Research Article

Computational Application on English Translation System Based on Intelligent Image Text Recognition

Wang Haijie and Zhang Li

Western Languages Department, Hebei Institute of International Business and Economics, Hebei, Qinhuangdao, China

Correspondence should be addressed to Zhang Li; spoonwhj@hbiibe.edu.cn

Received 31 March 2022; Revised 9 April 2022; Accepted 18 April 2022; Published 2 May 2022

Academic Editor: Muhammad Arif

Copyright © 2022 Wang Haijie and Zhang Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This work offers an enhanced intelligent picture text recognition algorithm based on the intelligent image text recognition method to increase the impact of English image text translation. Texture blocks of adaptable size are used to successfully increase the accuracy and efficiency of restoration due to the varying texture information present in various photos. Furthermore, the repair sequence is altered as a result of the improved priority calculation algorithm, and the weight of the structural information is enhanced at the same time. In addition, in order to reduce the overall structural complexity and calculation amount of the system, the gated loop unit is also selected as the RNN structure. Finally, this article constructs an English translation system based on intelligent image text recognition according to the requirements of intelligent image text recognition and designs experiments to evaluate the performance of the system constructed in this article. The experimental statistical results show that the English translation system constructed in this article can basically meet the needs of English image text recognition.

1. Introduction

With the rapid development of human society and economy, people from different regions need to communicate more and more frequently, but language barriers are still obstacles to human communication. In the early days, in order to solve communication barriers, manual translation was mainly used. However, manual translation is not only costly but also slower and gradually cannot meet people’s communication needs. Therefore, with the rise of computers, machine translation has attracted more and more attention, which aims to achieve natural interaction between two languages through computers. Machine translation systems have been separated into two categories in recent decades: statistics-based machine translation systems and neural network-based machine translation systems. The statistical machine translation system uses a probability model to convert one language into another. Statistical machine translation, neural machine translation uses an end-to-end model architecture and vectorizes the sentence to extract the semantic features and information of the sentence, resulting in a smoother translated translation that is more in line with the grammatical characteristics of the language and higher translation quality. Therefore, the neural machine translation model has become the mainstream translation system [1].

With the rapid development and renewal of technology, existing research on text translation may not be applicable to the latest translation systems. Therefore, we take image text translation as the theme and do research on the latest translation system. Moreover, we combine the image text information to generate a more suitable translation so that the translation can maintain the consistency of the translation style and the accuracy of the translation in the entire document or semantic environment [2].

Based on intelligent image text recognition technology, this article studies an English translation system that can be used for image text recognition. This system can be used for various English recognition in daily life and provides a
theoretical reference for subsequent research on translation systems.

2. Related Work

For low-resource neural machine translation, transfer learning provides an effective method to alleviate the lack of data [3]. Under the premise that the structure does not change and the parameters do not increase, it uses the shared word vector space to find common features between the two domain corpus. After observing the vector distribution, translation, and model training parameters, it is observed that transfer learning is still not effective for some niche or very low-resource translation tasks. The main reason is that there is no similar source domain vector distribution. That is to say, to a certain extent, transfer learning still has relevance requirements for the transfer and the transferred domain knowledge [4]. This requirement is not obvious in the languages of the same language system because the common characteristics will be learned through a large number of source domains. However, in the transfer of cross-language or unrelated domain knowledge, transfer learning has become very obvious for the distance and feature similarity requirements of the vector distribution [5].

