Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

Weld Seam Detection Method with Rotational Region Proposal Network*

He Zhang, Wei Song and Zheng Chen

Institute of Ocean Electronics Engineering
& Intelligent Systems
Zhejiang University

Zhoushan, Zhejiang Province, China
{21834137 & weisong & zheng_chen}@zju.edu.cn

Shiqiang Zhu

Institute of Ocean Electronics Engineering
& Intelligent Systems
Zhejiang University
Zhoushan, Zhejiang Province, China
Zhejiang Laboratory
Hangzhou, Zhejiang Province, China
sqzhu@zju.edu.cn

Cunjun Li and Huadong Hao

Zhoushan Institute of Calibration and Testing
for Quality and Technology Supervision

Zhoushan, Zhejiang Province, China
licunjun@sina.com, gentlehhd@163.com

Jason Gu

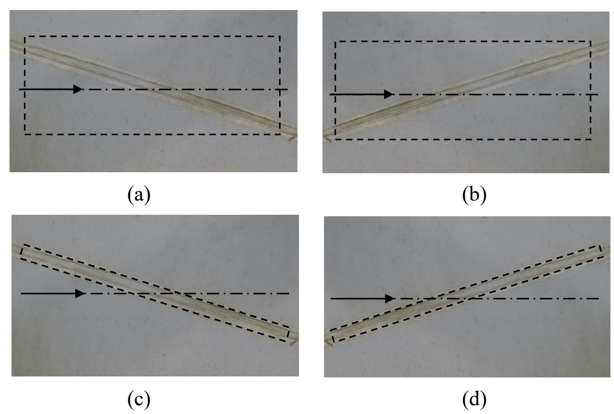
Zhejiang Laboratory
Hangzhou, Zhejiang Province, China
Dalhousie University
Halifax, NS B3H 4R2, Canada
jason.gu@dal.ca

***Abstract* - The traditional weld seam detection methods in passive vision are usually realized by detecting edges or textures of the weld seam. Since illumination conditions, weld seam types and backgrounds vary with tasks, these methods only work with specific weld types or environment. This paper raises a weld seam detection method that replaces the four main steps in traditional weld seam detection (pretreatment, thresholding, seam detection, seam fitting) with an end-to end neural network system. The method eliminates the ambiguity of original horizontal bounding box by adding an inclination parameter to the region proposal network (RPN). Compared with other methods in passive vision, our method is appropriate for accurate and fast detection of various types of weld seams in complex environment and meets the requirements for online seam detection of industrial robot.**

Index Terms - Weld Seam Detection, Inclined Region Proposal, Convolutional Neural Network.

I. Introduction

Frameworks of large equipment like oil tanks and ships are usually made with welded steel plates. The welding condition deteriorates with daily usage, which would reduce the strength of steel structure. Considering equipment’s safe and reliable running, it is necessary to inspect welding condition regularly. As weld inspection is mainly completed by workers with inspection tools nowadays, this manual operation is inefficient, risky and usually needs assistant equipment like scaffold, which make it more complicated. Therefore, it is an advance to use robots instead of human to enter workspace and complete weld inspection tasks. The method could expand the application scenarios of weld inspection, and complete those inspection tasks impossible for manual operation with higher accuracy and efficiency.



(a) Weld seam is on the left but considered as middle
(b) Weld seam is on the right but considered as middle
(c) Weld seam is on the left and considered as left
(a) Weld seam is on the right and considered as right
(Arrows show the direction that robot is approaching)
Fig. 1 Importance of Inclination Information to Robots’ Weld Navigation

In fact, there is still a series of problems unsolved in weld inspection on robots which is mainly reflected in robotic autonomy, specifically in weld detection, path planning and obstacle avoidance. Among them all, weld seam detection which effects the inspection result and accuracy of path planning, has these problems: weld image collection depends on processing illumination and shadow areas; seam target extraction effected by interference from rust or stains; multi-type weld operation demands adaptive capacity to different weld. Specifically, inclination of weld seam is important to weld seam navigation. Fig. 1 shows that the ideal method of

* This research is supported by the Scientific and Technology Cooperation Project of Zhoushan City and Zhejiang University (Grant No. 2017C8221) and the Quality Technology Infrastructure Foundation of Zhejiang Quality Supervision System (Grant No. 20180130).

