The Augmented Agronomist Pipeline and Time Series Forecasting

George Onoufriou, Marc Hanheide and Georgios Leontidis

March 31, 2020

The Augmented Agronomist Pipeline and Time Series Forecasting

George Onoufriou
MLearn, LCAS. University of Lincoln
School of Computer Science
Lincoln, UK
0000-0002-9316-3196

Marc Hanheide
LCAS. University of Lincoln
School of Computer Science
Lincoln, UK
0000-0001-7728-1849

Georgios Leontidis
University of Aberdeen
School of Natural and Computing Sciences
Aberdeen, UK
0000-0001-6671-5568

***Abstract*—We propose a new pipeline to facilitate deep learning at scale for agriculture and food robotics, and exemplify it using strawberry tabletop. We use this multimodal, autonomously self-collected, distributed dataset for predicting strawberry tabletop yield, aiming at informing both agronomists and creating a robotic attention system. We call this system the augmented agronomist, which is designed for agronomy forecasting, and support, maximizing the human time and awareness to areas most critical. This project seeks to be relatively protective of both its neural networks, and its data, to prevent things such as adversarial attacks, or sensitive method leaks from damaging the future growers livelihoods. Toward this end this project shall take advantage of, and further our existing distributed-deep-learning framework Nemesyst. The augmented agronomist will take advantage of our existing strawberry tabletop in our Riseholme campus, and will use the generalized robotics platform Thorvald for the autonomous data collection.**

***Index Terms*—deep learning, database, agriculture, nemesyst, thorvald, strawberries**

I. Introduction

Machine/Deep learning is becoming a bigger and more important part of our daily lives through the rise of an ever-increasing quantity of available data. 3rd-party services use machine learning in combination with user data for tasks ranging from, natural language processing [5], image recognition, diagnosis [3], detection, classification [6], generation, imputation, broadly prediction; medical diagnosis [2], self-driving cars [8], facial recognition [7], etc. However one area with which deep learning has remained relatively stagnant is in agriculture, where data is scarce, forcing the use of remote sensing datasets or the like, as well as the existing research using classical techniques without many of the recent advances. [1, 4, 13] The primary reason why agriculture has remained relatively constant for this long is likely the lack of, and consistency of data, but also the lack of willingness, and trust of the growers/ agriculturalists to release their potentially sensitive techniques latently in any data they provide. Thus if there is little to no data there can be little advancement with deep learning techniques, meaning prospective research will require self collected data to find any meaningful relations between the features and targets with which to predict accurately and far enough ahead to facilitate timely and effective actions.

This work is supported by a BBSRC CTP-FCR studentship and InnovateUK under grant agreements #105151 and #104587

We contribute our work-in-progress methods and results towards creation of a larger and more accurate plant yield prediction framework which both automates but crucially involves experts in a time-effective and prescribed manner. Our work is facilitated by the RASberry research programme[1], which is a collaboration effort between UoL, Saga Robotics, and BerryGardens, funding autonomous strawberry data collection, under our direct control. This involves the generic expandable Thorvald platform, which is an autonomous robot ready for use in many terrains. Thorvald is an ideal candidate platform to use for our own experiments thanks to its autonomy, and available resources. The only drawback of using strawberries is that they are only grown from late June to early October.

II. Plant Yield Prediction Literature

As it stands there are many existing methods that have been used to attempt to predict crop yield, using data such as remote sensing [16, 4], satellite image, climate conditions, geolocation data, etc. [9] However, there is high variation in the type, quality, and quantity in the datasets used, with very little from a standard dataset with which to use. [14, 16, 15] The vast majority of papers use remote models relying primarily on: temperature, humidity, precipitation, and soil moisture. Some others attempt image based approaches but lack of data is a serious problem for them [13]. This means as far as yield prediction is concerned it is necessary to create a consistent, and granular dataset [14]. All these papers use many different techniques, with a wide variety of data types such that they only marginally narrow the focus for our data collection efforts to things such as climate conditions, [10] meaning we will have to collect a large variety of data and thereafter assess the correlation to achieve the best results.