Literature [6] proposed applying the encoder-decoder structure to the translation system. This method uses CNN to encode the input sentence into a continuous vector and then decodes the state vector into the target language through RNN. This can be seen as the beginning of NMT, but the process of using CNN to map sentences to vectors has the problem of word order information loss. Literature [7] proposed to apply long- and short-term memory networks to translation systems. LSTM is a variant of RNN, which can alleviate the problem of gradient disappearance of RNN in long-distance transmission. The literature [8] suggested an attention approach to produce a variable length hidden layer vector. The attention mechanism was initially developed to handle classification difficulties in the picture field. This approach converts the input sentence to a vector sequence, removing the need for the encoding end to convert all source sentences to fixed-length vectors. At the same time, the decoding end uses the attention mechanism to adaptively select the source information for decoding, so this method effectively solves the problem of long sentence translation and greatly improves the translation performance of NMT. Literature [9] proposed a local attention mechanism and a global attention mechanism, and literature [10] proposed a positive and negative attention mechanism. Literature [11] proposed a machine translation model based entirely on convolutional neural networks. This method uses a multilayer convolutional neural network on both the encoding side and the decoding side, uses gated linear units to control the gradient transfer, and uses an attention mechanism in each layer of the decoder. The model proposed in literature [12] only uses the attention mechanism to construct the encoder and decoder so that the model can handle the information transmission of long-distance sentences well. The experimental results show that this method not only has good parallelism and reduces training time but also greatly improves the translation performance of the model, so it was later widely used in other fields of natural language processing. Literature [13] used reverse translation technology to reverse translation of monolingual target data to form corresponding parallel data to expand the internal domain corpus. Literature [14] proposed two methods to make full use of source-side monolingual data. The first method uses a self-learning algorithm to generate a large-scale synthetic parallel corpus for NMT training. The second method uses a self-learning algorithm to enhance the encoder so that two NMT systems can simultaneously predict translation and reorder the source-side monolingual sentences. Literature [15] proposed to expand the training corpus by calculating the vector similarity to select sentences with a strong correlation with the internal domain from the external domain corpus. Literature [16] used a dynamic data selection method; that is, during the training process of the system, the dataset of each round of training is adjusted to be different. The method of model enhancement mainly changes the structure and training steps of the translation system to obtain a high-quality domain translation model. Literature [17] proposed a fine-tuning training method. This method first trains the translation system with the foreign domain parallel corpus and then uses the inner domain training corpus to fine-tune the translation model obtained in the previous step. Literature [18] proposed a weight training method. The method mixes the parallel corpus of the two domains to jointly train the translation system, changes the proportion of internal and external domain data in each round of the iterative training process, and increases the loss proportion of the internal domain instances when calculating the loss function. Literature [19] used a general encoder to encode sentences in the inner domain and outer domain and added a discriminator to the encoder to ensure that only domain-invariant information is transmitted to the decoder. In each area, literature [20] presented domain-specific encoders, domain-specific classifiers, and generic encoders. Furthermore, it adjusts the weight of the target word in the training target based on the attention weight produced by the domain classifier so that domain-related words have a higher effect on the model training process.

3. English Text Image Recognition Algorithm Based on Texture Information

An English text image restoration method based on texture information uses the texture features contained in the English text image to select local blocks with similar texture information for splicing and finally complete the restoration of the entire English text image. The specific idea of the algorithm is as follows: first, select the appropriate texture block size claw according to the texture information of the English text image, and then calculate a rectangle with the point as the center and the claw as the side length for each pixel on the boundary of the area to be repaired block’s priority, the size of the priority determines the repair order of the block to be repaired, and then find the texture block closest to the texture of the block to be repaired in the undamaged area of the English text image, and finally fill the
block to be repaired with the information in the texture block and unknown pixel information in the block.

The Criminisi algorithm is the most typical of the texture-based synthesis algorithms. The fundamental concept is to calculate the priority of the border pixels to establish the repair order and then assign greater priority to repairing the target block and filling the block to be fixed using an appropriate matching approach. It can be seen that the key steps of the Criminisi algorithm are as follows: the determination of the block to be repaired, the determination of the search strategy, and the matching strategy. The algorithm model is shown in Figure 1 [21].

Among them, I represents the entire English text image, P represents the area to be repaired, ∂Ω represents the boundary of the area to be repaired in the English text image, Φ represents the undefective area of the English text image, point p is the target pixel point, n_p is the normal vector of point p, and ∇I_p represents the tangent direction of the isoline direction. The algorithm includes the following three steps.

3.1. Calculate the Block to Be Repaired with the Highest Priority. The calculation of priority determines the repair order of subsequent pixel blocks, which is the key to the entire algorithm. In order to meet human visual requirements, the texture and structure of the English text image must be consistent during restoration. Therefore, these two aspects should also be considered when calculating the priority.