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

weld seam detection should generate bounding boxes surrounding weld area and reflect the inclination of it.

There are two types methods to detect weld seam through vision technology: active vision and passive vision methods. Compared to active vision methods[1][2] represented by structure light, passive vision methods decrease the cost and complexity of instrument by directly capturing weld images under natural light with camera. The advantage of passive vision expands the work environment of mobile robots, and improves visualization of inspection. As a result, we choose passive vision method to detect weld seams.

Traditional passive vision methods to detect weld seam are mostly based on morphological image processing and expert algorithms, which could be summarized into four main steps: pretreatment, thresholding, seam detection and seam fitting.

a) Pretreatment. Pretreatment transfers original images into gray images, and eliminate noises in images. Different types of filters to eliminate noises should be chosen based on noise types correspondingly. For example, median filter is suitable for eliminating salt and pepper noise and pulse noise; mean filter and morphological filter are suitable for eliminating noise around weld areas[3]. b) Thresholding. Thresholding divides images into two types of areas: object areas and background areas. By analyzing the image gray distribution, the appropriate threshold is selected. The threshold value could divide the gray distribution histogram into two parts[4]. c)Seam detection. Weld edge inspection[5][6] is mostly used in weld detection, as Hough transform is usually applicable only to the weld line extraction in welding operation[7][8]. d) Fitting weld. The characteristics of welding seams are usually extracted by calculating the coordinates of the weld center through the coordinates of the weld edges, and fitting results are obtained by combining Hough line detection with least square method.

However, results in thresholding for those welds of which grayness is similar to the backgrounds would be very bad. Especially, interference from rust and stains would be hard to eliminated even using adaptive thresholding. What's more, edge detection relying on texture of objects would cause obvious error for weld edge location, or even miss the weld seam target.

According to the passage above, traditional methods are limited by the work condition and could not generate one policy for multiple scenes or weld types. Therefore, an advanced method should conquer this main shortage. For past years, automation industry pays more attention to the deep convolutional neural network[9]. A neural network could complete multiple targets detection from complex background, which advances from traditional methods obviously.

R-CNN[10][11][12] proposes selective search to extract region proposals for the first time, which avoids to generate proposal boxes on each pixel in old algorithms and greatly reduced the computation of convolution network. However, R-CNN needs to cut or stretch the proposal boxes of different sizes into a same fixed size, then extract the feature maps inside the proposal boxes, and finally send them into the classification and bounding box regression network, which not only consumes more time, but also causes image data loss or geometric distortion.

SPP-Net[13][14] introduces spatial pyramid pooling, and transfers proposals of different scales into the same scale through SPP layer without losing original image information. In addition, SPP-Net performs only one convolution operation on the original image to extract the feature map, which is about 100 times faster than R-CNN in this corresponding step. However, there are still too many layers in SPP-Net for feature extraction and the key steps are not simplified substantially, and the calculation amount is still the main problem affecting the detection speed.

Fast R-CNN[15][16][17] refers to the idea of SPP-Net and uses a pooling layer called region of interest (RoI) pooling layer, which is actually a single layer SPP-Net. The RoI pooling converts the features inside regions of interest into a series of small feature maps with a fixed spatial extent. In addition, Fast R-CNN shares part of the convolution layers of object classification and bounding box regression, which reduces the total computation amount of the two tasks. Compared with R-CNN, Fast R-CNN accelerates the detection by about 25 times. By then, selective search becomes the bottleneck restricting the speed of Fast R-CNN object detection.

Faster R-CNN[18][19][20] proposed a region proposal network (RPN) to replace selective search and integrated all the steps of object detection task into a neural network. Through the RPN, the proposal box information of the image could be generated directly. By introducing the RPN, Faster R-CNN improves the detection speed by about 10 times to Fast R-CNN.

However, RPN of original Faster R-CNN could only generate proposal boxes surrounding the outer contour of objects in horizontal or vertical directions, which could not describe inclined information of objects. Therefore, the convolutional neural network structure similar to Faster R-CNN is used in this paper, but rotation parameters of bounding boxes are added into the regional proposal network to realize the weld seam detection of different angles, so as to meet the task requirements of the robotic weld seam detection during movement.

