Optimal view selection method based on object detection

Fanwu Meng*, Yichen Xu, Zhipeng Huang, Lizhong Wang
School of Mechanical Engineering Beijing Institute of Technology Beijing China

*Corresponding author: 529847958@qq.com

Abstract. Industrial parts detection are often studied in three-dimensional space, because of subparts occlusion, the choice of view has a great impact on the detection results. In this paper, the method of object detection combined with the 3D CAD model is proposed, the object detection model based on convolutional neural network was used to screen the visual angle, so as to eliminate the influence caused by the occlusion of parts, which provides a general method for the selection of the optimal view.

1. Introduction
The traditional detection method is the manual sampling inspection method to detect the workpiece quality, the detection workers mainly through the visual way to observe whether there is abnormal workpiece, once the abnormal is found, then alarm and take out the abnormal workpiece, if there is a continuous abnormal, the production needs to stop, check the processing equipment or workers. In the highly efficient automated processing process, it is a great burden and pressure for the detection workers to monitor the production process and deal with faults, both mentally and physically. Because the abnormal situation of complex workpiece does not often occur and some small and micro parts are not easily detected, the detection workers are easy to fatigue and negligence, and it is difficult to take timely response measures to avoid failure.

Machine vision mainly studies the simulation of human visual function through the combination of mechanical devices and computer image processing. As a widely used industrial detection technology in recent years, the efficiency and accuracy of machine vision based detection technology are unmatched by manual detection. Compared with traditional manual detection, machine vision detection has the advantages of high accuracy, good repeatability, high speed and low cost.

Traditional machine vision algorithm in defect detection field made remarkable achievements, most existing algorithms apply to surface detection. when it comes to 3D complex workpiece, due to the existence of multiple two-dimensional surfaces, the choice of different viewpoints has a great impact on the detection effect. How to choose viewpoints has become an urgent problem to be solved. Zhu et al. proposed a novel viewpoint selection approach that was capable of selecting best viewpoints for 3D models based on a feature points detection process. [1]. Polonsky el al analyzed some of the best view-selection algorithms and linked the measurements to the candidate views, comparing them by showing some of the best views of the models [2]. Yamauchi used grid salience to measure the quality of a view [3]. Yang et al. proposed that the feature entropy of the apparent plane can be used as a measure to measure the quality of the viewpoint, which mainly analyzes the curvature features on the apparent plane and measures the distribution and embodiment of the three-dimensional geometric features of the model on the apparent plane[4].Liu et al proposed a new selection algorithm of optimal view samples for all kinds of 3D objects in the model library which will be set up based on the user’s knowledge[5]. Jin et
al proved that the new skeleton map can effectively integrate into viewpoint selection and provide the functionality of rapid selection for the final best and worst viewpoints [6].

The above viewpoint selection methods basically need to extract the manual features in the three-dimensional space, which makes the robustness weak. In recent years, the convolutional neural network has been greatly developed, and its ability to extract features is very strong. In this paper, the method of object detection based on convolutional neural network combined with the 3D CAD model is used to select the optimal view which not influence by occlusion.

2. The algorithm of view selection method

In the industrial situation, the defect detection is often studied in three-dimensional condition. The selection of viewpoints has a great influence on the detection results. The viewpoint selection method proposed in this paper mainly includes two parts: data set construction, object detection network.

2.1. Data set construction

The training of deep CNN often needs a lot of data. In the production process of many workpieces, due to the complex process and high cost, it is difficult to obtain physical products for experiments in advance. The use of 3D CAD model for testing can provide theoretical guidance for future physical testing. Input 3D CAD model of the workpiece and its N subparts. Make the 3D model center of the object coincide with the XYZ virtual axis center. The virtual camera is placed in the positive direction of the X-axis, away from the 3D model center of the object. In order to ensure the consistency of scale, the placement distance d of the virtual camera is fixed (d is the diameter of the minimum circumscribed circle of the object), that is, the coordinate is (d, 0, 0), and 4 virtual point light sources are set above and below the virtual camera and at its symmetrical position, that is, the coordinate is (d, d, d), (d, -d, d), (d, -d, -d), and each 3D CAD model is rotated along the X, Y, and Z directions respectively. After each rotation of θ, a rendering image is taken once to obtain a total of (360/θ)2 2D rendering images with no duplicate from different perspectives of each 3D CAD model. Data augment is carried out by means of flipping and random cutting. After labeling, the enhanced images are stored as classes divided into training sets and validation sets.

2.2. Object detection network

Object detection methods based on CNN [7-8] can be divided into one-stage methods and two-stage methods. Typical two-stage methods are a series of region-based convolutional neural networks which include RCNN [9], Fast-RCNN [10], and Faster-RCNN [11]. These networks include two stages: region proposal and object detection. Due to the presence of region proposal process, the detection accuracy is improved, but a certain detection speed is sacrificed. The typical methods of one-stage methods are YOLOs [12-14] and SSD [15], etc. These methods abandon the region proposal steps and greatly improve the speed, but sacrifice the detection accuracy to some extent.