For any pixel p on the boundary to be repaired, the corresponding formula for its priority P(p) is as follows:

\[ P(p) = C(p) \times D(p). \]  

Among them, C(p) represents the confidence item, which reflects the amount of information available in the neighborhood of point p. The calculation formula is as follows:

\[ C(p) = \frac{\sum_{q \notin P \cap (I_p - \Omega)} C(q)}{|\psi|}. \]  

In the above formula, \( \psi_p \) represents the area of the texture block to be repaired, and the initial value of C(p) is calculated according to the following formula:

\[ C(p) = \begin{cases} 
0, & \forall p \in \Omega, \\
1, & \forall p \in \partial \Omega. 
\end{cases} \]  

C(p) represents the credibility of the pixel block in the neighborhood of the pixel point p, which is mainly used to measure the amount of information available in the neighborhood of the point. The higher the confidence is, the more useful the information provided is when searching for the best matching block, and the earlier it will be repaired. The introduction of C(p) ensures that the area containing more known pixels in its neighborhood is preferentially repaired. In areas with relatively similar structural information, the priority calculation mainly depends on C(p), so the repair order is transformed into a repair order that gradually shrinks from the outside to the inside. D(p) represents the data item, and the calculation formula is as follows [22]:

\[ D(p) = \frac{\|\nabla I_p \times n_p\|}{\alpha}. \]  

Among them, \( \alpha \) is the normalization coefficient. For grayscale English text images, its value is usually 255. \( \nabla I_p \) represents the tangent direction of the isoline direction at the pixel point, which is expressed by the following formula:

\[ \nabla I_p = \frac{(-I_y, I_x)}{\sqrt{I_x^2 + I_y^2}}. \]  

Among them, \( I_x \) and \( I_y \) are the partial differentials of pixel p in the x and y directions, respectively. D(p) reflects the strength of the structural information of the pixel p. When the angle between the isoline direction of point p and the normal vector at the boundary of point p is smaller, the corresponding data item is larger, indicating that the part with strong linear structure will be repaired first. When the number of known information contained in the pixel block is roughly the same, the calculation of the priority depends on the data item D(p); that is, the direction with strong structural information is preferentially repaired.

3.2. Search for the Best Matching Block and Fill in the Block to Be Repaired. After calculating the priority of all the blocks to be repaired on the boundary of the damaged area according
to step (1), the pixel block with the highest priority is set to \( \psi_p \), and then the global search strategy is adopted. That is, in the known area of the entire English text image, according to the SSD (mean square deviation between pixels in the English text image block) matching criterion, the English text image block \( \psi \) that is most similar to the pixel block to be repaired is searched out, which is used to fill the defect area in the block to be repaired. The calculation formula of SSD is as follows:

\[
SSD(p, q) = \arg\min_{\psi_p, \psi_q} d(\psi_p, \psi_q).
\]  

Among them, \( d(\psi_p, \psi_q) \) is the sum of squares of the \( R, G, \) and \( B \) color differences of the known pixels in the corresponding area of \( \psi_p \) and \( \psi_q \), and the formula is as follows:

\[
d(\psi_p, \psi_q) = \sum (I_{pR} - I_{qR})^2 + (I_{pG} - I_{qG})^2 + (I_{pB} - I_{qB})^2.
\]  

3.3. Update the Confidence of the English Text Image.

Generally speaking, every time a block to be repaired appears on the boundary, and each time a repair is completed, the boundary will change, so the boundary needs to be continuously updated. After the texture block with the highest priority is repaired, the previously unknown pixels become known pixels, and the confidence level has changed. Therefore, the confidence value needs to be continuously updated, which is updated according to the following formula:

\[
C(p) = C(\bar{p}), \quad \forall p \in \psi_p \cap \Omega.
\]  

Among them, \( C(p) \) refers to the confidence of the unknown pixel in the block to be repaired, and \( C(\bar{p}) \) refers to the confidence value of the center pixel of the block to be repaired. That is, after the repair of the pixel block with the highest priority is completed, the confidence value at the center of the pixel block is used to update the confidence value at the pixel just repaired.

The above three steps are repeated until all areas to be repaired in the English text image are repaired.

Figure 2 shows the specific flow chart of the Criminisi algorithm.

Although the Criminisi algorithm considers both structural information and texture information during the restoration process, it has great advantages compared to other restoration algorithms. However, after repairing different types of English text images, it is found that for English text images with strong structure, or when the area to be repaired is large, the repaired English text image will have obvious faults in the area to be repaired. “Phenomenon, and the entire repair time is longer and the efficiency is low.” The main reasons for these situations are as follows:

(1) When the algorithm calculates the priority of the block to be repaired, it uses the form of multiplication of the confidence item and the data item so that when one item is 0, the priority is 0, even if the value of the nonzero item is large. It is impossible to get priority repairs from time to time, so there may be an incorrect repair sequence, which will affect the final repair effect.

