China Steel Technical Report, No. 34, pp.58－65, (2021)

Research of Character Recognition on Billet Surface Based on 3D Sensing

TUNG-YING WU

Green Energy & System Integration Research & Development Dept,
China Steel Corporation, Taiwan (R.O.C)

In this work, a method to detect and recognize 3D ID characters that are hard stamped on the cross section surface of a billet is proposed. Instead of the 2D camera mostly used in traditional optical character recognition (OCR), a displacement sensor is used to detect the depth variation over a cross sectional area to obtain 3D information to generate a depth map in this work. By means of a data transforming method, the obtained depth map can be transformed into a grayscale image, and the transformed image is then used for image recognition. For character recognition, YOLO architecture, which is based on a deep neural network (DNN) algorithm is applied, and two models are ultimately trained. In order to build a model of robustness against unpredictable variation and disturbance on billet manufacturing, a pseudo multi-brightness data augmentation method is proposed for generating images with various degrees of noise for model training in this work. Consequently, the experiment shows that the accuracy can exceed 98%. This work has been applied to the billet manufacturing control.

Keywords: Billets, optical character recognition, artificial intelligence

1. INTRODUCTION

Billets in CSC are made from blooms, and in the following process are then manufactured as rod and bar products. During the manufacturing process, each billet has its own specific requirement, such that to ensure the manufacturing flow and distribution of the product correctly, each billet should have its own ID for identification. As shown in figure 1, in CSC the ID is marked by hard stamping on the head cross section surface of the billet with a string of characters. After the intense bumping process, the characters are stamped with concavity, called 3D ID. Therefore, the features of the characters can be seen by means of the shadow caused from non-uniform optical reflection or diffusion on the cross section surface.

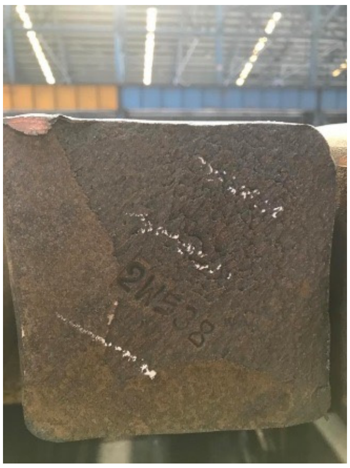

Fig.1. ID is marked by hard stamping on the head cross section surface of the billet

3D ID has the advantage that it can last longer than painted marks, being able to resist surface rusting and natural decaying and/or scratching of the cross section face. Besides, the 3D ID process does not need consumable materials like paint, so it is more economical. However, 3D ID does not discriminate from the background unlike the color contrast of painted characters, so to read 3D ID characters, whether by human eyes or camera sensor, requires shadow resulted from a light field to demonstrate the concave features of the characters. As shown in figure 2, we can see the projected light is occluded by the wall at the concave region, and it means that a shadow will occur at this region.

However, in figure 1 as well as the 3D ID characters, there may also be natural textural fluctuations and man-made painted marks on the surface of the cross sections, and can possibly result in obstruction when reading the 3D ID characters. Because of these unpredictable factors, the optical reflection-based imaging system does not have a high reliability when capturing the features of hard stamped characters. Therefore, in order for industrial application, an imaging system has to be able to sense more exclusive characteristics of the hard stamped characters and eliminate unwanted disturbances.

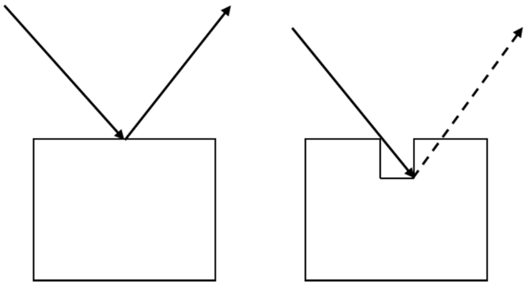

Fig.2. Reflection on the plane surface (left) and the concave surface (right).

