International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 1562-1575

https://doi.org/10.70517/ijhsa464139

Computer Vision Technology Explains the Representation System and Communication Mechanism of Visual Symbols in British and American Literature

Zhongbao He[1,*]
[1] School of Foreign Studies, Anhui Normal University, Wuhu, Anhui, 241002, China
Corresponding authors: (e-mail: hzbjason@163.com).

Abstract With the acceleration of globalization, cultural exchanges between China and the West have become more and more frequent, and English literature has been widely introduced into China. In this paper, the visual symbols in English literature are studied. In the pre-processing process of literary images, tilt correction, text region extraction and segmentation are carried out on the scanned images in turn. Based on DBNet algorithm, a text detection algorithm SMA for English literature is proposed, and Res2Net is used as a feature extraction network to extract multi-scale features to realize text extraction in literary images. Using the literary image preprocessing method and text extraction method proposed in this paper, we obtain the text of the novel "The Picture of Dorian Gray" by the British writer Oscar Wilde, and explore the visual symbols embedded in it. "Life", "soul", and "love" are the three keywords with the highest word frequency in the novel, which reach respectively 385, 366, and 340, implying the novel's thematic information and visual symbols to a greater extent. By analyzing the visual symbols of the three main keywords of "life", "soul", and "love", the novel "The Picture of Dorian Gray" is the crystallization of the fusion of Wilde's moral and artistic views.

Index Terms DBNet algorithm, Res2Net, text detection, image processing, English literature, "The Picture of Dorian Gray"

I. Introduction

Under the background of the strengthening trend of globalization, cross-border cultural interactions and exchanges have become more and more intensive, while literary works, as a core part of culture, have always received extensive attention in the process of multicultural dissemination and acceptance [1]-[4]. As a key part of global literature, British and American literary works not only play a decisive role in their respective cultural inheritance, but also have a wide and far-reaching influence on global culture [5]-[7]. The creation of Anglo-American literature is deeply influenced by cultural elements such as geographic positioning, local folklore and art forms, and these literary works not only reflect the social realities of the time, but also convey the emotions and values shared by all human beings [8]-[10]. British and American literature, through dramatic monologues and rich metaphorical symbolism, combines a variety of narrative techniques, showing its incomparable artistic appeal and deep cultural heritage [11], [12].

With the development of computer vision technology, English literature has realized a wider range of communication through the form of visual conformity [13]. Visual symbols are symbols that convey specific meanings or arouse emotional resonance through visual elements, these symbols can be graphics, images, shapes, colors, etc., which convey information in an abstract way, can transcend the limitations of space and language, and directly touch the audience's perception and feelings, and plays an important role in the dissemination of British and American literature, which can realize the communication of different levels of British and American literature with the audience, and is a medium for expressing deep meanings and emotions [14]-[17].

This paper takes English literature as the subject of research and takes its visual symbols and communication mechanism as the theme of this research. In order to obtain the text data of English literature, the preprocessing method of literary images is proposed. Firstly, tilt correction is performed on the literary images, and after completing two erosion operations on the binarized image, the target horizontal line is located and the affine transformation matrix is calculated to obtain the new coordinates of the corrected image. Segment the image and perform two erosion operations and expansion operations on the binarized image, retain the text region and index image portion, and statistically count the pixel values to achieve text region extraction. Further elimination of the binarized image, removing the foreground image, setting different segmentation thresholds to segment the image, and finally completing the preprocessing of literary images. The DBNet-based text detection algorithm SMA for English literature is proposed, and Res2Net is used as the feature extraction network to extract multi-scale features with finer granularity. The feature pyramid enhancement module is constructed, and a new bottom-up feature fusion path is added, so that the localization information at the lower level can be transmitted to the higher level more easily and enhance the features at the higher level. Introducing channel attention and spatial attention mechanisms, we adaptively fuse multi-scale features by learning the importance degree of features in channel, space and scale, so as to obtain a more robust fusion feature map and realize text detection and extraction. Simulation experiments of text detection are carried out to test the effectiveness of the text detection method proposed in this paper. The work “The Picture of Dorian Gray” by British writer Oscar Wilde is selected to carry out the visual symbol analysis of the whole work after realizing text detection and extraction. Finally, combined with the background of the continuous development of new media technology nowadays, the optimization strategy of the communication path of English literature is proposed.

II. Literary Image Preprocessing

With the strengthening of globalization and cultural exchanges, English literature has gradually gone global and attracted more and more readers. This paper combines computer vision technology to study the visual symbols in English literature and its dissemination mechanism, aiming to explore how to better convey the rich connotation and aesthetic ideas of the works, spread them to a wider range of people, and bring the enjoyment of beauty to readers around the globe, which also helps to deepen the cultural exchanges, and to promote the understanding and tolerance among different cultures. In this chapter, we will first preprocess the images of English literature to provide a data base for the study of visual symbols in English literature in this paper.

II. A. Picture tilt correction

The original paper English literature will inevitably be tilted in the electronic scanning process, due to the large size of the scanned image, the text area is dense and other factors, so in the process of layout analysis, the smaller tilt angle will lead to the occurrence of serial lines between adjacent text lines, affecting the layout of the subsequent data restoration, so it is necessary to tilt correction of the image.

Step 1, detect the target horizontal line.

1) Image binarization [18].

The print body image contains numerous elements within it, and it is necessary to accurately detect the target horizontal line in the image. Using the weighted average method, the color image is converted into a single-channel grayscale image, after which the grayscale image is binarized by setting the image background to white and the image foreground to black.

2) Morphological processing

Firstly, two corrosion operations are performed on the binarized image to corrode the array of text regions as a whole. Although the target horizontal line is connected by a number of short straight lines, after the image corrosion operation has been connected to a whole. Then the image of the two expansion operations, due to the text area between the columns and columns of the existence of large gaps, and the target cross-hairs no longer exist gap, so after the second image expansion operation, the image in addition to the target cross-hairs outside the elements are erased.

