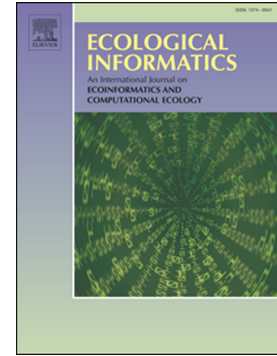# Accepted Manuscript

Columnar cactus recognition in aerial images using a deep learning approach

Efren López-Jiménez, Juan Irving Vasquez-Gomez, Miguel Angel Sanchez-Acevedo, Juan Carlos Herrera-Lozada, Abril Valeria Uriarte-Arcia

Columnar Cactus Recognition in Aerial Images using a Deep Learning Approach

Efren López-Jiménez[b], Juan Irving Vasquez-Gomez[a,b,*] jivasquezg@conacyt.mx, Miguel Angel, Sanchez-Acevedo[c], Juan Carlos Herrera-Lozada[b], Abril Valeria Uriarte-Arcia[b]

[a]Consejo Nacional de Ciencia y Tecnología (CONACYT), Ciudad de México, México
[b]Instituto Politécnico Nacional (IPN), CIDETEC, Ciudad de México, México
[c]Universidad de la Cañada (UNCA), Teotitlán, Oaxaca, México

[*]Corresponding author

Abstract

Tehuacán-Cuicatlán Valley is a semi-arid zone in the south of Mexico. It was inscribed in the World Heritage List by the UNESCO in 2018. This unique area has wide biodiversity including several endemic plants. Unfortunately, human activity is constantly affecting the area. A way to preserve a protected area is to carry out autonomous surveillance of the area. A first step to reach this autonomy is to automatically detect and recognize elements in the area. In this work, we present a deep learning based approach for columnar cactus recognition, specifically, the *neobuxbaumia tetetzo* species, endemic of the Valley. An image dataset was generated for this study by our research team, containing more than 10,000 image examples. The proposed approach uses this dataset to train a modified LeNet-5 Convolutional Neural Network. Experimental results have shown a high recognition accuracy, 0.95 for the validation set, validating the use of the approach for columnar cactus recognition.

1. Introduction

Tehuacán-Cuicatlán Valley is a semi-arid zone between the states of Puebla and Oaxaca in the south of Mexico. It contains one of the richest biodiversity in North America and it was inscribed in the World Heritage List as a mixed site (cultural and natural) by the UNESCO in July fourth, 2018 [1]. This area is filled by endemic vegetation forming a unique landscape. Unfortunately, human activity is constantly affecting the region, as a result, the natural land area has decreased in size, as well as in vegetation density. According to the records, by the year 2000, more than 72,063 Ha (14.69 % of the total area) had been affected [2]. In spite of the government efforts, the area is still in degradation.

A first step to preserve the natural areas is to carry out surveillance and to keep a register of the human activity and its effects on the vegetation. Such surveillance not only will help to preserve the area but it will also provide a knowledge base for future researches on the ecosystem evolution. Field investigation, carried out by experts, can provide accurate information in terms of plant morphology, the number of plants, human activity and weather conditions. However, field investigation requires a large number of resources not only economic but also in terms of time. Manned aerial vehicles for surveillance or information gathering [3], such as helicopters or small planes, could cover wider areas however its use is very expensive. In the last years, several works have been proposed to automatize the surveillance activities, for instance, using hyperspectral images perceived by remote sensors (satellite imagery) [4] or the use of digital photometrics [5]. Even though such remote sensor methods increase the coverage of the surveillance, they are limited by temporal constraints (in case of public information), low spatial resolution or budget constraints (in case of high frequency sampling images). Therefore, there is still a need for a low cost autonomous surveillance.

