IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 18, 2019, accepted August 31, 2019, date of current version September 18, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2940411

PolishNet-2d and PolishNet-3d: Deep Learning-Based Workpiece Recognition

FUQIANG LIU AND ZONGYI WANG
College of Automation, Harbin Engineering University, Harbin 150001, China
Corresponding author: Fuqiang Liu (fqliu@hrbeu.edu.cn)

ABSTRACT In recent years, there have been many deep learning research projects based on two-dimensional object detection and three-dimensional point cloud recognition. However, relatively few of these combine the two, and the number of projects based on a three-dimensional workpiece recognition method for the industrial field is even fewer. In this paper, to perform the recognition task for polishing workpieces in manufacturing, we use RGB-D images as input, and propose our designed PolishNet-2d for two-dimensional workpiece detection and our designed PolishNet-3d for three-dimensional workpiece recognition. The two deep neural networks are employed together in series to first detect and then recognize polishing workpieces in an industrial environment. For this paper, a large number of experiments have been carried out on deep learning datasets of polishing workpieces. These datasets were created by us to contain diverse combinations of real and simulated workpieces in real and simulated industrial environments. The contributions of this paper are as follows: (1) A rotation parameter learning network is proposed and a two-dimensional workpiece detection neural network, named PolishNet-2d, which was constructed by integrating our designed algorithms with the backbone networks ResNet101 and region proposal network (RPN), is introduced; (2) A hierarchical feature extraction network is proposed and a three-dimensional workpiece recognition neural network, named PolishNet-3d, which was constructed by integrating our designed algorithms with the backbone network PointNet, is introduced; (3) PolishNet-2d and PolishNet-3d are employed in series, with the detection output of PolishNet-2d being used as the input for PolishNet-3d for its recognition tasks: the workpiece regions are detected in the RGB image; the workpiece point cloud is segmented in the corresponding regions in the depth image, and lastly the segmented point cloud is placed into PolishNet-3d to identify the workpiece; (4) For the experiments in this paper, datasets containing rich and diverse data types of polishing workpieces for industrial fields have been constructed; (5) Numerous experimental results on polishing workpiece datasets show that the conjunction of PolishNet-2d and PolishNet-3d can achieve exemplary recognition results on polishing workpiece datasets.

INDEX TERMS Deep learning, 2d workpiece detection, 3d workpiece recognition.

I. INTRODUCTION

Nowadays, in the manufacturing industrial field, it is still a frontier topic and a difficult problem to recognize three-dimensional workpieces stably, robustly and accurately. At present, the more robust method focuses on using traditional two-dimensional image descriptors or three-dimensional point cloud descriptors to describe the features of a workpiece, and then match and recognize the workpiece using those descriptors. Because the traditional descriptor is designed by hand, when the shape of the workpiece is complicated, and there are only small differences between different workpieces, the traditional processing method can be inadequate. As a consequence, three-dimensional point cloud recognition methodology based on deep learning has become a hot and emerging research field; however, the recognition rate is relatively low, unable to meet the requirements of high stability and robustness in the industrial field. Considering the current research situation of this subject, there are still some problems and difficulties in the research of three-dimensional workpiece recognition:

1) Traditional two-dimensional image descriptors and three-dimensional point cloud descriptors cannot well represent the characteristics of different workpieces with complicated structures, especially when there are only minor structural differences between different

The associate editor coordinating the review of this manuscript and approving it for publication was Venkateshkumar M.

workpieces. Therefore, how to detect and recognize low-texture three-dimensional workpieces stably and efficiently is still an unsolved problem.

2) Recognition methods based on deep neural networks require a high consistency of data distribution between training data and testing data, which also means that a deep neural network trained on a dataset generated from an ideal computer-aided design (CAD) model can hardly be applied in real scenarios, while collecting and labeling a large number of real scene data is very labor-intensive and time-consuming.
3) The accuracy of existing methods based on deep learning cannot meet the requirements of the manufacturing industrial field.
4) Directly segmenting and recognizing the three-dimensional point cloud data in a scene has a very low accuracy rate. Effectively integrating the mature two-dimensional deep learning based object detector into three-dimensional workpiece recognition methods is very valuable.

Three-dimensional data are mainly divided into a point cloud ([1]–[3]), voxels ([4], [5]), a multi-view image ([6],[7]), and an RGB-D image ([8]–[12]). With the three-dimensional point cloud deep learning based processing methods, there are some methods to recognize single object point cloud, and some methods to semantically segment scene point cloud. However, direct recognition of a single object in the scene cannot achieve good results.

In order to improve the recognition accuracy of the three-dimensional workpiece in the industrial field, we use RGB-D images as input; the proposed two-dimensional workpiece detection neural network, PolishNet-2d, and the proposed three-dimensional workpiece recognition neural network, PolishNet-3d. This paper mainly does the following:

1) The two-dimensional workpiece detection neural network, PolishNet-2d consists of the backbone network, ResNet101; the region proposal network, RPN; and a sub-network, X-Net, proposed in this paper for learning the rotation transformation of workpiece objects in the input images. X-Net can learn the rotation matrix of the workpiece itself in an unsupervised way according to the training data. After the transformation of X-Net, the success rate and accuracy rate of the workpiece detection are improved compared with those not using X-Net;
2) A hierarchical feature extraction network is proposed. Combining with the backbone network, PointNet, a three-dimensional workpiece recognition network PolishNet-3d is constructed. With the addition of the hierarchical feature extraction network, the three-dimensional point cloud deep learning based method can extract the features of workpieces at different levels, such that the network can learn more detailed and global features;
3) A dataset containing abundant image types of polishing workpieces for the industrial field is constructed, making it possible for the deep neural network to learn more abundant features;
4) The two-dimensional workpiece detection neural network, PolishNet-2d, and the three-dimensional workpiece point cloud recognition neural network, PolishNet-3d, are combined to detect and recognize the workpiece images collected in real industrial scenes. Good recognition results were obtained on the dataset proposed in this paper.

II. RELATED WORK

A. DEEP LEARNING BASED TWO-DIMENSIONAL OBJECT DETECTION METHOD

The local image descriptor based recognition method had been the mainstream method for a long time, until the deep neural network method achieved a breakthrough in image classification in 2012 [13], when people began to explore ways to use the deep neural network method for object recognition [14]. Girshick et al. first proposed a landmark convolutional neural network, R-CNN. Since then, image classification and image recognition competitions have been held in large numbers (such as ImageNet [15] and MSCOCO [16]). Many deep learning based image classification and image recognition algorithms have been proposed one after another. Ever enlarging datasets enable deep neural networks to solve increasingly more complicated real world problems. DetectorNet [17], OverFeat [18], MultiBox [19], and R-CNN [20] are several deep neural networks proposed almost simultaneously for object detection and recognition. Girshick et al. combined a convolutional neural network, AlexNet [13], with a selective search method, and proposed a new convolutional neural network, R-CNN, for object detection. Recent research results [21] show that the convolution layer of the convolutional neural network has the strong ability to locate objects, which is weak in the fully connected layer. Ren Shaoqing and others put forward Faster R-CNN [22] on this basis, using a region proposal network instead of a selective search method to generate the recommendation region.

B. DEEP LEARNING BASED THREE-DIMENSIONAL OBJECT RECOGNITION METHOD

The traditional point cloud processing method mainly describes the three-dimensional objects by hand-designed features. This presents a problem: for the three-dimensional objects with complicated features, it is impossible to extract abundant three-dimensional information by hand-designed features, a factor which contributes to making the recognition process difficult. Therefore, a data driven approach is needed to understand and process three-dimensional data, that is, a three-dimensional deep learning method.

