Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 38 (2019) 142–147

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

29th International Conference on Flexible Automation and Intelligent Manufacturing (FAIM2019), June 24-28, 2019, Limerick, Ireland.

Object Detection using Convolutional Neural Networks for Smart Manufacturing Vision Systems in the Medical Devices Sector

Kelly O'Brien, Jacqueline Humphries
Limerick Institute of Technology, Ireland

Abstract

Industry 4.0 has opened the doors for Deep Learning to enter into the manufacturing arena with a bid to improve efficiency and quality check process. In many assembly lines Vision Systems are applied that can identify anomalies, read labels, count components and such like. However these systems are sensitive to lighting and setup conditions, and in many cases the technology is unable to read or classify, leaving gaps in the assembly process where human validation is a necessity. A typically manufacturing response is to add further quality control check layers onto the backend of the process. An ideal Industry 4.0 Smart Manufacturing vision system would keep track of components being used, identify anomalies and identify processes successfully during the production stage providing efficient quality checks in real-time, thus creating a more efficient Quality Control process, and move closer to Zero-Defect scenario.

One area in which Vision Systems are rarely used is the medical technology sector, due to the high standards required to approve a line. Because current Vision Systems can fail in different setup conditions, this makes them a risk and so, Quality Control is not in any way aided or improved upon. This study examines the application of Deep Learning with neural networks on components from a medical technology company, to demonstrate how they can be used as a more reliable and less prone to error vision system, that can track the components in real time regardless of lighting conditions and other constraints and perform other Quality Control checks.

Peer-review under responsibility of the scientific committee of the Flexible Automation and Intelligent Manufacturing 2019 (FAIM 2019)

Keywords: *Artificial Intelligence; Deep Learning; Object Detection; Vision Systems*

Peer-review under responsibility of the scientific committee of the Flexible Automation and Intelligent Manufacturing 2019 (FAIM 2019)
10.1016/j.promfg.2020.01.019

1. Introduction

With the introduction of Smart Manufacturing under Industry 4.0, Machine Learning and Vision Systems are beginning to see an increase in their role on the manufacturing floor. Assistive technologies, robotics and smart quality check systems are being rolled out in manufacturing plants worldwide. It is only natural that these would eventually find themselves in the Medical Devices sector. One of the biggest hurdles for vision systems and medical devices is the fact that regulation is so strict in this sector, and vision systems find it difficult to meet these regulations due to the fact that they are still not perceived as being entirely trustworthy and prone to error. However, it has been found that computer vision has an advantage over human vision with some tasks where variability is low [1] and there are opportunities for Vision Systems to be incorporated into the manufacturing line, as an assistive technology or quality control.

1.1. Literature review

Smart Manufacturing is a growing trend in the world of industry having first being conceived in 2010 in Germany as Industry 4.0, or the next Industrial Revolution. The idea behind the concept is that IoT devices will change the manufacturing process completely, outputting large amounts of data at every stage and layer of the process. This data can be utilised to make changes or inspect for error and so improving efficiency tenfold. Machine learning and Vision systems play a very large role in Smart Manufacturing, as the technologies can be used at numerous different stages and instances. Traditional vision systems have found use over the years on the manufacturing floor, however have been proven to be limited in their abilities. With the recent rise in popularity of Deep Learning, improved and more intelligent vision systems are beginning to surface. Assistive technologies are also being introduced to aid human workers which incorporate vision systems as well as AR, VR, etc [1]. They are a more human-centred approach to vision technology usually found in manufacturing. While machines are still not quite trusted to complete every task, they have been proven to be more efficient in some cases.

Deep Learning has begun to play a much more important role in Smart Manufacturing vision systems and analytics thanks to a recent trend in available libraries and packages and improved neural networks such as the convolutional neural net (CNN) which is particularly efficient with image data [2]. One of the biggest issues with applying deep learning to the manufacturing floor however is the dataset problem; the data can sometimes be quite difficult to collect (e.g. samples of defect products). It is, therefore, an important area to be considered when introducing deep learning to the manufacturing process.