(2) The size of the texture block window utilised in the repair process is fixed, and the English text image’s local texture information is ignored. The error produced is huge if the block is too large, and the block is too tiny if the block is too small, slowing down the repair process.

(3) When determining the best matching block, the algorithm uses a global search strategy to search in the known area of the entire English text image. When the size of the English text image is large, it is very time-consuming conducive to practical applications.

(4) After determining the best matching English text image block, the pixels of the block to be repaired are directly updated. However, in actual situations, the two blocks cannot be completely matched. The error effect may not be obvious when the first repair is performed, but as time goes on, the error gets bigger and bigger, making the final repair effect very poor.

This article selects the appropriate template size based on the texture distribution information of the English text image. For English text images with rich texture, we choose a smaller window, and for English text images with weaker texture, we choose a larger window. Generally speaking, the gradient change of the pixels of the English text image can represent the amount of texture information contained in the English text image. The gradient of a certain pixel \((i, j)\) is calculated as follows:
$$\nabla g(i, j) = \left( g_x(i, j), g_y(i, j) \right). \quad (9)$$

Among them, $g_x$ and $g_y$ are calculated by the central difference method, as shown in the following formula:

$$g_x(i, j) = \frac{g(i + 1, j) - g(i - 1, j)}{2}$$

$$g_y(i, j) = \frac{g(i, j + 1) - g(i, j - 1)}{2} \quad (10)$$

The modulus of the gradient operator is calculated with the following formula:

$$|\nabla g(i, j)| = \sqrt{g_x^2 + g_y^2}. \quad (11)$$

The area with the larger gradient change indicates that the texture information contained is richer. At this time, a smaller window needs to be selected. On the contrary, the area with the smaller gradient change indicates that the texture information contained is less. At this time, it is necessary to select a larger window. Therefore, this article sets a method for selecting the template size based on the gradient information of the pixels of the English text image:

$$\text{size}(p) = \begin{cases} 
3, & |\nabla g_p^x| \geq 1.0, \\
5, & 0.3 < |\nabla g_p^x| < 1.0, \\
7, & |\nabla g_p^x| \leq 0.3.
\end{cases} \quad (12)$$

It can be seen from the above formula that this article uses different template sizes for different gradient English text images, including three sizes of $3 \times 3, 5 \times 5, 7 \times 7$. The reason for choosing an odd number is to ensure that the boundary pixel is located in the center of the template block.

Figure 3 is a comparison diagram of the English text image restoration result of the Criminisi algorithm and the English text image restoration result after the template block size is improved. Before proceeding with the image restoration of the English text, this article will locate the text area.

The calculation of priority is a crucial step in the repair process. It determines the repair order of the blocks to be repaired and directly affects the final repair result. The Criminisi algorithm calculates the priority of the block to be repaired by multiplying the confidence item and the data item. Although both texture information and structural information are considered, the following problems still exist:

1. The human eye is more sensitive to the structural information of English text images than the texture information, so in the calculation of priority, it is necessary to ensure that the data item $D(p)$ has a higher weight than the confidence item $C(p)$.

2. As the restoration of the English text image proceeds, the value of the confidence level $C(p)$ will quickly drop to 0. At this time, no matter what the value of the data item is, the calculation result of the priority tends to 0, which leads to the wrong repair sequence. The main reason is that multiplication is used to calculate the priority, which makes it too sensitive to the zero factor so the calculation form of the priority must be adjusted.

In response to the above-mentioned problems, this article has improved the original priority calculation formula, and the improved formula is as follows:

$$P(p) = mC(p) + nD(p). \quad (13)$$

Among them, $m + n = 1$. When $C(p) > 0$ and $D(p) = 0$, $m = 1$ and $n = 0$. When $C(p) = 0$ and $D(p) > 0$, $m = 0$ and $n = 1$. When $C(p) \neq 0$ and $D(p) \neq 0$, $m = 0.325$ and $n = 0.675$.

It can be seen from the formula that the formula for the priority of the block to be repaired has changed from multiplication to addition. In this way, when one of the data items and the confidence level are 0, the priority value will not be 0, which effectively solves the sensitive problem of the zero factor for multiplication. At the same time, the weight of the data item in the entire calculation is increased, and the maintenance of the structural information of the English text image is strengthened. At the same time, the texture information of the English text image is also retained, which is in line with the visual effect of the human eye.

