Automatic Tobacco Plant Detection in UAV Images via Deep Neural Networks

Zhun Fan*, *Senior Member, IEEE,* Jiewei Lu, Maoguo Gong, *Senior Member, IEEE,* Honghui Xie, and Erik Goodman

***Abstract*—Tobacco plant detection plays an important role in the management of tobacco planting. In this paper, a new algorithm based on deep neural networks is proposed to detect tobacco plants in images captured by unmanned aerial vehicles (UAVs) (called UAV images). These UAV images are characterized by a very high spatial resolution (35mm), and consequently contain an extremely high level of detail for the development of automatic detection algorithms. The proposed algorithm consists of three stages. In the first stage, a number of candidate tobacco plant regions are extracted from UAV images with the morphological operations and watershed segmentation. Each candidate region contains a tobacco plant or a non-tobacco plant. In the second stage, a deep convolutional neural network is built and trained with the purpose of classifying the candidate regions as tobacco plant regions or non-tobacco plant regions. In the third stage, postprocessing is performed to further remove the non-tobacco plant regions. The proposed algorithm is evaluated on a UAV image dataset. The experimental results show that the proposed algorithm performs well on the detection of tobacco plants in UAV images.**

***Index Terms*—Detection, tobacco plants, unmanned aerial vehicles (UAVs), convolutional neural network.**

I. Introduction

VERY high resolution (VHR) satellites, such as GeoEye-1 and Worldview-2, are important platforms for the acquisition of VHR images with less than $1m$ spatial resolutions. The acquired VHR images contribute to the development of many new applications such as quantifying bird migration [1], object tracking [2], counting roofless buildings [3] and automated detection of arbitrarily shaped buildings [4].

Besides VHR satellites, in recent years, unmanned aerial vehicles (UAVs) are increasingly becoming a new effective acquisition platform in the remote sensing community [5]. UAVs are small aircraft controlled either by attached microprocessors or by an operator on the ground. They have several important features: low cost, high mobility, flexibility, safety and customizability. The observation and monitoring of the earth have become much easier and faster because UAVs can reach an area of interest in a short time. UAVs were initially developed for military applications. However, due to the great potential and the rapid developments in technology, UAVs have become a practical solution for many civilian applications, such as car counting [6], [7], vegetation monitoring [8], [9], precision farming [10], [11], anomaly detection in archaeology sites [12], [13] and the detection of nonforested areas [14].

UAVs can obtain information at a low altitude, which allows us to collect images with very high spatial resolution (on the order of few centimeters). The detection and recognition of specific objects or a particular class of objects in an image play an important role in various applications. In [15], Fergus et al. proposed an algorithm to learn and recognize objects from unlabeled and unsegmented cluttered scenes. In [16], Agarwal et al. presented a technique to automatically learn to detect instances of the object class in new images. One of the classes of objects which needs particular attention is the tobacco plant. The interest in performing tobacco plant detection arises for several reasons: (1) The tobacco plant is an important economic crop in China, India, Brazil and the United States. (2) The detection of tobacco plants contributes to the management of tobacco planting. (3) Information about the number of tobacco plants is essential for yield estimation [17], [18]. (4) Current methods of counting tobacco plants are based on site inspection. Skilled inspectors go to the site and calculate the number of tobacco plants, which is a tedious and time-consuming task.

In order to complete the detection of tobacco plants in UAV images, deep neural networks [19] are employed in our work. Deep neural networks have been developed rapidly and have become a popular machine learning method due to the introduction in 2006 of an effective new algorithm to learn deep neural networks [20]. Deep neural networks have an impressive record of applications in image analysis and interpretation due to their powerful capacity [21]–[23]. Convolutional neural networks (CNNs) are the early proposed architectures of deep neural networks, which were built in the 1970's [24]. CNNs are inspired by the organization of the animal visual cortex [25], and initially applied to solve the challenging tasks like the recognition of handwritten characters [26]. With the rapid development of deep neural networks, both the frameworks and training algorithms of CNNs have been developed and improved rapidly [27], [28]. The powerful capabilities of CNNs allow them to be successfully applied

Manuscript received September 15, 2017; accepted January 9, 2018. This work was supported by the National Natural Foundation of Guangdong Province, Integrated Platform of Evolutionary Intelligence and Robot, support no.(2015KGJHZ015).

Zhun Fan, Jiewei Lu and Honghui Xie are with the Guangdong Provincial Key Laboratory of Digital Signal and Image Processing, College of Engineering, Shantou University, Shan'tou 515063, China. (email: zfan, 12jwlu1, 14hhxie@stu.edu.cn)

Maoguo Gong is with the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education, Xidian University, Xi'an 710071, China. (e-mail: gong@ieee.org)

Erik Goodman is with the BEACON Center for the Study of Evolution in Action, Michigan State University. East Lansing, Michigan, USA. (email: goodman@egr.msu.edu)

Fig. 1. The framework of the proposed algorithm

Fig. 2. A UAV image containing tobacco plants

for a large spectrum of vision tasks, including remote sensing images. In [29], Maggiori et al. proposed an end-to-end framework for the dense, pixelwise classification of satellite imagery with convolutional neural networks. In [30], Zhang et al. achieved aircraft detection with convolutional neural networks.

In this paper, a new algorithm based on deep neural networks is proposed to perform the detection of tobacco plants in UAV images. To the best of our knowledge, this is the first research on detecting tobacco plants in UAV images. The proposed algorithm is evaluated on a UAV image dataset. The experimental results demonstrate the effectiveness of the proposed algorithm.

The rest of this paper is organized as follows: Section II details the proposed methodology. Section III introduces the UAV image dataset and the evaluation metrics. Section IV presents the experimental results. In Section V, the conclusions of this paper are provided.

II. The Proposed Methodology

Let us consider a high-resolution image $I(x, y)$ (where (x, y) represents pixel coordinates in image I) captured by a UAV over an agricultural region of tobacco planting (Fig.2). UAVs can obtain images with high spatial resolutions when they fly at relatively low altitudes (several hundred meters). The high level of detail in UAV images provides much useful information, and contributes to the development of new analysis approaches. The objective of this study is to develop an automatic algorithm for the analysis of tobacco planting in UAV images. More specifically, our work focuses on the detection and counting of tobacco plants in UAV images.