The recent development of inexpensive micro air vehicles (MAV), also called drones, has allowed scientist and practitioners to build 2D and 3D maps of relatively large areas at a low cost [6]. In the last decade, mission planning [7, 8] and MAV control [9] have been widely studied so nowadays MAVs can perform surveillance missions with almost no human intervention. On the other hand, deep learning (DL) based on Convolutional Neural Networks (CNN) has become a powerful technique to solve many pattern recognition problems [10], for example, object recognition[11], facial key-point detection [12] or image captioning [13]. Unlike other machine learning techniques, where several manually designed features are extracted (for example, edges,

shapes or feature points) and passed to a classifier, DL automatically extracts relevant features and performs a logistic regression [33]. A branch of DL is the supervised DL, which uses already labeled examples to train a CNN by minimizing the error provided by a differentiable loss metric that compares the network prediction against the ground truth label. Despite the fast spreading of deep learning, such techniques require a large amount of data to perform well. This is a limitation for the recognition of all kind of vegetation. Our research focuses on the detection of endemic plants in the Tehuacan-Cuicatlan Valley. The first challenge we faced was that it did not exist an available dataset of images for this particular task. Some limited examples were available as illustrations used in a previous work [14]. Namely, neither a database of terrestrial images from vegetation nor a dataset of aerial images was available.

Figure 1: DJI Phantom 3 flying over a cactus forest during the dataset acquisition.

In this work, we present a deep learning based approach for recognizing the cactus species in the Tehuacán-Cuicatlán Valley. Our target is to start by recognizing these species and later extend the system to segment and recognize other ones. Due to the rich diversity of flora species and lack of a specific dataset, we have selected the *neobuxbaumia tetetzo* as the target to be recognized in this first study. The *neobuxbaumia tetetzo*, that we call columnar cactus for reading purposes, is an endemic plant of the Valley and it is spread in the majority of the terrain extension. In addition, we have focused our method in recognizing cactus from aerial images taken by small drones due to their low cost, see Fig. 1. Several challenges are presented in this images, one of them is the small area that the cactus occupy in the images, another one is that they are limited to the RGB spectrum, compared with other approaches where multispectral images are available. Under those circumstances, the proposed approach was developed to work with low resolution examples and RGB channels.

Our contributions are: i) a labeled dataset which required many on site flights along with a large number of human work hours for data labeling; available at [15], ii) the validation of a deep learning based approach for cactus recognition in aerial images, and iii) the analysis of the relevant features for cactus recognition.

The rest of the paper is organized as follows: section 2 presents a study of recent techniques for flora recognition in aerial images. Section 3 describes the methodology for image gathering,

labeling and dataset construction. Section 4 describes the proposed approach for cactus recognition. In section 5, we present the training of the neural network model and the validation of the approach. Finally, in section 6, we present the conclusion of the study as well as future research directions.

2.   Related Work

Automatic plant recognition has been an active research topic for several years, given its importance for different application fields such as preservation of natural areas, precision agriculture, disease detection, among others. A wide variety of techniques have been applied to address this problem, among which stand out classical pattern classification algorithms (Multilayer Perceptron, Support Vector Machines, Random Forest), image segmentation for feature extraction (shape, texture, vein pattern), multispectral/hyperspectral data, and in recent years deep learning.

Several works focused in plant identification use single leaf images and its characteristics for classification, this is mainly because leaves are more easily identifiable than other plant parts. The most commonly used features are shape, texture and vein patterns. Naresh and Nagendraswamy [16] introduced an approach for classification of medicinal plants. The proposed solution is based on a concept called Modified Local Binary Patterns to extract leaf texture. After processing, each image is represented by a histogram, considered a texture descriptor, that will be used for identification purposes. The chi-square distance combined with a nearest neighbor classification technique is used for matching unknown leaf sample of a species with the reference leaf stored in the knowledge base.

