2020 5th International Conference on Computer and Communication Systems

# Packaging Defect Detection System Based on Machine Vision and Deep Learning

Jiming Sa
School of Information Engineering
Wuhan University of Technology
Wuhan, China
e-mail: jimingsa@whut.edu.cn

Zhihao Li*
School of Information Engineering
Wuhan University of Technology
Wuhan, China
e-mail: lzhgd@whut.edu.cn

Qijun Yang
School of Information Engineering
Wuhan University of Technology
Wuhan, China
e-mail: yangqijun@whut.edu.cn

Xuan Chen
School of Information Engineering
Wuhan University of Technology
Wuhan, China
e-mail: 275663@whut.edu.cn

***Abstract*—Detecting packaging defection with high accuracy and efficiency is of great significance in product quality. We use OpenCV to preprocess images which come from damaged package according to characteristics of the image. The processed data is combined with deep learning and based on neural network model ResNet. Meanwhile the processed image data is sent to a convolutional neural network (CNN) for model training. We establish a detection system for product packaging. The detection system provides a solution for automatic detection of package defection, which realizes rapid and accurate detection of product packaging.**

***Keywords-packaging defection detection; image preprocessing; convolution neural network; TensorRT; ResNet; Keras***

## I. INTRODUCTION

Packaging is the most direct and effective advertising carrier for a brand, and is the consumer's first impression of the brand. In food, medicine, cosmetics and other industries, the impact of product packaging on product quality and sales is particularly obvious. Packaging detection has already become an indispensable and important link in the production process. Because the detection requires repeating and precise work, manual inspection adds huge labor and management cost to the factory compared with machine inspection. Also, manual inspection is difficult to ensure accuracy and standardization, and cannot obtain satisfactory inspection results. Therefore, it is significant to develop a fast, accurate, effective and intelligent packaging inspection system.

With the continuous development of image recognition technology, it is widely used in various fields. Medical field is one of the most widely applied fields of image recognition technology. Fan Wenxia et al.[1] used image processing methods to identify unqualified packages with only 10% percison and finally achieve 100% accuracy.Zhang Ming et al.[2] employed optical fiber sensors to detect of the paper package with flying bag problem and obtained 100% accuracy. Qi Ruizheng[3] based on machine vision to identify cigarette product packaging with accuracy better than 95%. For the conventional rule-based machine vision system mentioned above, the system can perform fast and reliable detection of fixed-shape product packaging. Because multiple variables (lighting, color changes, and curvature) are hard to separate, some defect detection is very difficult to program and solve by traditional machine vision systems. Detecting irregular product packaging with complex appearance changes is a serious challenge.

Convolutional neural networks (CNN) can perform image recognition without mage specific features and templates. For complex inspections involving extensive deviations and unpredictable defects, deep learning is the most appropriate method. This paper applies machine vision to conduct image recognition processing on product packaging. Meanwhile, this paper is based on convolutional neural networks to deal with the interference of defect detection on background and packaging appearance. Through the combination of machine vision and deep learning, the compatibility problem of packaging defect detection in complex backgrounds can be solved as well. It provides a reference for the identification of packaging defection.

## II. METHODS

### *A. Image Segmentation Based on Machine Vision*

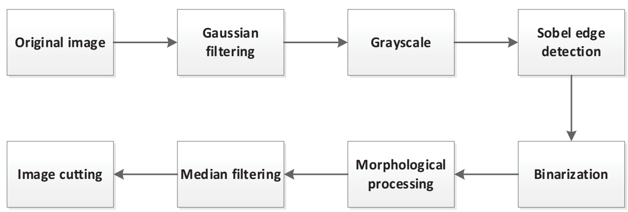


Figure 1. Flowchart of drug interception operation.

Properly preprocessing the images collected by the camera can effectively reduce the complexity of the images to be detected, reduce the difficulty of the machine learning

978-1-7281-6136-5/20/$31.00 ©2020 IEEE

training and improve the accuracy of recognition. This experiment applies OpenCV to intercept the image and adjust the brightness and size. Figure 1 displays the process of interception of the drug to be tested:

In the Sobel edge detection step, the Sobel operator is used to derive the image, which can effectively extract the outline of the package. Assuming the image is 1, the image is derivative in horizontal direction and vertical direction.