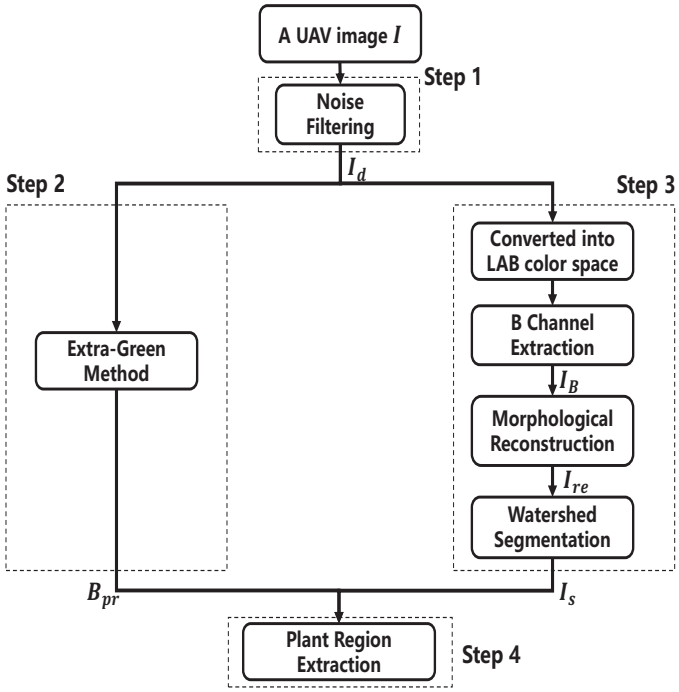


Fig. 3. The framework of candidate tobacco plant region extraction

The proposed algorithm is composed of three stages. In the first stage, a number of candidate tobacco plant regions are extracted from UAV images. Each candidate region contains a tobacco plant or a non-tobacco plant. In the second stage, a convolutional neural network is built and trained with the purpose of classifying the candidate regions as tobacco plant regions or non-tobacco plant regions. In the third stage, postprocessing is performed to further remove the non-tobacco plant regions. The framework of the proposed algorithm is shown in Fig.1. The image notations and their definitions used in the proposed algorithm are shown in Table I.

TABLE I
DEFINITIONS OF IMAGE NOTATIONS

Notations	Meanings
I_d	The denoised UAV image
B_{pr}	The binary image containing the plant regions
I_B	The B channel image
I_{re}	The morphologically reconstructed image
I_s	The region segmented image
I_e	The image used for plant region extraction
I_{rc}	The region-classified image
I_{plant}	The tobacco-plant-detected image

A. Candidate Tobacco Plant Region Extraction

Candidate tobacco plant region extraction aims at extracting candidate tobacco plant regions from UAV images. Each region contains a tobacco plant or a non-tobacco plant. The extraction of candidate tobacco plant regions consists of four steps: 1) noise filtering, 2) soil region detection, 3) plant region segmentation and 4) plant region extraction. Fig.3 shows the framework of candidate tobacco plant region extraction.

1) Noise Filtering: In order to smooth the noise in UAV images, each UAV image $I(x, y)$ is convolved with a Gaussian kernel $w(x, y)$

$$I_d(x, y) = I(x, y) \star w(x, y) \tag{1}$$

where I_d is the denoised UAV image. $w(x, y)$ is of dimensions $m \times m = 3 \times 3$, mean $\mu = 0$, and variance $\sigma^2 = 0.25$. $\star$ represents the convolution operation.

2) Soil Region Detection: UAV images generally contain soil regions. In order to reduce the influence of soil regions, the extra-green method [31] is applied to remove the soil regions and preserve the plant regions in UAV images, which is defined as follows:

$$B_{pr}(x, y) = B_{gr}(x, y) \cap B_{gb}(x, y) \tag{2}$$

$$B_{gr}(x, y) = \begin{cases} 1 & I_{gr}(x, y) > \omega_1 \\ 0 & else \end{cases} \tag{3}$$

$$B_{gb}(x, y) = \begin{cases} 1 & I_{gb}(x, y) > \omega_2 \\ 0 & else \end{cases} \tag{4}$$

$$I_{gr}(x, y) = I_g(x, y) - I_r(x, y) \tag{5}$$

$$I_{gb}(x, y) = I_g(x, y) - I_b(x, y) \tag{6}$$

where B_{pr} is the resultant binary image containing the plant regions. I_g, I_r and I_b are the green channel, red channel and blue channel of image I_d. I_{gr} is the difference image between I_g and I_r, and I_{gb} is the difference image between I_g and I_b. B_{gr} is the binary image obtained by thresholding the difference image I_{gr}, and B_{gb} is the binary image obtained by thresholding the difference image I_{gb}. ω_1 and ω_2 are depth control parameters. In our experiment, ω_1 and ω_2 are set as 0.05 and 0, which were selected according to the RGB values of lawn green ($R = 0.486, G = 0.988, B = 0$) and spring green ($R = 0.235, G = 0.702, B = 0.443$).

3) Plant Region Segmentation: The central regions of tobacco plants are generally brighter than the leaf regions [32]. In order to make use of this available property to divide the UAV images into a number of candidate tobacco plant regions, the denoised UAV image I_d is first transformed from RGB color space to LAB color space since LAB color space can describe all the colors visible to human eyes and was created to serve as a device-independent model [33], [34]. Second, the B channel image I_B is extracted because I_B provides the best contrast between central regions and leaf regions, while the lightness channel is the brightest color channel, and the A channel offers poor dynamic range. In order to reduce the influence of background regions in I_B, the morphological reconstruction by erosion of I_B from a marker image F is performed, resulting in the morphologically reconstructed image I_{re}. F is obtained by eroding I_B with a disk structuring element of size 5×5. Finally the watershed segmentation algorithm [35] is applied to divide the morphologically reconstructed image I_{re} into a number of candidate tobacco plant regions, resulting in the region segmented image I_s. The process of plant region segmentation is shown in Fig.4.

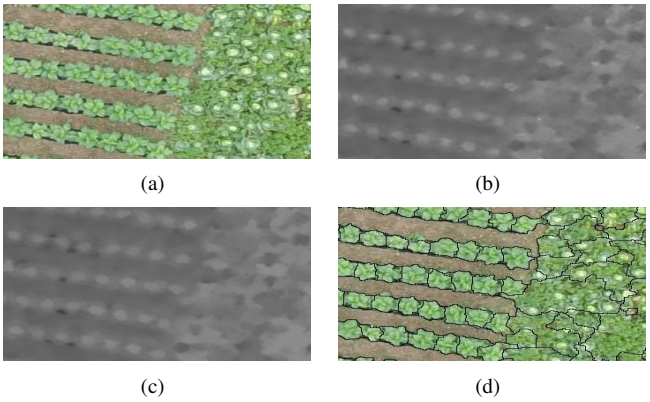


Fig. 4. The process of plant region segmentation: (a) A fragment of the denoised UAV image I_d. (b) A fragment of the B channel image I_B. (c) A fragment of the morphologically reconstructed image I_{re}. (d) A fragment of the region segmented image I_s.

