Label Design and Identification in High Temperature Logistics Based on Concave Coding and DeepLabV3+ Model

Xiaoyan Zhao[*1,2], Pengfei Zhao[*2], Yuguo Yin [*3], Luqi Tao[*1] and Zhaohui Zhang[*1,2]

[*1] School of Automation and Electrical Engineering, University of Science and Technology Beijing 30# Xueyuan Road, Haidian District, Beijing 100083, China
E-mail: zhaoxiaoyan@ustb.edu.cn
[*2] Shunde Graduate School, University of Science and Technology Beijing Fo Shan 528399, China
E-mail: zpf18810075175@163.com
[*3] Shandong Start Measurement and Control Equipment Co., Ltd, Weifang, Shandong, 261101, China
E-mail: ygyin@startdy.com

Abstract. With the development of the Internet of Things, the logistics tracking technology at normal temperature has been quite mature, but there are few tracking methods for products in the high-temperature production process. The main difficulties include that the label materials cannot withstand high temperature for a long time, and the detection devices are vulnerable to environmental impact. This paper takes the logistics tracking of carbon anode for aluminum electrolysis as the application background to carry out specific research. This paper designs a high temperature resistant label, that is, printing concave coding on the surfaces of the anode carbon blocks, which ensures the integrity of labels in high temperature environment without affecting product performance. There is no obvious color distinction between the coding area and the background, so the direct recognition is not effective. Therefore, based on DeepLabV3+ model, MobileNetV2 and Xception are selected as the backbone network to semantically segment the concave coding in order to realize the effective segmentation of the coding area and background, and lay the foundation for the subsequent recognition of the concave coding. The experimental results show that the DeepLabV3+ model based on MobileNetV2 is superior, with MIoU of 95.45% and MPA of 97.38%. In summary, the scheme is feasible and provides a new idea for the development of high-temperature logistics tracking technology in smart manufacturing.

Keywords: Semantic Segmentation, DeepLabV3+, High Temperature Logistics

1. INTRODUCTION

With the advancement of society, the development of science and technology, and the rapid changes in production management technology, in order to ensure production and further improve efficiency, logistics tracking becomes very necessary. Currently commonly used logistics label coding technology has the following kinds: RFID radio frequency identification[1], a typical short-range wireless communication technology, can realize the non-contact automatic rapid identification. Bar code and two-dimensional code[2] due to the large amount of information, high reliability and other advantages in the logistics tracking system is widely used. These technologies usually work in low temperature or normal temperature conditions, and have been quite mature. However, there are few product tracking methods in the process of high-temperature production. Most of the existing methods are based on RFID combined with the actual need to adjust or change the antenna structure. High temperature insulation filler is added so as to achieve high temperature resistance[3]. But the applicable temperature is usually below 500 ℃ [4]. Or tags are made by some high temperature-resistant special materials which are studied for a long time. And tags, on which bar code or two-dimensional code is printed, are fixed in the product by rivets or other means . In this paper, we take carbon anode for aluminum electrolysis as the research target. During its production and use, it not only has to endure long time roasting at up to 1200℃, but also has to face harsh environment such as scratching and dust. Therefore, the above logistics tracking technology is not fully applicable, mainly because the label material cannot resist high temperature for a long time; the label is not resistant to scratch, easy to fall off and wear; the detection device is vulnerable to environmental impact. So the study of a new label coding detection technology for high-temperature products can help solve the logistics tracking problem of these industrial products and lay the foundation for the full realization of intelligent manufacturing.

2. LABEL DESIGN AND IDENTIFICATION IN HIGH TEMPERATURE PRODUCTION PROCESSING

To realize the logistics tracking of anode carbon block in the whole manufacturing process, the label coding is very critical. The carbon anode has a stable geometry after forming, usually cuboid, and then is subjected to high temperature of about 1200° C during the roast stage. Since the parameters of the roast stage are very important, so the logistics tracking of the anode carbon block is considered from the forming stage. In addition, the concave coding is identified after each process, related information about this anode carbon block are viewed and recorded. Due to the raw materials and

manufacturing process used in the anode carbon block, it has the characteristics of low resistivity, compactness and pressure resistance.

