Eng Int Syst (2025) 3: 329–338

Engineering
Intelligent Systems

Performance Improvement Strategy for Few-shot Semantic Segmentation Assisted by Large Language Models

Xue Han[1,a], Shuang Zhang[2,b*] and Yu Li[3,c]

[1] *School of Foreign Language, Zhuhai College of Science and Technology, Zhuhai 519040, Guangdong, China*
[2] *Foreign Language Department, Zhuhai College of Science and Technology, Zhuhai,Guangdong, 519040,China*
[3] *School of Computer Science, Zhuhai College of Science and Technology, Zhuhai 519040, Guangdong, China*

Few-shot semantic segmentation is a computer vision technology that is used to segment pixel-level objects in images segment a small number into labeled samples. Traditional methods have limited generalization capability, are sensitive to background interference, susceptible to class imbalance, and have inadequate feature representation. This study explored the use of Large Language Models (LLMs) for few-shot semantic segmentation tasks. Using ChatGLM (Chat Generative Language Model), with its powerful semantic understanding and feature extraction, Named Entity Recognition (NER) with image context was achieved. The NER results were used to enhance few-shot semantic segmentation, improving its ability to process and understand image semantics. Experimental results showed significant improvements in the mean Intersection over Union (mIoU) evaluation metric using a few-shot semantic segmentation model assisted by LLMs. The small-sample semantic segmentation model assisted by the large language model has a higher mIoU ratio scores on 5 validation sets than the SETR (Segmentation Transformer) model and Mask R-CNN (Mask Region-based Convolutional Neural Network), indicating that the application of the large language model effectively improves the accuracy and generalization ability of small-sample semantic segmentation.

Keywords: few-shot semantic segmentation, large language models, named entity recognition, mean intersection over union, feature extraction

1. INTRODUCTION

Semantic segmentation is a key technology in computer vision, which categorizes each pixel in an image to achieve a detailed understanding of the image content [1–2]. With the continuous development of society, the application of artificial intelligence and deep learning technology has become increasingly widespread, promoting the development of fields such as autonomous driving, medical image analysis, and smart city monitoring. The increasing demand for precise analysis of image and video data in these fields has led to an increasing number of applications for semantic segmentation. The few-shot semantic segmentation method is applied to segment different categories in an image at the pixel level using a very small amount of annotated data [3–4]. Traditional deep learning models typically require a large amount of annotated data for training in order to learn sufficiently rich and generalized features. However, obtaining high-quality annotated data is very expensive and time-consuming, and the training of deep neural networks, especially complex models for semantic segmentation, requires a large amount of computing resources, including high-performance GPUs (Graphics Processing Units) and long training cycles [5–6].

*Corresponding author. Email: [a]22412285@qq.com, [b]zhangshuang@zcst.edu.cn, [c]jluzhliyu@zcst.edu.cn

The dependence on data augmentation and preprocessing is also a major issue. To improve the robustness of the model, a large number of data augmentation and preprocessing operations are required, such as rotation, scaling, cropping, etc. Although these operations help improve model performance, they also increase the complexity of data processing.

The large language model (LLM) is an artificial intelligence model constructed based on deep learning technology, whose core capability lies in processing and generating natural language text [7–8]. This model, through pre-training on large-scale datasets, can deeply understand the syntax structure, semantic relationships, and contextual information of language, thus demonstrating broad application potential in various fields [9–10]. In terms of natural language understanding, large language models can accurately understand and infer the meaning behind text, capturing the deep semantics and logical relationships of language, not just the superficial literal meaning. In generative tasks, they are able to produce coherent and logically consistent text based on input conditions, demonstrating excellent performance in applications such as automatic summarization, dialogue generation, and text translation. The application of large language models (LLMs) in few-shot semantic segmentation tasks is of great significance. An LLM helps to expand and enrich training datasets by generating a large number of semantically rich descriptions [11–12]. Its powerful feature extraction and generalization capabilities can significantly improve the learning effect of the model on a small amount of annotated data, thereby achieving more accurate and reliable pixel-level object segmentation [13–14]. In addition, through multimodal learning, the LLM integrates text and image information, improving the understanding and analysis of complex scenes. By considering both text and image content simultaneously, it is possible to understand the scene context more thoroughly, and accurately label object boundaries and attributes in semantic segmentation tasks.

In summary, this study explores strategies for enhancing few-shot semantic segmentation using LLMs. The semantic understanding and feature extraction capabilities of LLMs are used to transform complex textual information into valuable features. The LLM improves model performance by accurately identifying and locating key entities in images. The experimental results show that the mIoU scores are 74.32%, 79.01%, 76.17%, 77.36%, and 72.05%, respectively, which are significantly higher than those of the few-shot semantic segmentation model without the assistance of LLMs. This indicates that the assistance of LLMs enables few-shot semantic segmentation models to process image semantic information more accurately and effectively, improving the overall performance and generalization ability of the model, making significant contributions to multi-domain applications and research, and promoting the application and development of intelligent systems for complex tasks.

