Leaf Spot Attention Network for Apple Leaf Disease Identification

Hee-Jin Yu and Chang-Hwan Son
Kunsan National University
558 Daehak-ro Gunsan-si 54150, Republic of Korea
cson@kunsan.ac.kr

Abstract

Although new deep learning approaches have recently been introduced for leaf disease identification, existing deep learning models such as VGG and ResNet have been used previously. Therefore, a new deep learning architecture is proposed to consider the leaf spot attention mechanism. The primary idea is that leaf disease symptoms appear in the leaf area, whereas the background region does not contain any useful information regarding leaf diseases. To realize this, two subnetworks are designed. The first is a feature segmentation subnetwork to provide more discriminative features for the separated background, leaf areas, and spot areas in the feature map. The other is a spot-aware classification subnetwork to increase the classification accuracy. To train the proposed leaf spot attention network, the feature segmentation subnetwork is first learned with a new image set, where the background, leaf area, and spot area are annotated. Subsequently, the spot-aware classification subnetwork is connected to the feature segmentation subnetwork and then trained through early and later fusions to produce the semantic-level spot feature information. The experimental results confirm that the proposed network can increase the discriminative power by modeling the leaf spot attention mechanism. The results prove that the proposed method outperforms conventional state-of-the-art deep learning models.

1. Introduction

Information and communications technology has been applied into the existing farming practices to increase the quantity and quality of plants and crops. Smart agriculture sensors including optical sensors, accelerometer, electrochemical sensors, and airflow sensors have been used to measure a leaf's angle and colors, soil properties, pH, soil nutrient levels, etc. [1]. Continuous monitoring yields a vast amount of sensing data, from which a plant diseases and its growth condition can be evaluated through data analysis, thereby enabling an increase in yield while minimizing resources such as water and fertilizer. Particularly, plant disease diagnosis in a timely manner is important to prevent diseases from spreading at an immature state and prevent economic damages to farmers. A large team of experts and farmers can identify plant diseases based on the symptoms on the leaves; however, this manual observation is time consuming and costly. In addition, it is inefficient to continuously monitor all the plants on a large field area. Therefore, the automatic detection of plant diseases is necessary. With the rapid advance in computer vision enabled by deep learning, image-based plant disease detections have garnered particular attention. The deep convolutional neural network (DCNN) [2,3] introduced recently have demonstrated powerful performance for image classification and detection problems. Therefore, image-based approaches have been studied actively using mobile cameras or digital cameras built on autonomous agricultural vehicles for plant disease identification.

1.1. Related works

Regardless of the plant disease, computer vision technologies can be used directly for image-based plant disease identification. To characterize local image appearances, SIFT [4], LBP [5], sparse codes [6,7], and other handcrafted features [8] including color, entropy, and local homogeneity can be extracted from the preprocessed plant images, and subsequently pooled through bag-of-words (BOW) [9] and fisher vector encoding (FVE) [10,11] to aggregate those features and obtain image-level representations. Next, given the pooled features, a support vector machine (SVM) [12,13], which is a data analysis tools, can be trained to classify the plant diseases. Certainly, other tools such as decision trees [14], genetic algorithm [15], and dictionary learning [16], can be used for classification. Recently, the DCNN has replaced a series of steps that consist of handcrafted feature designs, pooling, and classification because the DCNN can automatically learn generic representations in a hierarchical manner for discriminative feature extraction. With the emergence of the DCNN, a profound knowledge in feature design, feature pooling, and classification are not necessary, thereby rendering it easier for nonexperts to handle the plant disease identification problems. If a new training dataset is provided

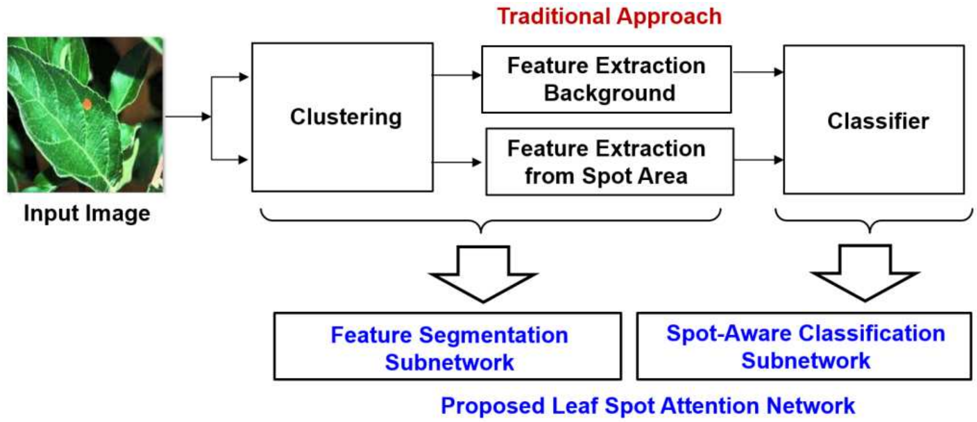

Figure 1. Traditional approach [28] versus proposed leaf spot attention network (LSA-Net).

