MLOp Lifecycle Scheme for Vision-based Inspection Process in Manufacturing

Junsung Lim, Hoejoo Lee, Youngmin Won, and Hunje Yeon, ***Samsung Research***

https://www.usenix.org/conference/opml19/presentation/lim

This paper is included in the Proceedings of the 2019 USENIX Conference on Operational Machine Learning (OpML '19).

May 20, 2019 • Santa Clara, CA, USA

ISBN 978-1-939133-00-7

Open access to the Proceedings of the 2019 USENIX Conference on Operational Machine Learning is sponsored by USENIX.

MLOp Lifecycle Scheme for Vision-based Inspection Process in Manufacturing

Junsung Lim, Hoejoo Lee, Youngmin Won, Hunje Yeon
Samsung Research
{junsung.lim, hoejoo.lee, ymin.won, hunje.yeon}@samsung.com

Abstract

Recent advances in machine learning and the proliferation of edge computing have enabled manufacturing industry to integrate machine learning into its operation to boost productivity. In addition to building high performing machine learning models, stakeholders and infrastructures within the industry should be taken into an account in building an operational lifecycle. In this paper, a practical machine learning operation scheme to build the vision inspection process is proposed, which is mainly motivated from field experiences in applying the system in large scale corporate manufacturing plants. We evaluate our scheme in four defect inspection lines in production. The results show that deep neural network models outperform existing algorithms and the scheme is easily extensible to other manufacturing processes.

1 Introduction

Machine learning(ML) have begun to impact various industrial fields and manufacturing is no exception. Manufacturers, in preparation for the smart manufacturing era to come, aim to improve their competitiveness by adapting new technologies that excel product quality, cut down production cost and reduce lead time in production [7]. Manufacturing industry is an attractive field for ML Operations(MLOps) for number of reasons. First, a huge volume of data is generated, forming the foundation for source of learning. Secondly, trivial and repeated tasks in production process opens up opportunities for ML models. For instance, consider a defect inspection task in which product surfaces are visually checked for scratches by a human inspector. While the task itself is trivial, thus susceptible to human errors, it is difficult to express a good set of rules for scratch detection. Given the recent advancement in deep neural network(DNN), MLOps have become natural selection for such tasks.

MLOps in production is more than just training and running ML models. Despite large volume of raw data collected, it needs to be cleaned and labeled to use them as a ML training dataset. Test data are generated from multiple devices on network edge and thus running inference on a single server is infeasible due to high latency caused by data communication and inference. Also, a use of off-premise services is not proper as every manufacturing data is confidential and should be stored securely on premise. Last but not least, there are multiple stakeholders with different roles in production process and thus, require different tools at each stage in MLOp lifecycle.

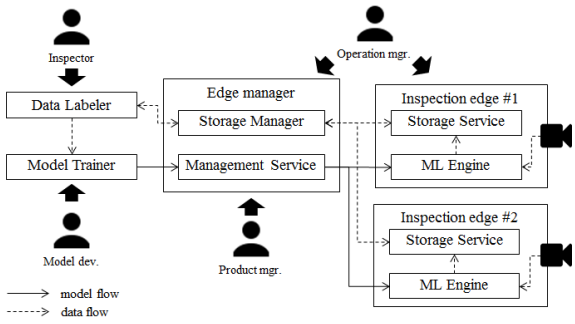


Figure 1: Overall system architecture of the proposed scheme with multiple stakeholders. A circular lifecycle is formed among the components.

In this paper, we propose a MLOp lifecycle scheme for vision-based inspection systems in manufacturing. Figure 1 describes overall architecture and components required for in-factory operations, ranging from data collection, ML model development and deployment on multiple edge devices. Based on the scheme, we developed a MLOp lifecycle solution called CruX. We have successfully set up CruX in Samsung Electronics' smartphone and home appliance plants for scratch, dent and missing part detection. Four DNN models of three different tasks(one-class adversarial net, multi-class classification, and object detection) are trained and deployed to a total of 158 edge devices for inspection. Compared to the existing rule-based algorithms, models achieved at least 32.8% improvement in defect detection accuracy and all inferences at edge took less than 2 seconds per image on CPU.

2 Related Work

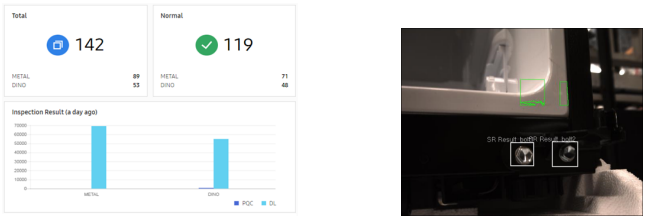
With the popularity of ML, a number of model versioning and serving solutions are available. Data Version Control [1], ModelDB [10] and ModelChimp [2] provide ML model and data pipeline versioning. These solutions, however, require model developers to control versions by either extending existing ML code or setting up extra infrastructure.TensorFlow Serving [8] is a solution to serve TensorFlow model, but it requires models to be accessible from its own file system, leaving the challenge of deploying the model across physically separated edge devices. Complex factors of real field requirements such as different stakeholders in the lifecycle, deployment needs, management and controllability of ML models on multiple edge devices, call for a new operational scheme in the manufacturing industry.

