Research on Image Recognition of Machine Tool Based on Semantic Segmentation

Ke Xu\textsuperscript{a,c}, Zekai Fei\textsuperscript{a}, Yuan Sun\textsuperscript{b,*} and Hao Ming\textsuperscript{a}

\textsuperscript{a}College of Computer Science, South-Central University for Nationalities, Wuhan, China
\textsuperscript{b}Hubei University of Traditional Chinese Medicine, Wuhan, China.
\textsuperscript{c}Shenzhen Research Institute of Huazhong University of Science and Technology, Shenzhen, China

Email: cole.xu@gmail.com

Abstract. In the task of image classification of machine tool machining, the coolant water column will block the tool and interfere with the image classification. To solve this problem, firstly, the coolant water column is removed from the image by semantic segmentation. Because the number of samples in the data set is insufficient, the effect of semantic segmentation model is not very good. These results are taken as the input number of image classification model According to, it will directly affect the accuracy of classification, but it is impossible to modify the parameters of semantic segmentation model during classification training. In this case, the image after semantic segmentation processing is combined with the original image according to certain rules, combined with the information of some original images, and then classified. This method can effectively reduce the interference to the classification model when the semantic segmentation model is wrong.

1. Introduction
In intelligent manufacturing, general requirements of high precision parts need to be conducted on machine tool cutting for the final processing, by cutting the artifacts are often expensive, if in the process of cutting on the machine tool, cutting tool case of damage and fracture of the situation, it is likely to lead to damage to the cutter error of work piece, cause irreparable damage to expensive artifacts, so in the case of tool case of damage or fracture, timely identifying whether the tool has been damaged, which shut down machine work, there is a big could save artifacts, prevent caused irreversible damage to the cutter.

In the process of the machine tool processing, usually need to use water cannons spray cooling liquid, cooling on cutting tool for operation, to prevent the friction in the process of cutting tool in cutting of overheating and melting, but ejected a column of water cannon coolant will be covered in cutting tools, thus influence the cutting tool out not appear damage criterion, so if you can before the image identification, the pretreatment of the coolant water removal, will greatly improve the identification accuracy of tool damage. Because there is no occluded dual data set of the tool, semantic segmentation and image repair cannot be used to remove the occluded water column directly.

2. Related Works
In the image classification task, there is often more than one person or object in the image, just like a picture of a desk. If it is not a newly bought desk, there are usually some books on it, maybe a water bottle, a computer, a socket, a pile of folders and so on. In this case, the computer "doesn't know"
which object we want it to classify, and other objects may interfere with it; For example in this paper will discuss the machine work in the image, it needs to be identified only for cutting tool, cutting tool is a silvery white metal color, but after cutting the workpiece, cutting tool, workpiece were scattered around the workpiece fixed metal pieces, metal instruments also are all the metal color of silvery white, so they can produce greatly interfere with the recognition of the cutting tool work, it is very important to get rid of the interference at this time. Or analogy in the operation of the human image classification, before we categorize things a judgment, first of all, of course, we will first make the judgement of the objects in the line of sight range, image input for computer, that is, and then humans tend to view focuses on the object, we need to decide which step is equivalent to a computer to semantic integral operation, the computer's "vision" should also focuses on the need to classify it distinguish objects, thus reducing, the interference of other irrelevant objects, on the other hand is also calculated pixels of unnecessary items in order to reduce processing costs.

In target detection was born on the basis of semantic segmentation, semantic segmentation will need to be as target detection, detection of objects in the box, but it's not just use a box will be detection of objects in the framed, but along the boundary of the object will need tested object to identify areas, such as when you need to tested object is a football, then semantic segmentation of the frame is a circle, when you need to tested object is a person's time, the frame semantic segmentation is humanoid box. Or the nature of semantic segmentation of image classification, but it is to classify each pixel in the image, or a picture of a table, just take pictures of a pixel point, semantic segmentation can tell us, the pixels, exactly is the pixels on the table, or pixels of books on the table, or computer on the desk of pixels. When we need to do the image classification task, if it is for the table, we can take out the pixels belonging to the table in the image; if it is for the computer classification task, we can take out those pixels belonging to the computer in the image. In addition, there are instance segmentation, panorama segmentation and other different items detection technology.