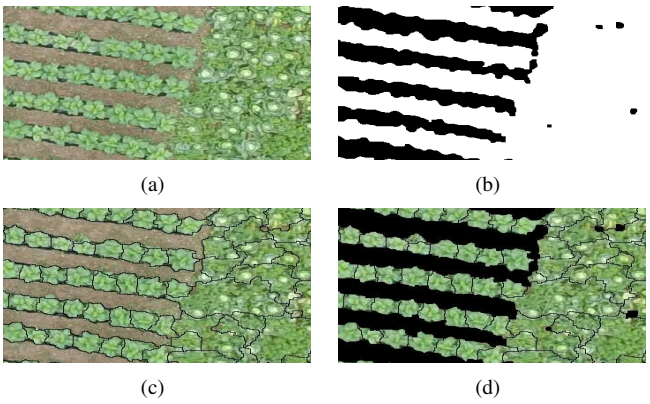


Fig. 5. The process of obtaining I_e: (a) A fragment of the original UAV image I. (b) A fragment of the binary image B_{pr}. (c) A fragment of the region segmented image I_s. (d) A fragment of image I_e.

4) Plant Region Extraction: The image used for plant region extraction (I_e) is obtained by carrying out a multiply arithmetic operation on image I_s and B_{pr}:

$$I_e(x, y) = I_s(x, y) \times B_{pr}(x, y) \tag{7}$$

The process of obtaining I_e is given in Fig.5. A number of candidate plant regions are extracted from I_e. Each candidate plant region is resized into $28 \times 28 \times 3$, and fed into a convolutional neural network. Each candidate region is associated with a class label 1 or 0. Label 1 means that the candidate region contains a tobacco plant, while label 0 means that the candidate region does not contain a tobacco plant. Fig.7 shows some exemplary instances of candidate regions with labels 1 and 0.

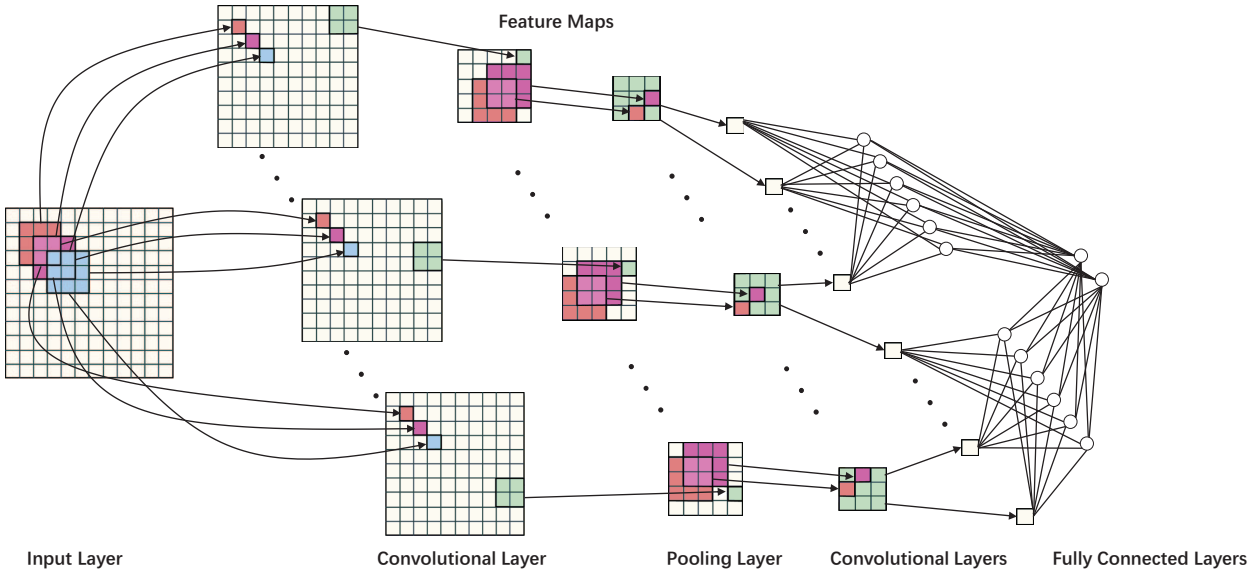


Fig. 6. A typical architecture of a convolutional neural network composed of three convolutional layers, one pooling layer and two fully connected layers. The network utilizes 3×3 convolutional kernels with stride 1 and 2×2 pooling kernels with stride 2. (The figure is modified from [23].)

(a)

(b)

Fig. 7. Some exemplary instances of candidate regions: (a) Examples of candidate regions with label 1. Each candidate region contains a tobacco plant. (b) Examples of candidate regions with label 0. Each candidate region does not contain a tobacco plant.

B. Deep Neural Network Establishment

Training a deep convolutional neural network (CNN) is the core component of the proposed algorithm. The CNN is composed of multiple elementary neurons (units), each performing convolution of the neurons' weights with the input volume and transforming a few weighted inputs into an output volume with some nonlinear functions. The neurons are spatially arranged in rectangular layers (grids) (Fig.6). The spatial arrangement of neurons controls the primary characteristics of the CNN and makes it suitable for a large variety of computer vision tasks. Some other important characteristics of CNNs are given as follows.

1) Sparse Interactions: Sparse Interactions mean that each neuron is connected to only a small region of the input volume or its receptive field (RF). This is accomplished by making the kernel smaller than the input. For example, when processing an image, although the input image may have a number of pixels, only small and meaningful features such as edges are detected with kernels. In other words, fewer parameters need to be stored, which greatly improves the statistical efficiency and memory requirement of the network compared with traditional fully-connected neural networks. More specifically, if a layer has 3×3 kernels and 1 stride, it only needs 9 neurons when applied to a 5×5 single-channel image. *Sparse Interactions* comply with certain aspects of natural visual systems [24] and greatly improve the performance of the neural network.

2) Parameter Sharing: Parameter Sharing refers to sharing the same parameters across neurons in the same layer. Compared with traditional neural networks, each member of the kernel is used at every position of the input in a CNN. *Parameter sharing* also means that only one set of parameters is learned for each new location, rather than learning a separate set of parameters. It further reduces the number of parameters and contributes to the equivariance of the extracted features. For example, when connected to a single channel image, a

layer of neurons with 3×3 kernels only has 10 parameters (nine for pixels in the RF and one for the neuron threshold).

3) Pooling (subsampling): Pooling means performing aggregations of the neurons' outputs by other means rather than convolution. *Max-pooling*, the most common pooling function, reports the maximum output within a rectangular neighborhood. *Pooling* contributes to the invariance to local translation and reduces the number of parameters.

A typical CNN consists of several convolutional and pooling layers optionally followed by one or more fully connected layers (Fig.6). The input to a convolutional layer is a $h \times w \times c$ image, where h and w are the height and width of the image, and c is the number of image channels. The convolutional layer will have k kernels of size $m \times n \times r$, where m and n are smaller than the dimensions of the image, r is less than or equal to c and may vary for each kernel. The kernels are convolved with the image to produce k feature maps of size [h-m+1,w-n+1]. Then each feature map is subsampled, typically with max pooling over $p \times q$ contiguous regions. p and q generally belong to $[2, 5]$. After the convolutional layers there is at least one fully connected layer, which maps the excitations into output neurons, each corresponding to one decision class.