2. RELATED WORK

With the increasing demand for efficient and accurate image analysis in the field of computer vision, semantic segmentation has become an important research direction aimed at achieving precise segmentation of new categories of images through annotated data. In order to address the challenges faced by Vision Transformer (ViT) in semantic segmentation tasks, There are scholars [15] proposed methods to explore the application and evolution of different ViT architectures in semantic segmentation. The research found that although ViT has achieved success in image classification, its image segmentation-oriented approach is not suitable for dense prediction tasks such as image segmentation and object detection. Some researchers [16] systematically reviewed and summarized deep learning-based image semantic segmentation methods and their current status and progress in the field. They detailed the representative algorithms, basic ideas, and advantages and disadvantages of each type of method, and emphasized the important contribution of deep learning in the field of image semantic segmentation. Some researchers [17] proposed a simple, efficient, and powerful semantic segmentation framework to address efficiency and performance issues in the field of semantic segmentation, combining Transformer with a lightweight multilayer perceptron (MLP) decoder. Better performance and efficiency were demonstrated on datasets such as ADE20K and Cityscapes, but the ability to generalize to specific datasets or scenarios was not extensively tested. Other researchers [18] found that current semantic segmentation models had a large number of parameters and were not suitable for mobile devices, while other models with small memory usage followed the idea of classification networks and ignored the inherent characteristics of semantic segmentation. To address this issue, a novel lightweight and efficient network called Context Guided Network (CGNet) was proposed for semantic segmentation. Experimental results showed that under the same number of parameters, the proposed CGNet significantly outperformed existing lightweight segmentation networks. However, the generalizability of the model on other datasets or tasks needs further evaluation and improvement.

LLMs have played an important role in the current social context, achieving rapid application and customized development in various fields through pre-training models and transfer learning, and bringing new application prospects to industries such as medical diagnosis, financial analysis, and educational intelligence. Some researchers [19] analyzed how LLMs can respond to free text queries without specific training, discussed the advantages and limitations of LLMs in clinical practice, and explored the potential for improving efficiency and effectiveness in the medical field. Some researchers [20] demonstrated that the LLM based on Transformer has made significant progress in various fields, especially in the application of natural language processing, biology, chemistry and computer programming, and analyzed the potential and effect of large language models combined with Internet and document search, code execution, experimental automation and other tools to accelerate research. The research results showed that artificial intelligence systems demonstrated multifunctionality, effectiveness, and interpretability in driving research, highlighting the potential of artificial intelligence systems in accelerating the research process. Some researchers [21] investigated

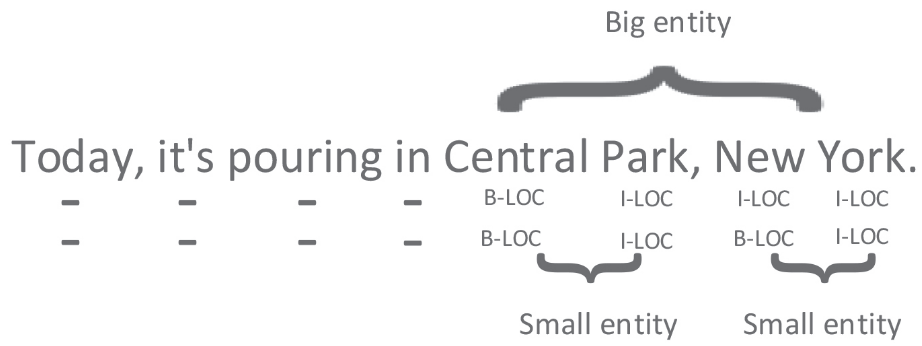

Figure 1 Flat named entities and hierarchical named entities.

whether large-scale language models can accurately infer the psychological tendencies of social media users, and whether different social population groups have different inference abilities. GPT-3.5 and GPT-4 were tested to derive Big Five personality traits from Facebook status updates. Although the results showed that their accuracy was similar to that of specially trained supervised machine learning models, their predictions were more accurate in the female and young population groups. Other researchers [22] systematically reviewed the specific applications of large language models represented by ChatGPT in multiple clinical scenarios through literature analysis, including clinical consultations, medical history collection, text writing, decision support, personalized medicine, doctor-patient communication, patient psychological support, academic research, medical education, and hospital management.

3. ENTITY RECOGNITION OF LARGE LANGUAGE MODELS IN ENGLISH TEACHING

In the context of college English teaching, few-shot semantic segmentation technology demonstrates multiple advantages. It can help students deeply understand and analyze the illustrations in literary works, accurately identify key scenes, characters, or symbols through fine image segmentation, and thus explore their symbolic meanings and thematic development in depth. Few-shot semantic segmentation technology can help students intuitively understand the context and meaning of different cultural backgrounds, promoting cross-cultural understanding. By applying modern technology, teaching becomes more attractive and interactive, not only enhancing learning outcomes, but also improving students' processing of information and their multicultural communication skills. Entity recognition can help segmentation models to capture and separate key objects in images more precisely by accurately identifying and locating specific entities in text, especially in situations where data is scarce, thereby improving segmentation accuracy and the capturing of details. Entity annotation not only enhances the model's understanding of semantic information, but also provides key contextual information, which helps the model better adapt to and interpret various elements and backgrounds in complex scenarios.