II. METHOD

To conquer shortages of existing methods, this paper proposes a method based on convolutional neural network that could precisely extract information of weld (location, inclination, etc.), and send it to the robot for weld navigation. The method completes the four main steps of weld seam detection in an end-to-end process. With one input of video signals captures by the camera, the network could output the real-time bounding box of the weld seams without any extra manual operation.

The backbone of network used in this paper is mainly inspired by Faster R-CNN raised by Shaoqing Ren et al. in 2015[18]. Considering the characteristics of weld seams, this paper changes the direction of region proposals, the size of anchor boxes and other parameters to design a network customized for weld seam detection. The details of network would be discussed in this section, especially for those parts different from the original Faster R-CNN.

A. Problem Definition

Precisely, the edge of the weld seam is not strict line and the shape of weld areas is not standard rectangle. However, for those mobile robots that navigate welds with visual sensors, describing weld seams as inclined rectangles could simplify the problem. To decrease storage space for weld seam information, and increase speed of neural network read dataset, we use x, y, w, h, and a five parameters to describe the location, shape and inclination of weld seams.

B. Network

Feature Extraction Layers

Like most neural networks for object detection, this network uses a combination of convolutional layers, ReLU layers and pooling layers to extract the feature of the original image and build a feature map accordingly. In this part, using a pre-trained network instead of an initial one could reduce the training time and speed up the deployment of the whole model. Faster R-CNN uses Zeiler and Fergus model (ZF)[21], which has 5 shareable convolutional layers and the Simonyan and Zisserman model (VGG-16)[22], which has 13 shareable convolutional layers[18]. Our network uses Resnet presented by Kaiming He et al. in 2016[23]. Compared with those two networks, Resnet are easier to optimize and can gain accuracy from considerably increased depth by introducing a residual network structure. A Resnet could have more than 100 layers without the degradation problem[23].

Region Proposal Networks

A Region Proposal Network (RPN) takes the $n \times n$ sized feature map as input and generates a set of rectangular object proposals, each with an objectness score[18]. In this part, a sliding network slides on the feature map, with each window mapped to a lower-dimensional feature. The features are input into two siblings fully connection layers, bounding box regression layer and bounding box classification layer.

Like Faster R-CNN, our network creates a series of anchor boxes at each sliding window position, with a maximum of k. This default parameters are according to the ordinary shape and size of the weld seam of robot perspective.

The most significant difference from the original RPN is that we introduce a rotation parameter θ into the anchor box. This change is designed for robot navigation of weld seams. As shown in Figure 4, the original RPN could only generate horizontal region proposals, which could not distinguish the two-symmetry situation so that the robot could not follow the weld seam properly.

The loss function of RPN is defined as below:

$$L\left(\left\{p_i\right\},\left\{t_i\right\}\right) = \frac{1}{N_{cls}} \sum_i L_{cls}\left(p_i, p_i^*\right) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}\left(t_i, t_i^*\right)$$

Here, i is the index of an anchor in a mini-batch and pi is the predicted probability of anchor i being an object. The ground-truth label p_i* is 1 if the anchor is positive, and is 0 if the anchor is negative. t_i is a vector representing the 4 parameterized coordinates of the predicted bounding box, and t_i* is that of the ground-truth box associated with a positive anchor. The classification loss L_{cls} is log loss over two classes (object vs. not object). For the regression loss, we use $L_{reg}(t_i, t_i^*) = R(t_i - t_i^*)$ where R is the robust loss function (smooth L_1) defined in Fast R-CNN[15]. The term $p_i^* L_{reg}$ means the regression loss is activated only for positive anchors ($p_i^* = 1$) and is disabled otherwise ($pi* = 0$). The outputs of the *cls* and *reg* layers consist of $\{p_i\}$ and $\{t_i\}$ respectively. The two terms are normalized by N_{cls} and N_{reg} and weighted by a balancing parameter λ.