After building the network's architecture, the parameters ω of the CNN, initialized with small signed random values, are learned by minimizing the cost function

$$J(\omega) = \frac{1}{n} \sum_{i=1}^{n} L(y_i, f(\mathbf{x}_i, \omega)) \tag{8}$$

where $\mathbf{x}_i$ is the feature vector of the i^{th} training example, y_i is the label of the i^{th} training example. n is the number of training examples. $f(\cdot)$ is the activation function, $L(\cdot)$ is the loss function expressing the penalty for predicting $f(\mathbf{x}_i, \omega)$ instead of y_i. *Stochastic Gradient Descent (SGD)* [36] is the most common optimization method applied to minimize the cost function. Instead of using all training examples, *SGD* updates the parameters of $J(\omega)$ with only a single or few training examples $(\mathbf{x}_s, y_s)$ (called batches)

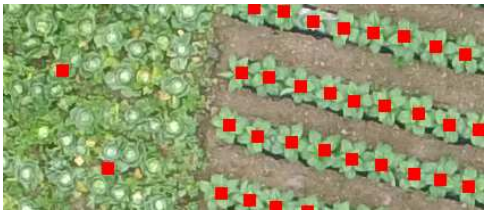
$$\omega^{t+1} = \omega^t - \alpha \frac{\partial}{\partial \omega^t} J(\omega; \mathbf{x}_s, y_s) \tag{9}$$

where t indicates the iteration index, α is the learning rate, $\frac{\partial}{\partial \omega^t} J(\omega)$ is the partial derivative of the cost function $J(\omega)$. The parameters tend to converge to the local optima after the update of each iteration.

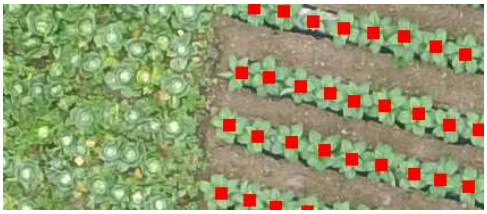
After training the network, the final CNN is established. The candidate regions are fed into the CNN, and the network outputs the class labels. For a more comprehensive description of CNN, [19], [37], [38] are recommended.

C. Postprocessing

The region-classified images I_{rc} (Fig.8.(a)) are obtained after classifying the candidate tobacco plant regions in UAV images with a CNN. However, some non-tobacco plant regions are misclassified as tobacco plant regions because they have similar features with tobacco plant regions. These non-tobacco plant regions are often far away from the tobacco plant regions and often appear spatially isolated. In order to remove the non-tobacco plants in I_{rc}, the local density p_t [39] of each classified tobacco plant t by the CNN is computed as:



(a)



(b)

Fig. 8. (a) is a fragment of the region-classified image I_{rc}. (b) is a fragment of the final tobacco-plant-detected image I_{plant}. Red regions represent the central regions of the detected tobacco plants.

Fig. 9. The UAV used for the acquisition of the tobacco plant images.

$$p_t = \sum_{j=1}^{v} \chi(d_{tj} - d_c) \tag{10}$$

where $\chi(x) = 1$ if $x < 0$ and $\chi(x) = 0$ otherwise. d_{tj} is the distance between plant t and plant j. d_c is a cutoff distance. v is the number of the closest tobacco plants of plant t. If $p_t > (\frac{v}{2})$, plant t is considered a tobacco plant; If $p_t \leqslant (\frac{v}{2})$, plant t is considered a non-tobacco plant. In our experiment, d_c and v are set as 120 and 7, which were selected based on an empirical study. The final tobacco-plant-detected image I_{plant} is shown in Fig.8.(b).

III. Dataset Description and Evaluation Metrics

In this section, a concrete description of the UAV image dataset is provided, followed by introducing the evaluation metrics used in our experiment.

A. Dataset Description

The UAV image dataset consists of 14 tobacco plant images. The images were obtained by using a UAV equipped with

TABLE II
THE ARCHITECTURE OF THE EVALUATED CNNS. LAYER NAMES ARE FOLLOWED BY NUMBERS OF FEATURE MAPS. SQUARE BRACKETS SPECIFY THE RECEPTIVE FIELD ($m \times n$ FOR CONVOLUTIONAL LAYERS OR $p \times q$ FOR POOLING LAYERS), AND STRIDE.

The 1st network architecture	Input	→	conv20 [5,5,1]	→	maxpool [2,2,2]	→	conv20 [5,5,1]	→	maxpool [2,2,2]	→	conv500 [4,4,1]	→	fc512	→	fc512	→	fc2
The 2nd network architecture	Input	→	conv20 [5,5,1]	→	conv20 [5,5,1]	→	conv500 [4,4,1]	→	fc512	→	fc512	→	fc2				

imaging sensors (Fig.9). The images were acquired over the agricultural regions of tobacco planting with different altitudes in Chongqing, China, on June 27, 2016. The images are characterized by three channels (RGB) and by a spatial resolution of $35mm$. All the acquired images are digitized to 4000×3000 pixels with 8 bits per color channel, and have been JPEG compressed. The set of 14 images were divided into a training and a test set.

1) Training Set: is composed of 7 images. It is used for the training of a CNN for the classification of candidate tobacco plant regions. 36857 candidate regions are extracted from the training set. 18302 candidate regions are labeled 0 while 18555 candidate regions are labeled 1.

2) Test set: is also composed of 7 images. It is used for evaluating the performance of the proposed algorithm for tobacco plant detection. 33539 candidate regions are extracted from the test set. 16773 candidate regions are labeled 0 while 16766 candidate regions are labeled 1.

B. Evaluation Metrics

In order to assess the performance of the proposed algorithm, three commonly used metrics are applied:

$$Sensitivity = \frac{TP}{TP + FN} \tag{11}$$

$$Specificity = \frac{TN}{TN + FP} \tag{12}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

where TP, TN, FP and FN indicate true positive (the number of correctly identified tobacco plants), true negative (the number of correctly identified non-tobacco plants), false positive (the number of incorrectly identified tobacco plants) and false negative (the number of incorrectly identified non-tobacco plants), respectively. *Sensitivity (Se)* reflects the algorithm's capability to detect tobacco plants while *Specificity (Sp)* is a measure of the algorithm's effectiveness in identifying non-tobacco plants. *Accuracy (Acc)* is a global measure of the performance of the proposed method.

In order to further evaluate the capability of our methodology to correctly identify and count the number of tobacco plants in UAV images, two useful evaluation metrics are also employed in our experiment.

The first one is the producer's accuracy (P_{acc}) [6], [40]:

$$P_{acc} = \frac{TP}{N} \tag{14}$$

where N indicates the actual number of tobacco plants in the UAV images. P_{acc} shows the percentage of correctly identified tobacco plants.