In order to ensure a relatively high accuracy, Faster-RCNN is chosen. The Faster-RCNN architecture consists of two networks: RPN (Region Proposal Network) and Fast-RCNN (Fast Region-based convolutional neural network). The function of the RPN module is to extract candidate regions, and the Fast-RCNN module is to classify objects and do bounding box regression in candidate regions. The deep features of the input image are extracted by CNN and output a feature map, on which the RPN network extracts the region proposal. Based on the extracted region proposal, Fast-RCNN performs multi-task learning on the feature map which includes object classification and bounding box regression.

Because industrial product images are less complex and have fewer categories than natural images, a relatively shallow CNN model feature extraction layer can be used to make the model free from overfitting and accelerate the speed of training and detection.

This paper adopts VGG16 model, which consists of 13 convolution layers and 3 fully connected layers. The convolution layer uses 1-pixel step size and 3×3 convolution kernel filled with 1 pixel. The same receptive field can be obtained by using multiple cascaded small convolution kernels as by using large convolution kernels, and the number of parameters can be reduced by using small convolution
kernels. The pooling layer is carried out on the basis of the maximum pooling, which is carried out on the core of 2\times 2 pixels, and the step size is 2. The VGG16 model used for feature extraction only needs to obtain the feature map output by the last convolution layer, so the last three fully connected layers are dropped in this case. At the same time, the shortcoming of fixed size of input image in VGG16 will be overcome by ROI pooling in the subsequent network. In actual training and testing, there is no limit to the size of the input image.

In order to achieve fast end-to-end training, the approximate joint training method is used to train the network. RPN network and Fast-RCNN network are fused into one network, which is called end-to-end neural network. In each iteration step, it is assumed that the candidate regions generated by forward propagation are fixed. During the back propagation, the influence of the gradient predicted by the frame coordinates is ignored and the loss values of the two networks are integrated.

The 2D rendering image was obtained from a dense Angle which defined as $\theta_2$ in the 3D model, and the Faster-RCNN detector with complete training above was used for screening, to filter out the image with no subparts can be detected. The remaining images are sorted according to the number of subparts that can be detected, and the higher the priority they have as a point of view, the higher the priority of the image in the order, the higher the priority of the image as the viewpoint.

![Fast-RCNN architecture](image)

**Fig 1.** Faster-RCNN architecture

3. **Experiment**

The experiments are performed on a Windows PC with Inter core i7-8750H, CPU (4.0 GHz), 8GB DDR3 and NVIDIA GTX1060 6G GDDR5 graphics card. The Tensorflow framework is chosen for deep learning.

Take $\theta_1$ as 30°. According to the method described in section 2, the data set is established. The images of the workpiece and the subparts are shown in Fig.2
Images are labeled by the minimum rectangular frame and saved as the XML format of VOC2007 which includes 4 coordinates of minimum rectangular and the category of the object in it. Use the Faster-RCNN network to perform training and parameter adjustment on the established data set, and save the trained network model parameters.

Take $\theta_2$ as $10^\circ$ to obtain the image to be detected, and use the trained network model for detection. Some test results are shown in the figure. In Fig.3, the detector cannot detect any sub-parts due to the occlusion between the subparts, so the view is not a good view. In Fig.4, when viewing from this Angle, it is impossible to distinguish the two types of subparts, which may easily lead to the false inspection of the subparts. Therefore, this view is also poor. In Fig.5, partial subparts can be observed, so this view is a general view. In Fig.6 all subparts can be completely detected when viewed from this Angle, so this view is the best one. According to the number of the correct sub-parts detected, the higher the ranking, the better the view point. The optimal viewpoint has more information and is more suitable for defect detection of other parts and other algorithms.
Fig 4. bad view

Fig 5. general view

Fig 6. optimal view

4. Conclusion
This paper proposes an optimal view selection method based on object detection. An object detector based on convolutional neural network is used to screen a large number of views generated based on 3D
CAD models. Under the condition of no need to design the manual features, the interference caused by the subparts blocking each other is effectively eliminated, and the optimal view point with more abundant information is selected, which provides a good prerequisite for the subsequent detection algorithm.

Reference
[1] F. Zhu and F. Yang, Best viewpoints selection based on feature points detecting[J]. Application Research of Computers. 2014, vol.31, pp.1242-1245
[2] O. Polonsky, Giuseppe Patané, S. Biasotti, C. Gotsman and M. Spagnuolo. What's in an image?[J]. visual computer, 2005, vol.21,pp. 840-847.
[3] H. auchi, W. Saleem, S. Yoshizawa, Z. Karni, A. Belyaev and H. P. Seidel. Towards Stable and Salient Multi-View Representation of 3D Shapes[C]. ieee international conference on shape modeling and applications, 2006, pp.40-40.
[4] L. Yang, W. Wang and E. Wu. Viewpoint selection by feature measurement on the viewing plane[J].Journal of Computer-Aided Design & Computer Graphics,2008, vol.20, pp.1097-1103
[5] Z Liu, P. Feng, X. Pan and C. Xu. Optimal view selection algorithm for 3D object based on similarity learning[J]. Journal of Computer-Aided Design & Computer Graphics,2012, vol24, pp.918-924.