Deep Convolutional Neural Network Feature Extraction for Berry Trees Classification

Jolitte A. Villaruz
Technology Department, Aklan State University - Kalibo Campus, Kalibo, Aklan, Philippines
Email: jvillaruz@asu.edu.ph

***Abstract*—To support biodiversity conservations, plant classification studies, particularly from images, are necessary. This study explores the use of the deep convolutional neural network as a feature extractor to a plant classification problem. An original dataset consisting of images of seedlings of the three most important berry trees belonging to the Philippine indigenous plants was used. The result shows that as the network layers are getting deeper, they are becoming better at extracting discriminative features, such that, irrespective of classifier used their prediction performance keeps on improving. When the different layers were individually visualized, the features extracted were far from random, uninterpretable patterns. Rather, they show relevant properties that are capable of sorting patterns progressively from low to higher level. Hence, for classification problems bounded with the limitation of data, time, and computational hardware, leveraging the representational power of the deep convolutional neural network is very useful.**

***Index Terms*—feature extraction, deep convolutional neural network, deep learning, AlexNet, plant classification, SVM**

I. Introduction

Plants, particularly berry types of Philippine indigenous trees are vital food sources and are known to have distinct properties, useful in specialized drug formulation [1]. Yet, despite their utmost economic and environmental importance, these berry trees are underutilized and less-known, even to Filipinos [2]. In addition, the richness of these plants is threatened and their extinction rates escalate due to climate disruptions [3] and elevated by the direct and indirect human exploitation [4].

According to the Philippine Bureau of Agricultural Research and the Department of Agriculture, utilizing and propagating these plants will help safeguard their rate of survival [2]. To support biodiversity conservations, plant classification studies, particularly from seedling images are helpful. This will help identify useful seedlings and prevent them from being weeded out. Furthermore, this will assist farmers to properly manage these plants in order to improve their agricultural productivity and sustainability [5].

Plant taxonomy or classification is the science of identifying, describing and naming plants. Classification can be done by associating one or more discriminating features of a plant to its common or scientific name. However, a plant belonging to the same class have subtle differences, making it hard for non-experts to classify them into different species. Besides this, taxonomic work is a highly technical skill, requiring expertise that can only be attained over intensive training and experience. In addition, there is a limited and yet declining number of these skilled taxonomists to classify the more than 450,000 plant species on earth [5], [6].

Today, manual taxonomic tasks have greatly improved by the recent advancement of technology. This innovation is greatly influenced by the application of computer vision and machine learning techniques. At present, the state-of-the-art solution includes classification based on digital images. In this context, classification is defined as the process of predicting as to in which category does the new and unseen images belong based on discriminating features learned from labeled training images. With the availability of more sophisticated and efficient way of image-based plant classification, indeed, automatic plant recognition nearly comes into reality [7].

A manifold of successful applications in the field of plant classification was found in the literature. Shallow machine learning techniques were used in [8]-[10]. Whereas easier implementations were found in [11]-[13] where deep learning via transfer learning was utilized. In addition, faster solutions using feature extraction using the deep Convolutional Neural Network (CNN) were implemented in [14]-[18]. Comprehensive reviews [5], [6], [19] were even conducted, offering solutions and presenting algorithms for the same purpose.

The general workflow for an image-based plant classification task is shown in Fig. 1. Accessibility to relevant technologies including smartphones, digital cameras and the remote access to databases allow easy acquisition of plant images. Also, advances in image processing provided various preprocessing techniques to make the images suitable for feature extraction. Feature extraction is the step undertaken in order to find discriminating features that will serve as the basis for classification. Once the discriminating feature has been extracted, the classification task can be carried out using several machine learning techniques like Support Vector Machines (SVM), Naïve Bayes, k-Nearest Neighbor (kNN), and CNN.

In this study, deep CNN, particularly the widely used pre-trained AlexNet [20] (detailed discussion is found in

Manuscript received November 23, 2020; revised April 22, 2021

doi: 10.12720/jait.12.3.226-233

Section II-B) model is being explored as a feature extractor. The extracted features are then applied to multiclass SVM to classify the images of the seedlings of the three most important berry trees belonging to the Philippine indigenous plants. Likewise, the end-to-end approach using the softmax activation function was also implemented.

A brief introduction to deep CNN and its application to plant classification is provided in Section II. Section III describes the details of the system design and implementation. Results and analysis are discussed in Section IV. Finally, Section V wraps up the paper with a discussion of the conclusion and future works.