The second one is the relative error ($Error$) [41], [42]:

$$Error = \frac{|N_p - N|}{N} \times 100\% \tag{15}$$

where $N_p = TP + FP$ is the number of detected tobacco plants. $Error$ measures the performance of the proposed method on yield estimation.

IV. THE EXPERIMENTS AND RESULTS

A. The Network Architecture and Training Parameters

A range of network architectures are considered in our experiments. Two representative network architectures are used in this work. In the first architecture, the input image is passed through a stack of convolutional layers and max-pooling layers, followed by three fully connected (FC) layers: the first and the second have 512 neurons, the third performs binary classification and thus contains 2 neurons (one for each class). Weights in each layer are sampled from a Gaussian distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$. The second network architecture abandons max-pooling, which is the only difference with the first one. The reason we chose this architecture for comparison is because it has been shown that networks without pooling layers may perform better when applied to small images [43].

Training is carried out by SGD. In each iteration, the training examples are passed through the network, which propagates excitations through the network and calculates errors committed by the neurons. Then the errors are backward propagated through the network, and used to calculate parameter corrections. The parameters are updated with 200 batches. The learning rate is set as 5×10^{-4}. The training was stopped after 11100 iterations. The implementation was based on Matcovnet [44], an effective and flexible toolbox of CNN.

B. Experimental Results

In order to assess the performance of the proposed algorithm, the following three experiments are conducted. In the first experiment, the performance of tobacco plant detection was analyzed. In the second experiment, the comparison between different classifiers was performed when CNNs were exploited just for feature generation. In the third experiment, the analysis of the sensitivity of the proposed algorithm to the number of training examples was given.

1) The 1st experiment: The performance of the proposed algorithm on the UAV test dataset is shown in Table III. In the first CNNs' network, the proposed algorithm achieves high scores on *Sensitivity* and *Specificity*, with average values of 0.9525 and 0.9159, respectively, which means that the proposed algorithm performs well on identifying both tobacco

TABLE III
THE PERFORMANCE OF THE PROPOSED ALGORITHM ON THE UAV IMAGE DATASET

The network architectures	Images	TP	TN	FP	FN	N	Se	Sp	P_{acc}	Acc	Error(%)
The 1st network architecture	Image Test 1	2202	2015	293	131	2506	0.9438	0.8731	0.8787	0.9086	0.44%
	Image Test 2	2681	1973	84	189	2924	0.9341	0.9592	0.9169	0.9446	5.44%
	Image Test 3	1498	3877	79	96	1660	0.9398	0.9800	0.9024	0.9685	5.00%
	Image Test 4	3238	2076	286	156	3551	0.9540	0.8789	0.9119	0.9232	0.76%
	Image Test 5	2623	2540	271	144	3026	0.9480	0.9036	0.8668	0.9256	4.36%
	Image Test 6	2591	1519	183	52	2658	0.9803	0.8925	0.9748	0.9459	4.36%
	Image Test 7	1127	1457	120	38	1163	0.9674	0.9239	0.9690	0.9424	7.22%
	Average	2280	2208	188	115	2498	0.9525	0.9159	0.9126	0.9370	3.94%
The 2nd network architecture	Image Test 1	2190	2004	304	143	2506	0.9387	0.8683	0.8739	0.9037	0.48%
	Image Test 2	2663	1966	91	207	2924	0.9279	0.9558	0.9107	0.9395	5.81%
	Image Test 3	1474	3810	146	120	1660	0.9247	0.9631	0.8880	0.9521	2.41%
	Image Test 4	3206	2072	290	188	3551	0.9446	0.8772	0.9028	0.9170	1.55%
	Image Test 5	2549	2549	262	218	3026	0.9212	0.9068	0.8424	0.9139	7.11%
	Image Test 6	2559	1516	186	84	2658	0.9682	0.8907	0.9628	0.9379	3.27%
	Image Test 7	1124	1445	132	41	1163	0.9648	0.9163	0.9665	0.9369	8.00%
	Average	2252	2195	202	143	2498	0.9414	0.9112	0.9015	0.9287	4.09%

TABLE IV
THE PERFORMANCE COMPARISON AMONG THE NN, SVM AND RANDOM FORESTS CLASSIFIERS

Network	Classifiers	Se	Sp	P	Acc	Error
The first CNNs' network	NN	0.9468	0.9196	0.9073	0.9361	4.50%
	SVM	0.9462	0.9215	0.9065	0.9365	4.39%
	Random Forests	0.9470	0.9199	0.9073	0.9363	4.49%
The second CNNs' network	NN	0.9484	0.9079	0.9081	0.9304	3.60%
	SVM	0.9454	0.9085	0.9054	0.9295	3.87%
	Random Forests	0.9361	0.9148	0.8971	0.9282	4.18%

plants and non-tobacco plants. The proposed algorithm also has good performance on *Accuracy* and P_{acc}, with average values of 0.9370 and 0.9126, respectively, which means that the proposed algorithm has good capability of correctly identifying and counting the number of tobacco plants in UAV images. Moreover, the proposed algorithm has less than 4% *Error* on average, which indicates that the proposed algorithm performs well on yield estimation. In the second CNNs' network, the proposed algorithm yields good results for *Sensitivity*, *Specificity*, *Accuracy*, P_{acc} and *Error*, with average values of 0.9414, 0.9112, 0.9287, 0.9015 and 4.09%, respectively. The proposed algorithm achieves slightly better results in the first CNNs' network than in the second CNNs' network. The final tobacco-plant-detected images achieved by the first CNNs' network are shown in Figures.11 and 12.

2) The 2nd experiment: when CNNs are used just for feature generation, a performance comparison among the nearest neighbor (NN), SVM and Random Forests classifiers is shown in Table IV. In this experiment, for the first CNNs' network, the values of neurons in the 8th layer are extracted as features; for the second CNNs' network, the values of neurons in the 6th layer are extracted as features. In both cases, the values of neurons in the second to the last layer in the CNNs' networks are chosen as features. Libsvm toolbox [45] is used in the experiment. From Table IV, it can be observed that the results among the NN, SVM and Random Forests classifiers are close in both CNNs' networks. The fact that all the three classifiers can obtain promising performance may indicate that CNNs can extract useful features.

3) The 3rd experiment: The analysis of the sensitivity of the proposed algorithm to the number of training examples is given in Fig.10, from which it can be observed that with the increase of the number of training examples, all the evaluation metrics become better in both CNNs' networks. When the number of training examples is less than 20000, the performance of the proposed algorithm improves significantly with the increase of the number of training examples. When the number of training examples is more than 20000, for both CNNs' networks, the performance of the proposed algorithm differs very slightly with the variation of training examples. The performance of the proposed algorithm becomes almost stagnant, which indicates that we can choose the value of training examples as close to 20000.

