Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 93 (2020) 1224–1229

www.elsevier.com/locate/procedia

53rd CIRP Conference on Manufacturing Systems

Image Processing based on Deep Neural Networks for Detecting Quality Problems in Paper Bag Production

Anna Syberfeldt*[a] and Fredrik Vuoluterä [a]

[a] University of Skövde, PO 408 SE-541 48 Skövde, Sweden

* Corresponding author. Tel.: +46(0)500448577. E-mail address: anna.syberfeldt@his.se

Abstract

It is critical for manufacturers to identify quality issues in production and prevent defective products being delivered to customers. We investigate the use of deep neural networks to perform automatic quality inspections based on image processing to eliminate the current manual inspection. A deep neural network was implemented in a real-world industrial case study, and its ability to detect quality problems was evaluated and analyzed. The results show that the network has an accuracy of 94.5%, which is considered good in comparison to the 70–80% accuracy of a trained human inspector.

Keywords: Deep Neural Networks; Image Processing; Quality Inspection; Industrial Vision Systems

1. Introduction

Detection and elimination of poor quality products is of critical importance to virtually all manufacturing companies [1]. They employ specialized personnel to undertake quality inspections in which one or more product characteristics are checked against product specifications. Products that do not meet the specifications are rejected or returned for improvement to avoid releasing poor quality products [2]. These inspections are vital to ensuring high-quality products, but performing them manually comes at a high cost. It has been suggested that automatic camera-based systems, so-called "vision systems," could replace manual quality inspections [3-4]. However, camera-based systems have not been widely adopted in manufacturing, mainly because of their poor performance and low flexibility when the environment changes. For example, a slight change in lighting conditions may cause a camera system to stop working [5]. The hardware needed is also expensive, and so is the cost of installing and maintaining the system [6].

In recent years, new techniques for image analysis based on artificial intelligence (AI) have been suggested. These techniques have been shown to be capable of overcoming the shortcomings of the traditional techniques used in industrial vision systems [7]. AI-based image processing is currently a hot research topic. The most successful implementations use so-called "deep learning", which is a machine learning method based on artificial neural networks [8]. Despite the great potential of deep neural networks, there are still only a few implementations of them in vision systems targeting the manufacturing industry, mainly because the technique and the available software platforms are still relatively new.

In this paper, we describe the implementation and testing of the technique of deep learning in real-world industrial production and evaluate and analyze its ability to detect quality problems. The company involved in the study was Jonsac AB, a Swedish manufacturer of paper bags that are used for packaging a variety of products including dry foods, waste, and animal feed. Their customers demand high-quality bags that will store their products in the best possible conditions. The

10.1016/j.procir.2020.04.158

bags also have to meet various production regulations relating to such things as food hygiene. The current manufacturing process maintains a high quality, but the occasional defective product is produced. When defects do occur, they are in the bottom of the bag and are caused by erroneous folding or gluing. To prevent any of the defective bags being delivered to a customer, every bag is inspected by an operator at an inspection station at the end of the production line where the bags are visually examined for any defects. This quality control requires the full attention of an employee when the production line is active. This inspection process is both costly and inadequate, for mistakes happen easily. The pace of production is such that one sack or bag is produced approximately every second, which means that even momentary distraction of the operator can result in a defective bag being missed. The company needs to find way to eliminate, or at least minimize, the manual quality inspection and is very motivated to try automatic inspection based on state-of-the-art AI.

In the next section of the paper, the concepts of deep learning and the specific type of deep neural network used in this study are described. Section 3 presents similar studies using deep learning techniques to detect production quality problems. Information about the data set used for training the deep neural network is provided in Section 4. Section 5 presents the implementation of the deep neural network, along with an evaluation. Finally, in Section 6 the results are discussed and future work is outlined.