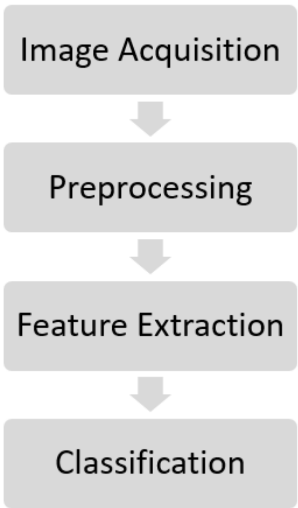

Figure 1. The general workflow of an image-based plant classification task.

II. RELATED LITERATURE

An introduction to deep CNN and AlexNet model is discussed in this section. Also, a brief overview of studies utilizing deep CNN as a feature extractor in plant classification problems is presented.

A. Deep Convolutional Neural Network

Basically, deep CNN is consist of a feature extractor network and classification neural network as shown in Fig. 2. The feature extractor is consist of alternating stacks of the convolution layer and the pooling layer pairs. Using the convolution operation, the convolution layer generates images known as feature maps that highlight the features of the input image. While the pooling layer combines the adjacent pixels as one, reducing the dimension of the image [21].

The convolutional layers are followed by one or more fully connected layers. Each fully connected layer combines all of the features learned by the previous layers across the images to identify the more intricate features. The last fully connected layer acts as the classifier.

Contrary to shallow machine learning techniques that require the feature extractor to be designed manually, deep CNN includes the feature extractor in the training process. This ability to turn the manual feature extraction to an automated one is the primary advantage of deep CNN.

Deep CNN provides the state-of-the-art solution in the areas of image recognition [20], [22]-[24], speech recognition [25]-[27], and exceptional capability in various natural language tasks [28]-[30]. As a specialized machine learning technique, deep CNN is best applied to problems where there is a substantial amount of training data. It is also vital that pattern should exist from these data, but there is no formula that can pin it down mathematically [31]. When these conditions are met, the application of deep CNN can bring tremendous success not only in plant classification tasks but even in detecting diseases like cancer [32].

B. AlexNet Model

AlexNet [20] model was trained with 1000 different classes of images from ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [33]. It has eight hidden layers, comprising of five convolutional and three fully-connected layers. Of the five convolutional layers, three of which are followed by max-pooling layers, making them accountable for feature learning. These layers are the first, second and fifth layers.

The first convolutional layer filters the $224\times224\times3$ input image with 96 kernels of size $11\times11\times3$ with a stride of 4 pixels. The output of the first convolutional layer is taken as the input to the second convolutional layer takes and filters it with 256 kernels of size $5\times5\times48$. The third, fourth, and fifth convolutional layers are connected to one another without any intervening pooling or normalization layers. The third convolutional layer has 384 kernels of size $3\times3\times256$ connected to the outputs of the second convolutional layer. The fourth convolutional layer has 384 kernels of size $3\times3\times192$, and the fifth convolutional layer has 256 kernels of size $3\times3\times192$.

The fully connected layers have 4096 neurons each. The last fully-connected layer act as the classification layer, ending with a softmax activation function. Fig. 3 shows AlexNet architecture. A more detailed overview of this architecture can be found for reference in [20].

The AlexNet model has a linear architecture that permits easy visualization of the different convolutional layers and also enables learning to occur in hierarchical ways.

C. Related Works

Leaf, stem, flower, and fruit are the common plant organs utilized in the plant classification tasks. Most studies consider one plant organ for classification, although more recent studies explored the use of the entire plant or multi-organ based classification. Since plant leaves are highly available at any time of the year, it's the most explored plant organ for classification [8]-[10], [14], [17], [34]-[36].

For any image-based classification task to be successful, a large collection of images at the rate of a thousand is needed. But, despite this number, these images are merely considered a collection of pixels associated with color information by machine learning algorithms. Further preprocessing is therefore required in order to extract discriminating features or useful patterns from these ambiguous data.

When using the traditional technique, the knowledge of a domain expert is needed to manually find and extract discriminating features. While yielding inconsistent level of performance, this task requires a significant amount of cost and time. Similarly, finding these features is not an easy task, more so that there is no single feature that can sufficiently distinguish plant species from another [6]. In fact, in the last decades, 90% of the development effort is devoted to the detection and extraction of useful features alone [37]. For the past years, this has gradually improved with the use of generic computer vision object recognition features algorithms, like Scale-Invariant Feature Transform (SIFT), Histogram of Gradients (HoG), Textons, Rotation-Invariant Generalization (RIFT), Speeded-Up Robust Features (SURF), and Gradient Location-Orientation Histogram (GLOH). Various plant classification studies successfully employed these algorithms [8], [9], [38]. However, the performance of these algorithms has plateaued in the years of 2010 to 2012 [39]. In 2012, Krizhevsky *et al.* [20] rekindled the forgotten deep CNN since its successful implementation in [40]. The use of deep CNN addresses the limitations of the previous generic algorithms. Since then, no other implementation has shown a better result than what deep CNN has accomplished in feature extraction, object recognition, classification tasks.

