



# A New Image Segmentation Method Based on the YOLO5 and Fully Connected CRF

Jian Huang<sup>1,2</sup> · Guangpeng Zhang<sup>1</sup> · Li Juan Ren<sup>1</sup> · Nina Wang<sup>1</sup>

Received: 12 July 2023 / Accepted: 1 November 2023  
© The Author(s) 2023

## Abstract

When manually polishing blades, skilled workers can quickly machine a blade by observing the characteristics of the polishing sparks. To help workers better recognize spark images, we used an industrial charge-coupled device (CCD) camera to capture the spark images. Firstly, the spark image region detected by yolo5, then segment from the background. Secondly, the target region was further segmented and refined in a fully connected conditional random field (CRF), from which the complete spark image obtained. Experimental results showed that this method could quickly and accurately segment whole spark image. The test results showed that this method was better than other image segmentation algorithms. Our method could better segment irregular image, improve recognition and segmentation efficiency of spark image, achieve automatic image segmentation, and replace human observation.

**Keywords** Image segmentation · YOLO5 · Fully connected CRF · CCD · Spark image

## 1 Introduction

Turbine compressors, industrial energy recovery turbines (TRT), and other equipment are widely used in petrochemical, metallurgy, wind tunnel tests and other national pillar industries. The blade is a key component with a complex structure. Its machining quality directly affects the performance of the turbine power system.

The blade size accuracy of manual polishing is completely dependent on the operator's experience, which leads to poor blade size consistency due to the big difference between personal experience and easy fatigue of operation; the dust pollution in the process of polishing is serious, the labor intensity is hard, which makes it difficult to hire young workers under 40 years old at present.

In manual blade grinding, workers control the material removal rate (MRR) by experience through observing the characteristics of grinding sparks. To achieve automatic

polishing, it is necessary to solve the problem of image processing in grinding sparks.

Fan [1] proposed an image low-level structure feature extraction method based on a deep neuron network and stacked sparse noise reduction auto-encoder (SSDA). It could extract the features of natural images and infrared image features. An unsupervised deep-learning training method was proposed by Chelsea Finn [2]. Combining predictive models obtained through raw perception with model predictive control (MPC) enabled the robot to show significant computer vision tasks under complex, unstructured environments. Avelino [3] addressed the problems of rapid visual movement and delays in image acquisition systems. By combining automatic stabilization mechanisms and advanced detection algorithms, the prediction accuracy of pedestrian position improved, and the human – robot interaction behavior strengthened.

Our team has studied the abrasive belt polishing work-piece process for many years. Among them, Ren [4] investigated the relationship between the spark field characteristics and the MRR. Ren [5] established a contact stress model on Hertzian contact theory, and developed a material removal rate mode to recognize the material removal in surface polishing. Wang Nina [6, 7] established an MRR model using images and sound. Wang Nina [8] proposed a new method using 2D-CNN learning to monitor material removal. Wang

✉ Jian Huang  
565200245@qq.com

<sup>1</sup> School of Mechanical and Precision Instrument Engineering, Xi'an University of Technology, Xi'an 710048, China

<sup>2</sup> School of Computer Science, XiJing University, Xi'an 710123, China

Nina [9, 10] investigated the effect of the wear state of abrasive belts on the MRR among the workpiece polishing, established a method for monitoring the wear state of belts based on the machine vision and image processing.

The above methods used a variety of sensors and investigated the effects of spark images and sound on the MRR. However, there are still the following issues. The segmentation of spark images is not fully automated; when the light is dim, it is hard to distinguish spark image from some background images; the processing of spark-image segmentation after a complete machining process is finished; an online system cannot be mentioned.

Huang [24] used the deep learning algorithm YOLO5 to perform target detection on spark images and obtain spark image regions. However, this paper is a further study after that one, to establish an automatic spark image segmentation system that can quickly detect target image regions in complex environments and separate them. The objective is to lay the foundation for studying the relevance between spark images and MRR and to achieve automatic workpiece processing.

The above processing methods for spark images are closely related to computer vision technology. The target detection and image segmentation are a research hotspot. Many new algorithms have proposed. Among the target detection algorithms, a typical one is the YOLO algorithm [11, 12]. The algorithm has developed through five generations, the accuracy and speed of target detection have greatly improved, and it can achieve multitarget detection [13, 14]. Its counterparts are Fast-RCNN [15] and Faster-RCNN [16] algorithms, with accurate target detection, but it still have some disadvantages. The YOLO algorithm is better by fast speed and high accuracy and is suitable for application in real-time target detection.

Image segmentation methods include semantic segmentation method, pixel segmentation method, and others. Many scholars have combined deep learning with image segmentation techniques and proposed new segment algorithms. Olaf R. proposed the U-Net algorithm [17]. He,

K. proposed the MASK R-CNN algorithm [18], which achieved semantic segmentation of images. In addition, many scholars have improved the classical algorithm. For example, Lian [19] proposed an attention feature-based method for finding small targets, achieved the detection of small targets in traffic scenes and improved the accuracy of target prediction. Wenkel [20] presented a methodology to evaluate the confidence level in target detection and obtained the best performance point to improve the efficiency. Wang [21] applied the YOLOv3 model to complete the experimental targets. Arunabha [22] used a YOLOv4 algorithm to improve the accuracy of plant disease detection.

