Precision Agriculture Based on Bayesian Neural Network

Lei Zhao[1] and Dugam Um[2]

[1]Department of Computing Sciences, Texas A&M University-Corpus Christi
[2]Department of Engineering, Texas A&M University-Corpus Christi
lzhao.islander@islander.tamucc.edu, Dugan.Um@tamucc.edu

Abstract

Precision agriculture, utilizing various information to manage crop production, has become the important approach to imitate the food supply problem around the world. Accurate prediction of crop yield is the main task of precision agriculture. With the help of neural networks, precision agriculture has progressed rapidly in past decades. However, neural networks are notoriously data-hungry and data collection in agriculture is expensive and time-consuming. Bayesian neural network, extending the neural network with Bayes inference, is useful under such circumstance. Moreover, Bayesian allows to estimate uncertainty associated with prediction which makes the result more reliable. In this paper, a Bayesian neural network was applied a small dataset and the result shows Bayesian neural network is more reliable under such circumstance.

1 Introduction

Producing enough food to fulfill the food requirement of the worldwide population is always one of the most challenging tasks related to the existence of human being. Moreover, because of the rapid increase of population and human being activities, the serious environmental pollution and the reduction of natural resources such as arable land and fresh water make this task even more challenging. Precision agriculture, also known as smart farming, is considered as a promising solution and has been drawn great attention and processed rapidly in these years. Precision agriculture can be defined as "the application of modern information technologies to provide, process, and analyze multi-source data of high spatial and temporal resolution for decision making and operations in the management of crop productions [1]". Precisely predicting the growth stage of crop is one of the core tasks in precision agriculture since farmers are able to apply precise number of resources, fertilizer and water as examples, to the specific area where and when they are needed for optimal crop growth and reduction of pollution. Due to the evolution of sensor technologies, taking a picture has become cost less and computer vision usage has grown to satisfy with the demand for fast and accurate methods in crop productions [2]. Meanwhile, as a result of the blossoming of machine learning technologies, Convolution neural network (CNN) has become the dominant methods in computer vision and many successful applications have been achieved. Romualdo et al. analyzed the nitrogen nutritional status of the maize plants through computer vision techniques. The result shows the improvement of detection efficiency and accuracy comparing to the traditional methods [3]. Oberti et al. conducted the research of the detection of powdery mildew on grapevine leaves based on computer vision techniques and the overall quality of the plants was improved by the significant improvement of accuracy [4]. However, CNN are notorious for data hungry. A CNN model requires large datasets to train and is prone to overfitting when trained with small datasets. Various regularization techniques such as early stop, dropout, etc., are widely employed to tackle overfitting. However, those techniques show limited effects when training with small datasets. Even

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

though increasing the dataset size is regarded as the most effective ways, collecting data could be extremely expensive and time-consuming in many domains, agriculture and medicine are as examples. Generative adversarial network (GAN), first introduced by Ian Goodfellow in [5], attracts many researchers. In computer vision, GAN has been proved as an efficient way to generate new data and CNNs are applied as the backbone. However, CNNs operate as block boxes and the uncertainty associated with the results are always missing which does not satisfy with the requirements of precision agriculture. Bayesian neural network (BNN) refers as the extension of CNNs with posterior inference. In contrast to CNNs, BNNs perform well in dealing with the problems where data is scarce and estimating the uncertainty in predictions. More reliable and explainable decisions could be made when considering the uncertainty associated with the predictions. In this paper, we predicted the growth stage of tomatoes from a single RGB image. Because of the shortage of data, a GAN model was applied to produce new data and a BNN model was utilized to generate predictions with uncertainty. The result shows that BNN performs more stable with scarce data and the predictions associated with uncertainty are more reliable.
The remaining paper is organized as following: in section 2, the methodologies of GAN and BNN are introduced. In section 3, experiment and result are presented. Finally, conclusions and feature work are discussed in section 4.

2 Methodology

2.1 Generative Adversarial Network (GAN)

Generative adversarial network was first introduced by Ian Goodfellow, et al. in 2014 []. The purpose of GAN is to generate new data with learning the latent distribution of training dataset. Generally, a GAN model is composed of two neural networks, named discriminator and generator respectively. The discriminator distinguishes the difference between the new data generated from the generator and the training data. The generator tries to fool the discriminator by generating the new data. The GAN model is trained iteratively until the discriminator cannot distinguish the difference.

