IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 41, No. 1, February, 2024, pp. 323-331
Journal homepage: http://iieta.org/journals/ts

Intelligent Recommendation System for Personalized Learning Resources for College Students Based on Image Processing

Yinle Zheng[1*], Jinfeng Miao[2], Sufang Ren[3]

[1] Department of Clothing Engineering, Hebei Vocational University of Technology and Engineering, Xingtai 054000, China
[2] Department of Economic Management, Hebei Vocational University of Technology and Engineering, Xingtai 054000, China
[3] Department of Electrical Engineering, Hebei Vocational University of Technology and Engineering, Xingtai 054000, China

Corresponding Author Email: zhengyinle@xpc.edu.cn

https://doi.org/10.18280/ts.410127

Received: 5 September 2023
Revised: 27 December 2023
Accepted: 8 January 2024
Available online: 29 February 2024

Keywords:
personalized learning resource recommendation, image semantic annotation, granular computing, second-order Conditional Random Field (CRF), product quantization sparse coding, image retrieval, intelligent recommendation system

ABSTRACT

With the rise of personalized learning, college students' demands for learning resources have become increasingly diversified. Traditional recommendation systems can no longer fully meet their needs for personalization and precision. Especially today, with an abundance of image resources, how to enhance the effectiveness of learning resource recommendation systems from a visual perspective has become a new challenge in the field of educational technology. This study proposes an intelligent recommendation system for personalized learning resources for college students, based on image processing. The system first implements semantic annotation of images that integrates contextual information through the granular computing concept and a second-order Conditional Random Field (CRF) model, improving the precision of annotations and the accuracy of semantic recognition. Secondly, the study explores an image retrieval method based on product quantization sparse coding, combined with edge feature descriptors and an optimized codebook, effectively enhancing the accuracy of learning resource retrieval and the relevance of recommendations. This research not only expands the application of image processing in the field of intelligent recommendation but also provides college students with more precise and personalized learning resource recommendation services.

1. INTRODUCTION

With the rapid development of information technology, personalized learning has become a focal point of interest in the field of education [1, 2]. College students, as the main force of social activity and innovation, urgently need a learning resource recommendation system that can provide customized services based on their specific needs [3]. Traditional recommendation systems rely heavily on text data processing, overlooking the rich resources of visual information. However, as an important medium for carrying information, the visual semantics contained in images have an undeniable value for accurately understanding and recommending learning resources [4-6]. Therefore, this paper aims to explore a new type of intelligent recommendation system for learning resources based on image semantic understanding, utilizing advanced image processing technology to meet the personalized learning needs of college students.

Personalized learning resource recommendation plays a significant role in promoting students' learning efficiency and quality [7]. Currently, with the integration of the internet and artificial intelligence technologies, as well as the advent of the big data era, personalized recommendation systems are increasingly applied in the educational field [8-10]. Image processing, as an important branch of intelligent recommendation systems, its research significance lies not only in technological innovation and frontier, but also in its practical impact on improving the match of learning resources and enhancing user experience [11, 12]. Utilizing image semantic understanding to assist in learning resource recommendation can significantly improve the accuracy and efficiency of the recommendation system, thus better serving the learning process of college students.

However, existing research still has some shortcomings in terms of image semantic annotation and learning resource recommendation [13, 14]. Although image recognition and processing technologies have become increasingly mature, how to efficiently integrate the contextual information of images to achieve precise semantic understanding and annotation remains a research challenge [14, 15]. Moreover, traditional image retrieval methods often suffer from limited feature representation when dealing with large-scale educational resources, resulting in low recommendation precision and recall rate [16, 17]. Therefore, developing a system capable of deeply mining image semantics and intelligently recommending based on the educational context is an urgent problem to be solved.

This paper's main research content revolves around two core parts: The first part is the learning resource image semantic annotation that integrates contextual information. By constructing multi-granularity context windows, semantic annotation can effectively capture the characteristics of local

areas and the symbiotic relationship between image semantic categories, thereby significantly improving the accuracy of image annotation. The second part is learning resource retrieval and recommendation based on image semantic understanding. A new method of product quantization sparse coding is proposed, which, by combining with edge feature descriptors, improves the feature vector representation of learning resource images. Through optimizing the construction of the codebook and quantization of residuals, the performance of the retrieval system is enhanced. In addition, this paper also delves into similarity calculation methods to achieve more accurate learning resource retrieval and recommendation. This research is not only significant for enhancing the performance of personalized learning resource recommendation systems but also provides a new research perspective and methodology for the application of image processing technology in the field of education.

2. INTEGRATION OF CONTEXTUAL INFORMATION IN SEMANTIC ANNOTATION OF LEARNING RESOURCE IMAGES

This study proposes an image annotation model that integrates multi-granularity contextual information, aiming to address the issue of traditional image annotation methods overlooking local details and the overall semantic relationships. This model can effectively identify and annotate the detailed features and contextual information within learning resource images, enhancing the granularity and accuracy of the annotations. By capturing the local transfer characteristics and co-existence relationships between semantic categories in images, the model can more accurately represent and understand image content, thereby providing college students with more relevant and precise learning resource recommendations. Specifically, the model adopts the concept of granular computing, through constructing multi-granularity contextual windows, it enables each granularity level to reflect the features of different areas and scales of the image, achieving comprehensive capture of image details and global information. Furthermore, it utilizes a second-order CRF model, considering not only the interactions between pixels but also the dependencies between labels, which further enhances the model's semantic understanding capability. Figure 1 shows a schematic diagram of the multi-granularity contextual windows in the model.

