IOP Conference Series: Materials Science and Engineering

Development of Machine Learning Applications for Chemical Objects

To cite this article: A Setiawan *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **990** 012001

IOP Conf. Series: Materials Science and Engineering **990** (2020) 012001 doi:10.1088/1757-899X/990/1/012001

Development of Machine Learning Applications for Chemical Objects

ASetiawan[1], S Rostianingsih[1], CI Halim[1]
[1]Informatics Department, Petra Christian University, Surabaya, Indonesia

E-mail: alexander@petra.ac.id

Abstract. Machine learning is a method used to create programs that can learn from data or applications. In contrast to ordinary static computer programs, machine learning programs are programs designed to be able to learn on their own. Machine learning technology has been widely applied in various fields, including in chemistry. Currently the introduction of basic chemical objects is a study that has many concepts that are easy for beginners to understand and imagine, such as atomic structure and molecular shape in chemical bonds. So that an alternative learning media is needed in order to improve understanding of the concept of chemical objects interactively. Machine learning technology is expected to provide a more realistic interaction and is an advancement of a promising technological method that can motivate users to engage in a more active learning system.

1. Introduction

The development of a technology is getting faster and faster. Technology and the development of machine learning is one of the most frequently developing fields of science today. Without realizing it, almost everyone uses products produced by machine learning applications. The term in machine learning is basically the computer's process of learning from data. All technology in machine learning will inevitably involve data. The data can be the same, but an algorithmic concept and a different approach to get the optimal level of ability and results. Presenting chemical processes for students with visualization is highly emphasized in chemistry education, because it can help students develop an understanding of chemical concepts to find the implications of using machine learning in chemistry learning, which is to encourage active, organized, and able learners to integrate the information obtained to understand the concepts and difficult principles and can be used to solve problems. Therefore, important steps that need to be taken are providing learning resources for students to overcome difficulties in understanding chemical concepts.

One technique for applying machine learning is supervised learning. As discussed earlier, machine learning without data will not work. Therefore, the first thing prepared is data. Data will usually be divided into 2 groups, namely training data and testing data. The training data will be used to train the algorithm to find a suitable model, while the testing data will be used to test and determine the performance of the model obtained at the testing stage.

Machine Learning is used to recognize a pattern as is done by our brains. Therefore, we need input, model, and output. When Machine Learning does a new job, it will see what work has previously been done so that it can provide the correct output. Google Cloud Platform has developed tools for users to

[1] alexander@petra.ac.id

IOP Conf. Series: Materials Science and Engineering **990** (2020) 012001 doi:10.1088/1757-899X/990/1/012001

develop Machine Learning in applications. Therefore, to continue previous research, namely 3D Modelling of Chemical Elements with Augmented Reality [1], an introduction to chemical objects was made using machine learning. The system has stored a set of datasets containing chemical objects and their element names. When a new object is scanned with this application, it will be matched with the previous dataset. If it turns out to be recognized it will appear the name of the chemical element. If not recognized, a new dataset can be added based on the results of the object recognition training. The desired end result is to display the chemical element names for the recognized objects.

2. Machine Learning

In machine learning, pattern recognition or feature extraction starts from the initial set of data and builds features that are meant to be informative, facilitating learning and subsequent generalization of steps, not superfluous. For some other cases, machine learning leads to better human interpretations. Several features are closely related to dimensional reduction. When inputting data into an algorithm, the size may be too large to be processed and may be overkill, the measurements are the same in meters and feet. Image repetition is represented as pixels. These can be turned into a reduced feature set (known as a vector pattern). In this process, it is known as pattern extraction. It is hoped that the extracted pattern will contain a relevant information from the input data. Therefore, the desired result is that it can be used for scaled down representations, not the initial complete data [2].

