IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 29, 2020, accepted January 9, 2021, date of publication January 14, 2021, date of current version January 22, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3051613

News Video Title Extraction Algorithm Based on Deep Learning

SHUYIN LI[1] AND YANG LIU[2,3]

[1]School of Art and Design, Zhengzhou University of Aeronautics, Zhengzhou 450046, China
[2]School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China
[3]The Jackson Laboratory for Genomic Medicine, Farmington, CT 06032, USA

Corresponding author: Yang Liu (ieyangliu@zzu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61303044 and Grant 61572444.

ABSTRACT In order to better serve the needs of news business, researchers apply the information extraction technology of news headlines to the news field to assist decision-making. This article designs the shared convolutional layer and RPN network in the detection network respectively, and improves the depth of the shared convolutional layer, using VGGNet instead of ZFNet. A classification method incorporating semantic enhancement is designed for fine-grained topic category news classification. Combined with the idea of modular fusion mechanism, a semantic-enhanced classification model Multiple-Fusion is proposed. The Bert module replaces the traditional Word2Vec for semantic vector characterization, and introduces Bi-LSTM to adaptively extract context features, strengthens feature expression through the self-attention mechanism and adjusts network weights, and finally makes the model achieve accurate classification. This article designs a novel word-level data enhancement strategy for text data enhancement, which solves the problems of fewer training corpus samples and model overfitting. This article proposes a video target extraction method based on deep learning. The algorithm framework includes particle filtering, pre-training features, convolutional neural networks, discriminative classifiers, and online parameter updates. This method deeply combines deep learning models and traditional target extraction methods and frameworks. The experimental results show that the method in this article has relatively outstanding performance, and can adapt to many interferences encountered in the extraction process and the change of the target itself, and has strong robustness.

INDEX TERMS Deep learning, convolutional neural network, title extraction, news video.

I. INTRODUCTION

With the increasing popularity of Internet technology in China, the role of relevant online media in people's lives has become ever more important [1]. Network media is called the ''fourth media'' after ''newspapers'', ''broadcasting'' and ''television''. It has brought profound changes in the field of communication. Nowadays, more and more people rely on the Internet for information, and online news is therefore very popular among the public [2]. Newspapers and periodicals needed in order to go through many steps such as editing, typesetting, printing, and distribution before they could meet with readers [3]. Radio and television transmission of news required a certain amount of time and space, and they could only listen to them for a period of time. Online news is spread by digital and published on websites, which can minimize the time it takes for news to speak to the public, so that real-time news updates can make the information arrive instantly [4]. Therefore, the timeliness of online news and the large amount of information also attract people to choose online channels to view news. Politically, government or social managers learn about people's conditions through online news, including the people's support or rejection of a certain policy and political attitudes, to assist social managers to better formulate and optimize policies. Economically, business learns about people's demand and preference for certain types of goods through news, and assists businesses in formulating marketing strategies and product promotion [5]. Culturally, colleges and universities assist in formulating better student training programs through the collection of cultural-related information, so that the students trained by colleges and universities better meet the need of society for talents [6].

This article classifies and introduces the structure of news information web pages, and analyzes the structural

The associate editor coordinating the review of this manuscript and approving it for publication was Jinjia Zhou.

characteristics of news headlines, mainly news list pages and news detail pages. The news list page usually displays all or part of the headline information about the news. Some news list pages contain a list page of another topic news. You are required to view the detailed information to click on the list page to see the detailed page of the news. No matter how the news website is laid out, news headlines can be classified into one of these two categories. After studying the organizational characteristics of web pages, the paper studies the current general methods of information extraction and analyzes the effects of these methods on news headlines. Most of these methods combine the characteristics of the news headlines themselves, eliminating more complicated links to achieve extraction tasks. After analyzing the difficulties of news information extraction, the paper proposes its own extraction scheme, and designs two extraction methods. One is to automatically extract news content, and the other is to create extraction rules by manual participation. In the extraction process, the two methods can complement each other to increase the extraction efficiency and accuracy. Specifically, the technical contributions of this article can be summarized as follows:

First: This article studies a series of region-based convolutional neural network algorithms, analyzes and compares the advantages and disadvantages of the network model, and chooses the Faster RCNN framework. This article designs a news video title extraction scheme, specifies the two major tasks of feature extraction and candidate regions, and designs them separately. The extraction task of the candidate area adopts the full convolutional network RPN network, and the feature extraction task needs to be improved to VGGNet on the basis of ZFNet.

Second: A data enhancement strategy based on fine-grained word level is designed to enhance text data, which solve the problem of the small number of training corpus samples and the model overfitting. The Bert pre-trained language model is utilized to replace the traditional Word2Vec for semantic characterization, and then combined with the improved self-attention mechanism-based Bi-LSTM to adaptively extract contextual features to improve the generalization and accuracy of the classification model. Finally, an experimental comparison between the model in this article and other typical methods was implemented on the nlpcc2017 data set and the THUCNews data set, which proved the effectiveness and superiority of the model.

Third: The proposed convolutional neural network for target extraction has a beneficial effect in adapting to the visual feature ability of the extracted target and the real-time performance of the algorithm. The filter used in the convolution of the convolutional neural network is provided by the pre-training process, which improves the translation invariance of the depth model. Through the collection of positive and negative samples and the training of the classifier combined with particle filtering, the problems in the target extraction process are effectively overcome, such as the extraction frame slowly shifts and finally stays in the background with stable features and the adaptive extraction of the target. This achieves robust extraction of moving targets.

The rest of this article is organized as follows. Section 2 discusses related work. Section 3 analyzes news information extraction and digital video retrieval. Section 4 designs a news video title extraction model based on convolutional neural network. Section 5 analyzes the experimental results. Section 6 summarizes the full text.

II. RELATED WORK

Deep learning is a sub-field of machine learning, which is based on multi-layer representation learning, and each layer corresponds to a specific feature, factor or concept [7]. High-level concepts depend on low-level concepts, and the same low-level concept helps to determine multiple high-level concepts [8]. Deep learning is one of many machine learning algorithms based on representation learning. An observation object can be shown in many ways, but some representations can make sample-based learning tasks easier [9]. This field of research attempts to solve a problem: what factors can produce better representations, and how to become familiar with these representations. The essence of deep learning is to calculate the hierarchical features or representations of observation data, in which the high-level features or factors are obtained from the low-level [10], [11]. Deep learning can be divided into unsupervised or generative deep learning models, supervised deep learning models and hybrid deep network models according to its structure and application fields [12]–[14]. As a new machine learning method, deep learning has gained wide attention and development in recent years. In recent years, there have been many advances in deep learning algorithms and application research.

Related scholars proposed Sparse Deep Belief Net (SDBN) combined with sparse coding [15]. Since the original layer-by-layer greedy pre-training algorithm cannot effectively train Gaussian visible units, the algorithm uses sparse coding to train the first layer of the resonant network, and then uses the original Restricted Boltzmann Machine (RBM) learning algorithm to train the deep network. Researchers propose a deep network of active units to understand [16]. The local facial action unit proposed can effectively extract facial expression changes. The algorithm uses convolution and fusion methods to encode the local expression features of all possible regions in the image, and then searches the subspace of the extracted expression features through the action unit understanding layer, and simulates the fusion of the action units, and finally uses the RBM-based model to recognize faces expressions [17].

Deep learning has made great achievements in intelligent processing problems. In speech recognition, deep learning has changed the monopoly of Gaussian mixture model in speech modeling. Speech recognition experts from Microsoft Research collaborated with deep learning experts to use deep neural networks to change the original framework of speech recognition. Baidu's research found that the use of deep

neural networks to intersect the traditional Gaussian mixture model can reduce the false recognition rate by 25%. In terms of image recognition, related scholars have used the CNN network proposed by Lecun to achieve a major breakthrough in the recognition of Image Net data, which has gained widespread attention. The improvement in image recognition is not only due to new algorithms, such as the proposal of dropout, but also related to the substantial increase in computer parallel computing performance.

In topic detection and tracking technology (TDT), the basic idea of topic detection and extraction was first proposed in DARPA (Defense Advanced Research Project Agency) [18]–[20]. In addition to DARPA, related researchers such as Carnegie Mellon University and the University of Massachusetts have defined it to help people deal with the problem of massive information and develop its preliminary technology [21]. The formal research process of the TDT project is gradually developed through repeated evaluation meetings and TDT research [22]. Thanks to the support of DARPA, the American Institute of Standards and Technology, the TDT International Conference is held by NIST every year, and TDT related systems are evaluated during the conference [23]. Participants in the conference are from some well-known universities or research institutions, such as Carnegie Mellon University, BBN, Massachusetts Institute of Technology, IBM Watson Research Center, etc. In the domestic research on hot spot discovery, the more outstanding system is the Founder Wisdom Public Opinion Early Warning Aided Decision Support System launched by the Founder Technology Research Institute of Peking University, which successfully realizes automatic real-time monitoring and analysis of the massive public opinion information on the Internet [24], [25]. For the implementation of public opinion monitoring by related departments in a traditional manual way, the system successfully solved this problem [26]. Other domestic products mainly include Autonomy system, TRS public opinion information monitoring system, etc. In addition to the systems mentioned above, some domestic commercial companies, such as Xiamen Meiya Biotech, Bangfu Software, and Guni International Software, found that online public opinion is inseparable from national government public opinion monitoring [27], [28].

