University of Salento

Faculty of Engineering

Degree Course in Computer Engineering

Master Thesis

in

Image Processing

Titolo tesi

Supervisor:
Ch.mo Prof. Nome COGNOME
Co-Supervisor:
Ch.mo Prof. Nome COGNOME

Candidate:
Davide Basile
Matricola 20034689

Academic Year 2020/2021

Lorem ipsum dolor sit amet, consectetuer adipiscing elit.
Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis.
Curabitur dictum gravida mauris.

— *Donald Ervin Knuthyu*

Acknowledgments

A conclusione di questo lavoro di tesi ringrazio...

G. M.

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.[1]

1.1 Lorem

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque

cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

1.1.1 Lipsum

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Chapter 2

Related Works

2.1 Plant Phenotyping

Plant phenotyping has been recognized as a bottleneck for improving the efficiency of breeding programs. In recent years, image-based approaches have shown great potential for high-throughput plant phenotyping, resulting in an increased focus on image-based plant phenotyping. To meet the world's growing demand, current crop productivity must be roughly doubled by 2050, due to global population growth that may exceed 9 billion by 2050 and 11 billion by 2100. Over the past years, various HTP solutions have been developed to improve phenotyping capability and throughput, including tower systems, gantry systems, mobile ground systems, and low and high-altitude aerial systems, one obvious trend has been noted in recent HTP systems, computer vision. Cameras have been used more frequently because of their extensive ability to extract complex features. 2D images can provide spatial information of a scene plus an additional dimension of data such as spectral information. 3D images can provide a 3D structure of a scene that can be used to calculate morphological features of objects. Finally, 2.5D images retain information about the image plane structure, which is similar to that of 2D images, and capture depth information about a scene, which can be used to reconstruct the 3D structure of that scene. one of the first operations to be performed for phenotyping is to identify and define the phenotypic traits to be measured, which largely determine the use of appropriate imaging modalities for plant detection. Measuring phenotypic traits usually requires one or more computer vision tasks that can be solved by developing new algorithms or improving existing ones through conventional image/signal processing, machine learning, or a combination of them. Conventional machine-learning based approaches have generally improved generalizability, but most still cannot meet the requirements for current phenotypic purposes. Deep learning is a subset of machine learning and enables hierarchical learning of data. The key benefit is that features will be learned automatically from the input data, thus

breaking down barriers to developing intelligent solutions for different of applications. A commonly used DL architecture is deep convolutional neural networks(CNNs), that were developed for image classification and eventually demonstrated better performance than humans on the same dataset.

2.2 Convolutional Neural Networks

Most imaging-based phenotyping applications essentially require solutions for one or more tasks related to image classification, object detection, and segmentation.

2.2.1 Image Classification

Image classification is one of the main tasks of computer vision and aims to assign images with predefined class labels. The development of modern CNNs for image classification can be grouped into three phases:

- emergence of modern CNNs;
- intensive development and improvement of the CNN architecture;
- reinforcement learning for the design of the CNN architecture.

After these improvements, representative CNNs can now usually outperform humans in image classification for various datasets. It should be noted that the performance improvement following the change in CNN architecture was highly dependent on human expertise and tuning efforts, which means that improving the CNN architecture could be as laborious as feature engineering in traditional ML. A reinforcement learning framework was introduced to search for optimal convolutional cells on a small annotated dataset, and the resulting cells were stacked in different ways and transferred to an unknown large dataset. In addition to performance improvement, studies have been conducted to understand the mechanism of CNNs. This leads to the development of techniques towards explainable artificial intelligence that helps in developing interpretable and inclusive machine learning models. The study also showed that learned features could be generalized to various classifiers, suggesting that CNNs could learn general representations of images rather than specific features for classification. Subsequent studies have continued in this direction and developed various gradient-based methods that can visualize the importance of features for classification results. Commonly used methods include guided backpropagation, gradient-weighted class activation mapping, and layered relevance propagation.

2.2.2 Object Detection

Object detection attempts to detect and classify all potential objects in a given image. The use of CNNs for object detection can be categorized into two groups: one-stage and two-stage CNN architecture. Two-stage models first detect candidate object regions and then classify the candidate regions into different object categories. The OverFeat framework was developed to use a single CNN to extract features for training classifiers and regressors separately. The trained classifiers and regressors were used to predict class labels and bounding box coordinates, respectively, for candidate region of interest generated using a sliding window method.