2.2 Bayesian Neural Network (BNN)

Bayesian neural network refers to the extension of neural network with Bayes inference. The motivation comes from the observations that aggregating the predictions of a large set of average-performance can lead to better predictions than a single well performance expert predictor. In contrast to CNNs which optimize the weight values in each layer, either each weights or activation are considered from some distributions. The problem converts to estimate those distributions by optimizing the parameters of those parameters. The parameters are updated based on the Bayes theorem which is described as (1).

$$P(w|\mathcal{D}) = \frac{P(\mathcal{D}|w)P(w)}{P(\mathcal{D})} \tag{1}$$

where $P(w)$ is the prior distribution, $P(\mathcal{D}|w)$ is the likelihood, $P(w|\mathcal{D})$ is the posterior, and $P(\mathcal{D})$ is the marginal distribution or evidence. The purpose of Bayes inference is to find w to maximize $P(\mathcal{D}|w)$. Since $P(\mathcal{D})$ is intractable, Markov Chain Monte Carlo (MCMC) and Variational Inference (VI) are two commonplace methods to solve function (1). Compared to MCMC, VI could achieve the close results of MCMC by consuming much less time [6]. To tackle the evidence problem, VI tries to estimate the posterior distribution from a simple distribution by minimizing the Kullback-Leibler (KL) divergence between the two distributions. The KL divergence function is defined as (2):

$$D_{KL}(q(w;\theta)||p(w|\mathcal{D})) = \sum_{i=1}^{N} q(w;\theta)(\log q(w;\theta) - \log p(w|\mathcal{D})) \tag{2}$$

where $q(w;\theta)$ is the simple distribution. Gaussian distribution is usually employed as the initialization for the simple distribution. Hence, the problem can be described as an optimization problem shown as (3):

$$q^{\star}(w;\theta) = argmin D_{KL}(q(w;\theta)||p(w|\mathcal{D}) \tag{3}$$

Additionally, (2) could be written as (4) by applying the conditional rule.

$$D_{KL}(q(w;\theta)||p(w|\mathcal{D}) = \mathbb{E}(\log q(w;\theta)) - \mathbb{E}(\log p(w,\mathcal{D})) + \mathbb{E}(\log p(\mathcal{D})) \tag{4}$$

The minimization problem changes to minimize the right side of function (4). A function called evidence lower bound (ELBO) is defined as (5):

$$ELBO = \mathbb{E}(\log p(w, \mathcal{D})) - \mathbb{E}(\log q(w; \theta)) \tag{5}$$

Substitute (5) in (4), the function is written as (6)

$$D_{KL}(q(w; \theta)||p(w|\mathcal{D})) = -ELBO + \mathbb{E}(\log p(\mathcal{D})) \tag{6}$$

In (6), $D_{KL}(q(w; \theta)||p(w|\mathcal{D}))$ is non-negative and $\mathbb{E}(\log p(\mathcal{D}))$ is constant. Hence, the minimization problem is able to be solved by maximizing the ELBO

3 Experiments and Results

3.1 Data Collection

All the images were collected in a greenhouse located at Agri Life Center, Corpus Christi, Texas. 54 tomatoes were planted for the experiment. Images were collected weekly by a GoPro camera adhered to a cablebot and labeled with the collection date. Due to the time-consuming and limited space, only 236 images were gathered including 5 weeks. Samples are shown as Fig. 1

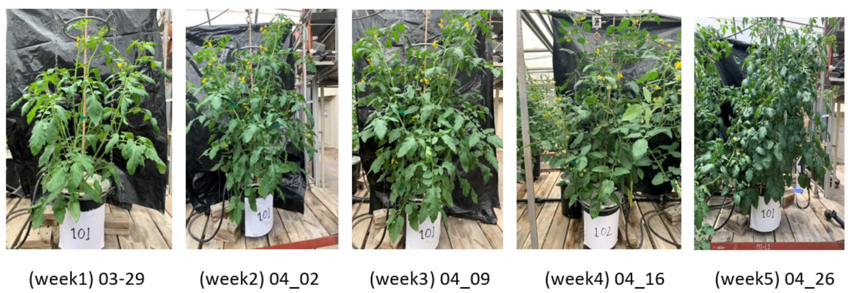

Figure 1: Samples of the dataset

3.2 Experiments and Results

Due to the shortage of data, a GAN model was adopted to increase the size of dataset. Since deep CNN models are prone to overfitting with small data, two shallow CNN model were created, defined as generator and discriminator. The generator consists of 3 transposed convolution layers following by LeakyReLU functions. A dense layer is used as the final output layer. The discriminator has a similar structure expect transpose convolution layers were substituted by convolution layers. The structure of GAN model is shown as Fig 2 1500 new images were generated containing 500 images for each week. For comparison, a shallow CNN model and a shallow BNN model were created separately shown as Fig 3. The results are shown in Fig 4 and 5, respectively. In Fig 4, the CNN model achieved 84% accuracy and in Fig 5, the BNN model reached a little lower accuracy. Meanwhile, because of the shortage of data, the curve of the CNN model wiggles heavily and the curve of the BNN model is stable. Moreover, the CNN model is easy to overfitting. To generate the prediction associated with uncertainty from the BNN model, 200 predictions were sampled from the distribution generated from the BNN model and the average was taken. Fig 6 showed the convergence of weights distributions and standard distribution and Fig 7 presents the posterior samples and the final predictive probabilities for each week.

