Crane Hook Detection Based on Mask R-CNN in Steel-making Plant

Shuai Kang and Hongbing Wang
University of Science and Technology Beijing, School of computer and Communication Engineering, No.30 Xueyuan Road,Beijing,China.
Email: wanghongbing0816@163.com

Abstract. This paper proposes a solution based on machine vision and deep learning for the hidden safety problems that the hook does not match the ladle trunnion correctly ,which is one of the hidden dangers during the crane lifting the ladle. Mask Region Convolutional Neural Network (Mask-RCNN) was introduced to segment the crane hook and find the bottom point of the hook contour. The trunnion center point can be directly located in image by painting a special color on it. Then determine whether the hook and trunnion match correctly by calculating the angle between the horizontal line and the line connecting the bottom point of the hook and the trunnion center point. According to the experimental results of 100 test images, that the average accuracy (AP) of our methods for hook segmentation can reach 92%. And the accuracy of the safety judgment algorithm has reached 96%.

1. Introduction
With the rapid development of the steel industry, bridge cranes are more frequently used in steelmaking process. When the crane is lifting the ladle, the driver of the crane will determine whether the hook is match with ladle trunnion correctly according to the feedback of the ground staff. If the ground staff makes a mistake, the lifting operation will be performed when the hook does not completely match the ladle trunnion and if may cause major safety accidents.

Mask R-CNN[1] algorithm is an image processing algorithm based on deep learning in machine vision, which can locate, classify and segment targets in images. In this paper, a method for crane hook detection based on Mask R-CNN was proposed. Mask R-CNN is used to accurately locate the position of the hook in the image, and pixel-level segmentation is performed to obtains hook’s binary mask. Then, the safety judgment algorithm based on the actual production is used to judge the matching state of the hook and the ladle trunnion.

In this paper, we used image data acquired from the 3D printing model of crane hook and ladle. By acquiring images under different conditions and augmenting data with noise, make the experimental data close to the actual data to the greatest extent.

2. Preparation of experimental data
The training of the Mask R-CNN network model requires a large number of images as input. The more the number and types of images, the higher the robustness of the network model. However, since the crane hook and the ladle are large industrial equipment that do not allow operation randomly, and accidents of the ladle overthrowing due to the incorrect matching of hook and ladle trunnion are rare in the actual production. Therefore, the actual data used for experiments is limited, and the training results through a small amount of data are difficult to verify the generalization and robustness of the
model. In order to solve this problem, we used 3D printing technology to print the model of the ladle and the hook, the size of which was reduced according to the actual size.

In an actual steelmaking workshop, the lighting conditions are not always the same. Due to the existence of various light sources, the images acquired by the image acquisition device may have uneven lighting. In addition, there are many large equipment in the steelmaking workshop, so the location of image acquisition device is limited, it may be far away or close to the crane hook. Uneven illumination will change the pixel values of image, and the placement distance of the acquisition equipment may make the scale of the hook in the image too small. These factors will reduce the image quality and affect the performance of the detection model. In order to verify the robustness of the model against the effects of uneven illumination and small targets, we collected images in three different conditions of lighting and distance and approximately 170 pictures were acquired in each condition. We also used three different noises to augment the data. The final data used for Mask R-CNN model training is shown in the table below. The group 1 of images was collected under normal illumination and normal distance, the group 2 was under strong illumination and normal distance, and group 3 was under normal illumination and long distance.

| Table 1. Experiment Data |
|--------------------------|
| Original image | Gaussian additive noise | Gaussian multiplicative noise | Sale and pepper noise | Total |
| Group 1         | 162                     | 162                         | 162                   | 748   |
| Group 2         | 178                     | 178                         | 178                   | 712   |
| Group 3         | 175                     | 175                         | 175                   | 700   |

3. Structure of Mask RCNN Hook Detection Model

Mask R-CNN is a two-stage algorithm in object detection and is the most advanced algorithm in a series of detection algorithms based on RCNN. Mask R-CNN is an extension of Faster R-CNN algorithm, by adding a Mask branch to the head of the model to achieve the function of instance segmentation, the network model can simultaneously perform classification, localization and segmentation tasks. The Mask R-CNN network model is mainly composed of three parts: backbone network, region proposal network (RPN) and the heads network with three branches. The image is first input to the backbone network, and a feature map of image is obtained after a series of operations such as convolution and pooling. Then the feature map is passed through the RPN network to generate regions of interest (RoIs). Finally, each branch of heads network completes its task by extracting features in the ROIs.