Three key techniques in the CNN architecture of the RCNN family were identified, including the region proposal network, the region-of-interest pooling operation, and the multitask loss function. A RPN was developed to generate ROIs of candidate objects using features extracted from CNNs, which simultaneously saved processing time and increased the accuracy of the region proposal. A ROI pooling operation was developed to extract a fixed number of features from ROIs of various sizes, thus avoiding repeated computation of features for different ROIs. A multitask loss function was used to unify the training process, which allowed for a training process for object detection.

Representative one-stage models include the you-onlylook-once family and the single-shot detector framework. A critical problem, however, has been discovered for these one-stage models: an extreme imbalance in the number of object and background regions. Most image regions contain only background information, providing little contribution to the model training process. However, if detection accuracy is the most important factor to consider, two-stage models would be the option; otherwise, single-stage models provide better computational efficiency for embedded systems and real-time applications.

2.2.3 Image Segmentation

Semantic segmentation seeks to provide masks for objects with the same semantic meaning, while instance segmentation seeks to provide individual objects in a given image. These in general can be divided into two categories: those encoder-decoder based, and those based on detection. The encoder-decoder-based is divided into two phases. The encoding phase uses CNNs to extract semantically meaningful feature maps from the input images, and the decoding phase uses transposed convolution for upsampling the extracted feature maps to labels. A detection-based framework is based on the CNN architecture for object detection. Several studies have explored the use of object detection models for instance segmentation, including concurrent detection and segmentation based on RCNN and DeepMask based on Faster RCNN.

2.3 CNN Applied to Plant Phenotyping

2.3.1 Plant Development

Morphological changes in plant shoots are critical in describing plant development. Canopy cover and leaf area are two commonly used parameters to quantify plant growth and development. In order to segment plants many studies used color-based features, but usually had imperfect segmentation because plant color could have large variations due to illumination, shade, occlusion, and so forth. Most studies conducted on CNN in plant phenotyping consider it a semantic segmentation task and use an encoder-decoder-based CNN architecture for processing. Although these studies have shown improved segmentation accuracy, annotation of training data can be extremely difficult. To solve this problem, one study attempted to generate synthetic images with semantic annotations automatically for training the CNN model. The combination of synthetic and real images would improve the generalizability of CNNs for plant segmentation and thus the accuracy of growth analysis. The researchers also combined CNNs with other DL methods for characterizing plant development. CNNs were used as a feature extractor to encode the spatial state of plants at individual growth stages, and RNNs were used to incorporate all spatial encodings to learn the temporal changes of plants. In this way, plant growth patterns could be fully encoded by neural networks to reveal differences between cultivars and treatment groups. This indirect phenotyping scheme could be particularly useful for selection-oriented programs, but explaining selection would be a significant challenge and barrier to many research studies that aim to understand the mechanism of many plant responses. In addition to morphological measurements, CNNs could be used to monitor some plant development events such as plant lodging. A new CNN architecture (LodgeNet) was developed by integrating a custom 7-layer CNN model with handcrafted features. Compared with 10 well-established CNN architectures, LodgeNet provided comparable or better performance on differentiation between allurement and regular plots, but with a significant improvement in processing speed.

2.3.2 Plant Stress Phenotyping

Plant stress phenotyping aims to identify and evaluate plant responses to abiotic and biotic stresses, providing information for selection of accession lines with high stress resistance and tolerance in breeding programs and understanding of intrinsic mechanisms in genetics/genomics studies. Plant stress phenotyping can be categorized into four stages:

- identification (presence of stress);
- classification (type of stress);
- quantification (severity of stress);

- prediction (possibility of stress occurrence).

