Region-Convolutional Neural Network for Detecting Capsule Surface Defects

Rixian Liu
College of Information Engineering, Zhejiang University of Technology, Hangzhou 310014, China
College of Information and Engineering, Jinhua Polytechnic, Jinhua 321007, China

Qinlong Gu, Xianbao Wang, Minghai Yao*
College of Information Engineering, Zhejiang University of Technology, Hangzhou 310014, China
*E-mail:ymh@zjut.edu.cn

Abstract
Learning the defect features with shift and rotation invariance is the main problem in the detection of capsule surface defects. This paper proposes a method based on region-convolutional neural network for detecting capsule surface defects. First, we obtain the capsule’s region proposals by selective search and then its features via a deep convolution network. Finally, we detect and recognize the defects. Two major defect types of scratch and stain in capsule surfaces are detected with a detection accuracy rate of 98.9%. The experimental results show that the proposed approach can obtain stable characteristics of surface defects, good detection speed, high accuracy, and good generality.

Key words: Region-Convolutional, Neural Network, Detecting Capsule, Surface Defects.

1. INTRODUCTION

Detection of capsule surface defects is important in the production of drugs. In the early stage, capsule surface defects are detected by manual visual inspection. However, manual sorting has the disadvantages of slow speed, high cost, and high error rate in tiny flaw detection. Machine vision inspection can achieve intelligent defect recognition in diversified and complex environments. Image analysis techniques are widely used for visual surface inspection in automated manufacturing to ensure product quality and yield. These techniques have the advantages of low cost and high detection accuracy.

Defect detection based on machine vision includes image acquisition, image feature extraction, and defect detection and recognition. Image feature extraction is an important step to ensure accuracy in defect detection. In practical applications, feature extraction is the most complicated step in defect recognition. In recent years, many features have been constructed to express the image concisely and efficiently, such as SIFT(Park, Kwon and Park, 2016), HOG(Feng and Chen, 2008), and LBP(Shanmugamani, Sadique and Ramamoorthy, 2015; Gong, Chu and Wang, 2015; He, Zhang, Ren and Sun, 2014). However, constructing and selecting features are considerable undertaking. Thus, the optimal solution is determined by numerous experiments, but not generally.

Deep learning is a new research area in machine learning. At present, extracting features by deep learning has elicited considerable research attention. As a widely used deep learning model, convolutional networks are driving advances in recognition and have achieved great success in many areas. Convolutional networks are powerful visual models that yield hierarchies of features. In 2012, Krizhevsky(Piras, Piras and Chiapponi, 2013) presented a deep convolutional neural network (CNN) to classify the ImageNet data set, and placed first in the ILSVRC-2012 championship. The network showed that features extracted from the activation of a deep convolutional network are better than a host of conventional visual representations and have good hierarchy and generality. Recently, the R-CNN achieved a great success on the object detection task(Krizhevsky , Sutskever, Hinton, 2012). The task of object detection is to find the precise locations of objects in a given image and identify the object categories. The traditional methods of object detection based on local features use the sliding window of the image first, then extract the feature of image window, and finally classify the image window. Given that we need many image windows and apply feature extraction and classification to each window, the amount of computation needed is large. However, the process is easy to apply and does not need considerable computing resources in the localization. As the sliding window extraction process is also time consuming and requires a large amount of computing, too many redundant windows will greatly decrease the accuracy and speed of the following recognition stage. So Girshick R et al.(Krizhevsky , Sutskever, Hinton, 2012) decomposes the object detection problem into two cascaded easier tasks: generating object proposals from images and classifying proposals into various object categories. The general object proposal algorithm can
provide around 2000 proposals per image to cover most of the objects, such as selective search used in R-CNN (Krizhevsky, Sutskever, Hinton, 2012).

Actually, this success benefits greatly from the good object classification capability of CNN in different situations. Motivated by this, we proposed a defect detection approach for capsule surface defects in this study. This method can directly learn the features of capsule surface defects; it can detect defective and non-defective capsules, as well as the category and specific location of defects.

First, the study analyzes the research status of capsule defect detection, and then describes the main idea, region proposal, and training process of R-CNN. Next, the R-CNN-based detection method for capsule surfaces is introduced. Finally, the relevant experiment analysis and comparison are carried out.