In order to establish the model between spark images and MRR, separating the spark images from the background is an important part. When we tried to segment the spark images, we found that the spark images were irregular in shape, similar to the shape of fireworks, with complex boundaries. Although using deep learning methods for segmentation, the target image had to be labeled first, which was very time consuming, had more manual marking errors, and the final segmentation effect after training was not very satisfactory. Therefore, we proposed a new method, first detected the target region with YOLO5, then used fully connected CRF to further segment and refine the sparkle image within the region, and finally extracted the sparkle image completely. Concretely, as shown in Table 1.

Section 2 introduces the experimental and the method of spark image acquisition. Section 3 discusses the process and steps of YOLO5 target detection, which achieves the separation of the spark image region from the background. Section 4 introduces the principle and method of fully connected CRF image segmentation, which further segments the complete spark image from the spark region. Section 5 shows the image segmentation using the above method and compares it with other algorithms. Section 6 discusses the result.

**Table 1** Introduction to the content of this article

Section	Title	Content
1	Introduction	Introduce the background and significance of the paper
2	Experiment setup	Grinding the workpiece with abrasive belt and collecting spark images
3	Steps of image segmentation	Introduce the steps of spark image segmentation
4	YOLO5 target detection process	Spark area detection
5	fully connected CRF image segmentation method	Further refine image segmentation
6	Discussion	Analysis results

## 2 Experiment Setup

### 2.1 Belt Grinding Mechanism

To realize the study of spark images, we built an experimental platform to record a complete workpiece machining process and capture the spark images collected during the machining process. Figure 2 shows the experiment, which mainly includes a three-axis machine, two cameras, and two computers. In Fig. 1, the workpiece selected GCr15 with a hardness of HRC58 and a size of 170 mm×41 mm×50 mm. The motor speed was 0~5000 rpm, the speed of the driven belt was 0~34 m/s. The contact wheel was rubber with a Shore A hardness of 85. The belt material was corundum and a width of 20 mm. When the motor rotates at high speed, it could cut the workpiece and produce a spark field.

A Beckhoff CX5130 embedded was selected as the controller; Table 2 shows the parameters.

The type of the camera is MT-E200GC-T CMOS; its parameters are shown in Table 3.

The CMOS camera can collect images in real-time with Mindvision software, its frame frequency is 100 Hz.

### 2.2 The Mechanism of Spark Generation

During the grinding process, a number of sparks generated due to the high-speed rotation of the abrasive belt. Two industrial CCD cameras were mounted on the front and side of the spark field to record and save the spark images at a



Fig. 1 Belt grinding system

Table 2 Beckhoff CX5130 controller

Content	Performance
MCU	INTEL E3827
Memory	4 GB DDR3
Interface	4×USB2.0, 1×
System	Windows Embedded
Power	20 w

frame rate of 100 Hz. Figure 2a shows the acquired side spark images, Fig. 2b shows the frontal spark images.

## 3 Steps of Image Segmentation

The whole processing steps of image segmentation are shown in Fig. 3. In the second section, we built an experimental platform and obtained spark images. Then, the acquired dataset was labeled and segmented and then trained with YOLO5 to the target region of the spark image. It will explain in a more in-depth explanation in Sect. 4. Finally, the images in this region were further subdivided and segmented with fully connected CRF, separate the complete spark image from it. The method used will describe detail in Part 5.

## 4 YOLO5 Target Detection Process and Steps

### 4.1 Image Annotation

YOLO5 training the dataset. We used software to label all images, as shown in Fig. 4, with only “fire” for labeling. The dataset divided into two parts: testing and training dataset, with a ratio of 0.2:0.8. We annotated the data in the training dataset.