III. Techniques

As depicted in Fig. 1, we use Nemesyst [11, 12] to manage MongoDB instances across all desired Thorvalds. We aggregate this data to the distributed database layer, where all the data is made available to offline back-end deep learning sites. These sites are responsible for training, and model evaluation of neural network models (NNs), along with packaging them

[1]https://rasberryproject.com

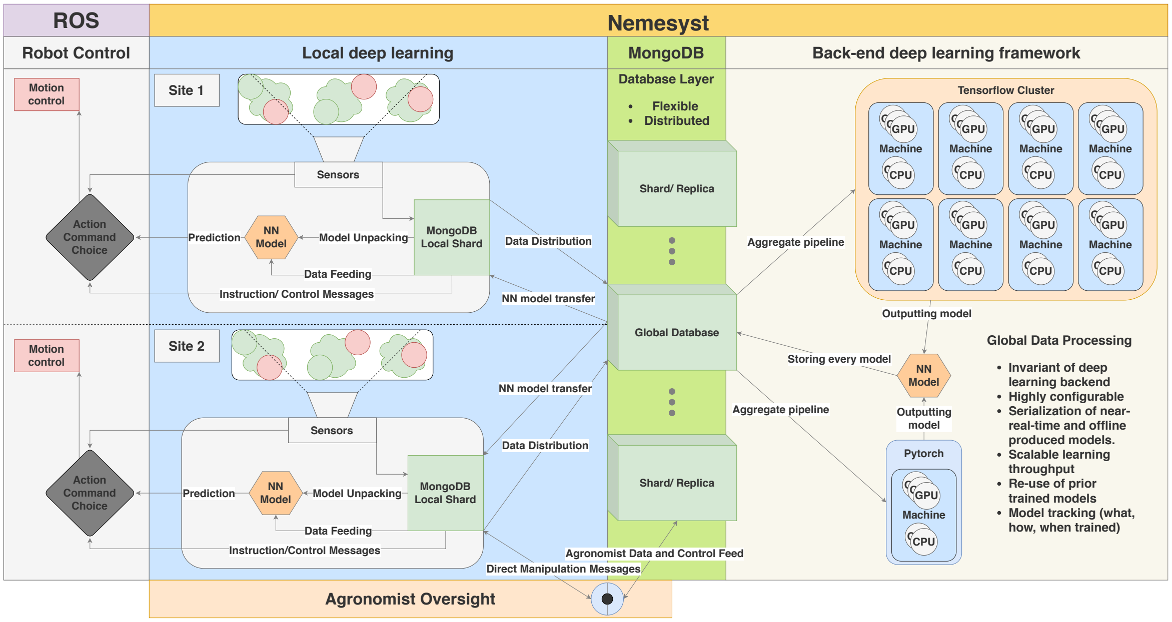

Fig. 1. RASberry data distribution and aggregation pipeline.[11]

back into the database for unpacking and use at a local level. These NNs can then be selected based on their performance and suitability to the application, such as the most performant yield prediction of strawberries versus other berries. The selected NNs used locally can then be used in future to inform decision making processes of the robot, such as attention mechanisms. Attention mechanisms can be used with our databases as a message passing interface to alert and request the attention of specialist agronomists to identify uncertain cases and help with learning along with any immediate control needs.

IV. Results

In the 10-12 fruitful weeks, before the first frosts in October, of June-bearing strawberries, we managed to collect in excess of 40GiB of compressed data (+100GiB if uncompressed), consisting of regular plant imaging, and environment sensing thanks to our Thorvald robots. This data represents a diverse set of scenarios containing:

- 3 still images (RGB, depth, IR, position) every 20cm at various plant angles, and the respective environmental data (temperature, humidity, etc) locally to the camera.
- Continuous environmental data every 15 min. throughout the year, including the lead up to the growing season.
- Stationary camera footage, and its local environment data every 30 minutes for select plants to provide in depth growth and performance data.
- Video based plant image capture during tabletop traversal.
- Yield/ picking data for the number of strawberry punnets collected over time.

However due to constraints in human time, we could only collect yield values twice a week, meaning this last dataset is still relatively small, and would in future need to use autonomous pickers to provide more, and consistent labels to our NNs, to train more advanced NNs. Table I thus shows some very early experimental results that demonstrate the ability of various recurrent networks to learn with this limited labeling. Due to the size of the data and how early on in the process we are our results (I) are split plainly 80% training, 20% testing, with around 10-13 epochs for saturation taking less than a few minutes to train using only environmental and yield data at this early stage.