Since the task of recognizing plants is complex, sometimes using only one characteristic is not enough to perform an effective plant identification. Due to this fact some authors have decided to combine several characteristics of the plants to achieve an improvement in the classification performance of the proposed models. Chaki *et al*. presented a methodology for plant leaves recognition using a combination of texture and shape features [17]. A Gabor filter and gray level co-occurrence matrix were used to model texture, while shape was captured using curvelet transform coefficients together with invariant moments. Classification was tested with 2 different classifiers: a neuro-fuzzy controller and a feed-forward back-propagation multi-layered perceptron, to discriminate between 31 classes of leaves. Another approach was presented in [18],

where a combination of shapes features and color histogram for plant leaf recognition using characteristics as: length, width, area, and perimeter of the leaf, a distance map along the vertical and horizontal axes, color histogram and centroid-based radial distance map. For classification a K-nearest neighbor classifier was used.

The step of feature extraction can be complex and time consuming, and in the end it may not extract the actual relevant characteristics. In deep learning, the model learns and extracts the relevant features directly from the images during the training process. CNNs have shown better performance that other machine learning techniques for the task of image classification [10], what has led to its application in several fields such as object recognition [11], facial key-point detection [12] or image captioning [13]. Several works have been developed in the area of plant recognition taking advantage of the strengths that these models offer. In [19], the authors proposed to solve the problem of plant identification with a approach based on leaf vein morphology. The original images were processed to obtain vein pattern segmentation, then the images were cropped for central patch extraction. With the modified images a CNN was trained to classify three different legume species: white bean, red bean and soybean. A relevant approach for deep learning is a form of transfer learning. It is well known that deep learning networks need large amounts of data to be trained, this has led to the use of pre-trained networks that are latter fine-tuned with specific data of the addressed problem. The aim is to take advantage of the knowledge learned from one problem to apply it to another closely related problem. In [20], three recognized deep learning architectures, namely GoogLeNet, AlexNet, and VGGNet, were used to identify the plant species captured in a photograph. The authors studied the impact of critical factors (iteration size, batch size, and data augmentation) affect the fine-tuning of pre-trained models. DeepPlant is another example of a system for plant identification based on a CNN that uses a pre-training approach [21]. Background textures it is always an element that complicates the classification task. Xiao *et al*. proposed a segmentation and crop method, based on Fourier transformations and a K-means algorithm, to generate a region of interest in the images; the cropped images are then used to train a CNN for different plant species classification [22]. CNN have not been used only for taxonomic classification of plants, but also for the detection of diseases that can affect them [23].

The aforementioned methods focus on plant recognition based on single leaf information, but for monitoring purposes in natural area preservation or precision agriculture, a broader field coverage is necessary. To address this kind of problems several authors have resorted to remote

sensing as a useful tool for vegetation identification. Remote sensing is accomplished through the use of images obtained by satellites or Unmanned Aerial Vehicles (UAV). Different technologies have been applied for vegetation mapping and classification such as LiDAR [24], and multispectral / hyperspectral [25, 26] analysis to data obtained from airborne platforms. An alternative approach is to combine some of these techniques to increase classification accuracy, such as the case of the work presented in [27], where the authors proposed to merge information from hyperspectral imagery and structural metrics extracted directly from a 3-D LiDAR point cloud acquire at crown-level, for mapping diverse urban forests. Once information has been fused, classification was performed using canonical linear discriminant analysis.

Recent developments of cheap UAVs have promoted its use for diverse monitoring tasks, including natural areas supervision. This technology allows acquisition of high resolution images at a low cost and in near real-time [28]. Its use in precision agriculture has aroused great interest. Comba *et al*. presented a method to automatically detect vine rows from high resolution gray-scale aerial images [28]. The method combines a dynamic segmentation, a cluster algorithm based on Hough Parameters Space and Total Least Square technique. Another application for agriculture was introduced in [29] to address the problem of identifying infected areas of grapevines. Their solution uses images obtained from an UAV to get information of different colorimetric spaces and vegetation indices to feed a CNN to detect disease symptoms in vine yards. Given the limited space in UAV the selection of spectral band is crucial; to overcome this problem, Ishida *et al*. [30], adopted the use of liquid crystal tunable filter, which can transmit selected wavelengths without the need to exchange optical filters. A high resolution classification map was then produced from the aerial hyperspectral images using a support vector machine model.