In addition to optical hardware, character recognition algorithm software, which is used to automatically locate and interpret the characters from the input images, also needs to be emphasized. Traditional OCR (Optical character recognition) methods such as shape context feature[(2)], utilize manually designed features based on human perceptions for recognition, but always encounter difficulties on deformation and disturbance from noise. Fortunately, in recent years, CNN (Convolutional neural network) related methods have been proposed[(1)], and have proven to be of high precision and reliability in image recognition. The advantages of CNN such as its ability to cope with deformation and noise, attributes better performance in recognition compared to traditional feature-based OCR methods and can be applied in fields requiring high robust reliability. RCNN (Regional convolutional neural networks)[(3,4)] methods were proposed, and these CNN-based methods can not only recognize their targets, but also predict the positions and sizes of them at the same time, instead of the traditional OCR algorithms which need to separate each character first before recognizing all the characters.

In order to detect the features of the 3D ID characters more in accordance with characteristics, in this work, we replace the traditional 2D imaging system with distance sensors to detect the concave region directly, which is more distinctly significant for concavity. As shown in figure 3, the distance sensor detects the concavity by measuring the distance variation on the surface.

By means of the bundles of distance data captured from the cross section surface, the distance data can be interpreted into grayscale values and the maps can be transformed into grayscale images. In this work, we propose a method to automatically transform the measured distance data into images. As a result, the image based CNN can be easily applied by using the distance map as input. It realizes the automation from the front-end distance sensor to the backend intelligence computing system.

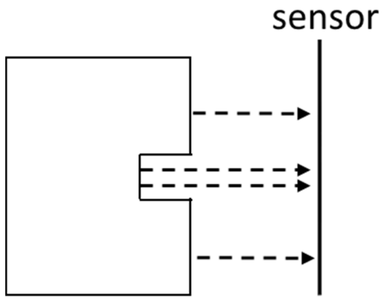

Fig.3. The distance sensor measures the distance from the target surface.

In this paper, the optical principle of laser distance sensor is introduced first, and the method to apply it on billet cross section surface scanning is also introduced. Then the algorithm to process the measured data and how to build YOLO models are in the following sections. Finally, the experiments demonstrate the performance of the method, and aside from YOLO model, Faster-RCNN are also built for comparison.

2. METHODS

2.1 Laser displacement measurement

Laser displacement sensor is one of the most used methods to evaluate 3D information of an object. As the geometric principle as shown in figure 4, a laser beam is projected vertically onto a plane, and by means of an image sensor such as CCD or CMOS, the laser light can be detected and the position where the laser light images on the sensor can also be determined. Geometrically, while the distance from the laser to the object S varies, ΔZ, the position where the laser light imaging on the sensor also varies (ΔH), in proportion to ΔZ,

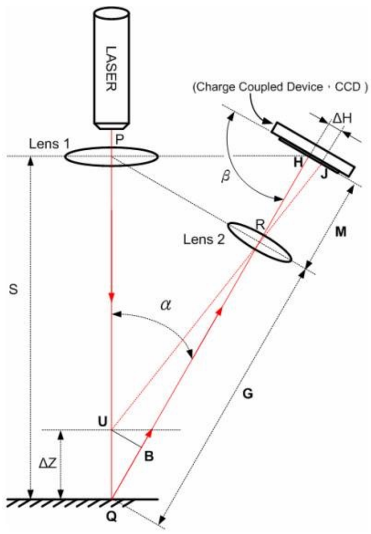

Fig.4. Optical geometry of the laser displacement measurement with the laser path vertical to the target plane.

$$\Delta Z = c \cdot \Delta H, \quad (1)$$

where c, is a constant depending on the parameters of the optical system.

While the laser path inclines and is not vertical to the target plane, the optical structure is shown in figure 5, and the optical structure in figure 4 can be regarded as a special case of figure 5.

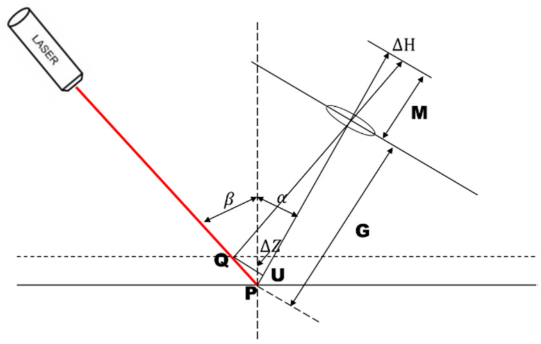

Fig.5. Laser displacement sensor optical structure with laser path tilting β degree.

