Oracle Bone Inscriptions Multi-Modal Knowledge Graph Construction and Applications

Jing Xiong, Guoying Liu, Shuanghao Chen, Yongge Liu and Mengting Liu

August 9, 2021

Oracle Bone Inscriptions Multi-modal Knowledge Graph Construction and Applications

Jing Xiong[*a, b], Guoying Liu[a, b], Shuanghao Chen[b], Yongge Liu[a, b], Mengting Liu[b]
[a]School of Computer and Information Engineering, Anyang Normal University, Anyang, China 455000; [b]Key Laboratory of OBI Information Processing, Ministry of Education, Anyang, China 455000

ABSTRACT

To solve the problems of the great learning difficulty, long learning period, wide range of knowledge points but weak knowledge connection, and low sharing of Oracle Bone Studies (OBS), Artificial Intelligence (AI) techniques are introduced into the traditional study of OBS. We have collected and organized over 120 years of basic Oracle Bone Inscriptions (OBI) data and academic literature to build the OBI big data platform. Based on this platform, we proposed the OBI information processing application pyramid model, which aims to realize the digitization, datafication and intelligence of OBS. And the basis of intelligence is OBS multimodal knowledge graph. The OBS multi-modal knowledge graph can provide a unified semantic space for multi-source heterogeneous data. Through multi-modal fusion and information complementation, the defects of a single modal in OBI information processing can be solved. This multi-modal knowledge graph organizes and manages the basic data better to serve OBI information processing research. Such as OBI detection and recognition, computer aided oracle-bones joint and knowledge question answering assistant. Taking OBI detection and recognition as examples, we studied the applications of the OBS multi-modal knowledge graph. The experimental results showed that the proposed method reached 81.3% accuracy in detection and 80.43% accuracy in recognition, and it had 3.7% in detection and 14.8% in recognition improved to the conventional methods.

Keywords: Oracle Bone Inscriptions, knowledge graph, deep learning, multi-modal data, detection and recognition

1. INTRODUCTION

Since Oracle Bone Inscriptions (OBI) was first discovered in 1899, there have been more than 120 years of research history. The research on OBI has also formed an internationally prominent science-Oracle Bone Studies (OBS). In the past 120 years, a large number of basic OBI data have been sorted out, such as photos, rubbings, facsimiles, etc., and thousands of academic papers and works have been published. More and more research institutions have gradually participated in the research of OBS. Since then, a large amount of basic OBI data has also been accumulated. Especially after the introduction of computer science and information technology into traditional OBS research, a variety of new data forms have emerged, such as images, 3D models, audios, videos, and animations. In fact, OBI basic data and materials have shown the characteristics of big data and multi-modality [1, 2], and the existing OBI data we have collected and sorted out have a certain scale, and we have built an OBI big data knowledge sharing platform [2]. At present, the platform has collected 122 descriptions, 220638 images, 31310 literature, and the OBI font library contains 4586 characters. They can be accessed on the website http://jgw.aynu.edu.cn. Although this platform greatly facilitates OBS researchers, it has not been able to break through the bottleneck of OBS research. These bottlenecks are the challenges encountered in traditional OBS research, and the outstanding ones are as follows:

1. OBI Experts need to read a lot of literature when they are interpreting OBI characters. But a large amount of knowledge is scattered in these massive works of literature. It is difficult for human beings to connect useful knowledge from the massive literature. This hinders the progress of the OBS research. So OBI experts need a strong knowledge network to help them find relevant knowledge.

2. Although computer technology has brought great help to OBS research, in the OBS field, computers can only be competent for the tasks of perceptual intelligence, such as image processing and literature retrieval, they lack the

[*] jingxiong125@gmail.com; phone 0372-3300039; http://jgw.aynu.edu.cn

advantages of cognitive intelligence. However, the research of OBS needs a lot of domain knowledge and knowledge reasoning.

3. OBS research is inseparable from experts' decisions, but most of these decisions are qualitative analyses, which also leads to inconsistent opinions. There is a lack of persuasive quantitative analysis.

4. At present, computer-assisted OBS research uses images and texts as research objects separately, and different methods are adopted for different objects, such as image processing, text mining, etc. In this way, the connection between images and texts is broken. However human experts exactly use images and texts at the same time, and they confirm each other. Therefore, the use of the computer to study OBS must take into account images, texts, and other types of data.

Taking OBI detection and recognition as examples. OBI automatic detection and recognition based on a single modality is very difficult, such as the high cost of labeling [3], the low accuracy of detection and recognition [4], the inability to process the OBI bones or shells with high noise [4], incomplete OBI characters usually appear on the edges of oracle bone or shell fragments. The strokes of these characters can easily be mistaken for the natural texture of the fragments [3], and when the image background contains a lot of flake noise or the color of the OBI is close to the background, it is impossible to obtain a high-quality segmented image [5]. In traditional OBS research, OBI experts often look for some auxiliary resources to supplement and verify. These resources often appear in other modalities, such as text, facsimile. If they are integrated, mutual supplementation and verification of information can be achieved. It is shown in Figure 1.