Charles et al. proposed a network architecture for learning the three-dimensional point cloud data directly, PointNet [1]. The network can be used for classification, component segmentation, and scene segmentation of a three-dimensional

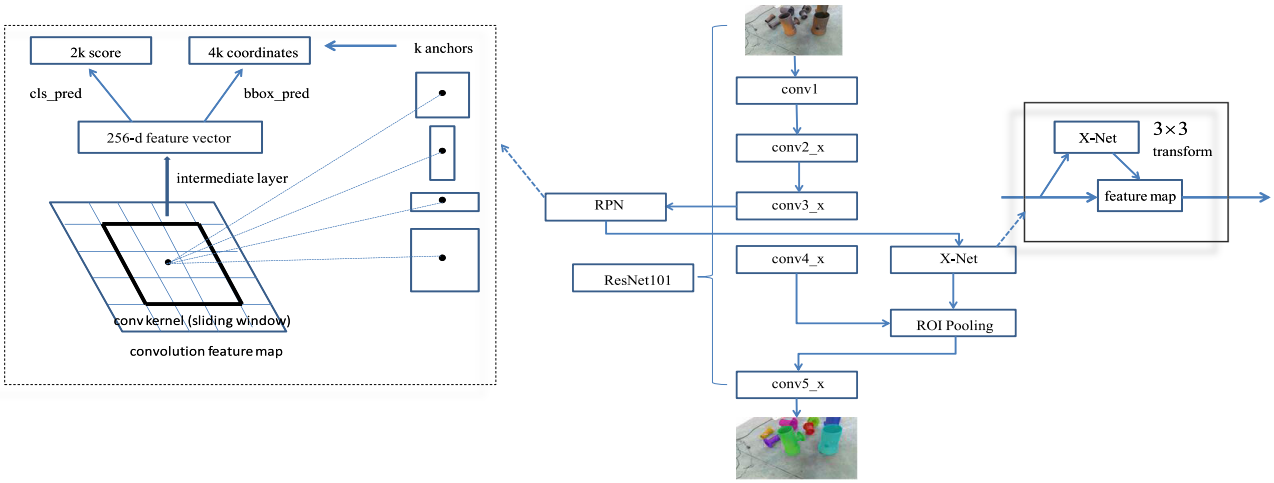

FIGURE 1. Network structure of PolishNet-2d.

point cloud. Song and Xiao [2] and Chen *et al.* [3] first made a region proposal in three-dimensional space. Wadim Kehl et al. proposed a hash method for large-scale three-dimensional object detection, Hashmod [8]. In the same year, they also proposed a local RGB-D data deep learning method for three-dimensional object detection and six-dimensional pose estimation [9]. Qi *et al.* [10] proposed a method that combines two-dimensional detection with three-dimensional point cloud recognition. The typical framework is Frustum-PointNets. Roman Klokov et al. proposed a deep neural network for 3-D point cloud model recognition, Kd-Networks [23]. Jiaxin Li et al. proposed a self-organizing network, SO-Net [24], for point cloud data analysis, which has a permutation invariance for disordered point cloud data. In recent years, three-dimensional workpiece recognition has attracted much attention. However, due to the limitations of the development of three-dimensional shape descriptors and machine learning technology, there has been little breakthrough in the work of three-dimensional workpiece recognition heretofore. Wu and Jen [25] proposed a three-dimensional pyramidal parts classification neural network, for which images were collected from three perspectives of three-dimensional parts, and edge information from the parts were extracted for the input of a polygon classifier. Ip *et al.* [26] and Ip and Regli [27] proposed a classification method for mechanical CAD parts based on machine learning. Ip and Regl [28] use a support vector machine (SVM) to classify prismatic machined parts and post-casting machined parts by using four types of surface curvatures as input vectors for a support vector machine.

III. PolishNet-2d

This paper presents a two-dimensional neural network, PolishNet-2d, for polishing workpiece detection. The network consists of a backbone network, ResNet101; a region proposal network, RPN; and a transformation network, X-Net, proposed in this paper for learning the rotation parameters of a workpiece in an input image. Fig. 1 shows the architecture diagram of PolishNet-2d. The input of PolishNet-2d is an RGB image, and the output of PolishNet-2d is the detected workpiece location. Because the rotation parameters of different workpieces are different in an image, this paper proposes the sub-network, X-Net, which can learn the rotation matrix of a workpiece.

A. NETWORK STRUCTURE COMPOSITION

1) ResNet101

He *et al.* [29] proposed ResNet. ResNet contains Identity Mapping and Residual Mapping, in which identity mapping is the input of the module itself, expressed as x, and residual mapping is the initial output part of the module. The residual network can obtain a new output module by adding the two mapping parts. The new output is represented as $y = F(x) + x$. The input of a module is added directly to the output of the module. ResNet is similar to VGG in that the size of the convolutional core used is mainly 3×3. Based on the VGG network, ResNet adds a shortcut connection to form a residual network module. In this paper, ResNet101 is used as the backbone network of PolishNet-2d.

2) REGION PROPOSAL NETWORK

The region proposal network, RPN, consists of a convolution layer and a fullly connected layer. The convolutional layer is responsible for outputting classification results. For each position in the input feature map, it outputs $2k$ confidence scores, where 2 represents the foreground or background, and k represents the number of anchor boxes at each location in the feature map. The fully connected layer is responsible for outputting the coordinate information of bounding box regression. For each position in the input feature map, it outputs $4k$ coordinate values, where 4 represents the upper-left coordinates of the bounding box, x and y, as well as the width

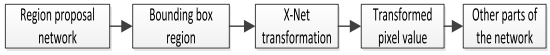

FIGURE 2. Flow chart of X-Net.

and height of the bounding box, w and h, and k represents the number of anchor boxes at each position in the feature map, as shown in Fig. 1.

3) X-Net

The uncertain angle of the workpiece in the image is a large factor in the workpiece detection. In this paper, the sub-network, X-Net, can learn the rotation angle information of the workpiece in an unsupervised way, in such a way that the rotation angle information of different workpieces can be normalized automatically, so as to improve the accuracy of the detection. Fig. 2 shows the flow chart that uses X-Net to transform the image bounding box area obtained by the region proposal network.

The matrix to be trained by X-Net is:

$$X = \begin{bmatrix} cos\theta & -sin\theta & 0 \\ sin\theta & cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

To optimize the transformation matrix X generated by X-Net, and in order to make the feature transformation matrix to be an orthogonal matrix, a regularization term is added after the training loss function of softmax:

$$H'_{y'}(y) = -\frac{1}{N}\sum_i [y'_i log(y_i) + (1 - y'_i) log(1 - y_i)] + \alpha L_{reg} \tag{2}$$

$$H_{y'}(y) = -\frac{1}{N}\sum_i [y'_i log(y_i) + (1 - y'_i) log(1 - y_i)] \tag{3}$$

$$L_{reg} = ||I - XX^T||_F^2 \tag{4}$$

$$y_i = softmax(y)_i = \frac{exp(y_i)}{\sum_j exp(y_j)} \tag{5}$$

Among them, $H'y'(y)$ increases the cross-entropy loss value of the regular constraint items, $Hy'(y)$ is the original cross-entropy loss value, α is the coefficient of the regular penalty term, L_{reg} is the regular constraint in the 2-norm form, X is the transformation matrix, y_i is the prediction label of the neural network output after the softmax function, and y'_i is the ground truth label. The softmax function is used to normalize the classification score into the range from 0.0 to 1.0. By adding regular terms, the optimization process becomes more stable, and the network can achieve better results.

Fig. 3 shows the comparison of the training results of the PolishNet-2d network using X-Net and of the PolishNet-2d network without using X-Net. It can be seen from the figure that the network loss value converges faster after using X-Net, and the final minimum loss value converges to the smaller value than that of PolishNet-2d without using X-Net.