2. RELATED WORKS

Quality inspection of capsules is a necessary step in the drug production process, and numerous scholars have studied the method of capsule defect detection and recognition. Zhengtao Z et al. (Luo and He, 2016) provided a 360° capsule surface image acquisition device; the capsule was divided into three regions by combining the projection and threshold segmentation. Islam MJ et al. (Trofimov, 2016) primarily used edge detection to detect defects. Karloff AC et al. (Lowe, 2004) obtained the defect edge of the collected capsule surface image by using median filter, Sobel operator, and image convolution, and then determined whether the capsule has defects. Hou H et al. (Liu, Xie, Wei and Lao, 2014) proposed a capsule defect feature extraction method based on RGB space, solved the time-consuming problem of the nonlinear transform from RGB color space to HSI color space, and obtained better results under the less time-consuming premise. Zuo Q et al. (Li, Liang, Huang and Shi, 2014) extracted the three capsule features of width, height, and area, as well as detected the capsule medicine board integrity by using the minimum Bayes classifier, to determine whether the capsule is abnormal and the medicine board is defective, thereby obtaining higher accuracy. The shape, perimeter, and area of the drug particles were extracted as features to identify the drug particles by establishing a neural network (Ojala, Pietikäinen and Harwood, 1996). By analyzing the drawbacks of classical threshold, dynamic threshold segmentation, and region-growing segmentation approach, Zhu Z et al. (Ojala, Pietikäinen and Mäenpää, 2002) proposed a method based on linear growth, and divided the capsule into three parts more effectively.

Following the current research on the detection approach of capsule defect, the following conclusions can be drawn:

1) We require a new method to extract the feature of capsule defect, which can extract the essential feature of the defect. Contemporary defect detection methods mostly extract features by using image processing methods, which are highly dependent on capsule image preprocessing, and certain defect information can easily be lost. In addition, this method is inappropriate for different types of defect detection, and certain disadvantages are observed in detection speed and defect information clustering.

2) We hope to identify the type of capsule surface defect in defect recognition results. At present, the detection method can only detect or recognize defects. If we can identify defect types further after defect detection and conduct statistics on the number of each defect category, then capsule manufacturers can improve and perfect the production process, as well as improve the overall quality of the capsule products.

3. MATERIALS AND MODELS

3.1. Study Model

Girshick et al. proposed an object detection R-CNN model based on deep learning, which increased the mean average precision (mAP) of the object detection task from 22.581% to 43.933%. The advantages of R-CNN are as follows:

1) Generating potential object candidate regions first to reduce the search space.

2) Applying high-capacity convolutional neural network to get internal features for bottom-up region proposals, which can get high-precision object detection based on general classification tasks.

3) Transferring the supervised pre-trained image representation for image classification to object detection.

R-CNN model has been used by many researchers for the realization of detection methods (Farris, Sanowar and Bader, 2010; Tsai and Lai, 2008; Uijlings, Sande, Gevers and Smeulders, 2013), such as face detection and pedestrian detection and so on. We think that R-CNN training method is highly appropriate for capsule defect detection.

If the region is well designed for the capsule image, the method can greatly reduce the complexity of the capsule defect detection. We need to research how to get high-quality candidates and how to get the trade-off between capsule candidates quality and quantity of object hypothesis and the defect detection accuracy.

After getting capsule regions, it is crucial to extract feature and identify the category of the regions. A deep convolutional network is built to extract the features of capsule candidate regions, in which the pixel signals of the capsule region directly pass through the deep CNN and the internal characteristic of the defect region is
extracted layer by layer through the deep convolutional network. Applying the R-CNN mode in capsule defect detection, we must design the structure of capsule deep CNN and parameters, such as number of layers, filter sizes, etc.

Since there often are a small amount of training dataset in capsule defect detection domain, pretraining is adopt to solve this problem. The deep CNN is pre-trained on the ImageNet dataset, and then applied to capsule domain. In addition, we use two distinct forms of label-preserving transformations to enlarge the capsule dataset. One is image translations and horizontal reflections. And the other is to generate multi-scale capsule image input for training of the deep CNN, which also obtain multi-scale features of capsule images.