The arithmetical geometry induction is shown in the following terms. ΔH, is the offset of the position where the laser images on the sensor while the height of the target plane varies ΔZ,. In Fig.5, based on similar triangle principles and geometric optics, ΔH, and $\overline{QU}$, are in proportion of γ, which denotes the magnification factor of the camera lens system, as written in the following equation.

$$\frac{G-\overline{PU}}{M} = \frac{\Delta H}{\overline{QU}} = \gamma, \quad (2)$$

The height variation ΔZ, and $\overline{QU}$, can both be derived from $\overline{PQ}$, with trigonometric functions.

$$\overline{QU} = \overline{PQ} \times sin(\alpha + \beta), \quad (3)$$

$$\Delta Z = \overline{PQ} \times cos\beta, \quad (4)$$

Combine (3) and (4), we can get

$$\frac{\overline{QU}}{\Delta Z} = \frac{\Delta H}{\gamma \Delta Z} = \frac{sin(\alpha+\beta)}{cos\beta}, \quad (5)$$

Because ΔH is obtained by the camera sensor, we can replace it with pixel size p, multiplying the number of pixels k.

$$\frac{\Delta H}{\gamma \Delta Z} = \frac{p \times k}{\gamma \Delta Z} = \frac{sin(\alpha+\beta)}{cos\beta}, \quad (6)$$

Finally, we can get

$$k = \Delta Z \times \frac{\gamma \times sin(\alpha+\beta)}{p \times cos\beta}, \quad (7)$$

In (7), it means that while the height of the target plane varies ΔZ, the position where the laser imaging on the sensor will also offset k. However, k is digital, so the minimum ΔZ, that can make k increase or decrease exactly 1 pixel, reflects the sensing resolution ability in height of this optical system. Besides, if β =0, it illustrates the displacement sensor with laser vertical to the target as shown in figure 4.

2.2 Billet cross section scanning

In order to make laser displacement measurement more effective, the point laser is replaced with a linear laser, so a row of range data (ΔZ,) can be captured once every shot by the image sensor. Therefore, similar to line-scan cameras, by means of synchronizing the shot with the motion control steps, multiple rows of range information can be combined into a so-called "range image". As demonstrated in figure 6, a row of range data reflecting the height information along the laser line can be captured once every shot.

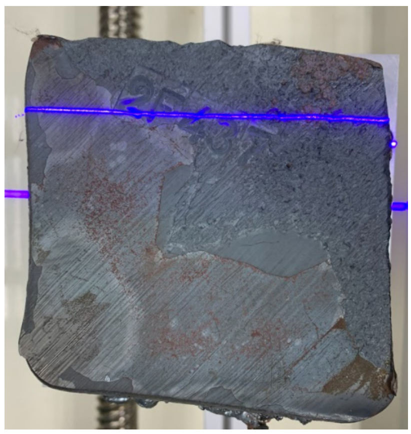

Fig.6. A laser line projected onto the cross section of a billet to detect the depth.

By shifting the laser line over the billet cross section, the range information of the cross section plane can be obtained. The corresponding range data values can be interpreted with a color palette or grayscale level, the range information can then be demonstrated as an image as shown in figure 7, and the 3D ID characters can be even more enhanced while the concave region is painted a darker color. In brief, the depth information of the cross section can be viewed.

Fig.7. Range map of a billet cross section and demonstrated in grayscale.

Obviously, by means of transforming the detected depth variation to grayscale level contrast, the concavity of characters on the billet cross section can be more salient even if there are inevitable natural textures as long as the characters are printed deep enough. Those manually painted marks in colors will disappear in this way. As a consequence, image recognition methods can be introduced to recognize the characters in the transformed gray scale images.

2.3 DNN based character recognition

Traditional character recognition methods include two main steps: segmentation and recognition. As shown in figure 8, the first step is to detect the characters and separate them into several bounding boxes. Then, the second step is text recognition of the character in each bounding box.