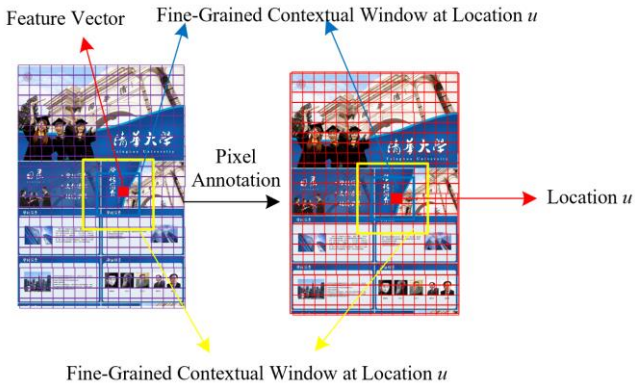

Figure 1. Multi-granularity contextual windows in the image annotation model integrating multi-granularity context

In the semantic annotation of learning resource images integrating contextual information, the main challenge is how to accurately understand and describe the complex dependencies between each pixel and its surrounding pixels, as well as the interactions between different semantic categories. Traditional image processing methods often fail to effectively integrate and utilize contextual information, leading to insufficient accuracy in semantic understanding. In the context of this study, the pixel-level semantic annotation task is defined as a high-precision image analysis process, aiming to assign accurate semantic labels to each pixel in the personalized learning resource recommendation system. This process maps visual features of images to corresponding semantic concepts through constructing a probabilistic graphical model, ensuring deep understanding and refined interpretation of image content. Suppose the observed data in the original image is represented by $b=\{b_u\}_{u\in T}$, where the visual feature vector extracted at location u is represented by $b_u=[b_{u1},b_{u2},\ldots,b_{uf}]$, and the set of all image locations is represented by $T=[1,2,\ldots,v]$. The total number of pixels in the image is represented by v, and the dimension of the feature vector is represented by f. The set of all semantic labels corresponding to the observed data is represented by M, and the number of label categories is represented by U. The probabilistic graphical model assigns a semantic label $a_u \in M$ to each location $u \in T$, and the annotation result is represented by $a=\{a_1,a_2,\ldots,a_{nv}\}$. Image annotation inference usually uses maximum posterior estimation to calculate the maximum probability of image annotation, represented by $\hat{a}=argmax_a O(a|b)$, then the posterior probability of image annotation results can be equivalent to:

$$O(a\,|\,b)\propto O(a,b)=O(a)O(b\,|\,a) \tag{1}$$

The research objective of this paper is to construct an advanced second-order CRF model that achieves mapping from image visual features to semantic labels by integrating different granularities of contextual information. Suppose the normalization function is represented by $C(b,\phi)=\sum a\Pi_{z\in Z}\Psi_z(a_z,b,\phi)$ the potential function on clique z is represented by φ_z, and the parameters of the potential function are represented by ϕ. If the posterior probability of the CRF model follows a Gibbs distribution, then there is:

$$O(a\,|\,b,\varphi)=\frac{1}{C(b,\varphi)}\prod_{z\in Z}\phi_z(a_z,b,\varphi) \tag{2}$$

In the semantic annotation of learning resource images integrating contextual information, a core issue is how to effectively capture and utilize contextual information at the pixel level to improve the accuracy of annotation. Typically, pixels in an image are related not only to their immediately adjacent pixels but also influenced by a broader neighborhood. A basic CRF model might only consider first-order neighborhood contextual information, limiting the model's ability to understand the complex structure of images. For learning resource recommendation systems, materials from different subjects may contain multi-level visual information and details, whose accurate semantic recognition requires the model to capture and integrate broader contextual information to accurately identify and recommend content highly relevant to students' learning needs. This paper chooses to establish an image annotation model that integrates a large amount of contextual information on the basis of a second-order CRF

model, because a second-order CRF model can consider a wider range of interactions between pixels, thus, in the analysis and inference process, it can more comprehensively consider the pixel relationships within a larger area. The integration of multi-granularity context also helps the model maintain performance when facing learning resources of different scales (such as detailed charts and macro conceptual diagrams).

The model consists of two types of potential functions. The unit location potential function primarily evaluates the likelihood of a single pixel receiving a specific semantic label. In the context of the learning resource recommendation system, this function can be designed to be sensitive to specific image features, thus more accurately identifying the core elements of learning materials. For example, in the image processing of mathematical resources, the unit location potential function would assign a higher probability to mathematical formulas or diagrams in the image, ensuring that semantic annotation is closely related to the content of the learning resources. To adapt to multi-class problems, this function can be extended in a binary expansion form, that is, by using a one-vs-all strategy, the multi-class classification problem is transformed into multiple binary classification problems, allowing each pixel to not only be classified into a single category but also independently judged on multiple binary problems, ultimately synthesizing these judgments to determine the final category of the pixel. Suppose the indicator function is represented by $\sigma(\cdot)$, if the semantic label of location u is $j(j \in M)$, then $\sigma(a_u=j)=1$; otherwise, $\sigma(a_u=j)=0$. Any specific domain classifier is represented by $o(a_u=j|b,\eta)$. The function expression is:

$$\prod_{u\in T}\phi_u\left(a_u,b,\eta\right)=\prod_{\substack{u\in T\\ j\in M}}\sigma\left(a_u=j\right)\log o\left(a_u=j\,|\,b,\eta\right) \tag{3}$$

When modeling $o(a_u=j|b,\eta)$, the logistic regression classifier can be extended to a *softmax* function. Suppose the parameters for the j-th semantic label are represented by $\eta=[\eta_{j1},\eta_{j2},\ldots,\eta_{jf}]$, a parameter vector of dimension $|M|$ is represented by η, consisting of $|M|$ semantic label parameters, that is, $\eta=\{\eta_j|j=1,2,\ldots,|M|\}$, then the expression is:

$$\begin{aligned} o\left(a_u=j\,|\,b,\eta\right)&=\frac{\exp\left(\eta_j^S b_u\right)}{\sum_{j=1}^{|M|}\exp\left(\eta_j^S b_u\right)}\\ &=\frac{\exp\left(\sum_f \eta_{jf}^S b_{uf}\right)}{\sum_{j=1}^{|M|}\exp\left(\sum_f \eta_{jf}^S b_{uf}\right)} \end{aligned} \tag{4}$$