2. Deep neural networks

A deep neural network is basically an artificial neural network with multiple layers between the input and output layers [9]. It transforms input into an output by mathematical manipulations based on linear or non-linear relationships. Each mathematical manipulation is considered a layer, and the term "deep" refers to the many layers in the network [9]. Currently the most successful deep neural networks for image processing use what is known as a convolutional neural network (CNN) architecture [10]. Interest in CNNs spiked in 2012 after a network called AlexNet solved a given problem with an error rate of 15.3%. In 2015 ResNet achieved an error rate of 3.6%, which surpassed the recognition ability of a human facing the same problem [11].

The name “CNN” comes from the fact that the network employs a mathematical operation called convolution, which is a specialized kind of linear operation. Basically, CNNs are neural networks that use convolution instead of general matrix multiplication in at least one of the layers. Instead of using fully connected layers where one neuron takes 1D data as input as in a traditional neural network, CNN uses convolutions that take 2D data [10]. This technique preserves the spatial information in an image, which is crucial for accurate predictions. Most CNN architectures consist of convolutions, rectified linear units (ReLUs) and pooling layers for feature detection, followed by fully connected layers and SoftMax for classification. This is the architecture we chose for the project. The concept of a CNN is presented in Figure 1. Further information about the technique can be found in [12].

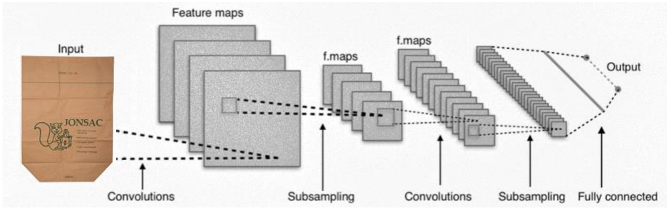


Fig. 1. Concept of a convolutional neural network.

In this study, we use a variant of CNN called Faster Region with CNN (R-CNN). Faster R-CNN further improves the processing pipeline by embedding a region proposal network after the CNN. It performs better than a traditional CNN when it comes to object detection and generally has a faster training process [10]. Faster R-CNN improves the concept of CNN by removing cropping/wrapping and introducing spatial pyramid pooling instead. Using region of interest (ROI) pooling further improves the performance as a Faster R-CNN performs the ROI pooling on the feature map instead of on the original image. The significantly lower resolution of the feature map decreases the required computation and allows Faster R-CNN to be much quicker and less resource intensive than competing methods [11].

To implement a Faster R-CNN in this study, the TensorFlow state-of-the-art machine learning library from Google is used (www.tensorflow.org). TensorFlow is extensively used today by many researchers and is considered to be highly competitive [13].

3. Related work

There are no previous studies on detecting defects in paper bag production using the same techniques as this study. Accordingly, this section of the paper focuses on the three most relevant studies using different variants of CNNs for detecting defective products.

Zhu et al. [14] present a study on detecting defects in the production of emulsion pumps. They evaluate the possibilities of using CNNs to replace the current manual inspection of the pumps. The main challenge in their implementation was that the limited number of images of defective products, making high-quality image samples very important. To increase the quality of the images, pre-treatment using slant correction was used. When validated on images not previously seen, the CNN achieved a 97% accuracy rate with a mean detection time of 0.18 seconds.

Jing et al. [15] present a study from the textile industry focused on finding defects in fabric. Like the study of Zhu et al. the purpose was to replace manual inspections with a CNN. The authors used a dataset of different fabrics with a range of colors and recurring patterns, and the CNN was trained to recognize six categories of common defects. The network utilized an automatic calculation of the patch sizes of the fabric, which improved accuracy and made it possible to identify very small defects. The CNN showed an average accuracy of over 97%.

A study on defect detection in nanofibrous materials presented by Napolentano et al [16] achieved similar results to Jing et al. Using scanning electron microscope (SEM) images, a regional CNN approach was used to discover defects in the materials. The accuracy achieved was also about 97%. The authors note that the accuracy was higher for coarse-grained defects but lower for fine-grain defects.