Based on the above characteristics of the anode carbon block, a design of a high-temperature label is proposed, that is, a preset workpiece coding groove is printed at the center of the surface of the anode carbon block, an example is shown in "**Fig 1**". The space occupied by the bit of the coding is about 5cm long square, and the number of digits of the coding is determined by the average annual total output of the anode carbon block of the enterprise. Every bit of the coding has 10 possibilities from 0 to 9. Large-capacity coding can be came true in a limited coding area, which greatly meets the needs of enterprises. Moreover, this solution utilizes the wear resistance and scratch resistance of the anode carbon block itself to ensure that the label is not easily affected to maintain detectability in an ultra-high temperature environment. In addition, the concave coding can be printed in the early stage of workpiece manufacturing, which is easy to operate and low in cost. As a part of the workpiece, the concave coding does not affect the product performance during the workpiece's subsequent entire life cycle, and does not need to be recycled.

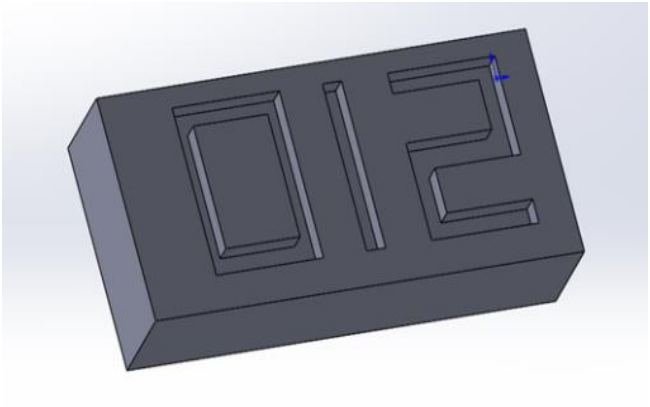


Fig. 1 Concave coding example.

For such a logistics label coding scheme, the original detection and identification means are no longer applicable, and a new decoding method needs to be considered. The anode carbon block is gray-black as a whole, and there is no obvious color distinction between the coding area and the background. It is worth considering that how to exclude the interference of similar background areas, accurately find the real coding area and complete the final concave coding identification. What is more, the grooves of the coding area may fall into dust and suffer from abrasion during the production and use process, which is also a problem that needs to be considered.

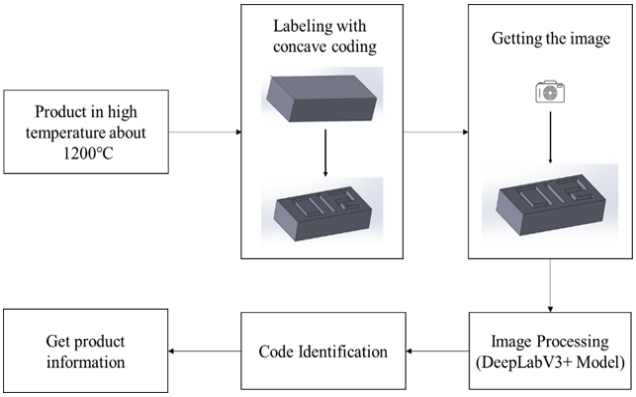


Fig. 2 Simulate vertical shot encoding area.

Based on the above analysis, it is considered to identify the concave coding on the anode carbon block through image recognition technology, so as to obtain the logistics information of its entire life cycle, the identification process is shown in "**Fig 2**". The ideal image recognition method is to obtain a clearly identifiable coded image through image processing, and then to achieve the decoding function through character segmentation and recognition. In this study, deep learning algorithm is used to perform semantic segmentation of coded and non-coded regions to obtain coded images.

3. METHODOLOGY

3.1. Experimental data

The sample used in the experiment is made of black nylon, which is used to simulate the anode carbon block. The concave coding images used in the experiment were all taken vertically by an iPhone. The shooting distances were 10cm, 15cm, and 20cm, and the shooting conditions were natural light and vertical illumination of auxiliary light source. The sampling data is quite different, which is convenient to enhance the robustness of the model. Artificial data enhancement, including flipping, rotating, changing brightness and so on, is used to expand the number of samples. Twice the amount of data is obtained by this way so that the model gets stronger generalization ability[5]. Among them, 550 concave coding pictures are selected for data labeling. An example of the test data is shown in "**Fig 3**". The picture on the right is the labeling picture of the original picture on the left, and the size of the picture is (512,512). The value of each pixel of the labeling picture represents the category it belongs to. In this paper, there are two categories, where category 0 is set to black (representing the background), and category 1 is set to red (representing the concave coding). All data is divided into training set and validation set according to the ratio of 9:1. Finally, unlabeled images are selected for experiment to test the performance of the model.



Fig. 3 An example of the test data.

3.2. Experimental model

The semantic segmentation model used in this study is DeepLabV3+. In order to integrate multi-scale information, it introduces the Encoder-Decoder form. The main structure is shown in "**Fig 4**".