good performance can be obtained through transfer learning, which uses pretrained models such as AlexNet [17], VGG [2], and ResNet [18], and subsequently updates the model's parameters. A large number of studies [19-23] have been performed based on transfer learning during the past few years for leaf disease identification. However, the difference is minimal between the transfer-learning-based approaches [19-23] because the architectures used are not new. Therefore, a new leaf spot attention network, which is hereafter referred to as LSA-Net, is proposed.

It is noteworthy that our goal is to solve the leaf disease identification problem, which is different from the leaf species identification problem [24]. It is necessary to model the leaf shapes for leaf species identification [24,25]. However, for the leaf disease identification, it is crucial to find the location of the leaf disease to extract discriminative features from background and spot areas separately. Additionally, the test images used for the leaf species identification [24] have solid background colors, whereas in this study, clutter backgrounds including leaves and branches have been considered. Moreover, this paper focuses on the state-of-the-art methods based on deep learning and FVE, and as such, details about traditional approaches are not discussed, for which the reader may refer to related literature [26,27].

1.2. Proposed approach

Fig. 1 shows the main differences between traditional approaches [28, 8, 13] and the proposed LSA-Net for apple leaf disease identification. As shown in Fig. 1, the traditional approach includes three steps, i.e., clustering, feature extraction, and classifier learning. First, the clustering step divides the input image into background and spot areas (i.e., areas with leaf diseases), and two types of features are extracted from the divided areas via a feature extraction step. Next, the classifier is trained using machine learning tools such as SVM and decision trees. However, this traditional approach has a few drawbacks. The first drawback is that the colors and textures used for clustering and feature extraction are handcrafted features, which are not better than deep features in terms of the discriminative power because deep learning can provide abundant and discriminative features via a layer-by-layer nonlinear transformation. The second drawback is that the employed classifiers, such as SVM and decision trees [13, 14], are not better than deep learning in terms of classification performance. To overcome the shortcomings of traditional approaches, the proposed method incorporates the three steps into a single deep learning architecture. Inspired by traditional approaches, the clustering and feature extraction are replaced by the feature segmentation subnetwork (FS-SubNet), and the classifier is replaced by the spot-aware classification subnetwork (SAC-SubNet), as shown in Fig. 1. Therefore, the main difference between traditional approaches and the proposed method is the incorporation of the three steps into a single deep learning architecture, which models a leaf spot attention mechanism for more accurate leaf disease identification.

The primary idea of the proposed LSA-Net is based on the observation that symptoms of leaf diseases can only be detected in the leaf area, whereas the background regions do not contain information about them. To realize this, the FS-SubNet is designed to first extract discriminative features from spot areas, leaf areas, and background, and subsequently the SAC-SubNet is combined through early and later fusions to be trained in an end-to-end manner. Unlike traditional approaches that conduct the three steps separately, the two subnetworks are incorporated into the proposed LSA-Net, and trained in an end-to-end manner. The FS-SubNet is used to transform the input image into a segmented feature map that includes three areas, i.e., the background, leaf area, and spot area. The SAC-SubNet is used to classify leaf diseases based on the segmented feature map and encoded high-level semantic spot information. During training, the FS-SubNet provides useful information for the determination of the features that are more important and those that have a decisive role in classifying leaf diseases. Therefore, the fusion of the FS-SubNet and SAC-SubNet plays a role in the attention mechanism [29, 30], which selects subsets of feature maps and emphasizes spatial regions in the feature maps. Owing to the leaf spot attention mechanism enabled by the early and later fusions of the FS-SubNet and SAC-SubNet, the proposed LSA-Net can increase its classification performance by using more discriminative features and high-level semantic spot information.

Even though deep learning approaches [19-23] have been recently introduced, the existing models, such as VGG and ResNet, have been directly used for leaf diseases identification. This means that recently introduced deep learning approaches have limitations in improving the discriminative power because these approaches do not consider the spatial attention mechanism to extract discriminative features from spot and background areas. This is the major difference between the proposed method and conventional state-of-the-art methods based on deep learning. Due to the major differences mentioned above, the proposed method surpasses traditional approaches and state-of-the-art methods based on deep learning.

