Nuclear Inst. and Methods in Physics Research, A 920 (2019) 88–94

Contents lists available at ScienceDirect

Nuclear Inst. and Methods in Physics Research, A

journal homepage: www.elsevier.com/locate/nima

Recognition of incorrect assembly of internal components by X-ray CT and deep learning

Yihong Li [a,b], Tong Wu [b], Yan Han [b], Ping Chen [b,*]

[a] School of Science, North University of China, Taiyuan 030051, China
[b] Key Laboratory of Instrumentation Science & Dynamic Measurement, North University of China, Taiyuan 030051, China

ARTICLE INFO

Keywords:
Assembly recognition
Convolution neural network (CNN)
X-ray CT
Projection sinogram

ABSTRACT

It is important to make sure that all components of a complex product are assembled correctly. Because in many cases, some components are enclosed in an opaque shell, X-ray imaging is currently used to extract their characteristics and match prior-known ones. However, X-ray imaging is not very robust in recognition of incorrect assembly of internal components, because some of them may overlap. To solve this problem, we propose a new method to detect internal component assembly fault, by X-ray computed tomography (CT) and convolutional neural network (CNN). Multi-view imaging is implemented by mechanical rotation of a product in respect with an X-ray CT machine to capture multiple projection information on each internal component, and then the component can be recognized by making use of deep learning. A CNN model is trained to classify the internal components and give the coordinates of each component. Based on the CNN recognition results and the CT projection sinogram, a projection corresponding to a reference in a projection data set of a standard product can be found. By comparing and matching the locations of each component, transposition or dislocation can be recognized. Both simulation and experiment show that this new method can effectively identify incorrect assembly, missing assembly, transposition, and other problems, improving the product quality.

1. Introduction

Some mechanical machines contain many complex internal components, which may be assembled by robots or manually. In some cases, the internal components are assembled incorrectly, such as transposition, short shipment, misalignment, and over-shipment, affecting the product quality. Therefore, it is important to identify any fault assembly of the internal components. Because in many cases, the internal components are inside of an opaque metal enclosure, it is difficult to check whether or not the internal components are assembled properly. Therefore, an effective detection method must be used to obtain the internal information.

Because of the opaque enclosure, the current detection and recognition method is based on digital radiography (DR) [1] and computed tomography (CT) [2,3]. DR can give attenuation information reflecting the thickness and density of an internal component. Considering DR can only give integral information, some complex components would cause the projected information to overlap. Therefore, DR cannot identify fault assembly of a complex product.

CT can give a cross-sectional image by multiple projections from different angles and can even give three-dimensional (3D) images of each component. Based on 3D CT images, the internal components can be accurately visualized. However, this technique is not applicable for online detection of internal components of some products because it requires intensive calculation and hence is time-consuming.

X-ray image recognition has found some successful applications in medicine and industry. Gedik used fast finite shearlet transform to extract feature vectors and applied a support vector machine (SVM) to achieve breast cancer recognition [4]. Some people attempted baggage inspection to recognize dangerous objects using Gabor texture features [5], bag-of-words-based techniques [6,7], pseudo-color, texture, edge, and shape features [8], and computer vision [9], based on a single view of mono-energy X-ray images or dual-energy X-ray images. To avoid the problem of overlapping and improve recognition, multi-view X-ray imaging was proposed. Yue et al. developed an online automatic detection system with X-ray to check assembly fault [10]. Mery proposed automated object recognition by multiple X-ray view [11]. However, these methods mainly used adaptive segmentation, feature extraction, connection area matching and classification. In some cases, due to mechanical precision and tolerance errors, slight displacement, rotation, and scale-zooming occur in multi-view DR, affecting

Abbreviations: 3D, three-dimensional; CNN, convolution neural network; CT, computed tomography; DR, digital radiography; GPU, graphics processing unit; IOU, inter-section over union; LRN, local response normalization; PRN, region proposal network; RCNN, regions with CNN features; ROI, region of interest; RPN, region proposal network; SVM, support vector machine

* Corresponding author.
E-mail address: pc0912@163.com (P. Chen).

https://doi.org/10.1016/j.nima.2018.12.031
Received 23 June 2018; Received in revised form 11 November 2018; Accepted 9 December 2018
Available online 13 December 2018
0168-9002/

classification and matching of the connection area. As a result, the internal components cannot be accurately recognized. Therefore, a more robust method is needed, which should be resistant to objective factors, such as slight part displacement or over-lapping components.