For instance, the study of Razavian *et al.* [41] yielded a very good result when deep CNN was used as a feature extractor in objects classification, scene recognition, and fine-grained recognition, attribute detection, and image retrieval applied to a diverse set of datasets.

Lee *et al.* [16], [42] implemented deep CNN to extract to classify 44 different plant species, collected at the Royal Botanic Gardens, Kew, England. Their result established that venation structure is an important feature to identify different plant species, and when fit to multilayer perceptron yielded 99.6% classification accuracy. Furthermore, they proved that combining both local and global features can better improve classification accuracy.

A novel approach to extract features from images based on deep CNN was proposed by Tan *et al.* [14] to classify 43 plant species of tropical trees collected from three locations in the University of Malaya, Kuala Lumpur, Malaysia. Their proposed model that ended with a softmax classification layer resulted in 94.88% classification accuracy.

Simon *et al.* [43] also used two deep CNN architectures, the AlexNet, and VGG19 as feature detector and extractor inside a part constellation modeling framework. Extracted features were then fit to SVM classifier. Using Oxford Flowers 102 dataset, they yielded 95.34% classification accuracy.

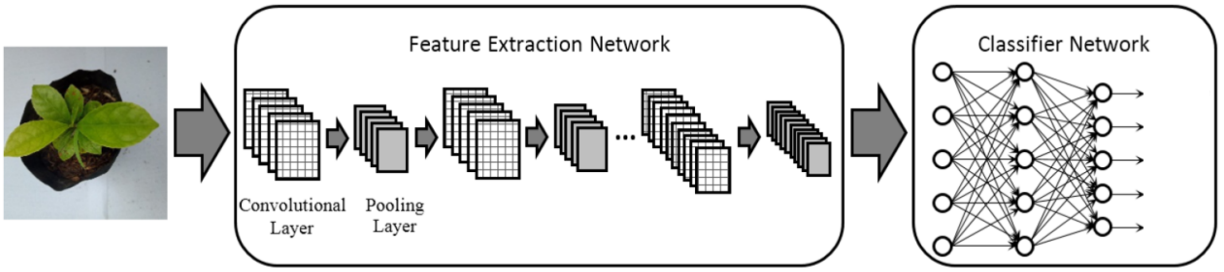

Figure 2. A typical deep CNN architecture.

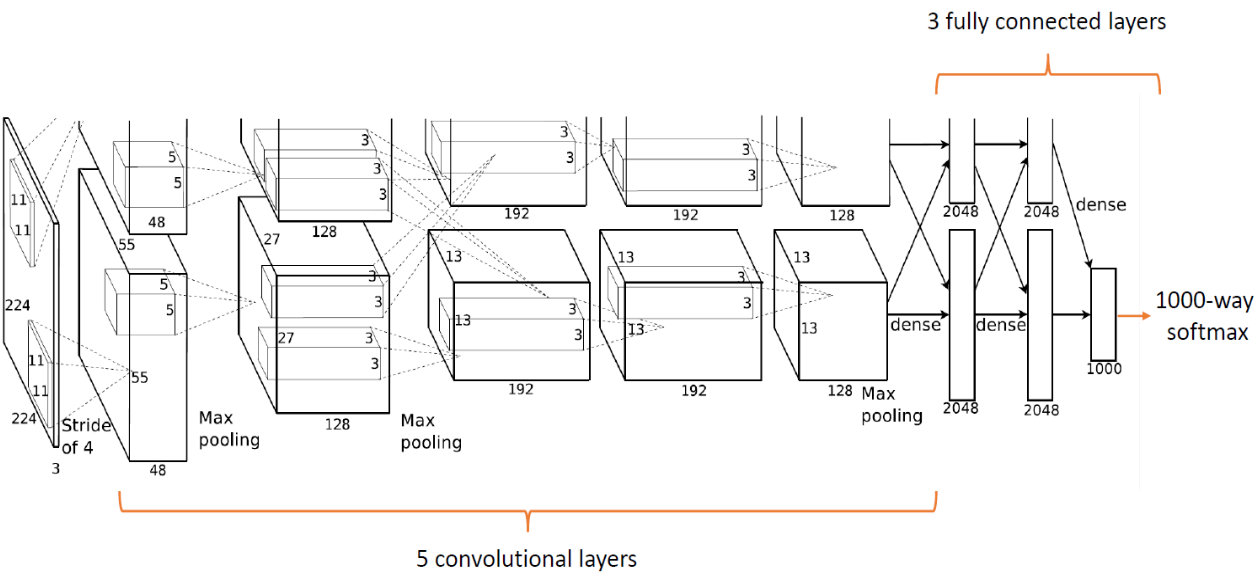

Figure 3. The AlexNet architecture [20].