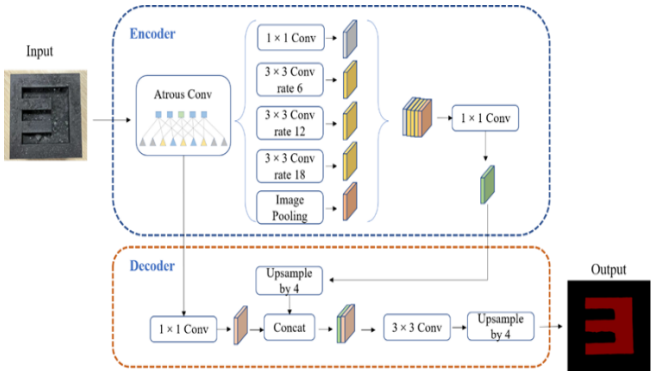


Fig. 4 DeepLabv3+ model structure.

Let's first look at the Encoder part, which has two core points:

1) Use serial atrous convolution in the backbone deep convolutional neural network (DCNN). The introduction of atrous convolution makes it possible to extract more effective features. In other words, without losing information, the receptive field is increased, so that each convolution output contains a larger range of information[6].

2) After the image passes through the DCNN, the result is divided into two parts, and one part is directly passed to the Decoder part. The other part uses parallel atrous convolution with different expansion rates to extract features. After merging the feature layers obtained, 1x1 convolution is undergone to compress the features. The latter is also known as Atrous Spatial Pyramid Pooling (ASPP) to achieve multi-scale feature extraction.

Let's look at the Decoder part. Its input consists of two parts: one part is the output of DCNN, and the other part is the result of DCNN output after ASPP. These two results are fused together after certain processing to obtain the final result[7].

3.3. Evaluation indicators

3.3.1. Training model evaluation indicators

For classification models, precision and recall are commonly used as evaluation metrics. Its definition is as follows:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

In the equation (1) and (2): TP indicates that it is predicted to be a groove and is actually a groove; FP indicates that it is predicted to be a groove, which is actually the background; TN indicates that it is predicted to be the background and is actually the background; FN indicates that it is predicted to be the background, but the actual is a groove.

The two indicators of Precision and Recall are often mutually restrictive. Ideally, it is hoped that both indicators are high, but often when Precision is high, Recall is low; when Recall is high, Precision is low. Therefore, it is needed to comprehensively weigh these two indicators, which leads to a new indicator F-score. F-score is the harmonic value of Precision and Recall, which is defined as:

$$\mathrm{F-score} = (1 + \beta^2)\frac{Precision \times Recall}{\beta^2 \times Precision + Recall} \tag{3}$$

Refer to equation (3), β represents the weight coefficient. When $\beta = 1$, it is called F1-score. At this time, the precision and recall weights are the same. When $\beta > 1$, it indicates that the model is evaluated with Recall; when $\beta < 1$, it indicates that the model is evaluated with Precision.

Another evaluation training metric is the loss function value, which consists of two parts: Cross Entropy Loss and Dice Coefficient Difference Function (Dice Loss).

$$loss = lossc + lossd$$
$$= -\sum_{i=0}^{N} |X| \log(|Y|) + 1 - \frac{2|X \cap Y|}{|X| + |Y|} \tag{4}$$

In the equation (4): N represents the number of categories of classification, $|X|$ and $|Y|$ respectively represent the number of elements of the labeled image and the predicted image, and $|X \cap Y|$ represents the number of intersection elements of the labeled image and the predicted image.

3.3.2. Semantic segmentation evaluation indicators

Generally speaking, the evaluation indicators of semantic segmentation include Mean Intersection over Union (MIoU) and Mean Pixel Accuracy (MPA).

MIoU is the average of the intersection and union ratio of each class in the data set. The so-called intersection and union ratio is the ratio of the intersection and union of the actual value and the predicted value. The calculation is shown in equation (5):

$$MIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}} \times 100\% \quad (5)$$

where k is the number of categories, there are $k+1$ categories in total (including one background category); p_{ii} is the number of correctly classified pixels; p_{ij} is the number of pixels predicted to be background; p_{ji} is the number of pixels predicted to be foreground.