Although the new priority calculation formula effectively adjusts the relationship between data items and confidence levels, it is known from their respective solving processes that $C(p)$ is obtained from the known pixel information in the area to be repaired and $D(p)$ is calculated only from the information of one pixel on the boundary, so the credibility is very lacking. Therefore, the data item $D(p)$ needs to be calculated by considering the relationship between the isoline of the entire block to be repaired and the boundary normal vector. The data item of the texture pixel block to be repaired is $D(\psi_p)$, and the corresponding formula is as follows:
\[ D(\psi_p) = \frac{\left| \nabla I_{\psi_p}^\perp \times n_{\partial \Omega_p} \right|}{\alpha} \]  

(14)

Among them,
\[ \nabla I_{\psi_p}^\perp = \frac{\sum_{q \in \psi_p \cap \Phi} \nabla I_q^\perp}{|\psi_p|}, \]
\[ n_{\partial \Omega_p} = \frac{\sum_{q \in \psi_p \cap \partial \Omega} P_q}{|\partial \Omega_p|}. \]  

(15)

In the formula, \( \nabla I_{\psi_p}^\perp \) represents the isoline tangent vector of the nondefective area in the texture pixel block to be repaired and \( |\psi_p| \) represents the area of the nondefective area in the to-be-repaired block, \( n_{\partial \Omega_p} \) represents the normal vector of the boundary in the texture block to be repaired, and \( |\partial \Omega_p| \) represents the total number of boundary pixels in the block to be repaired. Figure 4 is the corresponding texture restoration map.

Figure 5 is a comparison diagram of the repair result of the algorithm after improving the priority calculation of the block to be repaired and the repair result of the Criminisi algorithm.

The algorithm in this article not only improves the accuracy of the English text image restoration but also improves the efficiency of the restoration. The priority of each pixel on the boundary is calculated, and the pixel block with the highest priority is obtained. Then, it is necessary to search the most in the undamaged part of the English text image. The optimal matching block is used to replace the unknown pixel value in the block to be repaired. In the traditional Criminisi algorithm, a global search strategy is used, and SSD (Mean Square Error) is used as the matching criterion. Since location information is not considered during search and matching, the selected matching block may not be optimal, which leads to error repair; if such errors accumulate to a certain extent, it will affect the final repair effect.

The best matching block generally occurs in the zone that is comparable to the structure of the block to be repaired in an English text picture, which may be separated into multiple regions based on the different structure. The things in the neighborhood, according to the local similarity principle of the English text picture, are comparable. The structure and texture have a stronger association, but the correlation between items in other areas is less. As a result, there is a better chance that the matching block will emerge in the neighborhood. The standard for measuring the neighborhood is generally distance; the corresponding distance is smaller at this time. Therefore, in order to improve the accuracy of matching, the influence of distance needs to be considered in the matching calculation. When the SSDs between the matching blocks are the same, the one with the closest distance to the block to be repaired should be the best matching pixel block.

In summary, this article defines a new matching function as follows:
\[ D(\psi_p, \psi_q) = S(\psi_p, \psi_q) \times d(\psi_p, \psi_q). \] (16)

Among them, \( S(\psi_p, \psi_q) \) represents the sum of squares of the difference in color values of all pixels between the pixel block to be repaired and the source block of the known area, that is, the SSD matching function in the traditional Criminisi algorithm:

\[
S(\psi_p, \psi_q) = \sqrt{\sum (r_{\psi_p} - r_{\psi_q})^2 + (g_{\psi_p} - g_{\psi_q})^2 + (b_{\psi_p} - b_{\psi_q})^2}.
\] (17)

\[ d(\psi_p, \psi_q) \] represents the distance between the pixel block to be repaired and the coordinate point of the center pixel of the source block:

\[
d(\psi_p, \psi_q) = \sqrt{(x_{\psi_p} - x_{\psi_q})^2 + (y_{\psi_p} - y_{\psi_q})^2}.
\] (18)

Figure 6 is a comparison diagram of the repair result of the algorithm after the improved matching function and the repair result of the Criminisi algorithm.

In Figure 6, there is a texture repair error. For example, the color of the clothes on the legs of the person on the far right is filled, and the error is obvious. Secondly, there was a matching error in the place where the beach and the sea were grounded, which caused the error to accumulate and spread, and finally, a large piece of wrong texture appeared, which seriously affected the final repair effect. In contrast, there is no obvious texture repair error in Figure 6, including human legs and beaches, which have been correctly repaired.