3) Locate the target horizontal line

By counting the pixels within the binary image, the start and end coordinates $(x_{start}, y_{start}), (x_{end}, y_{end})$ of the target horizontal line within the image can be localized, and the tilt angle α of the line segment is calculated as:

$$\alpha = \arctan \frac{y_{end} - y_{start}}{x_{end} - x_{start}} \tag{1}$$

where $\alpha \in (-\frac{\pi}{2}, \frac{\pi}{2})$.

In step 2, the affine transformation matrix Mat is computed, and a rotation operation is performed to correct the image.

After obtaining the tilt angle α, compute the affine transform matrix Mat:

$$Mat = \begin{bmatrix} \cos\alpha & -\sin\alpha & dx \\ \sin\alpha & \cos\alpha & dy \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

Select the image center point as the rotation center:

$$\begin{cases} dx = \frac{w}{2}(1-\cos\alpha) + \frac{h}{2}\sin\alpha \\ dy = \frac{h}{2}(1-\cos\alpha) + \frac{w}{2}\sin\alpha \end{cases} \tag{3}$$

After the original image is rotated, the dimensions of the new image change. Calculate the width new_w and height new_h of the image after rotation:

$$\begin{cases} new\_w = w\cos\alpha + h\sin\alpha \\ new\_h = w\sin\alpha + h\cos\alpha \end{cases} \tag{4}$$

At the same time, the image center point of the rotated image changes compared to the original image, which may result in the loss of some image information, so it is necessary to calculate the image center position offset ($\Delta x, \Delta y$) in order to perform a panning operation on the rotated image:

$$\begin{cases} \Delta x = \frac{new}{2} - \frac{w}{2} \\ \Delta y = \frac{new}{2} - \frac{h}{2} \end{cases} \tag{5}$$

The affine transformation matrix Mat' is recomputed using the image center position offset ($\Delta x, \Delta y$):

$$Mat' = \begin{bmatrix} \cos\alpha & -\sin\alpha & dx + \Delta x \\ \sin\alpha & \cos\alpha & dy + \Delta y \\ 0 & 0 & 1 \end{bmatrix} \tag{6}$$

Step 3, the points in the original image are affine transformed using the transformation matrix Mat' of Eq. (7) to find the coordinates after rotation as:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = Mat' \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha & dx + \Delta x \\ \sin\alpha & \cos\alpha & dy + \Delta y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{7}$$

where (u, v) denotes the new coordinates of the point (x, y) in the original image obtained after affine transformation. After the operation of step 1 a step 3, the tilt correction of the printed body image is completed.

II. B.Image text area extraction

In this section, we will design the text region extraction method for the pictures of English literary works from the characteristics of printed pictures, mainly using morphological processing methods.

In the image tilt correction session, the tilt angle of the picture is calculated by detecting the horizontal line in the table head part of the picture of literary works, and then the tilted image is corrected. In this section, we continue to use the horizontal line of the table head as the entry point, and detect and record the position of this horizontal line through the image tilt correction session, so as to separate the table head part of the picture from the following text region.

1) Image binarization

The image of the printed body after splitting the table head is converted into a single-channel grayscale image, after which the grayscale image is binarized by setting the background of the image to white and the foreground of the image to black.

2) Morphological processing

Firstly, two corrosion operations are performed on the binarized image to corrode the array of text regions as a whole. Then one image expansion operation is performed on the image to retain only the text region and index image part.

3) Counting pixel values

Since the index image is on the leftmost side of the processed printed image, the pixel values within the image after image expansion are counted vertically to locate the index position, and the index part of the image can be segmented.

II. C.Image Text Area Splitting

After extracting the text region of the printed body image, the width of the text region of the type I printed body image is about 4,000 pixels and the height is about 2,000 pixels; the width of the text region of the type II printed body image is about 4,000 pixels and the height is about 3,000 pixels, and in order to ensure the correctness of the subsequent text detection, it continues to perform image segmentation on the text region of the image of literary works.

Step 1, index image processing.

1) Image binarization

The segmented image of the index part of the printed body picture is converted into a gray-scale image, and the gray-scale image is binarized by setting the background of the image to white and the foreground of the image to black.

2) Morphological processing

Firstly, one corrosion operation is performed on the binarized image to corrode the index part of the label as a whole. Then one expansion operation is performed on the image to eliminate the foreground image outside the indexed part, and then a second image corrosion operation is performed, and the parts inside the image except the indexed part are connected as a whole according to the arrangement characteristics.

Step 2, segment the end of the image.

The index image and the text area image have the same height, and the index line within the index image and the text line within the text area image correspond one-to-one, so it is possible to segment the tail elements within the text area of the type I printed image by counting the pixel values, with the help of the position of the last index line within the index image, and there are no tail elements within the type II printed image, which are not processed.

Step 3, segment the text area.

After segmenting the figure-tail image, the pixel values within the binary map are counted, and the specific position of the index of each group can be obtained, for example, the position of each 5 groups of date-indexed text rows can be obtained after this step of segmentation for type I printed body images, and the position of type indexes for each 4 groups of text rows can be obtained after this step of segmentation for type II printed body images. Afterward, a suitable segmentation threshold can be set to segment the text area of the printed body picture into a specified number. In this section, in order to facilitate the subsequent layout processing and restore the layout structure, different segmentation thresholds are set for two types of printed images: the text area of type I printed image is divided into upper and lower parts, and the text area of type II printed image is divided into a part of every 8 groups of text lines.

III. Text Detection of English Literature Based on DBNet

In this chapter, DBNet network will be introduced, and for the problems of layer-by-layer expression of multi-scale features, lack of quasi-localization information of features at the upper level of the pyramid, and insufficient fusion of multi-scale features in DBNet, we propose the text detection algorithm SMA for English literature based on multi-scale feature enhancement and adaptive fusion, and we build up a text detection model for English literature, and extract visual symbols in literary works.

