Measurement 183 (2021) 109821

Contents lists available at ScienceDirect

Measurement

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

BV-Net: Bin-based Vector-predicted Network for tubular solder joint detection☆

Chenlin Zhou [a,b], Xiaofei Shen [a,b], Peng Wang [a,b,c,*], Wei Wei [a,b], Jia Sun [a], Yongkang Luo [a], Yiming Li [a,b]

[a] Institute of Automation, Chinese Academy of Sciences, Beijing, China
[b] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China
[c] CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai, China

ARTICLE INFO

Keywords:
Object detection
Defect detection
Quality inspection
Tubular solder joint detection
Deep learning

ABSTRACT

Tubular solder joint detection is an important and challengeable issue in industry, due to the small objects, rarely collected datasets and real-time and high precision requirements. Traditional methods on defect detection cannot solve tubular solder joint detection due to lacking of angle estimation. In this paper, we propose a tubular solder joint detection method named Bin-based Vector-predicted Network (BV-Net), which combines the framework of state-of-the-art deep-learning-based object detector (YOLOv4) with specific characteristics and requirements of tubular solder joint detection. BV-Net could effectively estimate both the center point and the direction of tubular solder joints. Firstly, To regress the center point, we propose a Circle-based Distance-Intersection over Union (CirDIoU) loss, which gets better learning performance for the center point of tubular solder joint than Distance-Intersection over Union (DIOU) loss. Secondly, to estimate the direction, we introduce a bin-based angle regression method, which transforms a regression task into a classification and regression task, improving the precision of direction estimation greatly. Thirdly, we establish a tubular solder joint dataset and design a new evaluation index: mAP (δ_d, δ_θ) for tubular solder joint detection, combining the relative deviation of center point positioning δ_d and the relative deviation of angle regression δ_θ. Finally, comparison experiments on the dataset are carried out. BV-Net achieved 85.5% mAP (0.5%, 3%) with 34.4 FPS, meeting the requirements of industrial system. In direction estimation, bin-based angle regression method promotes 4.3% mAP (-, 3%), compared with the baseline. In center point positioning, BV-Net outperforms YOLOv4 by an improvement of 0.4% mAP (0.5%, -). The experimental results verified the effectiveness of our method.

1. Introduction

Tubular solder joint detection usually comes from industrial defect detection or quality inspection. To prevent refrigerant from leakage, the quality inspection of tubular solder joints in compressor is an important step to ensure the quality of refrigerator production. The traditional quality control process of tubular solder joints performed by inspectors usually demands many workers to operate leak detectors without protection (shown in Fig. 1). Manual quality inspection has the disadvantages of high cost and low efficiency. Therefore, automatic quality inspection of tubular solder joints contributes significantly in the quality control of refrigerator products. As is shown in Fig. 1, tubular solder joints are usually at the junction or at the end of the metal pipe in the compressor of refrigerator. During the quality inspection of automatic system, manipulator needs to obtain the center point coordinates of tubular solder joint, and then move to the vicinity of tubular solder joint. A suction leak detector is installed at the end of the manipulator. In order to avoid or reduce collisions with metal pipe, suction leak detector needs to wrap the tubular solder joint along the metal pipe, and then check the quality of tubular solder joint. Therefore, Tubular solder joint detection in the perception module of manipulator system includes estimating both the center point (x, y) of solder joint and the direction θ of solder joint along the pipe. At present, industrial tubular solder joint detection with various background is a worthwhile area.

Most researches of defect detection related to solder joint detection are about printed circuit board (PCB) solder joints [1–6] and metal

☆ This work was supported in part by the National Natural Science Foundation of China under Grants (91748131, 62006229, 61771471), the Beijing Municipal Natural Science Foundation, China under Grant 4204113, and the Strategic Priority Research Program of Chinese Academy of Science under Grant XDB32050100.
* Corresponding author.
E-mail address: peng_wang@ia.ac.cn (P. Wang).

https://doi.org/10.1016/j.measurement.2021.109821
Received 22 March 2021; Received in revised form 7 June 2021; Accepted 25 June 2021
Available online 29 June 2021

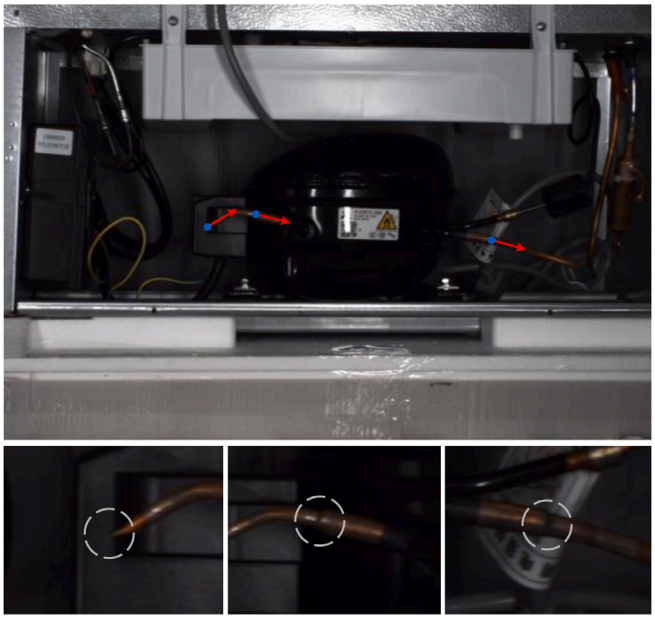


Fig. 1. The results of tubular solder joint detection based on BV-Net. In the first row, blue dot and red arrow respectively represent the center point and direction of tubular solder joint. The second row shows the enlarged view of tubular solder joints.

surface weld [7]. The method on tubular solder joint detection is rarely proposed. Compared with other defect detection tasks, tubular solder joint detection is more challengeable issue. They are small objects, which usually account for less than 1/5000 of the whole image. The background of tubular solder joints is so complex that objects are hard to identify or locate. Furthermore, the detection of manipulator system in industrial production line requires real-time performance and high precision. Moreover, previous defect detection tasks mainly aimed at horizontal bounding box regression which is with the form of (x, y, w, h). (x, y), w, h represent center point coordinate, width and height of horizontal box, respectively. Tubular solder joint detection needs to estimate the direction θ of solder joint along the pipe. And this requirement never appeared in the field of solder joint detection and other defect detection tasks which lead to these current defect detection cannot solve tubular solder joint detection. All in all, these factors bring great difficulties to tubular solder joint detection.

