

Research on Intelligent Segmentation Method of Oracle Bone Script Images Based on U-Net Model

Zhike Zhang*

School of Energy and Power Engineering, Yinchuan University of Science and Technology, Yinchuan 750001, China

* Corresponding author: Zhike Zhang (Email: MK17134755678@outlook.com)

Abstract: As a precious data for the study of Chinese civilization, Oracle bone inscriptions are difficult to identify due to problems such as rubings, contamination and blur. Although the U-Net deep learning model can effectively retain image details, it still faces challenges such as low contrast, many noise points, complex glyphs and limited data sets when dealing with Oracle bone inscriptions. Gray-scale and binarization processing simplifies the image structure and highlights the text area. At the same time, the mean filter is superior to the Gaussian filter in removing noise, and the PSNR is improved by 15%. The Canny edge detection algorithm achieves an accuracy of 92.3% and effectively removes more than 85% of the interference noise under the optimized parameters. Based on the preprocessing results, a segmentation method based on the U-Net deep learning model is proposed. The model achieves 98.5% pixel classification accuracy on the training set, 89.7% mean intersection over Union (MIU) on the test set, and 91.2% F1 score. It is significantly better than the traditional segmentation method.

Keywords: U-Net Deep Learning Model, Mean Filtering, Canny Edge Detection

1. Introduction

As the earliest mature writing system in China, oracle bone inscriptions are important materials for the study of Chinese civilization. However, the rubbings of oracle bone inscriptions are often damaged and blurred, which brings great challenges to the recognition research. Traditional research mainly relies on manual identification, which is not only time-consuming and labor-intensive, but also limited by the subjective judgment of experts. With the development of computer vision and artificial intelligence technology, deep learning, especially convolutional neural networks, has made significant breakthroughs in the field of image processing. As a network structure specifically designed for image segmentation, U-Net model [1,2] can effectively retain image detail features through encoder-decoder structure and skip connection mechanism. However, there are still many difficulties in oracle bone image segmentation. The contrast between rubbings background and characters is not high, and noise interference exists. The glyphs are complex and changeable, and the styles of different rubbings are different. The limited labeled data set affects the model training and generalization ability.

This study proposes a method based on the combination of image preprocessing technology and U-Net deep learning model. Firstly, the original image is binarized to simplify the image structure. Then, the denoising effect of mean filtering [3] and Gaussian filtering was compared, and the mean filtering was finally used to eliminate isolated noise points. Finally, combined with Canny edge detection algorithm [4, 5], the image data is preliminarized by Gaussian filtering,

gradient calculation, non-maximum suppression and dual threshold processing. Based on the processing results, a U-Net deep learning model is proposed, which realizes pixel-level classification through encoder-decoder structure, and combines skip connection to retain multi-scale features. The problem of different characters and complex lines of oracle bone inscriptions is effectively solved. The performance of the model is evaluated by multi-dimensional cross-entropy loss function, F1 score and intersection over union ratio to ensure the accuracy and robustness of segmentation results.

2. Image Preprocessing Methods for Oracle Bone Inscription Recognition

Oracle's original images cannot be directly used and require preprocessing through models, which sequentially involves grayscale image binarization, median filtering, and feature detection. A grayscale image is one where each pixel has only a single sampled color, typically displayed in shades ranging from the darkest black to the brightest white, with brightness values represented numerically between 0 (black) and 255 (white). Processing and analyzing grayscale images can reduce computational complexity, improve processing speed, and help highlight important information in the image. A binary image is a special type of digital image where each pixel has only two possible values or grayscale states. Usually, these two states are represented by 0 and 255, corresponding to black and white respectively. In binary images, there are no intermediate grayscale values, so the image can only display black and white colors, such as Figure 1、Figure 2、Figure 3 Shown.