III. A. SMA Overall Network Architecture

The benchmark model DBNet in the text detection process, the backbone network ResNet characterizes the multi-scale features in a layer-by-layer manner, which is insufficient for image multi-scale feature extraction [19], [20]. The high-level features of the fusion network FPN lack the strong localization information of the lower layers. Multi-scale features are fused by simple up-sampling and splicing, resulting in poor segmentation model scale robustness. To address the above problems, a scene text detection (SMA) algorithm with multi-scale feature enhancement and adaptive fusion is proposed using DBNet as a benchmark.The SMA text detection model can be divided into three parts: backbone network, fusion network, and detection head.

Backbone: a multiscale backbone network Res2Net is introduced to extract finer-grained multiscale features from the input scene graph, and features $C_2 \sim C_5$ at different scales are extracted as inputs to the feature fusion network.

Neck: design the feature pyramid enhancement module (FPR), which takes the extracted feature maps of four scales as inputs, and enhances the high-level features by fusing the accurate localization information of the lower pyramid levels through the constructed bottom-up path; construct the attention feature fusion module (AFF), which takes the multi-scale feature maps output from the feature pyramid enhancement module as inputs, and uses the channel attention and spatial attention mechanisms to learn the weights of features on channel, space, and scale to obtain a robust cross-scale fusion feature map.

Head: The SMA model follows the detection head in DBNet, and the text detection result map is obtained by differentiable binarization post-processing.

III. B. Feature extraction network Res2Net

Feature Extraction Network for DBNet Algorithm ResNet, as a widely used backbone network in text detection algorithms, utilizes bottleneck blocks to represent image multi-scale features in a layer-by-layer fashion. This obtains a rich feature representation to a certain extent, however, the network itself has limited sensory field and the ability to extract image multi-scale features cannot fully meet the requirements of the model to detect multi-scale text instances.

To address the above problems, the SMA model introduces a multiscale backbone network, Res2Net, to enhance the learning and representation of image multiscale features. This network constructs intra-layer grouped class-residual connections within each bottleneck block of ResNet to characterize image multiscale features at a finer-grained level. This increases the receptive field of each layer of features and improves the model's ability to extract information about multiscale features.

If F_{in}, F_{out} are the input and output feature maps of the ResNet bottleneck block, respectively, and $Conv_1(\cdot), Conv_3(\cdot)$ are the standard convolution operations with convolutional kernel sizes of 1×1 and 3×3, respectively, then the ResNet bottleneck block can be given by Eq. (8) Representation:

$$\begin{cases} X = Conv_1(F_{in}) \\ Y = Conv_3(X) \\ F_{out} = F_{in} + Conv_1(Y) \end{cases} \tag{8}$$

Res2Net's residual block decouples the single 3 × 3 convolution operation in Eq. (8) at multiple scales, adapting it to a hierarchical class of residual connections. First, the X in Eq. (8) is chunked along the channel dimension. If:

$$x_i = \begin{pmatrix} x_{1,\frac{(i-1)c}{4}+1} & \cdots & x_{1,\frac{i}{4}c} \\ \vdots & \ddots & \vdots \\ x_{h,\frac{(i-1)c}{4}+1} & \cdots & x_{h,\frac{i}{4}c} \end{pmatrix} \in R^{h\times w\times \frac{c}{4}} \tag{9}$$

Then $X = (x_1, x_2, x_3, x_4) \in R^{h\times w\times c}$.

The output subfeature map y_i is obtained by a series of operations in Eq. (9). Then cascade y_1, y_2, y_3 and y_4 along the channel direction to obtain the convolutional output feature map Y' as shown in Eq. (10). Y' is added element by element with the input feature map F_{in} after 1×1 convolution operation to obtain the output feature map F'_{out} of the Res2Net residual block:

$$y_i = \begin{cases} x_i, i = 1; \\ Conv_3(x_i), i = 2; \\ Conv_3(x_i + y_{i-1}), 3 \le i \le 4 \end{cases} \tag{10}$$

$$Y' = Concat(y_1, y_2, y_3, y_4) \in R^{h\times w\times c} \tag{11}$$

From Eq. (10) and the formula for calculating the sensory field, it can be seen that the sensory field of the output sub-feature maps y_i increases with the increase of i , i.e., each group of output sub-feature maps possesses a larger sensory field than the previous group of output sub-feature maps. From Eq. (10) and Eq. (11), it can be seen that Y' has different numbers and combinations of receptive fields. Compared with Y , Y' has more number and variety of receptive fields. Therefore, Res2Net obtains features containing combinations of multiple sizes and numbers of receptive fields, which enables the SMA model to learn rich global and local features, enhances the model's ability to characterize multi-scale features, and improves the model's ability to detect text boxes of different scales.

III. C. Feature Pyramid Enhancement Module

Aiming at the problem of insufficient localization information of high-level features, the feature pyramid reinforcement (FPR) method is proposed, which constructs a bottom-up feature fusion path and enhances the high-level features of the pyramid by using the accurate localization information of the lower levels.

FPR uses depth-separable convolution to replace the standard convolution, so the computational consumption of FPR is lower than that of FPN.Depth-separable convolution can be decomposed into 3×3 depth convolution and 1×1 point-by-point convolution. Equation (12) and Equation (13) represent the computational consumption of depth separable convolution and normal convolution, respectively.The ratio of the computational consumption of FPR and FPN is shown in Equation (14). From Equation (14), the computational consumption of FPR is about 1/5 of FPN:

$$C_1 = C_{in} \times {D_K}^2 \times {D_H}^2 + C_{in} \times {D_H}^2 \times C_{out} \tag{12}$$

$$C_2 = C_{in} \times D_K^2 \times D_H^2 \times C_{out} \tag{13}$$

$$\frac{C_{FPR}}{C_{FPN}} = \frac{C_1 \times 6}{C_2 \times 4} = \frac{3}{2}(\frac{1}{C_{out}} + \frac{1}{D_K^2}) \tag{14}$$

where D_K is the convolutional kernel size, D_H is the height or width of the input feature map, and C_{in}, C_{out} is the number of channels of the input and output feature maps, respectively.

