., F. Flores-de-Santiago, J. M. Kovacs, and F. Flores-Verdugo. 2018. "An Assessment of Commonly Employed Satellite-Based Remote Sensors for Mapping Mangrove Species in Mexico Using an NDVI-based Classification Scheme." *Environmental Monitoring and Assessment* 190 (Jan.): 13. doi:10.1007/s10661-017-6399-z.

Wan, H. W., Q. Wang, D. Jiang, J. Y. Fu, Y. P. Yang, and X. M. Liu. 2014. "Monitoring the Invasion of Spartina Alterniflora Using Very High Resolution Unmanned Aerial Vehicle Imagery in Beihai, Guangxi (China)." *Scientific World Journal* 2014: 7.

Wan, L. M., H. S. Zhang, T. Wang, G. Li, and H. Lin. 2018. "Mangrove Species Discrimination from Very High Resolution Imagery Using Gaussian Markov Random Field Model." *Wetlands* 38: 861–874. doi:10.1007/s13157-017-0925-1.

Wang, H. Z., Y. C. Gong, Y. Wang, L. F. Wang, and C. H. Pan. 2017. "DeepPlane: A Unified Deep Model for Aircraft Detection and Recognition in Remote Sensing Images." *Journal of Applied Remote Sensing* 11 (Sep.): 10. doi:10.1117/1.JRS.11.042606.

Wang, L., W. P. Sousa, and P. Gong. 2004. "Integration of Object-Based and Pixel-Based Classification for Mapping Mangroves with IKONOS Imagery." *International Journal of Remote Sensing* 25 (Dec.): 5655–5668. doi:10.1080/014311602331291215.

Wang, L., W. P. Sousa, P. Gong, and G. S. Biging. 2004. "Comparison of IKONOS and QuickBird Images for Mapping Mangrove Species on the Caribbean Coast of Panama." *Remote Sensing of Environment* 91 (Jun.): 432–440. doi:10.1016/j.rse.2004.04.005.

Wang, T., H. Zhang, H. Lin, and C. Fang. 2015. "Textural–Spectral Feature-Based Species Classification of Mangroves in Mai Po Nature Reserve from Worldview-3 Imagery." *Remote Sensing* 8: 24. doi:10.3390/rs8010024.

Wong, F. K. K., and T. Fung. 2014. "Combining EO-1 Hyperion and Envisat Asar Data for Mangrove Species Classification in Mai Po Ramsar Site, Hong Kong." *International Journal of Remote Sensing* 35: 7828–7856. doi:10.1080/01431161.2014.978034.

Xiao, Z. F., Y. P. Gong, Y. Long, D. R. Li, X. Y. Wang, and H. Liu. 2017. "Airport Detection Based on a Multiscale Fusion Feature for Optical Remote Sensing Images." *IEEE Geoscience and Remote Sensing Letters* 14 (Sep.): 1469–1473. doi:10.1109/LGRS.2017.2712638.

Yao, Y., Z. G. Jiang, H. P. Zhang, D. P. Zhao, and B. W. Cai. 2017. "Ship Detection in Optical Remote Sensing Images Based on Deep Convolutional Neural Networks." *Journal of Applied Remote Sensing* 11 (Sep.): 12. doi:10.1117/1.JRS.11.042611.

Zeiler, M. D., and R. Fergus. 2014. "Visualizing and Understanding Convolutional Networks." In *European*

Conference on Computer Vision, Cham, September 6–12, 818–833. Zurich: Springer.

Zhang, F., B. Du, L. P. Zhang, and M. Z. Xu. 2016. "Weakly Supervised Learning Based on Coupled Convolutional Neural Networks for Aircraft Detection." *IEEE Transactions on Geoscience and Remote Sensing* 54 (Sep.): 5553–5563. doi:10.1109/TGRS.2016.2569141.

Zhang, P., X. Niu, Y. Dou, and F. Xia. 2017. "Airport Detection on Optical Satellite Images Using Deep Convolutional Neural Networks." *IEEE Geoscience and Remote Sensing Letters* 14 (Aug.): 1183–1187. doi:10.1109/LGRS.2017.2673118.

Zhang, Y. H., K. Fu, H. Sun, X. Sun, X. W. Zheng, and H. Q. Wang. 2018. "A Multi-Model Ensemble Method Based on Convolutional Neural Networks for Aircraft Detection in Large Remote Sensing Images." *Remote Sensing Letters* 9: 11–20. doi:10.1080/2150704X.2017.1378452.

Zhang, H., T. Wang, M. Liu, M. Jia, H. Lin, L. M. Chu and T. D. Adam. 2018. "Potential of Combining Optical and Dual Polarimetric SAR Data for Improving Mangrove Species Discrimination Using Rotation Forest." *Remote Sensing*. 10 (3): 467. doi:10.3390/rs10030467.

Zhong, Y. F., F. Fe, and L. P. Zhang. 2016. "Large Patch Convolutional Neural Networks for the Scene Classification of High Spatial Resolution Imagery." *Journal of Applied Remote Sensing* 10 (Apr.): 20. doi:10.1117/1.JRS.10.025006.

Zou, Z. X., and Z. W. Shi. 2016. "Ship Detection in Spaceborne Optical Image with SVD Networks." *IEEE Transactions on Geoscience and Remote Sensing* 54 (Oct.): 5832–5845. doi:10.1109/TGRS.2016.2572736.