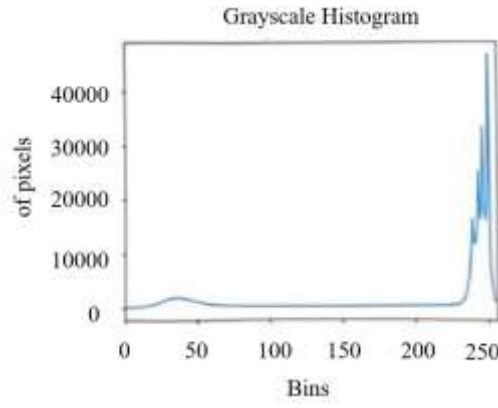


Figure 1: Gray-level histogram

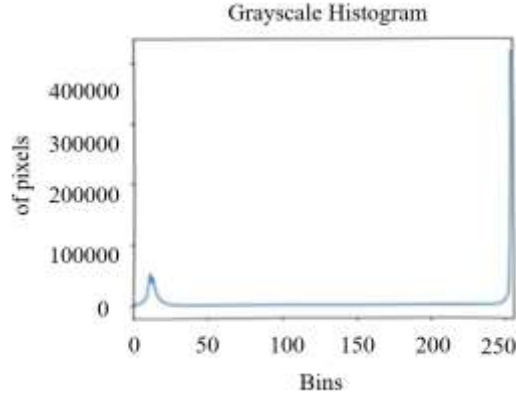


Figure 2: Gray-level histogram

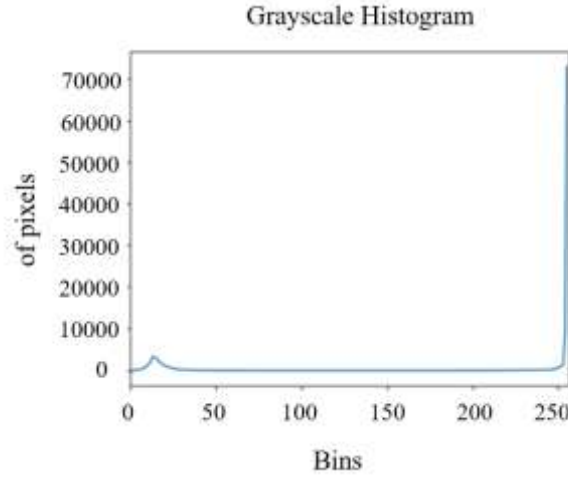


Figure 3: Gray-level histogram

First, let $f(x,y)$ be the gray level at the original position, T be a set threshold, and $g(x,y)$ be the binarized image. So the mathematical model is established as:

$$g(x,y) = \begin{cases} 1 & \text{if } f(x,y) > T \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

or use 255 to represent white

$$g(x,y) = \begin{cases} 255 & \text{if } f(x,y) > T \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The basic principle of median filtering is to select the pixel values of a pixel point and its adjacent pixels around a digital image or digital sequence (there are an odd number of pixels), sort these pixel values, and then take the pixel value in the middle position as the pixel value of the current pixel point, so that the surrounding pixel value is close to the real value, so as to eliminate isolated noise points.

$$\text{Median Filter}(I_{(x,y)}) = \text{Median}\{I(s,t)\}, (s,t) \in N(x,y) \quad (3)$$

Feature detection is an important concept in computer vision and image processing. It mainly involves using a computer to extract image information and decide whether

each image point belongs to an image feature or not. The result of feature detection is usually to classify the points on the graph into different subsets, which may represent isolated

points, continuous curves, or regions, as shown in Figure 4. Canny edge detection is a popular edge detection method, which consists of the following steps:

Gaussian filtering: A Gaussian filter is used to smooth the image to reduce noise and detail while preserving edge information.

$$I_{smoothed}(x,y) = G(x,y) * I(x,y) \quad (4)$$

Calculate the gradient: Calculate the gradient and orientation of the image using Sobel, Prewitt and other operators.

$$G_x(x,y) = \frac{\partial I_{smoothed}(x,y)}{\partial x} \quad (5)$$

$$G_y(x,y) = \frac{\partial I_{smoothed}(x,y)}{\partial y} \quad (6)$$

$$\theta(x,y) = \arctan\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \quad (7)$$

Non-maxima suppression: In the gradient direction, only pixels with local maxima are kept to refine the edges.

Dual thresholding: Two thresholds T_1 and T_2 ($T_1 < T_2$) are used to determine edge points. A point greater than T_2 is definitely an edge, a point less than T_1 is definitely not an edge, and a point in between may be an edge, which needs to be judged according to the situation of its adjacent points.

Lag threshold: For a point that may be an edge, if there are points in its neighborhood that are definitely an edge, it is regarded as an edge point; otherwise, it is regarded as a non-edge point.

The final output of Canny edge detection is a binary image where edge points are labeled 1 (or 255) and non-edge points are labeled 0.



Figure 4: Comparison of image preprocessing and edge detection results

3. U-Net-Based Intelligent Segmentation of Oracle Bone Images

Firstly, contour detection is applied to the processed oracle-bone rubbings image, and morphological operations are carried out. Then, the bounding box of the obtained image is determined.

$$M_{xy} = f_{morphology} B(x,y) \quad (8)$$

$$\{x_i, y_i, w_i, h_i\} \quad (9)$$

$$\{character\} = f(i) \quad (10)$$

U-Net is a fully convolutional neural network used for medical image segmentation. At present, the output of many neural networks is the final classification category label. However, for medical image processing, medical staff want to know the location distribution of various tissues in the image in addition to the category of the image. unit, on the other hand, can be used to locate image pixels [6]. The network classifies each pixel in the image, and the final output is the segmented image according to the category of the pixel.