The development of image classification-based approaches can be divided into two phases. In the first phase, studies have intensively investigated known and customized CNN architectures because of the availability of annotated datasets and the simplicity of implementing and training CNN for image classification. Data annotation for image classification is also relatively easy, so a large number of images in a newly collected dataset can be annotated in reasonable time and cost, especially when a proper data collection procedure is used. As a result, studies related to plant stress detection typically have a sufficient number of images annotated for model training. In the second phase, pioneering studies sought to understand the reasons leading to the high performance of CNNs for stress identification and classification, because understanding would not only help to improve CNNs but also ensure the biological correctness of the results obtained. Although some studies have adopted deconvolution layers to visualize activated pixels in different convolution layers, the visualization results have not been used to compare with human evaluation or correlate with biological knowledge. Compared with the first phase studies, the two pioneering studies demonstrated the importance of understanding the mechanism of CNNs for stress phenotyping, as well as the potential for quantifying stress severity. Image annotation is still recognized as a limiting factor for the use of many DL algorithms, so the researchers investigated the use of generative adversarial networks to generate synthetic images for training CNN models for plant stress detection and classification. A very recent study explored the use of a custom CNN architecture to detect plant diseases in hyperspectral images. The novelty of the custom architecture is the use of a 3D convolution operation that can directly convolve spatial and spectral information into hypercubes. This would not only inspire future studies related to plant stresses, but also allow the reanalysis of many previous hyperspectral images collected for plant stress analysis. With improved detection accuracy, subtle differences in stress between cultivars/treatments can be revealed to improve our understanding of plant responses to stresses.

2.3.3 Plant Counting

Counting plants and plant organs is critical to characterizing plant development. Regression or image classification is the simplest and most straightforward way for fruit/organ counting from the perspective of technical development. For regression-based methods, a major modification was made that replaced the softmax layer of a CNN with a single neuron for regression of numerical values. This simple end-to-end counting solution provided high accuracy for counting fruit and leaves from plants. A particular challenge of regression-based solutions is the limited availability of annotated images, which leads to many potential problems such as poor generalizability of the model. To address this issue, one study attempted to generate synthetic data of tomatoes to increase data availability and diversity. Although the trained

CNNs achieved 91% counting accuracy on real images, the study tested only red tomatoes, which have distinctive color characteristics from the background. The generalizability of this approach should be further validated for challenging situations such as detecting green tomatoes from leaves. GANs were also used to generate synthetic data for model training.

An alternative approach was to use patch-based training. TasselNet was developed to count corn tassels in two phases. In the first phase, a local CNN regression model was established to predict the number of tassels in each patch of an image. In the second step, the estimated count in each image patch was averaged over the individual pixels in that patch to create a count map with the same spatial dimension as the original image. The sum of all pixel intensities in the count map represents the final count of tassels in that image. Experimental results showed that TasselNet achieved counting accuracies from 74.8% to 97.1%, which were 2 to 5 times higher than those of conventional methods. TasselNet uses the patch-based training method, which substantially increases the number of training images. For classification-based methods, plant/organ counts were treated as a discrete counting problem and, therefore, a predefined score or grade was assigned to a given image rather than an exact count. An example of a classification-based method is WheatNet, which was developed to predict the percentage of flowering in wheat images. Multiple images were acquired for each plot. A total of 11 classes were annotated for each plot, corresponding to 11 visual scores with a percentage header from 0 to 100% with a 10% interval. The average prediction of all images in a plot was the final percent header for that plot, which reduced counting errors due to inaccurate classification.

Object detection is an intuitive approach to counting plants and plant organs in still images: accurate object detection ensures accurate object counts. DeepFruits was the first study to explore the use of modern CNN architecture for fruit detection. Several key contributions were recognized in this study. First, transfer learning was used to train a Faster RCNN model with 100 labeled images, demonstrating the potential of using limited labeled images to train the CNN architecture. Second, when using RGB images, the trained Faster RCNN model provided a 1% improvement in F1 score over that of the CRF model. Third, data fusion was conducted at the level of the raw data and at the decision level for the Faster RCNN models. A custom two-step framework was proposed using the superpixels generated by the simple iterative linear clustering algorithm as proposed regions. A CNN model was used to classify each superpixel as a flower or a non-flower object. While this approach showed higher performance than conventional ML methods, it has a potential limitation in region proposal. The advantage of the end-to-end CNN architecture is that they are able to use richer features for accurate localization, especially when the images vary greatly. However, superpixels are subject to image variation and may not provide optimal region proposals. The generalizability of this approach, therefore, is most likely lower than that of end-to-end methods.