Relevant scholars analyzed the related technologies of text mining, and put forward a mining analysis model for online public opinion information [29]. The experimental analysis shows that text mining is feasible for public opinion analysis. Researchers have designed a topic-based network public opinion analysis model, which is designed after detailed analysis of the basic conditions of network public opinion [30]. Reports have proposed an incremental hierarchical clustering algorithm for topic discovery, which is an association of the main advantages of divided clustering and agglomerative clustering [31]. On the basis of thorough research on TDT, relevant scholars put forward a practical and effective single-granular topic identification method for the characteristics of the event and a hierarchical organization of topics based on the MLCS algorithm [32]. In view of the large scale and obvious periodicity of online news, scholars have designed a hot event discovery system for Internet news report streams, which can automatically discover hot events on the Internet in any time period [33]. After analyzing the needs of online public opinion, the researchers constructed a hot issue discovery and analysis system, which are based on clustering. Considering the discovery of hot information, the researcher proposed a hot information data mining network [34]. It uses the relevant characteristics of a complex network to aggregate and classify information, and this network is based on a scale-free topology. Related scholars extract the voiceprints of news anchors as features, put them into a neural network to learn, and then use the trained neural network to segment news videos [35]. I think this method is more accurate, but it is universal enough. It can only train one station, and then divide this station, which cannot achieve widespread effects. Researchers use neural networks to classify pixels in images into two categories: text-based pixels and non-text-based pixels. First, the gray level of each pixel around each pixel is input into the neural network for training as a feature. The neural network will adjust its weight through these samples. After the neural network training is completed, it can be used to test whether additional samples are text-based pixels. After the neural network test, the image must be reprocessed: the candidate text block is screened again by characteristics such as the length and width of the candidate text block, because the candidate text block whose height and width do not meet certain requirements cannot accommodate subtitles.

At present, there are a lot of research documents on the method of extracting web page text information. For example, relevant scholars have analyzed a large section of web page text information, and the text information is preceded by some format information (such as navigation information, interactive information, Java Script Script etc.) web page structure, and from this, a fast Fourier transform-based web content extraction algorithm is proposed. The algorithm uses window segmentation technology, uses statistical principles and FFT to obtain the possible weight of each possible area, and solves the best text interval. Experiments have shown that this method can more accurately extract ''text-style'' web page information. However, this method must be limited to a set of web pages with the same template. Because there are so many web page templates on the Web, this method cannot be promoted. Based on the document object model DOM, for the semi-structured features of HTML and the lack of semantic description, the path of the information to be extracted in the DOM hierarchy is regarded as the information extraction ''coordinates'', and structure-based filtering and semantic-based pruning can extract the topic information more accurately. They expand the hierarchical structure of the DOM tree, and re-integrate the visual characteristics and semantic information at the same time; finally determine the subtree containing the information block, and deeply tra-

verse the DOM tree to realize the web page text information extraction.

III. NEWS INFORMATION EXTRACTION AND DIGITAL VIDEO RETRIEVAL

A. SELECTION OF NEWS INFORMATION EXTRACTION METHODS

The news list page is a way for major portal websites to display numerous news. The news list page arranges news titles and release time in a certain order into a list, and sorts them according to the news content. In this approach, users can see the latest news at a glance and quickly find the news they want to read. Although the news list page contains noise such as navigation bars and advertisements like other web pages, the page structure is clear, and the addresses of news links have certain rules to be followed.

The news detail page refers to a web page that provides detailed information about news. For the definition of news details, there is currently no uniform and precise definition. Some studies on news extraction include links to news-related reports, and some even include evaluation information on this news. The News details defined by the paper only include the headline of the news, source, time, body information and editing information.

The title of the webpage is usually decorated with a more prominent style to achieve the effect that stands out in the surrounding environment and is eye-catching; the source, author/editor and time information of the webpage are also obvious in the webpage. It is found that these news elements have a fixed position in the news template, and they are often between the title and the main text, or the edited information is after the main text.

The rule-based information extraction method means that before information extraction; the system has existing rules for this field or type of webpage, and searches for target data points on the webpage for extraction according to the existing rules. The information extraction rules are mainly used to specify the context constraints that constitute the target data. As long as the information that meets the constraints contained in the rules is stated on the web page, this is the final result of information extraction. Information extraction using this method is mainly divided into two parts: rule learning and application rule extraction data. If the rules are not learned correctly, it will directly affect the ultimate extraction result. Therefore, the learning of rules has become the core part of the entire system, and data extraction is the future second.

This method can also be called Wrapper-based learning. The wrapper refers to the generation of extraction rules through semi-automatic or automatic means. A set of extraction rules and applications comprise a wrapper, thereby reducing the heavy work of constructing an extractor. Wrapper is a program that extracts structured data.

Using this method for information extraction, the extraction efficiency and accuracy are relatively high, but its high accuracy is generally limited to a specific field, and the portability is relatively poor. Once a new domain is used or the original domain template is changed, it is necessary to re-learn the rules, generate new rules and then extract them.

B. NEWS INFORMATION EXTRACTION MODEL

Based on the results of the web page clustering, the commonality analysis is performed on the web page sets of the same cluster, and the best extraction rule for this cluster is extracted. Generally, news headlines include not only news content, but also advertisements, related links, questionnaire surveys, participation in interactions, comments and other information. These are the parts that users who read the news do not care about. This is not extracted, only applicable to the main news. The news title, time, source and text information of the news, generate the corresponding extraction rules, and mark the semantic information of the data when generating the rules, such as which part is the news headline, which part is the news time, etc.

According to the generated extraction rules, the function functions provided by Html Parse are used to extract information from the corresponding web pages, and the extracted information is stored in a relational database to facilitate future data integration and subsequent value-added services.

The paper designs two schemes for extracting news: one is a semi-automatic extraction method, and the other is an automatic extraction method. The design diagram of the information extraction system is illustrated in Figure 1. The difference between the two schemes lies in the different methods used in rule generation. One is that manual participation is required to generate extraction rules, and the other is to automatically generate extraction rules.

C. DIGITAL VIDEO DATA

Video data is composed of a series of ordering images. This group of images does not constitute an isolated expression of irrelevant content, but a description of the same event that is continuous in time. Unlike machines, when humans observe images through vision, there is a phenomenon of persistence of vision, that is, when an image suddenly disappears from the field of vision, the human optic nerve will not immediately perceive this change. The residual illusion remains at 0.1 to 0.4 seconds. This makes that if a group of images are played at a certain speed, human vision cannot perceive the interval between images, creating a continuous visual effect. In the video field, this speed is at least 24 frames per second. Once the speed is less than this speed, humans will perceive a freeze, causing a feeling of frame dropping.

In a computer system, grayscale images are stored in the form of a two-dimensional matrix p(x,y), whose dimensions depend on the resolution of the image. The value of any position in the matrix corresponds to the pixel value of the corresponding position in the image. For example, p(x,y) represents the pixel value of the image at (x,y). Similarly, a three-dimensional matrix f(x,y,t) can be used to represent video data, and the image data can be arranged along the time

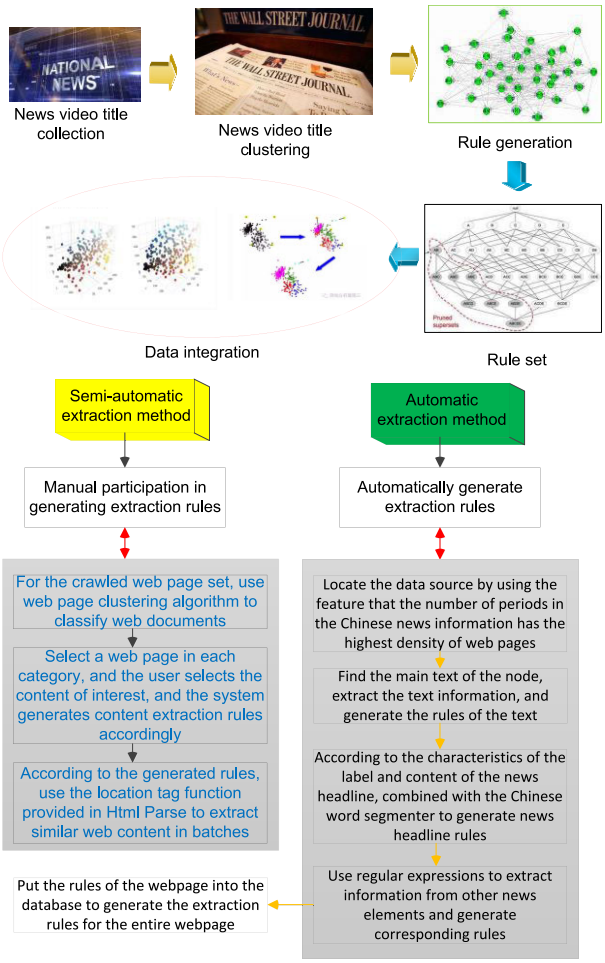


FIGURE 1. Schematic diagram of information extraction system design.