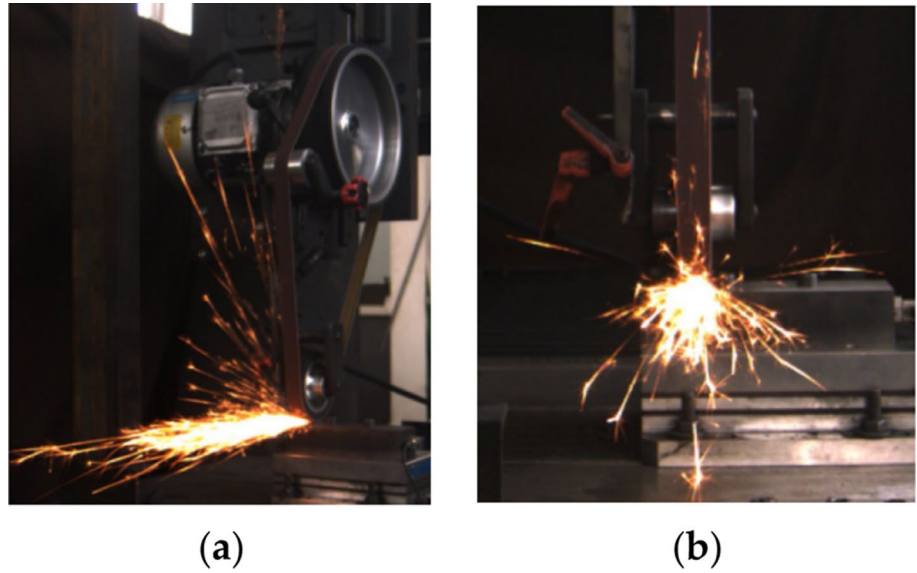
### 4.2 YOLO5 Model

The YOLO target detection algorithm has undergone many years of development. The latest version is YOLO5, which

Table 3 The indicators of cameras

Performance	Parameter
Camera	HT-GE200GC
Sensor	Colour COMS industrial camera
Pixels	2 million
Resolution	1200×1600
Lens sight distance	6–12 mm
Angle	28–53
Size	$\phi$ 32 (mm)×41 (mm)

**Fig. 2** a Side spark image. b Frontal spark image



has several versions. The minimum architecture is YOLO5S, which is only 14 Mb and can detect targets quickly and accurately. The basic architecture of YOLO5S shown in Fig. 5.

It mainly includes four parts: Data input side, Backbone, Neck, and Prediction. Among them, the data input side adopts Mosaic data enhancement to improve the effect of target detection. Backbone adopts a Focus structure and Cross-Stage Partial (CSP) structure. The original 608\*608\*3 image is input to the Focus structure. It becomes a 304 \* 304 \* 12 feature map after slicing. Through 32 convolution kernels, it finally becomes a feature map of 304\*304\*32. At the end of Backbone, SPPF (Fast Spatial Pyramid Pooling) is used to fuse feature maps of different receptive fields, thereby improving the expression ability of feature maps. The CSP1\_X structure applied to the Backbone core network, and the other CSP2\_X structure applied to Neck. Neck adopts the form of Feature Pyramid Networks (FPN) and Pixel Aggregation Networks (PAN). The FPN obtains the feature map for making predictions. In the PAN structure, two feature maps combined using concat operation and the feature.

### 4.3 Train

The computer configuration for training YOLO5 is shown in Table 4. The computer is equipped with an RTX3060 GPU, could significantly improve the training speed. The training performed on the frontal and side spark image datasets.

The results of frontal and side spark image target detection are shown in Figs. 6 and 7. The complete spark image area with an accuracy of 0.9 or higher detected.

The software used to crop the above image according to the bounding box of the YOLO5 detection region to obtain

the front and side spark image regions. As shown in Fig. 8, the region will be further segmented below to separate the spark images from the background.

## 5 Fully Connected CRF Image Segmentation Method

### 5.1 Fully Connected CRF Image Segmentation Principle

Fully connected CRF uses an efficient full-connected conditional random field model for image semantic segmentation. In the fully connected pairwise CRF model, the corresponding Gibbs energy [23] is shown in Eq. (1).

$$E(x) = \sum_i \psi_u(x_i) + \sum_{i < j} \psi_p(x_i, x_j). \quad (1)$$

The range of  $i$  and  $j$  is from 1 to  $N$ . The unary potential  $\psi_u(x_i)$  is computed independently for each pixel. The unary potential incorporates shape, texture, and location. The pairwise potentials are shown in Eq. (2).

$$\phi_p(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^K \omega^{(m)} k^{(m)}(f_i, f_j). \quad (2)$$

The  $k^{(m)}$  is a Gaussian kernel, the vector  $f_i$  and  $f_j$  are feature vectors for pixels,  $\omega^{(m)}$  is a linear combination weight, and  $\mu$  is a label compatibility function.

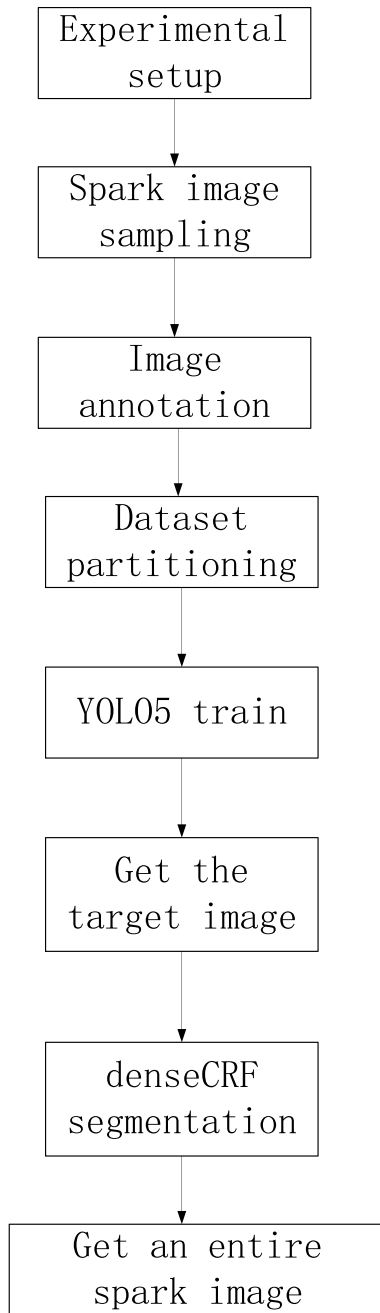


Fig. 3 Image segmentation steps

The mean field approximation computes a distribution  $Q(\mathbf{X})$  that minimizes the KL divergence using the following iterative update equation, as shown in Eq. (3).