The traditional feature extraction method cannot achieve robust recognition, because the extracted feature is unique (only focus on color, texture, shape, etc.), and that cannot fully describe all features of the components. Therefore, finding more efficient identifying features is the key to effective matching. In recent years, machine learning has become increasingly popular for solving feature recognition problems. In 2006, Hinton et al. proposed the concept of deep learning [12]. Later, Girshick et al. proposed a series of target detection methods combining the regional merge algorithm with a deep convolutional neural network (CNN) [13–15], such as Regions with CNN features (RCNN) [13], Fast-RCNN [14], or Faster-RCNN [15]. They improved the detection accuracy by optimizing the CNN structure, from an initial 66.0% [13] to 68.1% [14] and subsequently, to 73.2% [15]. They also improved the detection speed. The above methods were applied to solve some practical problems. For example, Sun et al. proposed an automatic fault recognition system based on CNN models for recognizing four typical faults about the running freight train. [16]. However, CNN cannot solve the internal assembly recognition problem, because the internal components can overlap each other in X-ray images.

Deep learning has demonstrated its advantages in target detection [17,18]. This paper presents a new method that combines X-ray CT and CNN, which can give classification results and coordinates of each internal component, so that missing component(s) or incorrect location can be detected.

2. Classification and recognition based on CNN

2.1. Principle

To detect internal components, a CNN is used to perform feature extraction. The extracted features are used to train a classifier. The trained model is applied to classify the internal components and detect the missing components. The input of CNN consists of three parts: an image, ground truth and object proposals. An "object proposal" can be defined as the area where a target exists. CNN calculates the ground truth coverage and the object proposals, obtains the region of interest (ROI) in each image, and extracts the corresponding feature vector for each ROI. A method based on the CNN model is shown in Fig. 1. The first stage is target recognition, in which the CNN model outputs classification and recognition probability for each component. The second stage is fault detection, where any displaced (e.g. missing or transposed) components are recognized.

2.2. Design of CNN

CNN consists of a feature extraction layer and a full connection layer. The feature extraction layer consists of convolutional and pooling layers. Each layer has a different convolutional kernel size, and each convolutional kernel processes an image by extracting different features. Then, the convolution results undergo an activation function, after which an output feature map is obtained. Each output feature map may be a combination of various input feature mapping values.

$$Outputmap_j^l = f\Big(\sum_{i\in M_j} Inputmap_j^{l-1} c_j^l + b_j^l\Big) \tag{1}$$

where $Inputmap_j^{l-1}$ is the feature map of the previous layer, l is the number of feature maps, c_j^l is the jth kernel in the lth layer, and b_j^l is the additive bias of the jth kernel in the lth layer.

These results are input into a nonlinear function $f()$ to obtain down-sampled results.

$$x_j^l = f(down(Inputmap_j^{l-1})). \tag{2}$$

A pooling layer usually follows the convolutional layer and is intended to merge similar features. It reduces the feature map dimension, helps avoid over-fitting, and creates invariance to small shifts and distortions. Here, *down*() represents a down-sampling function.

A local response normalization (LRN) layer follows the max-pooling layer. In an LRN layer, the output maps and input maps have identical channels, widths and heights. The local input area is normalized to accelerate the process of training the CNN model. In CNN, all the convolutional and pooling layers are followed by several fully connected layers to satisfy the output layer. The number of feature extraction layers and fully connected layers can be adjusted as required. The network structure of multiple convolutional and pooling layers enables feature extraction with a high degree of invariance. To reduce over-fitting in the fully connected layer, a regularization method called "dropout" [19] is commonly used without large training datasets. The dropout technique enables skipping the weights of some hidden layers in CNN, forcing the network to learn more robust features in an image, which helps to limit over-fitting. For the assembly recognition of the complex product, it is difficult to obtain good training result with gradient extinction from the deeper network because of limited data. Therefore, the ZFNet model is used for training [20].

2.3. Detection of missing component using ZFNet model

In general, a training process can be divided into four steps:

(1) candidate region generation, by splitting the input image into many small regions and extracting the object proposal regions

(2) feature extraction, by computing the features for each object proposal region using the CNN model

(3) classification, by analyzing each region using a class-specific linear SVM to determine whether a part belongs to a given class (positive samples) or not belongs to that class (negative samples)

(4) location refinement, by training a linear regression model to predict a new detection window given the last pooling layer features for a selective search region proposal.