3.   Dataset

Deep learning approaches require a vast amount of examples to perform well, otherwise the models can overfit to the data. For example, COCO (common objects in context) dataset has more than 330,000 of images [31]. Under those circumstances, it is essential to build our own dataset due to the lack of previous information. In this section, we present the methodology that we follow to build our dataset coupled with an analysis of it.

3.1.   *Case of study*

The Tehuacán-Cuicatlán Valley is located in the south of Mexico, between the states of Puebla and Oaxaca. The center of the area is located at 18.1605° latitude and -97.4202° longitude (decimal degrees) and covers an area of 10,000 km^2. See Fig. 2. Its arid condition is due to natural walls that surround it, by east the east mountain chain (*sierra madre oriental*) and by west and south the south mountain chain (*sierra madre del sur*). These walls inhibit the pass of fresh water from the Atlantic Ocean, leaving only annual precipitations between 400 and 800 mm [3].

Figure 2: Geographic location of the Tehuacán-Cuicatlán valley. The red polygon encapsulates the valley. The blue star indicates the location of the data collection.

Its location between two tectonic plates has motivated the flourish of a rich and unique ecosystem. In fact, it is the arid and semi-arid ecosystem with the bigger biodiversity from North-America [3], having a high ratio (30%) of endemic plants [32]. Among the diverse flora, there are some families which stand out from the rest because they have found an appropriate environment for diversification, they are the *asteraceae*, *cactaceae* and *poaceae* families. In particular, the *columnar cactaceae* family has in the region 45 species from the 70 in total that live in the Mexican country[3]. Furthermore, the columnar cactaceae family existing in this valley has been the focus of several researches given that many of its species are endemic to the region, which is why monitoring and conservation activities are of the utmost importance.

In several places of the Tehuacán-Cuicatlán Valley, the *columnar cactaceae* plants are very common so that they form cacti forests with a density of 1500 individuals per Ha. (See background of Fig. 1). Such forests are named in local language as "Tetechera" or "Cardonal". The cacti forests are spread in the Valley covering hundreds of square kilometers. According to Valiente *et al.*, [3] the forests can be classified depending on the contained flora. In their study 18 cacti forests are described.

In this study, we have investigated the area at the coordinates 18.1288799 Lat., -97.1639087 Lon.. This area is dominated by the Tetechera forest [3]. The species that are found in this area are: *Mimosa luisana, Agave karwinskii, Agave marmorata, Neobuxbaumia tetetzo, Verbesina neotenorensis, Bursera aloexylon, Cordia curassavica, Fouquieria formosa, Calliandra eriophylla, Ipomoea arborescens, Myrtillocactus geometrizans, Sanvitalia fruticosa,*

Ferocactus jlavovirens, Mammillaria spp, Ruellia rosea, Kanvinskia humboldtiana, resine calea, Opuntia pilifera, Ceiba parvifolia, Cathestecum brevifolium, Tillandsia makoyana and *Plumeria rubra.* See Fig. 3 where we present an example of the Tetechera forest landscape.

Figure 3: "Tetechera" (Spanish word) forest. The most visually outstanding plant in this forest is the *Neobuxbaumia tetetzo* which is the target of the object recognition.

Due to its endemic characteristic, in this study we have restricted the plant recognition to the *Neobuxbaumia tetetzo*, species that belongs to the *Cactaceae* family. From now on, we will refer to the *Neobuxbaumia tetetzo* as cactus in order to simplify the reading of the paper.

3.2.   Methodology

To collect the information, we did several flights over the coordinates (18.1288799 Lat., -97.1639087 Lon., 851 m. Alt.) near the Teotitlán district. The drone used was a DJI Phantom 3 Advanced. The mounted camera records video with $2704{\times}1520$ pixels at 24 fps. The flights were done manually at a flight altitude of 100 m. Once the videos were recorded they were split into images. Time period for sampling was 5 s. Then, the images were manually labeled. For each saved image, cacti found in the images were manually identified and marked. Next, for every marked cactus a patch is generated, which is a region of interest (ROI) containing the marked cactus. This ROI is a square area centered around the cactus. The right upper corner of the ROI is manually selected such that the entire cactus is covered. An example of the saved images and data is shown in Fig. 4. Next, a patch was generated for each marked cactus. Fig 5 shows some example patches. The different size of the objects in the image provoke patches of different size, so, we filtered the patches and we kept patches with a minimum resolution of $32{\times}32$ pixels. At the end of the labeling, a total number of 16,136 cactus examples were saved. All the captures images underwent this process to generate the examples belonging to the cactus class.