$$Q_i(x_i = l) = \frac{1}{Z_i} \exp\{-\psi_u(x_i) - \sum_{l' \in L} \mu(l, l') \sum_{m=1}^K \omega^{(m)} \sum_{j \neq i}^K k^{(m)}(f_i, f_j) Q_j(l')\}. \tag{3}$$

This updated equation leads to the following inference algorithm [23].

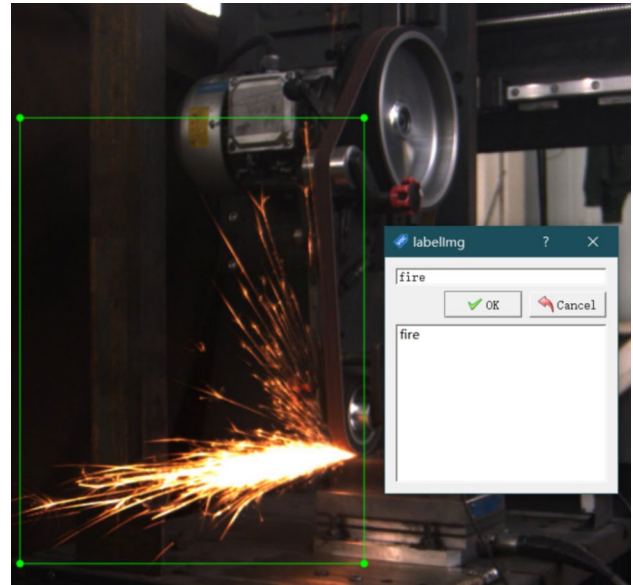


Fig. 4 Image annotation

### 5.2 Fully Connected CRF Image Segmentation Results

We further subdivided the area map of the spark image obtained after YOLO5 detection in Fig. 8 with a fully connected CRF. Then, the extracted complete spark image is shown in Fig. 9. It can be seen from the figure that the spark image processed by this method removes the complex background and obtains the complete spark image.

## 6 Discussion

### 6.1 U-net Algorithm for Image Segmentation

In addition, we also segmented the spark images with the U-Net algorithm. Figure 10 shows the segmentation results.

Table 4 Configuration for Yolo5

Project	Content
CPU	6core 11th Gen Intel(R) Core(TM) i5-11400F
RAM	32 GB
GPU	RTX3060 12 GB
Operating systme	Windows 10
CUDA	Cuda11.3
Software	Python3.8.8
Framework	Torch1.12