4. Training data set

The paper bag product line at Jonsac is subject to frequent changes. The company delivers products to a vast assortment of customers in very different quantities and at different intervals. Each customer has the option of having a custom print, and can also choose among an array of colors for each area of the bag. There are also structural differences in the bags, such as length. As in any market, Jonsac may gain new customers or lose old ones, and established customers may change their preferences. These factors make it unfeasible, if not impossible, to gather data on every single product variant for training the deep neural network. Even if it were possible to collect enough images of all the current product series, as soon as a new product series was introduced the deep neural network would need new data. It is also doubtful that a single network would be able to distinguish between faulty print/color in variant X and the correct print/color in variant Y. Accordingly, this study focused on the geometry of the bags, which is almost identical between variants except for length. Thus, the network ignores faults in print, coloring, and other purely aesthetic defects.

A dataset of 1,729 images was collected for the study, including six different product variants as shown in Table 1. Each variant is clearly distinguishable from the others by some factor, be it color or print. They share the same geometry and "folding lines", and are identical from a structural point of view, except for variants 5 and 6 which are slightly longer than variants 1–4. As previously mentioned, only the bottom of the bag needs to be checked for quality; thus only the bottom part of the bag is used for the training data.

To properly train the network, all the images in the dataset need to be classified as "OK" (the product has no flaws) or "NOK" (the product is of inferior quality). More precisely, NOK denotes an anomaly, a divergence from the normal geometry. The images were classified as "OK" or "NOK" manually, as each had to be evaluated by a human who could determine the NOK or OK assignment from experience. The images were also annotated to indicate the region of interest and preferable traits. Figure 2 shows an example of an annotation.

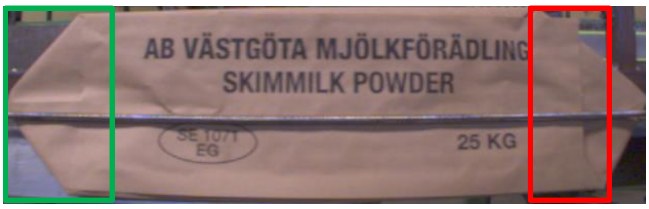
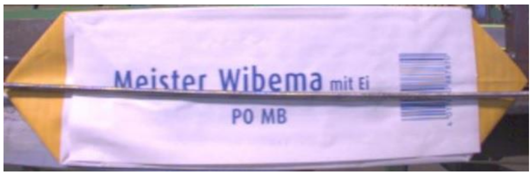
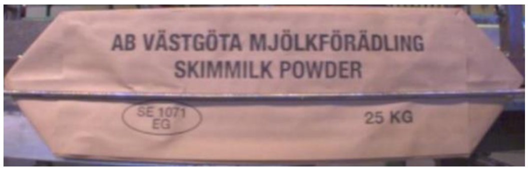


Fig. 2. Example of annotation (green box = OK, red box = NOK).

From the classification process, it became clear that there were five different categories of anomalies: skewed side (right), skewed side (left), crushed side, tearing, and offset bottom. Examples of products exhibiting these anomalies are shown in Table 2 (anomalies marked with yellow circles).

Table 1. Product variants included in the study.