The pairwise location potential function evaluates the likelihood of specific label combinations received by adjacent pixel pairs in the image, typically capturing multi-granularity contextual information to describe the interactions between pixels. In personalized learning resource recommendation systems, the pairwise location potential function can utilize contextual information to differentiate between various learning elements, such as distinguishing text areas from image areas. This means the system considers not just the relationships between pixels within a local neighborhood (fine-grained neighborhood) to maintain annotation continuity and consistency but also analyzes broader relationships between pixels (coarse-grained neighborhood) to capture more complex structures and patterns between pixel collections. Figure 2 shows the clique structures for expanded location pairs in fine-grained and coarse-grained neighborhoods.

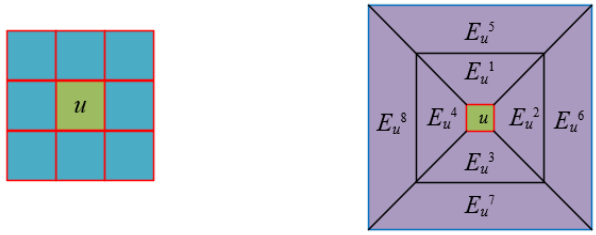

Figure 2. Clique structures for expanded location pairs in fine-grained and coarse-grained neighborhoods

(1) Fine-Grained Context

The pairwise location potential function constructed from fine-grained context focuses on analyzing and utilizing the subtle connections between adjacent pixels within the image to achieve local smoothness and continuity of semantic labels. The core of this approach is to ensure that semantic annotations remain consistent within the same object area of the image, reducing noise and incorrect annotations, while simplifying the computation process. By statistically analyzing the label continuity of adjacent pixels, the pairwise location potential function enhances the annotation accuracy of specific areas of the image, especially the edges of target objects, allowing the system to finely delineate image content, such as accurately distinguishing between images and text.

Suppose the smoothing parameter for fine-grained contextual information is represented by β. Any adjacent location of u in the fine-grained neighborhood $\lambda^1{}_u$ is represented by k, forming a location pair (u,k) within $\lambda^1{}_u$. Assuming the semantic annotation of u, a_u, is the j-th semantic label, and the semantic annotation of k is represented by a_k, if the semantic labels corresponding to the location pair (u,k) are different, i.e., $a_u \neq a_k$, then the smoothing parameter $\beta_{jm}=0$. The fine-grained contextual descriptor for (u,k) is represented by d_{uk}, and the following formula gives the calculation for the label transition probability from u to k:

$$\begin{aligned} &o\left(\left(a_u,a_k\right)\overset{\Delta}{=}j\,|\,d_{uk},\beta,k\in\lambda_u^1\right)\\ &=\begin{cases}\dfrac{\exp\left(\beta_{jj}^T d_{ukT}\right)}{1+\sum_{j=1}^{|M|}\exp\left(\beta_{jj}^T d_{ukT}\right)}, & \text{if } j\leq|M|\\ \dfrac{1}{1+\sum_{j=1}^{|M|}\exp\left(\beta_{jj}^T d_{ukT}\right)}, & \text{if } j=|M|+1\end{cases} \end{aligned} \tag{5}$$

For location pairs with similar visual features, the label transition probability calculated based on the above formula models the pairwise location potential function as:

$$\begin{aligned} &\prod_{u\in T,k\in\lambda_u^1}\phi_{uk}\left(a_u,a_k,d_{uk},\beta\right)=\\ &\prod_{\substack{u\in T,k\in\lambda_u^1\\ j\in M}}\log o\left(\left(a_u,a_k\right)\overset{\Delta}{=}j\,|\,d_{uk},\beta,k\in\lambda_u^1\right) \end{aligned} \tag{6}$$

(2) Coarse-Grained Context

The pairwise location potential function formed by coarse-grained context aims to capture and utilize the contextual information within larger areas of the image. This macro-level contextual analysis goes beyond the perspective of a single pixel, focusing on the distribution of different semantic categories within larger areas and their spatial relationships, thereby revealing the symbiotic relationships between different learning resources and their interactions in spatial layout. For example, the system might identify the layout patterns of textbooks with illustrations or infographics, where these macro features help the system more accurately identify and recommend learning materials that structurally match specific educational content. Specifically, in the coarse-grained neighborhood λ^2_u, the coexistence relationships between different semantic labels are considered. Suppose the coarse-grained neighborhood location of u is represented by E^o_u, abbreviated as p, and the semantic label coexistence parameter is represented by α, the coarse-grained contextual descriptor for the neighborhood location p is represented by g_{up}. g_{up} contains $|M|$ elements, i.e., $g_{up}=\{\omega^v_{up},v=1,2,...,|M|\}$, where the v-th element of g_{up} is represented by ω^v_{up}, indicating the maximum value of the likelihood mapping for the v-th semantic category when p implies semantic label $v(v\in M)$. Therefore, the pairwise potential function formed by u and p is:

$$\prod_{u\in T,k\in\lambda^1_u}\phi_{up}\left(a_u,a_p,d_{up},\alpha\right)=\prod_{\substack{u\in T,p\in\lambda^1_u\\ j\in M}}\alpha_{lv}\omega^v_{up}\sigma\left(l\neq v\right) \tag{7}$$