Derivative of horizontal direction can be expressed as:

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * I \quad . \tag{1}$$

Vertical of horizontal direction can be expressed as:

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * I \tag{2}$$

For each point on the image, combine the above results to find the approximate gradient:

$$G = \sqrt{G_x^2 + G_y^2} \tag{3}$$

To accelerate the calculating speed, Equation (3) can be replaced by:

$$G = |G_x| + |G_y| \tag{4}$$

Figure 2-4 display some key steps in image processing:

Figure 2. Original image.

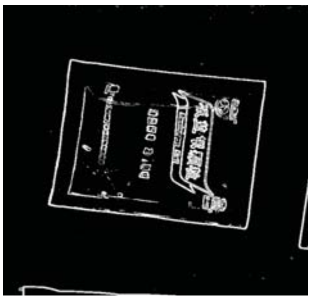

Figure 3. Binary image.

Figure 4. Intercepted image.

## B. Convolutional Neural Network

Convolutional neural network is a neural network specially used to process data with similar grid structure, such as image data in this experiment. Compared to a fully connected network, CNN introduces a convolution layer structure and a pooling layer structure. Its basic structure of it is shown in Figure 5. The convolution layer completes the extraction of image features through the convolution operation. By increasing the number of convolution kernels and convolution layers, more comprehensive features can be extracted. The pooling layer mainly reduces the dimensionality of the image obtained by the convolution layer to simplify network parameters and complexity.

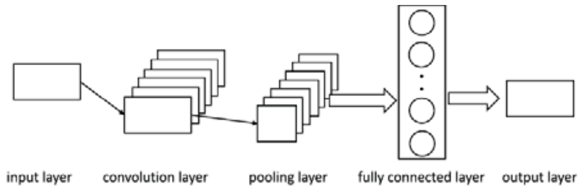


Figure 5. Basic structure of CNN.

# III. System Design

## A. Hardware Composition

In order to realize the functions of acquisition, identification, and transportation, the designed system consists of a camera, an embedded device, and a conveyor belt. The hardware composition is shown in Figure 6.

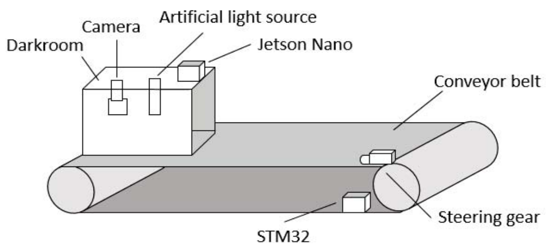


Figure 6. Hardware composition.

The model of camera is IMX219, and the output pixel can be up to 8 million. In order to reduce the external interference of the captured image, the system uses dark room and supplemental light to collect the image. Embedded devices include NVIDIA's Jetson Nano and ST's STM32.

Among them, Nano is used to provide a system interactive interface, perform image processing and run a network model to make decisions. STM32 carries out the control of the conveyor belt. The conveyor belt employs a DC geared motor as the power, which can be programmed to control the speed and direction. At the end of the conveyor belt, a steering gear is installed for product classification.

### B. *Sample Acquisition*

The object of this research is product packaging. Due to the diversity of product packaging, representative products need to be selected as the specific object in this experiment. This experiment chooses to test the granules because its bag-type packaging is universal and the requirements for packaging of drugs are relatively stricter.

The front and back of the package are collected as samples, and the actual number of sample is 800. In order to obtain better training results, the number of samples in different categories should be equal, including 200 sheets on the front and back of the right packaging and 200 sheets on the front and back of the damaged packaging (the damaged packaging is obtained by manual destruction). In order to increase the generalization of the model, the collected samples are subjected to transformation operations such as offset, rotation, and scaling. And the sample set is eventually expanded to the number of 8,000. In addition, another 200 untransformed images are collected as a validation set. The samples are divided into two grades according to the intactness of the packaging: intact and damaged. The samples are labeled with 0 or 1 according to the grade. Some samples are shown in Figure 7-8:


Figure 7. Broken font.