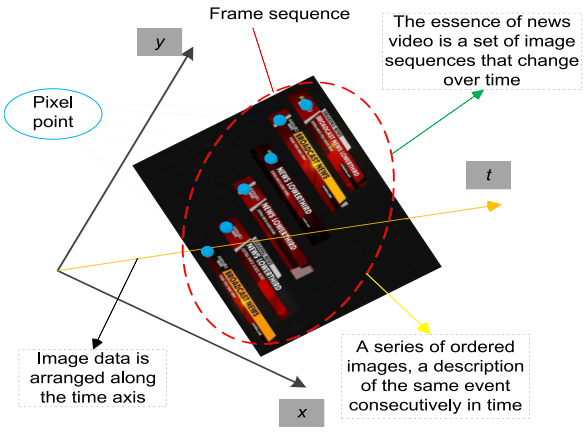


FIGURE 2. Schematic diagram of digital video.

axis, as shown in Figure 2. It can be seen that the essence of video is a set of image sequences that change over time.

Video is composed of a set of sequential images in form, and it also contains a variety of information such as text and audio. With the development of the time, human beings are no longer satisfied with monotonously accepting images or sound information. Only a complete video can fully describe a business or event. From a video, people can not only understand appearance characteristics, but more importantly, they can discover dynamic information about affairs. For example, the cause and effect of a story, as well as the various details of the process, is important information that other multimedia information cannot provide.

In a computer system, the storage space occupied by a single image or a text file is very limited, but if a video is stored, the storage space occupied by it will increase geometrically. For a 1080P ultra-definition movie, the resolution size of each frame of the image is 1920*1080, and each pixel occupies 24bit. Then if you follow the playback speed of 24 frames per second, the amount of data per second will reach 142M. Secondly, if data compression is not considered, a 1-minute ultra-clear short video will reach 8.3G data volumes. At the same time, video data has high requirements for storage space and transmission channels. Even if a small segment of the video is transmitted, a large amount of bandwidth and memory is required.

A group of sequential images in time constitutes the content of the video, and they often describe the same scene in the same shot. Therefore, this is a collection of image sequences with complex temporal and spatial relationships. In contrast, images or text files do not get such a complicated relationship.

D. CONTENT-BASED VIDEO RETRIEVAL FRAMEWORK

The previous system usually uses manual identification and tagging of video content. When searching, users can only use keywords to describe the general content of the video they need. The system bases on these keywords to find similar tags in the database. Content-based video retrieval is distinct from text-based video retrieval. It no longer uses a few separate keywords or descriptions for retrieval. Instead, it can use images or videos that describe in the content. The system automatically analyzes the content supplied by the user. The content-based news video retrieval framework mainly includes three parts: feature generation, feature management, and feature retrieval, as shown in Figure 3.

After the video retrieval system receives the video data, it uses the feature generation part to extract the features of the video. Video is a multimedia form that contains a variety of information, so the feature generation part not only needs to extract the feature information of the image in the video, but also consider the impact of the editing process on the video content. In addition to extracting from image frames, feature extraction can also start with other structural elements. For example, the event characteristics described by the shots and the organizational characteristics of the scene can be used in the process of video retrieval to improve the retrieval effect and make the retrieval results more in line with user expectations. These audio features and video summary form a video feature for retrieval.

In order to achieve efficient retrieval of large amounts of data, an index must be created on the feature database. In the feature management separate, the various features of the video are usually organized together in a special structure, which can not only effectively save storage space, but also

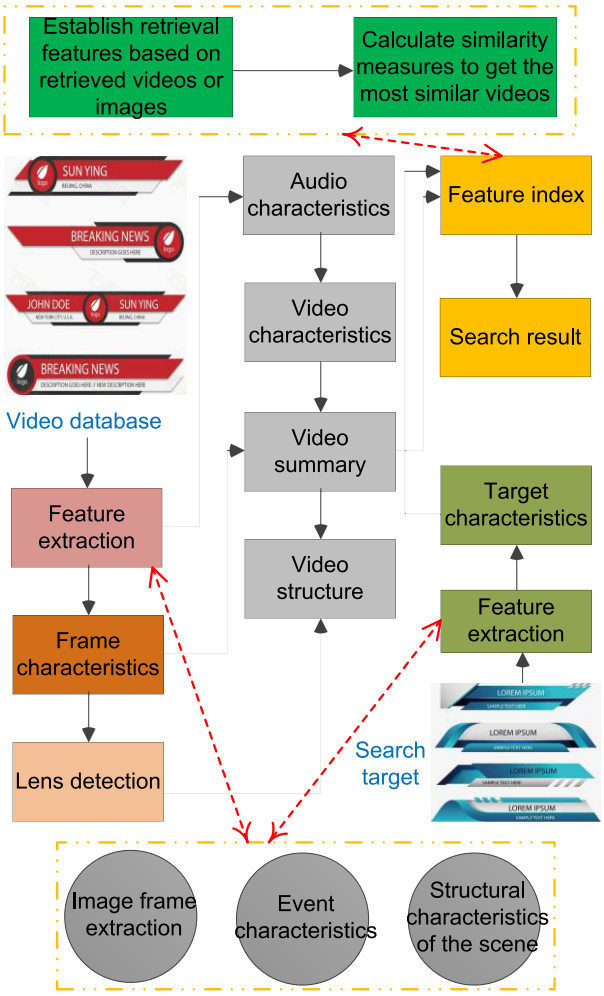


FIGURE 3. News video retrieval framework.

improve the user's retrieval experience. In commonly used databases such as My SQL, the structure of the index is mostly realized by establishing a B-tree or B+ tree, which greatly reduces the time for users to view.

E. VIDEO IMAGE FEATURE EXTRACTION

SIFT feature is currently one of the most important image features in the field of computer vision. Because it has many good features, it is widely used in many fields. The SIFT feature has good invariance to image scaling and rotation, brightness and three-dimensional viewing angle changes. At the same time, it can also be well positioned in the frequency and space domains, reducing the bad effects of occlusion, clustering and noise. In addition, this feature has good distinguishability, and can achieve good results in scene or object matching in large databases. Through a large number of filtering methods, the computational cost of extracting these features is minimized. Among them, the most consuming operation is the initial positioning operation.

The scale space of the image can be expressed as S (x, y, θ), which is convoluted by the variable scale function Gauss H (x, y, θ) and the input image function I (x, y).

$$S(x, y, \theta) = I(x, y) \otimes H(x, y, \theta) \tag{1}$$

$$H(x, y, \theta) = \frac{e^{\frac{2x^2+y^2}{2\theta^2}}}{2\pi\theta^2} \tag{2}$$

θ is called the scale space factor, which determines the degree of image blur and smoothing.

After the scale space is established, the algorithm is required to search for feature points with scale invariance in the space. Search for feature points usually uses the Laplacian of Gaussian. The SIFT algorithm has been improved. LoG is approximated by the difference of Gaussian (Do G) image D (x, y, θ), where L (x, y, θ) is the Gaussian scale space of the image, and k is the scale multiple.

$$D(x, y, \theta) = I(x, y) \otimes H(x, y, k\theta) - H(x, y, \theta) \tag{3}$$

The SIFT algorithm initially only locates candidate points at the location and scale of the pixels that satisfy the requirements. After that, improvements to the method were proposed. It uses a quadratic function to fit the limited data of the sampling point to find the precise location of the maximum value in the region, which also improves the stability of the algorithm. The method is to move the origin to the position of the candidate point, and then Taylor expands the scale space function D (x,y, θ):

$$D(x) = 0.5\Delta x^T \Delta x \frac{\partial^2 D}{\partial x^2} + \frac{\partial D^T}{\partial x}\Delta x^T + 0.5D \tag{4}$$

You solve the offset of the extreme position relative to the candidate position:

$$\Delta x = \frac{\partial D(x)}{\partial x}\frac{\partial^2 D^{-1}}{\partial x^2} \tag{5}$$

You substitute the obtained x into the Taylor expansion of D(x):

$$D(\vec{x}) = 0.5(\vec{x}\frac{\partial D^T}{\partial x} + D) \tag{6}$$

Only having scale invariance is not enough just to make SIFT so widely used, and it is also necessary to achieve rotation invariance for this feature. The main direction of the feature points is calculated and the rotation transformation is performed through a unified standard. For the rotated image, the approximate ratio of their gradients will not change. Once the main direction is rotated in the same direction, their gradient feature values will also be the same, which achieves rotation invariance. The modulus M (x,y) and direction Z(x,y) of the gradient of each point S (x,y) can be obtained by the following formula:

$$M(x, y) = \sqrt{[S(x+1, y)-S(x-1, y)]^2+[S(x, y+1)-S(x, y-1)]^2} \tag{7}$$

$$Z(x, y) = \arctan\frac{S(x, y+1) - S(x, y-1)}{S(x+1, y) - S(x-1, y)} \tag{8}$$

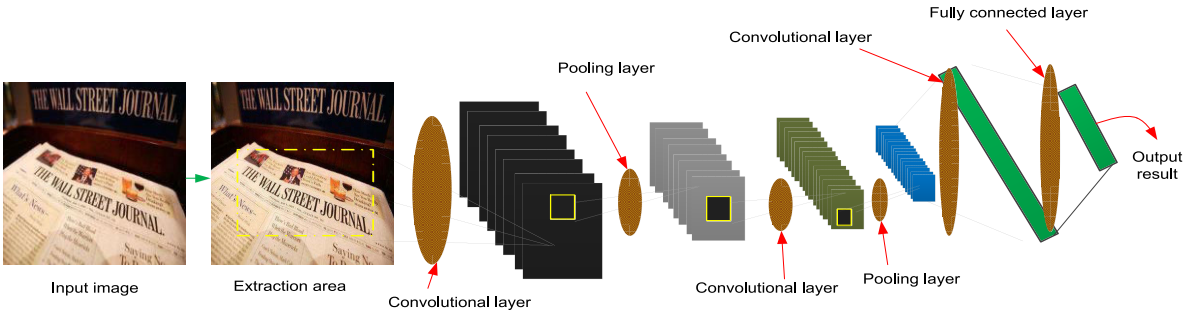


FIGURE 4. RCNN flow chart.