CNNs have found popularity due to the fact that they do not require any manual feature extraction unlike traditional machine learning models [3]. They are also a lot better at working in highly variable environments which is imperative in situations where conditions such as lighting, placement, component variation, etc are not stable or continuous. However, CNNs require an immense amount of data and proportional processing power in the form of GPUs (Graphical Processing Units). Although they are much more advantageous to the manufacturing process for their accuracy, the high processing and training means that they are slower to be introduced.

It cannot be ignored that CNNs have significantly improved vision systems but it has also been found [4] that evaluating visions systems statistically in a very uncommon practice due to how difficult it is. This is mostly due to the fact that it is very difficult setup a control and the only way to truly evaluate different vision system is to ensure that they have been trained on the same dataset. This, in itself may cause problems, as the dataset may not suit the system being trained. However, for the purpose of the study in this paper, both systems were compared using the same dataset and McNemars test.

2. Method

Object Detection in dynamic environments is best obtained with use of Convolutional Neural Networks (CNNs) [3]. The most basic of these can be implemented in Python with Keras and OpenCV libraries, while more complex CNNs can be configured using Tensorflow. For this study, two different models were trained.

2.1 Single Shot Multibox Model

The first was a complex model called the Single Shot Multibox (SSD) which is pre-trained on the COCO dataset [5]. This dataset is a large-scale object detection dataset with labelled images of up to 80 object categories. SSD is incredibly powerful and accurate, however it requires a large number of data to train. Therefore it is necessary to pre-train the SSD model on a large dataset before you train it on your own data. Training with a high-end GPU took roughly 6-8 hours. The images used to train the model were 370 x 280 pixels and Region of Interest information was assigned to each image depending on where the component was located. To acquire the images for the dataset, the component to be detected was photographed in various lighting conditions and angles (Fig 1). Roughly 300 images were captured to train the model. A further 30 were captured for testing.

Fig 1 Sample of dataset.

2.2 Keras Model

The second model is a simpler cNN developed using Keras deep learning library. The model was trained on a very small number of samples (images) in comparison to the SSD. It does not require pre-training on a larger dataset. The images for the dataset were resized to 64 x 64 pixels. The model was trained with 2 classes – one where the object to be detected is present, and another where it is not. Training takes roughly 30 seconds with a high-end GPU.

3. Evaluation & Results

The two models trained by the CNNs were evaluated by attempting to detect the component in various settings. This included various lighting, objects placed beside/ on top of the component, hands in place with the component, component placed on different surfaces which varied in colour and matte/shiny surface, and shadows were also applied. The models were tested on 50 different instances and the result was recorded whether the classifier was right ("Yes") or wrong ("No") in a table (Table 1). The table was then converted into a 2x2 contingency table (Table 2) so that the Null Hypothesis 0 (H0) could be tested. McNemars Test [6] states that H0 should show that both classifiers have the same error rates and are therefore as effective as each other.

Table 1 Results.

Instance	SSD CNN Correct?	Keras CNN Correct?
1	Yes	Yes
2	Yes	Yes
3	Yes	Yes
4	Yes	No
5	Yes	Yes
6	Yes	Yes
7	Yes	Yes
8	Yes	No
9	Yes	Yes
10	Yes	Yes
..N..		
50	Yes	Yes

Table 2 Contingency Table.

	Keras Correct	Keras Incorrect
SSD Correct	36	14
SSD Incorrect	0	0

Statistic = 0.000, p-value = 0.000

Different proportions of errors (reject H0)

The results obtained by the team demonstrated a clear difference in error rate between the two models. The SSD proved to be extremely accurate and effective, detecting the object at all times under every tested circumstance. The Keras CNN yielded a much higher error rate and was more easily fooled by varying environment configurations.

4. Conclusion

It is clear from the results that the SSD model is much stronger and more accurate than the Keras CNN. It was found that it did not take much variance in the environment setup to confuse the Keras-based classifier and yield the wrong results. The SSD could not be fooled into yielding the wrong result, successfully detecting the component under all settings, including detecting the component through a transparent Perspex ruler overlaying the component (Fig 2).