V. CONCLUSION

The tobacco plant is an important economic crop in China, India, Brazil and the United States. Tobacco plant detection is of great significance to the management of tobacco planting. However, current methods of tobacco plant detection are based on site inspection, which is tedious and time-consuming. In order to achieve automated detection of tobacco plants, tobacco plant images are collected by means of UAVs. These images have high spatial resolution and contain a high level of detail for the detection of tobacco plants. Then a new algorithm

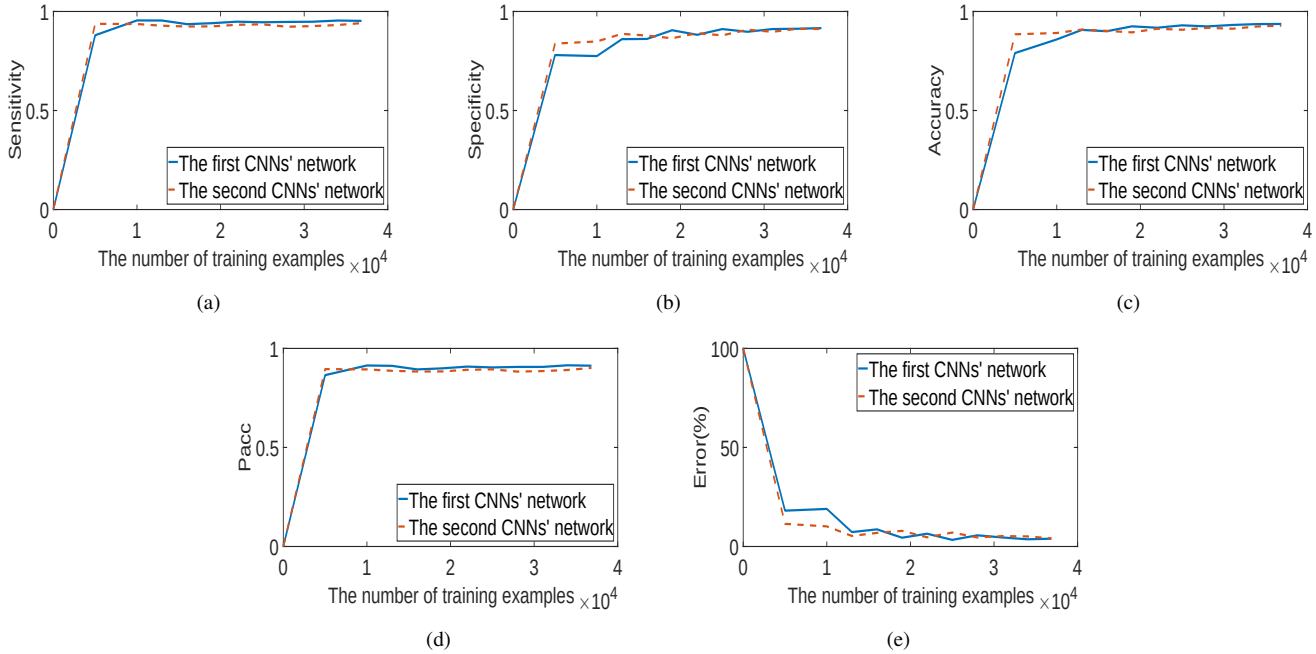


Fig. 10. Variations of different evaluation metrics with the increase of the number of training examples. (a) The variation of average *Sensitivity*. (b) The variation of average *Specificity*. (c) The variation of average *Accuracy*. (d) The variation of average P_{acc}. (e) The variation of average *Error*.

based on deep neural networks is proposed for the automated detection of tobacco plants in UAV images. To the best of our knowledge, this is the first research aimed at detecting tobacco plants in UAV images. The proposed algorithm has three stages. In the first stage, a number of candidate tobacco plant regions are extracted from UAV images using morphological operations and watershed segmentation. Each candidate region contains a tobacco plant or a non-tobacco plant. In the second stage, a deep convolutional neural network is trained and established in order to classify each candidate tobacco plant region as a tobacco plant region or non-tobacco plant region. In the final stage, postprocessing is performed with the purpose of further removing non-tobacco plants.

In order to evaluate the performance of tobacco plant detection, the proposed algorithm is tested on a UAV image dataset. The proposed algorithm performs well on the detection of tobacco plants and achieves an average $Accuracy$, P_{acc} and $Error$ of 0.9370, 0.9126 and 3.94%, respectively. The experimental results demonstrate that the proposed algorithm has a good capability to correctly identify and count the number of tobacco plants in UAV images. Future research can be performed by adapting the algorithm to detect other food crops, such as corn, rice and rapeseed.

ACKNOWLEDGMENT

The authors would like to thank the Key Lab of Digital Signal and Image Processing of Guangdong Province for providing the UAV image dataset.

REFERENCES

[1] R. Nebuloni, C. Capsoni, and V. Vigorita, "Quantifying bird migration by a high-resolution weather radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 6, pp. 1867–1875, 2008.

[2] L. Meng and J. P. Kerekes, "Object tracking using high resolution satellite imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 1, pp. 146–152, 2012.

[3] L. Gueguen, M. Pesaresi, A. Gerhardinger, and P. Soille, "Characterizing and counting roofless buildings in very high resolution optical images," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 1, pp. 114–118, 2012.

[4] A. O. Ok, C. Senaras, and B. Yuksel, "Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 3, pp. 1701–1717, 2013.

[5] A. Y. Chen, Y.-N. Huang, J.-Y. Han, and S.-C. J. Kang, "A review of rotorcraft Unmanned Aerial Vehicle (UAV) developments and applications in civil engineering," *Smart Structures and Systems*, vol. 13, no. 6, pp. 1065–1094, 2014.

[6] T. Moranduzzo and F. Melgani, "Automatic car counting method for unmanned aerial vehicle images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 3, pp. 1635–1647, 2014.

[7] M. Thomas and M. Farid, "Detecting cars in UAV images with a catalog-based approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 10, pp. 6356–6367, 2014.

[8] J. A. Berni, P. J. Zarco-Tejada, L. Suárez, and E. Fereres, "Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 3, pp. 722–738, 2009.

[9] K. Uto, H. Seki, G. Saito, and Y. Kosugi, "Characterization of rice paddies by a UAV-mounted miniature hyperspectral sensor system," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 2, pp. 851–860, 2013.

[10] H. Xiang and L. Tian, "Development of a low-cost agricultural remote sensing system based on an autonomous unmanned aerial vehicle (UAV)," *Biosystems Engineering*, vol. 108, no. 2, pp. 174–190, 2011.

[11] S. Candiago, F. Remondino, M. De Giglio, M. Dubbini, and M. Gattelli, "Evaluating multispectral images and vegetation indices for precision farming applications from UAV images," *Remote Sensing*, vol. 7, no. 4, pp. 4026–4047, 2015.