FPR adds a new bottom-up feature fusion path on the basis of the feature pyramid network, which enriches the localization information of the high-level features of the FPN by fusing the low-level features and shortens the information transfer path from the low level to the high level, so that the information of the low level is more easily transmitted to the high level of the pyramid. And thanks to the use of deep separable convolution with low computation, FPR improves the model detection performance while making the computational consumption of the model decrease.

III. D. Attention feature fusion module

Feature Pyramid Network FPN fuses cross-scale features by up-sampling plus cascading. This improves the scale robustness of the segmentation model to a certain extent, but ignores the semantic differences between text features at different scales, which affects the model's effectiveness in detecting multi-scale text. To address this problem, the Attention Feature Fusion (AFF) module is constructed. This module introduces channel attention and spatial attention mechanisms to adaptively learn the corresponding weights for multi-scale input feature maps with different sensory fields, in order to more robustly fuse cross-scale feature maps.

Let X_k, P be the input feature maps of AFF, and the feature maps obtained from the input cascade, respectively. The cascade feature map P is obtained by channel attention mechanism CA with summation operation to get the channel-weighted feature maps F_1, F_1 by spatial attention mechanism SA and summation operation to get the global feature maps F_2 as shown in Eqs. (15)-(17):

$$P = Concat(X_5, X_4, X_3, X_2) \tag{15}$$

$$F_1 = P \oplus CA(P) \tag{16}$$

$$F_2 = F_1 \oplus SA(F_1) \tag{17}$$

where $P, F_1, F_2 \in R^{n\times c\times h\times w}, X_k = (x_{ij}^k) \in R^{c\times h\times w}, k \in \{2,3,4,5\}$.

The global feature map F_2 is subjected to 1×1 convolution and Sigmoid activation operations to obtain the weight matrices W_k corresponding to the feature maps at different scales. After W_k is obtained from Y_k by dimension expansion and dot product operation, cascade Y_5, Y_4, Y_3 and Y_2 in order to get the output feature maps F of the AFF as shown in Equations (18)-(20):

$$W_k = \sigma(Conv_1(F_2)) = (w_{ij}^k) \tag{18}$$

$$Y_k = (y_{ij}^k) = (w_{ij}^k x_{ij}^k) = (w_{ij}^k) \square \ (x_{ij}^k), w_{ij}^k = \begin{pmatrix} w_{ij}^k \\ w_{ij}^k \\ \vdots \\ w_{ij}^k \end{pmatrix} \tag{19}$$

$$F = Concat(Y_5, Y_4, Y_3, Y_2) \tag{20}$$

Of these, $W_k \in R^{1\times h\times w}, (w_{ij}^k), Y_k \in R^{c\times h\times w}, F \in R^{n\times c\times h\times w}, k \in \{2,3,4,5\}$.

The AFF module improves the cross-scale feature fusion approach of FPN by adaptively fusing multi-scale feature maps by learning the importance of features on channel, space and scale. Among them, the channel attention module highlights the channels containing features related to text instances and suppresses the channels containing a lot of noise and background information by learning the weights of the input feature maps on different channels. And the spatial attention module emphasizes the feature information related to the target and suppresses the useless background, noise features by focusing on the features on different locations of the feature map. As a result, the AFF module learns the weight matrix containing a large amount of valid weight information, obtains a more robust multi-scale fusion feature map, and improves the detection performance of the SMA algorithm for multi-scale text boxes.

IV. Text Detection Simulation Experiment

In this chapter, the text detection performance of the proposed text detection model for English literature will be tested to analyze its utility in text detection work.

IV. A. Data sets

1) Icdar2015 dataset

Icdar2015 is a multidirectional dataset which consists of 1500 (1000 for the training set and 500 for the test set) plain English labeled natural scene images, labeled in the form of top left, top right, bottom right and bottom left in that order. The dataset is taken with Google glasses in unfocused scenario, mainly for shopping malls and street scenes, proposed by Incidental scene text competition 2015.Icdar2015 dataset is a standard dataset widely used in the field of text detection and recognition, and a lot of text detection and recognition algorithms are based on this dataset for evaluation and comparison.

2) MSRA-TD500 dataset

MSRA-TD500 is a benchmark dataset for text detection and recognition, produced by Microsoft Research Asia, containing 500 images of natural scenes that cover a variety of environments and conditions, such as different lighting, shadows, noise and blur. Each image contains one or more lines of English and Chinese text, totaling 2233 text instances. Following this method an additional 400 training images are included from the TR400 from Huazhong University of Science and Technology (HUST), a dataset jointly published by HUST and Microsoft Research Asia.

IV. B. Experimental setup

The experimental platform settings for model training and testing in this experiment are shown in Table 1. The experimental platform for the algorithms in this chapter is Ubuntu18.04 operating platform, server integrated with 6 NVIDIA RTX2080Ti graphics cards, Intel Xeon 6146 CPU, 12 cores and 24 threads.

Table 1: Table Experiment platform configuration parameters

Tools	Name	Version number
GPU	NVIDIA RTX2080Ti	-
Operating system	Linux Ubuntu	18.04
Deep learning framework	Pytorch	1.4.0
Operation platform	CUDA	10.2

Because part of the public dataset training set is only about one thousand images, for the enhancement of the dataset is rotated, cropped, and flipped in the range of (-10°,10°). In this paper, the network is based on PyTorch framework, following the multiple learning rate strategy, the input image size is 640*640, the iterative learning rate is set to lr=0.0007, the batch size is set to 16, and the loss function image is specifically shown in Figure 1. It can be seen that after 1200 rounds of model weight adjustment, the final convergence is around 0.4.