To improve the accuracy and speed of the training process, a region proposal network (RPN) is introduced, which takes an image (of any size) as the input and outputs a set of rectangular object proposals with a negative value. A previous RCNN algorithm used a selective search process to generate object proposals [21]. In the design of CNN, RPN computation shares features with the CNN model and is connected to the 5th convolution layer (*Conv Layer* 5). The structure of the CNN model is shown in Fig. 2.

During the training process, the convolution operation is performed by a fixed-size sliding window in a feature map after *Conv Layer* 5. The fixed size is called the anchor. At each sliding position, 3 scales ({128, 256, 512}) and 3 aspect ratios ({1:1, 1:2, 2:1}) are set. For all manually-marked ground-truth rectangles, the inter-section over union (IOU) with the anchor is calculated. An IOU above 0.7 is considered as a positive label. For an IOU below 0.3, a negative label is assigned. Thus, a proposal that is too small or beyond the thresholds is eliminated. This approach significantly accelerates the training and detection speed. In Fig. 2, the CNN model inputs the convolutional feature maps into two fully connected layers and outputs the classification results and the coordinate rectangle.

2.4. CNN training data

Assembly recognition is a specific engineering problem. Because no public dataset exists, specialized training data must be generated manually. For X-ray imaging, the projected gray and structural information of each internal component differs if the projection angle changes. However, the projection at any angle can be found in all CT projections when the product is rotated. Thus, CT images can be used to build a training data set. Specifically, the X-ray projection is captured by rotating a product. Then, all the CT projections are used as the data set.

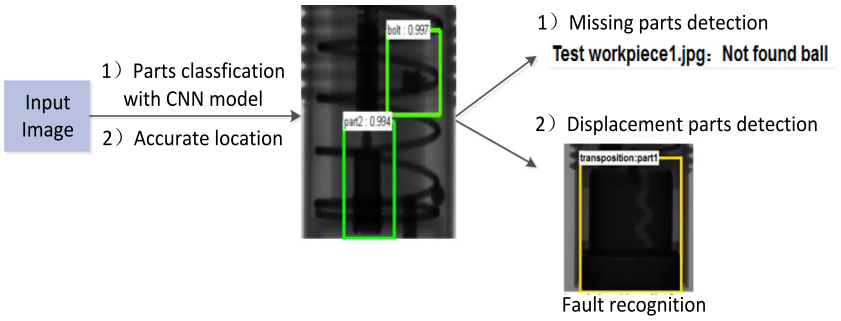


Fig. 1. Fault detection system.

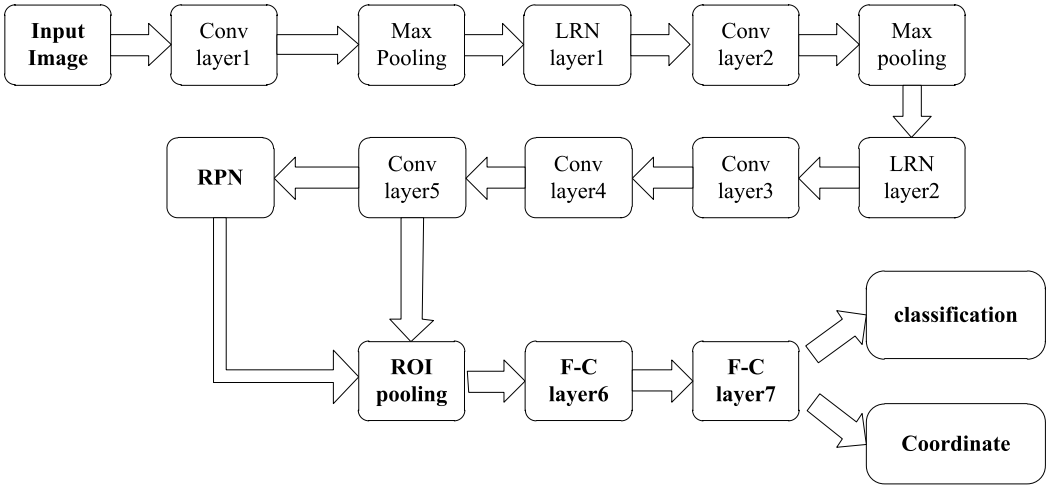


Fig. 2. Structure of CNN model.

