Distributed Deep Learning for Persistent Monitoring of agricultural Fields

Yasaman Esfandiari[1], **Koushik Nagasubramanian** [1], **Fateme Fotouhi** [1],

Patrick S. Schnable [1], **Baskar Ganapathysubramanian** [1], **Soumik Sarkar** [1]
[1]Iowa State University
{yasesf,koushikn,fotouhif,schnable,baskarg,soumiks}@iastate.edu

Abstract

Distributed deep learning algorithms have shown eminent performance in learning from data that are privately allocated between several agents. Recent advances in sensor technology have enabled the cheap collection of spatial and temporal high-resolution data for agriculture across a wide geographical area. This continuous increase in the amount of data collected has created both the opportunity for, as well as the need to deploy distributed deep learning algorithms for a wide variety of decision support tasks in agriculture. Distributed deep learning algorithms are typically divided into two major categories: centralized vs decentralized learning algorithms, depending on whether a central parameter server exists for gathering information from participating agents. In the case of rural agriculture applications, transferring a large amount of high-resolution data (e.g., images, videos) collected with IoT devices to a central server/cloud could be very expensive especially with limited communication infrastructure. This suggests the need for decentralized learning approaches, which also naturally provide some measure of privacy. Here, autoencoders are trained using a decentralized optimization algorithm to create a latent representation of growing maize plants in a large-scale field experiment involving several hundred cameras deployed in a maize genome diversity growth experiment. We trained the autoencoders for different communication network topologies of the field-deployed cameras. The feature representations from these autoencoders are then utilized to solve downstream tasks such as anomaly detection and image retrieval. Experimental results show that distributed deep learning is effective in learning from large datasets distributed among several learning agents associated with different cameras. Anomaly detection in particular was useful to make course corrections in imaging protocol and identify localized crop management.

Introduction

Recent dramatic advances in sensor technology have enabled remote (drone, ground) as well as proximal (soil sensors, touch sensors) data acquisition that can target a large range of features (coarse vs. fine spatial resolution, high vs. low temporal frequency, visible vs. hyperspectral wavelengths, chemical vs. physiological attributes) for different decision support tasks in agriculture. Distributed AI-enabled advances are helpful to extract informative agronomic and physiological traits from the large amounts of raw sensor data collected from agricultural fields. Typically distributed deep learning algorithms are divided into two main categories; centralized McMahan et al. [2017], Kairouz et al. [2019] and decentralized Lian et al. [2017], Nedić et al. [2018] learning depending on whether a central parameter server is participating in the learning process Tang et al. [2020]. Centralized learning refers to the class of algorithms which generally contain a parameter server that aggregates information i.e. gradients, model parameters, etc. from the participating agents and performs parameter updates. On the other hand, in decentralized learning, agents communicate

NeurIPS 2021 AI for Science Workshop.

based on some graph topologies to update parameters Lyu et al. [2021]. However, in agriculture applications, transferring a large amount of data (e.g., images, videos) which are collected with IoT devices to a central server/cloud is computationally expensive. Additionally, the central parameter server can act as a single point of failure. Therefore, decentralized deep learning algorithms Kempe et al. [2003], Xiao and Boyd [2004], Boyd et al. [2006], Dekel et al. [2012], Lian et al. [2017], Jiang et al. [2017], Yu et al. [2019], Koloskova et al. [2019], Nadiradze et al. [2019], Balu et al. [2021], Esfandiari et al. [2021] appear to act as a better alternative. As a result, we consider *DPMSGD* Lian et al. [2017] which is a decentralized learning algorithm as the method to train deep learning models in a distributed fashion.