3.1 Flat Named Entity Recognition

Named Entity Recognition (NER) tasks can be divided into two main types: flat naming and hierarchical naming [23–24]. In Flat Named Entity Recognition (Flat NER), each entity is treated as an independent label and there is no hierarchical relationship between entities. The system assigns a separate label to each entity, regardless of its relationship or hierarchy [25–26]. Hierarchical NER (Hierarchical Named Entity Recognition) considers the hierarchical relationships between entities. The system not only recognizes individual entities, but also identifies the containment or hierarchical relationships between entities [27–28]. The difference between flat naming and hierarchical naming is shown in Fig. 1. Here, B-LOC denotes the location name's beginning and I-LOC denotes the location name's inner part.

The flat NER model is simpler and more direct than the hierarchical NER model, as it does not need to consider the hierarchical relationships between entities; it only needs to assign an independent label to each entity [29–30]. Moreover, flat naming does not involve complex hierarchical structures and relationships, making the design, implementation, and training process of flat NER models relatively simpler and more efficient [31–32]. Flat NER is suitable for many common entity recognition tasks, such as personal names, place names, organization names, etc., and is more suitable for college English teaching. Therefore, in this study, flat naming was chosen for the entity recognition of the large language model in English teaching.

The BIO (Begin, Inside, Outside) annotation approach is commonly utilized in flat named entity recognition datasets. The position of the entity's first word or character in the text is denoted by B, which stands for the beginning of the entity [33]. I represents the internal part of the entity, that is, the position of the words or characters in the middle of the entity in the text. O represents the non-entity part, that is, the position of words or characters that do not belong to any entity in the text. Assuming a text is T, containing n words, $T = \{t1, t2, t3 \dots tn\}$, all entities in the text T need to be identified and classified into the pre-set possible types Y, as shown in Fig. 2. Among them, PER (Person) represents a person's name; ORG (Organization) represents an organization's name; LOC (Location) represents a place name.

The original label-based NER dataset contains named entity information for both text and annotations. The dataset is

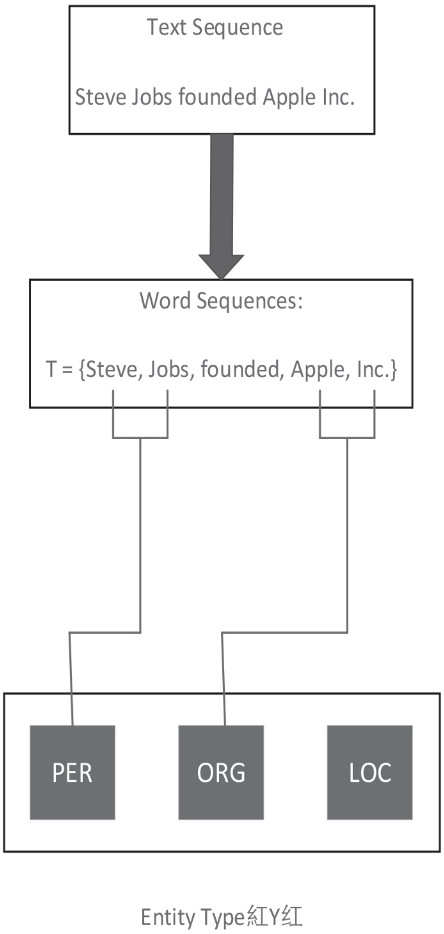

Figure 2 Schematic diagram of named entity recognition.

Table 1 Building triplets.

Data	Entity type	Q	A
Steve Jobs founded Apple Inc. in Cupertino.	Bureau	What organizations are included in the data	Apple Inc
	Geographic location	What locations are included in the data	Cupertino
	Character	What characters are included in the data	Steve Jobs
	Event	What events are included in the data	No such entity

transformed into a triplet containing data, questions, and answers, with each category of data. A question Q is constructed, where $Q = \{q1, q2, q3, \ldots qm\}$. m represents the number of categories. The annotated entity list is A, where $A = \{a1, a2, a3, \ldots am\}$, and each am represents an entity in the data corresponding to the question qm. If there is no corresponding entity in the data, "no entity of this category" is output. The extraction of entity information from data is achieved by constructing a system that includes data, qm, and am, as shown in Table 1.

Table 1 shows the triplet constructed for the original data "Steve Jobs founded Apple Inc. in Cupertino." For entity types (Bureau, Geographic location, Character) that appear in the data, corresponding answers (Apple Inc, Cupertino, Steve Jobs) are output based on the question. For entities that do not appear in the data, "No such entity" is output.

3.2 Fine-tuning of the Model

In NER tasks, when constructing triplets, the Q is crucial, equivalent to the prompt in natural language processing (NLP) tasks. Using different prompts for the same task can yield different results, so it is necessary to build a Prompt class to facilitate its modification.

Pre-trained models can be adapted effectively to new tasks by fine-tuning them on specific application tasks, as large language models' scale and performance increase. However, if the weights of the pre-trained model are changed during the fine-tuning process, this may decrease in the universality of the model. This is because the weights of the pre-trained model have already learned a wide range of knowledge and features when training on large-scale data, and changing these weights may cause the model to lose these characteristics, resulting in

Table 2 Entity recognition prompt table.