Figure 4: Manual labeling from the aerial images. The figure display an image captured by the micro aerial vehicle. Red circles mark the manually identified cacti by a human.

In order to later recognize the cacti in the images, we added a non-cactus class. It was build

by manually selecting the ground that do not have a cactus. All patches of the non-cactus class are the same size (32×32). Fig. 6 shows some examples of the non-cactus class. In total, 5,364 non-cactus examples were added to the dataset.

The built cactus dataset contains in total 21,500 examples with two classes. This dataset is available at [15] under GPL v3 license. We have observed that the dataset presents several challenges with respect to other plant datasets, i) low resolution examples, besides the drone camera has a high resolution, the space occupied by the objects are small due to the flight altitude and fish-eye camera effect, ii) reduced object shape, since the images are captured from a superior view, in most cases only the tree canopy is visible and iii) reduced spectral information, in many studies multispectral information is available, however, due to the low cost of the cameras, in this dataset only the RGB channels are available.

Figure 5: Examples from the extracted patches corresponding to the cacti found. These patches were manually labeled. Note that each patch has a different size.

Figure 6: Examples of the non-cactus class.

4.   Cactus Recognition

In this section, we present a deep learning based approach for automatic cactus recognition. Deep learning is a machine learning technique for pattern recognition [10]. Unlike other machine learning approaches, deep learning can perform automatic extraction of the features that describe the object or task. The supervised learning branch, applied in this work, uses a large set of examples for training a multilayer neural network (MNN). The MNN can be defined as a stack of layers composed by single nodes, called perceptrons, where each node first performs a linear combination of the inputs given a set of weights and biases, and second pass the result to an activation function. Therefore, for a given input, raw image pixels, the MNN outputs a vector of scores, one for each class. Since a random set of parameters (weights and biases) is highly unlikely to produce a correct output, the set of parameters is trained by adjusting the values given the error reported by the loss metric and the contribution of each parameter to that error (the contribution of each parameter is calculated by back-propagating the error gradient).

Figure 7: Custom LeNet-5 network for cactus recognition.

The proposed approach uses a type of MNN called convolutional neural network (CNN) [33], $\Phi_w : I \to \overline{y}$, whose input is an RGB image, I, the output is a predicted class label represented by a multinomial distribution, $\overline{y}$, and the parameters of the network are defined by w. Unlike the MNN, the CNN reduces the number of parameters by means of applying a stack of convolutions between the inputs and the parameters grouped in kernels. For this work, we define Φ as a modified version of LeNet-5 network proposed by LeCun *et al.* [34]. See Fig. 7. Our custom LeNet-5 network is configured as follows. It receives as input a 3-channel image with resolution $32 \times 32 \times 32$. Then, 6 convolution filters with size 5×5 are applied with stride one. Next, a max pooling operation is performed, using a kernel of 2×2 and stride two. Then, a new set of 16 convolution kernels with size 5×5 and stride one is applied. Again a max pooling operation is performed, using a kernel of 2×2 and stride two. Next, the features are flattened to a one dimension vector of size 400. Later, three fully connected layers are applied with 120, 84, and 2 nodes respectively. Until this point, the output of the CNN is a vector of real numbers called logits. Therefore, a *LogSoftMax* function is applied to convert the logits into a normalized probability distribution:

$$LogSoftMax(x_i) = \log\left(\frac{\exp(x_i)}{\sum_j \exp(x_j)} \right) \tag{1}$$

We are using *LogSoftMax* because, during training, it usually shows a better numerical performance with respect to *SoftMax* given that the log operation can undo the exp term [33].