Deep learning methods have achieved good performances on image classification [8] and object detection tasks [9], and have extended to some interdisciplinary fields, such as digital currency forecasting [10], medical image diagnosis [11], wind speed forecasting [12]. They are also widely used in industry, such as defect detection, quality inspection. A fast intelligent inspection model based on the YOLO v3 is presented for exterior substance inspection [13]. DeepInspection, a novel end-to-end attention-based fully convolutional neural network framework, is proposed for automated defect inspection on specular surface [14]. [15] proposes a deep convolutional neural network to detect wood features and automatically classify defects from wood images. [16] proposes a generic deep-learning-based approach for automated surface inspection. This method extracts feature using pre-trained deep learning network. Then, it generates the defect heatmap based on patch features, and predicts the defect area by thresholding and segmenting the heatmap. [17] proposes Efficient Defect Detectors (EDDs) for fabric defect detection. EDDs are sufficiently optimized with consideration of the characteristics of fabric surface images, such as resolution, defect appearance, etc. Because of powerful feature extraction and feature expression capability, we reasonably believe that deep learning methods have great advantages in tubular solder joint detection.

In general object detection, deep learning method YOLOv4 [18] is the state-of-the-art object detector which has optimal speed and accuracy. But YOLOv4 cannot be directly applied in tubular solder joint detection, because of lacking the mechanism which could estimate both the center point and the direction of a general object, and lacking the evaluation index for tubular solder joint detection. To deal with these problems, we propose a tubular solder joint detection method named Bin-based Vector-predicted Network (BV-Net), which combines the framework of YOLOv4 with the specific characteristics and requirements of tubular solder joint detection. BV-Net discards the anchor mechanism of YOLOv4 and adds new loss function and evaluation index for tubular solder joint detection. It could effectively estimate both the center point and the direction of tubular solder joints. In BV-Net, we propose a circle-based DIOU (CirDIoU) loss, whose circle is more consistent with the geometric information of tubular solder joint than box of DIOU loss. It gets better learning performance for the center point of tubular solder joint than DIOU loss. To estimate the direction of tubular solder joint, we introduce a bin-based angle regression method, which reduces the difficulty of the regression task indirectly and works much better than the way of directly regressing. In addition, we establish a tubular solder joint dataset and design a new evaluation index: mAP $(\delta_d, \delta_\theta)$ for tubular solder joint detection, which combines the relative deviation of center point positioning δ_d and the relative deviation of angle regression δ_θ. We conduct sufficient experiments to verify the effectiveness of our method, BV-Net achieves 85.5% mAP (0.5%, 3%) with 34.4 FPS on the tubular solder joint dataset, meeting the requirements of industrial system.

The remainder of this paper is organized as follows. Section 2 introduces the related work of this paper. Next, Section 3 describes the details of our proposed BV-Net. Section 4 carries out extensive experiments, where the experimental results and related analysis are provided. Finally, Section 5 gives the conclusion of this paper.

2. Related work

The work related to tubular solder joint detection can be summarized into the following three parts: object detection models, solder joint detection and introduction of YOLOv4.

2.1. Object detection models

Most object detection models based on deep learning can be divided into anchor-based models and anchor-free models. Anchor-based models are usually categorized into one-stage models and two-stage models. The core idea of two-stage models is generating candidate box firstly, and then using Convolution Neural Network (CNN) to regress and classify, the most representative two-stage model is the R-CNN [9] series, including fast R-CNN [19], faster R-CNN [20], R-FCN [21], and Libra R-CNN [22]. One-stage models directly use CNN to regress and classify. The most representative one-stage models include YOLO [23,24,18], SSD [25], and RetinaNet [26]. Anchor-free models include CenterNet [27], CornerNet [28], FCOS [29], RepPoints [30], etc. Compared with anchor-free models, anchor-based models need extra prior knowledge for parameter configuration of anchor in general. Two-stage models are generally more accurate but slower than one-stage models, therefore, they are hard to apply in industrial scenes with high real-time requirement. YOLO series have been the most widely used detection algorithm in industry. Meanwhile, YOLOv4 is the state-of-the-art model which has optimal speed and accuracy on the MS COCO dataset [31]. The proposed method BV-Net is based on YOLOv4, but belongs to anchor-free one-stage object detection model.

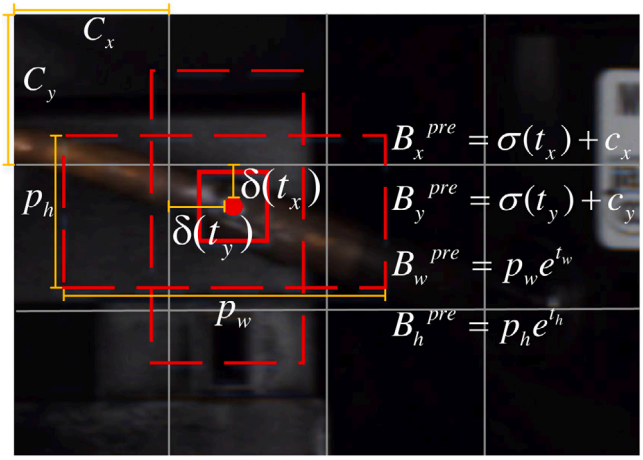


Fig. 2. Bounding box prediction process with anchor in YOLOv4.

2.2. Solder joint detection

Most researches related to solder joint detection are about PCB solder joint and metal surface weld. PCB solder joint detection in a visual way mainly is applied to the quality inspection of solder joint. It is essentially a classification task of different defect types, which represent the different types of connection between component and solder joint on the PCB after mounting. The main methods are template matching methods and deep learning algorithms [3–5]. Metal surface weld detection which is also applied for quality inspection, but its main role is to locate weld, then, some special instruments like ultrasonic flaw detector are used to inspect the quality of weld. The background of metal surface weld is usually simple, and it is easy to identify or locate. The main methods are traditional structured localization methods and deep learning algorithms [6,7]. Tubular solder joint detection includes estimating both the center point and the direction of tubular solder joint. The background of tubular solder joint is complex and the target is very small. Traditional structured location methods are difficult to solve, and advanced detection algorithms are urgently needed.