3. Theory and Method

A CNN utilizes the convolution operation, which can preserve the spatial information of original images. CNN architecture shares parameters over the entire image region, thus, it is naturally robust to overfitting and has a transition invariant attribute. If an input of a certain layer \( x \) in Convnet has the dimensions \( h \times w \times c \), where \( h \) is the height, \( w \) is the width, and \( c \) is the number of channels, the output of the layer \( z \) is calculated as

\[
\{ z_n = f \left( \sum_{m=1}^{c} W_{(m,n)} \times x_m + b_n \right) \mid 1 \leq n \leq c_{out} \}
\]  

(1)

Where \( W(m, n) \in \mathbb{R}^{h \times w} \) is the kernel for the \( n \)-th output from the \( m \)-th input, \( x_m \in \mathbb{R}^{h \times w \times c_{in}} \) is the \( m \)-th channel of the input, \( b_n \in \mathbb{R} \) is the bias, and \( f \) is an activation function. The height and width of the output of the convolutional layers can be maintained by padding input features, or reduced by adding strides to a convolution operation.

4. Experiment

4.1. Machine Tool Dataset

First of all, machine tool working video frame by frame extraction of images, because the data is extracted by frame, the original video can be extracted more than 30 images in seconds, and that more than 30 pictures almost no difference, that is to say, this is 30 almost identical pictures, in the subsequent model training, the training too much exactly the same as the picture, also is the equivalent of a picture repeated training a lot of time, this will cause model fitting; Image data redundancy in the image, prevent the classification model training when fitting, and reduce the production of semantic segmentation data set of work, because the semantic segmentation is to classify each pixel of the image, so the production data set needs to be two different categories separated along the boundary pixels, than making general image data sets, a lot more complicated, in this case, reduce the production of semantic segmentation data set the amount of work it is very important; Watch video,
cutting tool and the work pieces have been in the position of the close to the center of the image, so the selection of image (1280, 720) directly from the center area the size of the cut out for (572, 572), after cutting work again, in order to be prudent, the images were given after the cut, found that really cuts after all include the tool pixels in the image, both undamaged cutter knives have been damaged. Then each pixel of image tag, will be used in cutting tools and cutting tools to fixed object pixel, tag names are written for the knife, cutting tool this kind of pixels to tag all is white, the other to another area of the pixel, tag name other, belongs to all other classes of pixel labeled black, this is a semantic segmentation of data sets. Secondly, images without semantic segmentation marks, images with tool damage are put into one category, and images without tool damage are put into another category. This is the data set of image classification.

![Image](image.png)

a) The original image of the tool without damage  

b) Sample image without damaged tool marking

c) The original image of the damaged tool  

d) Sample image of damaged tool marking

**Figure 1.** Tool processing images

### 4.2. Experimental Results

The error of the trained model in the training set reached 0.0624, and the accuracy reached 98.42%. The output results of the images of the two test sets after model operation are shown in Fig. 2.
Figure 2. Semantic segmentation results

Fig. 2 there is no damage on the four knives, picture 3-2 under four is damaged knives, for cutting tools, and without any damage model of the effect is good, although achieved a very low error, also reached a high precision, but after all is not error reached 0, precision has not reached 100%, will have some error, just like the picture on the right is a model calculated the damage of the cutting tool, model recognition becomes a tool, the part of water obviously nozzle jet of water also can have certain effect on semantic segmentation model. Of course, this may also be because in this experiment, only one picture is used in one batch during training. Due to the limitation of hardware and the large video memory required by this model, only one picture can be trained in one batch in the environment of this experiment. In the next chapter, how to process the image input to improve the accuracy of image classification in the case of some errors in semantic segmentation will be discussed.

For different processing methods, the training results of deep learning model are shown in Table 1.

| Type of image processing                  | loss  | accuracy | precision | recall | F1    |
|------------------------------------------|-------|----------|-----------|--------|-------|
| Original image unprocessed               | 0.709 | 0.573    | 0.628     | 0.571  | 0.598 |
| Image semantic segmentation              | 0.452 | 0.868    | 0.791     | 0.709  | 0.748 |
| Image semantic segmentation processing   | 0.361 | 0.952    | 0.924     | 0.729  | 0.815 |

The image data set without any processing is disturbed by the water column of the coolant or the background, and the fitting accuracy is only 57.3%, with an error of 0.709, accuracy rate of 0.628, recall rate of 0.571, and F1 value of 0.598. Through semantic segmentation model to get rid of coolant water column, has markedly improved the classification model, accuracy increased to 86.8%, the error is reduced to 0.452, accuracy rate increase to 0.791, the recall rate is increased to 0.709, F1 value increased to 0.748, compared with the control group without any processing of the data set, the model precision is improved by 31.5%, the error reduced 0.295, accurate rate increased by 0.163, the recall rate increased by 0.150. And after dealing with the semantic segmentation model of image and the original image are synthesized, in addition to the disturbance of a certain semantic segmentation model error, further improve the classification model, the accuracy is 95.2%, the error is reduced to 0.361, the accurate rate is increased to 0.952, the recall rate is increased to 0.729, F1 value increased to 0.815, compared with the control group, without any processing error reduced 0.348, reduced nearly half, accurate rate increased by 0.296, the recall rate increased by 0.158, F1 value increased by 0.216, compared with using semantic segmentation of data set, Error decreased by 0.9, accuracy increased by 0.9, accuracy increased by 0.13, recall rate increased by 0.02, F1 value increased by 0.7. Figure 3 shows the ROC curve obtained by the three pretreatment methods of training classification model.
Figure 3. The ROC curve of Semantic segmentation results