All of this means it's possible to quickly and automatically generate models that can analyst larger, more complex data, and provide faster and more accurate results - even at very large scales. And by building the right model, organizations stand a better chance of identifying profitable opportunities - or avoiding unknown risks. Lots of machine learning the researchers performed their work on single-purpose and general-purpose computers [3,4], and multiple single-machine frameworks supported this scenario. Caffe [5] is a high performances framework for training that is described ina declarative manner neural networks on multicore and general purposes computer purposes. As discussed above, the programming model is similar to DistBelief, so it is easy to model from existing layers, but relatively difficult to add new layers or optimizers. Torch [6] offers a powerful imperative programming model for scientific computation and machine learning. This allows fine grained control over command execution and memory usage, allowing power users to optimize performance their program. While this flexibility is beneficial for research, Torch does have the advantages of graphical data flow as a portable representation across small-scale experiments, production training, and deployment [7].

3. Research Methodology

The research methodology of machine learning application is very important for the success. Therefore, machine learning applications at student must be designed and developed so that they can be used. Testing machine learning must be done, and to received feedback from participants so that the level of application. Synthetic chemistry methodology refers to the methods used for the synthetic dataset objects can be seen in Figure 1.

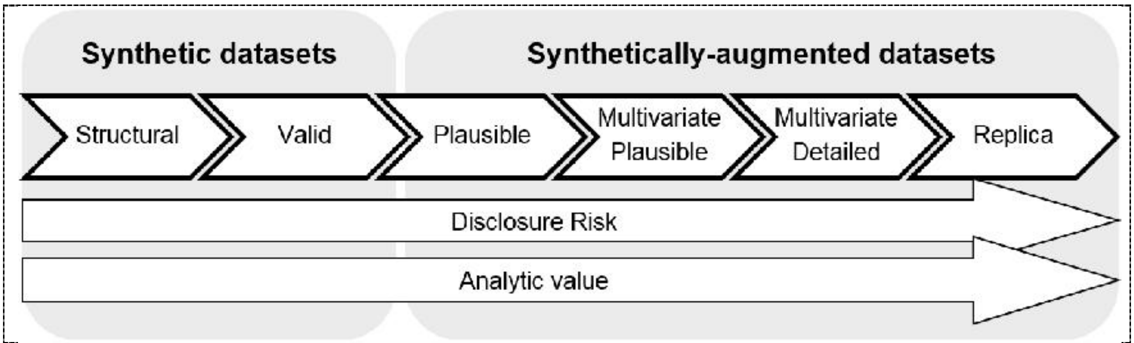

Figure 1. The synthetic datasets.

4. Implementation System

The implementation can be used Jupyter Notebook, Tensorflow, Python 3.6, CUDA Toolkit 10.0 and CUDNN 7.4.1. For machine learning development, Google Cloud Platform provides a machine learning API, while for image or voice recognition it uses TensorFlow. The programming language used is Python. Machine learning API is used to detect images (Cloud Vision API) such as logos or photos, video, natural language (text), voice, and translators. The five of them can be combined into one. From a set of datasets, machine learning (training, deployment, server) is carried out, then predictions are generated using the REST API.

The training dataset, you can use two ways, namely transfer learning or training from scratch. Transfer learning means updating existing data using and conducting training with your own data. While training from scratch is doing training from scratch with only your own data. After the training is complete, two hooks are tested: Testing of synthetic images / combined images, and testing with real images / real photos. Displays several sample images to confirm whether the neural network can read the image correctly can be seen Figure 2.