2.3. YOLOv4

One-stage object detection models have been widely used in practice. YOLOv4 is the latest, most advanced iteration of YOLO series which is one of the most famous series in one-stage models. In YOLO series, input image is divided into $S \times S$ grids, and each grid is responsible for the detection of an object. As is shown in Fig. 2, YOLOv4 predicts 4 coordinates for each anchor which is a preset box in each grid, t_x, t_y, t_w, t_h. Respectively, (t_x, t_y) and (t_w, t_h) represent the network-predicted center offset of the anchor and the scaling of width and height. If the grid is offset from the top left corner of the image by (c_x, c_y) and the anchor prior has width and height p_w, p_h. The conversion from initial anchor to predicted box can be described as $B_x^{pre} = \sigma(t_x) + c_x$, $B_y^{pre} = \sigma(t_y) + c_y$, $B_w^{pre} = p_w e^{t_w}$, $B_h^{pre} = p_h e^{t_h}$. $\sigma(\cdot)$ is activation function.

For feature extraction, the backbone of YOLOv4 use CSPDarknet53 [32], which combines the residual block in Darknet53 with the CSP model. The neck of YOLOv4 are SPP [33] and PANet [34]. SPP is used behind CSPDarknet-53 as it increases the receptive field, differentiates the most significant feature and does not cause a reduction in speed [33]. PANet [34] is used for enriching the semantic and geometric information of extracted features by concatenating lower-level features maps with higher-level feature maps, A bottom-up path is also added to make it easier to disseminate information at lower levels. YOLOv4 uses the way of YOLOv3 which has three heads for multi-scale prediction.

Apart from the architecture, YOLOv4 uses some training strategies to get better accuracy with little or no computation cost during inference. These strategies include some new data augmentation methods like CutMix, Mosaic, etc., selection of hyper-parameters while applying general algorithms, and a new activation technique called Mish activation, etc.

3. Proposed method

In this section, the details of our method are discussed. It contains BV-Net, loss function and evaluation index for tubular solder joint detection.

3.1. BV-Net

The label of a tubular solder joint is (V_x, V_y, V_θ). V_θ is encoded as bin classification target bin_θ : $[bin_\theta^1, bin_\theta^2, \ldots, bin_\theta^m]$ and residual regression target res_θ : $[res_\theta^1, res_\theta^2, \ldots, res_\theta^m]$. The output of BV-Net is a form of vector: $(V_x^{pre}, V_y^{pre}, V_\theta^{pre})$. As is shown in Fig. 4, the prediction process of BV-Net can be described as $V_x^{pre} = \sigma(t_x) + c_x$, $V_y^{pre} = \sigma(t_y) + c_y$, $V_\theta^{pre} = g(\hat{bin}_\theta, \hat{res}_\theta)$. t_x, t_y, c_x, c_y, $\sigma(\cdot)$ are the same in YOLOv4. $g(\cdot,\cdot)$ is shown in formula (9), and indicates the decoding process of angle prediction.

The pipeline of BV-Net is shown in Fig. 3. The structure of BV-Net is based on YOLOv4, containing CSPDarknet53 [32], SPP [33] and PANet [34]. CSPDarknet53 is a powerful feature extraction module, which is used to extract the features of tubular solder joints. SPP and PANet belong to feature fusion module. They are used to enrich the expression ability of feature maps, which can further improve the detection ability of tubular solder joints. YOLOv4 uses three heads for multi-scale prediction due to the sparsity in object sizes. Different from YOLOv4, object sizes relatively regular in tubular solder joint detection, therefore, BV-Net only uses one head for prediction. In addition, BV-Net uses multi-scale training to learn multi-scale features, and uses Non Maximum Suppression (NMS) in post-processing. In NMS, object confidence threshold and IoU threshold are set as 0.2 and 0.4 separately. For improving the object detection training, BV-Net uses Mish activation, and also uses image flip, CutMix and Mosaic in data augmentation.

The loss function of BV-Net includes location loss of center point, classification loss, confidence loss, and direction loss of tubular solder joint.

$$loss_{vector} = loss_{loc} + loss_{class} + loss_{conf} + loss_{dir} \tag{1}$$

In our model, the classification loss uses softmax cross-entropy and confidence loss uses binary cross entropy.

3.2. Location loss

To regress a point (x, y), Mean Squared Error (MSE) loss is widely used in CNN-based model. Although tubular solder joint is very small and we only focus on its center point in industrial system, regression with a certain area is helpful for the regression of the point from the perspective of detection.

In general object detection task, Intersection over Union (IoU) loss, GIoU [35] loss, DIoU [36] loss are widely used, and they are all based on box. CircleNet [37] proposes circle IoU (CirIoU) for detection of the ball-shaped glomerulus. In tubular solder joint detection, objects are small and the boundary is difficult to define precisely, circle is more consistent with the geometric information of object than box. Therefore, for small object or round object detection, we extend the way of circle IoU to other IoU-based loss, and propose Circle-based DIoU (CirDIoU) loss and Circle-based GIoU (CirGIoU) loss, respectively recorded as $L_{CirDIoU}$ and $L_{CirGIoU}$. The difference between IoU and

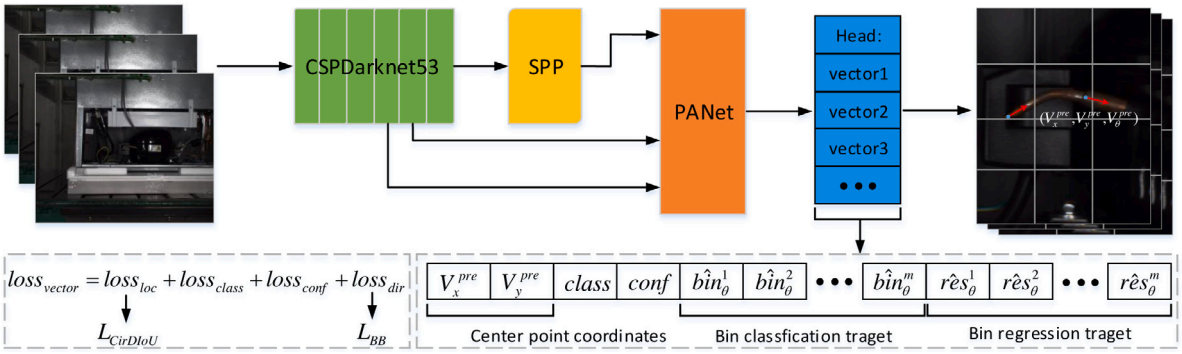


Fig. 3. The pipeline of BV-Net. The prediction of BV-Net is vector containing center point positioning and angle regression of tubular solder joints. V_θ^{pre} can be decoded from bin classification target $\hat{bin}_\theta : [\hat{bin}_\theta^1, \hat{bin}_\theta^2, \dots, \hat{bin}_\theta^m]$ and residual regression target $\hat{res}_\theta : [\hat{res}_\theta^1, \hat{res}_\theta^2, \dots, \hat{res}_\theta^m]$ by formula (9). Center point (V_x^{pre}, V_y^{pre}) positioning and angle regression respectively use CirDIoU loss ($L_{CirDIoU}$) based on fixed circle and bin-based loss (L_{BB}). The evaluation index for tubular solder joint detection is mAP (δd, $\delta\theta$).