Task Description	Example Display	Test Sentence Input	Answer
Description: Analyze the text below and identify the entity type mentioned in the text.	Text: Shakespeare's works have been studied extensively in academia.	Test Sentence: Who is studied extensively in academia?	Answer: Shakespeare
Description: Read the sentence below and identify the name of the organization mentioned.	Text: The University of Oxford is renowned for its academic excellence.	Test Sentence: What university is renowned for academic excellence?	Answer: The University of Oxford
Description: Find the information below to understand the location or place mentioned in the text.	Text: London is a vibrant city with a rich cultural heritage.	Test Sentence: What city has a rich cultural heritage?	Answer: London
Description: Explain the date below and determine when the event occurred.	Text: The Declaration of Independence was signed on July 4, 1776.	Test Sentence: When was the Declaration of Independence signed?	Answer: July 4, 1776

poor performance on other tasks or new data. LoRA (Low-Rank Adaptation) is a technology used for fine-tuning pre-trained models by adding low-rank matrices between the linear layers of the model to adapt to downstream tasks, rather than changing the original weights. Therefore, in this study, LoRA is used for supervised learning. Assuming the initial weight of the pre-trained model $P_\Phi(y|x)$ is Φ, the NER task after converting the raw data into triplets is a series of dialogues: $D = \{(xi, yi)\}, i = 1, \ldots, N$. Here, xi is the Prompt and training data, and yi is the named entity in the data. By training the data, the initial weight Φ_0 of the pre-trained model is upgraded to $\Phi_0+\Delta\Phi$, to maximize the conditional language modeling objective, as detailed in Formula (1):

$$max_\Phi \sum_{(x,y)\in Z} \sum_{t=1}^{|y|} log\,(P_\Phi\,(y_t|x,\,y<t)) \tag{1}$$

During the global fine-tuning process, all parameters of the pre-trained model are adjusted, and the newly-learned parameter increment $\Delta\Phi$ and the original pre-trained model parameter Φ_0 together form the final model. A large number of parameters need to be adjusted and stored, which greatly increases the demand for computing and storage resources. In contrast, by using LoRA, only a small number of parameters Θ need to be adjusted to achieve the same effect as global fine-tuning, that is, $\Delta\Phi=\Delta\Phi\,(\Theta)$, where $|\Theta| \ll |\Phi_0|$. Therefore, the optimization objective is converted to Θ, as shown in Formula (2):

$$max_\Theta \sum_{(x,y)\in Z} \sum_{t=1}^{|y|} log\left(P_{\Phi_0+\Delta\Phi(\Theta)}\,(y_t|x,\,y<t)\right) \tag{2}$$

Assuming the length of the input vector z is l, the weight W of the model is adjusted and optimized through the training data of downstream tasks, as shown in Formula (3).

$$h = W_0 z + \Delta W z = W_0 z + KFz \tag{3}$$

In Formula (3), $W_0 z$ represents the original calculation result of the model, which is adapted to the new task through $\Delta W z$; W_0 is the weight of the pre-trained model; ΔW is the weight increment learned during the fine-tuning process; KF is two low dimensional matrices, $K \in R^{d\times r}$, $F \in R^{r\times d}$, and r is much smaller than d.

3.3 Contextual Learning of Large Language Models

Contextual learning of large language models refers to the ability of models to learn and understand information based on context when processing text or other types of data [34–35]. In natural language processing, contextual learning enables models to understand and generate more accurate and contextually-relevant text. The key to contextual learning is the ability of the model to capture long-term dependencies and contextual information in the text. Contextual learning relies on finding training data that is similar to the predicted data and providing it as a reference to the model. The advantage of this method is that it does not directly modify the structure or parameters of the model itself, but adapts to specific tasks or data by utilizing the potential of existing models. Taking the scenario of college English teaching as an example, a prompt is constructed as displayed in Table 2.

The entire prompt is divided into four parts: Task Description, Example Display, Test Sentence Input, and Answer. In the Task Description, different words can be used to achieve better results. The quality of cases directly affects outcomes [36–37]. This study is based on the ChatGLM large language model, with college English teaching as the background, and establishes a large language model based on training data. Using the K-Nearest Neighbors algorithm, the distance between different features is measured to determine the category to which a data point belongs, and the K closest cases are selected. The accuracy of model recognition is determined by comparing the actual answers with the model answers.

3.4 Entity Recognition Experiment of Large Language Model

The experimental dataset for entity recognition of the large language model in this study includes three sets of data, as shown in Table 3. SemEval 2010 Task 7 contains approximately 3000 Twitter messages of 6 entity types, including PER, LOC, Organizational ORG, Miscellaneous

Table 3 Large language model entity recognition dataset.

Network Error	**Scale (documents/ sentences)**	**Number of entity types**	**Number of entity instances (k)**	**Main entity types**
SemEval 2010 Task 7	3,000 Twitter messages	6	8	PER, LOC, ORG, Misc, Hashtag, URL
SciERC	500 papers	6	20	Author, Periodical, Task, Method, Material, Others
TAC KBP 2010 English Pilot Data	300 documents	5	10	PER, LOC, ORG, GPE, Misc