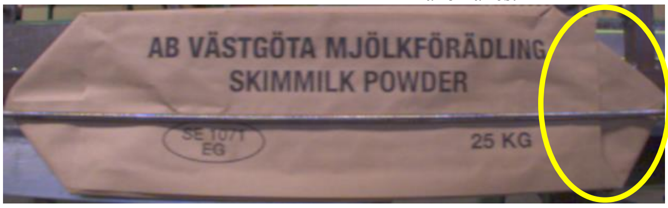
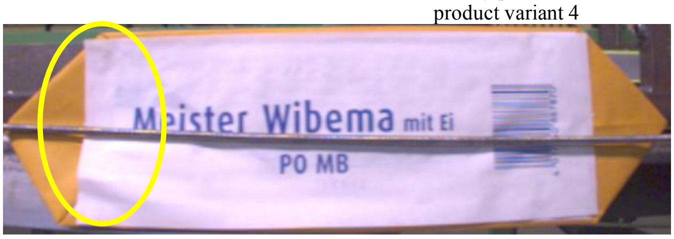
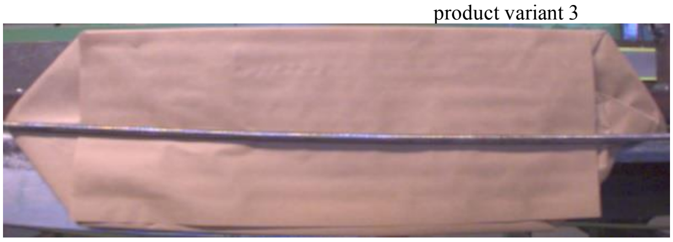
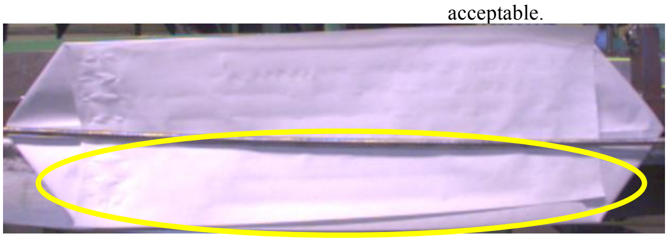
Variant number	Product
1	
2	
3	
4	
5	
6	

4. Training and evluation

This study used a pre-trained network with transfer learning based on the publicly available model "faster_rcnn_inception_v2_coco_2018_01_28". The specification for the model can be found at https://github.com/opencv/open_model_zoo/blob/master/models/public/faster_rcnn_inception_v2_coco/faster_rcnn_inception_v2_coco.md.

There are six prediction layers, decreasing in feature-map size: 38^2, 19^2, 10^2, 5^2, 3^2, and 1^2, with the following numbers of default boxes: 4, 6, 6, 6, 4, and 4. Thus there are 8732 detections per class ($38^2\times4+19^2\times6+10^2\times6+5^2\times6+3^2\times4+1^2\times4$). However, many of these are unlikely candidates with very low probability, and so they are removed using non-maximum suppression (NMS).

Table 2: Anomalies present in the training data dataset.

Description	Common	Comment
Skewed side (right)	Yes	Represents most of all anomalies.
		
Skewed side (left)	No	Mostly present for product variant 4
		
Crushed side	No	Mostly present for product variant 3
		
Tearing	No	Very easy to identify
		
Offset bottom	No	A small offset is acceptable.
		

Before starting the training of the network, 20% of the annotated images from the training data set were set aside as a test set. This evaluation set was used once the network had been trained in order to test the network on situations it has not been exposed to during training. To add another layer of validation to the training process, 50% of all NOK and OK images included in the training data were left unannotated. Using the AI classify function, the software automatically generated annotations for these images. These annotations were clearly marked as automatically generated and could therefore be inspected manually. The classification results aided in modifying the training data to better fit the products and find erroneous behaviors, often caused by mistaken manual annotations. Figure 3 shows the procedure used.

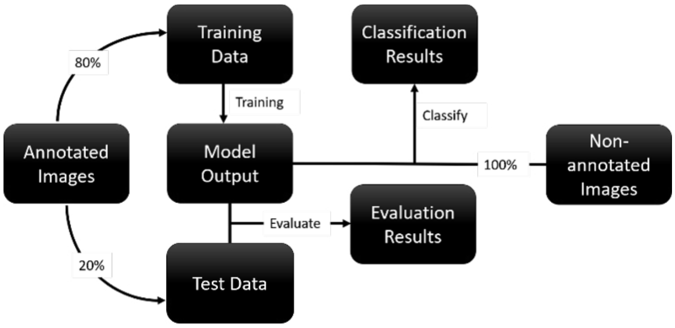


Fig. 3. Flowchart showing the process of training and evaluating the network.