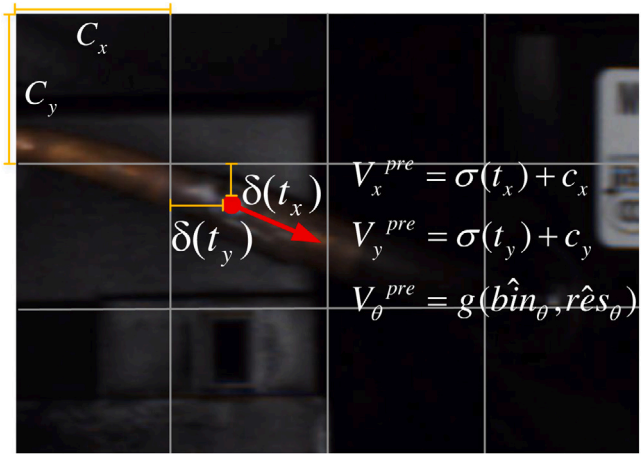


Fig. 4. Vector prediction process without anchor in BV-Net.

CirIoU is shown in Fig. 5, and these mathematical expressions are as follow:

$$L_{CirIoU} = 1 - \frac{|C \cap C^{gt}|}{|C \cup C^{gt}|} \tag{2}$$

where $C^{gt} = (x^{gt}, y^{gt}, r^{gt})$ is the ground-truth, and $C = (x, y, r)$ is the predicted circle.

$$L_{CirGIoU} = 1 - \frac{|C \cap C^{gt}|}{|C \cup C^{gt}|} + \frac{|C_1 - C \cup C^{gt}|}{|C_1|} \tag{3}$$

where C_1 is the smallest circle covering C and C^{gt}.

$$L_{CirDIoU} = 1 - \frac{|C \cap C^{gt}|}{|C \cup C^{gt}|} + \frac{\rho^2(c, c^{gt})}{d^2} \tag{4}$$

where c and c^{gt} denote the center points of C and C^{gt}. $\rho(\cdot)$ is the Euclidean distance, and d is the diameter of the smallest enclosing circle covering the two circles.

In tubular solder joint detection, the boundary of solder joint is difficult to define precisely and we only focus on its center point in industrial system. Therefore, for the center point positioning of tubular solder joint, we chose CirDIoU loss based on fixed circle. That is $r = r^{gt} = R$, center point (x, y) is encoded as (x, y, R), R is a hyperparameter.

3.3. Direction loss

Directly regressing the direction of tubular solder joint is a difficult regression task, experiments demonstrated that way does not work well. To regress the direction of tubular solder joint, we introduce a bin-based angle regression method [38] from 3D detection field. It transforms a regression task into a classification and regression task.

In tubular solder joint detection, the direction is defined as solder joint along the right side of metal pipe, and the angle range is $[0, \pi]$. As is shown in Fig. 6, we divide angle π into m bins, and calculate bin classification target $bin_\theta : [bin_\theta^1, bin_\theta^2, \dots, bin_\theta^m]$ and residual regression target $res_\theta : [res_\theta^1, res_\theta^2, \dots, res_\theta^m]$. Each residual regression head $res_\theta^j (1 \le j \le m)$ is only responsible for bin_θ^j. m is a hyperparameter. The encoding process of angle θ is shown as follows:

$$bin_\theta^j = \begin{cases} 1 & , j = \lceil \theta/m \rceil \\ 0 & , j \neq \lceil \theta/m \rceil \end{cases} \tag{5}$$

$$res_\theta^j = \begin{cases} \theta \% m & , j = \lceil \theta/m \rceil \\ 0 & , j \neq \lceil \theta/m \rceil \end{cases} \tag{6}$$

In loss function, bin classification target uses cross-entropy cost function, residual regression target uses mean squared error cost function. The mathematical expression of bin-based loss (L_{BB}) is as follows:

$$L_{BB} = \frac{1}{N} \sum_i^N \sum_j^m \left[f\left(res_\theta^{ij}, \hat{res}_\theta^{ij}\right) - bin_\theta^{ij} \log\left(\hat{bin}_\theta^{ij}\right) \right] \tag{7}$$

$$f\left(res_\theta^{ij}, \hat{res}_\theta^{ij}\right) = \begin{cases} 0, bin_\theta^{ij} = 0 \\ \left(res_\theta^{ij} - \hat{res}_\theta^{ij}\right)^2, bin_\theta^{ij} = 1 \end{cases} \tag{8}$$

where $\hat{bin}_\theta^{ij}$ and $\hat{res}_\theta^{ij}$ are the predicted bin assignments and residuals of angle, bin_θ^{ij} and res_θ^{ij} are the encoded ground-truth targets, N is the number of samples in a batch.