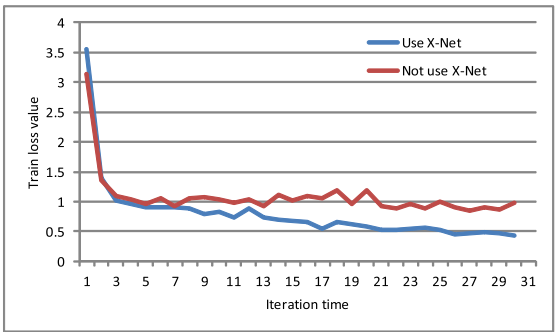

FIGURE 3. Results comparison between the network with and without X-Net.

B. EXPERIMENTAL RESULT AND ANALYSIS

1) EXPERIMENTAL NETWORK INITIAL PARAMETERS

In the training process, the optimization strategy of a weight decay is adopted, with the parameter of the weight decay being 0.0001. In the region proposal network, the length-width ratios of the anchor boxes are 0.5, 1 and 2. The sizes of the anchor boxes are 32, 64, 128, 256 and 512 pixels. The intersection over union (IoU) threshold value of non-maximum suppression is 0.7. During the training process, the number of region of interest (ROI) in each image is 200.

2) DATASET

In this paper, a set of industrial datasets is constructed for polishing workpieces, consisting of eight different kinds of datasets:

1) single simulated workpiece with no background dataset
2) single simulated workpiece with real environmental background composite dataset
3) single real workpiece with real environmental background composite dataset
4) single real workpiece with a real environmental background non-composite dataset
5) multiple simulated workpieces with no background image dataset
6) multiple simulated workpieces with real environmental background composite dataset
7) multiple real workpieces with real environmental background composite dataset
8) multiple real workpieces with real environmental background non-composite dataset.

The above datasets are shown in Fig. 4, in which each row represents a kind of dataset.

3) INTERMEDIATE PROCESSING RESULTS

In this paper, we randomly select one of the images from one of the datasets described above, and then demonstrate the intermediate results of the detection process, including input image, region proposal network, filtered out low confidence detection results, bounding box refinement, non-maximum suppression, final detection results, and the feature map obtained by the region proposal network.

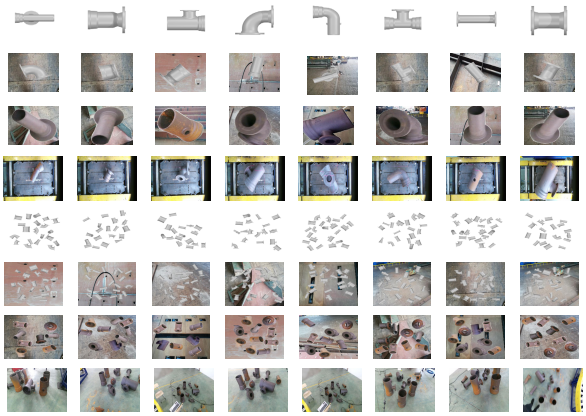

FIGURE 4. Proposed eight kinds of datasets.

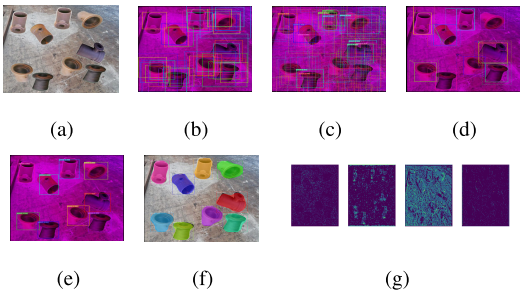

FIGURE 5. Intermediate results of multiple real workpieces with real scene background composite dataset. (a) Input image. (b) Region proposal. (c) Filtered out test results with low confidence. (d) Bounding box refinement. (e) Non maximum suppression. (f) Final detection results. (g) Intermediate feature map of RPN.

Fig. 5 shows the intermediate results of the multiple real workpieces with real scene background composite dataset.

4) TRAINING AND TESTING RESULT

Following is a comparison result between Faster R-CNN and PolishNet-2d, including the training loss curve, testing set accuracy, receiver operating characteristic (ROC) curve, and area under curve (AUC) value.

a: TRAINING LOSS VALUE CURVE

The loss value curve of PolishNet-2d in the training process is shown in Fig. 6 and Fig. 7. Fig. 6 shows the loss value curve on the four datasets with only one workpiece in the image. The Fig. 7 shows the loss value curve on the four datasets with multiple workpieces in the image. In the training set, there are 66, 000 images corresponding to the above 8 datasets, each of which has 8, 250 images.

Among them, dataset 1 in Fig. 6 and Fig. 7 represents the simulated workpiece(s) with no background dataset, dataset 2 represents the simulated workpiece(s) with real environmental background composite dataset, dataset 3 represents the real workpiece(s) with real environmental background composite dataset, and dataset 4 represents the real workpiece(s) with a real environmental background non-composite dataset. It can be seen that the training loss values of the single

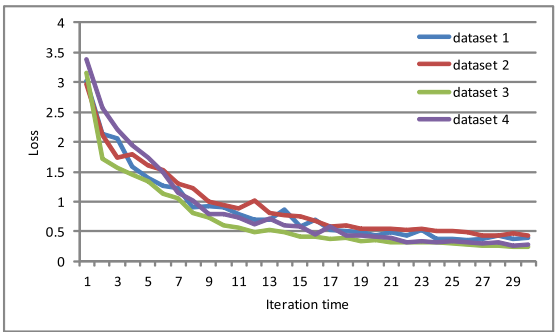

FIGURE 6. Training statistics data for four single workpiece datasets.

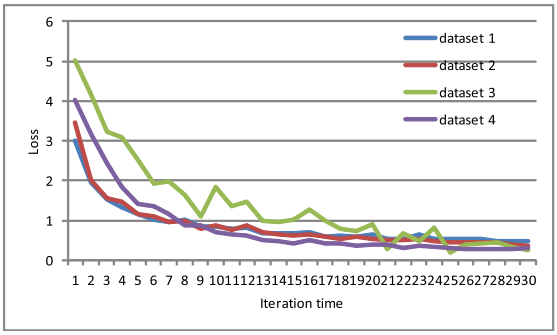

FIGURE 7. Training testing statistics data for four multiple workpiece datasets.

workpiece datasets and of the multi-workpiece datasets converge well for PolishNet-2d.

b: PRECISION, ROC CURVE AND AUC VALUE

Precision is defined as the proportion of true positive samples to all positive samples. The ROC curve is called the receiver operating characteristic curve. The abscissa of the ROC curve is the false positive rate and the ordinate is the true positive rate. AUC is an abbreviation for area under the curve, which is the area under the ROC curve. In this paper, we use Faster R-CNN, the current mainstream object detection network, and PolishNet-2d to test the test set of 2, 500 single workpiece images and 2, 500 multiple workpiece images. The number of workpieces in the test set images is 84. In the multiple workpiece test set, the number of possible workpieces in each image is at least 1 and at most 15. Table 1 shows the detection accuracy of Faster R-CNN and of Polishnet-2d trained on the polishing workpiece dataset constructed in this paper. Fig. 8 shows the ROC curves for Faster R-CNN and for PolishNet-2d on the test set. Table 2 shows the corresponding AUC value of Faster R-CNN and of PolishNet-2d.

c: ANALYSIS OF TRAINING AND TESTING RESULTS

According to the results in Table 1, Fig. 8 and Table 2, it can be seen that PolishNet-2d achieves higher accuracy and AUC values than Faster R-CNN, both in the case of where there is a single workpiece in the image and in the case of where there are multiple workpieces in the image. Faster R-CNN achieves

TABLE 1. Precision statistical data of workpiece detection experiments.