1.3. Contributions

- To the best of the author's knowledge, this study is the first attempt to introduce a new deep learning architecture that considers spot attention mechanism for leaf diseases identification. In previous studies, the existing models such as VGG [2], ResNet [18], and SSD [31] have been used without any modifications. However, in the proposed architecture, a new deep learning architecture that includes the FS-SubNet and SAC-SubNet is proposed to indicate the location of the spot areas and effectiveness of features in them for improving classification accuracy. Recently, attention networks [29,30] that select subsets of feature maps and emphasize spatial regions in these maps are a popular research topic. These networks have exhibited better performance than existing models. The proposed architecture includes the FS-SubNet and SAC-SubNet to predict locations of spot areas in feature maps and encode semantic-level spot information. Thus, the proposed method can be regarded as one of the spatial attention networks. Therefore, the major contribution of this study is the suggestion of a new spot attention network for the identification of leaf diseases.
- Studies related to leaf disease identification have not compared their methods with state-of-the-art methods such as multiscale-based deep feature extraction [32,33], SqueezeNet [29], and bilinear models [34]. This means that the literature only provides the classification performances of existing deep learning models. However, in this study, we quantitatively prove that the proposed method is capable of improved classification accuracy, as compared to existing deep learning models. Also, it is shown that the proposed method surpasses state-of-the-art deep learning methods such as VGG, ResNet, SqueezeNet, multiscale-based deep feature extraction and pooling, and bilinear models. This is the first study to test state-of-the-art methods based on deep learning and analyze them.
- This study shows the method of incorporating the traditional approach [28, 8, 13] (including three steps, i.e., clustering, feature extraction, and classifier learning) into a deep learning architecture. In the proposed LSA-Net, the clustering and feature extraction steps are replaced by the FS-SubNet, and the classifier learning is replaced by the SAC-SubNet. The difference between the proposed method and the traditional approach is that the traditional approach performs the three steps separately, whereas the proposed method is trainable in an end-to-end manner. In addition, the clustering method in the traditional approach can be applied to simple leaf images where background colors are nearly solid. However, in our image database, the background colors are similar to the spot colors of a certain disease; thus, the color-based clustering algorithm might fail to extract the spot colors from leaf images. In contrast, the proposed method incorporates the clustering algorithm into the deep learning architecture; thus, a more accurate feature clustering can be obtained. Further, this study considers a real environment to some degree, and the leaf images are more complicated in the background than those tested in [28].

2. Conventional Approaches

This section presents an overview of recently introduced state-of-the-art leaf disease identification [2,19-23,32-34]. This will facilitates readers to understand the differences between the proposed method and the conventional methods.

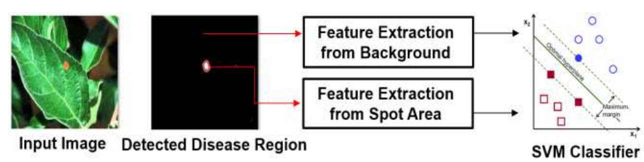

Figure 2. Clustering-based handcrafted feature extraction.

2.1. Clustering-based handcrafted feature extraction

For leaf disease identification, it is important to find the location of leaf diseases. In [28], color-based clustering algorithm is used to define leaf disease colors. More specifically, spot areas with leaf diseases are manually marked and subsequently averaged to define the leaf disease color. Given the test image, leaf disease region is detected by comparing the distance between the predefined leaf disease color and all the pixel colors in the test image. If the distance is smaller than the threshold value, the pixel belongs to the leaf disease region. Subsequently, handcrafted features, such as RGB histogram and LBP [5], are extracted from the background and leaf disease region separately, which are subsequently fed into the SVM classifier. Fig. 2 shows the procedure of the clustering-based feature extraction for leaf disease identification [28]. This approach can work efficiently if the leaf disease colors are significantly different from the leaf colors and background. However, in our image database, leaf colors are similar to spot colors for a certain disease type. This implies that the clustering-based feature extraction might fail to detect the leaf disease region, and this might lead to reduced recognition accuracy.

2.2. Deep feature extraction and pooling methods

Fig. 3 shows the procedure for the multiscale-based deep feature extraction and pooling method (MDFEP) using a pretrained VGG network in [32,33]. Initially, multiscale images are generated to be invariant to scales, and subsequently fed into the pretrained VGG network. Feature maps are extracted from the truncated VGG network at a convolutional layer including nonlinearity (e.g., relu4_3 denoted by [2]). In other words, feature vectors of 512 dimensions can be extracted by gathering all values at the same locations in the feature maps. For example, red circles in Fig. 3 constitute one feature vector. Unlike handcrafted features such as SIFT [4] and LBP [5], the features are learned in a hierarchical manner; thus, these features are referred to as deep features. It has already been verified in [35] that deep features can be generic representations to be applied to different tasks. Next, for all training images in different classes, feature vectors are extracted similarly as mentioned above; subsequently, FVE [10,11] is conducted to pool the feature vectors and characterize the global image appearance based on the Gaussian mixture model (GMM). FVE obtains the gradient vector by taking the derivative of the log likelihood with respect to the GMM's parameters. Instead of FVE, other pooling methods, e.g., BOW [9] and its variant [6] can be used. The FVE feature vectors are fed into the SVM for leaf disease identification. Two choices are available for local feature types and pooling methods, respectively. In other words, either deep features or handcrafted features are used for local feature extraction. Similarly, one of the FVE and BOW can be used for feature pooling. By altering the local feature type in Fig. 3, the discriminative power between the deep features and handcrafted features can be evaluated.