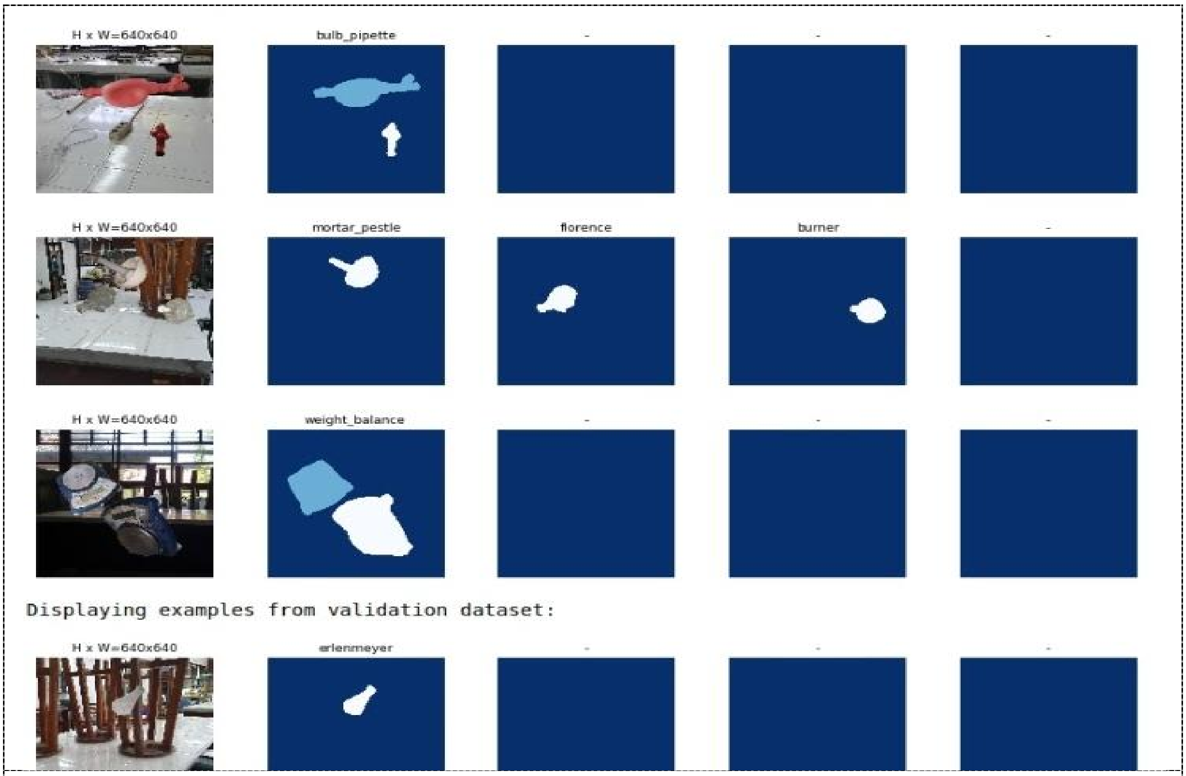

Figure 2. The testing of synthetic dataset.

Training Process:
Training: 4 Epochs, each 1000x cycles (total 4000 cycles)
Fine-tuning: 4 Epochs, each 1000x cycles (total 4000 cycles)
Total 8 epochs or 8000 cycle7.

The prototype that uses Google Machine Learning as the basis of this research, keep in mind that this neural network is being trained using a synthetic dataset, and according to result The Synthetic of Burner and Spektrofotometer can be seen Figure 3, The Synthetic of Weight Balance can be seen Figure 4, The Synthetic of Bulb Pipette can be seen Figure 5, and The Real Synthetic of Weight Balance can be seen Figure 6, when tested to recognize real photos, the results are quite satisfying, namely being able to recognize some of the objects/objects correctly.

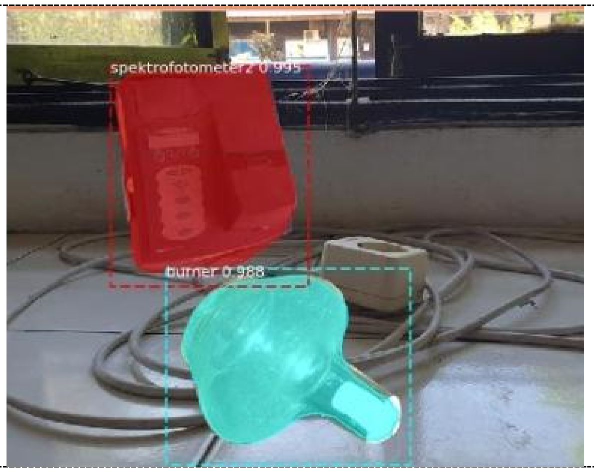

Figure 3. The synthetic of burner and spektrofotometer.