	Single workpiece in the scene	Multiple workpieces in the scene
Workpiece type number	84	84
Workpiece number	1	1 15
Training set image number	66, 000	66, 000
Test set image number	2, 500	2, 500
Faster R-CNN detection precision	94.33%	92.85%
PolishNet-2d detection precision	96.25%	95.40%

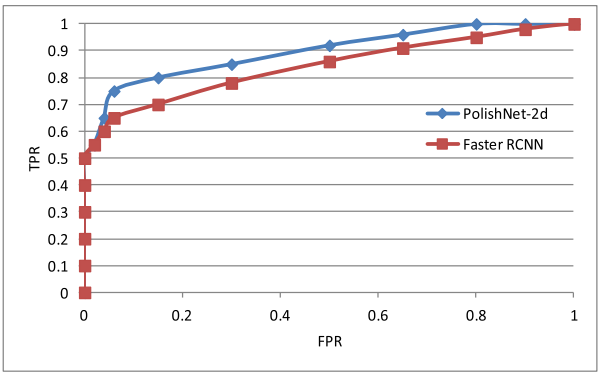

FIGURE 8. ROC curve of Faster R-CNN and PolishNet-2d on testing dataset.

TABLE 2. AUC statistical data on testing dataset.

	Faster R-CNN	PolishNet-2d
AUC	0.865	0.917

an accuracy value of 94.33% and PolishNet-2d achieves an accuracy value of 96.25% in the case of where there is a single workpiece in the image. Faster R-CNN achieves an accuracy value of 92.85%, and PolishNet-2d achieves an accuracy value of 95.40% in the case of where there are multiple workpieces in the image. PolishNet-2d has an AUC value of 0.9165 and Faster R-CNN has an AUC value of 0.865 on the test set. Since PolishNet-2D and Faster R-CNN have the same backbone network, the difference between the two detection networks is that PolishNet-2d introduces the transformation network X-Net, which enables unsupervised learning of the workpiece transformation in the proposed regions obtained by the region proposal network, thus making PolishNet-2d robust for workpiece rotation. The robustness of the system is better than Faster R-CNN in the same dataset, and the accuracy of the system is 96.25% and 95.40% for the single workpiece dataset and the multiple workpiece dataset, respectively.

d: PolishNet-2d NETWORK PERFORMANCE ANALYSIS

PolishNet-2d can achieve an inference time of 287 ms per image (including the CPU time of 30 ms to change the output resolution to the original resolution) on NVIDIA Geforce 1080 GPU. Under such running time, it can also ensure high detection accuracy. Although PolishNet-2d has reached a relatively fast inference speed, in order to ensure the detection accuracy, this paper does not optimize its computation further, for in the industrial workpiece detection area, the requirement for detection accuracy is higher than the requirement for running time. There are also some methods to optimize the running time, for example, changing the size of the input image and the number of proposed regions in the region proposal network. The training of PolishNet-2d on the datasets proposed by this paper can achieve a better result of training 46 per hour on NVIDIA Geforce 1080 GPU.

e: PolishNet-2d STRUCTURAL RATIONALITY ANALYSIS

PolishNet-2d primarily corresponds to the current mainstream two-stage network Faster R-CNN. ResNet101 extracts the features of two-dimensional images. The feature map is then input into the region proposal network to extract the proposed regions. Finally, the proposed regions extracted from the region proposal network are processed by the transformation network X-Net. A large number of experiments show that the pipeline structure of the network has a very good effect on two-dimensional object detection with high accuracy and robustness. At the same time, the introduction of the transformation network X-Net does not interfere with the overall structure of the two-stage network, and X-Net is integrated into the backbone network as a plug-and-play module. In conclusion, the network structure of PolishNet-2d is reasonable, which can guarantee detection accuracy and better robustness for workpiece transformation.

IV. PolishNet-3d

In this paper, a convolutional neural network, PolishNet-3d, is proposed for the three-dimensional industrial workpiece point cloud recognition. The backbone network is PointNet. In order to extract the features of the workpiece with different scales from the three-dimensional point cloud based deep neural network, a hierarchical feature extraction network is proposed, which can extract the corresponding features in different layers for different scales of the workpiece point cloud. R-Net is proposed to learn the rotation parameters of the point cloud. Experiments show that PolishNet-3d can extract feature information with different scales in the point cloud, and it can achieve high recognition accuracy for actual collected point cloud data.

A. NETWORK STRUCTURE

PolishNet-3d, a point cloud recognition network for three-dimensional polishing workpiece, is proposed in this paper. The network consists of the backbone network, PointNet; a hierarchical feature extraction network; and the sub-network, R-Net. Fig. 9 is a data flow diagram for PolishNet-3d to process three-dimensional point cloud data. The original input data is a collection of $n \times 3$ point clouds. After being transformed by R-Net, a transformation matrix with a shape of 3×3 is generated.

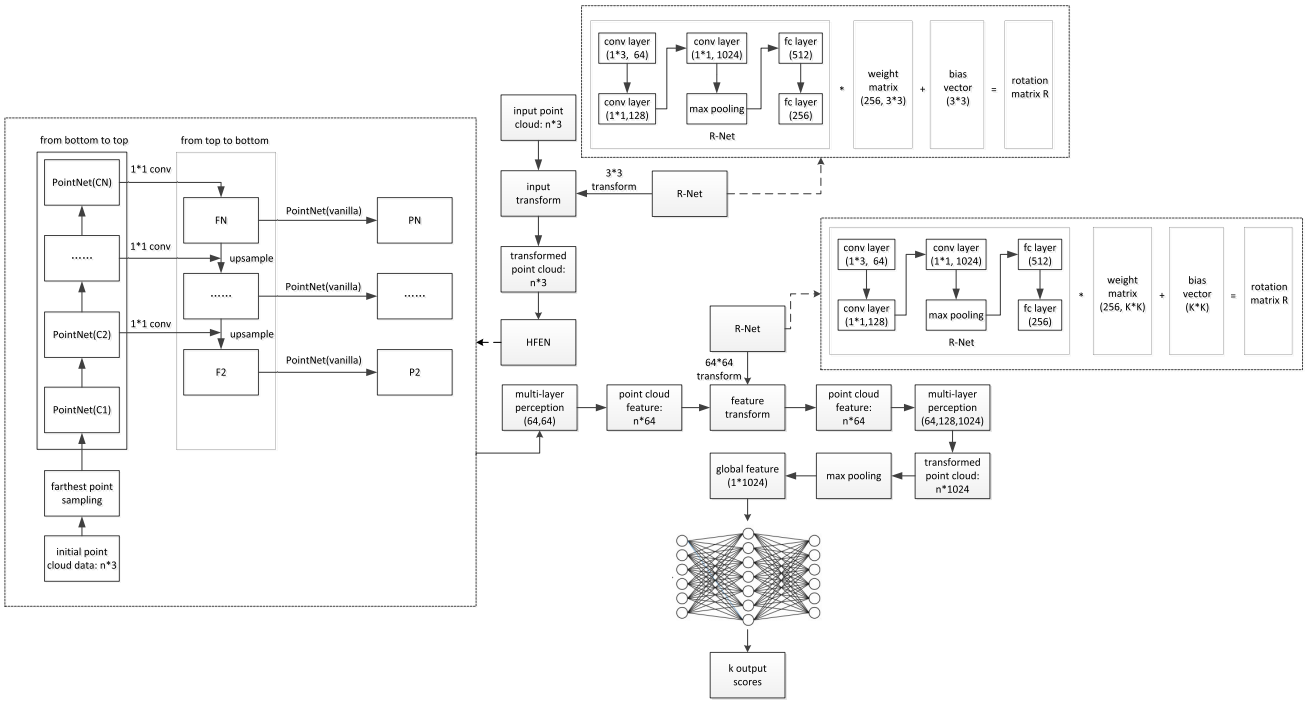

FIGURE 9. Network structure of PolishNet-3d.