Combining the two types of potential functions defined above with Formula 2, and letting the model parameter set be represented by $\phi=\{\eta,\beta,\alpha\}$, we obtain the following pixel labeling model expression:

$$O(a\mid b,\varphi)=\frac{1}{C(b,\varphi)}\prod_{u\in T}\theta_u\left(a_u,b,\eta\right)\prod_{u\in T,k\in\lambda^1_u}\theta\left(a_u,a_k,d_{uk},\beta\right)\prod_{u\in T,p\in\lambda^2_u}\theta_{up}\left(a_u,a_p,d_{up},\alpha\right) \tag{8}$$

3. IMAGE SEMANTIC UNDERSTANDING-BASED LEARNING RESOURCE RETRIEVAL AND RECOMMENDATION

In the personalized learning resource recommendation system for college students, a core issue is how to accurately retrieve image resources that match students' learning interests and course content from a vast resource library. Given the complex and diverse content of learning resources, including diagrams, schematics, photographs, etc., traditional retrieval methods based on keywords or simple feature matching struggle to understand the deep semantic information of images, leading to inaccurate and irrelevant resource recommendations. Therefore, a method capable of deeply analyzing image semantic content and extracting features closely related to learning themes is needed to improve retrieval accuracy and recommendation quality. The method proposed in this paper, based on product quantization sparse coding, uses advanced edge feature descriptors to transform learning resource images into feature vectors. This not only improves the expressiveness of features but also significantly enhances discriminability through the use of smaller sub-codebooks constructed by the Cartesian product. These advantages make the recommendation system more efficient and precise in understanding and retrieving learning resource images, thus providing college students with more personalized and relevant learning materials. Figure 3 shows the principle of the method proposed in this paper.

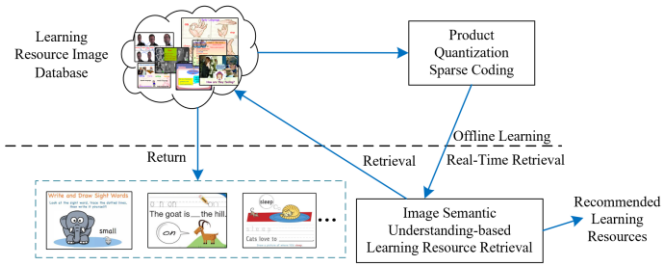

Figure 3. Principle of the image semantic understanding-based learning resource retrieval and recommendation method

3.1 Codebook construction using sparse coding method

The codebook construction using the product quantization sparse coding method aims to address the distinctiveness and robustness issues of feature representation. By combining smaller sub-codebooks through the Cartesian product, a large codebook is generated that can finely divide the feature space, enhancing the distinctiveness of feature representation. This method effectively deals with the high intrinsic dimensionality and visual diversity of learning resource images, making it easier for the system to differentiate between different image resources and improving the accuracy of retrieval. At the same time, this method avoids the problem of computational complexity growing exponentially with the size of the codebook, making the algorithm more suitable for retrieval from large-scale learning resource libraries. Figure 4 compares vector quantization and product quantization.

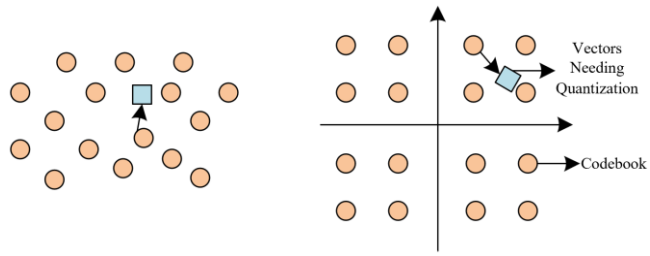

Figure 4. Comparison between vector quantization and product quantization

The coding process needs to be formalized first. In this step, each image is processed into a series of feature vectors, which need to be effectively encoded for subsequent retrieval tasks. Suppose a feature vector is represented by A, the codebook by Z, and the encoding of A by $B\in E^l$, with l_1 and l_2 norms represented by $\|.\|_1$ and $\|.\|_2$ respectively. The formalized representation of the encoding process is as follows:

$$\underset{B}{\operatorname{argmin}}\|A-ZB\|_2^2+\eta\|B\|_1\ s.t.B\geq 0 \tag{9}$$

In the feature space, each feature vector may be high-

dimensional, leading to computational burden if processed directly. Through the product quantization method, the feature vector is divided into L sub-vectors, transforming the original high-dimensional problem into L low-dimensional problems, thus reducing computational complexity. Each sub-vector can be seen as a segment of the original feature, and together they constitute the entire feature representation of the image. Suppose A is divided into L sub-vectors according to product quantization, represented by $A=[a_1,a_2,\ldots a_L]$, and the Cartesian product of A by $Z=Z_1*Z_2*...*Z_L$, with each sub-codebook represented by Z_u, then we have:

$$\begin{aligned}&\underset{B}{\operatorname{argmin}} \| A-ZB \|_2^2 +\eta \| B \|_1 \ s.t. B \geq 0, Z \\ &= Z_1 \times Z_2 \times ... \times Z_L\end{aligned} \tag{10}$$

After dividing the feature vector into sub-vectors, the coding problem is decomposed into L sub-problems. Each sub-problem corresponds to the encoding of a sub-vector and can be conducted independently. This decomposition not only simplifies the problem but also makes parallel processing possible, further enhancing the efficiency of the coding process. Moreover, each sub-problem can be individually optimized, ensuring the quality of encoding. The decomposed L sub-problems are given by the following:

$$\begin{aligned}&\underset{b_1}{\operatorname{argmin}} \| a_1 - Z_1 b_1 \|_2^2 +\eta \| b_1 \|_1 \ s.t. b_1 \geq 0 \\ &\vdots \\ &\underset{b_L}{\operatorname{argmin}} \| a_L - Z_L b_L \|_2^2 +\eta \| b_L \|_L \ s.t. b_L \geq 0\end{aligned} \tag{11}$$