Since pixels in the same area have different effects on feature points, the longer the distance, the smaller the effect. Therefore, when calculating the gradient value of the feature point, it is not a simple addition of the gradients of neighboring pixels, and the influence of distance must be considered.

In the case of using the distribution histogram for statistics, all angles are divided into 36 equal parts, that is to say, the sum of the gradients is counted every 10 degrees. The sum here is not a simple addition, but a Gaussian weighted result. Finally, the overall gradient distribution of the restricted area of the feature point is obtained.

IV. NEWS VIDEO TITLE EXTRACTION MODEL BASED ON CONVOLUTIONAL NEURAL NETWORK

A. A SERIES OF ALGORITHMS OF REGION-BASED CONVOLUTIONAL NEURAL NETWORK

1) RCNN

The core of the detection problem is the image classification problem, and the core of the classification problem is the feature expression. RCNN breaks through the limitation of the number of convolutional layers and sets the number of convolutional layers to five layers, which is a great improvement compared to the commonly used two-layer convolution. Unlike image classification, the detection task needs to locate multiple targets in a picture. One method that has been used for at least two decades is the sliding window detector, which makes windows of different sizes slide on the image. If a certain window happens to contain a certain target, the position of the target is obtained. A similar pattern is still used in RCNN, but instead of exhaustive sliding search, a selective search method is used. This method quantifies the similarity of regions based on indicators such as color, texture and size, and gradually merges the small regions with the greatest similarity into large regions, and finally obtains candidate regions. The number of such candidate regions is far lower than the number obtained by exhaustive search.

When using RCNN for image detection, Selective Search will get about 2000 candidate region boxes from a single input image, and then the convolutional neural network will extract a fixed-dimensional feature vector from each candidate region, and then linear support vector machine (SVM) classifies the candidate regions. Finally, according to the scoring results of all candidate regions output by SVM, the target category in the candidate region is judged. The exact process is shown in Figure 4.

2) SPPNet

In the convolutional neural network, the convolution operation does not have a fixed requirement on the size of the input data, which are related to the principle of the convolution operation. The output after the convolution operation corresponds to the input size. If the end of the network structure is connected to a fully connected layer, then the input size of the image has to be set. This is explained by the fact that the input of the fully connected layer has a fixed dimension, otherwise the entire convolutional neural network cannot work normally. In response to such problems, the advent of SPPNet introduced the Spatial Pyramid Pooling (SPP) layer, which can be used to remove the limitation that the size of the network input must be fixed. The working mechanism of the SPP layer is to input the output result of the preceding convolutional layer into the SPP layer, generate a fixed-dimensional feature vector after pooling, and then send it to the first fully connected layer. A new SPP layer is inserted between the convolutional layer and the fully connected layer. This layer processes input of different sizes into outputs of the same size, so that any image input scale can be accepted.

The SPP mechanism has some good features. The first is its meaning, that is, it can produce the same size output regardless of the size of the input data. In addition, the unexpected benefit of the SPP mechanism is that the entire picture can be convolved at one time to reduce the time cost. In the SPP layer, spatial blocks of different scales are used to pool the response of each filter. The input size of the original image is 224×224, and the output of the last convolutional layer is $13 \times 13 \times 256$. It can be considered that there are 256 filters, and each filter corresponds to a 13×13 characteristic response graph. You pool the feature response map into three sub-feature maps of 4×4, 2×2, and 1×1, and then obtain a 21-dimensional feature vector. The final feature is a fixed-length $(16+4+1) \times 256$ feature vector. If the size of the image changes, the SPP layer can still output $(16+4+1) \times 256$ dimensional feature vectors. The pooling part of the fifth-layer network has the property of adaptive window size to a certain extent. No matter what size the initial input is, the feature vector with the required dimension can be obtained.

Candidate regions obtained by RCNN are identified one by one, and then the target with the highest score is selected as the detected target. Compared with RCNN, SPPNet is employed to detection. The candidate area is first selected by a selective search strategy. The difference is that SPPNet does not input each candidate area to the convolutional network, but only rolls the entire image once. Because the size and scale of the candidate regions are diverse, the feature response maps obtained by the mapping are still different. This requires the SPP layer to normalize the feature response maps of different sizes to the feature vector of the same dimension, and then perform classification and regression. Obviously, the computational load of RCNN becomes very large due to the processing of each candidate area, while SPPNet only needs to process the original image, plus some mapping operations, which greatly reduces the time and computational cost.

3) FAST RCNN

RoI pooling layer is the first to appear in Fast RCNN. It uses maximum pooling to convert the features in the region of interest into a small-scale feature map with a fixed spatial size, such as 7×7. Because the size of the region of interest is not fixed, the size of the pooling window where Ro I is located is also not fixed. Only when the two correspond to each other, a fixed-size feature map can be obtained. Compared with the SPP pooling layer, the Ro I pooling layer has only become a layer in the number of layers, and the function has not changed.

When each region of interest is manually labeled, it is classified as u, and is marked with the coordinate v of the target frame. Then you can combine the two factors of classification and positioning to form a multi-task loss function:

$$L(u, v, p, t^u) = L_{loc}(v, t^u) + 0.5L_{cls}(u, p) \tag{9}$$

The multi-task loss function of Fast RCNN brings convenience to network training. The main performance is that the two major functions of classification and positioning do not need to be trained step by step, and do not need to follow the sequence of training. The most important thing is why you do not need to save all the calculation results for the classifier. The disadvantage of Fast RCNN is that the acquisition of candidate regions still has to be done separately, which will still cause inconvenience when used.

The two methods, RCNN and SPPNet, first select the Selective Search method to extract the candidate regions, then use the convolutional neural network to achieve feature extraction, and finally train the SVM classifier. On this basis, further regression can be used to obtain the location frame of the detection target. The training process of RCNN and SPPNet is not completed at one time. The training process of multiple stages is more complicated, so the time cost and space cost are high. When SPPNet extracts features, it only needs to perform a forward convolution operation on the picture. The feature map corresponding to each candidate area can be obtained through spatial mapping, and it does not need to be obtained by convolution on each candidate area like RCNN. In terms of space storage, RCNN and SPPNet consume more space. This is because all the features used to train the SVM classifier need to be saved, which is also caused by the SVM training mechanism. In addition, the training of the convolutional network and the classifier are independent. Generally, the convolutional network is trained first to provide data for the training of the classifier, so the loss function of the classifier cannot be used for the parameter update of the convolutional layer. This will result in low utilization and slow training speed. Therefore, even if a deeper convolutional network is used for feature extraction, the accuracy of the classification network cannot be guaranteed to be improved.

B. NEWS VIDEO TITLE EXTRACTION SCHEME DESIGN

This article will design a network structure for news video title extraction based on the RCNN algorithm. When inputting a video, we extract one frame of the image and input it into the network portal. First, we extract the candidate area from the frame of the image, and the candidate area includes the news video title as much as possible. The traditional method is to evenly divide the image into several small blocks and adopt an exhaustive search strategy, which will result in a huge amount of calculation. Later, by selectively searching the entire image using the characteristics of the image's color, texture, object outline, etc., a small number of multi-scale candidate regions can be obtained, but this needs to be run on the CPU first. In this article, it is necessary to detect news video titles quickly or even in real time. In response to this need, a convolutional neural network is proposed to extract the candidate regions of news video titles. This will not only further reduce the number of candidate regions, but also reduce the number of regions. Choosing this step to incorporate GPU computing can achieve a true end-to-end deep network structure, which is conducive to networking training. The RPN network is maintained in circumstances under this demand. Instead of Selective Search, Edge Boxes and other methods, it uses the feature map after the convolution operation to generate the news video title candidate area, and the speed is significantly improved. This article uses the respective advantages of the RPN network and the Fast RCNN network to design the news video title extraction scheme shown in Figure 5.