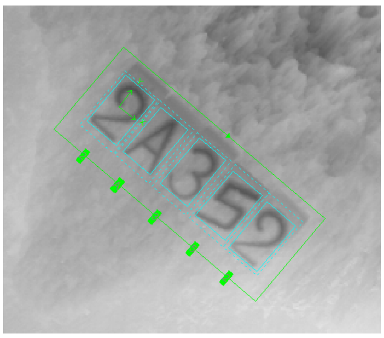

Fig.8. Character recognition with traditional method.

In other words, if the segmentation process fails, recognition will also eventually fail. Segmentation often fails because of a complex background, and unfortunately, the natural textures and scars on the billet cross section often affect the segmentation seriously, like the case in figure 9.

It is that the segmentation is searching for relatively strong features, and the unwanted features on the background will certainly interfere with the recognition process.

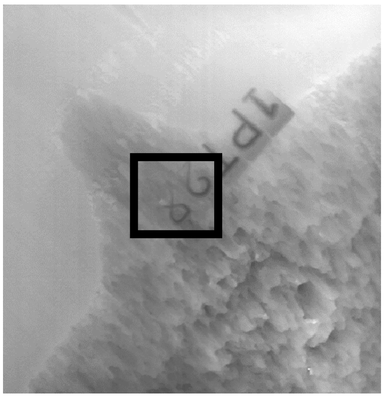

Fig.9. Characters affected by the complex background on the billet cross section surface.

One of the drawbacks of traditional character recognition algorithms is the lack of tolerance and robustness. Therefore, to solve the problems, methods based on convolutional neural network (CNN) are proposed. CNN has been proven that it can perform well on recognition even under noise and deformation as long as the model is well trained. Furthermore, in recent years, Regional convolutional neural network (RCNN) methods are also proposed. Besides recognition, RCNN-based methods involve a localization algorithm into the architecture, so the outputs of RCNN-based methods include not only the class of the input but also the position and size of the anchor boxes of the targets. Several methods have been proposed, including RCNN which is proposed first, fast RCNN, faster RCNN and YOLO[5].

YOLO is a the method known for speeding up the detection process of RCNN. In the first proposed RCNN or even the faster RCNN, the scanning process will inevitably put the predefined anchor boxes to test if this region contains a target or not. However, for fear of loosing targets, the denser the search is, the higher the precision is, but eventually the more time consuming it becomes. Therefore, YOLO gave up the scan-and-test process, and by means of the multi-scale anchor box method and bounding box regression inherited from the Faster-RCNN, the targets can be found and bounded. After the third version of YOLO, called YOLOv3, was proposed, YOLO has been proven to have better performance on precision and speed. The architecture of YOLO algorithm is illustrated below in figure 10.

At first, several anchor boxes are defined preliminarily, and these anchor boxes are put on to the tiles which the image is divided into. Then by means of the convolutional neural network in the YOLO architecture, each tile is evaluated for targets inside. If the tile contains a target, the class of the target is output, also the size of the anchor box most fitted to the target is also output. Finally, the class of the target and the location where the target is both obtained by this single stage

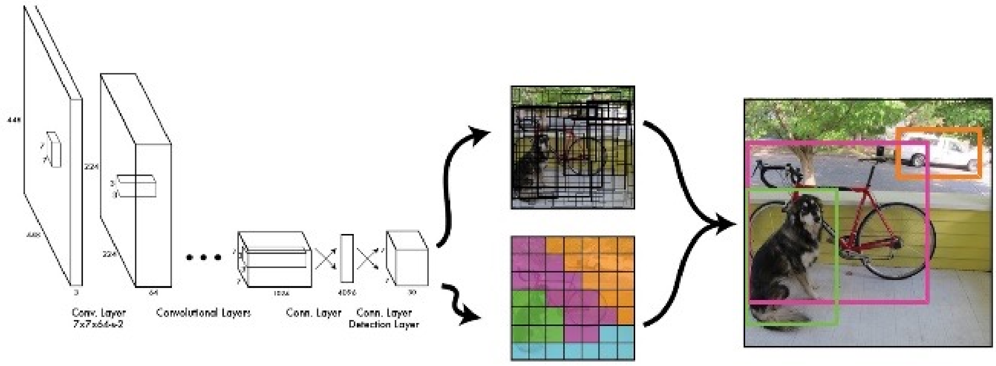

Fig.10. The process flow of YOLO

CNN-based architecture. The loss function of YOLO includes three parts:
(1)Classification loss
(2)Localization loss
(3)Confidence loss