For bounding box regression, there are 5 parameters t_x, t_y, t_w, t_h, t_θ as following:

$$\begin{aligned}
&t_x = \left(x - x_a\right)/w_a, \quad t_y = \left(y - y_a\right)/h_a \\
&t_w = \log\left(w/w_a\right), \quad t_h = \log\left(h/h_a\right), \quad t_\theta = \theta \ominus \theta_a \\
&t_x^* = \left(x^* - x_a\right)/w_a, \quad t_y^* = \left(y^* - y_a\right)/h_a \\
&t_w^* = \log\left(w^*/w_a\right), \quad t_h^* = \log\left(h^*/h_a\right), \quad t_\theta^* = \theta^* \ominus \theta_a
\end{aligned}$$

where x, xa and x^* refer to the central point abscissa of the prediction box, anchor box and ground truth box respectively. Similarly, y's refer to their central point ordinates, h's and w's refer to their width and height. θ's refer to the angle from the positive direction of the abscissa to the direction parallel to the long side of the inclined anchor box. Operation $\theta \ominus \theta_a = \theta\text{-}\theta_a + k\pi, k \in Z$, ensuring that:

$$\theta \ominus \theta_a \in \left[-\frac{\pi}{4}, \frac{3\pi}{4}\right).$$

After each iteration, RPN would generate a list of $[tx, ty, tw, th, t\theta]$ to each proposal box for regression. At last, RPN would output a series of inclined rectangular bounding boxes, in which are the region that are supposed to be weld seams judged by the network. As is mentioned above, these bounding boxes are presented by $[x, y, w, h, \theta]$.

RoI Pooling

After RPN, bounding boxes are of various sizes, which makes classification operation complex and slow. In RoI pooling layer, the network extracts a fixed-sized feature box from the original feature map for each region proposal. Then, the final classification layer would only have to compute once of each image rather than every proposal generated in RPN.

RoI pooling in Faster R-CNN could only deal with the proposal boxes that are axis-aligned. So, we use the angle parameter θ in RPN to rotate each proposal box to axis-aligned temporarily. The angle parameter would not be aborted during RoI pooling, which would not affect the final detection and location result.

III. EXPERIMENT

Since there is no open source dataset of weld images, we make a weld image dataset which contains 5454 images of different views and distance. We divide the dataset into 1083 images as train-valid dataset and 4371 images as test dataset. The annotation contains location, shape and inclination of each weld in the corresponding image. The precision of test result is described with three indexes: precision, recall and F-measure.

Implement Details

Use Resnet pretrained on VOC2007 as our feature extracting network. Weights of network update at a learning rate of 10^-4, weight decay as 10^-4 and momentum as 0.9. To find the appropriate scales of proposal regions and decrease the iteration loops, we set a series of scales and angles of default bounding boxes.

Data enhancement

To expand image dataset as many as possible, we use two methods to enhance data. First, we use a series of operation of rotation and flip on the original images. Then, we randomly add noise into these images. Correspondingly, we calculate the updated annotations of these new generated images. After these operations, we obtain the enhanced dataset.

Border padding.

As the weld intersects with the image boundary, there are often proposal regions beyond the image boundary, which cannot be roughly screened out. Therefore, as long as the proposal box beyond the image boundary is within a certain range, it will still be aligned and enter the grading stage. Remove only proposal boxes that exceed the image boundary. Similarly, for the output of the test set, the boundary box not exceeding the boundary threshold is retained to cover as many weld areas as possible.

After experiment, the neural network method is found out to have a good performance of weld seam detection. The method has immunity from interference of complex illumination, rust and stains (Fig. 2). In the experiment, welds in the middle of images are more likely detected than those on the edge of images. The weld with a larger inclination Angle is more likely to have contour deviation than the weld with a smaller inclination angle (Fig. 3). According to the analysis, the shape of the edge weld in the picture is not complete and there is a big gap between it and the typical weld, which is a special case. In addition, due to insufficient training samples, the test result is not as good as that of other welds. The inclination Angle of the welding seam with a larger inclination Angle differs greatly from that of the neural network proposal box, so more gradient descent calculations are needed to obtain better training results. In the actual detection, it is often difficult to reach the ideal number of calculations, so the detection results are not as good as those with a smaller inclination Angle.