Because of the small differences between adjacent angles, the training set and testing set do not use the minor interval samples. However, using the CT projections only as the training data set would result in over-fitting. To solve the insufficient data problem, the projected gray transformation and symmetry inversion are used to extend the dataset manually. During the testing process, the product projection is captured at random angles.

3. Location recognition of internal component based on CT projection sinogram

CNN recognition can obtain the target classification and location coordinates of internal components, and can recognize presence or absence of the components. Because location recognition, such as missing assembly, misloading, and transposition, is required, assembly defects cannot be obtained by CNN directly. Therefore, for assembly recognition, a location and recognition process must be applied.

3.1. Projection sinogram

According to CT imaging, the projection information differs at different rotation angle. A projection data set occupies a 2π angular range from fan beam geometry. A projection trajectory of one point in the object is a so-called sinogram, as shown in Fig. 3. In a sinogram space, the horizontal axis represents the detector channels, and the vertical axis represents the projection angle. Therefore, a single projection is represented in a sinogram as a set of samples located along a horizontal line. The projection of a point in the object is specified by its polar coordinate (r, φ) in the sinogram space. To calculate the location of a point on the detector plane, a rotating coordinate system (x', y') is defined, where the y' axis corresponds to the CT rotation angle β. The x' coordinate of the point clearly satisfies the following relationship:

$$x' = r \times \cos(\phi - \beta). \tag{3}$$

In Fig. 3, the same projection sinogram exists in 3D-CT scans. In the ray projection space, the trajectory of projection location for a given point in the 3D space is a sinusoidal line because of the rotation of the ray source and the detector position. Therefore, during CNN training and testing, for each component to be recognized, the output location coordinates for all CT projections form a unique sine line. During CNN training, the recognition model is saved. The location of each component is also saved, which serves to capture the location information of all components.

3.2. Location recognition based on the sinogram

Assembly defect recognition involves in incorrect location of internal components, which can be recognized by the trained CNN model, and the location of all recognized components can be captured. When the assembly is correct, the recognized location of all the components match those in the CNN training set based on their CT projection sinograms. Based on the saved location of each component to be recognized and the CT sinograms, the test projection can find an optimal matching projection of the standard product according to the location coordinates. To save location information, the projection of the standard product is consistent with the current detection image when a matched result is found. When no matched result can be found, the component may be missing or transposed. During the matching projection, the coordinate information is further used to identify assembly defects, such as missing assemblies, misloading, and transposition.

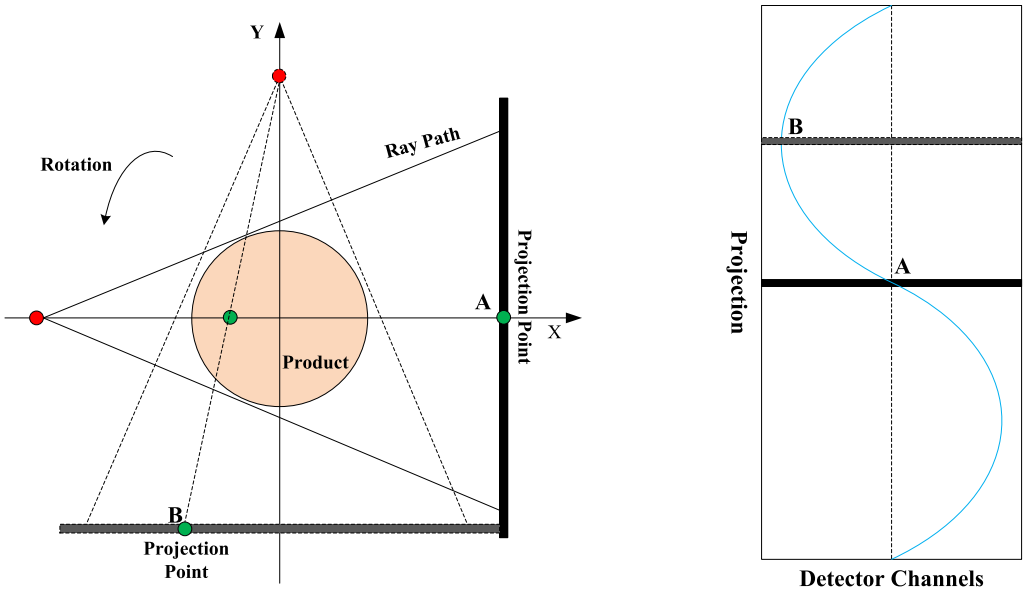


Fig. 3. Mapping between object space and sinogram space.