The network was then trained with the following parameter settings:

- Training steps: 60,000
- Learning rate: 0.0002 up to 30,000 training steps, 0.00002 between 30000–40000 training steps, and 0.000002 for more than 40 000 training steps
- Dropout: 60%
- No regularization
- Aspect ratio of pictures kept intact

After the training, the network was tested on the evaluation data set. Results from the evaluation are shown in Table 3 below.

Table 3. Results from evaluating the network.

Name	Result
Total training data	1,729
Total OKs	858
Total NOKs	871
Total evaluation data	346
Accuracy	94.51%
Miss rate (no annotation)	2.89%
False OK portion	1.73%
False NOK portion	0.87%
Total duplicate annotations	27
Duplicate portion	7.80%

Duplicate annotations refer to images for which the network provided inconsistent classifications, as shown in the example in Figure 4 below. The duplicate annotations fall into two categories: those that reaffirmed the results, and those that contradicted the results (as in Figure 4). While the occurrence of duplicate annotations is concerning, reaffirmation is not catastrophic as the action taken would be the same regardless of the number of annotations. Contradictory duplicate annotations are much more of a problem as there is no clear principle for which annotation to trust.

Missed annotations also occur, but it should be noted that missing an annotation may be better than providing an inaccurate one. This argument is strengthened by the fact that all the missed images had an OK classification, which is an indication that there would be little to no risk of defective products being approved. Should a missed product be regarded as OK, the evaluation would have generated an accuracy of 97.40%.

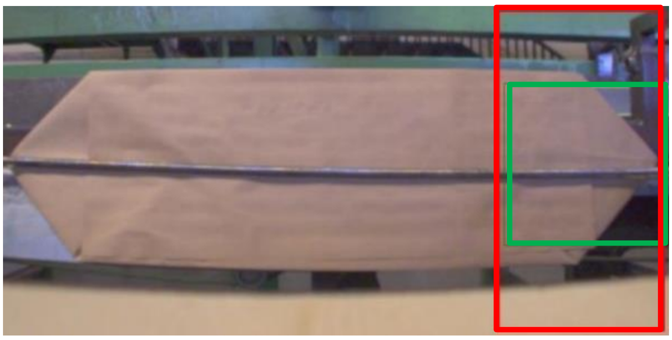


Fig. 4. Duplicate annotations made by the network. The human inspector has classified this bag as OK, but being a border case (a minimal skewness in the bottom right corner).

5. Conclusions and future work

This paper investigated the use of deep neural networks for performing automatic quality inspections based on image processing to eliminate the current manual inspection process. The focus of this study is a real-world industrial case study of paper bag production using a Faster R-CNN to detect defective bags. It is difficult for even a trained human to detect defective bags as the defects may be small and can vary considerably. An evaluation of the deep neural network shows, however, that it has an accuracy of 94.5%. This is at least as good as the accuracy of the operator currently undertaking the quality inspection manually. Follow-ups from the manual inspection that has been undertaken by the company show that the accuracy of a human inspector usually lies between 70–80% after a period of training, which is not surprising given the high rate of production.

The evaluations performed by the network. network were not only accurate; they were also rapid. This is important given the high rate of production. With 1–2 bags being produced per second, there is at most 0.5 seconds to take a picture, evaluate the bag quality, and notify the production system that a defective product has been detected. Evaluation of the implementation shows that the network can solve the problem in 0.3 seconds, meaning that it can be used in real production. In fact the performance of the network was considered so promising that the company has decided to introduce it into their production line.

Future work could focus on improving the solution so that it can assess not only the geometry of bags but also faults in print, coloring, and other purely aesthetic defects that are important to customers even though such flaws do not affect the function of the bag. The personnel at the company did indicate that issues with print and color are somewhat predictable, often happening when refilling printing materials or switching between product variants. Although it might not be critical to automatically detect aesthetic defects, doing so would reduce the burden on the operators in the line. It is thus worth investigating adding this capability as an extra feature of the network.