2.3. Transfer-learning-based method

If only the pretrained VGG network, a part of Fig. 3, is applied and subsequently trained with a new training set, the VGG network is finetuned to be adopted to a new task. This type of approach is known as transfer learning (TL) [2]. In

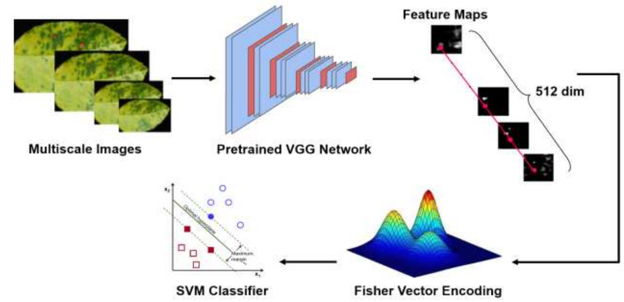

Figure 3. Multiscale-based deep feature extraction and pooling.

this study, a new training set contains leaf disease images and the corresponding label information. Given the new training set, the VGG network can be trained using mini-batch stochastic gradient descent optimization [36]. Instead of the VGG network, ResNet [18] and AlexNet [17] can be used. Compared to the TL method, in the MDFEP method [32,33], as shown in Fig. 3, a series of steps that consist of the local feature extraction, pooling, and classification should be performed. This approach appears more complicated. However, in the TL method, input images should be scaled to a fixed size, thus implying that the TL method can overlook texture information at various scales, i.e., leaf textures and spot areas, which are important for leaf disease identification. Meanwhile, multiscale images are allowed to be used with the MDFEP method. It is worth comparing the performance of the TL method with that of the MDFEP method to demonstrate the effectiveness of using multiscale images.

3. Proposed Leaf Spot Attention Network

Symptoms can be detected only in the leaf area whereas the background region contains no information regarding leaf diseases. Therefore, a leaf spot attention model is needed to predict which local areas in the feature maps are more important and should have a decisive role in classifying leaf diseases. Fig. 4 shows the proposed LSA-Net to identify apple leaf diseases. The proposed architecture consists of two subnetworks: FS-SubNet and SAC-SubNet. The FS-SubNet is to transform the input image into the segmented feature maps that include the background, leaf area, and spot area. In the FS-SubNet, the crop layer applies two-dimensional cropping to the input feature maps. Two input feature maps are required. One is to be cropped and the other is the reference to determine the size of the cropped feature map. The transposed convolution layer applies the transpose of convolution to the input feature maps for upsampling, and $\oplus$ indicates the addition layer that adds the input feature maps by element. The concatenation layers are used to connect the FS-SubNet with the SAC-SubNet. As shown in Fig. 4, there are two concatenation layers. The first one, shown on the left side of Fig. 4, stacks the segmented feature map on top of the input