TABLE I
Time series forecasting of yield by number of punnets.

Technique	Mean Absolute Error (Test set)
Vanilla Recurrent Neural Networks	0.210
Long Short-Term Memory	0.381
Gated Recurrent Units	0.155

V. Conclusion

It has been shown that using distributed Nemesyst database pipelines for data aggregation and modelling as well as distribution in more complex scenarios such as autonomous agricultural data collection, how this can augment the ability of growers to collect data and predict outcomes such as crop yields. A need has been identified for more autonomous data collection to collect more data along with more consistency to feed to NNs to learn more complex representations. Lastly our pipeline can also be used as message passing interfaces for agronomists to monitor, be alerted of any uncertain/ unusual cases, label difficult examples, and potentially control the robots to support their efforts. Our next step is to provide certainty metrics to assess how certain the deep learning models are of the result in such that this can be used for more effective decision making.

References

[1] R Alvarez. "Predicting average regional yield and production of wheat in the Argentine Pampas by an artificial neural network approach". In: *European Journal of Agronomy* 30.2 (2009), pp. 70–77.

[2] Rachel Anderson et al. "Evaluating deep learning techniques for dynamic contrast-enhanced MRI in the diagnosis of breast cancer". In: *Medical Imaging 2019: Computer-Aided Diagnosis*. Vol. 10950. International Society for Optics and Photonics. 2019, p. 1095006.

[3] M Biswas et al. "State-of-the-art review on deep learning in medical imaging." In: *Frontiers in bioscience (Landmark edition)* 24 (2019), pp. 392–426.

[4] Anna Chlingaryan, Salah Sukkarieh, and Brett Whelan. "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review". In: *Computers and electronics in agriculture* 151 (2018), pp. 61–69.

[5] Hai Ha Do et al. "Deep learning for aspect-based sentiment analysis: a comparative review". In: *Expert Systems with Applications* 118 (2019), pp. 272–299.

[6] Hassan Ismail Fawaz et al. "Deep learning for time series classification: a review". In: *Data Mining and Knowledge Discovery* 33.4 (2019), pp. 917–963.

[7] David Güera and Edward J Delp. "Deepfake video detection using recurrent neural networks". In: *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE. 2018, pp. 1–6.

[8] Brody Huval et al. "An empirical evaluation of deep learning on highway driving". In: *arXiv preprint arXiv:1504.01716* (2015).

[9] Konstantinos G Liakos et al. "Machine learning in agriculture: A review". In: *Sensors* 18.8 (2018), p. 2674.

[10] Gniewko Niedbała. "Application of Artificial Neural Networks for Multi-Criteria Yield Prediction of Winter Rapeseed". In: *Sustainability* 11.2 (2019), p. 533.

[11] George Onoufriou. *Nemesyst; Generalised and highly customisable, hybrid-parallelism, database based, deep learning framework*. 2019. URL: https://github.com/DreamingRaven/nemesyst (visited on 11/10/2019).

[12] George Onoufriou et al. "Nemesyst: A Hybrid Parallelism Deep Learning-Based Framework Applied for Internet of Things Enabled Food Retailing Refrigeration Systems". In: *arXiv preprint arXiv:1906.01600* (2019).

[13] Anup K Prasad et al. "Crop yield estimation model for Iowa using remote sensing and surface parameters". In: *International Journal of Applied Earth Observation and Geoinformation* 8.1 (2006), pp. 26–33.

[14] Maryam Rahnemoonfar and Clay Sheppard. "Real-time yield estimation based on deep learning". In: *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping II*. Vol. 10218. International Society for Optics and Photonics. 2017, p. 1021809.

[15] Anna X Wang et al. "Deep transfer learning for crop yield prediction with remote sensing data". In: *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*. ACM. 2018, p. 50.

[16] Jiaxuan You et al. "Deep gaussian process for crop yield prediction based on remote sensing data". In: *Thirty-First AAAI Conference on Artificial Intelligence*. 2017.