III. System Design and Implementation

While there are better deep CNN architectures such as ResNet [23] that have provided nearly or beyond human classification accuracy, this study still prefers to employ the pre-trained AlexNet model as feature extractor due to the linearity of its architecture. By using this implementation, visualization to the different network layers can easily be performed. In addition, this provides the fastest way of using the representational power of pre-trained deep CNN's as compared to transfer learning and training the model from scratch.

This study follows the general workflow of an image-based plant classification task as presented in Fig. 1, consisting of four steps, namely image acquisition, preprocessing, feature extraction and classification. In the first step, images of seedlings of the three types of berry trees belonging to Philippine indigenous plants were acquired. These images were then preprocessed. This is then followed by extracting discriminating features using deep CNN. Finally, the extracted features were fit to a multiclass SVM for classification.

To further evaluate the functionality of the extracted features using the AlexNet model, it was also implemented in an end-to-end approach using softmax activation function as the classifier. The result of the classification of both the multiclass SVM and softmax activation function were then compared.

The implementation was done using MATLAB 2018a on a computer equipped with Intel Core i5-7200U CPU @2.50 GHz and 2.71 GHz, with NVIDIA GEFORCE 940MX GPU with CUDA enabled and 16GB of RAM.

A. Image Acquisition

An original dataset consisting of images of seedlings of the three most important types of berry trees belonging to the Philippine indigenous plants was utilized in this study. Plant species include Bignai (*T. Antidesma bunius Spreng*.), Agosip (*Symplocos cochinchinensis*), and Lipote (*Szygium curranii*).

Bignai seedling is cotyledons elliptic to almost orbicular, about 12-14 × 9-10 mm, base cuneate to obtuse, apex obtuse. At the tenth leaf stage: leaves are hairy on both the upper and lower surfaces along the midrib; scattered hairs visible with a lens elsewhere; petiole hairy; stipules filiform, about 10-12 mm long, hairy. Agosip seedling is cotyledons linear, also about 12-15 × 2 mm. First pair of leaves ovate, about 10-15 × 6-7 mm, margins with 1-4 teeth on each side of the leaf blade. At the tenth leaf stage: leaf blade elliptic, apex acuminate, base cuneate to attenuate, margin serrate, glabrous on the upper surface; petiole glabrous. Lipote leaves are alternate, oblong-lanceolate or obovate, acuminate, 6 to 20 centimeters long, 4 to 7 centimeters wide, with 14 to 16 pairs of secondary veins.

Each plant was raised individually from seeds at the Clonal Nursery of Aklan State University, Banga, Aklan, Philippines. During image acquisition, plants were one to four months old. In this study, each species of berry tree comprised of 500 different images. These were captured at varying periods at daytime from January to March 2018 by different smartphones in a natural environment with variable elevation and lighting conditions. Images of the whole plant seedlings were captured and were saved in jpg format. The horizontal and vertical resolution of the images ranges from 72 to 96 dpi, all in RGB format, and sizes vary from 2448×3264 to 3120×4160.

Full annotation of these images was conducted by three domain experts. A sample of resized images that were utilized in the study is shown in Fig. 4.

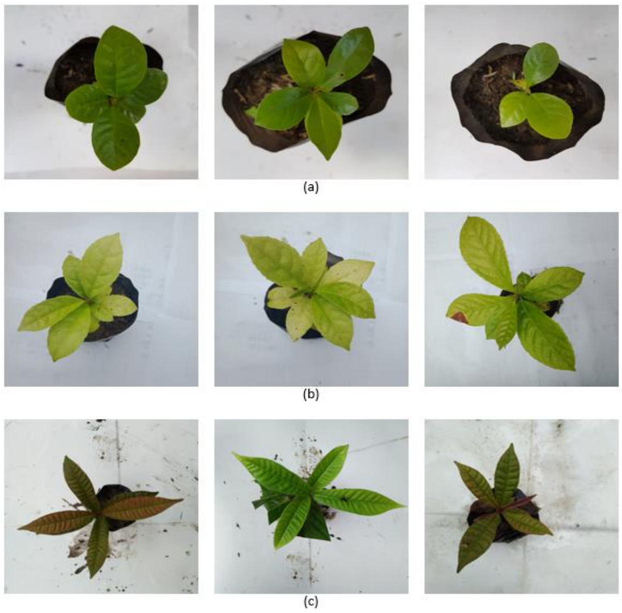

Figure 4. Sample images of the three types of berry trees belonging to the Philippine indigenous plants that were utilized in the study. (a) Bignai (*T. Antidesma bunius Spreng.*); (b) Agosip (*Symplocos cochinchinensis*); and (c) Lipote (*Szygium curranii*).