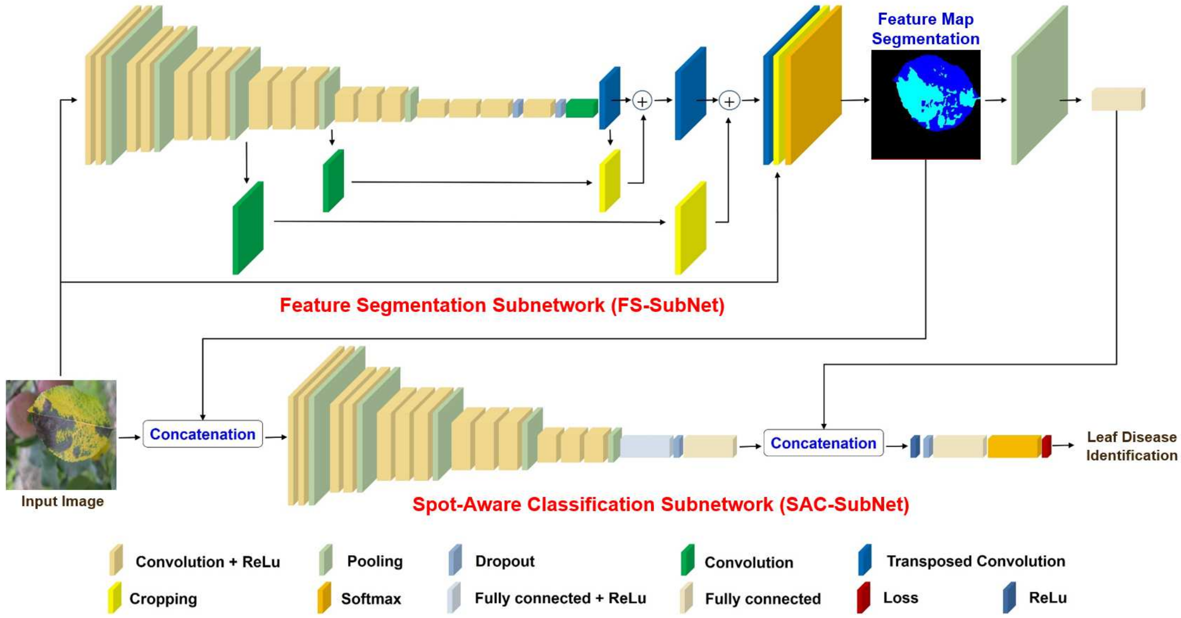

Figure 4. Proposed leaf spot attention network for apple leaf disease identification.

image, thereby forming a multi-dimensional input tensor for early fusion. Through the second one, shown on the right side of Fig. 4, the encoded semantic-level spot information is sent to the SAC-SubNet for later fusion. Other layers such as pooling, rectified linear unit (ReLU), convolution, fully connected, and softmax are described in [2].

The architecture of the FS-SubNet in Fig. 4 is inspired by the semantic segmentation in [37]. However, the goal of this study is different from that in [37]; in other words, our goal is not to divide the input image into multiple regions, but to achieve apple leaf disease identification. In the proposed architecture, the FS-SubNet is pretrained with a new training set that contains ground truth segmentation maps; subsequently, this subnetwork is combined with the SAC-SubNet through the concatenation layers to complete a whole network to be trained in an end-to-end manner. Therefore, the purpose of using the FS-SubNet is to encode the semantic-level spot information from the predicted segmentation feature map and subsequently teach the SAC-SubNet regarding which areas in the segmented feature map should have a decisive role in classifying leaf diseases. The FS-SubNet serves as a guide to achieve a more accurate leaf disease identification.

Naturally, the proposed architecture is different from those of TL-based methods [19-23] because two types of subnetworks are connected to create a whole network that is subsequently trained in an end-to-end manner. In other words, conventional TL-based methods do not include the leaf spot attention model. In Fig. 4, the architecture of the SAC-SubNet is the same as that of the VGG network except the concatenation layers. If the FS-SubNet is excluded from Fig. 4, the proposed architecture becomes identical to the conventional VGG model. Therefore, whether recognition accuracy can be increased must be verified by comparing the performance between the proposed LSA-Net and the conventional VGG network [2]. Certainly, other pretrained networks, such as ResNet and SqueezeNet [29, 30], can also be considered for the SAC-SubNet.

4. Experiments

In this study, the proposed LSA-Net was implemented using Matlab and trained with four Titan-XP GPUs on a windows operating system. To compare the proposed method, state-of-the-art methods, i.e., FVE with SIFT [10,11], clustering-based feature extraction [28], TL methods using VGG [19,21], ResNet [20], and SqueezeNet [29], DCNN-based bilinear model [34], and MDFEP [32,33] were tested. Correct recognition accuracy, defined as the ratio of the correctly classified images to the total images, was used for performance evaluation. For training the FS-SubNet as well as the entire network, the ratio of the training set to the validation set was set to 0.5, and the test set was identical to the validation set. The VGG network is a pre-trained model for a largescale image set called ImageNet [2]. Thus, the VGG model was used without transfer learning. In addition, data augmentation was not performed during the training. The number of Gaussians for FVE is 256 and the VGG-16 model in [2] was used as the VGG network in all the methods: TL methods, bilinear model, MDFEP, and the proposed method.

4.1. Image collection

All apple leaf images used in this study were provided by the Apple Research Institute in our country. The apple leaf images were categorized into three groups, according to two types of leaf diseases and normal leaf. This implies that a test image was categorized into one of three groups to determine apple leaf diseases. Fig. 5 shows the example of the apple leaf images. The first row shows the normal leaf images, and the second and third rows show the diseased leaf images. Particularly, for the diseased leaf images with marssonia blotch, as shown in the second row, the blotch colors are similar to the normal leaf colors in the background. Therefore, the color-based clustering algorithm [28] might fail to extract the blotch colors from the normal leaves. This reveals that the leaf areas, background, and spot area must be divided. In addition, a real environment was considered to some degree, in that the leaf images were more complicated in the background than those tested in [28], where the background colors were nearly solid. In our database, the total numbers of normal leaf images, diseased leaf images with marssonia blotch, and diseased leaf images with alternaria leaf spot were 118, 120, and 166, respectively.