It can be concluded that the matching criterion adopted in this article takes into account the influence of distance and guarantees the optimality of matching while ensuring the minimum SSD, thereby effectively avoiding the accumulation of errors and improving the repair effect.

4. Construction and Performance Testing of English Translation System Based on Intelligent Image Text Recognition

The input scene text image has been converted into a vector sequence containing local static feature information and context-related feature information. On this basis, it can be decoded in sequence to complete the positioning and recognition of the text according to the relevant feature information, and it will be completed content recognition and analysis of Chinese and English texts in images. The improved decoder is mainly composed of three parts: the local attention mechanism module, the cyclic neural network model, and the classifier. Among them, the local attention mechanism module will quantify and compare the vectors associated with the current position in the feature information vector sequence through the soft attention mechanism, measure the relative priority of each vector according to the quantized results, and screen out the vectors that are highly associated with the current position. By using normalized weight averaging, the input vector is utilized to synthesize the current position’s attention mechanism vector. The cyclic neural network will use the synthesized vector obtained through the local attention module as the input of the network, and combine the state of the previous moment in the network to complete the update of the state of the hidden layer at that moment. In order to reduce the overall structural complexity and computational complexity of the system, the gated recurrent unit (GRU) is also selected as the RNN structure. The classifier takes the current hidden state of the recurrent neural network as input, uses a 3-layer fully connected network to complete the classification of English text, finally completes the normalization of the probability distribution through a softmax layer, and selects the text with the highest probability. The recognition result of the current position is obtained.

Figure 6: Comparison of repair effects before and after the matching function improvement.
Table 1: Statistical table of the effect of image text feature recognition of English translation system based on intelligent image text recognition.

| No. | Image text feature recognition | No. | Image text feature recognition |
|-----|--------------------------------|-----|--------------------------------|
| 1   | 98.89                          | 34  | 98.28                          |
| 2   | 98.09                          | 35  | 97.57                          |
| 3   | 96.25                          | 36  | 98.71                          |
| 4   | 94.63                          | 37  | 96.99                          |
| 5   | 95.61                          | 38  | 96.18                          |
| 6   | 95.42                          | 39  | 96.25                          |
| 7   | 96.58                          | 40  | 98.86                          |
| 8   | 97.90                          | 41  | 94.14                          |
| 9   | 95.74                          | 42  | 96.43                          |
| 10  | 94.33                          | 43  | 95.18                          |
| 11  | 94.14                          | 44  | 93.38                          |
| 12  | 94.69                          | 45  | 98.95                          |
| 13  | 97.69                          | 46  | 95.97                          |
| 14  | 97.30                          | 47  | 97.70                          |
| 15  | 97.13                          | 48  | 93.94                          |
| 16  | 98.08                          | 49  | 95.02                          |
| 17  | 95.10                          | 50  | 95.98                          |
| 18  | 95.39                          | 51  | 96.28                          |
| 19  | 95.94                          | 52  | 95.61                          |
| 20  | 97.70                          | 53  | 98.39                          |
| 21  | 94.46                          | 54  | 95.58                          |
| 22  | 96.45                          | 55  | 95.03                          |
| 23  | 93.02                          | 56  | 95.78                          |
| 24  | 98.44                          | 57  | 96.34                          |
| 25  | 97.00                          | 58  | 94.55                          |
| 26  | 94.56                          | 59  | 94.92                          |
| 27  | 93.16                          | 60  | 95.33                          |
| 28  | 96.25                          | 61  | 95.09                          |
| 29  | 94.28                          | 62  | 95.74                          |
| 30  | 96.14                          | 63  | 98.86                          |
| 31  | 98.86                          | 64  | 98.62                          |
| 32  | 95.09                          | 65  | 95.51                          |
| 33  | 98.66                          | 66  | 96.73                          |

Figure 9: Statistical diagram of the effect of image text feature recognition of English translation system based on intelligent image text recognition.
The overall flow of the decoding process is shown in Figure 7.

In the classification process, because the English text in the natural scene in the selected dataset in this article mainly includes uppercase letters A–Z, lowercase letters a–z, numbers 0–9, and space characters, there are a total of 63 categories. Therefore, this is equivalent to a classification problem with 63 categories. Therefore, the output of the classifier is 63 dimensions, which is suitable for the 63 classification problems involved in this article. After that, the probability is normalized through a layer of softmax; that is, the probability distribution diagram of each possible character at the current moment is obtained, and the result with the highest confidence level can be selected as the output character at this time, as shown in Figure 8.