Classification loss is used to evaluate the classification results, and localization loss is used to evaluate the accuracy of the target position and size it predicts. Confidence loss is to evaluate if a tile contains a target. By combining the three parts, we can get

$$\lambda_{coord}\sum_{i=0}^{s^2}\sum_{j=0}^{B}1_{ij}^{obj}[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2]$$

$$+\lambda_{coord}\sum_{i=0}^{s^2}\sum_{j=0}^{B}1_{ij}^{obj}\left[\left(\sqrt{w_i}-\sqrt{\hat{w}_i}\right)^2\right.$$

$$\left.+\left(\sqrt{w_i}-\sqrt{\hat{w}_i}\right)^2\right]$$

$$+\sum_{i=0}^{s^2}\sum_{j=0}^{B}1_{ij}^{obj}\left[\left(C_i-\hat{C}_i\right)^2\right]$$

$$+\lambda_{noobj}\sum_{i=0}^{s^2}\sum_{j=0}^{B}1_{ij}^{noobj}\left[\left(C_i-\hat{C}_i\right)^2\right]$$

$$+\sum_{i=0}^{s^2}1_{ij}^{obj}\sum_{c\epsilon class}[(p_i(c)-\hat{p}_i(c))^2], \qquad (8)$$

where λ ,means learning rate. 1_i^{obj},=1 if and object appears in cell *i*, otherwise 0, and 1_i^{noobj},is the complement of 1_i^{obj} . C_i, is the box confidence score of the box in cell *i*. The loss function considers including the predicted position (*x, y*) and size (*w, h*) of the bounded box, the class of the target (p(c)) belonging to and the confidence whether the divided sub-region of the image contains a target. The loss function implies that both bounding box information and the class of the target are trained at the same stage, instead of the two stage architecture like Faster-RCNN requiring scanning-and-test searching procedure.

In Redmon ed al. (2016)(6), the speed of YOLO is tested to be more than 5 times faster than Faster-RCNN, and even about 100 times faster than the original RCNN. With YOLO, the traditional character recognition algorithm with two steps can be substituted. YOLO provides a single stage method that can divide the characters and recognize each divided character at once.

2.4 Model training and data augmentation

YOLO model training is the same procedure as training the traditional CNN model. By labeling a number of images and inputting them into the network, the parameters in the network will be iteratively revised by back-propagation based on the loss function until convergence.

YOLO is said to have poor performance on an objenc which is too small or too large when when comparing to the anchor boxes. Fortunately, the size of the characters on the billet are fixed. However, the difficulty that may be encountered is the complex background from the scars or textures of the billet cross section surface. In order to train a model which is robust enough to resist the unpredictable disturbance, in this work, a pseudo multi-brightness method is proposed.

2.4.1 Pseudo-brightness transform

In section 2.2, we mentioned that the range information captured from a billet cross section will be transformed into an image. It means that the range data value within a certain period will be converted into color or grayscale levels to illustrate the difference in value with grayscale level contrast. The corresponding relationship can be demonstrated with a curve, as shown in figure 11., and the higher the contrast of the image, the steeper the slope of the transforming curve is.

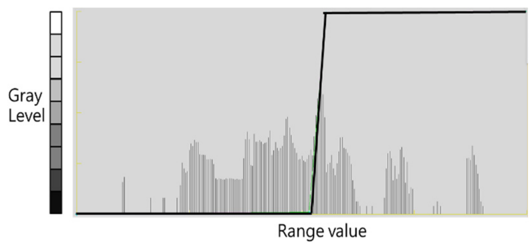

Fig.11. Range data transforming and the transforming curve.

In other words, if the transforming curve is changed, the brightness and contrast of the transformed image will also be different. Therefore, in order to improve the robustness of the model against noise, in a training step, we can make data augmentation by creating multiple images with various transforming curves from a range image. For example, three transformed results from the same range image using different transforming curves are shown in figure 12, and we can see the textures in the background make different degrees of disturbance on the characters.

In practice, the parameters of the corresponding curve are randomly adjusted while each range image is processed and multiple transformed images can be finally obtained.