Through applying the transformation matrix to the initial input point clouds, a new set of point clouds can be obtained. The new set of point clouds is processed by a multilayer perceptron, and the original three-dimensional point coordinates are transformed into the 64-dimensional high-dimensional feature space to obtain the intermediate data with a shape of $n \times 64$. The 64-dimensional feature vectors are transformed continuously by R-Net, and more normalized 64-dimensional feature vectors are obtained. The multilayer perceptron is used repeatedly to transform 64-dimensional feature vectors into 128-dimensional feature vectors, and then into 1024-dimensional feature vectors. The 1024-dimension feature vectors are symmetrically operated by using a maximum pooling operation, and a 1024-dimension global feature is obtained. The global feature with the shape of 1024-dimension is taken as input, and k output values are obtained through a fully connected network. PolishNet-3d is used to solve the recognition task of the three-dimensional workpiece's point cloud.

1) PointNet

PointNet [1] is a network architecture for direct end-to-end learning on irregular data such as a point cloud. It is used to handle different three-dimensional vision tasks with a unified framework, including object classification, object segmentation, and semantic scene understanding. This paper takes PointNet as the backbone network.

2) HIERARCHICAL FEATURE EXTRACTION NETWORK

In practical engineering application scenarios, it is possible to have both large and small workpieces. In order to recognize workpieces with different scales, a hierarchical feature extraction network is proposed based on the idea of a feature pyramid network for three-dimensional workpiece point cloud recognition. In this paper, the farthest point sampling algorithm is used to divide the point cloud into different regions. PointNet is used to extract the features of a point cloud in each local area, and a new point set is generated. In the new point set, PointNet is used iteratively to generate a new and smaller point set. In the learning process, the invariance of translation can be achieved by using a local coordinate system, and the invariance of permutation can be guaranteed by using PointNet in the local area. After the point cloud feature extraction of several layers, the density of the point cloud decreases gradually, and the features of the point cloud extracted from each layer are preserved to form the feature vector of the point cloud layer. In high resolution point cloud features, more accurate position information of the workpiece can be detected, but the semantics information identified is not rich enough. Low-resolution point cloud features can recognize higher-level semantic information, but the detected location information is not very accurate. Therefore, single-scale point cloud features may not be able to identify smaller-scale workpieces. Point cloud features on hierarchical features are sampled to obtain higher resolution point cloud features. Fig. 9 shows the network structure of the hierarchical feature extraction network in PolishNet-3d.

3) R-Net

In order to ensure that the three-dimensional point cloud recognition system can identify the type of point cloud stably and robustly, regardless of the pose in which the workpieces

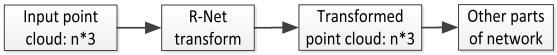

FIGURE 10. Initial feature vector data flow diagram of R-Net.

are located, the rotation transformation sub-network, R-Net is used. R-Net can learn the rotation transformation of point cloud data through training data. For point cloud data, it can rotate around three axes in space:

$$R_x(\alpha) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos\alpha & -sin\alpha & 0 \\ 0 & sin\alpha & cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6}$$

$$R_y(\beta) = \begin{bmatrix} cos\beta & 0 & sin\beta & 0 \\ 0 & 1 & 0 & 0 \\ -sin\beta & 0 & cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{7}$$

$$R_z(\gamma) = \begin{bmatrix} cos\gamma & -sin\gamma & 0 & 0 \\ sin\gamma & cos\gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8}$$

Therefore the total rotation matrix is:

$$R_{xyz}(\alpha, \beta, \gamma) = R_x(\alpha) \times R_y(\beta) \times R_z(\gamma) \tag{9}$$

In order to ensure the orthogonality of the rotation matrix, a regular term is added after the softmax loss function when R-Net is used to optimize the rotation matrix.

$$H'_{y'}(y) = -\frac{1}{N}\sum_i [y'_i log(y_i) + (1 - y'_i) log(1 - y_i)] + \alpha L_{reg} \tag{10}$$

$$H_{y'}(y) = -\frac{1}{N}\sum_i [y'_i log(y_i) + (1 - y'_i) log(1 - y_i)] \tag{11}$$

$$L_{reg} = ||I - RR^T||_F^2 \tag{12}$$

$$y_i = softmax(y)_i = \frac{exp(y_i)}{\sum_j exp(y_j)} \tag{13}$$

Among them, $H'y'(y)$ increases the cross-entropy loss value of the regular constraints, $Hy'(y)$ is the original cross-entropy loss value, α is the coefficient of the regular penalty term, L_{reg} is the regular constraint term in the form of 2-norm, R is the rotation matrix, y_i is the predictive label after the softmax function, and y'_i is the ground truth label. Fig. 10 shows a data flow diagram of R-Net. The network structure detail of R-Net is shown in Fig. 9, and the detailed network parameters of R-Net are listed in Table 3.

R-Net cannot only transform the original point cloud, but also process the feature vectors from the middle layer of the network. If the dimension of the feature vector in the middle layer is K, then R-Net can generate a transformation matrix with a dimension of size $K \times K$, as shown in Fig. 11.

TABLE 3. Network structural design parameters of R-Net.

Network layer name	Conv layer 1	Conv layer 2	Conv layer 3
Conv kernel size	1×3	1×1	1×1
Conv layer output channel number	64	128	1024
Conv stride	[1, 1]	[1, 1]	[1, 1]
Network layer name	max-pooling layer	fc layer	fc layer
Input channel number	N_{point}	1024	512
Output channel number	1	512	256

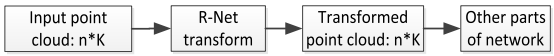

FIGURE 11. Intermediate feature vector data flow diagram of R-Net.

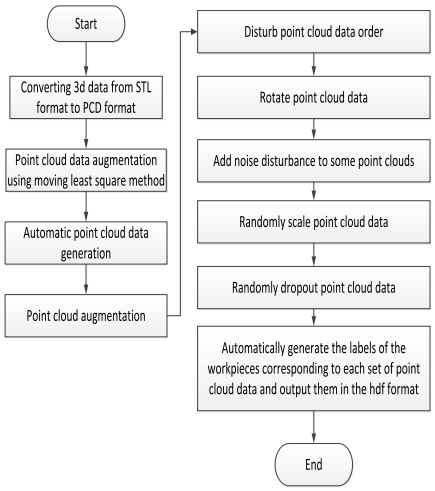

FIGURE 12. Flow chart for automatically constructing 3d workpiece dataset.

4) DATASET

Until now, there are very few public datasets of three-dimensional workpiece CAD models for the manufacturing field. For this paper, polishing workpiece datasets have been constructed, including theoretical CAD models and a large number of RGB-D images of real workpieces collected in actual industrial scenes. After point cloud data format conversions and point cloud augmentation, a three-dimensional polishing workpiece point cloud dataset is generated. Fig. 12 shows the flow chart for automatically constructing 3d workpiece dataset. Fig. 13 shows the original three-dimensional model of the workpiece. Fig. 14 and Fig. 15 show the point cloud data transformed from the theoretical model of the three-dimensional workpiece and the actual collected point cloud data of the workpiece, respectively.

B. ANALYSIS OF EXPERIMENTAL RESULTS

In this paper, the results of original PointNet, of PolishNet-3d added with the hierarchical feature extraction network, and of PolishNet-3d added with the hierarchical feature extraction

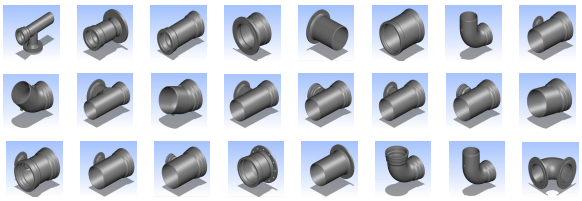

FIGURE 13. Original 3d workpiece model.

FIGURE 14. Typical samples of theoretical point cloud dataset.