The solution to each sub-problem can be equated to a least squares problem with quadratic constraints, as shown in the following formula. Specifically, it involves finding the optimal encoding coefficients under given constraints to minimize the reconstruction error.

$$\begin{aligned}&\underset{Z_1}{\operatorname{argmin}} \| a_1 - Z_1 b_1 \|_2^2 +\eta \| b_1 \|_1 \ s.t. \| Z_1 \| \leq 1 \\ &\vdots \\ &\underset{Z_L}{\operatorname{argmin}} \| a_L - Z_L b_L \|_2^2 +\eta \| b_L \|_1 \ s.t. \| Z_L \| \leq 1\end{aligned} \tag{12}$$

The final solution step involves the use of the Lagrangian multiplier method. By constructing the Lagrangian dual problem, the original problem can be transformed into a dual problem, which allows the recovery of the optimal solution of the original problem, i.e., the optimal encoding coefficients we need, through Lagrangian duality.

3.2 Feature representation

Sparse coding typically results in some loss of information, as it attempts to reconstruct the original signal with as few non-zero coefficients as possible. The reason for using residual construction to optimize feature representation in this paper is to reduce the inevitable information loss during the quantization step. By incorporating quantization residuals, the system can more accurately approximate the original image features, retaining more detail information, which is particularly important for understanding fine-grained image semantics. Suppose a sub-vector is represented by a_u, the following formula provides the residual expression:

$$r_u(a_u) a_u - Z_u b_u \tag{13}$$

To better represent a_u, using the expression of a_u under Z_u along with residual information, the final feature vector expression is:

$$b_u = \begin{bmatrix} b_u \\ r_u \end{bmatrix} \tag{14}$$

3.3 Similarity calculation

In similarity calculation, using a feature histogram method allows for a comprehensive assessment of the similarity between the query image and images in the database. This method considers multiple feature dimensions of the image, rather than simply relying on differences in a single feature or pixel, providing a more comprehensive reflection of the image's overall visual and semantic content. Through the comparison of feature histograms, the system can effectively identify learning resources that are closest to the query image at the feature level, thereby recommending them to the user. This method of similarity measurement based on the global distribution of features, compared to simple Euclidean distance or cosine similarity, can offer a more refined assessment of similarity, helping to improve the performance of the recommendation system. Suppose the feature histogram of the query learning resource image is represented by g, and the feature histogram of an image data is represented by g', the similarity between g and g' can be calculated using the following formula:

$$t(g,g') = \frac{\langle g, g' \rangle}{\| g \| \| g' \|} \tag{15}$$

The retrieval result is a collection of learning resource images sorted by similarity, ultimately leading to a real set of learning resource recommendations corresponding to this collection of images.

4. EXPERIMENTAL RESULTS AND ANALYSIS

The training method employed allows for the independent training of the model's parameter set $\phi=\{\eta,\beta,\alpha\}$. By adopting a model decomposition strategy, parameters can be updated independently, reducing dependencies between parameters and enabling parallel computation. This significantly enhances training efficiency and shortens training time. Experimental results, as described and supposedly depicted in Figure 5, demonstrate that the gradients of all three parameters converge within 100 iterations. This indicates that the model reaches a stable state after relatively few iterations, validating the convergence of the parallel-style segmented training approach. It can be concluded that by constructing multi-granularity context windows, the proposed model effectively captures details and relationships between semantic categories within learning resource images, thus improving the accuracy of image annotation. Since parameters can be trained independently, the entire model's training process can be parallelized, improving training efficiency and significantly

reducing training time.