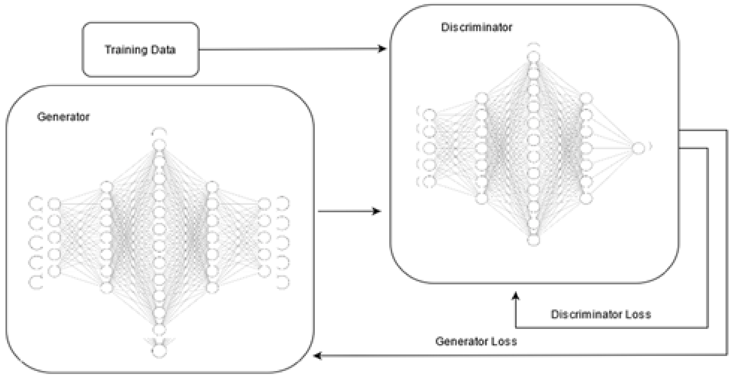

Figure 2: The structure of GAN

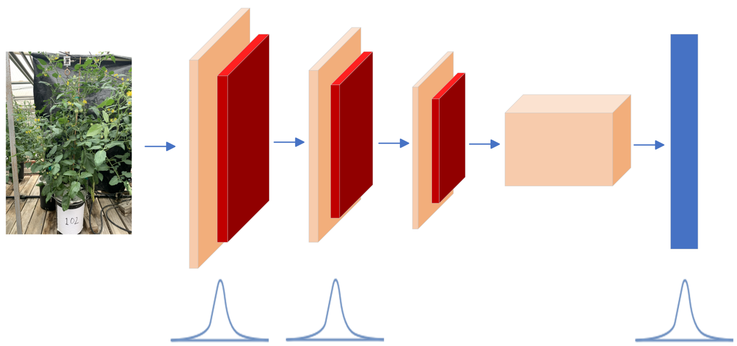

Figure 3: The structure of BNN

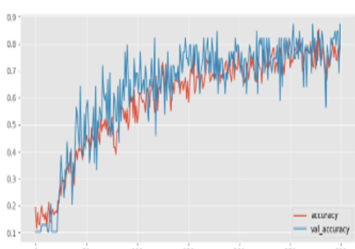

Figure 4: Accuracy For CNN

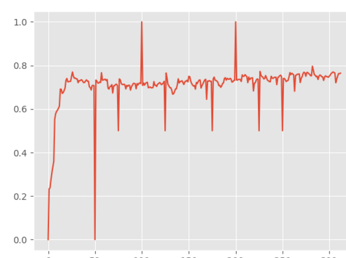

Figure 5: Accuracy For BNN

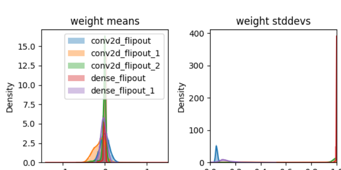

Figure 6: Weights Converge of BNN

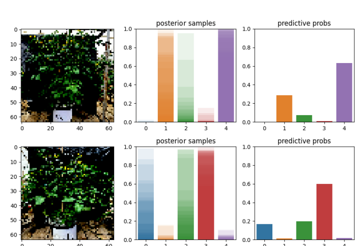

Figure 7: Prediction of BNN

4 Conclusions

In this paper, we studied an approach to deal with the data shortage problem when training a neural network model. Tomato growth stage prediction is used for the study. During the study, although CNN model achieve a little higher result, the curve is wiggling all the time which means the result is unstable. In contrast, BNN shows its potential ability when dealing with the problems were data is scarce. Even though the accuracy is lower than CNN model's, the curve is much more stable. Further, with the distribution generated by the BNN model, confidence level could be computed which could offer the help for farmers to make more reliable and explainable decisions.

References

[1] N. R. Council, *Precision Agriculture in the 21st Century: Geospatial and Information Technologies in Crop Management*. Washington, DC: The National Academies Press, 1997. [Online]. Available: https://www.nap.edu/catalog/5491/precision-agriculture-in-the-21st-century-geospatial-and-information-technologies

[2] D. I. Patrício and R. Rieder, "Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review," *Computers and electronics in agriculture*, vol. 153, pp. 69–81, 2018.

[3] L. Romualdo, P. Luz, F. Devechio, M. Marin, A. Zúñiga, O. Bruno, and V. Herling, "Use of artificial vision techniques for diagnostic of nitrogen nutritional status in maize plants," *Computers and electronics in agriculture*, vol. 104, pp. 63–70, 2014.

[4] R. Oberti, M. Marchi, P. Tirelli, A. Calcante, M. Iriti, and A. N. Borghese, "Automatic detection of powdery mildew on grapevine leaves by image analysis: Optimal view-angle range to increase the sensitivity," *Computers and Electronics in Agriculture*, vol. 104, pp. 1–8, 2014.

[5] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, 2018.

[6] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley, "Stochastic variational inference." *Journal of Machine Learning Research*, vol. 14, no. 5, 2013.