As shown in Figure 5, the news video is simply extracted to obtain a single frame image, and there is no need for normalization here. The original RCNN algorithm needs to be normalized because the fully connected layer in the network requires fixed-dimensional input, and here the idea of a spatial pyramid pooling layer is adopted, and the Ro I pooling layer before the fully connected layer produces a fixed-dimensional output vector, thereby avoiding this problem and saving time for pre-processing the video. The single frame image is input to the shared network, and is forwarded to the previous shared convolutional layer through the convolutional neural network. The feature map obtained at this time has to be input to the RPN network, but also to continue forward with the Ro I convolutional layer. Feature response

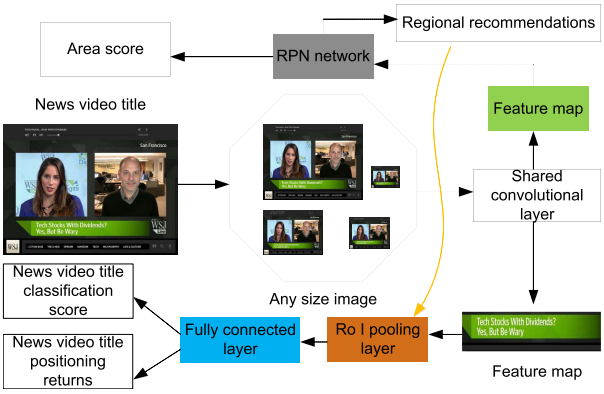


FIGURE 5. The specific scheme of news video title extraction.

map input to the RPN network is processed to obtain the score of the candidate regions and corresponding regions, and uses non-maximum value suppression to delete some inaccurate candidate regions. The scoring area is given to the Ro I pooling layer. The feature map output by the shared convolutional layer and the region proposal output by the RPN network are input to the Ro I pooling layer at the same time, and the higher-level features of the corresponding candidate region are extracted. After these higher-level features pass through the fully connected layer, the classification score of the news video title in the area and the location of the returned news video title can be output.

In general, the news video is processed frame by frame; the features are extracted through the shared convolutional layer, and the candidate regions are extracted by the RPN network. After the features are extracted again, the classification and positioning of the news video titles in each frame of pictures are finally output.

C. NEWS VIDEO TITLE EXTRACTION NETWORK DESIGN

Next, for the workflow of obtaining the classification and location of news video titles from news videos, specific news video title extraction network design is carried out, mainly for the design of the shared convolutional layer and the regional suggestion network.

1) SHARED CONVOLUTIONAL LAYER

The shared convolutional layer is to initially extract lower-level features in the image, such as contours, edges, and colors. The rapid improvement in the ability of convolutional neural networks to adaptively extract feature information is inseparable from the emergence of convolutional neural networks such as Le Net, Alexnet, ZFNet, VGGNet, Goog Le Net, and ResNet. In common target detection tasks, convolutional part in ZFNet acts as a shared convolutional layer and it is sufficient to deal with it.

The selection of the activation function is very acute in the convolutional network. The main function of the activation function is to ensure the network with nonlinear modeling capabilities. A general network lacking an activation function can only express a linear mapping. Even if there

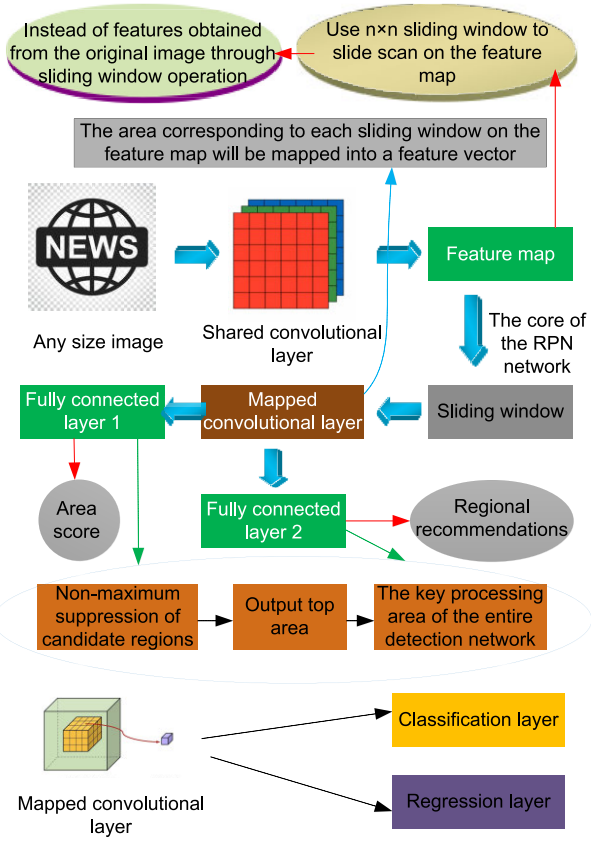


FIGURE 6. Structure diagram of RPN network.

are more hidden layers, the entire network is equal to a single-layer neural network. Widespread activation functions include Sigmoid function, Tanh function and Re LU function. The derivatives of the first two on both sides gradually approach 0, with soft saturation. If the convolutional layer is deep, the gradient disappears easily, resulting in poor network training effects or even stop. It happens that the Re LU function addresses this issue. When $x < 0$, Re LU is hardly saturated, and when $x > 0$, there is no saturation problem. Therefore, Re LU can keep the gradient unchanged and not attenuate when $x > 0$, thereby effectively alleviating the problem of gradient disappearance. This advantage can directly use the supervised training method instead of relying on unsupervised layer-by-layer pre-training, which greatly accelerates the training process. Therefore, there are many layers based on the convolutional neural network, and the Re LU function is invoked as the activation function here.

2) REGIONAL PROPOSAL NETWORK

The design of the sliding window is the core of the RPN network, and the RPN network is the core of the entire network structure, as shown in Figure 6. The RPN network essentially implements the functions of selective search and other methods, that is, regional recommendations. The mechanism of RPN is to add sliding window operation, mapped convolu-

tional layer and two fully connected layers after sharing the convolutional layer. Mapping convolutional layer maps the area corresponding to each sliding window on the feature map into a feature vector. The output of fully connected layers 1 and 2 respectively correspond to the area score of each sliding window position and the area suggestion after the position regression correction. After the non-maximum suppression operation is performed on the candidate area, the areas with the highest score are output, and these output areas are the areas that the entire detection network should focus on.

For the feature map generated by the shared convolutional layer, an n × n sliding window is used to slide and scan on the feature map, instead of obtaining features from the original image through the sliding window operation. In the design of this article, considering that the operation is a relatively high-level feature map, the sliding window can be designed as 3 × 3. The size of the sliding window is not large, but each sliding window can perceive a wide range of receptive fields. The area corresponding to each sliding window position is mapped to a 256-dimensional feature vector through the mapping convolution layer. The size of the convolution kernel of this mapping convolution layer is 3 × 3 × 256, and then the Re LU activation function is used. Each sliding window considers k possible reference window anchors. The reason for this is that a small window actually corresponds to a large receptive field. Specifically, a w × h feature map can generate w × h × k region suggestions, which not only guarantees the quality of the candidate regions, but also ensures the number of candidate regions. The low-dimensional feature vector is input to two fully connected layers, namely the reg layer positioning regression layer and the cls layer classification layer, which are used to regress the candidate area to generate the bounding box and score whether the candidate area is a target or background. In terms of quantity, each sliding window generates k candidate regions, the reg layer generates 4k translation and zoom parameters, and the cls layer gets 2k scores.

The anchor mechanism is to map a certain point on the feature map to a precise pixel on the original image, and use this pixel as the center to expand and design different regional suggestions. The specific method is to first find the reference point used to expand around, generally select the center point of the receptive field mapped from each sliding window, and then use this reference point to expand and select k anchors of unusual sizes and side lengths. There are usually 3 pixel scales, 128 × 128, 256 × 256, and 512 × 512, and 3 side length ratios, 1:1, 1:2, and 2:1. The anchor mechanism and the SPP mechanism are contradictory to a certain extent. Different pixel scales are selected for different target sizes, and different side length ratios are selected for different target shapes.

The Anchor mechanism can generate a large number of anchors, but in order to subsequently train the classifier, the anchor needs to be divided into positive and negative samples. The anchor with the largest overlap with the calibration area is set to a positive sample to ensure that each calibration area has at least one positive sample corresponding to it. After finding a positive sample for a certain calibration area, you continue to look for positive samples from the remaining anchors. This can set a threshold. If the overlap ratio of an anchor with this calibration area is higher than the threshold, you continue to set it as a positive sample. Therefore, there may be such a situation that a calibration area may have multiple positive sample anchors, that is, one-to-many, but each positive sample anchor can only belong to one calibration area, that is, only one-to-one. The definition of a negative sample is of course that the overlap ratio is lower than the set threshold.

3) IMPROVE CONVOLUTIONAL NETWORK

The design of the news video title extraction network mainly uses the advantages of the RPN network, but the design of the RPN network is very mature and can cope with the extraction of candidate regions in general scenarios. In view of the variety and shape of news video titles on real news videos, this article attempts to deepen the depth of the shared convolutional layer to obtain more news video title features, and it also helps the RPN network to extract candidate regions. Considering that shared convolutional layers are mainly convolution operations, you can consider using a deeper network VGGNet to replace the shared convolutional layer. Here we choose 16-layer VGGNet.

The network structure of VGGNet-16 consists of 13 convolutional layers, 5 pooling layers, and 3 fully connected layers. The specific improvement method is that the fifth Pool layer of VGGNet-16 is replaced by the Ro I layer. The 13 convolutional layers in VGGNet-16 are used as part of feature extraction.