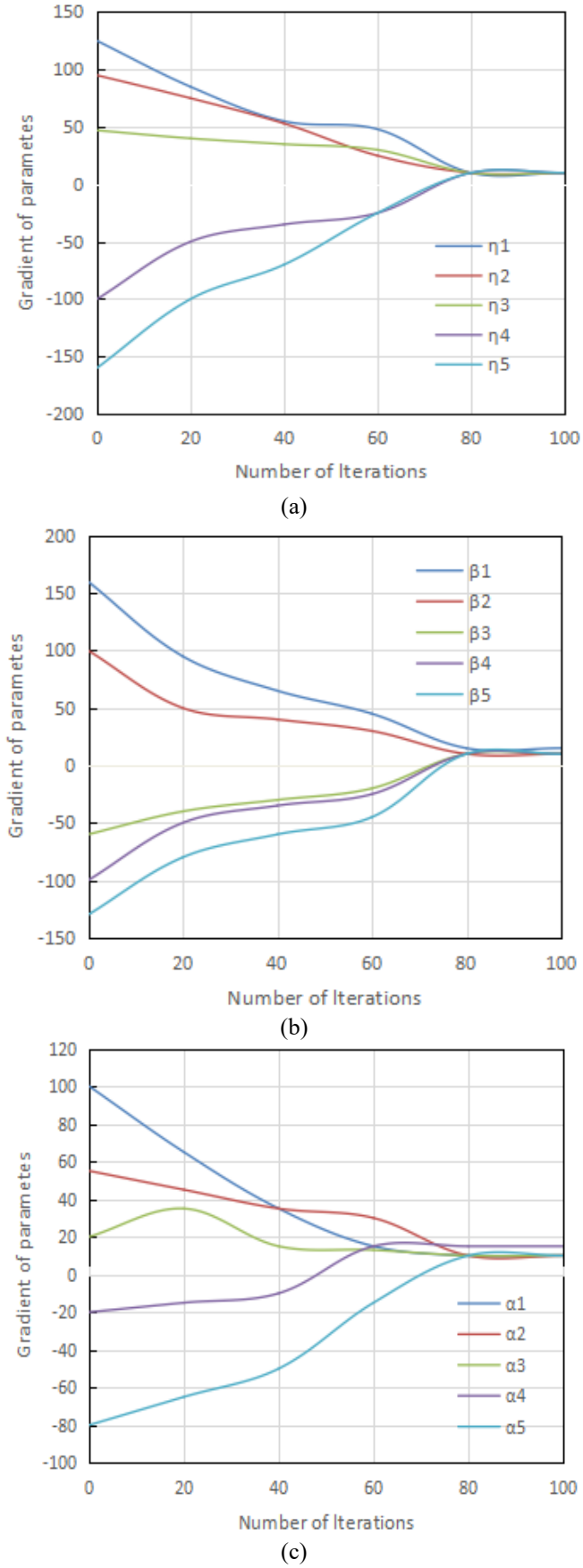

Figure 5. Changes in gradients of parameters in the learning resource image semantic annotation model with the number of iterations

Figure 6 shows the performance of five different image semantic annotation methods in terms of recall rate, including the proposed method, PSPNet, DGCNN, LSTM-CRF, and DeepLab. In the context of image semantic annotation, recall rate refers to the proportion of relevant image annotations correctly identified by the system. According to the description, the recall rate of the proposed method increases steadily from 75.0% to 89.0% from rank 0 to rank 25, showing stable improvement. Throughout the entire rank range, the recall rate of this method consistently leads the other four methods, especially showing significant growth in the latter half (from rank 15 to rank 25), indicating its effectiveness in handling large datasets. The recall rate performance of PSPNet grows from 75.0% to 87.0%, showing overall good performance but with both growth speed and final results lower than the proposed method. The recall rate increase for DGCNN is more gradual, from 67.0% to 86.5%, although it gradually improves, the gap with the proposed method widens. The performance of LSTM-CRF improves from rank 0 to rank 15 but then slows down, reaching only 84.0% at rank 25. The recall rate increase for DeepLab is the slowest, from 62.5% to 81.5%, showing the weakest performance among all methods. Based on the above analysis, it can be concluded that the proposed method is very effective in improving the recall rate of learning resource image semantic annotation after integrating contextual information. It not only shows a higher recall rate from the beginning but also maintains stable growth in recall rate as the dataset size increases. The proposed method continuously leads other methods at all rank points, indicating its clear advantage over existing technologies. The continuous growth trend in recall rate for the proposed method suggests that it may perform particularly well in identifying complex or ambiguously edged semantic labels, especially in scenarios that require recognizing a large amount of detail to improve recall rates.

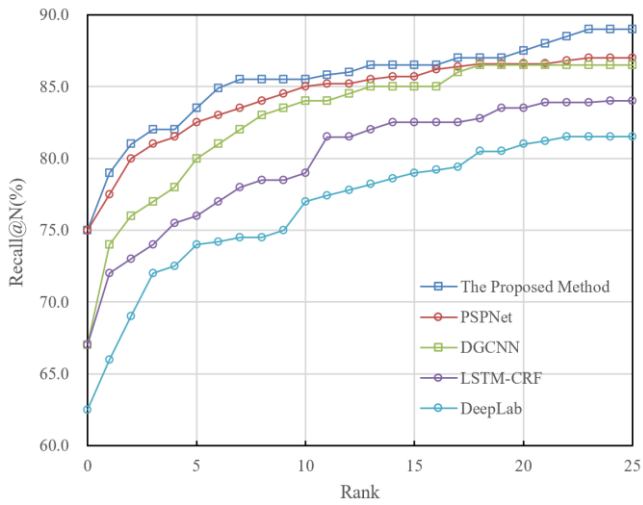

Figure 6. Line graph of recall rate in the learning resource image semantic annotation experiment

Table 1 showcases the performance of various image semantic annotation methods for learning resources across two evaluation metrics: METEOR and Intersection over Union (IoU). METEOR, a metric rooted in natural language processing, gauges the similarity between generated text and reference text. IoU is a computer vision metric assessing the overlap between predicted and actual object bounding boxes.

Table 1. Experimental results of semantic annotation for learning resource images

Source	*METEOR*			*IoU*		
	Online Course Content Image Set	Interactive Learning Interface Image Set	Learning Activity Record Image Set	Online Course Content Image Set	Interactive Learning Interface Image Set	Learning Activity Record Image Set
PSPNet	0.112	0.147	0.189	2.14	1.23	12.36
DGCNN	0.214	0.223	0.234	0.98	3.89	18.97
LSTM-CRF	0.218	0.227	0.268	4.56	3.14	23.45
DeepLab	0.227	0.235	0.289	5.23	5.24	25.36
The Proposed Method	0.254	0.245	0.312	5.46	5.87	26.87

From Table 1, it's evident that the proposed method achieves higher METEOR scores across all three image sets compared to other methods, indicating a higher similarity of its generated annotation texts with the reference texts. This significant improvement in METEOR scores underscores the proposed method's effectiveness in understanding image content and generating accurate annotations. On the online course content image set, the method leads with an IoU score of 5.46, indicating more precise prediction of image object bounding boxes. Similarly, it showcases the highest IoU scores on interactive learning interface images (5.87) and learning activity record images (26.87), further confirming its advantage in identifying key information like interactive elements. Overall, considering both METEOR and IoU metrics, the context-aware image semantic annotation method proposed in this paper surpasses comparative methods in annotation accuracy and target localization precision. Especially in handling diverse learning resource image sets, this method demonstrates high robustness and applicability, highlighting its effectiveness and potential value in practical applications.