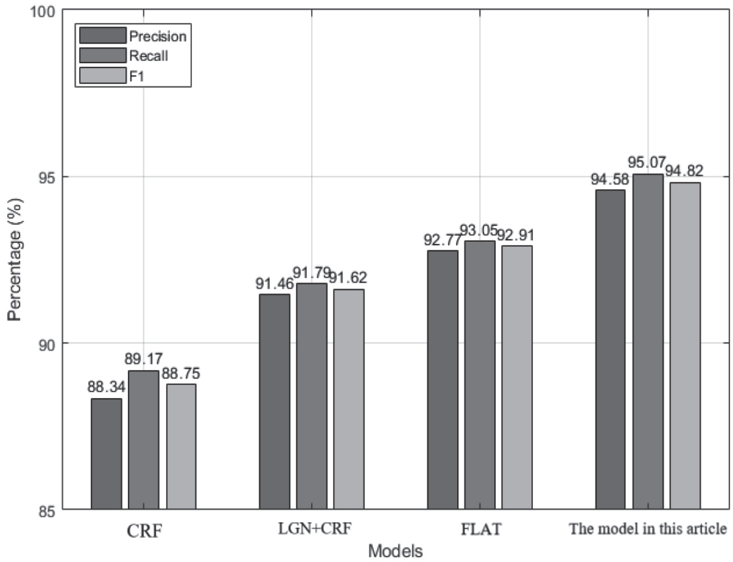

Figure 3 SemEval 2010 Task 7 dataset results.

(Misc), Hashtag, and URL (Uniform Resource Locator). The SciERC dataset contains approximately 20000 entity instances, distributed across approximately 500 scientific literature items, covering six entity types: Author, Periodical, Task, Method, Material, and Others. The entity instances of TAC KBP 2010 English Pilot Data are approximately 10000, consisting of 300 news documents and 5 entity types: PER, LOC, ORG, Geographical Political Entities (GPE), and Misc.

To determine the effectiveness of the model in named entity recognition tasks, the precision, recall, and F1 value of the model recognition is calculated for each dataset, and performance is compared with that of other types of models. We compared CRF (Conditional Random Fields), LGN (lexicon-based graph neural network)+CRF, FLAT (flat-lattice Transformer) and the model used in this paper. The performance on the SemEval 2010 Task 7 dataset is shown in Fig. 3.

Fig. 3 denotes that the model has a precision of 94.58%, a recall of 95.07%, and an F1 value of 94.82% on SemEval 2010 Task 7 (a complex dataset containing social media text, spelling errors, abbreviations, and slang), all exceeding 94%, indicating that the model has both high accuracy and comprehensiveness when identifying specific entities.

The performance of the four models on the SciERC dataset is illustrated in Fig. 4.

Fig. 4 shows that the proposed model's various indicators on the SciERC dataset exceed 93%, surpassing the performance of the comparison model, demonstrating its deep understanding of scientific language and entities, and indicating its adaptability to the scientific research field. It is expected to enhance the effectiveness of automatic summarization, information retrieval, and knowledge graph construction, and improve the efficiency and quality of scientific research.

The performance of the four models on the TAC KBP 2010 English Pilot Data dataset is shown in Figure 5.

Fig. 5 shows that the precision, recall and F1 value obtained by the proposed model for the TAC KBP 2010 English Pilot Data (with complex sentence structures and diverse entities) are all over 93%, surpassing the other three models including CRF, thereby demonstrating a more efficient performance.

4. FEW-SHOT SEMANTIC SEGMENTATION MODEL ASSISTED BY LARGE LANGUAGE MODELS

4.1 Building Few-shot Semantic Segmentation Models

The large language model can extract rich semantic information contained in text by means of entity recognition. These pieces of information can be used to annotate relevant objects in images, and through entity recognition, large

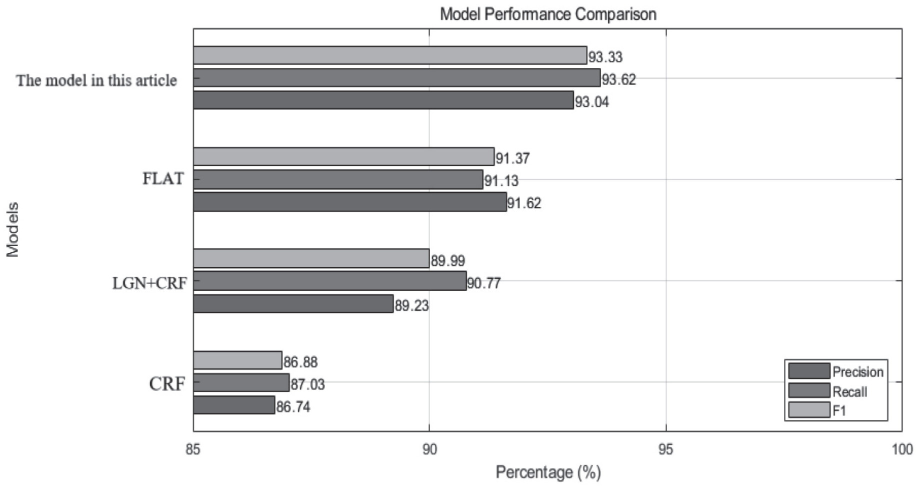

Figure 4 Experimental findings of the model on the SciERC dataset.