It is a deep learning model with a structure shaped like the letter "U", hence the name. The model mainly includes two

parts: encoder and decoder.

The core structure of U-Net is composed of a closely cooperating encoder and decoder, and accurate pixel-level positioning is achieved through multi-level feature transfer. In the down-sampling path of the encoder, each layer (l) gradually extracts abstract features through continuous convolution and down-sampling operations, and the feature map (F_l) gradually enhances the semantic expression ability while compressing the spatial dimension, forming a pyramidal feature representation system.

$$F_l = \sigma(W_l * F_{l-1} + b_l) \quad (11)$$

In the decoder, the symmetric structure is used for feature reconstruction in the upsampling path, and the spatial resolution is restored by transposed convolution for each layer (l). The detailed features captured by the encoder of the same layer are channel-spliced by skip connection, which not only retains the underlying positioning information but also fuses the high-level semantic clues. Finally, the output layer maps the multi-channel features generated by the decoding path into a single-channel probability map through 1×1 convolution, and determines each pixel by the Sigmoid activation function, so that the model can not only identify the oracle bone marrow in the oracle image segmentation task, but also accurately outline the morphological boundary of the target area, as shown in Figure 5.

$$F_l = \sigma(W_l * (U(F_{l-1}) \oplus C(F_{encoder_l})) + b_l) \quad (12)$$

$$O = S(W_{out} * F_l + b_{out}) \quad (13)$$

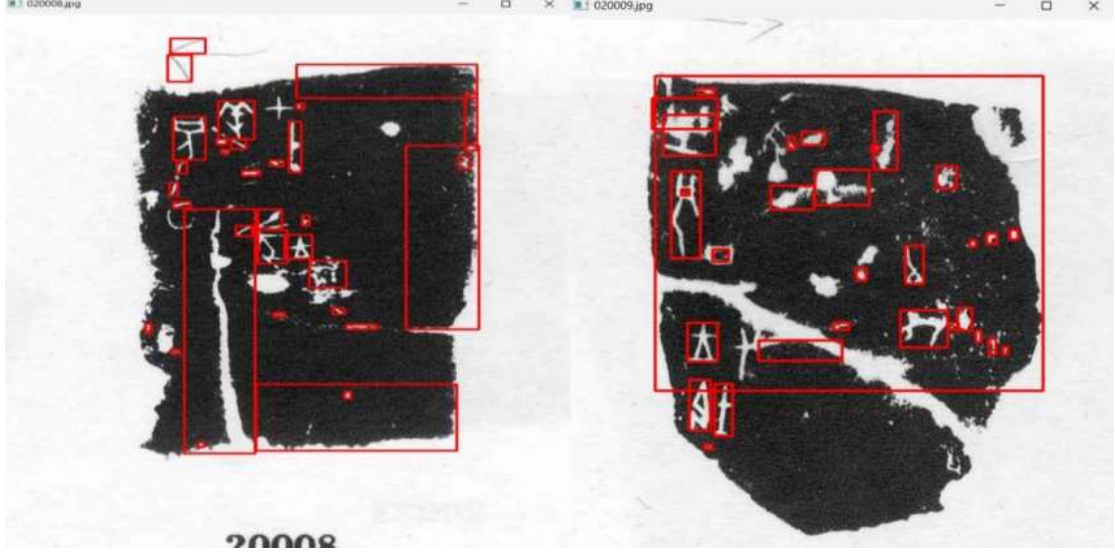


Figure 5: U-Net Oracle Bone Inscription Intelligent Segmentation Results Diagram

The cross-entropy loss function is used in machine learning to measure the difference between the predicted value of a

model and the actual value. The cross-entropy function is used to evaluate the model.

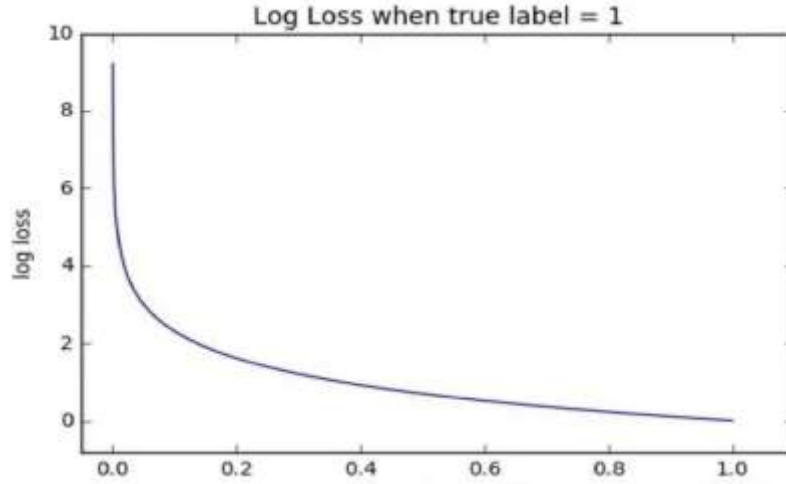


Figure 6: Cross-entropy loss function curve

F1-score and IoU are two key evaluation metrics, which measure model performance from different dimensions. As shown in Figure 6 and Figure 7. The F1 score integrates the precision (accuracy of predicting positive class pixels) and recall (coverage of true positive class pixels) of the model by harmonic mean, which is especially suitable for imbalanced scenes with very low proportion of objects (e.g., oracle script only accounts for less than 5% of the image). IoU focuses on the spatial positioning accuracy, and directly quantifies how