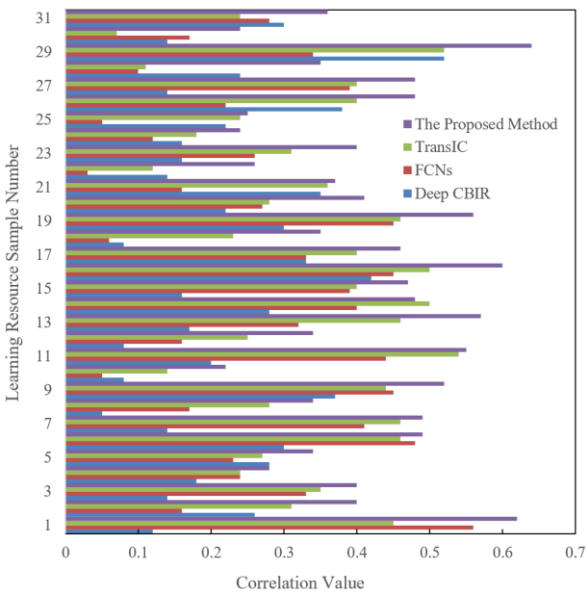

Figure 7. Kendall correlation values for different learning resource retrieval and recommendation methods

According to the data provided in Figure 7, it can be seen that the proposed method generally has higher Kendall correlation values across all 31 given learning resource sample numbers compared to Deep CBIR, FCNs, and TransIC methods. This means that the ranking by the proposed method in these samples has higher consistency with the true annotation ranking. Specifically, across all samples, the Kendall correlation coefficient of the proposed method never falls below 0.22, and in most samples, it exceeds 0.4, indicating that this method maintains high ranking consistency across different images and contexts. The Kendall correlation values for the Deep CBIR method fluctuate significantly, dropping to as low as 0.05, which may indicate its poor performance in some types of images or annotation tasks. Although it performs well in some samples, overall, its performance is not as stable as the proposed method. The performance of FCNs also shows variability, with correlation values ranging from 0.05 to 0.56. Although it excels in some samples, its performance on other samples is less satisfactory. The performance of TransIC is relatively stable but still generally lower than the proposed method. TransIC's performance exceeds 0.2 in most samples but rarely surpasses 0.5, indicating it has good consistency but may lack accuracy compared to the proposed method. Based on the above analysis, it can be concluded that the image semantic understanding-based learning resource retrieval and recommendation method proposed in this paper outperforms the other three compared methods in terms of performance, its generally higher Kendall correlation values suggest that the proposed method can maintain better ranking consistency across different scenarios, thus offering higher reliability. Such consistency and accuracy are crucial for learning resource retrieval and recommendation systems because they can ensure that users receive more relevant and accurate recommendation results.

Based on the data provided in Table 2, it is observable that different learning resource retrieval methods exhibit varying initial retrieval times and retrieval time costs when processing fine-grained features (such as texture details, edge directions, key point descriptors, and local colors) and coarse-grained features (such as regional shapes, color distributions, and scene categories). According to the table, the proposed method demonstrates significantly faster initial retrieval times for all features compared to other methods. This indicates the efficiency of the proposed method in processing preliminary image data, which is crucial for enhancing the overall system's response speed. When handling fine-grained features, the retrieval time cost of the proposed method is significantly lower than that of other methods in all instances. Notably, in processing key point descriptors, the retrieval time of the proposed method is only 5.64 seconds, in stark contrast to at least 12.48 seconds for TransIC and up to 98.41 seconds for FCNs. Similarly, in retrieving coarse-grained features, the proposed method also shows a significant speed advantage. For instance, in processing regional shapes, the retrieval time of the proposed method is merely 4.98 seconds, far below the fastest of the other methods, TransIC, which takes 25.47 seconds. Therefore, considering both initial retrieval and retrieval time costs, the image semantic understanding-based learning resource retrieval and recommendation method proposed in this paper excels not only in accuracy but also in

efficiency. This increase in efficiency is crucial for practical applications, especially regarding user experience and system performance. By optimizing feature vector representations and effectively handling quantization residuals, the proposed method provides a viable and robust solution for the rapid and accurate retrieval and recommendation of learning resources.

Table 2. Comparison of initial retrieval and retrieval time costs among different learning resource retrieval and recommendation methods

	Features	Initial Retrieval	*Deep CBIR*	*FCNs*	*TransIC*	Proposed Method
Fine-Grained	Texture Details	0.41*s*	31.24*s*	145.23*s*	23.36*s*	11.23*s*
	Edge Direction	0.31*s*	32.26*s*	147.23*s*	22.36*s*	11.54*s*
	Key point Descriptors	61.23*s*	18.69*s*	98.41*s*	12.48*s*	5.64*s*
	Local Color	0.34*s*	35.46*s*	65.47*s*	11.24*s*	8.91*s*
Corse-Grained	Regional Shape	0.34*s*	83.21*s*	105.65*s*	25.47*s*	4.98*s*
	Color Distribution	3.12*s*	26.31*s*	104.25*s*	22.17*s*	3.91*s*
	Scene Category	3.12*s*	26.35*s*	10.2.41*s*	38.47*s*	2.35*s*

5. CONCLUSION

This paper introduces a multi-granularity contextual window method that integrates contextual information to enhance the accuracy of semantic annotation. By constructing multi-granularity contextual windows, the study successfully captures the characteristics of local regions in images and the symbiotic relationships between semantic categories. This approach significantly improves the precision of image annotations, laying a solid foundation for subsequent retrieval and recommendation tasks. A novel product quantization sparse coding method is proposed to enhance the feature vector representation of learning resource images. Combined with edge feature descriptors, the feature representation has been improved, and retrieval performance has been enhanced through optimized codebook construction and quantization residual processing. This method not only strengthens the accuracy of the retrieval system but also significantly increases retrieval speed.