4.2. Network training

Before training the entire network in an end-to-end manner, as shown in Fig. 4, the FS-SubNet was first pre-trained. To train the FS-SubNet, ground truth segmentation maps are required. In this study, ground truth segmentation maps were generated manually through image editing to divide them into three areas: background, leaf area, and spot area. During image editing, the leaves without diseases were classified by background in the image and those with diseases were classified by leaf area in the image. This simplifies the labeling process. Given the ground truth segmentation maps, the FS-SubNet was trained using mini-batch gradient descent optimization [36].

Next, to connect the pretrained FS-SubNet with the SAC-SubNet, the last loss layer of the FS-SubNet was removed, and two concatenation layers were added. The first one, as shown on the left side of Fig. 4, stacks the predicted segmentation map on top of the input image and the second one, shown on the right side of Fig. 4, transforms the predicted segmentation map into high-level semantic spot information. Subsequently, the entire network was trained in an end-to-end manner. Therefore, the FS-SubNet changed the pretrained parameters to be adopted for a new task. In other words, the predicted segmentation map was adjusted during the entire network training for a more accurate leaf disease identification.

In Fig. 6 the first row shows the leaf images, and the second and third rows show the ground truth segmentation images and corresponding predicted segmentation feature maps, respectively. Here, the predicted segmentation feature maps are extracted from the entire trained network.

Figure 5. Example of apple leaf images: normal leaf images (first row), diseased leaf images with marssonia blotch (second row), and diseased leaf images with alternaria leaf spot (third row).

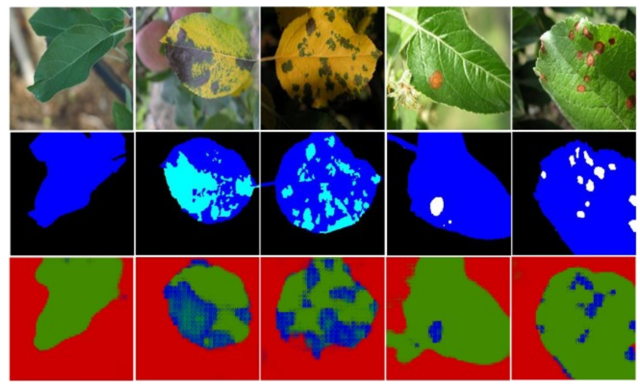

Figure 6. Leaf images (first row), ground truth segmentation images (second row), predicted segmentation feature maps after training entire network (last row).

In Fig. 6, the feature maps are visualized using different colors. During retraining of the FS-SubNet, the output of the FS-SubNet inevitably changes to help increase the accuracy of the SAC-SubNet. Nevertheless, the predicted segmentation feature maps can still possess discriminative features for different areas, i.e., background, leaf areas, and spot areas. As shown in the last row of Fig. 6, the predicted segmentation feature maps divide the input images into background and leaf areas. In particular, it is observed that the spot areas in the leaves have blue colors, which are clearly distinguishable from the background's red colors and leaf's green colors. This indicates that the FS-SubNet can provide discriminative features from different areas, i.e., background, leaf areas, and spot areas, thereby enabling the improvement of the SAC-SubNet.

It is noteworthy that the ultimate objective of this study is not accurate segmentation but leaf disease identification. This implies that the FS-SubNet is sufficient to serve as a guide to achieve a more accurate leaf disease identification. Even though there are a few segmentation errors, as shown in the last row in Fig. 6, the FS-SubNet can provide the SAC-SubNet with more discriminative features and high-level semantic spot information through the early and later fusions.

The mean accuracy of the FS-SubNet, defined as the ratio of correctly classified pixels to total pixels for each class, was ~86% and mean IoU (Intersection over Union), also known as the Jaccard similarity coefficient, is ~69% [38].

As shown in Fig. 4, if the FS-SubNet and concatenation layers were excluded from the entire network, the proposed architecture would be identical to the conventional VGG network. Thus, whether the correct recognition accuracy can be improved by comparing the performance between the proposed LSA-Net and the pretrained VGG network must be verified.