V. EXPERIMENTAL RESULTS AND ANALYSIS

A. VALIDATION OF NEWS DATA ENHANCEMENT STRATEGY

In order to verify the effectiveness of the word-based fine-grained data enhancement strategy proposed in this article, the nlpcc2017 data set with moderate news type, refined content, and annotated authority is selected as the experimental data for testing. In the data enhancement experiment, the α parameter and β parameter are significant factors that affect performance changes. Among them, the α parameter represents the percentage of the number of words in the text that changes with each expansion; the β parameter represents the number of texts generated by each original text.

1) α PARAMETER DETERMINATION

First we determine the optimal parameter of α. The specific method is to separate the five operating methods of the data enhancement strategy and conduct an experimental analysis one by one, and then confirm its ability to improve performance by changing the α parameter. For these five operation

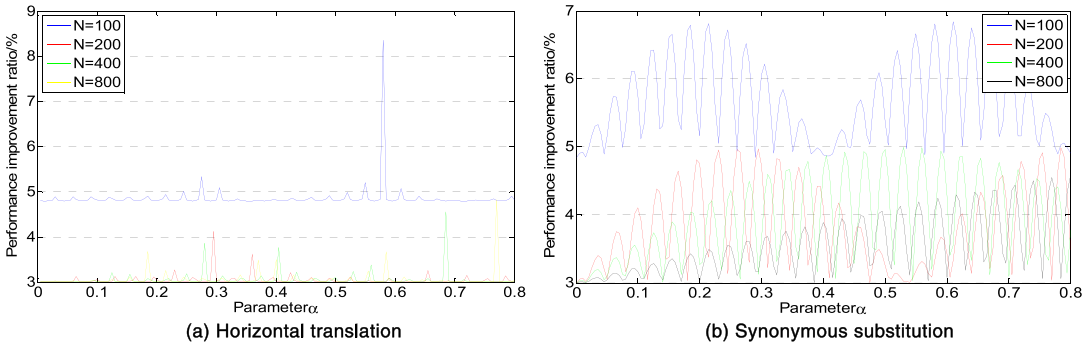


FIGURE 7. Proportion of each operation method to improve classification performance.

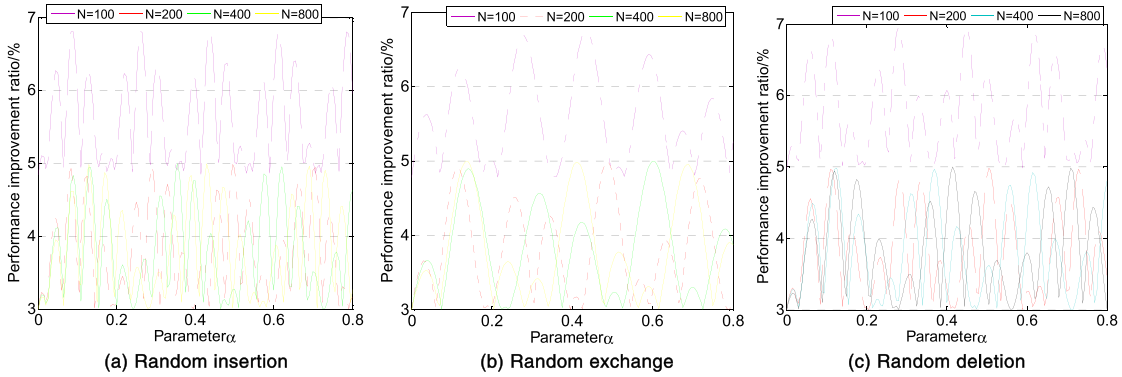


FIGURE 8. Performance improvement ratio of each operation classification.

methods, this experiment uses a single operation method to execute the algorithm code, and the results are presented in Figure 7 and Figure 8.

The vertical axis in Figure 7 represents the performance improvement ratio, that is, the percentage of change in classification accuracy before and after the data enhancement strategy is adopted. If more than half of the words are translated, the sentence will be incomprehensible. For synonymous replacement operations, the curve trend in the figure and the horizontal translation curve tendency are quite different, because if too many words are replaced, the semantics of the original sentence will be lost.

As shown in Figure 8, for the random insertion operation, setting different α values, the overall performance improvement ratio is in a relatively stable trend. Because in this operation, the original words and relative order of the sentences in the text are still maintained. For random exchange operations, because too many unplanned exchanges of sentence words will change the original structural order of the text grammar, resulting in poor performance. For random deletion operations, the performance improvement ratio fluctuates as the value of α increases. The reason is that too many words are randomly deleted and the sentence will be meaningless.

In summary, all the operating methods of the data enhancement strategy proposed in this article have a greater performance improvement on a smaller data set, and the peak in the figure is the best performance value.

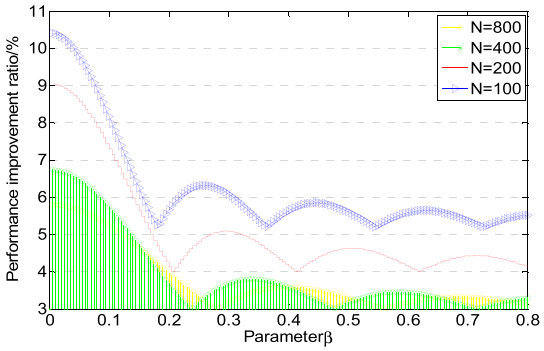


FIGURE 9. Data enhancement strategy generation quantity performance improvement change graph.

2) DETERMINATION OF β PARAMETERS

We determine the optimal parameter of β, that is, determine the number of expanded text generated by each creative text. The experiment sets the value range of β from 0 to 0.8, and tests on data sets of different sizes. The experimental results are shown in Figure 9.

By comparing N = 100 with data sets of other sizes, it can be seen that the overall performance improvement is more obvious for the experiment with the training set of N = 100 scales. For training sets with N = 200 or larger, the performance improvement is relatively insignificant. Built on the analysis of all the above experiments, data enhancement is performed on data sets of different scales.

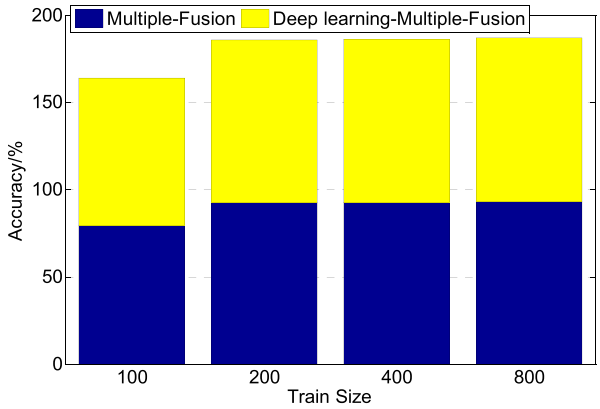


FIGURE 10. Comparison of model training experimental results of fusion data enhancement strategy.

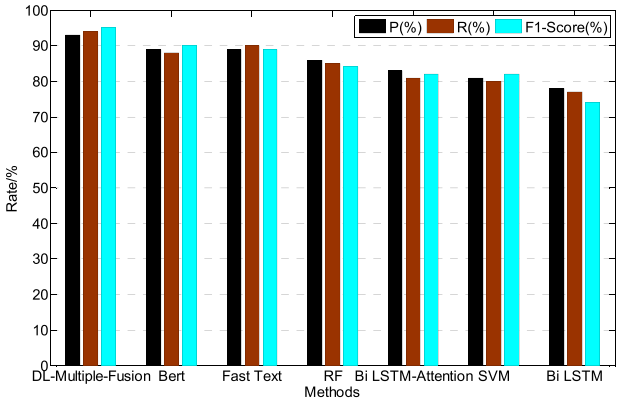


FIGURE 11. Performance comparison of various algorithms in the nlpcc2017 data set.

As the size of the data set is determined, the two α and β parameters that affect the classification performance are also determined.

3) COMPARATIVE VERIFICATION

Finally, the experiment set up a set of comparative experiments to compare the proposed classification model based on data enhancement and experiments without data enhancement strategies, and explore the impact of different scales of data on the classification effect. Among them, Deep Learning-Multiple-Fusion represents the proposed based on the deep learning classification model of data enhancement strategy fusion mechanism, Multiple-Fusion represents the proposed classification model without data enhancement strategy fusion mechanism. The experimental results are shown in Figure 10.

It can be observed in Figure 10 that the model training is performed after the data is expanded by using the deep learning data enhancement strategy. The classification result is linked to the size of the data set. When N = 100, the accuracy of Deep Learning-Multiple-Fusion is about 4% higher than that of Multiple-Fusion. When N400, the two have generalized due to the large training set, resulting in the classification effect is not clear. In summary, the data enhancement strategy enriches the data set, and the effect of improving the classification performance is more obvious, especially when the size of the data set is smaller.

B. VALIDATION OF FUSION NEWS SEMANTIC ENHANCEMENT MECHANISM

In order to highlight the effectiveness and cutting edge of the semantic enhancement-based classification model proposed in this article, this experiment compares the proposed model with some predictable text classification algorithms. For the comparison algorithm involved in this experiment, this article will strictly follow the source code disclosed in the paper for parameter configuration, training and analysis.