With these transformed images of various brightness, two YOLO models are trained in this work. The first is used to find the region where the characters are localized and then the characters can be rotated to a horizontal arrangement. The second model is used to recognize each character within the extracted region in the first model. Because the characters are hard stamped diagonally on the billet cross section surface, the first model is also used to figure out the orientation of the characters. As a result, the characters can be rotated based on the result. Four determined classes corresponding to the four types of orientation of the characters are illustrated below.

In figure 14, the characters are localized and bounded, and also the direction of orientation is also recognized. figure 15. shows the localized region in Fig.14 after being rotated.

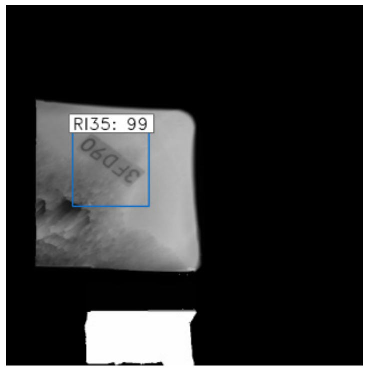

Fig.14. Characters on the billet cross section are located and the orientation direction is recognized.

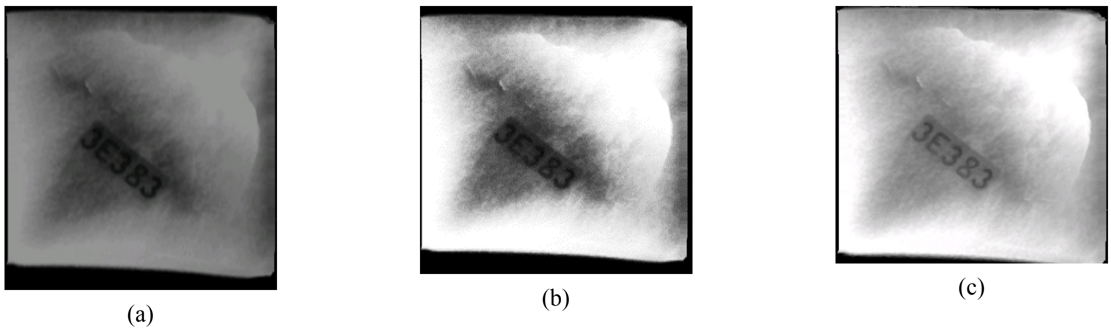

Fig.12. Transformed images with different transforming curves.

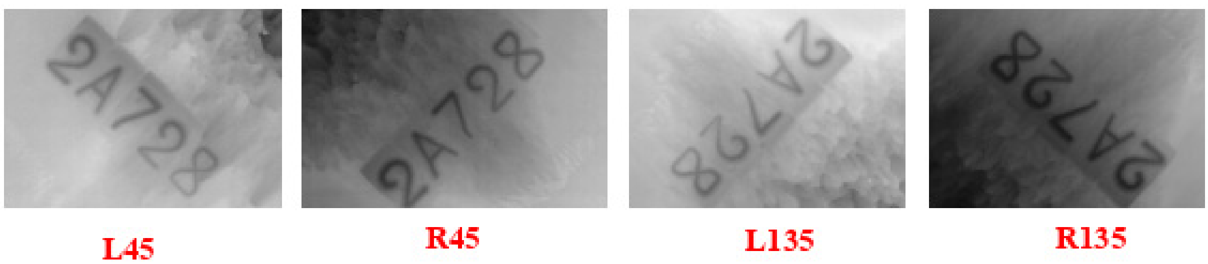

Fig.13. Four examples of orientation of the characters.

Fig.15. Characters found in figure 14 after rotation.

Fig.16. A bad result of the range data brightness transforming process.

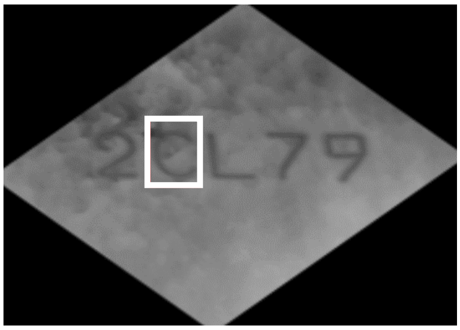

Fig.17. Recognition error of character caused by the background noise.