Identification of inputs that lie far away from the training distribution (in-distribution) is called Out-of-distribution (OOD) detection, anomaly detection or outlier detection Grubbs [1969]. OOD detection has been used in broad range of safety-critical applications including medical diagnosis Caruana et al. [2015], autonomous driving Evtimov et al. [2017], cyber-security Kruegel and Vigna [2003], and biometric authentication Gunther et al. [2017]. In this paper, we deal with the challenging scenario of fully unsupervised outlier detection where the goal is the goal is to detect outliers from data containing both normal and outlier patterns. We are using LOF Breunig et al. [2000] and cluster-conditioned detection Sehwag et al. [2021] methods for performing anomaly detection using the autoencoders which are trained in a distributed fashion. Content-based image retrieval (CBIR) technique Knorr et al. [2000] is the computer vision based process of retrieving images that are similar to visual content of a query image from an extensive archive. In the case of anomaly detection, the query image would be a anomaly example. Early research for CBIR considered the global (shape, color, and texture) and local descriptors Mikolajczyk and Schmid [2005], Arampatzis et al. [2013] of an image as a feature vector to perform the retrieval procedure Flickner et al. [1995], Huang et al. [2001]. Recently, representations from deep learning models have been highly efficient for image retrieval tasks Chen et al. [2021]. In this work, we use the nearest neighbors algorithm on the features from the CNN model for the image retrieval task.

Contributions: Specifically, in this paper (i) the use of distributed learning algorithms to train models from real-world agricultural datasets is investigated. (ii) the efficacy of the autoencoder models are shown, (iii) downstream tasks such as anomaly detection, and image retrieval using these models are performed followed by the discussion on the results and future work.

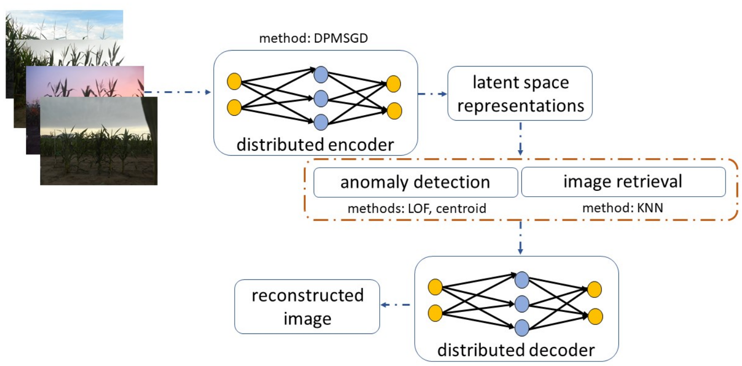


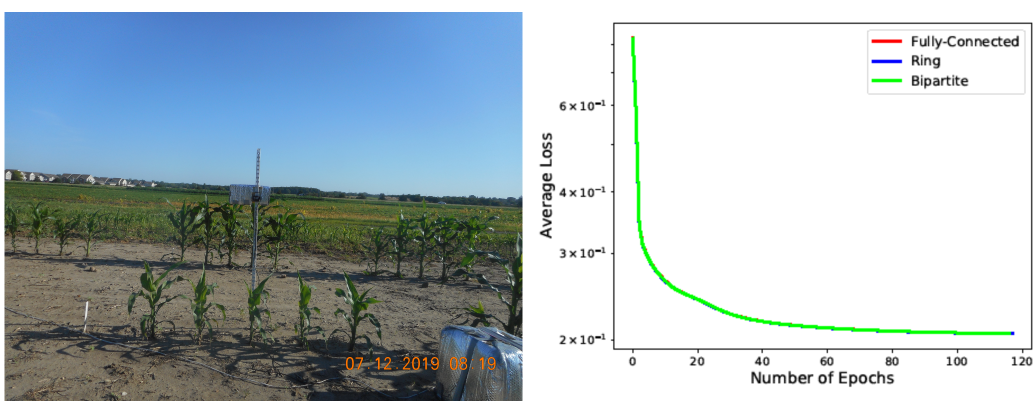
Figure 1: Workflow of the proposed method