First, this part conducts algorithm comparison experiments on the nlpcc2017 Chinese news headline data set. The data set contains 18 news types such as military, current affairs, and technology, with a total of 228,000 sample data. This article extracts part of the data as experimental data, selects 1000 of each type, and divides the training, testing and validation sets according to the ratio of 6:2:2. Then we use precision (Precision, P), recall (Recall, R) and F1 value (F1-Score, F) three performance indicators to evaluate the algorithm. The experimental results are shown in Figure 11.

It can be seen from the experimental results in Figure 11 that the effects of these algorithms are quite different. The three performance indicators of the traditional SVM algorithm are all near 80%, while the deep learning model is above 80%. The reason is that the data set is a short text of Chinese news headlines, and the short text has the characteristics of sparse content and strong context dependence. When the SVM algorithm is used for text representation, the characteristics are relatively sparse. The CNN model uses the convolution kernel to fully extract the features around the word vector, so that its classification precision P is 80.12%. Since Bi LSTM and Bi LSTM-Attention have the ''gated'' structure of long and short-term memory, they can better solve the problem of context dependence of text sequences, so the performance indicators have achieved good results, and the F1 value of Bi LSTM-Attention has reached 86.58%. The Fast Text algorithm uses n-gram training word vectors and combined with hierarchical Softmax classification to ensure extremely fast training and testing speed while maintaining 87.96% precision value. Although Bert is a single structure of the classification model Multiple-Fusion, it still guarantees an extremely high level of classification.

The semantic enhancement classification model Multiple-Fusion proposed in this article has achieved the best results in all performance indicators. The fusion mechanism combines the advantages of the Bert and Bi LSTM methods, and utilizes the self-attention mechanism to enhance the expression of important features, making the classification effect the most effective good. In summary, experiments show the superiority of the Multiple-Fusion method in the application of Chinese news headline classification.

The THUCNews data set contains 10 news types such as finance and current affairs, with a total of 65,000 sample data.

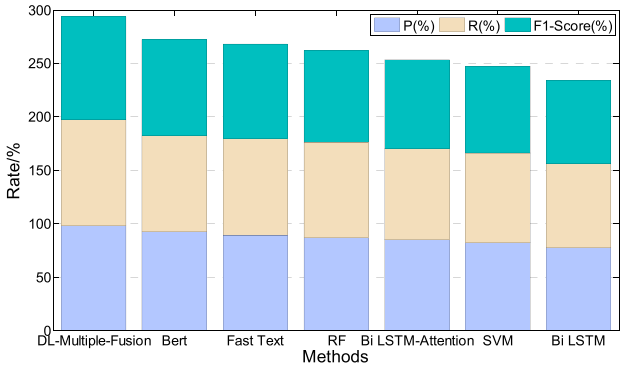


FIGURE 12. Performance comparison of various algorithms in the THUCNews data set.

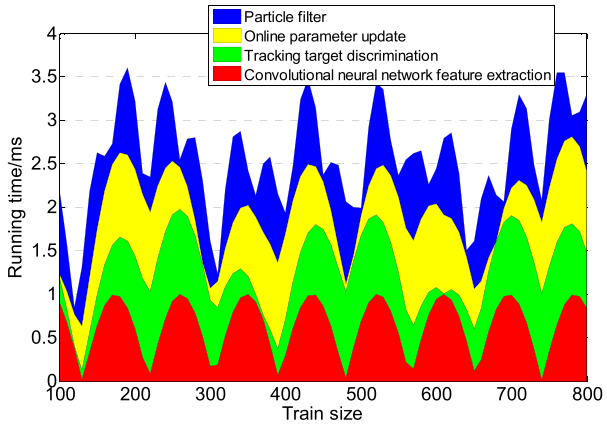


FIGURE 13. The running time of each part of the extraction algorithm in this article.

This experiment will select a 1000-scale training set for each class to verify the performance of each algorithm on the data set. Classification performance indicators are consistent with the above. The result is shown in Figure 12.

It can be seen from Figure 12 that by comparing and analyzing the various indicators of each algorithm in a single data set. It is known that the performance of each algorithm shows a decreasing trend. By comparing the various indicators of the data lay down in each algorithm, the training set has obvious advantages in the DL-Multiple-Fusion algorithm. Compared with the nlpcc2017 data set, all indicators are further increased by about 4%. This is because the THUCNews data set provides the content of the news text. Compared with the news headline data set, the content of each sample of the data set is richer.

C. REAL-TIME EXPERIMENT AND ROBUSTNESS ANALYSIS

1) REAL-TIME EXPERIMENT

Figure 13 counts the running time of the algorithm in this article, and gives the time made by each part of the algorithm in each frame of the experiment. The algorithm runs on a PC, CPU i7 3770 (3.4Hz), Memory 8GB. The statistical data is the average time performed for each frame of each video sequence of the experiment.

It can be seen from Figure 13 that the average processing time of one frame is about 5ms, which can basically meet the real-time requirements in practical applications. Judging from the time distribution of each part of the algorithm, the execution time of the extraction algorithm is mainly consumed by the particle filter of the convolutional neural network. In the experiment, the calculation of the convolutional neural network is run on the CPU, which contains many matrix operations. If the GPU can be used to optimize the parallel calculation, the calculation time for the convolutional neural network will be greatly reduced. For the parameter online update process, because the gradient descent method is used to converge the optimized function, the time consumption is relatively high. Note that the parameter online update is not performed every frame. The time indicated in the above figure is the average time, so occasionally, the execution time of the current frame will become longer when the parameters are updated.

For this kind of time performance, a compromise between running speed and accuracy has been made in many parts of the algorithm, such as the number of particles delivered by the particle filter, the number of pre-training features and the number of hidden layer neurons in the convolutional neural network. Adjusting these numbers, such as increasing the number of particles and the number of pre-training features, can further improve the accuracy of the algorithm, but the execution time will increase, so the algorithm in this article considers the accuracy and running speed in the selection of these parameters.

2) ROBUSTNESS ANALYSIS

In many realistic scenes, the extraction of moving targets will be subject to a lot of interference, such as lighting changes, partial occlusion, motion blur, etc. The target itself will also undergo many changes such as in-plane and out-of-plane rotation, shape changes, and posture changes. The emergence of these situations brings certain difficulties with video target extraction. Therefore, in order to make a target extraction algorithm strong, it must be able to adapt and solve these interference and changes in the extraction process. The video sequence used in the experiment in this article contains many interference and target changes. The following experiment is to verify and analyze the ability of the extraction algorithm to deal with these interference and changes.

When the target is irradiated by natural light or other light sources during the movement, when the intensity of the light source changes or the appearance and disappearance of the light source will cause the target's illumination to change. This kind of interference will have a greater impact on the target extraction algorithm that depends on the gray value. Here we use the news video title sequence for experiments, and the experimental results are shown in Figure 14. It can be seen from Figure 14 that the extraction result of the target is relatively accurate, and the position of the target can be accurately captured. However, the target area is usually slightly smaller, but the target location is accurate.

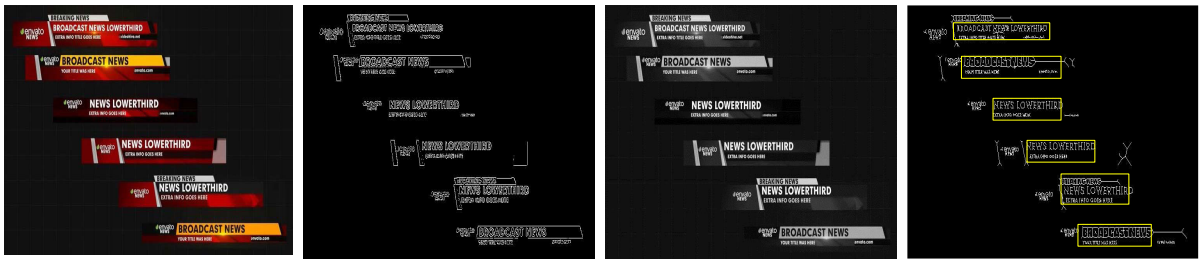


FIGURE 14. Extraction results of news video titles.

FIGURE 15. Sylvester video sequence extraction results.

Changes in the target itself are very common, and there are many kinds of changes in the target itself, including changes in the pose, shape changes, in-plane rotation and out-of-plane rotation of the target. The various rich changes of the target itself also bring great challenges to the video target extraction task, because the initial conditions of the extraction task contain relatively little information, and only a small amount of information about the target, such as the target information of one angle or one of the poses. To adapt to the various changes of the target itself during the extraction process, the extraction algorithm needs to have effective learning and model online update capabilities. We also choose Sylvester news videos for experiments, and the experimental results are shown in Figure 15. From the results, the algorithm in this article can basically extract the target without losing the extracted target or deviating too far from the target, but there is still a slight deviation in the extraction process.