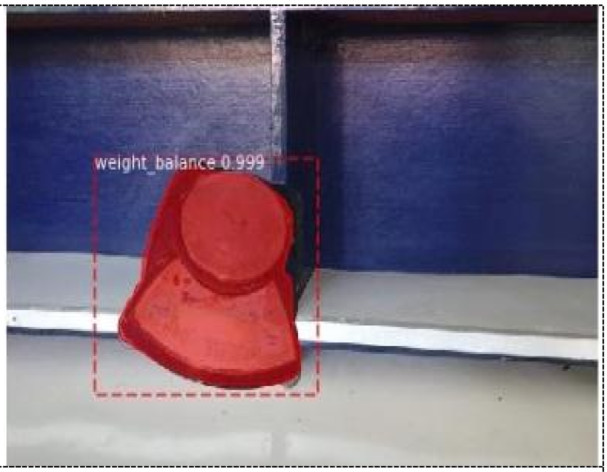

Figure 4. The synthetic of weight balance.

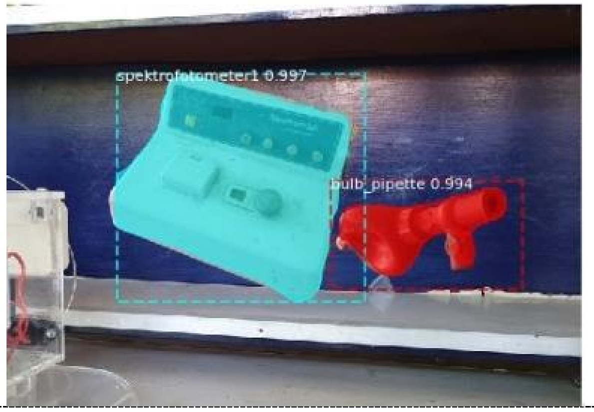

Figure 5. The synthetic of bulb pipette.

Figure 6. The real synthetic of weight balance.

5. Conclusion

The machine learning applications for chemical object have the problem is that most of the lab equipment is made of glass / transparent, such as: erlenmeyer flask, florence flask. It is more difficult for the neural network to recognize correctly, because the atmosphere / colour behind the glass objects affects the performance of the training data. The manual annotation process requires a lot of labour (human labour) and takes a long time to get meaningful results. Because machine learning requires at least thousands of unique images. Using synthetic datasets can speed up the creation of datasets, when training, can produce results that are not bad.

Acknowledgments

This Research Project was funded by a Internal Research Grant, received in 2020. We thank the Centre of Research Petra Christian University for the supports and guidance.

References

[1] Setiawan A, Rostianingsih S and Widodo T R 2018*MATEC Web Conf.* **24805007**

IOP Conf. Series: Materials Science and Engineering **990** (2020) 012001 doi:10.1088/1757-899X/990/1/012001

[2] Yadaf S K and Tiwari A K 2015*Mach. Learn. Appl.***2(3/4)**30-49
[3] Krizhevsky A 2014 One weird trick for parallelizing convolutional neural networks arXiv:1404.5997
[4] Krizhevsky A, Sutskever I and Hinton G E 2012 *Adv. Neural Inf. Process. Syst.* **25(2)**1097-1105
[5] JiaY,Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S and Darrel T 2014*Proc. 22nd ACM International Conference on Multimedia* 675-678
[6] Collobert L, Bengio S and Marithoz J2002*IDIAP Research Report*
[7] Abadi M et al. 2016*Proc. 12th USENIX Symposium on Operating Systems Design andImplementation* 265-283.