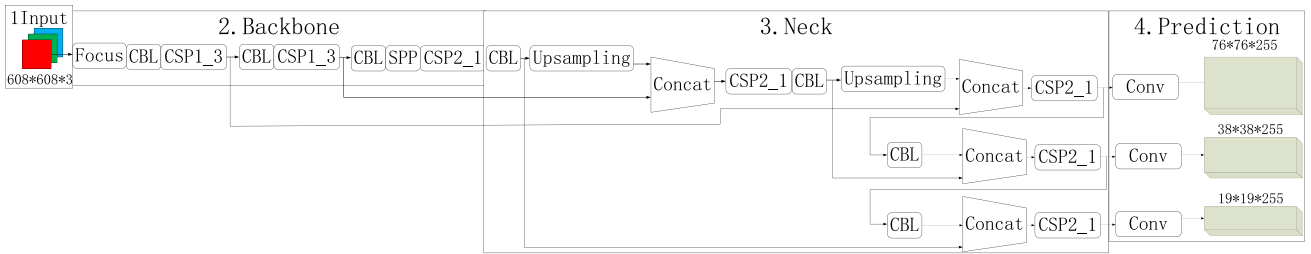


Fig. 5 YOLO5s architecture

Fig. 6 YOLO5 frontal spark image training results

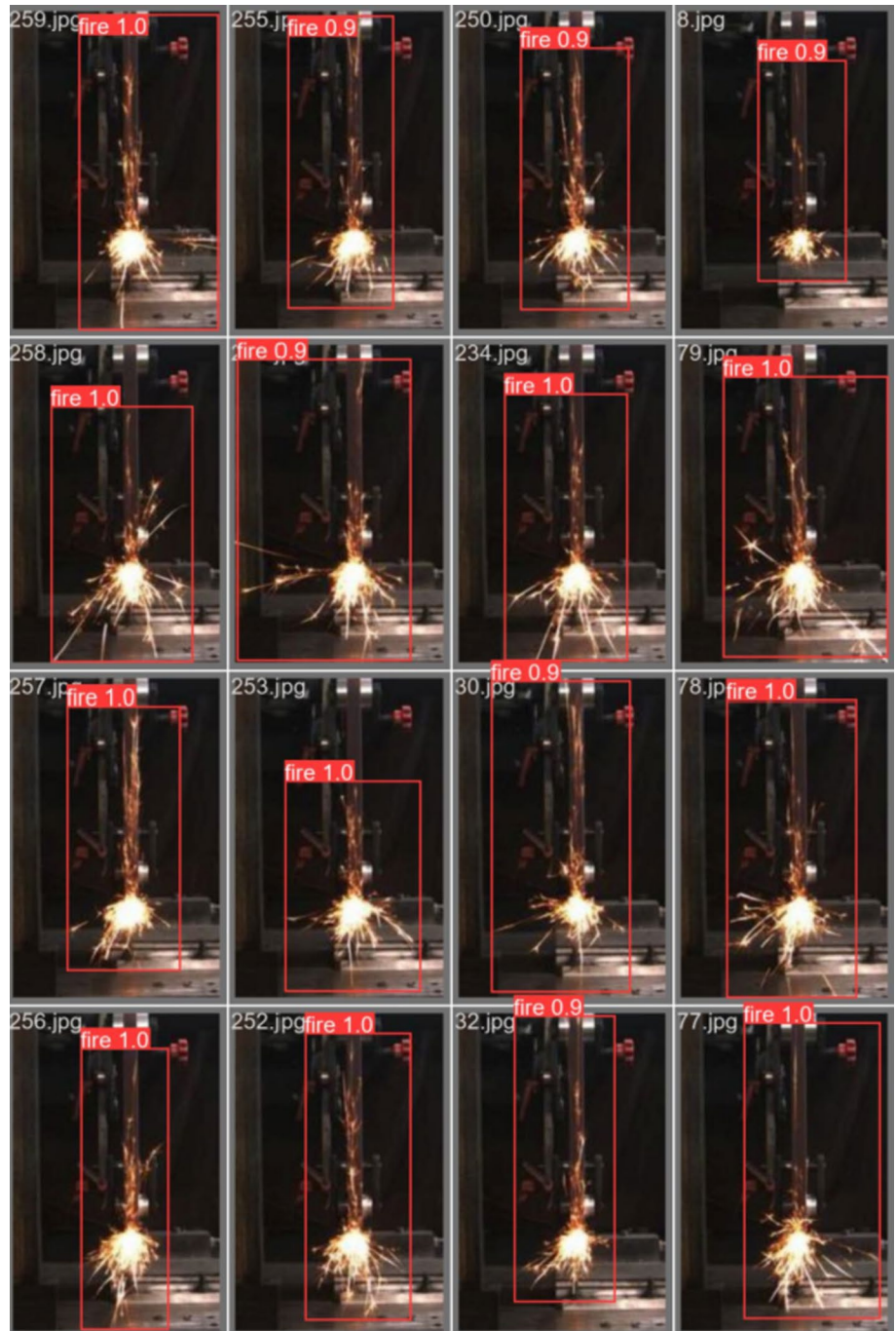




Fig. 7 YOLO5 side spark image training results

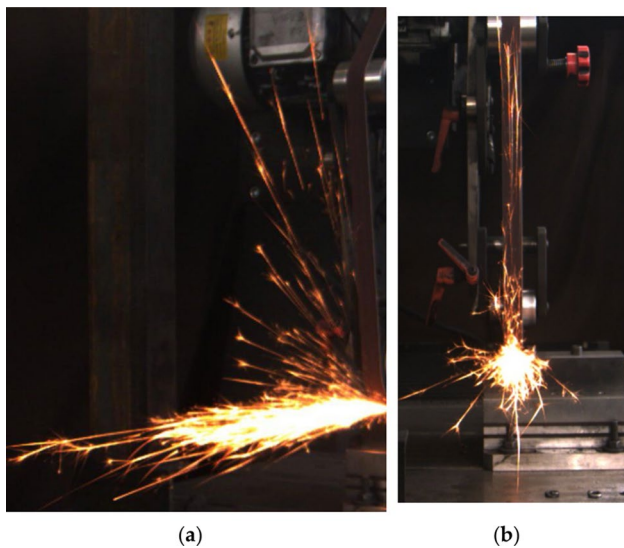


Fig. 8 a YOLO5 crop side spark image. b YOLO5 cropped frontal spark image

**Algorithm 1** Mean field in fully connected CRF [23]

---

Initialize  $Q$   $Q_i(x_i) \leftarrow \frac{1}{Z_i} \exp\{-\Phi_u(x_i)\}$

**While** not converged **do**

$\tilde{Q}_i^{(m)}(l) \leftarrow \sum_{j \neq i} k^{(m)}(f_i, f_j) Q_j(l)$  for all  $m$