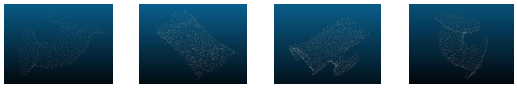

FIGURE 15. Typical samples of real workpieces point cloud dataset.

TABLE 4. Test results of proposed network.

	Average test error	Test average classification accuracy
PointNet(vanilla)	1.622	0.643
PolishNet-3d(HFEN)	0.927	0.776
PolishNet-3d(HFEN+R-Net)	0.128	0.955

network and with R-Net are listed. The test accuracy and error data of 84 polishing workpieces in different networks are shown in Table 4 and Fig. 16. Among them, the formula for calculating the average error value is as follows:

$$AE = \frac{\sum loss}{N_{batch}} \tag{14}$$

The formula for calculating the accuracy value is as follows:

$$A = \frac{N_{TotalCorrect}}{N_{TotalSeen}} \tag{15}$$

The formula for calculating the average classification accuracy value is as follows:

$$ACE = \frac{\sum \frac{N_{TotalCorrectClass}}{TotalSeenClass}}{N_{batch}} \tag{16}$$

Fig. 16 shows the test accuracy of each workpiece in different networks. Due to the length of this article, only the test results of the first 15 types of workpieces are shown here. Table 4 shows that with the addition of the hierarchical feature extraction network and of R-Net, the average error of the PolishNet-3d test decreases rapidly, and the average classification accuracy of the PolishNet-3d test increases rapidly.

The hierarchical feature extraction network enables PolishNet-3d to learn the point cloud features at different resolutions, while the transformation network R-Net can learn the corresponding rotation parameters of point cloud

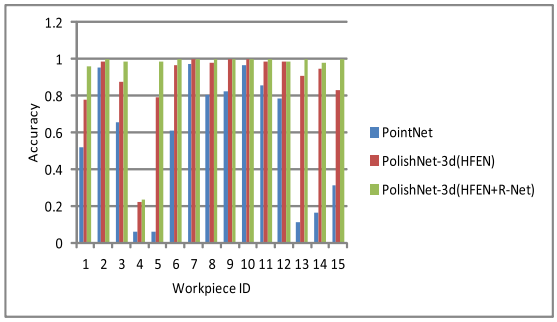

FIGURE 16. Test accuracy and error data of PolishNet-3d.

data through unsupervised learning, which makes point cloud recognition under different positions and poses more robust, thus gradually achieving a higher point cloud recognition rate. Fig. 16 shows the test accuracy of the first 15 types of workpieces. Among them, the blue data represent the recognition rate of different workpieces trained with the original PointNet network, the red data represent the recognition rate of different workpieces trained with the PolishNet-3d network combined with the hierarchical feature extraction network; and the green data represent the recognition rate of different workpieces trained with the PolishNet-3d network, which is combined with the hierarchical feature extraction network and with R-Net. Consistent with the data in Table 4, PolishNet-3d achieves the highest point cloud recognition rate.

I. Train with theoretical point cloud data and test with real point cloud data

The three-dimensional polishing workpiece CAD model is used to create the sampled point cloud data. After 5 augmentation methods have been performed on the point cloud data, the total number of samples in the dataset is 25, 200, all of which are used as training set data. The point cloud data collected in the real environment are used as the testing dataset. The initial number of samples in the testing dataset is 1, 500. The number of samples is expanded to 7, 500 through the same 5 augmentation methods. The results of training with the theoretical point cloud data and testing with the real point cloud are shown in Table 5.

II. Train with theoretical point cloud data and real collected point cloud data and test with real point cloud data

Among them, there are 25, 200 samples of theoretical point cloud data and 1, 500 samples of real collected point cloud data. After 5 augmentation methods are applied on the point cloud data, the number of samples is expanded to 37, 500. Of this, 20, 000 point cloud data are mixed with theoretical point cloud data for training, and the remaining 17, 500 point cloud data are used as the testing set.

Table 5 shows the accuracy value of the test set under different training methods. From the results in the table, it can be seen that the average classification accuracy can reach 0.9726 while using the theoretical point cloud data for training and testing on the actual collected point cloud data. This correct rate basically meets the requirements in the actual

TABLE 5. Point cloud recognition result on the testing dataset under different training modes.

		Theoretical data	Mixed data
Train set	Theoretical data	25, 200	25, 200
	Actual data	0	20, 000
Test set	Actual data	7, 500	17, 500
Average test error		0.0593	0.0421
Test accuracy		0.9793	0.9932
Test average classification accuracy		0.9726	0.9925

industrial field. The reason why the network trained with only the theoretical point cloud data can achieve such good testing results on real point cloud data is that: (1) Unlike RGB image data, point cloud data only contains the coordinate information of the workpieces, and coordinate information will not change according to the change of external light, or the different surface materials; (2) In this paper, the point cloud data obtained from the theoretical CAD model is augmented with 5 methods, so that PolishNet-3d can be robust to noise; differences in scale, translation, and rotation; and other changes in the actual point cloud data.

V. EXPERIMENTS AND RESULTS

A. NETWORK WORKFLOW DIAGRAM

For the workpiece in the industrial environment, it is difficult to extract rich key points and descriptors because of the low texture features of the workpiece surfaces, thus contour information gathering is an important task. We extract the contour in the workpiece image, and then use the contour information to detect the workpiece. Combining the previous two-dimensional workpiece detection network, PolishNet-2d; the three-dimensional workpiece point cloud recognition network, PolishNet-3d; the random sample consensus (RANSAC) based spatial point cloud segmentation method; and a new gray contour image and depth contour image fusion algorithm, the workpiece images captured in the actual industrial scenes are processed to obtain the two-dimensional workpiece detection and the three-dimensional workpiece point cloud recognition. Fig. 17 shows the algorithm structure of low texture three-dimensional workpiece detection and recognition based on the spatial plane segmentation method.

1) RANSAC BASED SPATIAL POINT CLOUD SEGMENTATION ALGORITHM

a The spatial plane equation is as follows:

$$Ax + By + Cz + D = 0 \tag{17}$$

b Randomly select three points from the point cloud, $P_1\{x_1, y_1, z_1\}$, $P_2\{x_2, y_2, z_2\}$ and $P_3\{x_3, y_3, z_3\}$. The spatial plane parameters can be calculated using these three point coordinates.

$$\begin{cases} A = y_1(z_2 - z_3) + y_2(z_3 - z_1) + y_3(z_1 - z_2) \\ B = z_1(x_2 - x_3) + z_2(x_3 - x_1) + z_3(x_1 - x_2) \\ C = x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2) \\ D = -[x_1(y_2z_3 - y_3z_2) \\ \quad + x_2(y_3z_1 - y_1z_3) + x_3(y_1z_2 - y_2z_1)] \end{cases} \tag{18}$$

c Iterate k times. When the number of the points whose distances to the spatial plane are less than the threshold value reach to the maximum value, record the spatial plane parameters. Those parameters are the best ones.

d Segment all the points beyond the spatial plane. For every spatial point $P_i\{x_i, y_i, z_i\}$, the distance between this point and the spatial plane is:

$$d_i = \frac{|Ax_i + By_i + Cz_i + D|}{\sqrt{A^2 + B^2 + C^2}} \tag{19}$$

e If d_i is larger than d_{thresh}, then insert the $i'th$ point into the vector; if d_i is less than d_{thresh}, then discard this point.

f Return the final point cloud set after the segmentation process.

For the low texture workpieces, the most important local feature is contour information, whereby the shape of the

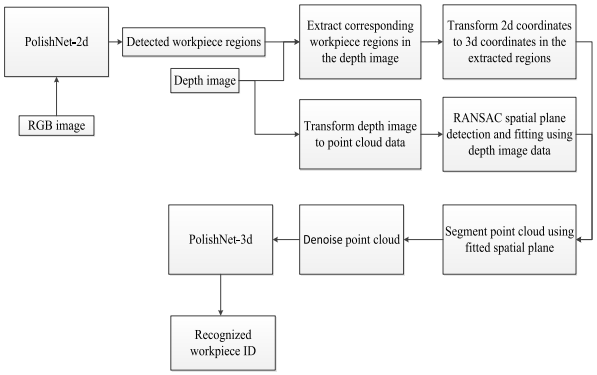

FIGURE 17. Structure diagram of low texture three-dimensional workpiece detection and recognition system based on spatial plane segmentation.