3.2. Capsule Defect Detection Based on Region-Convolutional Neural Network

Capsule defect detection based on R-CNN first obtains the capsule region proposals by selective search, and then extracts the features of the capsule surface defects with translation and rotation invariance by deep convolution network. The pipeline of our proposed approach is shown in Fig 1: (1) annotating capsule image ground truth, (2) generating capsule region proposals, (3) computing capsule surface defect feature, (4) fine-tuning deep convolutional networks, and (5) training capsule defect classifier. Capsule defect detection and recognition will use the trained classifier.

![Figure 1. The capsule defect detection pipeline based on RCNN.](image)

Annotating capsule image ground truth: The ground truth for the capsule sample is the defect category and the bounding box of the defect, which is the position of the image. Ground truth annotation is the basis of the capsule defect detection. We use our capsule defect dataset, which we collected ourselves, so ground truth annotation is necessary before defect detection.

Generating capsule region proposals: capsule region proposals are generated by selective search, which can capture an accurate bounding box of almost any object with a small set of object localizations. The candidate regions of capsule image produced by selective search are shown in Fig 2:

![Figure 2. Capsule region proposals.](image)

Computing capsule surface defect feature: we first pre-trained the deep convolutional network through the 1000-category classification task of the ImageNet data set, and then obtained the initial parameters of deep CNNs. The architecture of pre-trained deep CNNs is AlexNet, which was proposed by Krizhevsky. AlexNet has eight layers. The first five layers are convolution layers, the next two layers are fully connected layers, and the last layer is softmax classification. The region is warped to a size of 227*227, which is input into deep CNNs. The second fully connected layer is activated as the feature of the candidate region, with a dimension of 4096. The feature extraction process is shown in Figure 3.
Fine-tuning deep convolutional networks using capsule samples: First, we replace the pre-trained deep CNN’s ImageNet-specific 1000-way classification layer with a randomly initialized (N+1)-way classification layer (where N is the number of capsule surface defects categories, plus 1 for background), as shown in Fig4. When fine-tuning the network, if the overlap between the candidate region and the ground truth is more than 60%, the region will be a positive sample to fine-tune the network, and all the regions whose overlap is below 40% will be negative samples. The overlap rate between regions and image ground truth is calculated following Formula 3-1:

$$\text{Overlap}(g,l) = \frac{\text{area}(g) \cap \text{area}(l)}{\text{area}(g) \cup \text{area}(l)}$$

where $g$ is the image ground truth, $l$ is the candidate region, and area() is the corresponding area.

**Figure 4.** the architecture of the fine-tuning capsule deep convolution network.

Training capsule defects classifier: This will be used to detect and identify the capsule defect. We use the image ground truth as a positive sample, and all the regions with an overlap rate of less than 30% are used as negative samples to train the support vector machine. Given a test capsule defect image, the trained capsule defect detection classifier is used to classify the candidate regions, as shown in Fig5.

**Figure 5.** The pipeline of capsule defect detection.
4. RESULTS

This experiment is conducted using a computer with an Intel Core i7-4770K 3.5 GHz, 8 GB memory, and under the Linux environment with Caffe and MATLAB R2014a. We first detect capsule images with scratch and stain respectively, and then mixed various defective capsules images to detect.

4.1. Detection Performance of Different Capsule Surface Defect

Scratch Defect Detection: Scratch is a common defect of capsule. All capsule samples are normalized to the size of 100*100. A total of 200 scratch capsule images are utilized as training data, in which the capsule position of training samples is random and with random translation and rotation; the scratches also appear randomly. A total of 100 capsule images are used as test samples, which contain defective and defect-free capsule images. Then, we adapt the class number of output softmax layer of the pre-trained deep CNN to 2 (scratch and background), and then generate the region proposals of training and test samples by selective search. When detecting the test samples, a detection threshold is set. The bounding box with the highest score is marked as the detection result. In the experiment, after obtaining the detected scores of all candidate regions, we determine whether a defect exists by comparison with the detection threshold. If the scores of all the regions are less than the threshold, the capsule image is determined to be defect-free. Different detection thresholds will directly influence the detection accuracy. With decreasing detection threshold, the missing rate become smaller until reaching 0, but the false positive detection rate will greatly increase because the capsule edge will be falsely detected to have scratch defects. With the increase of the detection threshold, the false positive detection rate decreases until 0, and the missing rate will rise; the optimal detection threshold is 0.474.