MPA is to first calculate the pixel accuracy (PA) of each class, that is, the number of correctly predicted pixels / the total number of predicted pixels, and then average the PA of all classes. The calculation is shown in equation (6):

$$MPA = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}} \times 100\% \quad (6)$$

4. EXPERIMENTAL

The experiment is based on the 64-bit Windows10 operating system , using the Pytorch deep learning framework with Python version 3.7. The hardware equipment used had the following configurations : the CPU is Intel(R) Xeon(R) W-2125, the main frequency is 4.0GHz, the memory is 32G , and the graphics processor is NVIDIA GeForce GTX 1080 with 8GB of RAM.

Using DeepLabV3+ semantic segmentation model, the backbone feature extraction network is Xception and MobileNetV2. Both Xception and MobileNet use depthwise separable convolution, but the purpose of the two is different. Xception uses depthwise separable convolution while increasing the amount of network parameters to improve effectiveness of this structure. MobileNet uses depthwise separable convolution to compress and speed up, and the number of parameters is significantly reduced.

The change of the loss value with epoch during the model training process is shown in "**Fig 5**". It can be seen that the loss value decreases as the epoch increases. The loss value of DeepLabV3+ / MobileNetV2 model converges faster than DeepLabV3+ / Xception; when the epoch exceeds 70, both curves tend to be flat; when the epoch is 100, the loss value of DeepLabV3+ / MobileNetV2 is 0.0584, the loss value of DeepLabV3+ / Xception is 0.0767. The closer the loss value converges to zero, the larger the overlap area between the labeled image and the predicted image.

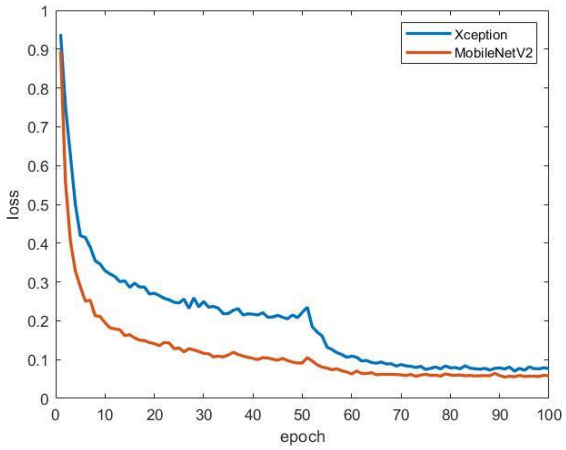


Fig. 5 Changes in loss during model training.

The change of the F1-score value with the number of iterations epoch during the model training process is shown in "**Fig 6**". It can be seen that the F1-score increases with the increase of the epoch. The F1-score value of DeepLabV3+ / MobileNetV2 model converges to 1 faster than DeepLabV3+ / Xception; when the epoch exceeds 70, both curves tend to be flat; when the epoch is 100, the F1-score value of DeepLabV3+/ MobileNetV2 is 0.975, the F1-score value of DeepLabV3+ / Xception is 0.966. The closer the F1-score value is to 1, the more accurate the model output, and the less the number of misclassifications.

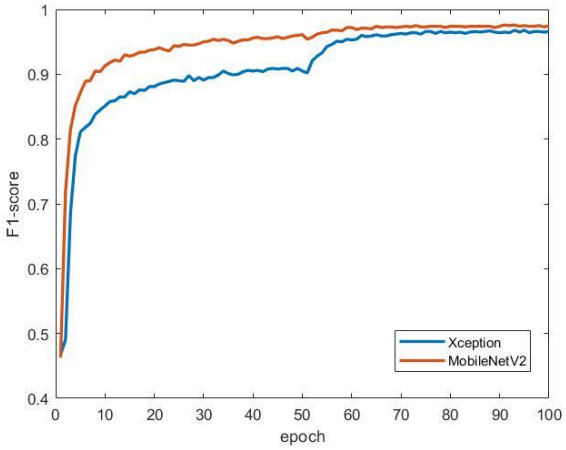


Fig. 6 Changes in F-score during model training.

In order to further evaluate the accuracy of the two segmentation methods, the MIoU and MPA values of DeepLabV3+ / MobileNetV2 and DeepLabV3+ /Xception are calculated respectively. The results are shown in Table 1. From the evaluation indicators, the segmentation accuracy of the two backbone networks for the background is very high, both of which are about 97%; for the recognition of grooves, DeepLabV3+ /

MobileNetV2 is slightly better than DeepLabV3+ / Xception.

Table. 1 Quantitative evaluation data of segmentation performance based on two extraction networks.