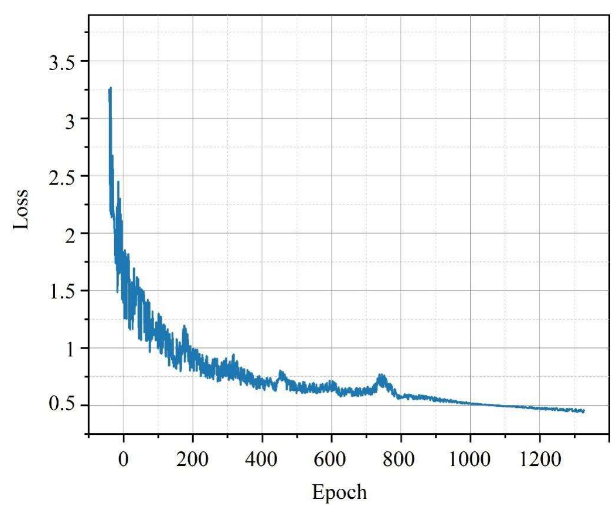

Figure 1: Loss function image

IV. C. Comparative analysis of model parameters

The performance parameters of the text detection model proposed in this paper are compared with those of the DBnet model, as shown in Table 2. The weight size of the model designed in this paper is 23.9MB, which is 6 times compressed compared to the original DBnet model, effectively reducing the storage cost. The model in this paper more than doubles the detection speed, significantly shortens the inference time, and is suitable for deployment in mobile device platforms. DSConv has a smaller number of parameters while ensuring that the entire network maintains a high detection accuracy. Overall, the improved overall model outperforms the DBnet network in terms of memory size and computing speed.

Table 2: Table Comparison of model performance parameters

Method	Weight/MB	FPS
DBnet	136	25
Ours	24.1	54

IV. D. Analysis of the results of comparative experiments

For the validity of the model in this paper, the model performance is discussed on the Icadr2015 dataset. On the Icadr2015 dataset, the performance detection results of different modeling methods are specifically shown in Table 3. Where P represents the accuracy rate, R represents the recall rate,F is the F-value to assess the text detection ability. Representative algorithm Corner proposes a new scene text detection algorithm that combines the ideas of target detection and semantic segmentation, one branch is used for edge point detection to extract the text region, and the second branch divides the sensitive locations by segmentation. The two-branch structure allows the model to observe the text details, and it achieves 94.2% accuracy on this dataset which is far better than other algorithms, but it performs poorly in recall, the reason for analyzing this is that the ambiguous text is judged as negative samples, which causes ambiguity to the model and leads to a decrease in the recall rate. The PSE model uses a progressive expansion algorithm to generate multiple subgraphs of the original segmentation map for prediction, which increases the computational volume of the model and lacks performance in FPS. The method in this paper has a substantial improvement in each evaluation index, P, R, F are 91.2%, 81.7%, 86.2%, respectively. From the experimental results, it is clear that this paper's method has better performance improvement on the noisy and complex Icdar2015 dataset. Comparing with other models, this paper's model obtains better experimental results on both public datasets, which proves that this method is able to detect more complete text under complex background.

Table 3: Icdar2015 data set detection and comparison results

Model	P	R	F	FPS
CTPN	74.4	51	61.9	7.1
EAST	82.7	74	77.9	13.2
SSTD	79.3	74.7	76	7.7
WordSup	80.1	76.4	78	-
Corner	94.2	71.7	79.7	3.6
TB	87.1	76.5	82.3	11.6
RRD	85.5	79.1	83.1	6.5
MCN	71.3	80.7	75.7	-
TextSnake	84.2	79.9	81.7	1.1
PSE	88.5	80	83.9	1.6
DB-ResNet18	87.3	77	83.3	48
DB-ResNet50	88.8	79.9	84.3	26
Ours	91.2	81.7	86.2	52

The TD500 dataset is dominated by multi-language and long text, which is affected by the imbalance of positive and negative samples, and the model is not effective in detecting multi-language. The methods involved in the comparison experiments include Corner, CRAFT, DBnet and many other classical text detection models, as shown in Table 4. In the TD500 experimental results, this paper's algorithm lacks in text recall compared to SAE, but there is a significant improvement in accuracy and F-value, and the algorithm in this chapter has better performance in a comprehensive view. The algorithm in this chapter achieves 92.8% accuracy and 81.5% recall, both of which are at a high level, proving that this method is significantly improved compared to other algorithms, implying that the algorithm in this chapter has a better detection effect when dealing with multilingual contexts such as English literature.

Table 4: TD500 data set detection results

Model	P	R	F	FPS
DeepReg	77.7	70	73.3	1.2
RRPN	81.6	67.2	74.3	-
RRD	86.7	72.3	79.6	10
MCN	88.5	78	84	-
PixelLink	82	72.8	77.5	3
Corner	87.9	75.6	80.9	5.8
TextSnake	82.4	74.5	78.9	1.2
CRAFT	89.2	77.5	82.4	8.5
SAE	83.7	82.1	83.6	-
DB-18	90.4	76.2	82.3	61
DB-50	92.1	79.1	85	30
Ours	92.8	81.5	87.2	66

V. Analysis of Visual Symbols in “The Picture of Dorian Gray”

Oscar Wilde was an advocate, poet and dramatist of the Aestheticism movement in the late 19th century in England. This chapter will take Wilde's literary work “The Picture of Dorian Gray” as the research object, and after completing the preprocessing of the work's images, apply the DBNet-based text detection model proposed in this paper to detect and extract the keywords in the novel of “The Picture of Dorian Gray”, and to design the visual symbols within the work to be analyzed.