To train the network Φ_w, the inner parameters, w, (weights and bias) have to be adjusted so that the output resembles the ground truth output. In this training process, the dataset images, **I**, are introduced to Φ_w in batches and the outputs are compared to the ground truth labels, y, using a Loss function. Such loss function measures the 'distance' of the output with respect to the ground truth. In this work, we are using as the loss function the negative logarithmic likelihood [33]:

$$L(\hat{y}, y) = \frac{1}{n} \sum_{i=1}^{n} -\log \hat{y}_{i,c=y} \tag{2}$$

where i is an element of the batch and c is the class index. Once the loss is calculated for a given batch, the inner parameters are adjusted using the back propagation algorithm. The iterative update for minimizing the loss is known as gradient descent optimization.

Training a neural network requires to set several hyper-parameters such as the learning rate, number of epochs and batch size. The learning rate is a special one since it defines how much the weights are 'moved' to decrease the loss. A bigger one could cause the network not to learn, on the other hand, a smaller one could require much more steps for learning. To decrease this issue, we train the network with Adam optimizer [35], a variant of the stochastic gradient optimization where the learning rate is adjusted automatically. The rest of the parameters are set empirically, more details are presented in the experiments section.

5.   Experiments

In this section, we present the experimental characterization of the proposed approach. The implemented network was trained with Adam optimizer [35] on a Inter Core i7 machine with NVIDIA GeForce 1080 GPU. The hyperparameters were set as follows: learning rate 0.01, number of epochs 150, batch size 2500.

5.1.   Data Augmentation Accuracy

A common practice to improve the accuracy of a deep learning model is to do data augmentation. In this section, we compare the accuracy of the network when data augmentation is included during training. The compared training variations are the following:

- No augmentation. The patches only are resized to 32×32 and normalized.
- Vertical and horizontal flip. In addition to the resize and normalization, the patches are flip vertically or horizontally with probability 0.5. Both flips are independent events.

The comparison is summarized in Figures 8, 9 and 10. We can observe that using data augmentation the loss decreases faster and the accuracy overcomes the accuracy of the no augmentation strategy for train and validation sets. In the validation test, the accuracy is very similar for both variations, however, at the end of the training, the use of vertical and horizontal flip overcomes the precision of no data augmentation. A peculiarity of this dataset is that in the first epochs of training (less than 20), even though the loss is decreasing the accuracy is not increasing, we believe that this phenomena is due to the random initial weights, which require

several updates before having impact on the accuracy.

Figure 8: Training loss.

Figure 9: Training accuracy.

Figure 10: Validation accuracy.

5.2.   *Accuracy Analysis*

As we have shown in the previous section, the best accuracy on the validation set was obtained using data augmentation, 0.98 and 0.95 for train and validation sets respectively. It is important to realize that the classes are imbalanced, cactus class has almost twice number of examples. Observing the non-normalized confusion matrix (Fig. 11), we note that the false positives for the cactus class are almost twice the false positive for the non-cactus class, this effect is due to the imbalanced classes. However, taking into account the normalized confusion matrix (Fig. 12) where false positives for both classes have the same value, we conclude that the errors are equally distributed among the classes.

Figure 11: Confusion matrix

Figure 12: Normalized confusion matrix

Figure 13: Example of a cactus image. The axes indicate pixel indices.

5.3.   *Learned features*

During training, CNN adjust their weights in order to decrease the loss. Weights from final layers are usually difficult to make sense for humans but weights in first layers are more easy to comprehend. In figure 14, we present the first layer's filters that were learned by the network along with the feature maps for the example cactus presented in Fig. 13. As can be seen, the first and second kernels of the CNN highlight some features such as the vertical edges, the third and fourth kernels perform a noise removal, the fifth kernel sharp some features and finally the sixth kernel

erase the background.

Figure 14: Learned kernels and feature maps. The axes indicate pixel indices.