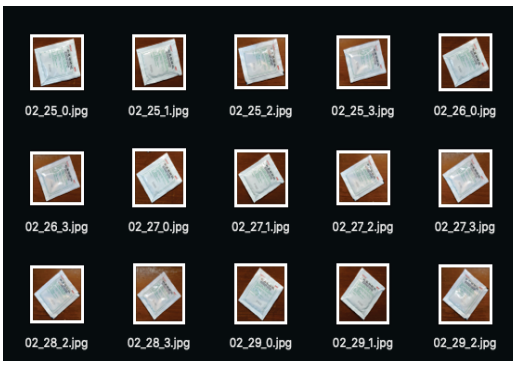


Figure 8. Intact back.

### C. *Convolutional Network Design*

CNN has a variety of typical models. For the same data set, the training results obtained by different models vary greatly. It is generally recognized that increasing the depth of the network can improve the performance of the network to a certain extent. But it will also increase the difficulty of training the network. Because ResNet introduces a residual learning unit, as shown in Figure 9. It can maintain a small amount of parameters while the network depth is greatly increased, the effect has a more efficient effect than other models[5].

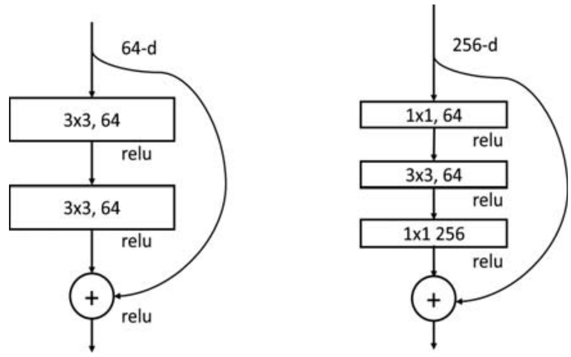


Figure 9. Residual learning unit

Because the surface of the test object has lots of text, and the general damage is small, it is difficult to distinguish and it belongs to fine-grained classification. Therefore, a deeper network should be selected. In this experiment, ResNet50 is selected as the basic network model. The neuron activation function is ReLU to achieve the effect of de-linearization. The optimization method is set to Stochastic Gradient Descent. The classifier uses Softmax. For the output of the neural network $y_1, y_2, y_3 \cdots y_n$, the output after Softmax regression processing is:

$$softmax = y_i' \frac{e^{y_i}}{\Sigma_{j=1}^{n} e^{y_i}} \quad (5)$$

In the Equation (5), $y_i$ represents the value of each unit before input to Softmax.

After the Softmax layer, the number of units does not change, but the values become a probability distribution. By selecting the unit with the highest probability, the classification task can be completed.

## IV. Experimental and Results Discussion

### A. *Analysis of Model Training Result*

In this experiment, we used the development platform Colab provided by Google for model training. We set the number of training epoch to 80, and record the accuracy of each epoch. The validation and training accuracies during training are shown in Figure 10:

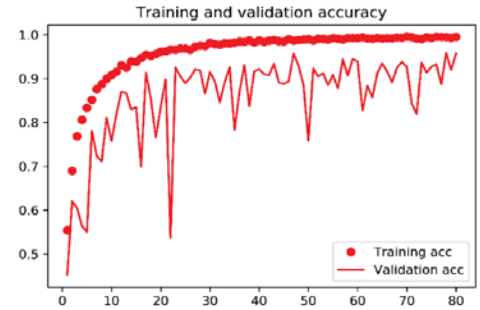


Figure 10. The accuracy results of the training.

As shown in Figure 10, the accuracy of the training increases with the increasing number of training epoch, and remains above 98% eventually. The accuracy of the validation set fluctuates greatly, but the overall trend continues to gradually increase. When the number of training rounds is 49, 95.2% accuracy has achieved.

### B. *Analysis of Model Actual Performance*

We used the tools provided by Nvidia to optimize the model. And the TensorRT engine was used to accelerate the model. The environment of model deployment is shown in Table 1.

TABLE I. ENVIRONMENT CHANGE

| Software | Version |
|---|---|
| operating system | Ubuntu 18.04 bionic |
| OpenCV | 3.31 |
| Tensorflow-gpu | 1.14.0+nv19.7 |
| Keras | 2.3.1 |
| TensorRT | 5.1.6.1-1+cuda10.0 |

We tested its performance with 100 images and recorded data of each image. The time cost on each image is shown in the Figure 11.

