GAN을 이용한 식물 병해 이미지 합성 데이터 증강
Synthetic Data Augmentation for Plant Disease Image Generation using GAN

나즈키 하십, 이재환, 윤숙*, 박동선**
전북대학교 전자공학과,
목포대학교 실감미디어기술연구소*,
전북대학교 IT융합연구센터**

Haseeb Nazki, Jaehwan Lee, Sook Yoon*,
Dong Sun Park**
Department of Electronics Engineering, Chonbuk National University,
Research Institute of Realistic Media and Technology, Department of Computer Engineering, Mokpo National University*,
IT Convergence Research Centre, Chonbuk National University**

ABSTRACT

In this paper, we present a data augmentation method that generates synthetic plant disease images using Generative Adversarial Networks (GANs). We propose a training scheme that first uses classical data augmentation techniques to enlarge the training set and then further enlarges the data size and its diversity by applying GAN techniques for synthetic data augmentation. Our method is demonstrated on a limited dataset of 2789 images of tomato plant diseases (Gray mold, Canker, Leaf mold, Plague, Leaf miner, Whitefly etc.).

I. Introduction

Diseases in plants have been extensively studied in the scientific area. Collecting plant disease related data is a complex procedure that requires the collaboration of people from different fields at contrasting stages. Researchers often come across the challenge of class imbalance which has been a general problem in machine learning as well as in computer vision. An effective way of synthesizing images to supplement training set may help boost accuracy in various fields. However, the diversity that can be gained from small modifications of the images is relatively small. This motivates the use of synthetic data examples; to introduce more variability and can possibly enrich the dataset.

A promising approach for training a model that synthesizes images is known as Generative Adversarial Networks (GANs) [1]. Recently, several imaging applications have applied the GAN framework including medical, however the application in the field of plant disease imaging has remained unexplored. In the current study we investigate the applicability of GAN framework to synthesize high quality plant disease images for data augmentation.

II. Generating Synthetic Plant Disease images

To preserve the plant characteristics, we used transformations that cause shape deformation (like shearing and elastic deformations). Each image was first rotated Xrot times at random angles with θ = [-10?, ..., +10?]. Afterwards, each rotated image was flipped Xflip times (up-down, left-right), translated Xtrans times where we sampled random pairs of [x, y] pixel values. As a result of the augmentation process, the total number of augmentations was X = Xrot × (1 + Xflip + Xtrans).

We employed the Deep Convolutional PGGAN [2] for improved quality, stability and variation to synthesize various labeled classes for each class separately. We follow their structure except for including the Network-in-Network layer and the low-resolution blocks, growing the output spatial resolution to 256 x 256 as a replacement for 1024 x 1024 output resolution of the generated images. The training starts with both the generator G and discriminator D having a low spatial resolution of 4×4 pixels. As the training advances, successive layers are incrementally added to G and D, thus increasing the spatial resolution of the generated images. This allows stable synthesis in high resolutions and speeds up training considerably.

Ⅲ. Experiments and Results

Our dataset contains images with several diseases and pests in tomato plants but our experiments were mostly carried on the PlantVillage dataset indistinguishable from our dataset in terms of class imbalance and lack of adequate training samples. For the implementation of our GAN architecture we used the TensorFlow framework. All training processes were performed using a setup with an NVIDIA Titan-X 12GB GPU and intel i7 processor.

We started with examining our results with the images of northern leaf blight in corn plants. The available training images are first augmented using techniques described in section 2. (Fig 1)

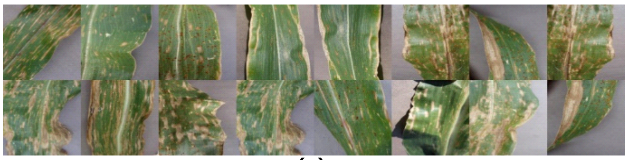
We further generated synthetic data using our diseased tomato plants dataset using the class involving the plants effected due to low temperature. We took the crops of the diseased portions in the plants. The results are displayed in Figure 2.

Ⅳ. Conclusion

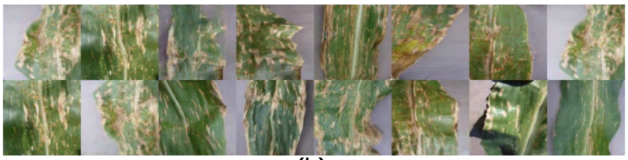
To conclude, in this work we demonstrated a method that uses the generation of synthetic plant disease images for data augmentation to improve performance on various ventures related to agricultural problems with limited data. In the future, we plan to extend our work to additional domains that can benefit from synthesis of this data for improved training and accuracy.

Acknowledgement

This work was supported by the Brain Korea 21 PLUS Project, National Research Foundation of Korea. We gratefully acknowledge the support of "Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ0120642016)" Rural Development Administration, Republic of Korea.



(a)



(b)

▶▶ Fig.1. Results for our synthetic augmentation on northern leaf blight in corn plants. (a) Real images (b) and the generated images.

(a)

(b)

▶▶ Fig.2. Results for our synthetic augmentation on the crops of the images for effect of low temperature on tomato plants. (a) Real images (b) and the generated images.

■ References ■

[1] Goodfellow, I., et al., “Generative adversarial nets,” in Advances in neural information processing systems, 2014.

[2] Karras, T., Aila, T., Laine, S., Lehtinen, J.,“Progressive Growing of GANs for Improved Quality, Stability, and Variation,” CoRR, 2017.