B. Image Preprocessing

When implementing the pre-trained AlexNet model, it is required that the input image should be resized to match with that of the model. Further, to build a powerful image classifier that avoids the effect of overfitting and prevents the network from memorizing the exact details of the training images, a variety of label preserving image transforms were performed. These are the random rotation of images with an angle up to 360 degrees, random horizontal reflections and random vertical reflections, and a random translation of up to 30 pixels horizontally and vertically. It should also be noted that during this process, no fine-tuning of the parameters of AlexNet was made. Other than resizing, no augmentation was done in the testing set.

C. Feature Extraction

Indeed, the remarkable success of the use of deep CNN for plant classification tasks is undisputable. However, even with this impressive classification performance, the use of this technique can perhaps be considered a black box [17] with no clear interpretation as to why this technique is performing so well [44]. Moreover, the implementation deep CNN is computationally expensive, requiring the use of a large dataset, coupled with powerful computational hardware, and significant training time.

To leverage on the representational power of pre-trained deep networks while using less powerful computing hardware and lesser training time, this study utilized AlexNet model as a feature extractor. Features maps formed from the five convolutional layers and two fully connected layers of AlexNet were individually extracted in this phase. The feature extraction approach employed in this study is shown in Fig. 5.

D. Classification

Images of seedlings of the three most important types of berry trees belonging to the Philippine indigenous plants were utilized in this study. The dataset was randomly partitioned into 70% training and 30% testing sets. In the classification phase, features that were extracted from five convolutional layers and two fully connected layers were fit to a multiclass SVM classifier. SVM is a powerful discriminative classifier based on supervised machine learning approach. Supplied with ample amount of labeled training data, it can be able to deal with high dimensional space and data points that are not linearly separable. SVM was chosen to act as classifier because it can be implemented faster even with low-storage devices.

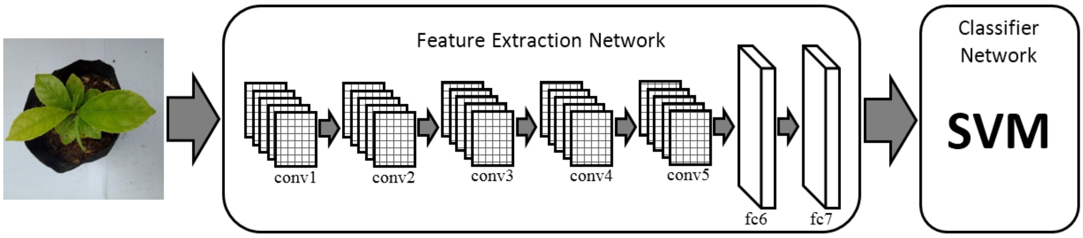

Figure 5. The outputs of the feature extraction network comprising of the five convolutional layers and two fully connected layers were fit individually to the SVM for classification.

IV. Results and Analysis

This study was able to extract features learned from the different layers of the pre-trained AlexNet model useful in classifying three of the most important types of berry trees belonging to the Philippine indigenous plants. To overcome the effects of overfitting, the 1500 images was increased to five times using different image augmentation techniques.

To evaluate how useful the extracted features are, each of the extracted features in every layer was fit individually to a multiclass SVM for classification.

The performance metric is defined as,

$$Accuracy = \frac{N_c}{N_t} x100\% \tag{1}$$

where N_c is the number of accurate prediction and N_t is the total number of test images.

Table I summarizes the classification accuracy on test images when discriminative features learned from the different layers of AlexNet including the five convolutional layers and two fully connected layers of the pre-trained AlexNet model were fit to a multiclass SVM classifier. It should be noted that conv1 to conv5 refer to convolutional layers 1 to 5, while fc6 and fc7 denote fully connected layers 6 and 7, respectively.

TABLE I. Classification Accuracy on Test Images when Discriminative Features Learned from the Different Layer of the Pre-trained AlexNet Model Were Fit to a Multiclass SVM Classifier

AlexNet Network Layers	Classification Accuracy (%)
conv1	59.30
conv2	87.60
conv3	71.60
conv4	89.80
conv5	95.30
fc6	97.80
fc7	97.10

As expected, classification accuracy improves with the increasing depth of the layers. Fifth convolutional layer “conv5" and the two fully connected layers “fc6” and “fc7” of the AlexNet model are powerful feature extractors, as when extracted features from these layers were classified using multiclass SVM, they yielded above 95% accuracy. As observed, deeper layers generally, are better at extracting the discriminant information. This finding agrees with [18], [39], [44]-[47], that the deeper the layer, the more abstractive and more dataset-specific features become. Further, the earlier layers have poorly discriminated plants as to species because so far, it has just learned the more generic and simple features, making it hard for the classifier to distinguish one berry species from another.