Framework

Figure 1 shows the different algorithms used in the training and inference pipeline.**Distributed Deep Learning**: As discussed before, there are several methods for decentralized distributed learning among them *DPMSGD* Lian et al. [2017] has shown to produce prominent results in learning from IID data distributions. In this algorithm, each agent is assigned a model for its portion of the data set. After each agent has computed a local stochastic gradient, it then fetches the optimization variables from neighbors and calculates the neighborhood weighted average. The local optimization variables are then updated using the neighborhood weighted average and the gradients. The consensus model is then achieved by averaging all the local model parameters. **Anomaly Detection**: **Local Outlier Factor**(LOF) is a density-based unsupervised outlier detection algorithm that detects outliers by calculating the local deviation of a given data point Breunig et al. [2000]. The points that contains substantially lower density than its neighbors are considered outliers. We used euclidean distance for identifying the nearest neighbors and used 20 nearest neighbor points to calculate the density of a data point. **Cluster-conditioned detection** Sehwag et al. [2021] method uses the feature representations learnt by the autoencoder for the outlier detection. We assume that majority of the samples in the data are inliers. Then we calculate the centroid of the feature representations of the complete data to represent the inlier data. Then, we use the Mahalanobis Mahalanobis [1936] distance metric to identify the outlier data points that are farthest away from the centroid feature representation. **Image Retrieval**: In the first step, feature maps of the dataset and query image are extracted using the trained CNN model. Then, we use the ball tree method Andoni and Indyk [2017] based nearest neighbors search on the feature maps to identify the k nearest neighbors of the query feature.

Experimental Setup and Results

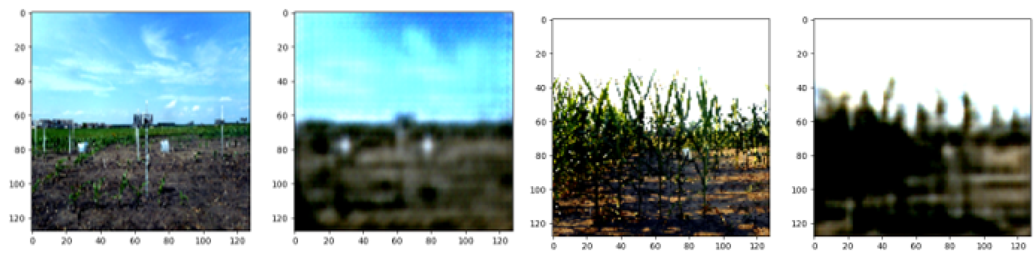
We empirically evaluate the effectiveness of our framework by using the trained model to perform downstream tasks such as anomaly detection and image retrieval on real-world agricultural dataset. To explore the algorithm performance under different topologies, the experiments are performed using 10 learning agents.The agricultural dataset consists of images of 655 rows of field-grown Maize plants. Figure 2a shows the imaging setup of the field-deployed stationary camera. The images were collected at an interval of 20 minutes from 8:00 AM to 5:00 PM over a period of two and half months in 2019 in the midwestern United States. Each row consists of 6 maize plants of a specific genotype. These field photos of maize plants are high-resolution 5152×3864 RGB images. A stationary camera was used to collect data from each row separately and a total of 655 cameras were used. In our experiments, data from a diverse subset of 30 cameras are used to create a dataset of $74\ k$ images. The images were resized to 128×128 pixels each for computational efficiency and the data collection process is described thoroughly elsewhere. We use a deep convolutional neural network (CNN) autoencoder (with 3 convolutional layers with $12, 24, 48$ filters in the encoder section and 3 layers with $24, 12, 3$ filters for the decoder, ReLU activation is used in convolutional layers). A mini-batch size of 128 is used, the initial step-size is set to 0.01. The step size is decayed with a constant 0.981. The stopping criterion is a fixed number of epochs and the momentum parameter (β) is set to be 0.98. We start our analysis in learning real-world agricultural datasets. In these experiments, data from 30 cameras are divided between 10 learning agents so each agent has access to data from 3 distinct cameras. Figure 2b shows that DPMSGD is converging for the agricultural dataset for all the three graph topologies. We then compare the original images in the dataset with their reconstructed counterparts to further analyze the model. Figure 3 shows that the autoencoder is moderately effective in reconstructing the images.



(a) (b)

Figure 2: *a) Illustration of imaging setup used for data collection from a single row of infield Maize plants b) Average training loss for DPMSGD method on agricultural dataset with 10 training agents*

The trained autoencoder model is then used to perform anomaly detection and image retrieval tasks. Figure 4 shows the anomaly predictions of cluster-conditioned detection and LOF methods. Identification of anomalies was particularly useful in making course corrections to imaging protocols of field-deployed stationary cameras. We then use the same model to perform image retrieval. Figure 5 shows some examples of image retrieval results by the nearest neighbors algorithm. These results show that the models trained in decentralized distributed fashion are effective to do image retrieval tasks on diverse images (normal weather, rainy weather, or from tilted cameras).