On the basis of the above analysis, the performance of the system constructed in this article is verified. The system in this article can extract English characters from image text and recognize English character features. In this article, 69 sets of image data are used to test the performance of the system, and the results are shown in Table 1 and Figure 9.

| No. | Image English translation effect | No. | Image English translation effect |
|-----|---------------------------------|-----|---------------------------------|
| 1   | 80.00                           | 34  | 79.84                           |
| 2   | 88.57                           | 35  | 86.71                           |
| 3   | 79.48                           | 36  | 85.46                           |
| 4   | 87.00                           | 37  | 80.67                           |
| 5   | 79.68                           | 38  | 78.03                           |
| 6   | 85.84                           | 39  | 85.16                           |
| 7   | 83.26                           | 40  | 80.69                           |
| 8   | 83.24                           | 41  | 84.84                           |
| 9   | 83.20                           | 42  | 87.72                           |
| 10  | 81.49                           | 43  | 83.57                           |
| 11  | 84.27                           | 44  | 86.68                           |
| 12  | 85.15                           | 45  | 79.96                           |
| 13  | 78.51                           | 46  | 85.60                           |
| 14  | 78.36                           | 47  | 88.59                           |
| 15  | 81.14                           | 48  | 82.69                           |
| 16  | 82.32                           | 49  | 82.13                           |
| 17  | 82.84                           | 50  | 82.65                           |
| 18  | 81.36                           | 51  | 84.55                           |
| 19  | 84.59                           | 52  | 80.50                           |
| 20  | 86.28                           | 53  | 78.14                           |
| 21  | 82.67                           | 54  | 83.73                           |
| 22  | 82.16                           | 55  | 81.71                           |
| 23  | 80.60                           | 56  | 88.73                           |
| 24  | 87.56                           | 57  | 85.37                           |
| 25  | 82.28                           | 58  | 84.33                           |
| 26  | 86.64                           | 59  | 81.03                           |
| 27  | 87.32                           | 60  | 82.52                           |
| 28  | 83.46                           | 61  | 84.58                           |
| 29  | 79.82                           | 62  | 81.85                           |
| 30  | 87.66                           | 63  | 87.23                           |
| 31  | 86.75                           | 64  | 78.42                           |
| 32  | 87.60                           | 65  | 80.69                           |
| 33  | 81.20                           | 66  | 81.92                           |

5. Conclusion

English recognition in natural settings has a wide range of applications, and it also involves the intersection of multiple disciplines. On one hand, the research on this problem can promote the development of application effects and popularity in various fields, and on the other hand, it will also lead to further theoretical development and technological innovation in related disciplines. In this article, considering the three aspects of texture template, block size, priority calculation, and optimal matching block search, an improved machine learning algorithm is proposed. In view of the different texture information contained in different images, texture blocks of adaptive size are used to effectively improve the accuracy and efficiency of repair. The restoration order is adjusted and the weight of structural information is enhanced by refining the priority calculation method, which effectively retains the image’s structural information and corresponds to the visual impact of human eyes. An English translation system based on intelligent visual text recognition is built based on the aforesaid analysis, and the performance of the English translation system in this study is confirmed by experimental research. The findings of this study back up the algorithm’s efficacy.
Data Availability
The data used to support the findings of this study are included within the article.

Conflicts of Interest
The authors declare that they have no conflicts of interest.