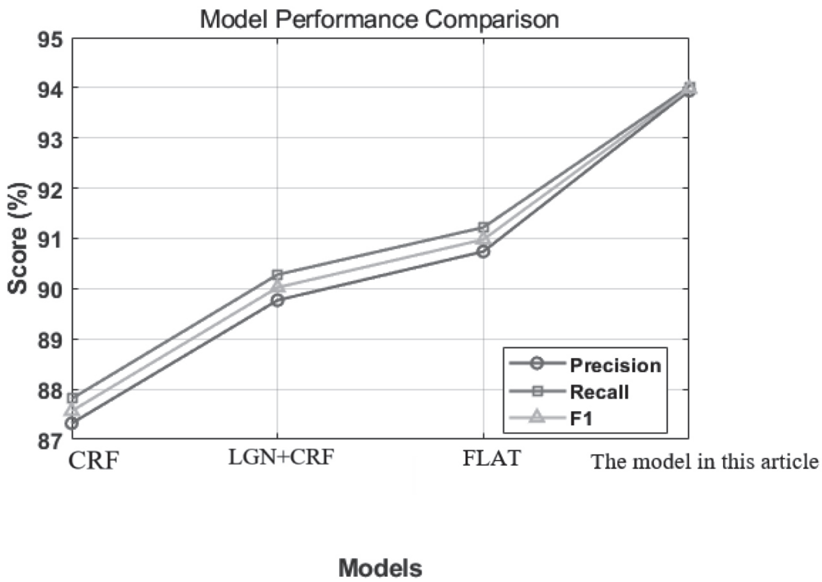

Figure 5 Experimental results of TAC KBP 2010 English Pilot Data dataset.

language models can transfer semantic information from the textual domain to the visual domain. Entity recognition provides more accurate feature representations for few-shot semantic segmentation by identifying the semantic categories of various objects in the image. In order to provide more semantic information for few-shot semantic segmentation, entity recognition technology in large languages was utilized in this study to construct a few-shot semantic segmentation model based on the large language model. Fig. 6 displays the proposed few-shot semantic segmentation model.

4.2 Experiments on Few-shot Semantic Segmentation Model Assisted by Large Language Models

This experiment was conducted to investigate ways to improve the visual content processing and comprehension abilities of college English students by combining few-shot semantic segmentation models and large language models. In this study, a dataset was constructed to provide rich and accurate semantic information for model training through network data collection and precise manual annotation. Firstly, contextual information was used to complete the entity recognition of a large language model, and then the recognition results were integrated as auxiliary inputs into a few-shot semantic segmentation model. The mloU was used as a performance evaluation metric, and it was compared with a few-shot semantic segmentation model that does not use an LLM to determine the effectiveness of integrating LLM to improve model performance. Fig. 7 displays the outcomes of the experiment.

Fig. 7 shows that the few-shot semantic segmentation model assisted by the large language model performs significantly better than the model without the assistance of the LLM in terms of the mIoU ratio of each fold. The auxiliary model achieves 72.05% to 79.01% on mIoU, while the model without auxiliary achieves only 62.81% to 69.32%. The large language model achieves precise entity recognition results by providing contextual information, significantly enhancing

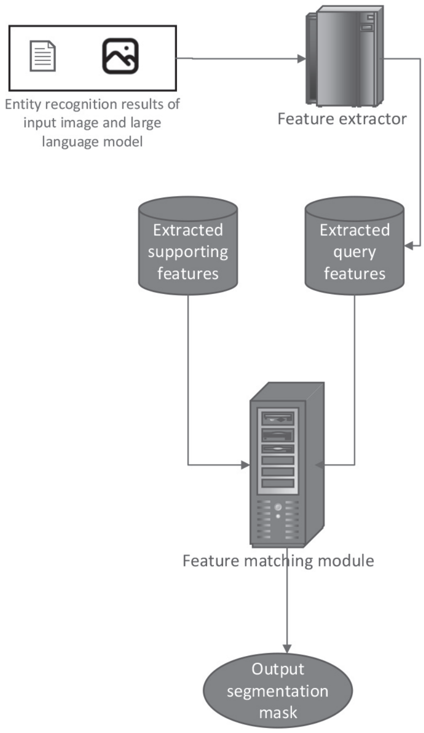

Figure 6 Framework diagram of few-shot semantic segmentation model based on large language model entity recognition.

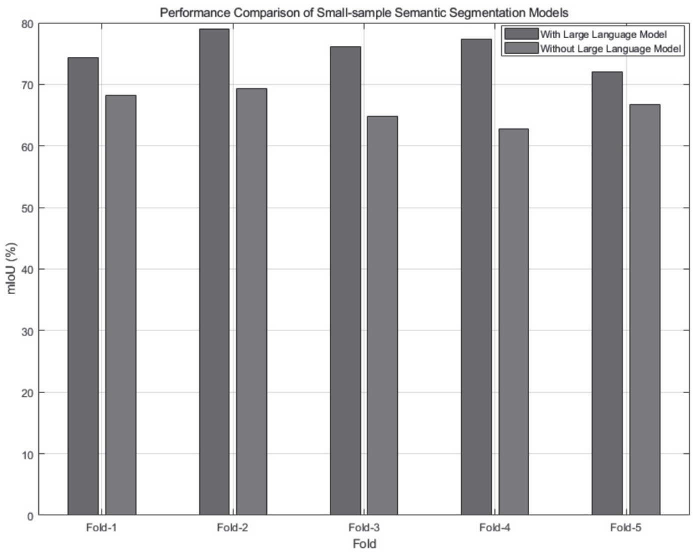

Figure 7 Experiment on few-shot semantic segmentation model assisted by LLMs.

the model's understanding of semantic information. This helps the model to better process and utilize a small amount of annotated data, thereby improving the accuracy and generalization ability of semantic segmentation. With the assistance of large language models, few-shot semantic segmentation models are better able to process and understand visual content in college English teaching, providing new pedagogical and research tools.