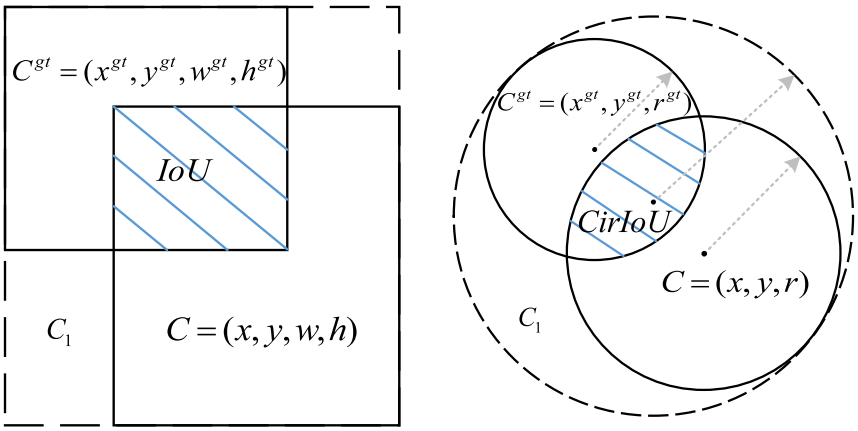
The predicted angle V_θ^{pre} can be decoded from the predicted bin classification target $\hat{bin}_\theta$ and the predicted residual regression target $\hat{res}_\theta$ by the following formula.

$$V_\theta^{pre} = g(\hat{bin}_\theta, \hat{res}_\theta) = (j-1) * (\theta/m) + \hat{res}_\theta^j \tag{9}$$

$$\{ j \mid \forall x : \hat{res}_\theta^x \le \hat{res}_\theta^j; x, j \in [1, 2, \dots, m] \} \tag{10}$$

3.4. Evaluation index for tubular solder joint detection

The evaluation method is a standard to analyze the performance of algorithms. In object detection task, IoU is used to analyze the performance of bounding box regression. IoU is the overlap rate between the candidate (C) and the ground truth bound (G), the mathematical description is $IoU = (C \cap G)/(C \cup G)$. When $IoU \ge threshold$, this



(a) IoU based on box (b) CirIoU based on circle

Fig. 5. Difference between IoU and CirIoU.

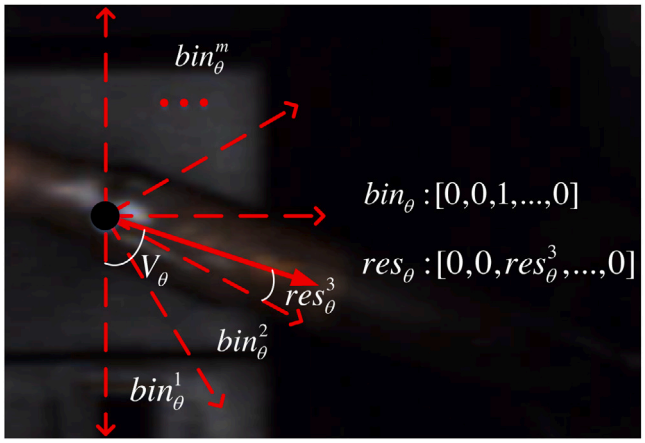


Fig. 6. Bin-based angle regression. V_θ is encoded as bin classification target bin_θ : $[bin_\theta^1, bin_\theta^2, \ldots, bin_\theta^m]$ and residual regression target res_θ : $[res_\theta^1, res_\theta^2, \ldots, res_\theta^m]$.

sample will be determined as True Positive (TP). But in tubular solder joint detection, IoU cannot measure the regression result of a vector. It is necessary to design a new evaluation index for tubular solder joint detection. Only when both center point positioning and angle regression are predicted within the relative deviation δ_d, δ_θ simultaneously, can the object be considered correctly detected. That is, when both formula (10) and formula (11) are met, the target in the sample will be determined as TP. In this way, False Positive (FP) and False Negative (FN) can also be defined.

$$\sqrt{\frac{\left(V_x^{pre} - V_x\right)^2 + \left(V_y^{pre} - V_y\right)^2}{W^2 + H^2}} \le \delta_d \tag{11}$$

$$\frac{\left|V_\theta^{pre} - V_\theta\right|}{180^0} \le \delta_\theta \tag{12}$$

where W, H are the width and height of whole image.

In object detection, main evaluation index Precision (P) and Recall (R) can be calculated by TP, FP and FN. Recall (R) and Precision (P) can be utilized to get Average Precision (AP) value of each category. The mAP is calculated by the mean value of AP over all categories. They are defined as follows:

$$\mathrm{P} = \frac{TP}{(TP + FP)} \tag{13}$$

$$\mathrm{R} = \frac{TP}{(TP + FN)} \tag{14}$$

$$AP = \int_0^1 P(R) dR \tag{15}$$

$$mAP = \frac{1}{N_{cls}} \sum_{i=1}^{N_{cls}} AP_i \tag{16}$$

where P(R) indicates the P–R function [39], N_{cls} represents the number of categories. In tubular solder joint detection, $N_{cls} = 1$. In general, P and R are one-sided for the same detection algorithm. Because of good balance between Precision and Recall, mAP is widely used to evaluate detection algorithms in some well-known public object detection dataset, such as ImageNet [8], PASCAL VOC [40], COCO [41]. In tubular solder joint detection, calculation of TP includes two constraint conditions of center point positioning and angle regression. In order to facilitate the experimental analysis, mAP for tubular solder joint detection is recorded as mAP (δ_d, δ_θ).

4. Experiments

4.1. Dataset

The dataset comes from the scene of industrial production line, we collected images of refrigerator on a production line in a day. The dataset contains 967 images from 967 refrigerators including 2872 tubular solder joints. Among the dataset, 677 pictures are randomly selected for training and 290 pictures for testing. The resolution of each image is 2448 × 2048, and a tubular solder joint is about 750 pixels. The dataset is shown in Fig. 7.

4.2. Experimental setup

All experiments are trained and tested with a 1080 Ti GPU. In post-processing, circle instead of box in NMS which is used in all experiments, the radius of circle is set as $r_0 = \sqrt{750/\pi}$. The training step is 34000. The step decay learning rate scheduling strategy is adopted with initial learning rate 0.01 and multiply with a factor 0.1 at the 25000 steps and the 30000 steps. The momentum and weight decay are respectively set as 0.9 and 0.0005. Batch size is set as 4.