Many studies have also investigated semantic segmentation-based approaches

for plant/plant organ counting. CNN architectures for semantic segmentation were first used to obtain plant/plant-organ masks. Then, the obtained masks were post-processed using conventional computer vision methods to isolate individual plants/plant organs so that the objects could be counted. An obvious concern is that although CNNs could provide accurate semantic masks, counting accuracy may still suffer from inaccurate post-processing. To address this concern, studies have explored the use of instance segmentation CNNs that can directly segment individual objects in images. These studies faced the same challenge in the lack of training data. To overcome this limitation, most of these studies developed algorithms to generate synthetic images for model training. Two types of image synthesizing methods have been proposed: rule-based and GAN. Rule-based methods use a predefined leaf model to generate a plant based on predefined growth rules. GAN-based approaches,however, could generate synthetic images without sacrificing leaf structure. Thus, a method combining rule-based and GAN-based methods was developed for image synthesizing.

2.4 3D Phenotyping Platform for Komatsuna dataset

In this paper, they built a platform to capture overhead views of a plant and measure the environmental conditions around it. They first selected a plant species, Komatsuna, which is a Japanese mustard spinach and a leafy vegetable. It can be grown indoors, and is known to be resistant to pests and grows very fast. Hydroponic culture is an alternative method.It is clean, irrigation and fertilization are automated so that plant growth is accelerated, fewer agricultural chemicals are needed than soil culture, and replanting failures do not normally occur. Another advantage is that plant root systems can also be measured because roots growing in water can be captured through cameras. For lighting, they used LED lights, with lighting durations and colors programmable using the software provided. To measure ambient temperature, humidity, and light intensity, they used a Sony MESH, where light intensity is measured, which can be measured in lux. However, lux may not be appropriate in a precise sense because it is based on the property of human eyes. To measure the pure energy of lights, photosynthetic photon flux density (PPFD) is more appropriate in biology. In their platform, lux is used as a simplified index in environmental information. An RGB-D camera consists of an RGB camera and a structured light or time-of-flight depth camera based on infrared lights. Since the leaf size of plants in the early stages of growth is small, depth images usually need to be captured at a closer distance from the plant to magnify the size in the captured images. However, some RGB-D cameras are not designed for such short depth ranges. In their platform, they used Intel RealSense SR300 which is specifically optimized for short-range acquisition.

They created two types of datasets using an RGB-D camera and a multiple RGB camera at approximately 2400 lux, 28 °C, and 30% humidity.

2.4.1 RGB-D Dataset

Since one RGB-D camera was attached to capture the entire part of the toolkit, each plant region was manually segmented as a plant image and labeled as a label image. In order to use the dataset for temporal leaf tracking, the same leaf label was assigned to the same leaf in images captured at different times. The original camera resolution was 640 × 480 and the resolution of the plant images was, for example, 166 × 190 pixels. Because the viewpoints of the RGB and depth images were aligned by the camera library, the labels were valid for both images. Due to the mismatch between RGB and depth camera, the resulting images were not aligned, so the depth image was aligned into an RGB image in our platform. For this reason, we provided the original depth image and a transformation matrix from the depth image viewpoint to the RGB image viewpoint so that the point cloud could also be transformed as needed without any interpolation.

2.4.2 Multi-View Dataset

For the multiview dataset, they captured five plants from three different viewpoints and manually segmented them into an individual plant like images. To use the dataset for spatiotemporal tracking of leaves, the same leaf label was assigned to the same leaf in images captured with different cameras at different times. The ground truth of the 3D shape was not measured because it was not easy to capture more accurate 3D shapes than when using multiple high-resolution images. This dataset will be useful for evaluating the segmentation of spatiotemporal instances for multiple views. In the relevant literature, instance segmentation has typically been performed using a single view image, and may be more difficult than using multiple views.

2.5

Chapter 3

Algorithm Design

Chapter 4

Implementation

Chapter 5

Experimental Results

Chapter 6

Conclusion

Bibliography

[1] A. Einstein, "Zur Elektrodynamik bewegter Körper. (German) [On the electrodynamics of moving bodies]," *Annalen der Physik*, vol. 322, no. 10, pp. 891–921, 1905.