In addition, to compare the outcome of the softmax activation function, with that of the multiclass SVM, the result of both classifiers in terms of testing and training time, including classification accuracy is reflected in Table II.

TABLE II. Comparison of Softmax Activation Function and Multiclass SVM Classifiers to Predict Species of Berry Trees

Classifiers	Training and Testing Time	Classification Accuracy (%)
softmax activation function	239 min 33 sec	97.80%
SVM	4 min 57 sec	97.10%

As revealed from this result, regardless of the classifier used, features that were extracted from the input images using deep CNN technique are useful in predicting as to in which species the new and unseen images of berry seedlings belongs. The result also shows that the classification accuracy is high for both softmax activation function and multiclass SVM classifiers. Also, both classifiers almost yielded the same classification accuracy. However, in terms of training and testing time, a significant difference was observed. This simply proves that using deep CNN as a feature extractor is a very useful technique as this provides the fastest implementation of deep CNN in image classification problems.

To be able to understand the internal behavior of the AlexNet model, the output of visualization per convolutional layer is illustrated in Fig. 6. Each figure depicts a montage of the images comprising of the output activations per convolutional layer. There are 96, 256, 384, 384, and 256 feature maps in the first, second, third, fourth, and fifth convolutional layers, respectively.

Several channels comprise areas of activation both in white and black pixels. White pixels depict strong positive activations while black pixels represent strong negative activations. Channels denoted by gray are not strongly activated on the input image. The position of a pixel corresponds to the same position as of the original image. Each layer is a new representation of an input image were discriminative features are gradually extracted.

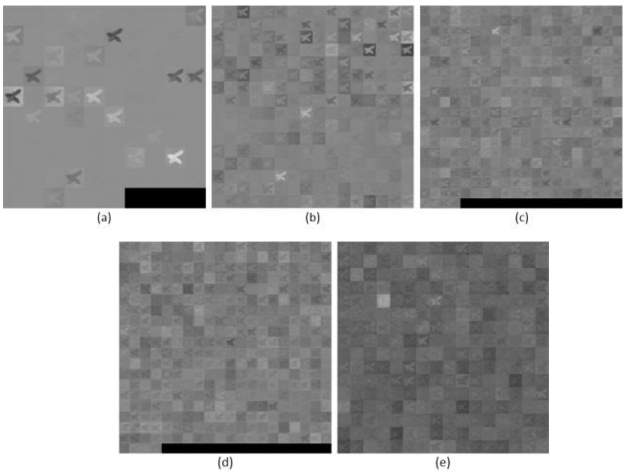

Figure 6. From left to right, top to bottom, (a) to (e) shows the internal behavior constructed by the output activations per convolutional layer of the AlexNet model.

Fig. 6 depicted too many images to investigate in details. For better understanding, the strongest activations within the different layers of AlexNet is shown in Fig. 7. As revealed, visualization clearly responds to the structure of the input image. When compared to the original image, channels in the first two layers learn more general, simple and low-level features like outlines and edges. Subsequent layers collect and combine the features in the earlier layers. Contour, base, and shape of the plant become more evident in the third to fifth layers. It can be also observed that dissimilar information is removed gradually from low to high layers. Consequently, channels in the deeper layers learn more-data specific or high-level features, suggesting that the network constructs a hierarchical representation of input images.

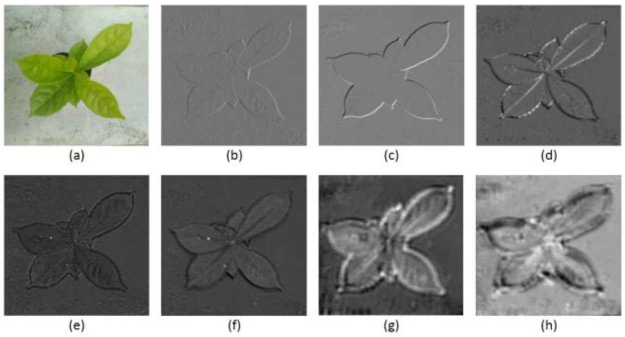

Figure 7. Visualization of the strongest activation from the different layers of AlexNet. (a) original image; (b) conv1; (c) conv2; (d) conv3; (e) conv4; (f) conv5; (g) fc6; (h) fc7.

