Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Transfer learning for the classification of sugar beet and volunteer potato under field conditions

Hyun K. Suh*, Joris IJsselmuiden, Jan Willem Hofstee, Eldert J. van Henten

Farm Technology Group, Wageningen University, P.O.box 16, 6700 AA, Wageningen, the Netherlands

ARTICLE INFO

Article history:
Received 23 December 2016
Received in revised form
11 June 2018
Accepted 19 June 2018

Keywords:
Automated weed control
Weed classification
Deep learning
Transfer learning
Convolutional neural network

Classification of weeds amongst cash crops is a core procedure in automated weed control. Addressing volunteer potato control in sugar beets, in the EU Smartbot project the aim was to control more than 95% of volunteer potatoes and ensure less than 5% of undesired control of sugar beet plants. A promising way to meet these requirements is deep learning. Training an entire network from scratch, however, requires a large dataset and a substantial amount of time. In this situation, transfer learning can be a promising solution. This study first evaluates a transfer learning procedure with three different implementations of AlexNet and then assesses the performance difference amongst the six network architectures: AlexNet, VGG-19, GoogLeNet, ResNet-50, ResNet-101 and Inception-v3. All nets had been pre-trained on the ImageNet Dataset. These nets were used to classify sugar beet and volunteer potato images taken under ambient varying light conditions in agricultural environments. The highest classification accuracy for different implementations of AlexNet was 98.0%, obtained with an AlexNet architecture modified to generate binary output. Comparing different networks, the highest classification accuracy 98.7%, obtained with VGG-19 modified to generate binary output. Transfer learning proved to be effective and showed robust performance with plant images acquired in different periods of the various years on two types of soils. All scenarios and pre-trained networks were feasible for real-time applications (classification time < 0.1 s). Classification is only one step in weed detection, and a complete pipeline for weed detection may potentially reduce the overall performance.

1. Introduction

Volunteer potato is a source of potato blight (*Phytophthora infestans*) and viral diseases. Volunteer potato in a sugar beet field can reduce the crop yield by 30% (O'Keeffe, 1980). There is a statutory obligation for sugar beet farmers in the Netherlands to control volunteer potato plants to no more than two remaining plants per m^2 by 1st of July (Nieuwenhuizen, 2009). For the automated control of volunteer potato in a sugar beet field, a vision-based and small-sized robot was developed within the EU-funded project SmartBot. Due to the small size of the robot and the required battery operation, the platform design had to refrain from additional infrastructure and needed to be able to robustly detect weeds in a scene that was

* Corresponding author.
E-mail address: davidsuh79@gmail.com (H.K. Suh).
https://doi.org/10.1016/j.biosystemseng.2018.06.017

fully exposed to ambient lighting conditions. Additional infrastructure such as a hood and lighting equipment, as used for instance by Nieuwenhuizen, Hofstee, and Van Henten (2010) and Lottes et al. (2016), was not considered viable. The robotic platform is shown in Fig. 1.

The classification of weeds amongst cash crops, i.e. weed/crop discrimination, is the core procedure for automated weed detection. In a pipeline for weed detection, vegetation segmentation is followed by classification of the segmented vegetation into weeds and crop. This classification step traditionally involves two aspects: selection of the discriminative features as well as selection of the classification techniques (Suh, Hofstee, IJsselmuiden, & Van Henten, 2016).

Regarding the features used for discrimination, many studies have used colour, shape (biological morphology) and texture on an individual basis or as a combination of multiple features (Ahmed, Al-Mamun, Bari, Hossain, & Kwan, 2012; Gebhardt & Kühbauch, 2007; Persson & Åstrand, 2008; Pérez, López, Benlloch, & Christensen, 2000; Slaughter, Giles, & Downey, 2008; Swain, Nørremark, Jørgensen, Midtiby, & Green, 2011; Zhang, Kodagoda, Ruiz, Katupitiya, & Dissanayake, 2010; Åstrand & Baerveldt, 2002). However, these features have shown poor performance under widely varying natural light conditions (Suh, Hofstee, IJsselmuiden, & Van Henten, 2018). Other features such as Scale Invariant Feature Transform (SIFT) (Lowe, 2004) and Speeded Up Robust Features (SURF) (Bay, Ess, Tuytelaars, & Van Gool, 2008) have shown their potential in recent studies in the classification of plant species (Kazmi, Garcia-Ruiz, Nielsen, Rasmussen, & Andersen, 2015; Suh et al., 2018; Wilf et al., 2016). However, the highest classification accuracy using SIFT and SURF obtained in Suh et al. (2018) is still not a satisfactory performance in view of the requirements set by the previous study of Nieuwenhuizen (2009): the resulting automatic weeding system should effectively control more than 95% of the volunteer potatoes as well as ensure less than 5% of undesired control of the sugar beet plants. Therefore, within the framework of the EU Smartbot Project, a solution was needed that achieves a classification accuracy of 95% or more as well as a misclassification of both sugar beet [false-negative (FN)] and volunteer potato [false-positive (FP)] of less than 5%. In addition, a classification time of less than 0.1 s per image was also needed because these algorithms should be used in a real-time field application.