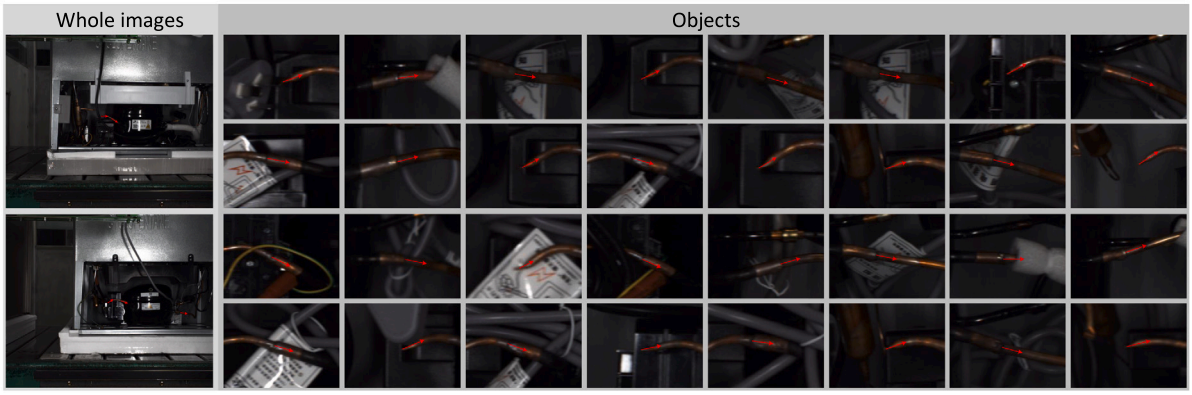


Fig. 7. Examples of tubular solder joint dataset. The dataset contains 967 images including 2872 tubular solder joints. The left part shows the whole images, whose resolution are 2448 × 2048. The right part shows the details of different tubular solder joints. Objects are small and illegible. Meanwhile, the background is complex and low contrast.

Table 1
Using different R for BV-Net training.

R	mAP (0.1%, –)	mAP (0.3%, –)	mAP (0.5%, –)
$0.5 \times r_0$	16.1%	89.9%	97.9%
$1 \times r_0$	19.8%	**92.2%**	98.0%
$2 \times r_0$	**20.1%**	89.5%	98.0%
$3 \times r_0$	15.7%	87.1%	**98.5%**
$4 \times r_0$	15.1%	84.3%	97.6%
$5 \times r_0$	10.6%	79.7%	96.8%

Table 2
Using different location loss for BV-Net training.

$loss_{loc}$	mAP (0.1%, –)	mAP (0.3%, –)	mAP (0.5%, –)
MSE	10.6%	82.3%	97.9%
L_{CirIoU}	19.5%	91.5%	97.8%
$L_{GIoU}(Squ)$	17.6%	89.5%	97.7%
$L_{GIoU}(Rec)$	17.8%	89.6%	97.9%
$L_{CirGIoU}$	**18.0%**	**89.8%**	**97.9%**
$L_{DIoU}(Squ)$	19.5%	92.1%	97.9%
$L_{DIoU}(Rec)$	19.6%	92.0%	97.9%
$L_{CirDIoU}$	**19.8%**	**92.2%**	**98.1%**

4.3. Hyperparameter:R in location loss

In this experiment, location loss of BV-Net uses $L_{CirDIoU}$. R is a hyperparameter that affects the center point positioning of tubular solder joints. The area of the circle with radius $r_0 = \sqrt{750/\pi}$ is approximately equal to the average area of tubular solder joints. We set different values of R to analyze the impact on center point positioning.

The results are shown in Table 1, mAP (∗, –) indicates only the center point positioning of tubular solder joints is concerned in the evaluation index, and the angle regression is ignored. BV-Net gets the best performance in mAP (0.1%, –), mAP (0.3%, –), mAP (0.5%, –) respectively when R = $2 \times r_0$, R = $1 \times r_0$, R = $3 \times r_0$. This experiment shows that including appropriate background information in the circle will help the center point positioning of tubular solder joints.

4.4. Influence of different location loss

In this section, in order to compare with box-based location loss, BV-Net use DIoU loss with two different shapes of fixed boxes, square and rectangle with an aspect ratio of 2: 1. These boxes have the same area with fixed circle which is used in $L_{CirDIoU}$. In this way, comparison experiments of $L_{CirGIoU}$ and L_{CirIoU} can be also established. R is set as $1 \times r_0$.

Table 3
Using different m for BV-Net training. The way of directly regressing (DR) uses mean squared error cost function.

$loss_{dir}$	mAP (–, 1%)	mAP (–, 3%)	mAP (–, 5%)
DR	19.4%	85.2%	98.8%
L_{BB} (m=3)	**25.8%**	88.5%	99.8%
L_{BB} (m=6)	23.5%	87.6%	**99.9%**
L_{BB} (m=9)	20.3%	**89.5%**	99.7%
L_{BB} (m=12)	23.5%	87.1%	99.5%
L_{BB} (m=15)	24.9%	86.0%	**99.9%**
L_{BB} (m=18)	19.5%	86.3%	99.3%
L_{BB} (m=36)	14.6%	69.0%	90.1%
L_{BB} (m=45)	9.1%	40.3%	77.6%

Table 4
Using different combinations of location loss and direction loss for BV-Net training.

$loss_{loc} + loss_{dir}$	mAP (0.5%, 3%)	mAP (0.3%, 3%)	mAP (0.3%, 5%)
$L_{CirDIoU} + L_{BB}$ (m=6)	85.3%	78.7%	89.6%
$L_{CirDIoU} + L_{BB}$ (m=9)	**85.5%**	79.1%	90.0%
$L_{CirDIoU} + L_{BB}$ (m=12)	84.6%	**79.3%**	**90.9%**
$L_{CirGIoU} + L_{BB}$ (m=6)	84.7%	78.7%	90.8%
$L_{CirGIoU} + L_{BB}$ (m=9)	85.1%	78.9%	90.3%
$L_{CirGIoU} + L_{BB}$ (m=12)	82.4%	76.0%	89.7%
$MSE + L_{BB}$ (m=6)	80.9%	70.5%	83.1%
$MSE + L_{BB}$ (m=9)	82.5%	69.3%	80.8%
$MSE + L_{BB}$ (m=12)	82.6%	69.8%	82.2%

The results are shown in Table 2, BV-Net gets best performance when using $L_{CirDIoU}$. On the whole, the proposed two loss functions $L_{CirGIoU}$, $L_{CirDIoU}$, which effectively utilize geometric feature of tubular solder joints, respectively get better performance than MSE loss, CirIoU loss and box-based location loss in tubular solder joint detection. This experiment shows that our proposed $L_{CirDIoU}$ can effectively optimize the network to center point positioning of tubular solder joints.