V. Conclusion and Future Works

This study supports the idea that the features extracted from within the layers of AlexNet were far from random, uninterpretable patterns. Rather, they show relevant properties that are capable of gradually sorting patterns from low to higher level. Through the comparison of classification accuracy as to the features learned with different depths, it shows that deeper CNN is better at discriminant information extraction, thus improves the prediction performance regardless of the classifier used (either SVM or softmax). Hence, when data, computational hardware, and training time are constrained, leveraging on the representational power of deep CNN such as AlexNet is very useful.

In the future, state-of-the-art techniques to detect and locate plant seedling images from the background will be explored, such that the deep CNN will only focus on the discriminative features rather than the background. The use of more sophisticated machine learning techniques such as CNN for classification will also be applied.

Conflict of Interest

The author declares no conflict of interest.

Acknowledgment

The author would like to thank the anonymous referees for their valuable comments and helpful suggestions. The support of Aklan State University through the Research and Development Services unit and the help extended by Dr. Rogelio L. Felizardo, Mr. Dennis M. Barrios II, and Prof. Julie Ann A. Salido is highly acknowledged. The field workers of the Clonal Nursery of Aklan State University and the efforts of Foresters Hanna Mae T. Ibuyan and Darlyn Joy I. Galicha in taking the photos of plant seedlings are particularly valued.

References

[1] R. T. D. Cruz, "Promoting the less-known, phytochemical-rich pinoy fruits," *DAR Res. Dev. Dig.*, vol. 14, no. 1, pp. 4-5, 2012.

[2] N. P. Eleazar, "Recognizing the importance of underutilized fruit crops," *DAR Res. Dev. Dig.*, vol. 14, no. 4, p. 3, 2012.