Fig. 3 shows the AUC of the unprocessed data set in training, the data set processed by semantic segmentation model and the data set processed by synthesis. The training effect of SVM model is shown in Table 2, and the ROC curve is shown in Fig. 5. When using the other two data sets to train the SVM model, no matter how to adjust the size of the image data, or the penalty coefficient, kernel function, kernel function parameter, residual convergence condition and other parameters of the SVM model, the training accuracy is only about 0.50. The SVM model was used to train the data set after semantic segmentation, and the accuracy reached 0.979, accuracy 0.915, recall rate 0.707 and F1 value 0.797. In terms of accuracy and AUC values, SVM is used to train the data set with semantic segmentation and image size adjusted to (28, 28), which has the best effect.

Figure 4. The AUC values of three cases

Figure 5. The ROC curve of SVM model

Table 2. Training results of SVM model

| Training results | loss | accuracy | precision | recall | F1    |
|------------------|------|----------|-----------|--------|-------|
| Semantic segmentation | -    | 0.979    | 0.915     | 0.707  | 0.797 |

5. Conclusion

This paper first introduces the related situation of machine tool processing, then the system introduced the image classification and semantic segmentation and related technology research, through semantic segmentation reduces the machine tool processing image, the coolant water interference, in the absence of shade and keep out the dual image data sets, semantic segmentation model to eliminate
obstructions enough problems effectively, provides the solution to deal with it. The effect of image classification is compared with that of three processing machine tool data sets, and its validity is verified by experiments.

The improvement proposed in this paper has been proved to be effective by the above experiments at the present stage, but due to the lack of time and the limitations of the experimental hardware conditions, there are still many areas to be improved. In the training of semantic segmentation model, the memory of experimental computer is insufficient, so only one picture can be trained in a batch. Moreover, the input of the image classification model is, to some extent, the output of the semantic segmentation model. Therefore, if the semantic segmentation model has been wrong in the first place, the image classification model can hardly be classified correctly. How to solve this problem is the next research focus. Finally for this problem this paper proposed a conjecture, direct use of semantic segmentation model classification task, since the semantic segmentation model is essentially classifying pixels, then maybe can direct classification machine image pixels which belonged to the damage of the cutter, which some pixels belongs to the undamaged cutter, thus avoiding the front after a model for the influence of a model. In this paper, no good results have been achieved in the training of SVM model by using unprocessed data set and synthesized data set. If appropriate parameters can be found later, some improvements may be made.

6. Acknowledgments
This research is supported by the Science and Technology Development Center Fund of the Ministry of Education of China (2018A02037), and the Industry-university Collaborative Education Project (Baidu Project, the second batch in 2018) for Higher Education Department of the Ministry of Education (A preliminary study on the training mode of artificial intelligence innovative talents based on the second class learning based on Paddle Paddle). Yuan Sun is the corresponding author.

7. Reference
[1] Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of Go without human knowledge[J]. Nature, 2017, 550(7676): 354-359.
[2] Criminisi A, Perez P, Toyama K, et al. Region filling and object removal by exemplar-based image inpainting[J]. IEEE Transactions on Image Processing, 2004, 13(9): 1200-1212.
[3] Suykens J A . Support Vector Machines: A Nonlinear Modelling and Control Perspective[J]. European Journal of Control, 2001, 7(2):311-327.
[4] Shelhamer E, Long J, Darrell T, et al. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
[5] Badrinarayanan V, Kendall A, Cipolla R, et al. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
[6] Chen L, Papandreou G, Kokkinos I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
[7] Niu Z, Liu W, Zhao J, et al. DeepLab-Based Spatial Feature Extraction for Hyperspectral Image Classification[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(2): 251-255.
[8] Shin H, Roth H R, Gao M, et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning[J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1285-1298.
[9] Krizhevsky A, Sutskever I, Hinton G E, et al. ImageNet classification with deep convolutional neural networks[J]. Communications of The ACM, 2017, 60(6): 84-90.
[10] Zhang K, Zuo W, Chen Y, et al. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising[J]. IEEE Transactions on Image Processing, 2017, 26(7): 3142-3155.