In a projection, angle matching mainly relies on the location coordinate. If assembly defects occur, all recognition components are used to find the matching angle based on the projection sinogram to avoid incorrect matching due to assembly tolerance and projection overlap. The most frequently occurring angle is selected for recognition of the location of a component. Because of X-ray projections overlap, some small components can be obscured by larger components in the projection direction. Therefore, it is necessary to recognize small components from other projections based on the above recognition process. The complete recognition flow chart is shown in Fig. 4.

4. Experimental results

4.1. Dataset construction

A dataset consisting of training images and test images is used. The experimental product is a cylindrical product as shown in Fig. 5a, which can be divided into three layers. The upper layer consists of a spring and a screw. The middle layer consists of a large spring, a small spring and a bolt. The bottom layer consists of a metal ball, a columnar metal and a screw.

The following are some key parameters:

- Model of a CT system: YXLON FF20
- Distance from the X-ray source to the object: 550 mm
- Distance from the object to the detector: 200 mm
- Voltage applied to the X-ray source: 170 kV
- Current: 60 μA
- Training database: 720 projections at 0.5° interval
- Resolution of each image: 261 × 561.

To show the location of every part in the product, 3D imaging was performed as shown in Fig. 5b, which is obtained by CT reconstruction and 3D visualization.

The training data set consists of 720 CT projections, the logarithm transformation of all projections, and the symmetric inversion of all projections. In the training data set, 70% of the data are selected as the training set, and 30% of the data as the validation set. 360 projections at 1° intervals were collected as the testing data for the standard product. Fig. 6 shows some samples from the projection data training set.

4.2. Training and component classification

An experiment was conducted on a computer with a Windows 7 64-bit operating system, Cuda 7.5 and Microsoft Visual Studio 2013. The Caffe framework is used for deep learning. A Quadro M4000 graphics processing unit (GPU) is used to speed up data processing. After the classification model was trained, the experimental dataset was applied for classification. During the verification process some projections were used, which were not involved in the training process, including a qualified product and an unqualified product with assembly defects. To classify the internal components, these data were input into the CNN model, which was trained during the classification process. The results of the qualified and missing-component test images are shown in Fig. 7. The two images with different projection angles indicate that the model detected all 10 components and completed the classification. The boxes show the different components. The label above each box shows the component name. The number reflects the degree of compliance between the classification results and the class. In Fig. 5, two components were artificially removed. The projection was captured at one angle, and the CNN model was used to recognize the missing components. The result is shown in Fig. 7c. At the bottom of Fig. 7c, the missing components (part1 and ball) are labeled.

4.3. Location recognition with incorrect assembly

Using the CNN model, the coordinates of the components being recognized can be obtained. The location between the current testing image and the CT projections of the standard product is compared based on the CT projection sonogram, to determine whether the internal components are displaced, transposed, or have other defects. The experimental results are shown in Fig. 8.

The CNN model was used to identify each component. The task is to find a matching image among the standard product projections, based on the CT projection sinogram. The matching result is shown in Fig. 8b. The coordinates of each recognized component in Fig. 8a is compared with the standard product to find the transposed components. The missing “ball” is recognized, which is labeled in the bottom of Fig. 8a. The transposed component is also recognized, as shown in Fig. 8c.