3 Proposed Scheme

We propose a MLOp lifecycle scheme for vision inspection systems, in which four key stakeholders and five components are identified and defined as shown in Figure. 1.

Raw image data are captured by camera which is usually located at the conveyor belt. While some of the images can be annotated by a non ML-domain expert (e.g. identify screw(s) from an image), some are not (e.g. classify scratches by type). Due to this reason, Data Labeler is designed and used by inspectors on site. An intuitive user experience is important as we do not want inspectors spending more time annotating than inspecting the product. Model developers use Model Trainer to train and test DNN models from annotated data. Model Trainer provides a function to train DNN models with different set of hyper-parameters to identify the best hyper-parameter set for the model. The trained model is then uploaded to Edge manager for configuration before deployment. We found this step to be important in production because no edge (or the inspected product) is the same. Model configurations, such as threshold, is adjusted per edge and deployed to edges under the supervision of operation manager. As the inspection continues, statistics are collected and visualized to the product manager.

All the components are modular but interconnected. This is important because it enables the process of training, deploying and running model possible through a single graphical user-interface without having to make any code-level changes.



(a) Edge manager (b) Inference result

Figure 2: (a) Manager monitors inspection status and deploys models to edges. (b) Inference result where detected objects(bolt) are located in white bounding boxes.

4 Evaluation

We implemented the proposed scheme called *CruX*, and applied in two different plants. Back-end components are developed in Python, Java and Go. Data are exchanged among the components using REST APIs and message queues. The proposed scheme supports three different DNN models, namely multi-class classification(ResNet50 [6]), one-class generative adversarial network(GAN [5]) and object detection(YOLO [9]). All are implemented with TensorFlow and fine-tuned from ImageNet [4] pretrained weights. Figure 2 shows a web-based GUI that is provided to end-users. Edge manager and Inspection edge run Windows 7 64bit with 8GB RAM, 2.60GHz CPU and no GPUs.

Table 1 shows the results on production lines. In prior to this, rule-based algorithms [3] are used to detect scratches, dents and missing parts. We noticed that the rule-based algorithms are very sensitive to small changes in data (e.g. image orientation and brightness) and difficult to update. On the other hand, DNN models showed higher defect detection accuracy, outperforming previous method by 32.8% 92.8%. All four production lines required inspection time to not exceed 3 seconds.

5 Conclusion

In this paper, we propose a MLOp scheme for vision inspection in manufacturing. We identify four key stakeholders and five components across in realizing MLOp lifecycle. We successfully applied it on four production fields of smartphone and home appliance plants. ML models trained and deployed by the scheme outperform existing inspection systems, and we aim to update the operation automated as the future work.

Table 1: Defect inspection results on four production lines (*: Defection detection accuracy).

Inspection area	Edges deployed	DNN model (Backbone)	DDA* improvement	Avg. inference time
Scratch (smartphone)	88	Multi-class (ResNet50)	32.8%	760 ms
Dent (smartphone)	52	One-class (GAN)	40.0%	998 ms
Missing part (refrigerator)	9	Object detection (YOLO)	92.8%	1416 ms
Missing part (washing machine)	9	Object detection (YOLO)	85.6%	1632 ms

References

[1] Data science version control system, 2019.

[2] Experiment tracking | modelchimp, 2019.

[3] Daniel Lélis Baggio. *Mastering OpenCV with practical computer vision projects*. Packt Publishing Ltd, 2012.

[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

[5] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[7] Hyoung Seok Kang, Ju Yeon Lee, SangSu Choi, Hyun Kim, Jun Hee Park, Ji Yeon Son, Bo Hyun Kim, and Sang Do Noh. Smart manufacturing: Past research, present findings, and future directions. *International Journal of Precision Engineering and Manufacturing-Green Technology*, 3(1):111–128, 2016.

[8] Christopher Olston, Noah Fiedel, Kiril Gorovoy, Jeremiah Harmsen, Li Lao, Fangwei Li, Vinu Rajashekhar, Sukriti Ramesh, and Jordan Soyke. Tensorflow-serving: Flexible, high-performance ml serving. *arXiv preprint arXiv:1712.06139*, 2017.

[9] J. Redmon and A. Farhadi. Yolo9000: Better, faster, stronger. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6517–6525, July 2017.

[10] Manasi Vartak, Harihar Subramanyam, Wei-En Lee, Srinidhi Viswanathan, Saadiyah Husnoo, Samuel Madden, and Matei Zaharia. Model db: a system for machine learning model management. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*, page 14. ACM, 2016.