[3] S. L. Pimm, *et al.*, "The biodiversity of species and their rates of extinction, distribution, and protection," *Science*, vol. 344, no. 6187, 2014.
[4] G. E. P. Murphy and T. N. Romanuk, "A meta-analysis of declines in local species richness from human disturbances," *Ecology and Evolution*, vol. 4, no. 1, pp. 91-103. 2014.
[5] E. Mata-Montero and J. Carranza-Rojas, "Automated plant species identification: Challenges and opportunities," in *Proc. IFIP World Information Technology Forum*, 2016, pp. 26-36.
[6] J. Wä ldchen and P. Mäder, *Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review*, Springer Netherlands, 2018, vol. 25, no. 2.
[7] J. Wä ldchen, M. Rzanny, M. Seeland, and P. Mäder, "Automated plant species identification — Trends and future directions," *PLoS Comput. Biol.*, vol. 14, no. 4, p. e1005993, 2018.
[8] Q. K. Nguyen, T. L. Le, and N. H. Pham, "Leaf based plant identification system for android using surf features in combination with bag of words model and supervised learning," in *Proc. International Conference on Advanced Technologies for Communications*, 2013, pp. 404-407.
[9] X. Y. Xiao, R. Hu, S. W. Zhang, and X. F. Wang, "HOG-based approach for leaf classification," in *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence*, Springer, 2010, pp. 149-155.
[10] S. Prasad, K. M. Kudiri, and R. C. Tripathi, "Relative sub-image based features for leaf recognition using support vector machine," in *Proc. Int. Conf. Commun. Comput. Secur.*, 2011, p. 343.
[11] J. A. Villaruz, J. A. A. Salido, D. M. Barrios, and R. L. Felizardo, "Philippine indigenous plant seedlings classification using deep learning," in *Proc. IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management*, 2018.
[12] M. M. Ghazi, B. Yanikoglu, and E. Aptoula, "Plant identification using deep neural networks via optimization of transfer learning parameters," *Neurocomputing*, vol. 235, pp. 228-235, 2017.
[13] M. Dyrmann, H. Karstoft, and H. S. Midtiby, "Plant species classification using deep convolutional neural network," *Biosyst. Eng.*, vol. 151, pp. 72-80, 2016.
[14] J. W. Tan, S. Chang, S. Abdul-kareem, H. J. Yap, and K. Yong, "Deep learning for plant species classification using leaf vein morphometric," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 17, no. 1, 2018.
[15] S. H. Lee, C. S. Chan, P. Wilkin, and P. Remagnino, "Deep-Plant: Plant identification with convolutional neural networks," in *Proc. IEEE International Conference on Image Processing*, 2015, pp. 452-456.
[16] S. H. Lee, C. S. Chan, S. J. Mayo, and P. Remagnino, "How deep learning extracts and learns leaf features for plant classification," *Pattern Recognit.*, vol. 71, pp. 1-13, 2017.
[17] S. Bertrand, *et al.*, "Visualization of leaf botanical features extracted from AlexNet convolutional layers," in *Proc. International Workshop on Image Analysis Methods for the Plant Sciences*, 2018.
[18] J. Donahue, *et al.*, "DeCAF: A deep convolutional activation feature for generic visual recognition," in *Proc. International Conference on Machine Learning*, 2014.
[19] J. S. Cope, D. Corney, J. Y. Clark, P. Remagnino, and P. Wilkin, "Plant species identification using digital morphometrics: A review," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7562-7573, 2012.
[20] A. Krizhevsky, I. Sulskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 60, no. 6, pp. 84-90, 2012.
[21] P. Kim, *MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence*, Apress, 2017.
[22] C. Szegedy, *et al.*, "Going deeper with convolutions," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1-9.
[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.
[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556, 2015.
[25] T. Mikolov, A. Deoras, D. Povey, L. Burget, and J. Černocký, "Strategies for training large scale neural network language models," in *Proc. IEEE Workshop on Automatic Speech Recognition and Understanding*, 2011, pp. 196-201.
[26] G. Hinton, *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82-97, 2012.
[27] T. N. Sainath, A. Mohamed, B. Kingsbury, and B. Ramabhadran, "Deep convolutional neural networks for LVCSR," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 8614-8618.
[28] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Advances in Neural Information Processing Systems*, pp. 3104-3112, 2014.
[29] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12, pp. 2493-2537, 2011.
[30] S. Jean, K. Cho, R. Memisevic, and Y. Bengio, "On using very large target vocabulary for neural machine translation," arXiv Prepr. arXiv1412.2007, 2014.
[31] Y. S. Abu-Mostafa, M. Magdon-Ismail, and H. T. Lin, "Learning from data. A short course," view article on Google Scholar, 2012.
[32] J. A. A. Salido and C. Ruiz, "Using deep learning to detect melanoma in dermoscopy images," *Int. J. Mach. Learn. Comput.*, vol. 8, no. 1, pp. 61-68, 2018.
[33] O. Russakovsky, *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211-252, Dec. 2015.
[34] S. G. Wu, F. S. Bao, E. Y. Xu, Y. X. Wang, Y. F. Chang, and Q. L. Xiang, "A leaf recognition algorithm for plant classification using probabilistic neural network," in *Proc. IEEE International Symposium on Signal Processing and Information Technology*, 2007, pp. 11-16.
[35] M. Sulc and J. Matas, "Texture-Based leaf identification," in *Proc. Conference on Computer Vision*, 2014, pp. 185-200.
[36] O. Söderkvist, "Computer vision classification of leaves from swedish trees," master thesis, Linkoping University, 2001.
[37] P. Barré, B. C. Stöver, K. F. Müller, and V. Steinhage, "LeafNet: A computer vision system for automatic plant species identification," *Ecol. Inform.*, vol. 40, pp. 50-56, 2017.
[38] P. Koniusz, F. Yan, P. H. Gosselin, and K. Mikolajczyk, "Higher-Order occurrence pooling for bags-of-words: Visual concept detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 313-326, 2017.
[39] R. Girshick, J. Donahue, T. Darrell, U. C. Berkeley, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2012.
[40] Y. Le Cun, *et al.*, "Handwritten digit recognition with a back-propagation network," *Adv. Neural Inf. Process. Syst.*, pp. 396-404, 1990.
[41] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 806-813.
[42] S. H. Lee, *et al.*, "Deep-Plant: Plant identification with convolutional neural networks," *Signal Process. Inf. Technol.*, vol. 5, pp. 452-456, 2016.
[43] M. Simon and E. Rodner, "Neural activation constellations: Unsupervised part model discovery with convolutional networks," in *Proc. the IEEE International Conference on Computer Vision*, 2015, pp. 1143-1151.
[44] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. European Conference on Computer Vision*, 2014, pp. 818-833.
[45] L. Ding, H. Li, C. Hu, W. Zhang, and S. Wang, "Alexnet feature extraction and multi-kernel learning for object-oriented classification," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 42, no. 3, pp. 277-281, 2018.
[46] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," arXiv preprint arXiv:1405.3531, 2014.
[47] W. Yu, K. Yang, Y. Bai, T. Xiao, H. Yao, and Y. Rui, "Visualizing and comparing AlexNet and VGG using deconvolutional layers," in *Proc. the 33rd International Conference on Machine Learning*, 2016.

Jolitte A. Villaruz obtained her bachelor's degree in computer engineering and master's degree in engineering from Western Institute of Technology in 2001 and 2005, respectively. She received her master's degree in information technology from Cebu Institute of Technology-University in 2012 as a Commission on Higher Education (CHED) Faculty Development Program (FDP) scholar. Her doctorate degree was obtained from the University of the Visayas in 2017. She has conducted research on fuzzy logic and deep learning. She is currently an associate professor at the College of Industrial Technology of Aklan State University, Kalibo, Aklan, 5600 Philippines.