The experimental results, depicted through the change in model parameter gradients over iterations, demonstrate the stable learning process of the proposed semantic annotation model. Recall rate graphs highlight the method's significant impact on image semantic annotation tasks, indicating its superior performance. Comparisons of Kendall correlation values among different retrieval methods show that the proposed method aligns more closely with user expectations in semantic-level retrieval tasks. Initial retrieval and retrieval time cost comparisons underscore the method's remarkable efficiency, especially in processing complex image features.

The research and experimental outcomes of this study collectively validate the effectiveness and efficiency of the proposed method. By integrating multi-granularity contextual information for semantic annotation, this paper not only improves annotation accuracy but also significantly enhances the efficiency of learning resource retrieval and recommendation through product quantization sparse coding and optimized feature vector representation. These improvements position the proposed method as highly valuable for practical applications, particularly in educational technology systems requiring rapid processing and high-quality recommendations. Overall, this work presents innovative academic methodologies and provides practical technical solutions for the management and distribution of learning resources. Future work could explore the model's generalizability, real-time processing capabilities, and its application effectiveness across different types of learning resources.

REFERENCES

[1] Liu, Q. (2023). Personalized learning resources recommendation for interest-oriented teaching. International Journal of Emerging Technologies in Learning, 18(6): 146-161. https://doi.org/10.3991/ijet.v18i06.38721

[2] Zhang, S., Diao, J. (2023). Personalized recommendation method of online education resources for tourism majors based on machine learning. In International Conference on E-Learning, E-Education, and Online Training, Yantai, China, pp. 222-235.

[3] Wu, W. (2023). Research on personalized recommendation method of preschool e-learning resources. Applied Mathematics and Nonlinear Sciences, 9(1): 1-18. https://doi.org/10.2478/amns.2023.2.00174

[4] Tsai, C.Z., Huang, H., Wei, C.J., Chiu, M.C. (2023). Apply deep learning to build a personalized attraction recommendation system in a smart product service system. In Advances in Transdisciplinary Engineering, 41: 151-160. https://doi.org/10.3233/ATDE230607

[5] Liu, F., Guo, W. (2022). Personalized recommendation algorithm for interactive medical image using deep learning. Mathematical Problems in Engineering, 2022: 2876481. https://doi.org/10.1155/2022/2876481

[6] Zhang, Q., Liu, Y., Liu, L., Lu, S., Feng, Y., Yu, X. (2021). Location identification and personalized recommendation of tourist attractions based on image processing. Traitement du Signal, 38(1): 197-205. https://doi.org/10.18280/ts.380121

[7] Xu, Y., Chen, T.E. (2023). The design of personalized learning resource recommendation system for ideological and political courses. International Journal of Reliability, Quality and Safety Engineering, 30(1): 2250020. https://doi.org/10.1142/S0218539322500206

[8] Du, H., Palaoag, T., Guo, T. (2023). Personalized learning resource recommendation framework based on knowledge map. In Proceedings of the 2023 International Conference on Advances in Artificial Intelligence and Applications, Wuhan, China, pp. 132-136. https://doi.org/10.1145/3603273.3634709

[9] Sun, P. (2023). Personalized course resource recommendation algorithm based on deep learning in the intelligent question answering robot environment. International Journal of Information Technologies and Systems Approach (IJITSA), 16(3): 1-13. https://doi.org/10.4018/IJITSA.320188

[10] Karthikeya, A., Kumar, A.A., Manicharan, C., Hariharan,

S., Kukreja, V., Prasad, A.B. (2023). Age based hybrid recommendation system using machine learning. In 2023 3rd International Conference on Innovative Sustainable Computational Technologies (CISCT), Dehradun, India, pp. 1-5. https://doi.org/10.1109/CISCT57197.2023.10351208

[11] Xiong, Z.W., Yu, H., Shen, Z.Q. (2023). Federated learning for personalized image aesthetics assessment. In 2023 IEEE International Conference on Multimedia and Expo (ICME), Brisbane, Australia, pp. 336-341. https://doi.org/10.1109/ICME55011.2023.00065

[12] Dubey, N., Verma, A.A., Setia, S., Iyengar, S.R.S. (2022). PerSummRe: Gaze-based personalized summary recommendation tool for Wikipedia. Journal of Cases on Information Technology (JCIT), 24(3): 1-18. https://doi.org/10.4018/JCIT.20220701.oa7

[13] Nitin Hariharan, S.S., Deepak, G., Ortiz-Rodríguez, F., Panchal, R. (2023). ITAQ: Image tag recommendation framework for aquatic species integrating semantic intelligence via knowledge graphs. In Iberoamerican Knowledge Graphs and Semantic Web Conference, pp. 135-150. https://doi.org/10.1007/978-3-031-47745-4_11

[14] Jian, M., Guo, J., Zhang, C., Jia, T., Wu, L., Yang, X., Huo, L. (2021). Semantic manifold modularization-based ranking for image recommendation. Pattern Recognition, 120: 108100. https://doi.org/10.1016/j.patcog.2021.108100

[15] Guo, G., Meng, Y., Zhang, Y., Han, C., Li, Y. (2019). Visual semantic image recommendation. IEEE Access, 7: 33424-33433. https://doi.org/10.1109/ACCESS.2019.2900396

[16] Hyun, C., Park, H. (2019). Image recommendation for automatic report generation using semantic similarity. In 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Okinawa, Japan, pp. 259-262. https://doi.org/10.1109/ICAIIC.2019.8669018

[17] Hur, C., Hyun, C., Park, H. (2020). Automatic image recommendation for economic topics using visual and semantic information. In 2020 IEEE 14th International Conference on Semantic Computing (ICSC), San Diego, CA, USA, pp. 182-184. https://doi.org/10.1109/ICSC.2020.00037