Table 1. Detection accuracy of capsule scratch defect detection.

<table>
<thead>
<tr>
<th>Detection threshold</th>
<th>Training data</th>
<th>Test samples</th>
<th>Defect-free samples</th>
<th>Defective samples</th>
<th>Accuracy</th>
<th>Positive false rate</th>
<th>False positive rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.474</td>
<td>200</td>
<td>100</td>
<td>40</td>
<td>60</td>
<td>97%</td>
<td>1%</td>
<td>2%</td>
</tr>
</tbody>
</table>

Stain Defect Detection: Stains are also common capsule surface defects. The experiment takes 100 capsule images of 100*100 size as training samples and 50 capsule images as test samples. Test samples contain defective and defect-free samples. The maximum accuracy of the capsule stain detection is 100%. In stain detection, without the interference of the capsule edge information, the detection threshold is relatively easy to set. When the detection threshold is within the range of 0.1–0.4, the accuracy of the stains detection is relatively stable.

Table 2. Detection accuracy of capsule stain defect detection.

<table>
<thead>
<tr>
<th>Detection threshold</th>
<th>Training data</th>
<th>Test samples</th>
<th>Defect-free samples</th>
<th>Defective samples</th>
<th>Detection accuracy</th>
<th>Positive false rate</th>
<th>False positive rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>100</td>
<td>50</td>
<td>20</td>
<td>30</td>
<td>100%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Mixed Detection of Scratches and Stains: the experiment mixes 300 scratch and stain capsule images, and sets the number of categories to three (scratch, stain, and background). A total of 145 capsule images are used as test samples. The contrast detection result is shown in Figure 6.

Figure 6. The Groundtruth and detection result of Stain and Scratch capsule images.
4.2. Detection Accuracy Analysis

4.2.1. Detection accuracy with different threshold

Detection accuracy, positive false rate, and false positive rate are used to measure the accuracy of detection. Detection accuracy is defined as follows:

\[
\text{Detection accuracy} = \frac{\text{Number of Sample Correctly Detected}}{\text{Total Number of Samples}}
\]

where total samples include defective and defect-free samples.

\[
\text{Positive False Rate (PF)} = \frac{\text{Number of Defective Samples Detected as Defect-free}}{\text{Total Number of Samples}}
\]

\[
\text{False Positive Rate (FP)} = \frac{\text{Number of Defect-free Samples Detected as Defective}}{\text{Total Number of Samples}}
\]

The accuracy of various defects of capsule is shown in Table 3.

### Table 3. Detection accuracy, PF rate, and FP rate of capsule defect detection.

<table>
<thead>
<tr>
<th>Detection threshold</th>
<th>Defect category</th>
<th>Test samples</th>
<th>Defect-free samples</th>
<th>Defective samples</th>
<th>Detection accuracy</th>
<th>PF rate</th>
<th>FP rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.474</td>
<td>Scratch</td>
<td>95</td>
<td>39</td>
<td>56</td>
<td>97.9%</td>
<td>2.1%</td>
<td>0.0%</td>
</tr>
<tr>
<td>0.3</td>
<td>Stain</td>
<td>50</td>
<td>15</td>
<td>35</td>
<td>100%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>0.1</td>
<td>Scratch and stain</td>
<td>145</td>
<td>54</td>
<td>91</td>
<td>98.6%</td>
<td>1.4%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

4.2.2. Detection accuracy with different number of proposals

Object proposals are a set of candidate regions or bounding boxes in an image that may potentially contain an object. Object proposals are used to reduce the search space as well as false alarms, which are an important component in the detection pipelines. In this paper, we use selective search (Sandell, 2012) to generate object proposals from images. It combines the strength of both an exhaustive search and segmentations.

We tested the detection performance with different number of proposals, as shown in Fig 7. To conclude, there is a trade-off between quality and quantity of object proposals and the defect detection accuracy. High quality object locations are necessary to detect a defect in the first place. Being able to sample fewer object hypotheses without sacrificing quality makes the defect detection problem easier and helps to improve results. Remarkably, for capsule defect detection at a reasonable 400 object proposals, our defect detection accuracy is close to optimal.