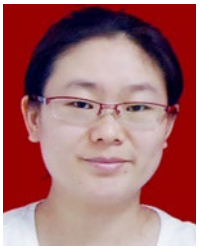
VI. CONCLUSION

When using artificially designed learning features to detect news video titles, there will be problems such as cumbersome learning feature design process and limited adaptation range. This article uses convolutional neural networks to automatically extract features. Based on the region-based convolutional neural network (RCNN), a news video title extraction scheme is designed, which combines the advantages of the Fast RCNN framework and the RPN region suggestion network. In view of the characteristics of the different shapes of the contours of news video titles, this article improves the shared convolutional network in the news video title extraction network, mainly to deepen the depth of the convolutional network, from 5 layers of convolution to 13 layers. Aiming at the problem of sparse news headline features and strong context dependence in news text classification, a news text classification method based on semantic enhancement is studied. The Bert language model is used to replace the traditional Word2Vec for semantic representation, and then combined with the improved self-attention mechanism-based Bi-LSTM to adaptively extract contextual features. Combining the idea of module fusion and based on the above structure, a semantically enhanced classification model Multiple-Fusion is proposed. In addition, this article also designs a word-level data enhancement strategy for text data enhancement to improve the generalization and accuracy of the classification model. Video target extraction algorithm based on deep learning is proposed. The algorithm framework includes particle filtering, pre-training features, convolutional neural networks, discriminative classifiers and online parameter updates. This method deeply combines the deep learning model and the mainstream target extraction framework, and replaces the appearance model in the original extraction framework with the deep learning model, making full use of the feature extraction capabilities of the deep learning model. This has outstanding performance in actual use, and can adapt to various interferences encountered in the extraction process and changes in the target itself, so it has strong robustness.

REFERENCES

[1] A. Bruns, A. Kornstadt, and D. Wichmann, "Web application tests with selenium," *IEEE Softw.*, vol. 26, no. 5, pp. 88–91, Sep. 2009.
[2] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018.
[3] Z. Qiu, D. Miller, and G. Kesidis, "Semisupervised and active learning with unknown or label-scarce categories," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 4, pp. 917–933, Apr. 2017.
[4] F. Z. Xing, E. Cambria, and R. E. Welsch, "Intelligent asset allocation via market sentiment views," *IEEE Comput. Intell. Mag.*, vol. 13, no. 4, pp. 25–34, Nov. 2018.
[5] E. N. Yılmaz and S. Gönen, "Attack detection/prevention system against cyber attack in industrial control systems," *Comput. Secur.*, vol. 77, pp. 94–105, Aug. 2018.

[6] M. Ahmad, D. I. U. Haq, Q. Mushtaq, and M. Sohaib, "A new statistical approach for band clustering and band selection using K-means clustering," *IACSIT Int. J. Eng. Technol.*, vol. 3, no. 6, pp. 606–614, Dec. 2011.
[7] H. Soleimani and D. J. Miller, "ATD: Anomalous topic discovery in high dimensional discrete data," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 9, pp. 2267–2280, Sep. 2016.
[8] B. Biggio, G. Fumera, and F. Roli, "Security evaluation of pattern classifiers under attack," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 4, pp. 984–996, Apr. 2014.
[9] F. Almonacid, E. F. Fernandez, A. Mellit, and S. Kalogirou, "Review of techniques based on artificial neural networks for the electrical characterization of concentrator photovoltaic technology," *Renew. Sustain. Energy Rev.*, vol. 75, pp. 938–953, Aug. 2017.
[10] H. Soleimani and D. J. Miller, "Parsimonious topic models with salient word discovery," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 3, pp. 824–837, Mar. 2015.
[11] P. Lopes and B. Roy, "Dynamic recommendation system using Web usage mining for E-commerce users," *Procedia Comput. Sci.*, vol. 45, pp. 60–69, 2015.
[12] H. Gunduz, Y. Yaslan, and Z. Cataltepe, "Intraday prediction of borsa istanbul using convolutional neural networks and feature correlations," *Knowl.-Based Syst.*, vol. 137, pp. 138–148, Dec. 2017.
[13] A. Picasso, S. Merello, Y. Ma, L. Oneto, and E. Cambria, "Technical analysis and sentiment embeddings for market trend prediction," *Expert Syst. Appl.*, vol. 135, pp. 60–70, Nov. 2019.
[14] M. Ahmad, M. A. Alqarni, A. M. Khan, R. Hussain, M. Mazzara, and S. Distefano, "Segmented and non-segmented stacked denoising autoencoder for hyperspectral band reduction," *Optik*, vol. 180, pp. 370–378, Feb. 2019.
[15] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.
[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[17] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018.
[18] L. Borges, B. Martins, and P. Calado, "Combining similarity features and deep representation learning for stance detection in the context of checking fake news," *J. Data Inf. Qual.*, vol. 11, no. 3, pp. 1–26, Jul. 2019.
[19] M. A. Beam and G. M. Kosicki, "Personalized news portals: Filtering systems and increased news exposure," *Journalism Mass Commun. Quart.*, vol. 91, no. 1, pp. 59–77, Mar. 2014.
[20] D. T. Davis and J.-N. Hwang, "Solving inverse problems by Bayesian neural network iterative inversion with ground truth incorporation," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2749–2757, Nov. 1997.
[21] M. Ahmad, D. I. U. Haq, Q. Mushtaq, and M. Sohaib, "A new statistical approach for band clustering and band selection using K-means clustering," *IACSIT Int. J. Eng. Technol.*, vol. 3, no. 6, pp. 606–614, Dec. 2011.
[22] J. Maillo, S. Ramírez, I. Triguero, and F. Herrera, "KNN-IS: An iterative spark-based design of the k-Nearest neighbors classifier for big data," *Knowl.-Based Syst.*, vol. 117, pp. 3–15, Feb. 2017.
[23] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.
[24] J. Bai, L. Li, D. Zeng, and Q. Li, "Associated activation-driven enrichment: Understanding implicit information from a cognitive perspective," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 12, pp. 2655–2668, Dec. 2017.
[25] H. Xiao, B. Biggio, B. Nelson, H. Xiao, C. Eckert, and F. Roli, "Support vector machines under adversarial label contamination," *Neurocomputing*, vol. 160, pp. 53–62, Jul. 2015.
[26] C. Kereliuk, B. L. Sturm, and J. Larsen, "Deep learning and music adversaries," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 2059–2071, Nov. 2015.
[27] A. Demontis, M. Melis, B. Biggio, D. Maiorca, D. Arp, K. Rieck, I. Corona, G. Giacinto, and F. Roli, "Yes, machine learning can be more secure! A case study on Android malware detection," *IEEE Trans. Depend. Sec. Comput.*, vol. 16, no. 4, pp. 711–724, Jul. 2019.
[28] A. Kurve, D. J. Miller, and G. Kesidis, "Multicategory crowdsourcing accounting for plurality in worker skill and intention and task difficulty," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 3, pp. 794–809, Mar. 2015.
[29] M. W. Graham and D. J. Miller, "Unsupervised learning of parsimonious mixtures on large spaces with integrated feature and component selection," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1289–1303, Apr. 2006.
[30] P. Ristoski, C. Bizer, and H. Paulheim, "Mining the Web of linked data with RapidMiner," *J. Web Semantics*, vol. 35, pp. 142–151, Dec. 2015.
[31] M. K. Sarma and A. K. Mahanta, "A DOM-tree based representation of Web document structure for Web mining applications," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 5, no. 6, pp. 1437–1439, Oct. 2017.
[32] P.-H. Chen, H. Zafar, M. Galperin-Aizenberg, and T. Cook, "Integrating natural language processing and machine learning algorithms to categorize oncologic response in radiology reports," *J. Digit. Imag.*, vol. 31, no. 2, pp. 178–184, Apr. 2018.
[33] S. Mullainathan and J. Spiess, "Machine learning: An applied econometric approach," *J. Econ. Perspect.*, vol. 31, no. 2, pp. 87–106, May 2017.
[34] L. Borges, B. Martins, and P. Calado, "Combining similarity features and deep representation learning for stance detection in the context of checking fake news," *J. Data Inf. Qual.*, vol. 11, no. 3, pp. 1–26, Jul. 2019.
[35] A. I. Taloba, D. A. Eisa, and S. S. Ismail, "A comparative study on using principle component analysis with different text classifiers," *Int. J. Comput. Appl.*, vol. 180, no. 31, pp. 1–6, Apr. 2018.



SHUYIN LI was born in Henan, China, in 1982. She received the bachelor's and master's degrees in landscape architecture from Huazhong Agricultural University, in 2004 and 2007, respectively. She visited the University of Social Sciences and Humanities, Poland, in 2018. She is currently a Lecturer with the Zhengzhou University of Aeronautics, Zhengzhou, China. Her current research interests include intelligence computing, machine learning, and landscape planning. In recent years, she has presided over and participated in many provincial and department level scientific research projects; and published more than ten peer-reviewed articles.

YANG LIU was born in Henan, China, in 1984. He received the bachelor's and Ph.D. degrees in computer science from Xi'an Jiaotong University, Xi'an, China, in 2004 and 2010, respectively. He visited the Department of Computer Science, Blekinge Institute of Technology, Sweden, from 2007 to 2009, under funding support of the China Scholarship Council. He is currently a Lecturer with Zhengzhou University, Zhengzhou, China, and a Visiting Scholar with The Jackson Laboratory for Genomic Medicine, Farmington, USA. His current research interests include deep learning, artificial intelligence, and DNA computing.