In order to more accurately demonstrate the improvement in performance achieved by using large language models to assist with few-shot semantic segmentation tasks, this study adopted a comparative analysis method. The proposed model is compared with two existing models to empirically demonstrate the performance of the few-shot semantic segmentation model with the support of a large language model in terms of efficiency and accuracy. The findings obtained are given in Table 4. SETR is a semantic segmentation model based on the Transformer architecture, which applies the self-attention mechanism of Transformer to semantic segmentation tasks, replacing traditional Convolutional Neural Networks (CNN) as encoders. Mask R-CNN is a deep learning model used for instance segmentation, which adds a branch for pixel level segmentation and can simultaneously perform object detection and semantic segmentation, thereby achieving instance segmentation.

Table 4 Results compared with other models (%).

Model	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Mean
SETR	58.14	53.71	56.03	55.69	57.42	56.198
Mask R-CNN	60.31	63.04	66.27	64.58	62.03	63.246
This study's model	72.04	77.41	74.95	78.34	74.22	75.392

Table 4 shows that the few-shot semantic segmentation model assisted by the large language model used in this study achieves the highest mIoU score in all folding validations, reaching 75.392%, indicating that using a large language model on a small sample dataset can significantly improve the accuracy and efficiency of semantic segmentation. The score of a single fold validation indicates that the fluctuation range of the score of the model is larger than the other two models. However, it maintains a high score overall, indicating that the application of a large language model can help the model to better generalize and learn semantic information in small sample situations.

5. CONCLUSIONS

The focus of this study was a few-shot semantic segmentation method that combined large language models. By utilizing the excellent semantic understanding and feature extraction capabilities of large language models, complex text information was transformed into useful features. From the experimental results, the mIoU ratio scores of the model assisted by the large language model in all five folds were 74.32%, 79.01%, 76.17%, 77.36%, and 72.05%, respectively. In contrast, the mIoU ratio scores of models without the assistance of large language models were 68.15%, 69.32%, 64.77%, 62.81%, and 66.72%. These scores were significantly lower than those of models assisted by large language models, and there was a noticeable performance gap in most folds. After comparing with other models, it can be seen that the average score of the few-shot semantic segmentation method combined with the large language model was significantly higher than that of the SETR model and Mask R-CNN. This indicates that the application of LLMs improves the accuracy of semantic segmentation models, enabling them to more accurately identify and segment semantic information in images. The auxiliary model can perform well even with a small amount of data, indicating its advantage in generalization ability. However, large language models rely strongly on the quality and diversity of input data in entity recognition and semantic understanding. If the data is insufficient or the labeling is inaccurate, the performance of the model may be affected. In the future, data quality and coverage can be improved by increasing data diversity, data augmentation techniques, or semi-supervised learning methods.

FUNDING

This work was supported by the "Three Levels" Talent Construction Project at the Zhuhai College of Science and Technology.

REFERENCES

1. Tsai T H, Tseng Y W. BiSeNet V3: Bilateral segmentation network with coordinate attention for real-time semantic segmentation [J]. Neurocomputing, 2023, 532: 33–42.
2. Zhang, X., Zhang, Y., Liu, X., & Wang, R. (2025). Blockchain-Based Intelligent Risk Management Decision Support System for Supply Chain Financing. International Journal of Intelligent Information Technologies (IJIIT), 21(1), 1–24.
3. Cheng B, Schwing A, Kirillov A. Per-pixel classification is not all you need for semantic segmentation [J]. Advances in Neural Information Processing Systems, 2021, 34: 17864–17875.
4. Mao B, Wang L, Xiang S, et al. Task-aware adaptive attention learning for few-shot semantic segmentation [J]. Neurocomputing, 2022, 494: 104–115.
5. Wang L, Li R, Zhang C, et al. UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2022, 190: 196–214.
6. Sehar U, Naseem M L. How deep learning is empowering semantic segmentation: Traditional and deep learning techniques for semantic segmentation: A comparison [J]. Multimedia Tools and Applications, 2022, 81(21): 30519–30544.
7. Kasneci E, Seßler K, Küchemann S, et al. ChatGPT for good? On opportunities and challenges of large language models for education [J]. Learning and Individual differences, 2023, 103: 102274.
8. Wang L, Ma C, Feng X, et al. A survey on large language model based autonomous agents[J]. Frontiers of Computer Science, 2024, 18(6): 186345.
9. Hoffmann J, Borgeaud S, Mensch A, et al. An empirical analysis of compute-optimal large language model training [J]. Advances in Neural Information Processing Systems, 2022, 35: 30016–30030.
10. Rong Guowei, Sun Bao, Jiang Xulei, et al. Construction and Implementation of Algorithm Platform Based on Large Language Model [J]. Network Security Technology & Application, 2024, (07): 48–51.
11. Chang Y, Wang X, Wang J, et al. A survey on evaluation of large language models [J]. ACM Transactions on Intelligent Systems and Technology, 2024, 15(3): 1–45.
12. Chen Jinying, Zhang Rongsheng, Ye Ayung. Exploration of Smart Laboratory Combining Large Language Model with WeChat [J]. Laboratory Science, 2024, 27 (03): 125–130+134.
13. Qi Siyang, Hu Huiyun, Li Hongbing, et al. Construction method of domain question answering system integrating large language model [J/OL]. Journal of Beijing University of Posts and Telecommunications, 1–7 [2400-07-11].
14. Shen Y, Heacock L, Elias J, et al. ChatGPT and other large language models are double-edged swords [J]. Radiology, 2023, 307(2): e230163.
15. Thisanke H, Deshan C, Chamith K, et al. Semantic segmentation using Vision Transformers: A survey [J]. Engineering Applications of Artificial Intelligence, 2023, 126: 106669.
16. Tian Xuan, Wang Liang, Ding Qi. Overview of Image Semantic Segmentation Methods Based on Deep Learning [J]. Journal of Software, 2019, 30(2): 440–468.