4. Security Judgment Algorithm

How to judge whether the hook and the ladle trunnion are correctly matched is a key problem in the crane detection system. We designed a security judgement algorithm on the basis of the actual production. It is angle-based and easy to calculate.

The security judgement algorithm needs to obtain the coordinates of two points, one is the lowest point of the hook’s outline and another is the center point of the ladle trunnion. By calculating the angle between the line connecting the two points and the horizontal line, it can determine whether the hook and the ladle trunnion are correctly matched.

The lowest point of the hook outline can be obtained by the binary mask of crane hook which is generated by Mask R-CNN model. And the center point of the ladle trunnion can be extracted from the image directly by painting a special color material to the ladle trunnion.

The security judgment algorithm flow is as in follow:
1) Find the lowest point of the hook through the binary mask
2) Connect the center point of the ladle trunnion and the lowest point of the hook, and calculate the angle between this line and the horizontal line.
3) Compare the angle with the pre-set threshold to determine matching state of hook and ladle trunnion

In our experiment, the threshold of the angle is set to 60 °, it is based on the shape of the hook and ladle. When the angle calculated by the security algorithm is greater than 60 °, we can judge that the hook and the ladle trunnion are correctly matched (see Figure 1), and when the angle is less than 60 °, it is judged that there is mismatched (see Figure 2).

![Figure 1. Results hook segmentation and safety judgment](image1)

![Figure 2. Results hook segmentation and safety judgment](image2)

5. Model Trainings

We divided all the data (original data and augmented data) into 6 groups and trained 6 models to verify the performance of the Mask R-CNN model under different illumination, distances, and interference by noise. In each dataset, the ratio of the training data to the validation data is 4:1. The detailed composition of the dataset is shown in Table 2.

| Type and number of datasets | Original data | Augmented data |
|----------------------------|---------------|----------------|
|                            | Training data | validation data | Training data | validation data |
| Group 1                    | 130           | 32             | 520           | 128            |
| Group 2                    | 144           | 34             | 576           | 136            |
| Group 3                    | 141           | 34             | 564           | 136            |

We used ResNet50 [2]+FPN[3] as the backbone network of Mask R-CNN model because the number of categories in our detection task is small (only hook and background), ResNet50 is enough to extract the feature of them, and using a pre-trained model based on the COCO dataset[4] can speed up the training of the model and optimize the learning rate of the model. The process for detecting and segmenting hooks using Mask R-CNN models as shown in Figure 3.
During the process of training six models, the training configs are completely consistent. FPN + ResNet50 is used as the backbone network. The base learning rate is 0.001 and was updated every 10k iterations by using the method of step with decay with $\Gamma$ is 0.1. The maximum number of iterations is 30k and batch size is 512. Since the size of the original images is too large (5472*3648), we resize them to 1024*684 before training in order to improve the training speed.

The training losses of model is shown in Fig 4. It can be seen that the losses are decrease during the training process. When iterations exceeds 10,000, the model's joint loss (green) tends to be stable and below 0.03, both the classification loss (red) and detection loss (yellow) are less than 0.01 and the segmentation loss is less than 0.03, it means that the accuracy of the model in classification and detection can reach more than 0.99, the accuracy of segmentation can reach more than 0.97. After iterating more than 13,000 times, all losses obtain convergence, indicating that the training process of the model is good and the model tends to be stable.

6. Result Analysis

To evaluate hook detection model, we used the evaluation indicators of the COCO dataset, average precision (AP) and average recall (AR). In addition, we also selected 100 images outside the training datasets and validation datasets to evaluate our model according the two indicators which are the integrity of hook segmentation and the accuracy of the security judgment algorithm.

We use the average precision (AP) and average recall (AR) as indicators to evaluate the hook detection model. The AP and AR of the six hook detection models are shown in Table 1 (detection task) and Table 2 (segmentation task). It can be seen from Table 1 that for the task of hook detection under the condition of normal illumination and normal distance, the AP0.5:0.95 of model is around 0.9 and when IoU = 0.75, it can still maintain an AP more than 0.9, which indicating that for 90% of the images, the detection model can correctly detect the hook target and determine its location.