4.5. Hyperparameter: m in direction loss

In bin-based angle regression, we set different values of m to analyze the impact on the angle regression of tubular solder joint. In this experiment, location loss uses $L_{CirDIoU}$, R is set to $1 \times r_0$. mAP (–, ∗) indicates only the angle regression is concerned in the evaluation index, and the center point positioning is ignored. $L_{BB}(m{=}6)$ means using six bin classification heads and using six residual regression heads in angle regression based on BV-Net.

As is illustrated in Table 3, BV-Net gets the best performance in mAP (–, 1%), mAP (–, 3%), mAP (–, 5%) respectively when using $L_{BB}(m{=}3)$, $L_{BB}(m{=}9)$, $L_{BB}(m{=}6)$. This experiment shows bin-based angle regression is much better than directly regressing, promoting

Table 5
Method extension and comparison on the tubular solder joint dataset. Frame-per-second (FPS) were measured on the same machine. The left three columns in mAP indicates only the center point positioning is concerned, the middle three columns in mAP indicates only the angle regression is concerned, and the right three columns in mAP indicates both the center point positioning and the angle regression of tubular solder joint are concerned. A dash means the method have no corresponding function. BV-Net* is the BV-Net without angle estimation.

Model	Backbone	Size	FPS	mAP								
				(0.1%, –)	(0.3%, –)	(0.5%, –)	(–, 1%)	(–, 3%)	(–,5%)	(0.5%, 3%)	(0.3%, 3%)	(0.3%, 5%)
YOLOv3	DarkNet-53	608	40.3	18.2%	91.5%	97.6%	–	–	–	–	–	–
YOLOv4	CSPDarknet-53	608	35.3	21.3%	93.0%	97.9%	–	–	–	–	–	–
CenterNet	Hourglass-104	608	4.5	16.9%	87.8%	97.5%	–	–	–	–	–	–
BV-Net* (Ours)	CSPDarknet-53	608	35.4	**21.5%**	**93.1%**	**98.3%**	–	–	–	–	–	–
BV-Net (Ours)	CSPDarknet-53	608	34.4	18.5%	90.1%	98.0%	**20.3%**	**89.5%**	**99.7%**	**85.5%**	**79.1%**	**90.0%**

6.4% in mAP (–, 1%), 4.3% in mAP (–, 3%), 1.1% in mAP (–, 5%) separately. Bin-based angle regression transforms a regression task into a classification and regression task, improving the precision of direction estimation greatly. In addition, different values of m greatly affect the precision of angle regression in tubular solder joint detection. The larger m is, the more difficult it is for bin classification target, while the smaller m is, the more difficult it is for residual regression target.

4.6. Influence of different combinations of location loss and direction loss

In the previous experiments, we respectively discussed location loss and direction loss for tubular solder joint detection. In this section, we use different combinations of location loss and direction loss to analyze the impact on tubular solder joint detection. R is set as $1 \times r_0$.

The results are shown in Table 4. BV-Net gets best performance in mAP (0.5%, 3%) when using $L_{CirDIoU}$ and L_{BB}(m=9). And it is when using $L_{CirDIoU}$ and L_{BB}(m=12) that BV-Net gets best performance in mAP (0.3%, 3%) and mAP (0.3%, 5%). This experiment further shows the effectiveness of our proposed $L_{CirDIoU}$. In addition, the selection of hyperparameter m in direction loss will be slightly affected by the addition of location loss.

4.7. Comparison with the-state-of-art horizontal object detection methods

These is no method which directly predicts the form of a vector (V_x, V_y, V_θ) in general object detection. In this section, To verify the effectiveness of our proposed BV-Net, we make a comparison with the-State-of-Art horizontal object detection methods on tubular solder joint detection in this experiment. We compare with YOLOv3 [24], YOLOv4 [18], CenterNet [42] and these methods experimented on the same machine, which is same as BV-Net. In testing, all the images are resized as 608 × 608.

Table 5 summarizes the experimental results of different methods. BV-Net achieves 85.5% mAP (0.5%, 3%), 79.1% mAP (0.3%, 3%), 90.0% mAP (0.3%, 5%) with 34.4 FPS, meeting high-precision and real-time requirements of industrial system. Tubular solder joints are small and illegible with complex and low-contrast background. BV-Net could effectively extract and identify the features of tubular solder joints because of its powerful feature extraction module and feature fusion module. Our proposed powerful loss function could effectively optimize BV-Net to detect the target parameters of tubular solder joints.

In BV-Net, adding the function of direction estimation slightly reduces the model location accuracy. When predicting the same amount of parameters, our proposed BV-Net gets best performance for location and achieves 21.5% mAP (0.1%, –), 93.1% mAP (0.3%, –), 98.3% mAP (0.5%, –) with 35.4 FPS. BV-Net is based on the framework of state-of-the-art object detector (YOLOv4), but works better than YOLOv4. The reason is that geometric feature of tubular solder joints is utilized adequately in location loss of BV-Net. The experimental results verified the effectiveness of our proposed method.

5. Conclusion

We propose a tubular solder joint detection method named BV-Net, which could effectively estimate both the center point and the direction of tubular solder joint. In BV-Net, we propose CirDIoU loss ($L_{CirDIoU}$) and introduce bin-based angle regression loss (L_{BB}) for tubular solder joint detection. Combining $L_{CirDIoU}$ and L_{BB} could effectively regress a vector with the parameter form of (x, y, θ). In particular, L_{BB} promotes 4.3% in mAP (–, 3%), compared with the direct regression method. In addition, we establish a tubular solder joint dataset and design a new evaluation index: mAP (δ_d, δ_θ) for tubular solder joint detection. On the dataset, BV-Net achieves 85.5% mAP (0.5%, 3%) with 34.4 FPS, meeting high-precision and real-time requirements of industrial system. The experimental results verified the effectiveness of our proposed method. In the future, we will investigate some model compression technology to make BV-Net operate in industry lighter, such as parameter pruning, knowledge distillation.