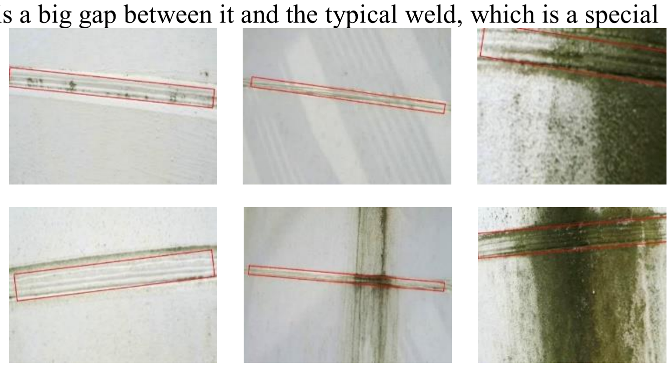


Fig. 2 Examples of results from our method

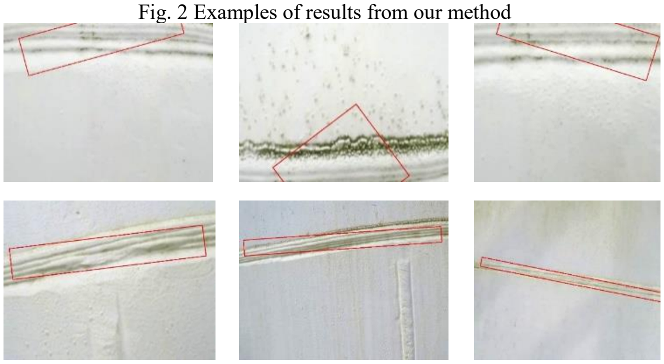


Fig. 3 Examples of bad detection results

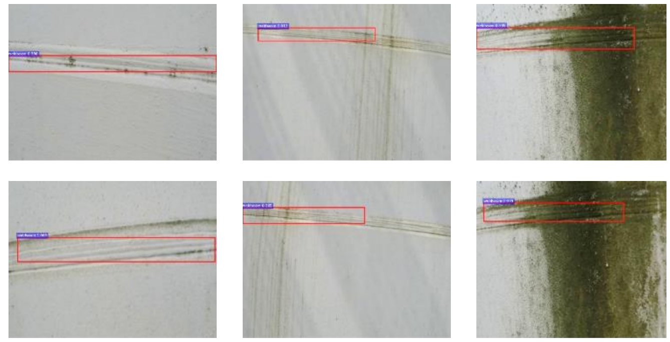


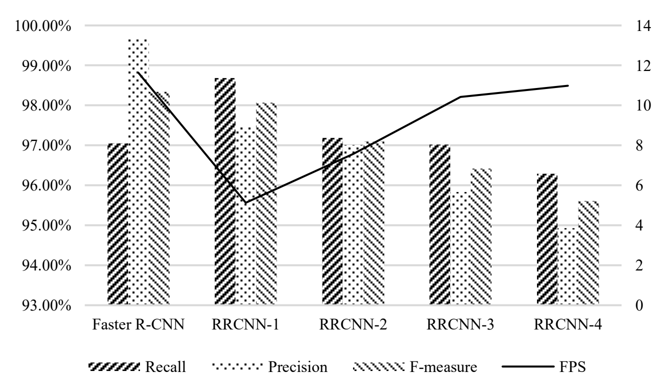
Fig. 4 Examples of Faster R-CNN results

Compared with the native Faster R-CNN (Fig. 4), the algorithm in this paper can accurately detect the direction and position characteristics of the weld, and extract more complete weld from the complex background environment. In order to find the best hyperparameters of the neural network, a series of anchoring boxes were selected in this paper and a comparative test was conducted. The results of the comparative test were shown in Table 1. To compare the performance of this network with other algorithms or networks in weld detection tasks, Sift, edge detection, Faster R-CNN, SSD and Yolov3 algorithms or networks were also tested on the weld image data set. The test results are shown in Table 2.