When the conditions of illumination (the second group) or distance (the third group) were changed, the AP of the model decreased slightly, but AP0.5: 0.95 and AP0.75 are still above 0.8 and 0.85.
Table 2 indicates that for the hook segmentation task, under the normal illumination and normal distance, AP0.5: 0.95 is 0.828 and 0.88, and AP0.75 = 0.930, indicating that for 93% of images, the accuracy of hook’s mask generated by models is 75% at least. When the illumination or the distance changes, the model's AP0.5: 0.95 remains above 0.75, while AP0.75 remains above 0.85.

Since the security judgement algorithm used in the subsequent steps only needs to find the lowest point of the hook’s outline, it is not necessary to set the IoU threshold to be very high. In the next section we will explain that the requirements of accuracy on the segmentation task are not high in security judgement algorithm.

Since our purpose is to judge the matching state of the crane hook and the ladle trunnion, we propose two clearly indicators and use them to evaluate the hook detection model. Integrity of hook segmentation refers to whether the model can segment the complete outline of the hook that include the correct position of the lowest point. The accuracy of the safety judgment algorithm refers to whether the model can calculate the angle correctly, so as to correctly determine the matching state, without having to consider whether the hook’s outline is complete.

Fig 5-b and Fig 5-d were judged to be able to segment the hook’s outline completely, while Fig 5-a and Fig 5-c were disable. But all of them are right on accuracy of the safety judgment algorithm.

The security judgement algorithm based on the actual production situation, and can determine the matching status correctly by using only the lowest point of the hook outline and the center point of the ladle trunnion. We selected 100 images outside the training set and the validation set. The results show that 92% of the images can obtain the complete hook outline; and 96% of the images can be judged the matching status correctly.

It can be seen from Fig 5 that even if the complete hook outline cannot be obtained from the hook mask generated by hook detection model, the matching status can still be judged correctly as long as the bottom of the hook mask is complete and the lowest point can be found.

![Figure 5. Results of security judgement algorithm](image)

Fig 6 are images those unable to apply security judgement algorithm among the 100 test images. It can be seen that for the fig 6-a, fig 6-b and fig 6-c, the detection model can not segment the complete mask of hook or can not detection the hook due to the dual effects of salt and pepper noise and strong illumination, that result in the inability to find the correct lowest point to calculate the angle. As for fig 6-d, although is does not effected by noise, the material of the 3D printing model of hook and ladle are the same, the colors of them are almost the same under strong illumination. It is difficult for the human to distinguish the hook from others, so the same for the detection model.

In actual production, the difference between the color of ladle and the crane hook is very large. So this kind of failure in segmentation due to color occurs rarely.

In actual production of steel-making, the process of lifting the ladle by the hook is very slow, at least more than 5s. Our model can detection 3 images per second. Once a frame of image is detected as a potential safety hazard, an alarm signal will be issued immediately. Therefore, the missed detection rate of the safety judgment algorithm is about 4.15% in actual production, and its accuracy rate is above 99.9%, which can meet the safety requirements of industrial systems.
6

Figure 6. Images unable to apply security judgement algorithm among the 100 test images

7. Test Results on Actual Images
We also verified the feasibility of our method through some actual images. Fig 7 are images collected from a steel-making workshop of a steel plant in China. It can be seen that the hook and the trunnion can be completely found, and the accurate lowest point of hook can be found correctly. The proposed method can be applied to actual production sites.

Figure 7. Results of security judgement algorithm on actual hook images

8. Concluding Remarks
In this paper, we propose a method to check whether the crane hook matches correctly with the ladle trunnion when the ladle is lifted. Compared with the current method of obtaining matching status by ground staff, our method based on machine vision can assist the crane driver to determine the matching state of the crane hook and the ladle turnnion quickly and accurately, thus improve production safety.

9. References
[1] Kaiming, H., Georgia, G., Piotr, D., et al., 2018. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence 1–1.
[2] K. He, X. Zhang, S. Ren, J. Sun. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition (2016), pp. 770-778
[3] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, S. Belongie. Feature pyramid networks for object detection. Proceedings of the IEEE Conference on Computer and Vision Pattern Recognition (2017), pp. 936-944, 10.1109/CVPR.2017.106
[4] Lin, T.Y., Maire, M., Belongie, S., Microsoft, C.O.C.O., et al., 2014. Common Objects in. Context [J].