V. A. Co-word analysis

V. A. 1) Co-occurrence matrix

Statistics on the keyword word frequency of “The Picture of Dorian Gray” is shown in Table 5. “Life”, “soul”, “love”, and “portrait” are the four keywords with the highest word frequency in the novel, and the word frequency reaches 385, 366, 340, and 123 respectively, which in a large extent implies the thematic information of this novel. Moreover, Wilde's language uses more expressive nouns, verbs and adjectives, which enhances the expressive effect of the language and covers a relatively wide range of themes, including “good”, “evil”, “beautiful” and “ugly”. However, in the nineteenth century, when Wilde lived under Victorian rule, the content of Wilde's novels could be said to be out of step with the times.

Table 5: Keyword

Keyword	Word frequency	Keyword	Word frequency
Life	385	Wonder	91
Soul	366	Romance	86
Love	340	Sins	72
Portrait	123	Had	67
Answered	118	Seemed	60
Beauty	115	Dreadful	56
Secret	108	Curious	44
Cried	106	Passed	30

Based on the contextual co-occurrence relationship, a co-occurrence matrix is generated, in which the data in the upper or lower triangular cells are the number of times two keywords appear in the same chapter at the same time, as shown in Table 6. The more times two keywords co-occur in the co-occurrence matrix, the closer the connection between these two keywords. Take the two antonyms of “love” and “sin” as an example, the number of co-occurrences of the two reaches 42 times, and combined with the word frequency of the two keywords in the above table are 340 and 72, respectively, the frequency of the appearance of “love” is much higher than that of the appearance of “sin”. That is to say, the theme of love occupies a greater proportion in this novel than evil, which seems to be contrary to the view held by early critics. It can be said that influenced by the traditional moral concepts in Victorian times, the traditional moral concepts are hidden in Wilde's subconscious mind, and in the artistic creation, Wilde reveals the morality unconsciously as the paint and visual symbols of artistic creation.

Table 6: Keyword co-occurrence matrix (part)

-	Life	Soul	Portrait	Love	Sins	Beauty	Secret	Romance	Dreadful
Life	0	174	139	142	44	49	39	56	5
Soul	174	0	185	96	42	21	55	69	16
Portrait	139	185	0	73	31	43	52	20	6
Love	142	96	73	0	42	6	16	1	26
Sin	44	42	31	42	0	20	33	23	13
Beauty	49	21	43	6	20	0	25	49	4
Secret	39	55	52	16	33	25	0	10	9
Romance	56	69	20	1	23	49	10	0	3
Dreadful	5	16	6	26	13	4	9	3	0

V. A. 2) Cluster analysis

R-type clustering is a variable based clustering method and Q-type clustering is a clustering method based on similarity or distance measure. R-type clustering does not consider the category of the data points and clusters them only by the similarity between the variables. Q-type clustering is usually applied to classification or grouping problems. In R-type clustering, different statistical methods are usually used to measure the similarity between the variables, such as correlation coefficient, chi-square test etc. In this paper, R-type clustering is used and the resulting co-occurrence matrix is imported into SPSS software. The systematic clustering method is used, and the clustering method selects the intergroup linkage, and the squared Euclidean distance is chosen. The degree of association between each keyword can be initially determined. The clustering results are shown in Figure 2. Domain one is verbs such as “cried”, “answered”, “passed” and adjectives such as “dreadful”, “wonderful”, which are the more expressive and graphic keywords in “The Picture of Dorian Gray”. Domain two is nouns and topic words such as “life”, “soul”, and “love”, which are relatively close to the central idea and point of view conveyed in the novel.

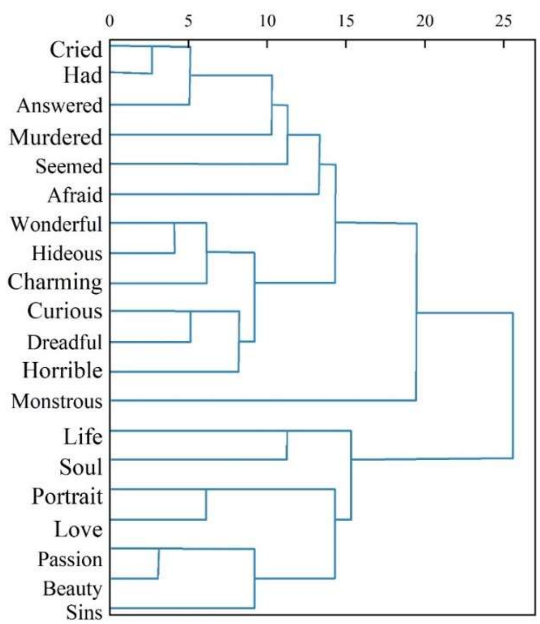

Figure 2: clustering results

V. B. Visual Symbol Analysis of Main Keywords

Among the keyword word frequencies counted in this paper, the three keywords with the highest word frequencies are “life”, “soul”, and “love”, and only three keywords account for 50.35% of the word frequencies of all the keywords, which largely imply the theme information and the main visual symbols of the novel “The Picture of Dorian Gray”. In this section, we will start from these three keywords to explore the main visual symbols of the novel.

1) “Life”

“Life” appears 385 times in the whole novel of The Picture of Dorian Gray. Statistics of adjectival possessive pronouns and qualifiers located in the left one of “life” are shown in Table 7. Third person singular/plural possessive pronouns such as “his”, “her”, “their” and -'s appear frequently with a frequency of 55 times, indicating that the topic talked about is not the speaker's own. The articles “the” and “a” appear relatively less frequently in the left position of life, with only 8 occurrences each, suggesting that the content of the talk is not general or specific. “Our” never appears in the left position of “life”, suggesting that the content of the talk is not shared by the speaker and the listener either, but the speaker's frequency of talking about himself is higher, as can be seen from the co-occurrence frequency of “my” and “life”. This shows that the speaker is more self-centered and more concerned with his own interests than those of others. In “The Picture of Dorian Gray”, at the beginning of the story, “life” tends to be paired with words with more positive meanings, and as the story progresses, the words paired with “life” become more and more negative and darker, which indicates a change in the character of the main character. Moreover, in order to emphasize the crimes committed by the main character, the verbs paired with “life” are also more destructive, such as “exhausted”, “wrecked”, “destroyed”, and so on.