4.3. Performance comparison

Table 1 presents the correct recognition accuracy results for the proposed method and conventional state-of-the-art methods. By comparing the proposed method and the TL method using the VGG network, it is verified that the use of the FS-SubNet increases the recognition accuracy by 14.7%. Hence, it is concluded that the FS-SubNet can teach the SAC-SubNet regarding which features in the segmentation map should have a decisive role in classifying apple leaf diseases. In other words, the FS-SubNet had served as a guide for a more accurate leaf disease identification. It is also shown that the proposed LSA-Net demonstrates the best performance among all the methods. As shown in Table 1, the performance of the MDFEP method is better than those of the TL methods. This reveals that it is important to use multiscale images for leaf disease identification. In this study, nine scales were used. The TL methods force the size of the input images to be fixed; thus, the leaf texture and spot areas can be removed, thereby resulting in a lower recognition accuracy. The bilinear model [34] demonstrates a better performance than the TL methods using VGG, ResNet, SqueezeNet networks. This indicates that a pairwise feature interaction model is effective in increasing the discriminative power, thereby improving the correct recognition accuracy. As expected, the performance of the FVE with the handcrafted SIFT features is worse than that of using deep features. The clustering-based feature extraction method shows the worst performance as the color-based leaf disease detection failed to extract spot colors from the leaf images, owing to the similarity between the both. Additionally, leaf disease detection is sensitive to the threshold value used for clustering. In other words, if the threshold value is high, the background and leaf colors also get included in the detected region. If the threshold value is low, the detected region is too small to extract sufficient features. SqueezeNet can obtain better accuracy than VGG and ResNet because a simple attention mechanism is incorporated using 1x1 convolutional filters, which are regarded as weighting values for feature maps.

Table 1. Performance evaluation.

Methods	Accuracy
Clustering-based feature extraction [28]	43.9%
TL method using VGG [19,21]	74.7%
TL method using ResNet [20]	76.6%
SqueezeNet [29]	79.3%
MDFEP method [32,33]	81.7%
FVE with SIFT [10,11]	71.7%
Bilinear model [34]	80.6%
Proposed LSA-Net	**89.4%**

In summary, the traditional approach has a few drawbacks. The first is that the colors and textures used for clustering and feature extraction are handcrafted features. Handcrafted features are known to be inferior to deep features because deep learning can automatically generate abundant and discriminative features via a layer-by-layer nonlinear transformation. The second is that the classifiers employed, such as SVM and decision trees, are not better than deep learning in terms of classification performance. Although existing deep learning models, such as VGG and ResNet, can be used for the identification of leaf diseases, these models have limitations in improving the discriminative power because they do not consider the spatial attention mechanism to extract discriminative features from spot and background areas. Even though the SqueezeNet includes a simple attention mechanism, its performance is not better than the proposed LSA-Net because the former does not involve the use of ground truth segmentation maps and thus it has limitation in encoding high-level semantic spot information. Therefore, the proposed LSA-Net surpasses traditional approaches and state-of-the-art methods based on deep learning.

5. Conclusion

A new leaf spot attention network was introduced in this study to identify apple leaf diseases. This study was motivated by the fact that leaf diseases exist only in the leaf area and the background contains no information regarding leaf diseases. In the proposed architecture, a new FS-SubNet was designed to divide input images into the leaf area, background, and spot area in a feature map. Subsequently, it was combined with another SAC-SubNet through early and later fusions to be trained in an end-to-end manner. During training, the FS-SubNet could teach the SAC-SubNet regarding which areas in the segmented feature map were more important and which features should have a decisive role in classifying leaf diseases. This FS-SubNet served as a guide for a more accurate leaf disease identification. The experimental results verified that the correct recognition accuracy could be improved by extracting more discriminative features and semantic-level spot information through the early and later fusions of the FS-SubNet and SAC-SubNet. It was also shown that the proposed LSA-Net yielded a better performance than state-of-the art methods: TL methods, MDFEP method, FVE with SIFT, and bilinear model.

Acknowledgment

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2020R1A2C1010405).

References

[1] https://www.mouser.com/applications/smart-agriculture-sensors/

[2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv: 1409.1556 [cs. CV]*, Sept. 2014.

[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 141-157, Jan. 2016.

[4] D. G. Lowe, "Distinct image features from scale-invariant key points," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, Nov. 2004.

[5] Z. Guo, L. Zhang, D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657-1663, Jun. 2010.

[6] J. Yang, K. Yu, Y. Gong, T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," In *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, Miami, FL, pp. 1794-1801, June 2009.

[7] S. Zhang, X. Wu and L. Zhang, "Leaf image based cumber disease recognition using sparse representation," *Computers and Electronics in Agriculture*, vol. 134, pp. 135-141, March 2017.

[8] V. Singh and A. K. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques," *Information Processing in Agriculture*, vol. 4, no. 1, pp. 41-19, March 2017.

[9] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," In *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, New York, NY, pp. 2169-2178, June 2006.