Another aspect that could be investigated is ways to make the training process for the network more efficient with respect to setting up the training data. It is time-consuming to check each image individually, classify it as OK or NOK, and annotate it. In this study we manually inspected 1,729 images for the training data set. On average, each image took about one minute to process. This means that we spent about 30 hours configuring the training data set (not including the time it took to take the pictures). Our training set was not particularly large, and considerably more images might be needed for solving other problems. In such cases, the time required to set up the training data becomes a real problem. One option could involve generating training data virtually. For example, CAD models of the products with superimposed defects could be used to generate training data. If it at least some virtual training data could be used, the time for setting up the training data set could be significantly reduced. We will therefore start investigating this, as we believe that many applications could benefit from such an approach if it can be proven to work.

Acknowledgments

The authors would like to thank Jonsac AB for their support in the study and for allowing us to work in their facility. The authors also want to thank Vinnova for financing the VISION project through the strategic innovation program Produktion2030, within which this work has been undertaken.

References

[1] Colledani, M., Tolio, T., Fischer, A., Iung, B. Lanza, G., Schmitt, R. & Váncza, J. Design and management of manufacturing systems for production quality. CIRP Annals 2014;63(2), 773-796.
[2] R.K. Rajput. A Textbook of Manufacturing Technology (Manufacturing Processes). ISBN 978-81-318-0244-1;2018.
[3] Forsyth, D., & Ponce, J. Computer Vision: A Modern Approach (2nd ed.). New Jersey, USA: Pearson;2012.
[4] Soini, A. Machine vision technology take-up in industrial applications. Proceedings of the 2nd International Symposium on Image and Signal Processing and Analysis (ISPA 2001);2001.
[5] Semeniuta, O., Dransfeld, S., Martinsen, K. & Falkman, P. Towards increased intelligence and automatic improvement in industrial vision systems. Procedia CIRP 2018; 67, 256-261.
[6] Davies, R. E. Machine Vision: Theory, Algorithms, Practicalities. 4th ed. London: Elsevier; 2012.
[7] Silva R.L., Rudek M., Szejka A.L., Junior O.C. Machine Vision Systems for Industrial Quality Control Inspections. In: Chiabert P., Bouras A., Noël F., Ríos J. (eds) Product Lifecycle Management to Support Industry 4.0. PLM 2018. IFIP Advances in Information and Communication Technology. 540. Springer, Cham;2018.
[8] Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y. & Alsaadi, F. A survey of deep neural network architectures and their applications. Neurocomputing 2017;234, 11-26.
[9] Goodfellow, I., Bengio, Y. & Courville, A. Deep Learning. First ed. Cambridge: MIT press;2016.
[10] Bosse, S., Dominique, M., Müller, K-R., Wiegand, T. & Wojciech, S. Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment. IEEE Transactions on Image Processing. 2018;27. 206-219.
[11] Du, J. Understanding of object detection based on CNN family and YOLO. Journal of Physics: Conference Series. 1004;2018.
[12] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J. and Chen, T. Recent advances in convolutional neural networks. Pattern Recognition. 2018;77. 354-377.
[13] Géron, A. Hands-on Machine Learning with Scikit-Learn and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems (2nd ed.). Sebastopol, CA: O'Reilly Media. ISBN: 978-1491962299;2019.

[14] Zhu, C., Zhou, W., Yu, H. & Xiao, S. Defect Detection of Emulsion Pump Body Based on Improved Convolutional Neural Network. Proceedings of the 2019 International Conference on Advanced Mechatronic Systems. Kusatsu, Shiga, Japan 26-28 August 2019, 2019;349-352.

[15] Jing, J-F., Ma, H. & Zhang, H-H. Automatic fabric defect detection using a deep convolutional neural network. Coloration Technology. 2019;135(3). 213-223.

[16] Napolentano, P., Piccoli, F. & Schettini, R. Anomaly Detection in Nanofibrous Materials by CNN-Based Self-Similarity. Sensors 2018;18(1).