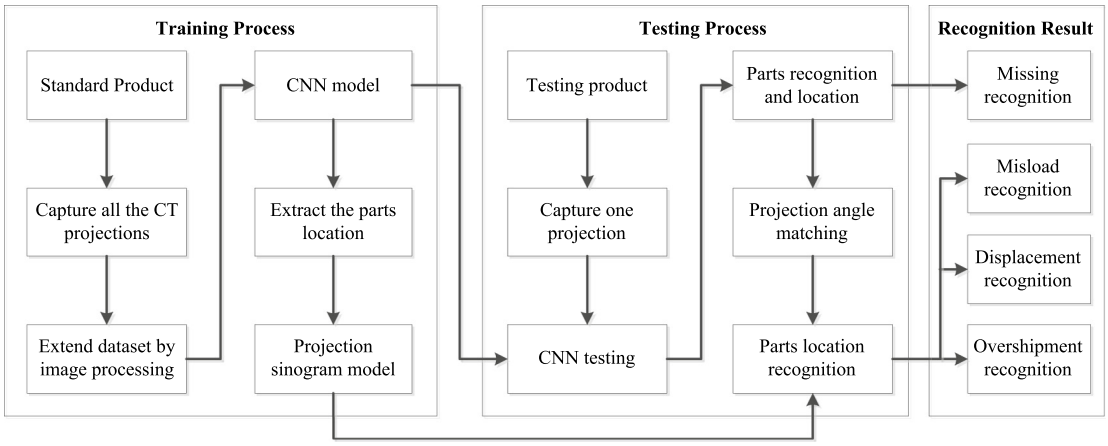


Fig. 4. Location recognition process for incorrect assemblies.

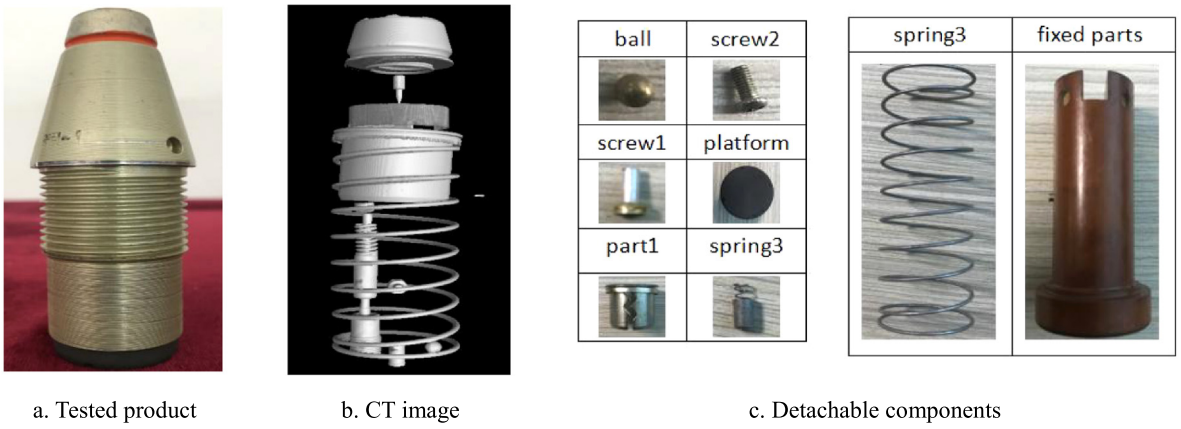


Fig. 5. Parts and structure of the test product.

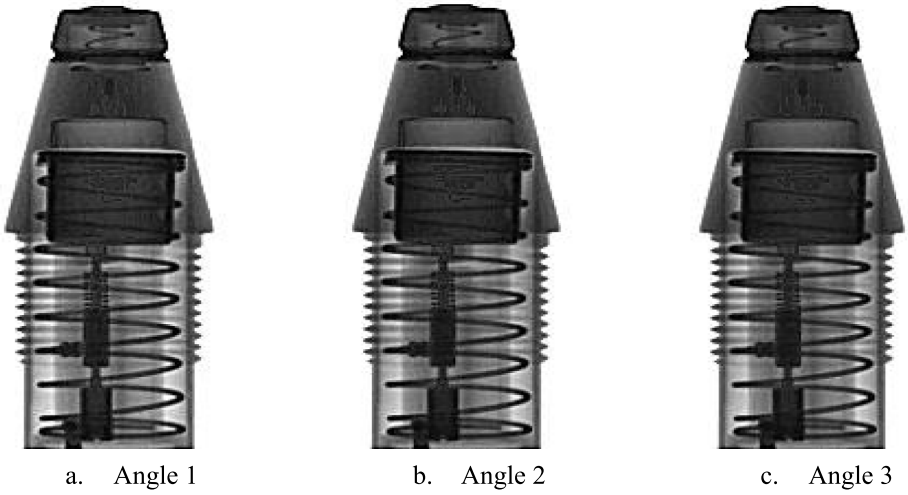
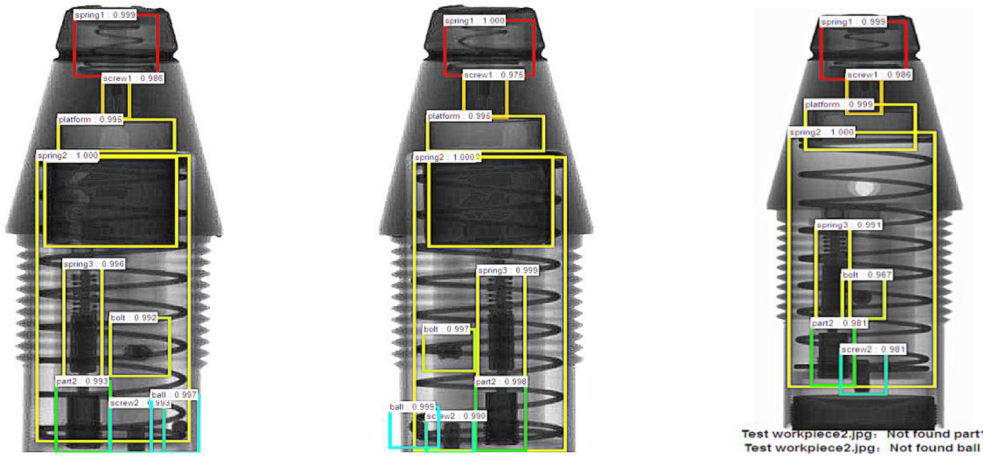