[10] V. Chandrasekhar, J. Lin, O. Morère, H. Goh, and A. Veillard, "A practical guide to CNNs and Fisher vectors for image instance retrieval," *Signal Processing*, vol. 128, pp. 426-436, Nov. 2016.

[11] F. Perronnin and C. Dance, "Fisher kernels on visual vocabularies for image categorization," In *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, June 2007, pp. 1-8.

[12] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.

[13] M. Islam, A. Dinh and K. Wahid, "Detection of potato diseases using image segmentation and multiclass support vector machine", In *Proc. IEEE 30th Canadian Conference on Electrical and Computer Engineering*, pp. 1-4, 2017.

[14] J. R. Quinlanm, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81-106, March 1986.

[15] Z. Chuanlei, Z. Shanwen, Y. Jucheng, S. Yancui, and C. Jia, "Apple leaf disease identification using genetic algorithm and correlation based feature selection method," *International Journal of Agriculture and Biological Engineering*, vol. 1, no. 2, pp. 74-83, March 2017.

[16] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[17] A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet classification with deep convolutional neural networks," In *Proc. Neural Information Processing Systems*, Harrah's Lake Tahoe, Nevada, pp. 1-9, Dec. 2012.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, pp. 770-778, June 2016,.

[19] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311-318, Feb. 2018.

[20] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park, "A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition," *Sensors*, vol. 17, no. 9, pp. 1-21, Sept. 2017.

[21] S. P. Mohanty, D. Hughes, and M. Salathe, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vo1. 7, article 1419, Sept. 2016.

[22] Y. Lu, S. Yi, N. Zeng, Y. Liu and Y. Zhang, "Identification of rice diseases using deep convolutional neural networks", *Neurocomputing*, vol. 267, pp. 378-384, Dec. 2017.

[23] E. Tetila et al., "Automatic recognition of soybean leaf diseases using UAV images and deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, DOI: 10.1109/LGRS.2019.2932385, 2019.

[24] N. Kumar, P. N. Belhumeur, A. Biswa, D. W. Jacobs, W. John Kress, I. Lopez, and J. Soares, "Leafsnap: A computer vision system for automatic plant species identification," In *Proc. European Conference on Computer Vision*, Florence Italy, pp. 502-516 Oct. 2012.

[25] C. Zhao, S. Chan, W.-K. Cham, and L. M. Chu, "Plant identification using leaf shapes- A pattern counting approach," *Pattern Recognition*, vol. 48, no. 10, pp. 3203-3215, Oct. 2015.

[26] R. Pydipati, T. F. Burks, W. S. Lee, "Identification of citrus disease using color texture features and discriminant analysis," *Computers and Electronics in Agriculture*, vol. 52, no. 1-2, pp. 49-59, June 2006.

[27] M. Bhange and H. A. Hingoliwala, "Smart farming: pomegranate disease detection using image processing", *Procedia Computer Science*, vol. 58, pp. 280-288, 2015.

[28] H. Ali, M. I. Lali, M. Z. Nawaz, M. Sharif, and B. A. Saleem, "Symptom based automated detection of citrus diseases using color histogram and textural descriptors," *Computers and Electronics in Agriculture*, vol. 138, pp. 92-104, June 2017.

[29] F. N. Iandola et al., "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5MB model size," *arXiv:1602.07360v4*, Nov. 2016.

[30] S. Zagoruyko and N. Komodakis, "Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer," *arXiv:1612.03928v3*, Feb. 2017.

[31] E. Jiang, Y. Chen, B. Liu, D. He and C. Liang, "Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks," *IEEE Access*, vol. 7, pp. 59069 – 59080, May 2019.

[32] M. Cimpoi, S. Maji, A. Vedaldi, "Deep filter banks for texture recognition and segmentation," In *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, Boston, MA, June 2015, pp. 3828 – 3836.

[33] C.-.H Son, "Recognizing species and conditions of street trees based on multiscale and deep Feature extraction", *Journal of the Institute of Electronics and Information Engineers*, vol. 55, no. 4, pp. 70-78, April 2018.

[34] T.-Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear CNN models for fine-grained visual recognition," In *Proc. IEEE International Conference on Computer Vision*, Santiago, Chile, pp. 1449-1457, Dec. 2015.

[35] J. Donahue et al., "DeCAF: A deep convolutional activation feature for generic visual recognition," *arXiv:1310.1531 [cs.CV]*, Oct. 2013.

[36] S. Ruder, "An overview of gradient descent optimization algorithms" *arXiv:1609.04747 [cs.LG]*, Sept. 2016.

[37] J. Long, E. Shelhamer, T. Darrell, "Fully convolutional networks for semantic segmentation," In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Boston, pp. 3431-3440, June 2015.

[38] https://kr.mathworks.com/help/vision/ref/evaluatesemanticsegmentation.html