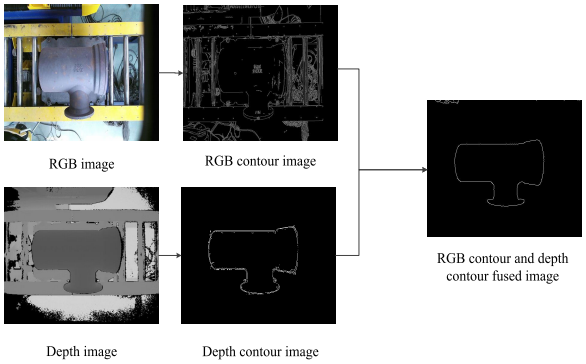

FIGURE 18. Schema of the fusion algorithm between color image and depth image.

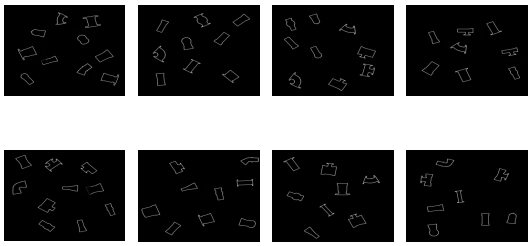

FIGURE 19. Multiple workpiece contour image dataset.

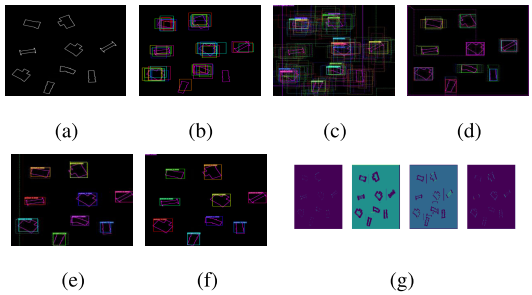

FIGURE 20. Training statistics data of a multiple workpiece contour image dataset.

workpiece can be ascertained. In this paper, based on contour features, a method of RGB image and depth image fusion is proposed. Firstly, the contours in the RGB image and in the depth image of the three-dimensional workpiece are extracted separately. The contour images obtained from the RGB image and from the depth image are fused. Secondly, the fused contour image is input into PolishNet-2d for workpiece detection. Experiments show that the fusion method can effectively solve the problem that the features in the RGB image are vague, due to the low texture characteristics of the workpiece. It can also eliminate the impact of different illumination conditions and different object surface reflective intensities. Fig. 18 shows an example of the fusion of the RGB contour image and the depth contour image.

B. DATASET

PolishNet-2d is trained and tested with the single workpiece contour dataset and with the multiple workpiece contour dataset. Fig. 19 shows the typical images in the multiple workpiece contour dataset. The tangent vector based iterative contour completion algorithm is used to fuse the information of the RGB and of the depth contour images, so as to achieve a higher detection rate.

C. EXPERIMENTAL RESULTS AND ANALYSIS

I. Experimental results and analysis of two-dimensional workpiece contour image detection

In this paper, the experiment of two-dimensional workpiece contour image detection is performed. The experimental results between PolishNet-2d and Faster R-CNN are listed in Table 6. There are two cases: one in which there is only one single workpiece in the image and another in which there are multiple workpieces in the image. In the case where there are multiple workpieces in the scene, the accuracy statistics data are obtained in three cases: no occlusion between the workpieces; no motion blur in the workpiece images; mutual occlusion between the workpieces and motion blur in the workpiece images. An image from the multiple workpiece contour dataset is chosen. The intermediate results of the detection process are shown in Fig. 20.

In this experiment, the total number of workpieces is 84. In the single workpiece dataset, the number of workpieces in each image is 1. There are 66, 000 images in the training set and 2, 500 images in the testing set. In the multiple workpiece dataset, the number of workpieces in each image is between 1 to 15. There are 66, 000 images in the training set and 2, 500 images in the testing set.

TABLE 6. Accuracy statistical data of workpiece detection experiments.

	Single workpiece	Multiple workpieces	
Number of workpieces	84	84	
Number of workpiece types	1	$1-15$	
Training image number	$66,000$	$66,000$	
Testing image number	$2,500$	$2,500$	
Faster R-CNN accuracy	96.25%	No occlusion or blur	96.33%
		Occlusion	84.50%
		Motion blur	91.25%
PolishNet-2d accuracy	97.50%	No occlusion or blur	97.25%
		Occlusion	88.68%
		Motion blur	95.25%

In Table 6, the detection rates of Faster R-CNN and PolishNet-2d for the single workpiece dataset and for the multiple workpiece dataset are calculated. The detection rate of Faster R-CNN and PolishNet-2d is 96.25% and 97.50%, respectively, when there is only a single workpiece in the image. On the same dataset, the detection rate of PolishNet-2d is higher than that of Faster R-CNN, which shows that R-Net can learn the transformation parameters of the workpieces, making it robust to different transformation parameters of the workpiece in the actual scene. In the case where there are multiple workpieces in the image, both networks can achieve high detection rates when there are no occlusions between the workpieces and no motion blurs of the workpiece images. Among them, the detection rate of Faster R-CNN is 96.33%, and that of PolishNet-2d is 97.25%. When the workpieces are occluded from each other, the detection rate of both networks decrease to a great extent. The detection rates of Faster R-CNN and PolishNet-2d are 84.50% and 88.68%, respectively. This result shows that when there are occlusions between multiple workpieces, the contours of the workpieces will be interlaced, which affects the detection rate of the workpiece by the convolutional neural network. When there is motion blurring in the image, the detection rates of

TABLE 7. Accuracy statistical data of the three-dimensional workpiece point cloud recognition experiment.

	Single workpiece in the image	Multiple workpieces in the image
Number of workpieces	84	84
Training point clouds	25, 200	25, 200
Testing point clouds	1, 500	1, 500
Testing images	250	150
PointNet accuracy	90.20%	88.33%
PointNet++ accuracy	95.85%	92.25%
PolishNet-3d accuracy	97.45%	94.50%

Faster R-CNN and PolishNet-2d are 91.25% and 95.25%, respectively. In the case of motion blurring, although the contours of RGB images are blurred and cannot be extracted accurately and steadily, for the depth image, contour features mainly come from structural changes, so motion blurring will not occur in the depth image. Therefore, the proposed RGB contour image and the depth contour image fusion algorithm can preserve the contour information in the depth contour image well, so the two networks still maintain a high detection rate in the case of motion blurring.

II. RANSAC based three-dimensional workpiece point cloud recognition experimental results and analysis

This paper compares PolishNet-3d with PointNet and PointNet++ respectively. The point cloud data transformed by 25, 200 theoretical CAD models are used for training PolishNet-3d, PointNet and PointNet++. There were 1, 500 workpiece point clouds collected from actual industrial scenes. In the test set, 250 RGB-D images are used to test the case of where there is only one workpiece in the image, and 150 RGB-D images are used to test the case of where there are multiple workpieces in the image, and the test results are shown in Table 7.