(a) (b)

Figure 3: *Sample original (left) vs reconstructed (right) images for agricultural dataset with 10 training agents*

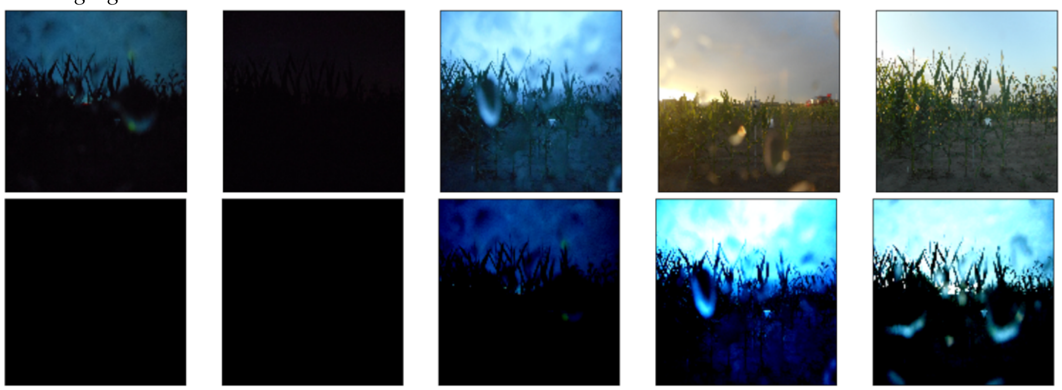


Figure 4: *Anomaly detection on agricultural dataset using (top) cluster-conditioned detection (bottom) LOF method*

Figure 5: *Image retrieval on agricultural dataset. The top images show the query images and the bottom ones are the similar images returned by nearest neighbors algorithm*

Conclusion

In this paper, the viability of distributed learning to train autoencoders which learn from real-world agricultural datasets gathered from field-deployed stationary cameras was discussed. Experiments conducted on this dataset showed that our proposed framework is effective to perform anomaly detection and image retrieval tasks. Future work includes: comparing different distributed learning algorithms, proposing graph topologies to optimize the communications between the cameras, and conducting experiments using larger datasets.

References

Alexandr Andoni and Piotr Indyk. Nearest neighbors in high-dimensional spaces. In *Handbook of Discrete and Computational Geometry*, pages 1135–1155. Chapman and Hall/CRC, 2017.

Avi Arampatzis, Konstantinos Zagoris, and Savvas A Chatzichristofis. Dynamic two-stage image retrieval from large multimedia databases. *Information Processing & Management*, 49(1):274–285, 2013.

Aditya Balu, Zhanhong Jiang, Sin Yong Tan, Chinmay Hedge, Young M Lee, and Soumik Sarkar. Decentralized deep learning using momentum-accelerated consensus. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3675–3679. IEEE, 2021.

Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Randomized gossip algorithms. *IEEE transactions on information theory*, 52(6):2508–2530, 2006.

Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. Lof: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 93–104, 2000.

Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noemie Elhadad. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1721–1730, 2015.

Wei Chen, Yu Liu, Weiping Wang, Erwin Bakker, Theodoros Georgiou, Paul Fieguth, Li Liu, and Michael S Lew. Deep image retrieval: A survey. *arXiv preprint arXiv:2101.11282*, 2021.

Ofer Dekel, Ran Gilad-Bachrach, Ohad Shamir, and Lin Xiao. Optimal distributed online prediction using mini-batches. *The Journal of Machine Learning Research*, 13:165–202, 2012.

Yasaman Esfandiari, Sin Yong Tan, Zhanhong Jiang, Aditya Balu, Ethan Herron, Chinmay Hegde, and Soumik Sarkar. Cross-gradient aggregation for decentralized learning from non-iid data. *arXiv preprint arXiv:2103.02051*, 2021.