References
[1] R. R. Kamble and D. V. Kodavade, “Relevance Feature Search for Text Mining using FClustering Algorithm,” International Journal of Computer Science and Engineering, vol. 6, no. 7, pp. 223–227, 2018.
[2] M. Pandi and V. D. K. Rajendran, “Efficient Feature Extraction for Text Mining,” Advances in Natural & Applied Sciences, vol. 10, no. 4, pp. 64–73, 2016.
[3] G. Wu, M. Zhao, and L. Han, “A fingerprint Feature Extraction Algorithm based on Optimal Decision for Text Copy Detection,” International Journal of Security and Its Applications, vol. 10, no. 11, pp. 67–76, 2016.
[4] M. Soleymanpour and H. Marvi, “Text-independent speaker identification based on selection of the most similar feature vectors,” International Journal of Speech Technology, vol. 20, no. 1, pp. 99–108, 2016.
[5] M. Mojaveriyan, H. Ebrahimpourkoleh, H. Ebrahimpourkoleh, and S. J. Mousavirad, “IGICA: A Hybrid Feature Selection Approach in Text Categorization,” International Journal of Intelligent Systems and Applications, vol. 8, no. 3, pp. 42–47, 2016.
[6] M. H. Aghdam and S. Heidari, “Feature Selection Using Particle Swarm Optimization in Text Categorization,” Journal of Artificial Intelligence and Soft Computing Research, vol. 5, no. 4, pp. 38–43, 2015.
[7] Z. Robati, M. Zahedi, and N. Fayazi Far, "Feature Selection and Reduction for Persian Text classification," International Journal of Computer Applications, vol. 109, no. 17, pp. 1–5, 2015.
[8] T. Zia, Q. Abbas, Q. Abbas, and M. P. Akhtar, “Evaluation of Feature Selection Approaches for Urdu Text Categorization,” International Journal of Intelligent Systems and Applications, vol. 7, no. 6, pp. 33–40, 2015.
[9] T. Zia, M. P. Akhtar, and Q. Abbas, “Comparative Study of Feature Selection Approaches for Urdu Text Categorization,” Malaysian Journal of Computer Science, vol. 28, no. 2, pp. 93–109, 2015.
[10] Li. De, Z. J. Xue, and L. H. Cui, "Text recognition algorithm based on text features," International Journal of Multimedia & Ubiquitous Engineering, vol. 11, no. 5, pp. 209–220, 2016.
[11] K. Yan, Z. Li, and C. Zhang, “A New multi-instance multi-label learning approach for image and text classification,” Multimedia Tools and Applications, vol. 75, no. 13, pp. 7875–7890, 2016.
[12] G. Kumar and K. Vivekanandan, “Intelligent Model view Controller based Semantic WebService Call Through Mishmash Text Featuring Technique,” Journal of Computational and Theoretical Nanoscience, vol. 14, no. 4, pp. 2021–2029, 2017.
[13] B. Shi, X. Bai, and C. Yao, "An End-to-End Trainable neural network for image-based Sequence recognition and its Application to Scene Text recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 11, pp. 2298–2304, 2017.
[14] M. Al-Tahraw, “Polynomial neural networks versus Other Arabic Text classifiers,” Journal of Software, vol. 11, no. 5, pp. 418–430, 2016.
[15] Y. Hu, E. E. Milios, and J. Blustein, "Document Clustering with dual Supervision Through Feature Reweighting," Computational Intelligence, vol. 32, no. 3, pp. 480–513, 2016.
[16] R. J. Mcgill and A. R. Spurgin, “Assessing The Incremental value Of Kabc-li Luria Model Scores in predicting Achievement: what do They Tell Us beyond The Mpi?” Psychology in the Schools, vol. 53, no. 7, pp. 677–689, 2016.
[17] T. Mohoric and V. Taksic, “Emotional understanding as a predictor of socio-emotional functioning and school achievement in adolescence,” Psychology, vol. 49, no. 4, pp. 357–374, 2016.
[18] D. L. Rabiner, J. Godwin, and K. A. Dodge, “Predicting Academic Achievement and Attainment: The Contribution of Early Academic Skills, Attention difficulties, and Social Competence,” School Psychology Review, vol. 45, no. 2, pp. 250–267, 2016.
[19] G. T. L. Brown and H. Eklöf, “Swedish student perceptions of achievement practices: The role of intelligence,” Intelligence, vol. 69, no. 3, pp. 94–103, 2018.
[20] A. M. Beigi and A. Maroosi, "Parameter identification for solar cells and module using a Hybrid firefly and Pattern Search Algorithms," Solar Energy, vol. 171, no. 4, pp. 435–446, 2018.
[21] A. Gotmare, R. Patidar, and N. V. George, "Nonlinear system identification using a cuckoo search optimized adaptive Hammerstein model," Expert Systems with Applications, vol. 42, no. 5, pp. 2538–2546, 2015.
[22] Z. Chen, X. Yuan, H. Tian, and B. Ji, "Improved gravitational search algorithm for parameter identification of water turbine regulation system," Energy Conversion and Management, vol. 78, no. 5, pp. 306–315, 2014.