Table 7: Adjective subject pronouns and determiners on the left side of life

Pronoun	Frequency
His,her,their,possessive 's	55
My	35
Our	0
The	8
A	8

2) “Soul”

The word frequency of the keyword “soul” is second only to “life”, reaching 366. Statistics on the predicate collocation of “soul” in “The Picture of Dorian Gray” are shown in Table 8. From the table, we can see that “soul” is usually collocated with some verbs and variants, such as “give”/ “given” and “cure” in the table, which can be seen to show that the main character of the novel has an inner struggle and refuses to accept his own soul. And the verbs counted in the table are more away from convergent verbs, indicating that the protagonist begins to loathe his own actions and feel shameful, the protagonist's inner struggle reveals that Wilde has a benchmark for measuring morality in his subconscious mind, which constantly corrects the faults of the protagonist, which has a certain degree of guidance to the subsequent plot development of the novel, and plays an important role in the overall expression of the novel's visual symbols.

Table 8: The predicate of soul

The predicate of soul	Frequency
give /given	10
cure	8
see	8
hide /hidden	4

3) “Love”

The keyword discussed in this section is “love”, which belongs to the good side of human nature and has a word frequency of 340 in the novel. Out of the 340 occurrences, “love” occurs 221 times as a noun, and out of the remaining 119 occurrences in the form of a verb, the occurrences of the object of “love” are as shown in Table 9. The object HIM occurs more frequently than HER and YOU occurs more frequently than ME, indicating an imbalance in the scales of “love”, which leads to subsequent tragedy. In the first half of “The Picture of Dorian Gray”, LOVE appears more frequently, especially when the main characters, Dorian and Sibel, meet and fall in love. In the second half of the novel, the frequency of “love” decreases sharply, which on the one hand indicates the disappearance of Dorian's love for Sibel, and on the other hand indicates the transformation of Dorian's character from an innocent and simple young man to a cold and heartless hedonist. The word “love” is an important constituent of the plot development of the novel and the visual symbols.

Table 9: The object of love

The object of love	Frequency
Him	43
Her	28
You	21
Me	15
Us	12

By analyzing the three main keywords of LIFE, LOVE, and SOUL in “The Picture of Dorian Gray”, a more detailed thematic information and visual symbols analysis of the whole novel is carried out. The study shows that morality is used by Wilde as the raw material for literary creation, and is the most central visual symbol in “The Picture of Dorian Gray”. On the one hand, he detests the hypocrisy of the upper class and the hypocrisy of Victorian morality, and hopes that beauty can be independent of utilitarianism, morality and propriety, which consciously shapes his own viewpoint of aesthetics, and on the other hand, it is inevitably affected by the traditional social norms and literary features of the Victorian period, subconsciously revealing the morality of his artistic creations. “The Portrait of Dorian Gray” is a fusion of his moral and artistic views.

VI. Suggestions for Optimizing the Dissemination Mechanism and Dissemination Path of English Literature

At the beginning of the reform and opening up, British and American literature flowed into China in the form of translation and re-publication, but with the continuous progress of new media technology, the dissemination of British and American literature has taken more forms.

VI. A. Mechanisms for the dissemination of English literature

The popularization of smart phones and computers has made people's access to information diversified, and the dissemination mechanism of English literature, as another way of information expression, has also changed greatly. Traditionally, English literature is disseminated through paper carriers, which has great disadvantages. Nowadays, the new media technology based on the Internet, cell phones and computers as the carrier for dissemination has great advantages compared with the traditional communication path, so that people from all walks of life can learn about the literature of each period.

The influence of the new media technology era on the dissemination of English literature is mainly in the following two aspects: firstly, the change of reading carrier; secondly, the change of the scope of readers. Nowadays, the popularization of computers and smart phones has changed people's way of life and reading.

VI. B. Optimization of English Literature Dissemination Paths

1) Carry out a very targeted website information dissemination.

Network information is characterized by relevance, efficiency and rapidity. The extensive use of network information makes the dissemination of literature more popular. At the same time, network information can provide inexhaustible literary materials for authors. In order to promote the dissemination of English literature, we can build a theme website of it, integrate the latest English literature and related news and information, so that more and more audiences can fully understand the charm of English literature.

2) Targeted development of APP for dissemination.

Targeted development of APP for dissemination of English literature can increase people's access to it, thus stimulating more people to actively understand and read English literature. At the same time, for the popular English literature, it is also possible to set up discussion topics in the APP, so that readers can communicate and share with readers from all over the world, and even talk with the authors directly.

3) Dissemination from digital resource platforms with a huge library of materials.

Developing a huge digital resource platform for the literary works of different countries, so that the digital resource platform includes the literary works of the whole world. Readers are not only able to make reading inquiries on the digital resource platform, but also can select English literature according to their own preferences. The digital resource platform includes English literary works of every period, so that the reading needs of readers can be satisfied to the greatest extent.

4) Multimedia communication is accepted by the public.

Sound, text and pictures are integrated into one, making English literature audiobooks, drama graphics, etc., providing readers with visual and auditory ways to understand the works, expanding the readership while also providing convenience for the blind audience to a certain extent.

VII. Conclusion

The visual symbol analysis of English literature needs to be built on the text data of literature. In order to efficiently obtain the text data of English literature, this paper proposes an image preprocessing method and a text detection model based on DBNet, which provide methods and approaches for obtaining the text data of English literature. The weight size of the text detection model proposed in this paper is only 23.9MB, which is 6 times more compressed than the traditional DBnet original model and can effectively reduce the storage cost. On the noisy and complex Icadr2015 dataset, the model in this paper has a substantial improvement in the evaluation indexes, P, R, F reached 91.2%, 81.7%, 86.2%, respectively, which are better than other comparative models and obtain better experimental results. In the TD500 dataset, which is mainly multilingual and long text, 92.8% accuracy and 81.5% recall are achieved, and both of them reach a high level, which is an obvious progress compared with other algorithms and models. The DBNet-based text detection method proposed in this paper is able to maintain excellent text detection performance in complex backgrounds and multilingual contexts.