![Figure 7. The performance of different number of object proposals](image)

4.3. Effect Comparison with Other Detection Method

4.3.1. Comparison with image subtraction method

We compare the proposed approach with the self-comparison surface defect detection method of fast structure image (Zuo and Shi, 2002). The experiment result of the proposed approach is compared with the experiment result of references (Zuo and Shi, 2002), as shown in Figure 8.
Figure 8. The experimental contrast of the proposed approach and detection method based on ICA.

The experimental comparison indicates the following:
1) Tsai DM et al. (Zuo and Shi, 2002) essentially present a simple image subtraction method to measure the difference between referenced pattern and reconstructed pattern from captured images. They first obtain a reference image in the training stage, obtain the reconstructed image of test sample, and then subtract the reference image with reconstructed image. Finally, the test results are obtained by threshold segmentation. This method is suitable for online defect detection, but the prerequisite is product image registration, and it is sensitive to illumination, translation, and rotation, which can only be applied to surface defect detection with fixed position and scale. By contrast, our proposed method aims to learn the essential characteristics of the product surface defects, an approach that is not only suitable for the fixed position and shape sample surface defect detection, but also appropriate for product surface defect detection with translation, rotation, and other spatial position changes.
2) Experiment result of Fig 8 shows that our proposed detection method can identify defective capsules and defect-free capsule images, as well as obtain the precise location of capsule surface defects clearly. By contrast, References (Zuo and Shi, 2002) can only identify samples as defective and defect-free, but cannot provide more detailed defect information such as defect position.
3) Defect detection method in reference (Zuo and Shi, 2002) need preprocessing images in the learning capsule template. Furthermore, after obtaining the residual image, setting the threshold, and doing threshold segmentation to determine whether the product has defect, the results depend on the preprocessing and residual image detection. The proposed defect detection method does not need any image preprocessing or image processing of detection results; it is a complete and independent learning and detecting process.

4.3.2. Comparison with Sparseness Representation Model for Defect Detection
Rouhana et al. (Rouhana, Vieira and Robertsgalbraith, 2012) has proposed defect decomposition method based on sparseness. In the framework of sparseness representation model for defect detection, a defect image is treated as combination of two components: background and defects. Background is coded by an over-complete sparse dictionary, while defect is sparsely coded by its dictionary that is exclusive to background. The experiment result of references (Rouhana, Vieira and Robertsgalbraith, 2012) is shown in Figure 9.
Comparing to sparse representation model (Rouhana, Vieira and Robertsgalbraith, 2012), the proposed approach has the following advantages:
Defect decomposition algorithm based on sparse representation is based on the principle of blind sources separation and block-coordinate relaxation. The background dictionary and defect dictionary are core elements of the model. For a particular detect object, it is difficult to learn a good dictionary. While the proposed approach in this paper can directly learn the feature of defect through a deep convolutional network.
Compared to our proposed method, the defect decomposition based on sparseness is only a shallow model. A deep learning network structure can simulate human visual perception principle. First, we generate the region proposals. Then, on the basis of acquiring the attention region, the hierarchy sensing system of deep learning of human brain is simulated; the pixel gray signals of the attention region directly pass through the recognition
network, which simulates the deep learning hierarchy model of the visual perception system. So, the internal characteristic of the suspicious defect region is extracted layer by layer through the convolution network.

Figure 9. the experimental result of Sparseness Representation Model for Defect Detection

5. CONCLUSIONS

In this paper, we propose capsule surface defect detection based on region-convolutional network. Instead of fixing the capsule’s position, our method can detect capsule surface defects with different scales and translations and rotations. Instead of using the CNN as a classifier, we use it as a feature extractor resorting to its strong ability of feature learning. Then we train a linear SVM to take on the final classification task. This detecting method can not only detect the defective capsule, but can also recognize the defect types and obtain the specific location of defects. Experimental results show that this method has good performance and outperforms the baseline methods of template matching and subtraction.

In the future, the performance and scalability of capsule defects detection can be improved. We will extend current method to handle more capsule surface defects. In addition, this model could be applied to other product surface defect detection. Although DL has achieved breakthrough in audio and image recognition, it does not have a considerable influence on defect detection yet. Therefore, designing a general DL model that is suitable for defect detection and analysis may become a trend in future research.

ACKNOWLEDGEMENTS

This study is supported by Zhejiang Provincial Natural Science Foundation of China, Grant No. LZ14F030001.

REFERENCES