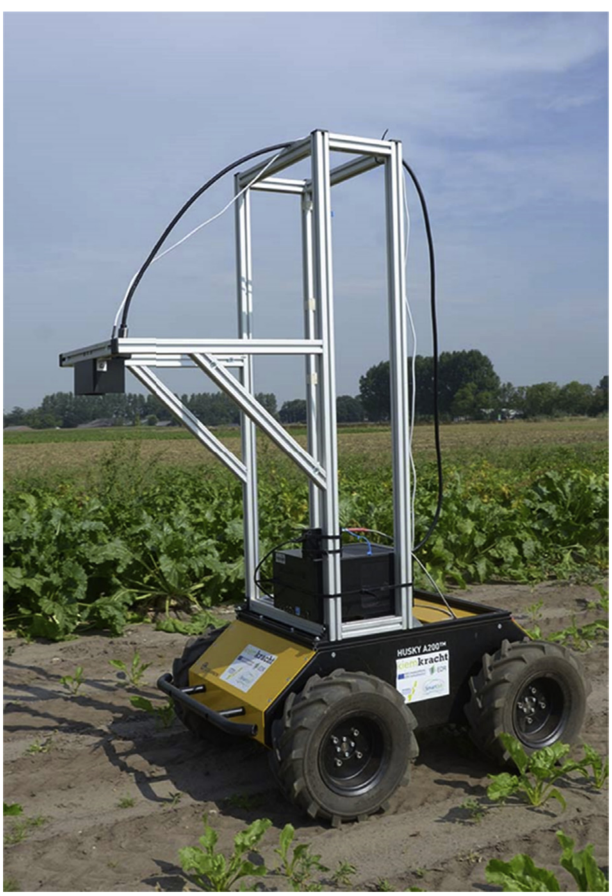

Fig. 1 – The robotic platform for volunteer potato control in a sugar beet field.

A promising way to meet these requirements is to use a deep learning approach. In recent studies, the deep neural network has shown its potential in an agricultural context for plant identification and classification. Grinblat, Uzal, Larese, and Granitto (2016) used a convolutional neural network (ConvNet, or CNN), a specific type of deep network, for plant identification from leaf vein patterns. Although the binary images of vein patterns were used, the study showed the potential of ConvNet for plant identification. Sun, Liu, Wang, and Zhang (2017) used a residual network (ResNet), one of the most common ConvNet architectures used for classification tasks, for plant species identification with images acquired by mobile phones. A 91.78% of classification accuracy was obtained, but they needed 10,000 images to train the network. Dyrmann, Karstoft, and Midtiby (2016) classified 22 plants species using a ConvNet and obtained 86.2% of classification accuracy. In their study, images were acquired under controlled conditions, a quite distinct difference from the conditions that confronted SmartBot, and the number of images needed to train the network from scratch was even more than 10,000. Obtaining such a large number of images, however, is a challenging task in agricultural fields (Xie, Jean, Burke, Lobell, & Ermon, 2016). Besides, training an entire ConvNet from scratch requires a substantial amount of time (Jean et al., 2016; Yosinski, Clune, Bengio, & Lipson, 2014) and is an expensive task that may be hard to realise in practice. Then, transfer learning can be a promising solution.

The objective and novelty of this paper are to deal with crop/weed classification under uncontrolled agricultural environments as well as to reduce the amount of data and time using transfer learning.

Transfer learning has gained its success in real-world applications (Jean et al., 2016; Shin et al., 2016; Yi; Sun, Wang, & Tang, 2014; Xie et al., 2016). Transfer learning, according to Goodfellow, Bengio, and Courville (2016), refers to exploiting what has been learned from one setting into another different setting. In transfer learning, the base network is trained on a base dataset and task, and then the (pre-)trained network is reused for another task (Yosinski et al., 2014). Interestingly enough, though the ConvNet is trained with a specific dataset to perform a specific task, the generic features extracted from ConvNet seem to be powerful and perform very well on other