[12] A. Y.-M. Lin, A. Novo, S. Har-Noy, N. D. Ricklin, and K. Stamatiou, "Combining GeoEye-1 satellite remote sensing, UAV aerial imaging, and geophysical surveys in anomaly detection applied to archaeology," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 4, no. 4, pp. 870–876, 2011.

[13] R. M. Cavalli, G. A. Licciardi, and J. Chanussot, "Detection of anomalies produced by buried archaeological structures using nonlinear principal

Fig. 11. Final tobacco-plant-detected images (Red regions represent the central regions of the detected tobacco plants. Black rectangles include the enlarged fragments of the images): (a) The first test image. The geo-coordinate of this image is ($N29°29^{'}9.2094^{''}$, $E107°40^{'}51.3395^{''}$). (b) The second test image. The geo-coordinate of this image is ($N29°29^{'}10.0000^{''}$, $E107°40^{'}51.1483^{''}$). (c) The third test image. The geo-coordinate of this image is ($N29°29^{'}11.0611^{''}$, $E107°40^{'}53.0032^{''}$). (d) The fourth test image. The geo-coordinate of this image is ($N29°29^{'}10.5399^{''}$, $E107°40^{'}50.8106^{''}$).

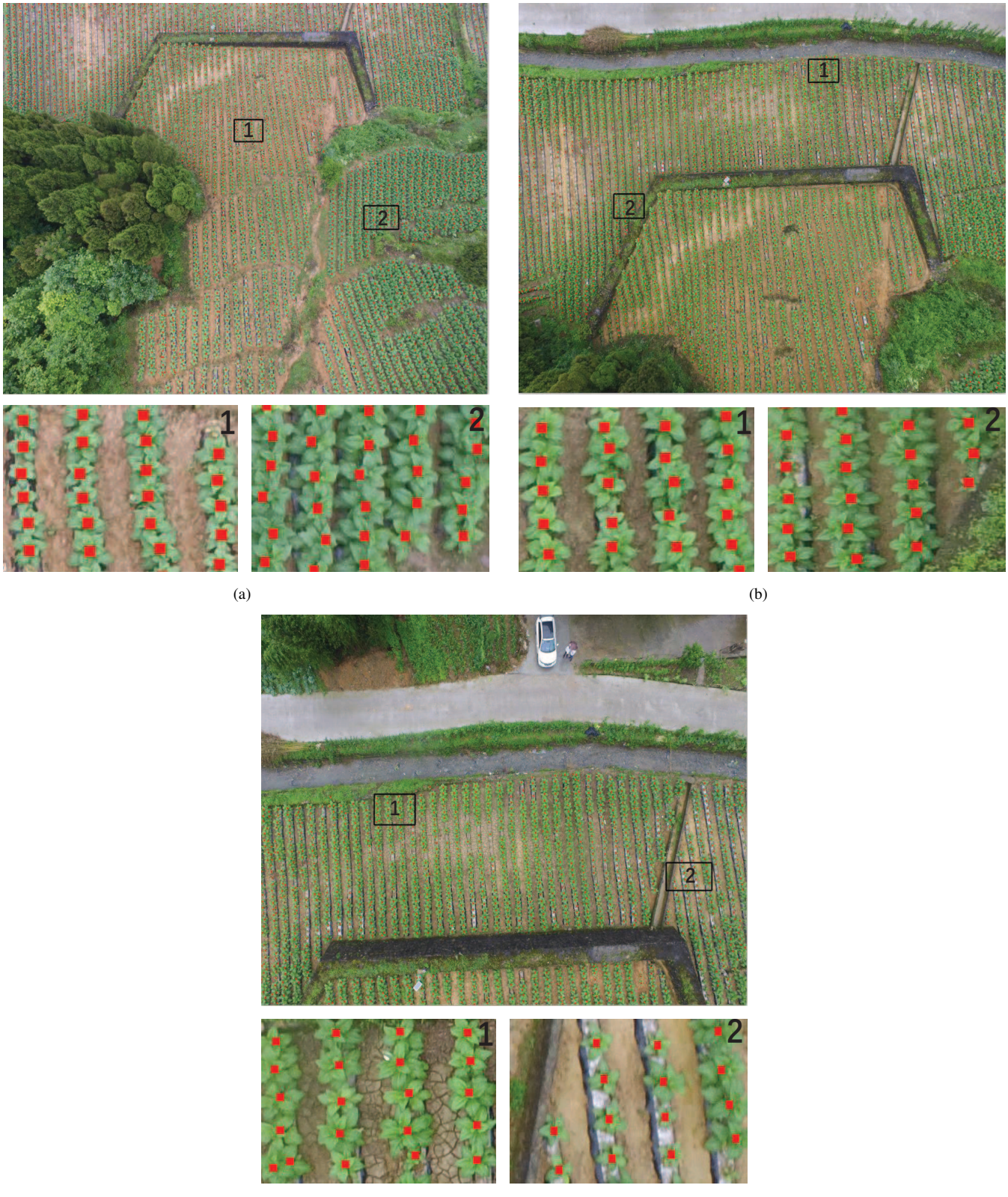


Fig. 12. Final tobacco-plant-detected images (Red regions represent the central regions of the detected tobacco plants. Black rectangles include the enlarged fragments of the images): (a) The fifth test image. The geo-coordinate of this image is ($N29°29^{'}10.8305^{''}$, $E107°40^{'}50.5756^{''}$). (b) The sixth test image. The geo-coordinate of this image is ($N29°29^{'}10.2505^{''}$, $E107°40^{'}51.0861^{''}$). (c) The seventh test image. The geo-coordinate of this image is ($N29°29^{'}9.8117^{''}$, $E107°40^{'}51.2844^{''}$).

component analysis applied to airborne hyperspectral image," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 2, pp. 659–669, 2013.

[14] H. O. Cruz, M. Eckert, J. M. Meneses, and J. F. Martínez, "Precise real-time detection of nonforested areas with UAVs," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 632–644, 2017.

[15] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 2. IEEE, 2003, pp. II–II.

[16] S. Agarwal, A. Awan, and D. Roth, "Learning to detect objects in images via a sparse, part-based representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1475–1490, 2004.

[17] L. S. Unganai and F. N. Kogan, "Drought monitoring and corn yield estimation in Southern Africa from AVHRR data," *Remote Sensing of Environment*, vol. 63, no. 3, pp. 219–232, 1998.

[18] P. Bose, N. K. Kasabov, L. Bruzzone, and R. N. Hartono, "Spiking neural networks for crop yield estimation based on spatiotemporal analysis of image time series," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 11, pp. 6563–6573, 2016.

[19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[20] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[21] G. E. Hinton, "Learning to represent visual input," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 365, no. 1537, pp. 177–184, 2010.

[22] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[23] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," *IEEE Transactions on Medical Imaging*, vol. 35, pp. 1–1, 2016.