5.4.   Dataset Extension

The proposed approach has shown good performance (accuracy) for cactus recognition in aerial images. One way to improve the current accuracy could be to make deeper the current CNN, however, improving the accuracy with a deeper network will tend to memorize the dataset due to the model capacity [33], causing model overfitting. A better improvement to the current work is to include more examples in the dataset. Such examples should be obtained under different conditions, some variations are different daylight conditions or different year seasons. Different season pictures will provide a distinctive context because the rainy season changes all valley's vegetation.

6.   Conclusions

We have presented a deep learning based approach for columnar cactus recognition. In the proposed approach, we have gathered on site thousands of images and we have labeled them manually. The built dataset has been used to train a modified version of the LeNet-5 neural network. Our experiments have shown that the proposed approach reaches a high recognition accuracy (0.95 for the validation set). This approach will be used to perform automatic surveillance of the Tehuacán-Cuicatlán valley with unmanned aerial vehicles. It is our hope that the automatic surveillance of this botanically and culturally rich region will stop its erosion due to the human activity, at the same time we hope that the acquired information will serve as a benchmark for future research on the natural evolution of the area. Our next step is to implement the proposed approach into a holistic surveillance system and to increase the number of species in the dataset.

Acknowledgments

References

[1] Unesco, Tehuacán-cuicatlán valley: originary habitat of mesoamerica, `https://whc.unesco.org/en/list/1534`, [online; visited on January 30 2019].

[2] B. P. V. Tapia, Estimación y actualización de la tasa de transformación del hábitat de las áreas naturales protegidas sinap i y sinap ii del fanp, Tech. rep., Fondo mexicano para la conservación de la naturaleza A. C. (2009).

[3] A. Valiente-Banuet, A. Casas, A. Alcántara, P. Dávila, N. Flores-Hernández, M. del Coro Arizmendi, J. L. Villaseñor, J. O. Ramrez, The vegetation of the valley of tehuacan-cuicatlan, Botanical Sciences (67) (2000) 25–74.

[4] E. Adam, O. Mutanga, D. Rugege, Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review, Wetlands Ecology and Management 18 (3) (2010) 281–296.

[5] J. S. Cope, D. Corney, J. Y. Clark, P. Remagnino, P. Wilkin, Plant species identification using digital morphometrics: A review, Expert Systems with Applications 39 (8) (2012) 7562–7573.

[6] F. Nex, F. Remondino, Uav for 3d mapping applications: a review, Applied geomatics 6 (1) (2014) 1–15.

[7] T. M. Cabreira, C. Di Franco, P. R. Ferreira, G. C. Buttazzo, Energy-aware spiral coverage path planning for uav photogrammetric applications, IEEE Robotics and Automation Letters 3 (4) (2018) 3662–3668.

[8] J. I. Vasquez-Gomez, M. Marciano-Melchor, J. C. Herrera-Lozada, Optimal coverage path planning based on the rotating calipers algorithm, in: International Conference on Mechatronics, Electronics and Automotive Engineering 2017 (ICMEAE), 2017.

[9] L. G. Carrillo, E. Rondon, A. Sanchez, A. Dzul, R. Lozano, Stabilization and trajectory tracking of a quad-rotor using vision, Journal of Intelligent & Robotic Systems 61 (1-4) (2011) 103–118.

[10] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (7553) (2015) 436.

[11] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.

[12] Y. Sun, X. Wang, X. Tang, Deep convolutional network cascade for facial point detection,

in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 3476–3483.

[13]   O. Vinyals, A. Toshev, S. Bengio, D. Erhan, Show and tell: Lessons learned from the 2015 mscoco image captioning challenge, IEEE transactions on pattern analysis and machine intelligence 39 (4) (2017) 652–663.

[14]   D. Hunt, et al., The new cactus lexicon: descriptions and illustrations of the cactus family, Milborne Port, Dorset: David Hunt Books 2.

[15]   J.   I.   Vasquez-Gomez,   Cactus   aerial   photos, `https://www.kaggle.com/irvingvasquez/cactus-aerial-photos`, [online; visited on February 10 2019] (2019).