Fig. 6. Training set of projection data from three adjacent angles.

4.4. Component recognition with overlap

The X-ray projections generate an accumulative image. In the X-ray direction, the attenuation of all components overlaps. At some projection angles, some small components are covered by larger components. As a result, the captured projection cannot show the small components. For example, in Fig. 9a, in the bottom of the projection, the "ball" and "screw1" cannot be seen. Based on the CNN result (see Fig. 9b), these two components cannot be recognized. However, based on the CT projection sinogram, the overlap problem was overcome by rotating the component by 90° to obtain a projection where the two components are not overlapped (see Fig. 9c). Subsequently, all parts are recognized by the CNN model, as shown in Fig. 9d.

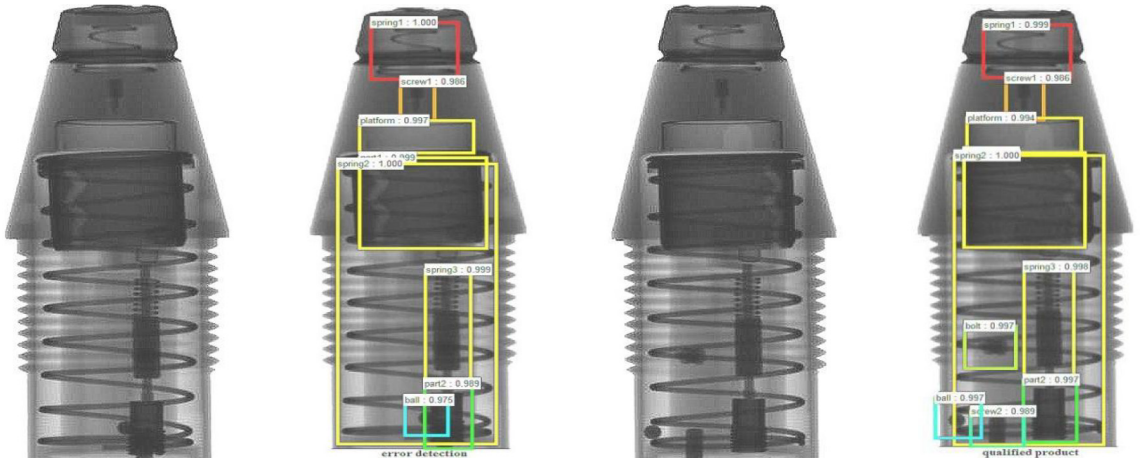


a. Qualified workpiece b. Qualified workpiece with the other projection angle c. Workpiece with missing internal components

Fig. 7. Test results of classification.

a. Recognition result b. Matching result c. Displacement detection

Fig. 8. Incorrect assembly recognition with a current-angle testing image matched with a standard product image.



a. Occlusion projection b. Recognition of occlusion c. Projection with 90° rotation d. Recognition of rotation

Fig. 9. Recognition results with components overlapped.

5. Conclusions

This paper presents a method for recognizing internal assemblies based on CNN classification and CT rotation imaging. In the candidate frame generation stage, RPN detection and the shared weights are combined, allowing the internal components in the X-ray image to be recognized. To overcome the limitations of the CNN model in incorrect assembly and overlap, CT projection sinograms are used. The CNN training process captures CT projections of a standard product, which is adopted as the training dataset for the CNN model. Then, the process evaluates whether the product contained missing, misplaced, or overlapped components via classification and comparison to the target coordinates of the CNN model. Based on the projection sinogram, the coordinate offset is analyzed and calculated to acquire location of internal components relative to the projection of the standard product.