17. Xie E, Wang W, Yu Z, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers [J]. Advances in Neural Information Processing Systems, 2021, 34: 12077–12090.
18. Wu T, Tang S, Zhang R, et al. Cgnet: A light-weight context guided network for semantic segmentation [J]. IEEE Transactions on Image Processing, 2020, 30: 1169–1179.
19. Thirunavukarasu A J, Ting D S J, Elangovan K, et al. Large language models in medicine [J]. Nature Medicine, 2023, 29(8): 1930–1940.
20. Boiko D A, MacKnight R, Kline B, et al. Autonomous chemical research with large language models [J]. Nature, 2023, 624(7992): 570–578.
21. Peters H, Matz S C. Large language models can infer psychological dispositions of social media users [J]. PNAS Nexus, 2024, 3(6).
22. Ma Wuren, Gong Mengchun, Dai Hui, et al. A review of the application of large language models represented by ChatGPT in clinical medicine [J]. Journal of Medical Intelligence, 2023, 44(7): 9–17.
23. Ehrmann M, Hamdi A, Pontes E L, et al. Named entity recognition and classification in historical documents: A survey [J]. ACM Computing Surveys, 2023, 56(2): 1–47.
24. Wang Y, Tong H, Zhu Z, et al. Nested named entity recognition: a survey [J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2022, 16(6): 1–29.
25. Mo, Yujian, et al. Review the state-of-the-art technologies of semantic segmentation based on deep learning. Neurocomputing, 493 (2022): 626–646.
26. Puccetti G, Giordano V, Spada I, et al. Technology identification from patent texts: A novel named entity recognition method [J]. Technological Forecasting and Social Change, 2023, 186: 122160.
27. Li J, Sun A, Han J, et al. A survey on deep learning for named entity recognition [J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(1): 50–70.
28. Nasar Z, Jaffry S W, Malik M K. Named entity recognition and relation extraction: State-of-the-art [J]. ACM Computing Surveys (CSUR), 2021, 54(1): 1–39.
29. Xu, Z., Jain, D.K., Neelakandan, S. et al. Hunger games search optimization with deep learning model for sustainable supply chain management. Discov Internet Things 3, 10 (2023).
30. Weston L, Tshitoyan V, Dagdelen J, et al. Named entity recognition and normalization applied to large-scale information extraction from the materials science literature[J]. Journal of Chemical Information and Modeling, 2019, 59(9): 3692–3702.
31. Yoon W, So C H, Lee J, et al. Collabonet: collaboration of deep neural networks for biomedical named entity recognition [J]. BMC Bioinformatics, 2019, 20: 55–65.
32. Li J, Liu R, Chen C, et al. An RG-FLAT-CRF model for named entity recognition of Chinese electronic clinical records [J]. Electronics, 2022, 11(8): 1282.
33. Wang X, Zhang Y, Ren X, et al. Cross-type biomedical named entity recognition with deep multi-task learning [J]. Bioinformatics, 2019, 35(10): 1745–1752.
34. Garg S, Tsipras D, Liang P S, et al. What can transformers learn in-context? a case study of simple function classes [J]. Advances in Neural Information Processing Systems, 2022, 35: 30583–30598.
35. Liu H, Tam D, Muqeeth M, et al. Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning [J]. Advances in Neural Information Processing Systems, 2022, 35: 1950–1965.
36. Irma Lindt, Wolfgang Gräther. Design thinking for sustainable transformation: an analysis framework for practitioners. Engineering Intelligent Systems, Vol 32 No 1, 2024.
37. Caiyan Chen. Design of intelligent education decision support system based on big data analysis. Engineering Intelligent Systems, Vol 32 No 4, 2024.

Xue Han obtained her PhD in Comparative Literature and World Literature from Jilin University. Currently, she is an associate professor at Zhuhai College of Science and Technology. Her research field and interests are comparative literature and English teaching.

Email: 22412285@qq.com

Shuang Zhang obtained her Bachelor's degree in English Education from Liaoning Normal University. Then she received her Master's degree in foreign linguistics and applied linguistics from Dongbei University of Finance and Economics. She has also obtained TEFL certification from Huntington University of USA. Currently, she is a lecturer at Zhuhai College of Science and Technology. Her specializations include English pedagogy and language education. Her current research interests are Second Language Acquisition and cross-cultural teaching.

Email: zhangshuang@zcst.edu.cn

Yu Li has a Master's degree and is currently is a professor at Zhuhai College of Science and Technology. Her research field and interests are machine learning and data mining.

Email: jluzhliyu@zcst.edu.cn