[16]   Y. Naresh, H. Nagendraswamy, Classification of medicinal plants: An approach using modified lbp with symbolic representation, Neurocomputing 173 (2016) 1789–1797.

[17]   J. Chaki, R. Parekh, S. Bhattacharya, Plant leaf recognition using texture and shape features with neural classifiers, Pattern Recognition Letters 58 (2015) 61–68.

[18]   T. Munisami, M. Ramsurn, S. Kishnah, S. Pudaruth, Plant leaf recognition using shape features and colour histogram with k-nearest neighbour classifiers, Procedia Computer Science 58 (2015) 740–747.

[19]   G. L. Grinblat, L. C. Uzal, M. G. Larese, P. M. Granitto, Deep learning for plant identification using vein morphological patterns, Computers and Electronics in Agriculture 127 (2016) 418–424.

[20]   M. M. Ghazi, B. Yanikoglu, E. Aptoula, Plant identification using deep neural networks via optimization of transfer learning parameters, Neurocomputing 235 (2017) 228–235.

[21]   S. H. Lee, C. S. Chan, P. Wilkin, P. Remagnino, Deep-plant: Plant identification with convolutional neural networks, in: 2015 IEEE International Conference on Image Processing (ICIP), 2015, pp. 452–456.

[22]   Q. Xiao, G. Li, L. Xie, Q. Chen, Real-world plant species identification based on deep convolutional neural networks and visual attention, Ecological Informatics 48 (2018) 117–124.

[23]   S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, D. Stefanovic, Deep neural networks based recognition of plant diseases by leaf image classification, Computational Intelligence and Neuroscience 2016.

[24] L. Cao, N. C. Coops, J. L. Innes, J. Dai, H. Ruan, G. She, Tree species classification in subtropical forests using small-footprint full-waveform lidar data, International Journal of Applied Earth Observation and Geoinformation 49 (2016) 39–51.

[25] E. Adam, O. Mutanga, D. Rugege, Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review, Wetlands Ecology and Management 18 (3) (2010) 281–296.

[26] M. Dalponte, H. O. rka, T. Gobakken, D. Gianelle, E. NÃ¦sset, Tree species classification in boreal forests with hyperspectral data, IEEE Transactions on Geoscience and Remote Sensing 51 (5) (2013) 2632–2645.

[27] M. Alonzo, B. Bookhagen, D. A. Roberts, Urban tree species mapping using hyperspectral and lidar data fusion, Remote Sensing of Environment 148 (2014) 70–83.

[28] L. Comba, P. Gay, J. Primicerio, D. R. Aimonino, Vineyard detection from unmanned aerial systems images, Computers and Electronics in Agriculture 114 (2015) 78–87.

[29] M. Kerkech, A. Hafiane, R. Canals, Deep leaning approach with colorimetric spaces and vegetation indices for vine diseases detection in uav images, Computers and Electronics in Agriculture 155 (2018) 237–243.

[30] T. Ishida, J. Kurihara, F. A. Viray, S. B. Namuco, E. C. Paringit, G. J. Perez, Y. Takahashi, J. J. Marciano, A novel approach for vegetation classification using uav-based hyperspectral imaging, Computers and Electronics in Agriculture 144 (2018) 80–85.

[31] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft coco: Common objects in context, in: European conference on computer vision, Springer, 2014, pp. 740–755.

[32] C. E. Smith, Agriculture, Tehuacan Valley., Vol. 31, Fieldiana Botany, 1965.

[33] I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, Deep learning, Vol. 1, MIT press Cambridge, 2016.

[34] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (11) (1998) 2278–2324.

[35] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980.

Highlights
- Supervised Deep Learning approach for cactus recognition in aerial images.
- Datasets generation methodology for aerial plant recognition.
- The collected dataset contains more than 10,000 examples.
- LeNet-5 Neural Network recognizes cactus with a high validation accuracy.
- Data augmentation improves the validation accuracy.