[24] K. Fukushima and S. Miyake, "Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition," in *Competition and Cooperation in Neural Nets*. Springer, 1982, pp. 267–285.

[25] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *The Journal of Physiology*, vol. 148, no. 3, pp. 574–591, 1959.

[26] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel, "Handwritten digit recognition with a back-propagation network, 1989," in *Neural Information Processing Systems (NIPS)*.

[27] J. Masci, J. Angulo, and J. Schmidhuber, "A learning framework for morphological operators using counter–harmonic mean," in *International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing*. Springer, 2013, pp. 329–340.

[28] J. Masci, A. Giusti, D. Ciresan, G. Fricout, and J. Schmidhuber, "A fast learning algorithm for image segmentation with max-pooling convolutional networks," in *Image Processing (ICIP), 2013 20th IEEE International Conference on*. IEEE, 2013, pp. 2713–2717.

[29] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 645–657, 2017.

[30] F. Zhang, B. Du, L. Zhang, and M. Xu, "Weakly supervised learning based on coupled convolutional neural networks for aircraft detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5553–5563, 2016.

[31] H. T. Søgaard, "Weed classification by active shape models," *Biosystems Engineering*, vol. 91, no. 3, pp. 271–281, 2005.

[32] V. Cokkinides, P. Bandi, E. Ward, A. Jemal, and M. Thun, "Progress and opportunities in tobacco control." *Ca A Cancer Journal for Clinicians*, vol. 56, no. 3, pp. 135–142, 2006.

[33] P. J. Baldevbhai and R. Anand, "Color image segmentation for medical images using L*a*b* color space," *IOSR Journal of Electronics and Communication Engineering*, vol. 1, no. 2, pp. 24–45, 2012.

[34] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[35] F. Meyer, "Topographic distance and watershed lines," *Signal Processing*, vol. 38, no. 1, pp. 113–125, 1994.

[36] L. Bottou, "Stochastic gradient descent tricks," in *Neural networks: Tricks of the trade*. Springer, 2012, pp. 421–436.

[37] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[38] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[39] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

[40] N. Lavrač, P. Flach, and B. Zupan, "Rule evaluation measures: A unifying view," in *International Conference on Inductive Logic Programming*. Springer, 1999, pp. 174–185.

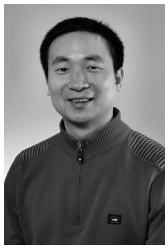
[41] J. S. Armstrong and F. Collopy, "Error measures for generalizing about forecasting methods: Empirical comparisons," *International Journal of Forecasting*, vol. 8, no. 1, pp. 69–80, 1992.

[42] B. E. Stine, C. Hess, L. H. Weiland, D. J. Ciplickas, and J. Kibarian, "System and method for product yield prediction using a logic characterization vehicle," Dec. 21 2004, uS Patent 6,834,375.

[43] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," *Eprint Arxiv*, 2014.

[44] A. Vedaldi and K. Lenc, "Matconvnet: Convolutional neural networks for matlab," in *Proceedings of the 23rd ACM International Conference on Multimedia*. ACM, 2015, pp. 689–692.

[45] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.



Zhun Fan received the B.S. and M.S. degrees in control engineering from Huazhong University of Science and Technology, Wuhan, China, in 1995 and 2000, respectively, and the Ph.D. degree in electrical and computer engineering from the Michigan State University, Lansing, MI, USA, in 2004. He is currently a Full Professor with Shantou University (STU), Shantou, China. He also serves as the Head of the Department of Electrical Engineering and the Director of the Guangdong Provincial Key Laboratory of Digital Signal and Image Processing. Before joining STU, he was an Associate Professor with the Technical University of Denmark (DTU) from 2007 to 2011, first with the Department of Mechanical Engineering, then with the Department of Management Engineering, and as an Assistant Professor with the Department of Mechanical Engineering in the same university from 2004 to 2007. He has been a Principle Investigator of a number of projects from Danish Research Agency of Science Technology and Innovation and National Natural Science Foundation of China. His research is also supported by the National Science Foundation. His major research interests include intelligent control and robotic systems, robot vision and cognition, MEMS, computational intelligence, design automation, optimization of mechatronic systems, machine learning and image processing.

Jiewei Lu is with the key lab of digital signal and image processing of Guangdong Province, Shantou University, Shantou, China, where he is currently pursuing the M.S. degree in information and communication with the School of Engineering. His current research interests include medical image analysis, image processing and machine learning.

Maoguo Gong received the B.S. degree in electronic engineering and the Ph.D. degree in electronic science and technology from Xidian University, Xian, China, in 2003 and 2009, respectively. He has been a Teacher with Xidian University, since 2006, where he was promoted to Associate Professor and Full Professor, both with exceptive admission, in 2008 and 2010. He has authored over 50 papers in journals and conferences, and holds 14 granted patents. His current research interests include computational intelligence with applications to optimization, learning, data mining, and image understanding.

Dr. Gong received the prestigious National Program for the support of Top-Notch Young Professionals from the Central Organization Department of China, the Excellent Young Scientist Foundation from the National Natural Science Foundation of China, and the New Century Excellent Talent in University from the Ministry of Education of China. He is the Vice Chair of the IEEE Computational Intelligence Society Task Force on Memetic Computing, an Executive Committee Member of the Chinese Association for Artificial Intelligence, and a Senior Member of the Chinese Computer Federation.

Honghui Xie received the B.S. degree from Hunan City University, Hunan, China, in 2010, and the M.S. degree from Shantou University, Guangdong, China, in 2017. His current research interests include image processing and machine learning.

Erik D. Goodman is PI and Director of the BEACON Center for the Study of Evolution in Action, an NSF Science and Technology Center headquartered at Michigan State University and funded beginning in 2010. His research centres on application of evolutionary principles to solution of engineering design problems. He received the PhD in computer and communication sciences from the University of Michigan, Ann Arbor, in 1971. He became Asst. Prof. of Electrical Engineering and Systems Science in 1972, Assoc. Prof. in 1978 and Prof. in 1984, all at Michigan State University, where he also holds appointments in Mechanical Engineering and in Computer Science and Engineering. He directed the Case Center for Computer-Aided Engineering and Manufacturing from 1983 to 2002, and MSUs Manufacturing Research Consortium from 1993 to 2003. He has co-directed MSUs Genetic Algorithms Research and Applications Group (GARAGe) since its founding in 1993. He is co-founder and vice president of Red Cedar Technology, Inc., a firm that develops design optimisation software for use in industry. He was chosen Michigan Distinguished Professor of the Year, 2009, by the Presidents Council, State Universities of Michigan.

Prof. Goodman was Chair of the Executive Board and a Senior Fellow of the International Society for Genetic and Evolutionary Computation, 20032005. He was founding chair of the ACMs Special Interest Group on Genetic and Evolutionary Computation (SIGEVO), serving from 2005 to 2007.