CRediT authorship contribution statement

Chenlin Zhou: Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing. **Xiaofei Shen:** Software, Methodology. **Peng Wang:** Supervision, Funding acquisition, Methodology. **Wei Wei:** Validation, Resources, Methodology. **Jia Sun:** Conceptualization, Data curation, Writing check. **Yongkang Luo:** Software, Investigation, Data curation. **Yiming Li:** Investigation, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. yuan Wang, Y. Zhao, L. Wen, PCB welding spot detection with image processing method based on automatic threshold image segmentation algorithm and mathematical morphology, Circuit World 42 (3) (2016) 97–103.
[2] H. Wu, X. Xu, Solder joint inspection using eigensolder features, Solder. Surf. Mount Technol. 30 (4) (2018) 227–232.
[3] N. Cai, G. Cen, J. Wu, F. Li, H. Wang, X. Chen, SMT solder joint inspection via a novel cascaded convolutional neural network, IEEE Trans. Compon. Packag. Manuf. Technol. 8 (4) (2018) 670–677.
[4] Y.-M. Chang, C.-C. Wei, J. Chen, P. Hsieh, An implementation of health prediction in SMT solder joint via machine learning, in: 2019 IEEE International Conference on Big Data and Smart Computing, BigComp, IEEE, 2019, pp. 1–4.
[5] D. Ma, X. Lei, H. Zhao, The research of qualification detection of cable joint solder joint based on DCNN, in: 2019 International Conference on Artificial Intelligence and Advanced Manufacturing, AIAM, IEEE, 2019, pp. 598–603.
[6] H. Wu, W. Gao, X. Xu, Solder joint recognition using mask r-CNN method, IEEE Trans. Compon. Packag. Manuf. Technol. 10 (3) (2019) 525–530.
[7] Z. Mo, L. Chen, W. You, Identification and detection of automotive door panel solder joints based on YOLO, in: 2019 Chinese Control and Decision Conference, CCDC, IEEE, 2019, pp. 5956–5960.
[8] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[9] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580–587.
[10] A. Altan, S. Karasu, S. Bekiros, Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques, Chaos Solitons Fractals 126 (2019) 325–336.
[11] A. Altan, S. Karasu, Recognition of COVID-19 disease from X-ray images by hybrid model consisting of 2D curvelet transform, chaotic salp swarm algorithm and deep learning technique, Chaos Solitons Fractals 140 (2020) 110071.
[12] A. Altan, S. Karasu, E. Zio, A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition methods and grey wolf optimizer, Appl. Soft Comput. 100 (2021) 106996.
[13] Z. Yao, D. He, Y. Chen, B. Liu, J. Miao, J. Deng, S. Shan, Inspection of exterior substance on high-speed train bottom based on improved deep learning method, Measurement 163 (2020) 108013.
[14] Q. Zhou, R. Chen, B. Huang, W. Xu, J. Yu, DeepInspection: Deep learning based hierarchical network for specular surface inspection, Measurement 160 (2020) 107834.
[15] T. He, Y. Liu, Y. Yu, Q. Zhao, Z. Hu, Application of deep convolutional neural network on feature extraction and detection of wood defects, Measurement 152 (2020) 107357.
[16] R. Ren, T. Hung, K.C. Tan, A generic deep-learning-based approach for automated surface inspection, IEEE Trans. Cybern. 48 (3) (2017) 929–940.
[17] T. Zhou, J. Zhang, H. Su, W. Zou, B. Zhang, EDDs: A series of efficient defect detectors for fabric quality inspection, Measurement 172 (2021) 108885.
[18] A. Bochkovskiy, C.-Y. Wang, H.-Y.M. Liao, YOLOv4: Optimal speed and accuracy of object detection, 2020, arXiv preprint arXiv:2004.10934.
[19] R. Girshick, Fast r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448.
[20] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, in: Advances in Neural Information Processing Systems, 2015, pp. 91–99.
[21] J. Dai, Y. Li, K. He, J. Sun, R-FCN: Object detection via region-based fully convolutional networks, in: Advances in Neural Information Processing Systems, 2016, pp. 379–387.
[22] J. Pang, K. Chen, J. Shi, H. Feng, W. Ouyang, D. Lin, Libra r-cnn: Towards balanced learning for object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 821–830.
[23] J. Redmon, A. Farhadi, YOLO9000: Better, faster, stronger, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7263–7271.
[24] J. Redmon, A. Farhadi, Yolov3: An incremental improvement, 2018, arXiv preprint arXiv:1804.02767.
[25] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, Ssd: Single shot multibox detector, in: European Conference on Computer Vision, Springer, 2016, pp. 21–37.
[26] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2980–2988.
[27] X. Zhou, D. Wang, P. Krähenbühl, Objects as points, 2019, arXiv preprint arXiv:1904.07850.
[28] H. Law, J. Deng, Cornernet: Detecting objects as paired keypoints, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 734–750.
[29] Z. Tian, C. Shen, H. Chen, T. He, Fcos: Fully convolutional one-stage object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 9627–9636.
[30] Z. Yang, S. Liu, H. Hu, L. Wang, S. Lin, Reppoints: Point set representation for object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 9657–9666.
[31] T.-Y. Lin, M. Maire, S.J. Belongie, L.D. Bourdev, R.B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Doll'a r, C.L. Zitnick, Microsoft COCO: Common objects in context, CoRR abs/1405.0312 (2014) arXiv:1405.0312.
[32] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, I.-H. Yeh, CSPNet: A new backbone that can enhance learning capability of cnn, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 390–391.
[33] K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (9) (2015) 1904–1916.
[34] S. Liu, L. Qi, H. Qin, J. Shi, J. Jia, Path aggregation network for instance segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8759–8768.
[35] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, S. Savarese, Generalized intersection over union: A metric and a loss for bounding box regression, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 658–666.
[36] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, D. Ren, Distance-IoU loss: Faster and better learning for bounding box regression, in: AAAI, 2020, pp. 12993–13000.
[37] H. Yang, R. Deng, Y. Lu, Z. Zhu, Y. Chen, J.T. Roland, L. Lu, B.A. Landman, A.B. Fogo, Y. Huo, CircleNet: Anchor-free detection with circle representation, 2020, arXiv preprint arXiv:2006.02474.
[38] S. Shi, X. Wang, H. Li, Pointrcnn: 3D object proposal generation and detection from point cloud, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 770–779.
[39] M. Everingham, S.M.A. Eslami, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman, The pascal visual object classes challenge: A retrospective, Int. J. Comput. Vis. 111 (1) (2015) 98–136.
[40] M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman, The PASCAL Visual object classes challenge 2007 (VOC2007) Results, http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html.
[41] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft coco: Common objects in context, in: European Conference on Computer Vision, Springer, 2014, pp. 740–755.
[42] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, Q. Tian, Centernet: Keypoint triplets for object detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 6569–6578.