In Table 7, the recognition rates of PointNet, PointNet++, and PolishNet-3d in the cases of where there is only one workpiece in the image and where there are multiple workpieces in the image are calculated. When there is only one workpiece in the image, the recognition rates of PointNet is 90.20%; PointNet++ is 95.85%; and PolishNet-3d is 97.45%. On the same dataset, the recognition rates of PointNet, PointNet++, and PolishNet-3d increase in turn, which shows that the hierarchical feature extraction network and R-Net can learn the point cloud features with different resolutions and poses of the workpieces, so that they can adapt to different transformations of workpieces in actual industrial scenes. When there are multiple workpieces in the image, the recognition rate of PointNet, PointNet++ and PolishNet-3d decrease compared with that of when there is only one workpiece in the image. The reason is that when there are multiple workpieces in the image, the segmentation accuracy of the workpieces from the point cloud data will decrease accordingly, thus affecting the recognition accuracy of the three-dimensional workpiece point cloud. PolishNet-3d doesn't directly recognize all of the point cloud data in the scene. It uses PolishNet-2d to detect two-dimensional workpieces in the contour image, and then it segments, preprocesses, normalizes and recognizes the point cloud data in the detected regions provided by PolishNet-2d. This allows the convolutional neural network to get a higher recognition rate. When PointNet, PointNet++ and PolishNet-3d process the image with multiple workpieces, the recognition rate decreases slightly compared with the case where there is only one workpiece in the image. Among them, the recognition rate of PointNet is 88.33%; that of PointNet++ is 92.25%; and that of PolishNet-3d is 94.50%.

VI. CONCLUSION AND FUTURE WORK

In this paper, PolishNet-2d and PolishNet-3d are proposed to detect and recognize three-dimensional workpieces in actual industrial scenes. In the workpiece regions detected by PolishNet-2d, the point cloud data transformed from depth images are segmented by the RANSAC algorithm. Finally, the pre-processed point cloud data are processed by PolishNet-3d, and the ID of the workpiece is then obtained.

In the comprehensive experiments of three-dimensional workpiece recognition, different experimental datasets are introduced, including single workpiece contour image datasets and multiple workpiece contour image datasets, as well as point cloud datasets generated by theoretical CAD models and RGB-D images collected in actual industrial scenes. In this paper, Faster R-CNN and PolishNet-2d are used to test the polishing workpiece contour image dataset. PointNet, PointNet++, and the PolishNet-3d are trained with the polishing workpiece dataset, and the testing results are analyzed in this paper. Experiments show that PolishNet-2d and PolishNet-3d proposed in this paper have achieved relatively ideal detection and recognition rates in the experiments of two-dimensional workpiece detection task and three-dimensional workpiece point cloud recognition.

In the future, we will collect multi-view images of the workpieces in the pre-processing stage, and add the function of three-dimensional point cloud reconstruction. Through using the mainstream RGB-D three-dimensional reconstruction framework, the workpieces can be represented more precisely, which will further improve the accuracy of the PolishNet-3d point cloud recognition result.

REFERENCES

[1] C. R. Qi, H. Su, L. J. Guibas, and K. Mo, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 652–660.

[2] S. Song and J. Xiao, "Deep sliding shapes for amodal 3D object detection in RGB-D images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 808–816.

[3] X. Chen, H. Ma, B. Li, T. Xia, and J. Wan, "Multi-view 3D object detection network for autonomous driving," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1907–1915.

[4] Z. Wu, S. Song, F. Yu, L. Zhang, X. Tang, J. Xiao, and A. Khosla, "3D ShapeNets: A deep representation for volumetric shapes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1912–1920.

[5] Y. Li, S. Pirk, C. R. Qi, L. J. Guibas, and H. Su, "FPNN: Field probing neural networks for 3D data," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 307–315.

[6] H. Su, S. Maji, E. Learned-Miller, and E. Kalogerakis, "Multi-view convolutional neural networks for 3D shape recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 945–953.
[7] A. Kanezaki, Y. Nishida, and Y. Matsushita, "RotationNet: Joint learning of object classification and viewpoint estimation using unaligned 3D object dataset," presented at the IEEE Conf. Comput. Vis. Pattern Recognit., 2018.
[8] W. Kehl, F. Tombari, S. Ilic, V. Lepetit, and N. Navab, "Hashmod: A hashing method for scalable 3D object detection," presented at the Brit. Mach. Vis. Conf., 2015.
[9] W. Kehl, F. Milletari, S. Ilic, N. Navab, and F. Tombari, "Deep learning of local RGB-D patches for 3D object detection and 6D pose estimation," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 205–220.
[10] C. R. Qi, W. Liu, and C. Wu, "Frustum PointNets for 3D object detection from RGB-D data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 918–927.
[11] G. Riegler, A. O. Ulusoy, and A. Geiger, "OctNet: Learning deep 3D representations at high resolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 3577–3586.
[12] W. Kehl, F. Manhardt, S. Ilic, N. Navab, and F. Tombari, "SSD-6D: Making RGB-based 3D detection and 6D pose estimation great again," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 1521–1529.
[13] A. Krizhevsky and I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[14] W. Ouyang, X. Zeng, X. Wang, S. Qiu, P. Luo, Y. Tian, H. Li, S. Yang, Z. Wang, C.-C. Loy, and X. Tang, "DeepID-Net: Deformable deep convolutional neural networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2403–2412.
[15] J. Deng, W. Dong, L.-J. Li, K. Li, L. Fei-Fei, and R. Socher, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[16] T.-Y. Lin, M. Maire, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, and S. Belongie, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.
[17] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 2553–2561.
[18] P. Sermanet, D. Eigen, M. Mathieu, R. Fergus, Y. LeCun, and X. Zhang, "OverFeat: Integrated recognition, localization and detection using convolutional networks," presented at the Int. Conf. Learn. Representations, 2013.
[19] D. Erhan, C. Szegedy, D. Anguelov, and A. Toshev, "Scalable object detection using deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2147–2154.
[20] R. Girshick, J. Donahue, J. Malik, and T. Darrell, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
[21] S. Ren, K. He, J. Sun, and R. Girshick, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
[22] B. Zhou, A. Khosla, A. Oliva, A. Torralba, and A. Lapedriza, "Object detectors emerge in deep scene CNNs," presented at the Int. Conf. Learn. Represent., 2015.
[23] R. Klokov and V. Lempitsky, "Escape from cells: Deep Kd-networks for the recognition of 3D point cloud models," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 863–872.
[24] J. Li, B. M. Chen, and G. H. Lee, "SO-Net: Self-organizing network for point cloud analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9397–9406.
[25] M. C. Wu and S. R. Jen, "A neural network approach to the classification of 3D prismatic parts," *Int. J. Adv. Manuf. Technol.*, vol. 11, no. 5, pp. 325–335, Sep. 1996.
[26] C. Y. Ip, W. C. Regli, A. C. Shokoufandeh, and L. Sieger, "Automated learning of model classifications," in *Proc. 8th ACM Symp. Solid Modeling Appl.*, Jun. 2003, pp. 322–327.
[27] C. Y. Ip and W. C. Regli, "Content-based classification of CAD models with supervised learning," *Comput.-Aided Des. Appl.*, vol. 2, no. 5, pp. 609–617, 2005.
[28] C. Y. Ip and W. C. Regli, "Manufacturing classification of CAD models using curvature and SVMs," in *Proc. Int. Conf. Shape Modeling Appl.*, Jun. 2005, pp. 361–365.
[29] K. He, X. Zhang, J. Sun, and S. Ren, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 630–645.

FUQIANG LIU received the M.S. degree in computer technology from Harbin Engineering University, Harbin, China, in 2013, where he is currently pursuing the Ph.D. degree in control science and engineering with the College of Automation. His research interests include computer vision, deep learning, artificial intelligent, deep neural networks, SLAM, and robotics.

ZONGYI WANG received the Ph.D. degree in control theory and control engineering from Harbin Engineering University, Harbin, China, in 2005, where he is currently a Professor with the College of Automation. His research interests include computer vision, robotics, welding, and cutting intelligence. He was a recipient of the First Prize of the Heilongjiang Provincial Scientific and Technological Progress Award, in 2004.