After the characters are rotated to the right orientation, the second model recognizes the characters in the extracted image output from the first model. Based on the position information of each character output bundled with the recognition results, the characters are arranged and the billet's ID can be obtained automatically.

3. EXPERIMENT METHOD

In this work, range image data captured from 5000 billets were used to train the two YOLO models. The 5000 range images were transformed to 15000 images with pseudo multi-brightness. Another 10000 billet range images were used to evaluate the model performance. The two models are evaluated respectively, and the input of the second model is the output of the first model.

4. RESULTS AND DISCUSSION

The experiment results are shown in Table 1.

In the first model of YOLO, most of the incorrectly predictions are because of the bad results from the range data transforming process, as shown in figure 16.

As to the second YOLO model, most of the incorrect results were because of the inevitable background noise, or the characters were not stamped deep enough for capturing their features. Figure 17 shows a recognition error of the second character "C", because the scars on the surface were too deep and created a disturbance on the features of the character such that the model wrongly recognized the character as "E".

The training and test data were also used to train a Faster-RCNN model, and the experiment results are shown in Table 2.

The experiment results show that the first Faster-RCNN model performs as well as the first model in YOLO, but the second model is obviously worse than YOLO. The reason is also because of the background noise. As mentioned above, The faster-RCNN scans the whole image to find where the possible targets are, but

Table 1 Experiment results of YOLO

Model	Precision rate
1^{st} Model	98.2%
2^{nd} Model	99.3%

Table 2 Experiment results of the Faster-RCNN

Model	Precision rate
1^{st} Model	97.4%
2^{nd} Model	90.3%

the salient features caused by the background noise may attract the anchor boxes to make meaningless recognition calculations. As a result, the number of located anchor boxes exceeds the upper limit, the searching process will stop, and some characters may be neglected while the scars or textures on the background are obvious in some images. In figure 18, the obvious features are regarded as possible targets and were found before the characters because the search pattern starts from the top-left of the image.

For fear of losing targets, the upper limit of the number of searched targets can be increased. However, this will increase time spent on the recognition process.

Fig.18. The salient features found by Faster-RCNN.

5. CONCLUSION

In this work, we propose a method for the purpose to automatically recognize the billet 3D ID which are concavely hard stamped on the cross section surface of the billet. The architecture of this method includes the distance sensor to detect the depth information of the billet cross section, the transforming curve to convert the obtained range data to grayscale images, and finally the YOLO models to recognize the ID characters. Because most modern methods for efficient manufacturing control such as barcode or surface printing, as used in the manufacture of small items such as integrated circuits or electronic devices, are not appropriate for steel manufacturing which are of high temperature, high humidity and dusty. The difficulties which may be encountered in the steel making fields are quiet different from electronic device manufacturing. Vibration and dust is the most challenging for optical devices, including cameras and laser displacement sensors. Vibration may cause distortion or blurring on the acquired images, and dust may scratch or occlude the lens. Steel billets are heavy, so it is almost impossible to control them as precisely as in electronic device manufacturing. Therefore, in order to approach high reliability for on-line application, not only optical design should consider the tolerance for position error but the software should be capable of dealing with the variations.

Consequently, the proposed method was successfully applied to reading the 3D IDs on cold billets. In the future, with necessary optical and software improvement and reforming, this architecture will be applied on other fields, such as hot billets, blooms and even slabs.

REFERENCE

1. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton., Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems, 2012, vol. 25, pp. 1097-1105
2. Belongie, Serge, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. IEEE transactions on pattern analysis and machine intelligence, 2002, vol. 24(4), pp. 509-522.
3. Girshick, Ross, et al., Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.
4. Ren, Shaoqing, et al., Faster r-cnn: Towards realtime object detection with region proposal networks, Advances in neural information processing systems, 2015, vol. 28, pp. 91-99.
5. Redmon, Joseph, et al., You only look once: Unified, real-time object detection, Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
6. Redmon, Joseph, and Ali Farhadi., YOLO9000: better, faster, stronger, Proceedings of the IEEE conference on computer vision and pattern recognition, 2017.