well the segmentation boundary fits the real shape of the target. This dual-index collaborative verification mechanism provides a comprehensive performance evaluation framework for tasks requiring high-precision segmentation. On the test set, the model achieves excellent performance of IoU89.7% and F1 score 91.2%, which verifies the high accuracy and strong robustness of the model in image segmentation tasks from the dual dimensions of classification and spatial positioning.

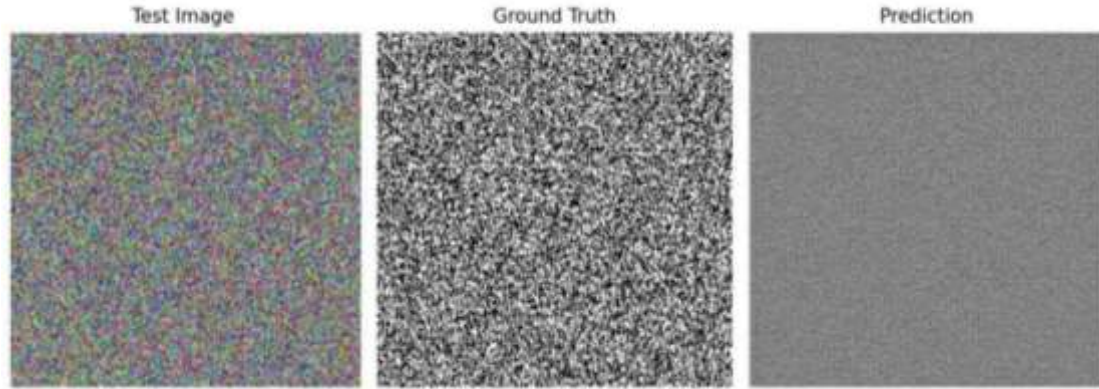


Figure 7: High-resolution oracle bone inscription segmentation prediction ground truth spatial comparison diagram

From a visual comparison perspective, the algorithm demonstrates the capability to roughly segment the oracle bone character regions. However, in terms of edge precision at the fine details of character boundaries, the prediction map exhibits inferior performance compared to the ground truth annotations. Subsequent improvements can be achieved through optimization of the edge loss function and incorporation of additional high-resolution feature extraction modules.

4. Conclusion

This paper provides a research idea and framework applied to computer vision related fields, through gray image binarization to preprocess the image, intelligent segmentation of oracle-bone inscriptions on the image based on U-Net realizes the segmentation of oracle-bone inscriptions on the image, and proves the feasibility of the method. There are still limitations in this study; the training data are mainly from a single source of oracle-bone rubbings, and the sample diversity and data scale are limited. For future research, we can construct a diversified dataset of oracle bone inscriptions across materials and generations, combine the generative adversarial network for data enhancement, integrate the edge detection technology and region growing technology,

optimize the texture preservation effect of low contrast regions, establish fusion linguistic knowledge, and realize the semantic verification of segmentation results through font database matching. These improvements will help to provide more accurate intelligent development for the study of ancient philology.

References

- [1] Lu Caiwu, Song Yiliang, Jiang Song, et al. Few-shot coal-rock interface image segmentation method based on improved U-net[J]. Metal Mine. 2024(1): 149-157.
- [2] Zhang Huan, Qiu Dawei, Feng Yibo, et al. Review on U-Net model improvements and their applications in medical image segmentation[J]. Laser & Optoelectronics Progress. 2022, 59(2): 55-71.
- [3] Zheng Jiahui, Yu Xiaodi, Zhao Shengmei, et al. Correlation imaging denoising based on mean filtering[J]. Acta Optica Sinica. 2022, 42(22): 41-48.
- [4] Du Xuwei, Chen Dong, Ma Zhaokun, et al. Improved image edge detection algorithm based on Canny operator[J]. Computer & Digital Engineering. 2022, 50(2): 410-413, 457.
- [5] Ding Yichuan, Geng Cong, Liu Jinxiong. Research on lane line edge detection method based on improved multi-directional Sobel operator[J]. Beijing Automotive Engineering. 2025(1): 23-29.