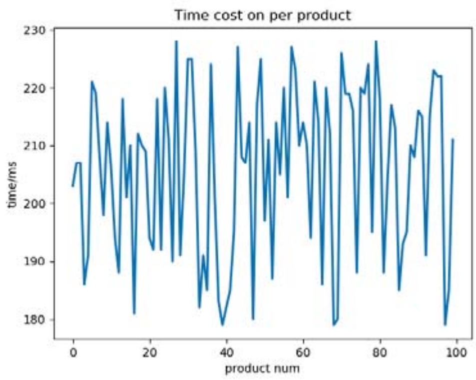


Figure 11. The curve of ideal results.

It takes the Jetson Nano about 200ms to finish each job, from image acquisition and contour extraction to model inference. The detailed result of each image is shown in Table 2.

TABLE II. RESULT OF EACH IMAGE

| ID | Label | Time/ms | Good | Broken | Result |
|---|---|---|---|---|---|
| 1 | Broken | 194.96 | 0.00014 | 0.999859 | Broken |
| 2 | Broken | 194.99 | 0.00000 | 1.000000 | Broken |
| 3 | Good | 195.31 | 1.00000 | 0.000000 | Good |
| 4 | Broken | 188.67 | 0.00098 | 0.999021 | Broken |
| 5 | Good | 200.51 | 1.00000 | 0.000000 | Good |
| 6 | Good | 199.96 | 1.00000 | 0.000000 | Good |
| 7 | Broken | 206.32 | 0.00000 | 1.000000 | Broken |
| 8 | Good | 197.85 | 0.99484 | 0.005155 | Good |
| 9 | Good | 199.41 | 0.13674 | 0.863264 | Broken |
| 10 | Broken | 188.66 | 0.00098 | 0.999021 | Broken |

The Label marks the actual result of each image, and the Result is the Jetson Nano's output of each image. The probabilities of good and bad of each image are also shown in the table. It can be seen that the probability of each correct judgment is very high. But there is also a misjudgment in this table.

## V. CONCLUSION

This research is based on machine vision to segment and extract features of damaged packaging images. The system keeps the packaging inspection from manual labor and provides a high-accuracy and high-efficiency automatic detection method. This provides a fast and effective solution for packaging product inspection. In addition, this system combines a deep learning framework with high portability. Model training can be applied to most of the product packaging inspections and can be widely used in the production process. It is suitable for packaging inspection of food, medicine, cosmetics and etc., which has broad market prospects.

However, due to the double-sided shooting method used in this research, the acquired image data is not sufficient to detect geometrically multi-sided product packaging, and can only be applied to double-sided packaging, such as granules and aerated puffed food. Therefore, the mechanical structure should be improved and a multi-camera device should be adopted to make it able to acquire multi-faceted images in the future detection system.

## ACKNOWLEDGMENT

This work is supported by National College Students Innovation and entrepreneurship training program in Wuhan University of Technology (Project No:S201910497049)

## REFERENCES

[1] Fan Wenxia, Zhang Min, Wang Xinming. Design of an Online Inspection System for the Quality of Pharmaceutical Packaging Based on Image Processing [J] .Journal of Xi'an University of Engineering Science and Technology, 2004,18 (2): 160-163 + 178.

[2] Zhang Ming, Sun Jianyun, Liang Guanghe. Design of flight package detection for packaging unit based on GD language [J] .Industrial Control Computer, 2019,32 (2): 78-79.

[3] Qi Ruihuan. Research on product packaging inspection system based on machine vision [D]. Dalian University of Technology, 2007.

[4] He, K. , Zhang, X. , Ren, S. , and Sun, J. . (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society.

[5] Atlason Hans E,Love Askell,Sigurdsson Sigurdur,Gudnason Vilmundur,Ellingsen Lotta M. SegAE: Unsupervised white matter lesion segmentation from brain MRIs using a CNN autoencoder.[J]. NeuroImage. Clinical,2019,24. Article in a conference proceedings:

[6] Ghosh Arna,Dal Maso Fabien,Roig Marc,Mitsis Georgios D,Boudrias Marie-Hélène. Unfolding the Effects of Acute Cardiovascular Exercise on Neural Correlates of Motor Learning Using Convolutional Neural Networks.[J]. Frontiers in neuroscience,2019,13.

[7] Kumarasinghe Kaushalya,Kasabov Nikola,Taylor Denise. Deep learning and deep knowledge representation in Spiking Neural Networks for Brain-Computer Interfaces.[J]. Neural networks : the official journal of the International Neural Network Society,2020,