$\tilde{Q}_i(x_i) \leftarrow \sum_{l \in L} \mu^{(m)}(x_i, l) \sum_m w^{(m)} \tilde{Q}_i^{(m)}(l)$

$Q_i(x_i) \leftarrow \exp\{-\Phi_u(x_i) - \tilde{Q}_i(x_i)\}$

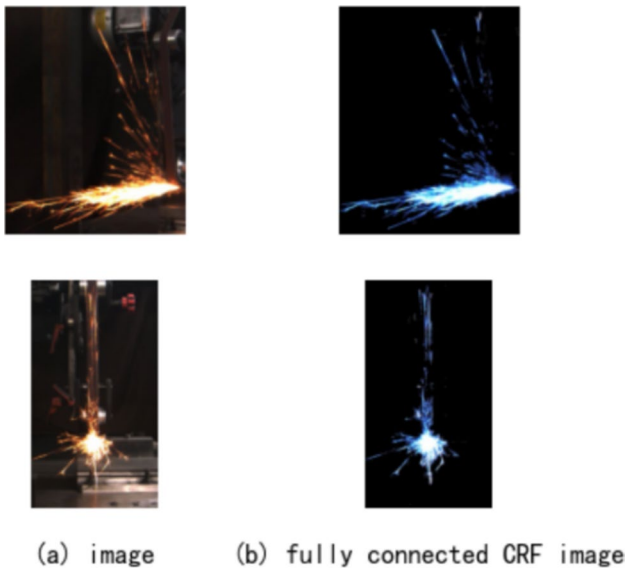
Normalize  $Q(x_i)$

**End while**

---

**Table 5** Qualitative and quantitative results on the PASCAL VOC dataset

	Runtime	Accurate ground truth	
		Global	Average
U-Net	1s	84.8 ± 1.5	82.4 ± 1.5
Fully connected CRF	0.2s	95.5 ± 0.7	94.8 ± 0.6



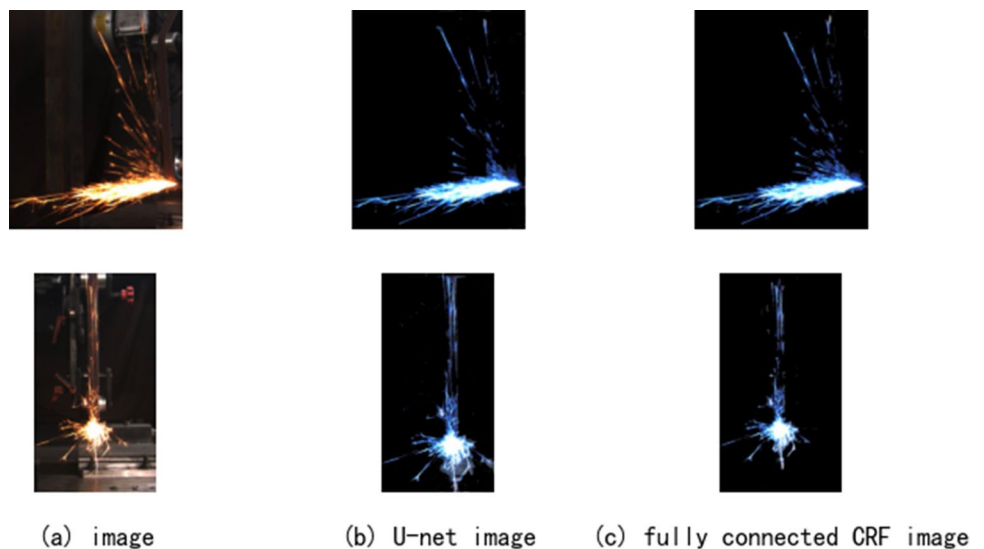
**Fig. 9** a Side spark image after image segmentation. b Frontal spark image after image segmentation

It can be seen that some of the backgrounds in the segmentation not completely removed cleanly, some of the spark images are lost. It indicates that the overall effect is not as good as using our method.

## 6.2 PASCAL VOC 2007

We also used the PASCAL VOC 2007 dataset to validate the algorithm in our experiment. We randomly partitioned the images into two groups: 80% is training set, and 20% is

**Fig. 10** Comparison of segmentation results between U-NET and CRF



the test set. Segmentation accuracy was measured using the standard VOC measure. The accuracy was shown in Table 5. Table 5 reports the segmentation accuracy against the U-Net. The segmentation images shown in Fig. 11 and in Table 5, our algorithm has improved accuracy by 10% as compared to the U-NET algorithm and faster.

## 6.3 Evaluation

Once the training completed, the optimal model was used to predict the images of the test set in the spark images dataset and PASCAL VOC dataset. The results of which are shown in Figs. 12 and 13.