The above proposed method was tested experimentally using a real cylinder product with 10 small components. Using this method, an incorrect assembly with a missing assembly and overlapped components can be perfectly recognized. These results show that the proposed method is robust in assembly recognition and projection overlap. It can solve the problems with CNNs only for recognizing the location of assembly defects. Future work is planned to carry out research into recognition of quantitative transposition. This process requires combining the imaging resolution ratio of the CT system and the deep neural network with precise location.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (61571404, 61871351 and 61801437).

References

[1] Pf van Der Stelt, Better imaging: The advantages of digital radiography, J . Am. Dent. Assoc. 139 (1) (2008) S7–S13.
[2] J.P. Kruth, M. Bartscher, S. Carmignato, R. Schmitt, L.D. Chiffre, A. Weckenmann, Computed tomography for dimensional metrology, CIRP Ann. Manuf. Technol. 60 (2) (2011) 821–842.
[3] H. Jiang, Computed tomography: Principles, design, artifacts, and recent advances, SPIE Int. Soc. Opt. Eng. (2009) 39–46.
[4] N. Gedik, A new feature extraction method based on multi-resolution representations of mammograms, Appl. Soft Comput. 44 (2016) 128–133.
[5] I. Uroukov, R. Speller, A preliminary approach to intelligent X-ray imaging for baggage inspection at airports, Signal Process. Res. 4 (5) (2015) 1–11.
[6] D. Turcsany, A. Mouton, T.P. Breckon, Improving feature-based object recognition for X-ray baggage security screening using primed visualwords, in: IEEE International Conference on Industrial Technology, February (2013) 25-27, Cape Town, pp. 1140-1145.
[7] D. Baştan, M.R. Yousefi, T.M. Breuel, Visual words on baggage X-ray images, in: Computer Analysis of Images & Patterns-international Conference, August (2011) 29-31, Seville, Spain, pp. 360-368.
[8] N. Zhang, J. Zhu, A study of X-ray machine image local semantic features extraction model based on bag-of-words for airport security, Int. J. Smart Sens. Intell. Syst 8 (1) (2015) 45–64.
[9] D. Mery, E. Svec, M. Arias, V. Riffo, J.M. Saavedra, S. Banerjee, Modern computer vision techniques for X-ray testing in baggage inspection, IEEE Trans. Syst. Man Cybern. Syst. 47 (4) (2017) 682–692.
[10] Y. Han, Y. Han, R. Li, L. Wang, Application of X-ray digital radiography to online automated inspection of interior assembly structures of complex products, Nucl. Instrum. Methods Phys. Res. 604 (3) (2009) 760–764.
[11] D. Mery, G. Mondragon, V. Riffo, I. Zuccar, Detection of regular objects in baggage using multiple X-ray views, Insight, Non-Destr. Test. Cond. Monit. 55 (1) (2013) 16–20.
[12] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.
[13] R. Girshick, J. Donahue, T. Darrell, J. Malik, Region-based convolutional networks for accurate object detection and segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 38 (1) (2016) 142–158.
[14] R. Girshick, Fast r-cnn, in: IEEE International Conference on Computer Vision, Santiago, Chile, 2015, pp. 1440–1448.
[15] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. 39 (6) (2017) 1137–1149.
[16] J. Sun, Z. Xiao, Y. Xie, Automatic multi-fault recognition in TFDS based on convolutional neural network, Neurocomput. 222 (2017) 127–136.
[17] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, 2012, pp. 1097–1105.
[18] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: IEEE Conference on Computer Vision & Pattern Recognition, Washington, USA, 2014, pp. 580–587.
[19] G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R.R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, Comput. Sci. 3 (4) (2012) 212–223.
[20] D. Matthew, R.F. Zeiler, Visualizing and understanding convolutional networks, in: Computer Vision-ECCV 2014-13th European Conference, Zurich, Switzerland, 2013, pp. 818–833.
[21] J. Uijlings, A. van de Sande, T. Gevers, M. Smeulders, Selective search for object recognition, Int. J. Comput. Vis. 104 (2) (2013) 154–171.