The novel “The Picture of Dorian Gray” is chosen as the research object, and the visual symbol analysis of it is carried out after applying the image preprocessing method and text detection method of this paper to obtain the text data needed for the research. The four keywords with the highest word frequency in the novel are “life”, “soul”, “love”, and “portrait”, and the word frequency reaches 385, 366, 340, and 123, respectively. The keywords are analyzed by clustering, and the clustering results are plotted to get the domains I and II. Domain one covers verbs and adjectives, which are expressive and graphic. Domain 2 is nouns and words, which are relatively close to the central idea and point of view of the novel. The three keywords “life”, “soul”, and “love”, which account for 50.35% of the word frequency in the novel, are further analyzed to explore the visual symbols conveyed in it. The keyword “life” is frequently used with third-person singular/plural possessive pronouns such as “his”, “her”, “their” and -'s, with a frequency of 55 times, indicating that the speaker is more self-centered. The keyword “soul” is usually paired with some verbs and variants, and the verbs are mostly far away from convergent verbs, indicating the inner struggle of the main character and revealing the role of Wilde's inner morality as a plot guide in the process of creation. As the object of the keyword “love”, “him” appears more frequently than “her” and “you” appears more frequently than “me”, symbolizing the tragedy caused by the imbalance of love in the novel. In the second half of the novel, the frequency of the keyword “love” decreases sharply, which is a sign of the transformation of the character of the protagonist, Dorian. On the whole, morality, as the raw material of Wilde's creation, is the most central visual symbol of his novel “The Picture of Dorian Gray”, and the novel as a whole is the crystallization of the fusion of his moral and artistic views.

Finally, by exploring the communication mechanism of English literature in the rapid development of new media technology, we propose the optimization strategies of English and American literature communication paths such as disseminating information on thematic websites, developing APPs, constructing digital resource platforms and multimedia dissemination, which provide certain methodological reference for the development of the dissemination of English literature.

Funding

Humanities and Social Science Research Project of Higher Education Institutions funded by Anhui Provincial Department of Education: “A Study of Imperial Consciousness in the Works of Samuel Johnson” (SK2020A0074).

References

[1] Kholodniak, O. (2023). Comparative literary theory: A cross-cultural approach to English literature. Research Journal in Advanced Humanities, 4(4), 255-273.
[2] OPREA, L. G. (2019). THE ROLE OF THE LITERARY TRANSLATOR IN ENHANCING CROSS CULTURAL COMMUNICATION. NEW PERSPECTIVES IN LITERATURE, LINGUISTICS AND ARTS, 1, 207.
[3] Zhang, Y. (2022, July). Cross-cultural literary comprehension: Theoretical basis and empirical research. In Interkulturelles Forum der deutsch-chinesischen Kommunikation (Vol. 2, No. 1, pp. 58-73). De Gruyter.
[4] Wang, W. (2020). Analysis on the significance of foreign literature translation to cross-cultural communication. Overview of Education Theory (English), 3(4), 98.
[5] Spitzer, L. (2015). Essays on English and American literature. Princeton University Press.
[6] Lawrence, D. H. (2019). Studies in classic American literature. Rosetta Books.
[7] Tadjibayev, M. S. (2020). WRITERS AND POETIC IDEAS OF AMERICAN REALISM IN XIX CENTURE LITERATURE. Mental Enlightenment Scientific-Methodological Journal, 08-17.
[8] Peck, J., & Coyle, M. (2017). A brief history of English literature. Bloomsbury Publishing.
[9] Blankenship, R. (2024). American literature as an expression of the national mind. Taylor & Francis.
[10] Blamires, H. (2020). A short history of English literature. Routledge.
[11] Swinfen, A. (2019). In defence of fantasy: A study of the genre in English and American literature since 1945. Routledge.
[12] Smith, R. G. (2015). Affect and American literature in the age of neoliberalism. Cambridge University Press.
[13] Mühlenbeck, C., & Jacobsen, T. (2020). On the origin of visual symbols. Journal of comparative psychology, 134(4), 435.
[14] Muayqil, T., Davies-Thompson, J., & Barton, J. J. (2015). Representation of visual symbols in the visual word processing network. Neuropsychologia, 69, 232-241.
[15] Mavroudi, E., Haro, B. B., & Vidal, R. (2020, August). Representation learning on visual-symbolic graphs for video understanding. In European Conference on Computer Vision (pp. 71-90). Cham: Springer International Publishing.
[16] Liu, W. (2022). Regional Cultural Symbols in Visual Communication Design. International Journal of Frontiers in Sociology, 4(5), 75-80.
[17] Barnard, N. (2019). The Shape of Words and the Voice of Visual Form: The Symbolism of Poetry and Painting. South Central Review, 36(1), 19-40.

[18] R. Cappaert,W. Yang,D.J. Ross,C. Johnston,C. MacLeod & C.A. White. (2024). Development of an image binarization software tool for net occlusion estimations. Aquacultural Engineering,107,102466-102466.

[19] Quanhong Ma,Shaohua Jin,Gang Bian,Yang Cui & Guoqing Liu. (2025). DBnet: A Lightweight Dual-Backbone Target Detection Model Based on Side-Scan Sonar Images. Journal of Marine Science and Engineering,13(1),155-155.

[20] Shaozhang Niu,Xiangxiang Li,Maosen Wang & Yueying Li. (2020). A Modified Method for Scene Text Detection by ResNet. Computers, Materials & Continua,65(3),2233-2245.