Table 1

Approaches	Inclination	Recall	Precision	F-measure	FPS
Faster R-CNN	FALSE	97.05%	99.68%	98.34%	11.6
RRCNN-1	TRUE	98.68%	97.47%	98.07%	5.1
RRCNN-2	TRUE	97.19%	97.01%	97.10%	7.6
RRCNN-3	TRUE	97.02%	95.83%	96.42%	10.4
RRCNN-4	TRUE	96.29%	94.93%	95.61%	11.0

Table 2



IV. CONCLUSIONS

This paper proposes a weld seam detection method with rotational region proposal network. Considering features of mobile weld detection, this paper introduces RRPN and studies the results of it. Furthermore, the structure and parameters of the network are optimized through series of experiments.

The method has a good performance on our weld image dataset: F-measure reaches higher than 96% with single image detection costing less than 0.1 second. Also, the method has immunity from complex illumination and interference of rust or stains. Compared with other deep neural network, method used in this paper could detect inclination of weld seams, locate weld seams more precisely and extract weld area more intact.

In further study, we would do more research and try to optimize the performance in those conditions of welds on edge of images and obliquely inclined. Also, we would use GAN to enhance dataset so that we could reduce our demand of original image quantities.

REFERENCES

[1] Fang, Zaojun, De Xu, and Min Tan. "A vision-based self-tuning fuzzy controller for fillet weld seam tracking." IEEE/ASME Transactions on mechatronics 16, no. 3 (2011): 540-550.

[2] Xue, Boce, Baohua Chang, Guodong Peng, Yanjun Gao, Zhijie Tian, Dong Du, and Guoqing Wang. "A Vision Based Detection Method for Narrow Butt Joints and a Robotic Seam Tracking System." Sensors 19, no. 5 (2019): 1144.

[3] Amruta, Rout , B. B. V. L. Deepak , and B. B. Biswal . "Advances in weld seam tracking techniques for robotic welding: A review." Robotics and Computer-Integrated Manufacturing56(2019):12-37.

[4] H. Chen, T. Lin, S. Chen, Seam tracking and dynamic process control for high precision arc welding, Robotic Welding, Intelligence and Automation, Springer,2011, pp. 193–201.

[5] Y. Xu, G. Fang, S. Chen, J.J. Zou, Z. Ye, Real-time image processing for vision based weld seam tracking in robotic GMAW, Int. J. Adv. Manufact. Technol. 73 (9–12) (2014) 1413–1425.

[6] Y. Xu, G. Fang, N. Lv, S. Chen, J.J. Zou, Computer vision technology for seam tracking in robotic GTAW and GMAW, Robot Comput. Integr. Manuf. 32 (2015) 25–36.

[7] H. Ma, S. Wei, Z. Sheng, T. Lin, S. Chen, Robot welding seam tracking method based on passive vision for thin plate closed-gap butt welding, Int. J. Adv. Manufact. Technol. 48 (9–12) (2010) 945–953.

[8] Fang, Zaojun, De Xu, and Min Tan. "Visual seam tracking system for butt weld of thin plate." The International Journal of Advanced Manufacturing Technology 49, no. 5-8 (2010): 519-526.

[9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105.

[10] Girshick, Ross, et al. "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation." IEEE Conference on Computer Vision & Pattern Recognition 2014.

[11] Russakovsky, Olga, et al. "Imagenet large scale visual recognition challenge." International journal of computer vision115.3 (2015): 211-252.

[12] Jia, Yangqing, et al. "Caffe: Convolutional architecture for fast feature embedding." Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in European Conference on Computer Vision (ECCV), 2014.

[14] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[15] Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE international conference on computer vision. 2015.

[16] He, Kaiming, et al. "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification." Proceedings of the IEEE international conference on computer vision. 2015.

[17] Liu, Zikun, et al. "Rotated region based CNN for ship detection." 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017.

[18] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).

[19] Jiang, Yingying, et al. "R2cnn: Rotational region cnn for orientation robust scene text detection." arXiv preprint arXiv:1706.09579 (2017).

[20] Ma, Jianqi, et al. "Arbitrary-oriented scene text detection via rotation proposals." IEEE Transactions on Multimedia 20.11 (2018): 3111-3122.

[21] Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." European conference on computer vision. springer, Cham, 2014.

[22] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[23] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.