Network Model	DeepLabV3+/ Xception		DeepLabV3+/ MobileNetV2	
	MIoU	***MPA***	***MIoU***	***MPA***
Groove	90.8%	95.18%	92.91%	96.62%
Background	96.99%	97.37%	97.99%	97.98%
Mean	93.9%	96.28%	95.45%	97.38%

The unlabeled original image is used to test the segmentation effect of the network model. The experimental environment and network parameters are set in the same way. Part of the segmentation results are shown in "**Fig 7**" and "**Fig 8**".

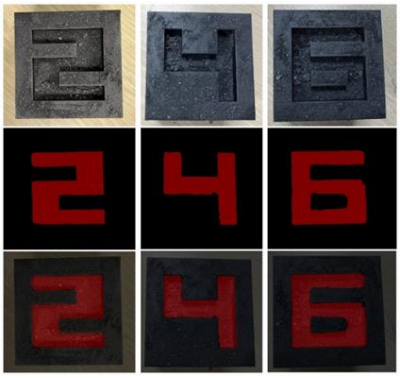


Fig. 7 DeepLabV3+/Xception segmentation example.

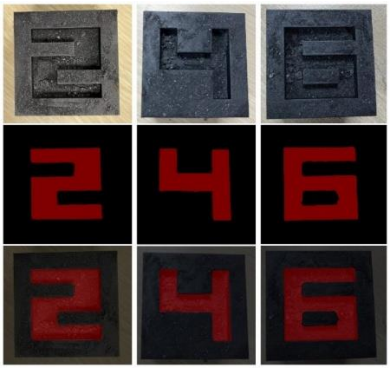


Fig. 8 DeepLabV3+/MobileNetV2 segmentation example.

"**Fig 7**" is an example of DeepLabV3+/Xception segmentation, and "**Fig 8**" is an example of DeepLabV3+/MobileNetV2 segmentation. The first row of the image is the actual captured image, the second row is the semantically segmented image, and the third row is a blend of the original image and the segmented region. Comparing "**Fig 7**" and "**Fig 8**", both DeeplabV3+/Mobilenetv2 and DeeplabV3+/Xception segmentation results are very good, but DeeplabV3+/Mobilenetv2 performs better in some details.

The MobileNet model is a lightweight deep neural network, and its requirements for computer computing power and video memory are significantly lower than those of the Xception model. To sum up, DeepLabV3+/MobileNetV2 performs better in the recognition of concave coding.

5. CONCLUSION

Based on the problem of concave coding detection, a related image dataset is made and the semantic segmentation model DeepLabV3+ with the two backbones feature extraction network of MobileNetV2 and Xception are selected to construct a concave coding detection model. It is a great method to improve the segmentation performance of coding region and noncoding region on the anode carbon block. The test results show that both models can correctly find the position of the concave coding, but the DeepLabV3+/MobileNetV2 model can achieve better results, MIoU of 95.45% and MPA of 97.38%. The follow-up work will expand more data sets , optimize the network to obtain a better result of detection and start the next step of identification work. At the same time, the feasibility of the scheme is verified by experiments, which provides a new idea for the logistics tracking of high-temperature industrial products and lays a foundation for the further realization of intelligent manufacturing.

Acknowledgements

This work was supported by Scientific and Technological Innovation Foundation of Shunde Graduate School (BK19BE020).

REFERENCES:

[1] R. Want, "An introduction to RFID technology," in IEEE Pervasive Computing, vol. 5, no. 1, pp. 25-33, 2006.

[2] X. He, "The Two-Dimensional Bar Code Application in Book Management," 2010 International Conference on Web Information Systems and Mining, pp. 409-411,2010.

[3] E. Mayer, I. Shrena, D. Eisele, J. Bardong, M. Schmitt and L. M. Reindl, "Characterization of langasite as a material for SAW based RFID and sensing systems at high temperatures," 2009 IEEE MTT-S International Microwave Workshop on Wireless Sensing, Local Positioning, and RFID, pp. 1-4, 2009.

[4] V. Franchina, A. Ria, A. Michel, P. Bruschi, P. Nepa and A. Salvatore, "A Compact UHF RFID Ceramic Tag for High-Temperature Applications," 2019 IEEE International Conference on RFID Technology and Applications（RFID-TA）, pp. 480-483, 2019.

[5] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," Convolutional Neural Networks Vis. Recognit, Vol.11, pp.1-8, 2017.

[6] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation,'' Proc. Eur. Conf. Comput. Vis., pp. 801–818, 2018.

[7] B. Baheti, S. Innani, S. Gajre, and S. Talbar, "Semantic scene segmentation in unstructured environment with modified DeepLabV3+," Pattern Recognition Letters, Vol.138, pp.223-229, 2020.