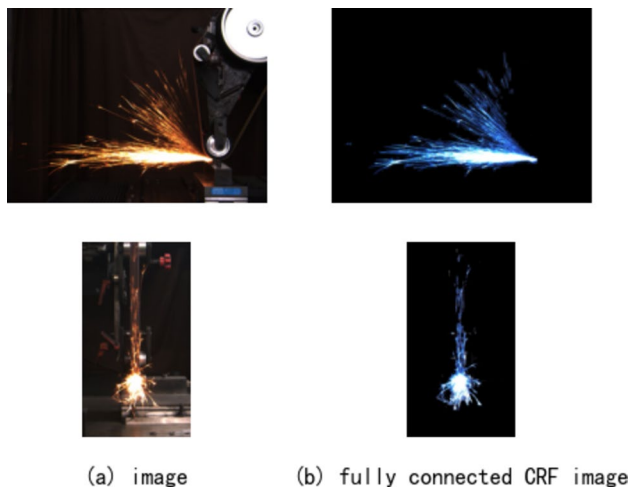
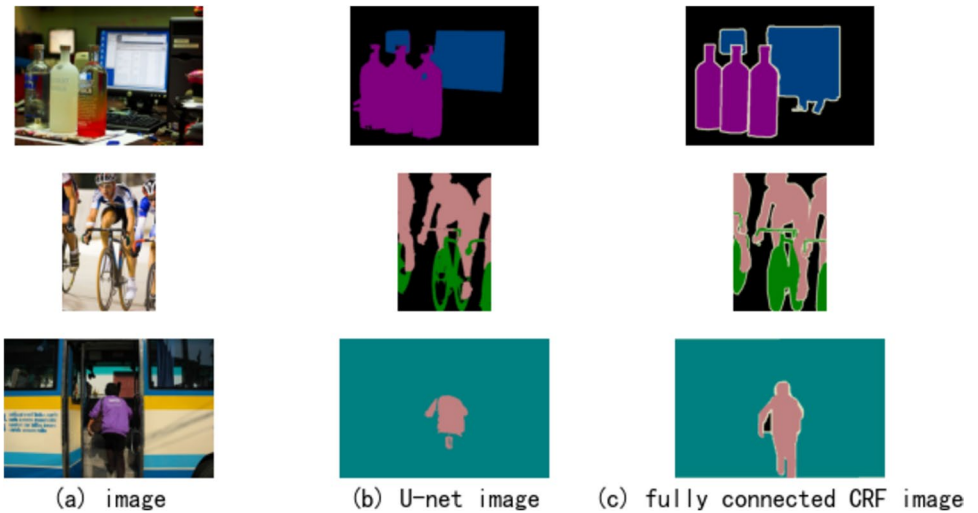
As shown in Figs. 12 and 13, when the optimal model generated by fully connected CRF was used to predict test images, the accuracy reached over 0.96, and the segmentation of a single image took 0.2 s.

## 7 Conclusion

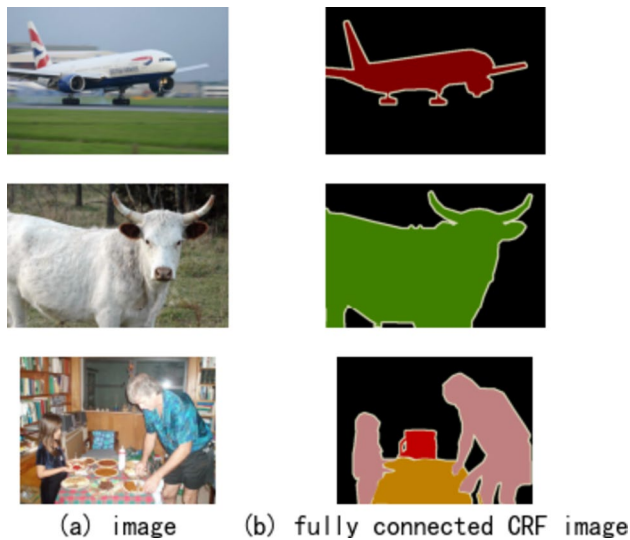
The paper uses YOLO5 + fully connected CRF to achieve the segmentation of images. The method suitable for graphics with irregular image edges and complex background images. Through experiments, we demonstrate that we successfully detected the target region image with YOLO5 to narrow the image segmentation and then further segment the region image with fully connected CRF to extract the spark image. This method provides a new idea for image segmentation. Later, we will further process the spark image, calculate the area of the spark region, and find the relationship between the spark image and the MRR.



**Fig. 11** Qualitative results on the PASCAL VOC 2010 dataset



**Fig. 12** The test results on the fire images dataset



**Fig. 13** The test results on PASCAL VOC dataset

**Author Contributions** Guang Peng Zhang planned the entire experiment. Jian Huang performed the analysis and summarized the experimental data and was a major contributor to the writing of the manuscript. Lijuan Ren and Nina Wang participated in grinding experiments. All authors read and approved the final manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding** This work was supported by the National Natural Science Foundation of China under Grant No. 52275511 (professor Guangpeng Zhang). This work was also supported by the Natural Science Basic Research Program of Shaanxi under Grant No. 2023-JC-QN-0428 (Doctor Lijuan Ren).