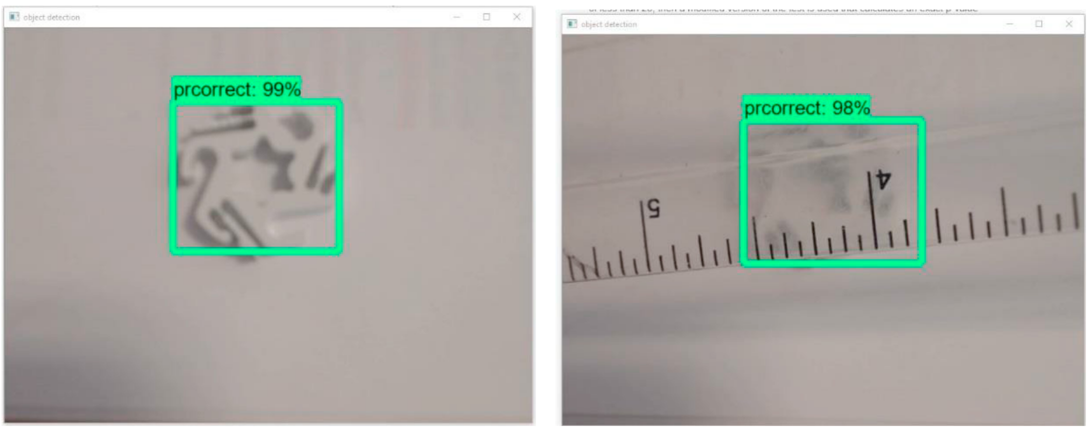

Fig 2 Classification.

That being said, the simpler Keras CNN was found to classify correctly when in a stable environment as close to the training data as possible, making the classifier quite effective given the right circumstances. The most significant disadvantage of the SSD model is the training time and processing power required to run it. This is in itself quite costly and so makes the CNN less desirable for real-world use. The fact that the simple Keras CNN can operate almost as accurately as the SSD given the right circumstances, and requires only a fraction of the training data, time and processing power gives it great advantage over the more accurate model. There is potential for this type of model to be used to drive systems that do not require as much accuracy as others, and to substitute in a stronger model such as the SSD when accuracy is of more importance. This allows such systems to gain the benefits of higher processing models, but with less training and processing time.

There are many risks involved in vision systems being introduced into the medical technology sector due to high regulations surrounding vision systems. Deep Learning has many advantages over traditional Image Processing, a more common approach in vision systems. The results from this study indicate that a very effective, high-powered CNN would yield very high accuracy in a medical technology plant's processing line and there is opportunity for further research into this area and how it can be best used.

5. Acknowledgments

This research is a project undertaken in Limerick Institute of Technology Research Development & Innovation section, and the authors acknowledge the support of the team, Liam Brown, Patrick Murray, Jenna Barry, Daragh Naughton, Ryan Sheehan, and Shauna Leahy.

The research work is co-funded by Enterprise Ireland, Cook Medical and Vistamed without which the research is not possible, and we acknowledge the support and feedback of the same.

6. References

[1] M. Leo, G. Medioni, M. Trivedi, T. Kanade, G.M. Farinella, Computer vision for assistive technologies, Computer Vision and Image understanding 154 (2017) 1-15.

[2] J. Wang, Y. Ma, L. Zhang, R.X. Gao, D. Wu, Deep Learning for smart manufacturing: Methods and applications, Journal of Manufacturing Systems 48(2018) 144-156.

[3] A.R. Pathak, M. Pandey, S. Rautaray, Application of Deep Learning for Object Detection, Procedia Computer Science 132(2018) 1706-1717.

[4] A. Borji, Negative results in computer vision: A perspective, Image and Vision Computing 69(2018) 1-8.

[5] COCO dataset, http://cocodataset.org/#home.

[6] T.G. Dietterich, Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms, Neural Computation Vol 10 Issue 7 1998 p.1895-1923.