Ivan Evtimov, Kevin Eykholt, Earlence Fernandes, Tadayoshi Kohno, Bo Li, Atul Prakash, Amir Rahmati, and Dawn Song. Robust physical-world attacks on machine learning models. *arXiv preprint arXiv:1707.08945*, 2(3):4, 2017.

Myron Flickner, Harpreet Sawhney, Wayne Niblack, Jonathan Ashley, Qian Huang, Byron Dom, Monika Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovic, et al. Query by image and video content: The qbic system. *computer*, 28(9):23–32, 1995.

Frank E Grubbs. Procedures for detecting outlying observations in samples. *Technometrics*, 11(1): 1–21, 1969.

Manuel Gunther, Steve Cruz, Ethan M Rudd, and Terrance E Boult. Toward open-set face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 71–80, 2017.

Jing Huang, Shanmugasundaram Ravi Kumar, Mandar Mitra, and Wei-Jing Zhu. Image indexing using color correlograms. Technical report, Cornell Univ., Ithaca, NY (United States), 2001.

Zhanhong Jiang, Aditya Balu, Chinmay Hegde, and Soumik Sarkar. Collaborative deep learning in fixed topology networks. *Advances in Neural Information Processing Systems*, 2017:5905–5915, 2017.

Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Keith Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and open problems in federated learning. *arXiv preprint arXiv:1912.04977*, 2019.

David Kempe, Alin Dobra, and Johannes Gehrke. Gossip-based computation of aggregate information. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pages 482–491. IEEE, 2003.

Edwin M Knorr, Raymond T Ng, and Vladimir Tucakov. Distance-based outliers: algorithms and applications. *The VLDB Journal*, 8(3):237–253, 2000.

Anastasia Koloskova, Tao Lin, Sebastian U Stich, and Martin Jaggi. Decentralized deep learning with arbitrary communication compression. *arXiv preprint arXiv:1907.09356*, 2019.

Christopher Kruegel and Giovanni Vigna. Anomaly detection of web-based attacks. In *Proceedings of the 10th ACM conference on Computer and communications security*, pages 251–261, 2003.

Xiangru Lian, Ce Zhang, Huan Zhang, Cho-Jui Hsieh, Wei Zhang, and Ji Liu. Can decentralized algorithms outperform centralized algorithms? a case study for decentralized parallel stochastic gradient descent. In *Advances in Neural Information Processing Systems*, pages 5330–5340, 2017.

Xueguang Lyu, Yuchen Xiao, Brett Daley, and Christopher Amato. Contrasting centralized and decentralized critics in multi-agent reinforcement learning. *arXiv preprint arXiv:2102.04402*, 2021.

Prasanta Chandra Mahalanobis. On the generalized distance in statistics. National Institute of Science of India, 1936.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics*, pages 1273–1282. PMLR, 2017.

Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *IEEE transactions on pattern analysis and machine intelligence*, 27(10):1615–1630, 2005.

Giorgi Nadiradze, Amirmojtaba Sabour, Dan Alistarh, Aditya Sharma, Ilia Markov, and Vitaly Aksenov. SwarmSGD: Scalable decentralized SGD with local updates. *arXiv preprint arXiv:1910.12308*, 2019.

Angelia Nedić, Alex Olshevsky, and Michael G Rabbat. Network topology and communication-computation tradeoffs in decentralized optimization. *Proceedings of the IEEE*, 106(5):953–976, 2018.

Vikash Sehwag, Mung Chiang, and Prateek Mittal. Ssd: A unified framework for self-supervised outlier detection. *arXiv preprint arXiv:2103.12051*, 2021.

Zhenheng Tang, Shaohuai Shi, Xiaowen Chu, Wei Wang, and Bo Li. Communication-efficient distributed deep learning: A comprehensive survey. *arXiv preprint arXiv:2003.06307*, 2020.

Lin Xiao and Stephen Boyd. Fast linear iterations for distributed averaging. *Systems & Control Letters*, 53(1):65–78, 2004.

Hao Yu, Rong Jin, and Sen Yang. On the linear speedup analysis of communication efficient momentum sgd for distributed non-convex optimization. *arXiv preprint arXiv:1905.03817*, 2019.