**Data Availability** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the data also forms part of an ongoing study.

**Declarations**

**Conflict of Interest** The authors declare no competing interests.

**Ethical Approval** All data in this paper comes from machining grinding experiments and does not involve ethical issues.

**Consent to Participate** Not applicable.

**Consent to Publish** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Fan, Z.: Low-level structure feature extraction for image processing via stacked sparse denoising autoencoder[J]. *Neurocomputing* **243**, 12–23 (2017)
2. Finn, C.: Deep visual foresight for planning robot motion[J]. *IEEE. Int. Conf. Robot. Autom.* [arXiv:1610.00696v2](https://arxiv.org/abs/1610.00696v2) [cs.LG] (2017)
3. Avelino, J.: On the perceptual advantages of visual suppression mechanisms for dynamic robot systems[J]. *Procedia Comput. Sci.* **88**, 505–511 (2016)
4. Ren, L.J., Zhang, G.P., Wang, Y., Zhang, Q., Huang, Y.M.: A new in-process material removal rate monitoring approach in abrasive belt grinding. *Int. J. Adv. Manuf. Technol.* **104**, 2715–2726 (2019)
5. Ren, L.J., Zhang, G.P., Zhang, L., Zhang, Z., Huang, Y.M.: Modelling and investigation of material removal profile for computer controlled ultra-precision polishing. *Precis. Eng.* **55**, 144–153 (2019)
6. Wang, N., Zhang, G., Pang, W., Ren, L., Wang, Y.: Novel monitoring method for material removal rate considering quantitative wear of abrasive belts based on LightGBM learning algorithm. *Int. J. Adv. Manuf. Technol.* **114**, 3241–3253 (2021)
7. Wang, N., Zhang, G., Pang, W., Wang, Y.: Vision and sound fusion-based material removal rate monitoring for abrasive belt grinding using improved LightGBM algorithm. *J. Manuf. Process.* **66**, 281–292 (2021)
8. Wang, N., Zhang, G., Ren, L., Li, Y., Yang, Z.: In-process material removal rate monitoring for abrasive belt grinding using multisensor fusion and 2D CNN algorithm. *Int. J. Adv. Manuf. Technol.* **120**, 599–613 (2022)
9. Wang, N., Zhang, G., Ren, L., Pang, W., Li, Y.: Novel monitoring method for belt wear state based on machine vision and image processing under grinding parameter variation. *Int. J. Adv. Manuf. Technol.* **122**, 87–101 (2022)
10. Wang, N., Zhang, G., Ren, L., Yang, Z.: Analysis of abrasive grain size effect of abrasive belt on material removal performance of GCr15 bearing steel. *Tribol. Int.* **171**, 107531 (2022)
11. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788
12. Redmon, J., Farhadi, A.: YOLO9000: better, faster, stronger. In: *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525
13. Redmon, J., Farhadi, A.: YOLOv3: an incremental improvement. 2018. [arXiv:1804.02767](https://arxiv.org/abs/1804.02767)
14. Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M.: YOLOv4: Optimal Speed and Accuracy of Object Detection. 2020. [arXiv:2004.10934v1](https://arxiv.org/abs/2004.10934v1)
15. Girshick, R.: Fast R-CNN. 2015. [arXiv:1504.08083v2](https://arxiv.org/abs/1504.08083v2)
16. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. 2016. [arXiv:1506.01497v3](https://arxiv.org/abs/1506.01497v3)
17. Olaf, R., Philipp, F., Thomas, B.: U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015. [arXiv:1505.04597v1](https://arxiv.org/abs/1505.04597v1)
18. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. 2018. [arXiv:1703.06870v3](https://arxiv.org/abs/1703.06870v3)
19. Lian, J., Yin, Y., Li, L., Wang, Z., Zhou, Y.: Small object detection in traffic scenes based on attention feature fusion. *Sensors* **21**, 3031 (2021). <https://doi.org/10.3390/s21093031>
20. Wenkel, S., Alhazmi, K., Liiv, T., Alrshoud, S., Simon, M.: Confidence score: the forgotten dimension of object detection performance evaluation. *Sensors* **21**, 4350 (2021). <https://doi.org/10.3390/s21134350>
21. Wang, J., Wang, N., Li, L., Ren, Z.: Real-time behavior detection and judgment of egg breeders based on YOLO v3. *Neural Comput. Appl.* **32**, 5471–5481 (2020). <https://doi.org/10.1007/s00521-019-04645-4>
22. Arunabha, R., Roy, M., Bose, R., Bhaduri, J.: A fast accurate fine-grain object detection model based on YOLO4deep neural network. *Neural Comput. Appl.* **34**, 3895–3921 (2022)
23. Philipp, K., Vladlen, K.: Efficient inference in fully connected CRFs with Gaussian edge potentials. 2012, [arXiv:1210.5644v1](https://arxiv.org/abs/1210.5644v1)
24. Huang, J., Zhang, G.: A study of an online tracking system for spark images of abrasive belt-polishing workpieces. *Sensors* **23**, 2025 (2023). <https://doi.org/10.3390/s23042025>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.