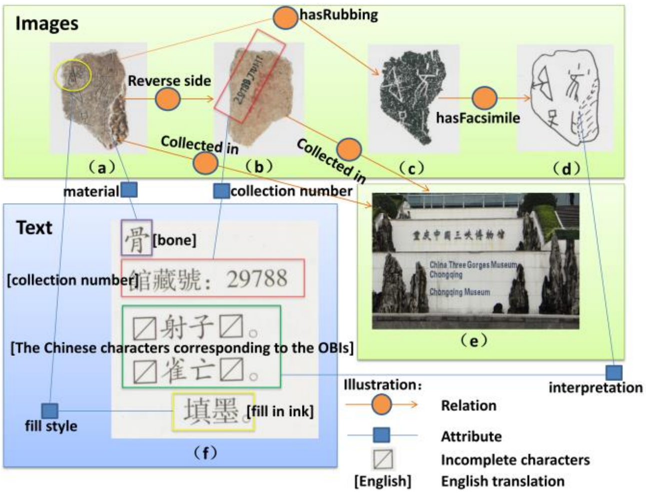

Figure 1. OBI Image-text multi-modal information integration: (a) OBI photo (front); (b) OBI photo (back), it is the back of (a). So it is the same piece of oracle bone (a); (c) OBI rubbing of (a); (d) OBI facsimile of (a); (e) The China Three Gorges Museum, Chong qing. This museum is the collection location of the oracle bone (a); (f) OBI interpretation text of (a), it was studied and confirmed by OBI experts, including the modern Chinese translation of OBI characters, as well as some archaeological information. It can be seen from (f) that (a) is an oracle bone, not a tortoise shell.

In Figure 1, (a) and (b) are OBI photos, and Figure 1 (b) is the back of Figure 1 (a). It is easy to see the material, incomplete situation, fragment edge, etc., but it is not easy to clearly recognize the OBI characters. There is a collection number arranged by archaeologists in Figure 1 (b). According to this number, some archaeological information can be traced, such as the excavation site and collection place of this oracle bone. Figure 1 (c) is the rubbing of Figure 1 (a). The OBI rubbings can clearly see the areas of OBI characters on the bones or shells, but they introduce new noises, such as the cracks may be recognized as strokes. Figure 1 (d) is the facsimile image of Figure 1 (a). The OBI facsimiles are the crystallization of the wisdom of OBI experts. Generally, they are the conclusions that OBI experts synthesize a variety of research data and have artificially removed the noise on photos or rubbings. However, because they are copied by experts, some of the character shape information is lost. Figure 1 (e) is a photo of the China Three Gorges Museum. This museum is where the real oracle bone of Figure 1 (a) is collected. Usually, the collection place can be obtained according to the collection number. For example, the sign 29788 shown in Figure 1 (b) is the collection number. Figure 1 (f) is the explanatory text corresponding to Figure 1 (a), which is called interpretation text. It records the modern Chinese characters corresponding to OBI characters on the oracle bones or shells and other important information. If there are incomplete or unclear characters, they will be described in the interpretation text. They are achievements of OBI experts.

It can be seen from Figure 1 that different types of data can provide different information, to meet different needs. And the confirmation of a lot of information relies on the experiential knowledge of OBI experts. How to make full use of basic OBI data in various modalities, and to share and reuse OBI experts' experiential knowledge is an important task in OBS research. Building a multi-modal OBS knowledge graph is an effective solution.

The main contributions of this paper are listed as follows:

1. Based on the OBI big data platform, the OBI information processing application pyramid model is proposed. It describes the foundations and applications of OBI information processing in terms of the data layer, the representation layer, and the application layer, respectively.

2. To put forward a solution to construct the OBS multi-modal knowledge graph and give the framework and process. The OBS multi-modal knowledge graph organizes and manages the multi-source heterogeneous basic OBI data better, and integrates the experience and knowledge of experts, which can better realize the sharing and reuse of OBS knowledge. It provides new research ideas in the field of OBI information processing and changes the previous method of separating images or texts as research objects.

3. Taking OBI detection as an example, we demonstrated the advantages of the OBS multi-modal knowledge graph. Through the comprehensive use of rubbings, facsimiles, and fonts, inefficient denoising was avoided and the detection and positioning of OBI characters were improved.

4. Taking OBI recognition as an example, we demonstrated the advantages of the OBS multi-modal knowledge graph. We map the feature vector of the image and the text vector to the same semantic space and improve the training efficiency of the image recognition model under the supervision of the text. And the CBOW is used to predict incomplete or unclear OBI characters.

The rest of the paper is organized as follows: In Section 2 we discuss the relevance of the multi-modal knowledge graph. Section 3 introduces the OBI big data platform and OBI information processing application pyramid. Section 4 introduces the construction process of the OBI multi-modal knowledge graph and shows its elements and scale. Section 5 introduces the advantages and applications of the multi-modal knowledge graph in OBI character detection and recognition in detail. Section 6 is the experiment and analysis. Finally, Section 7 presents our conclusions and the limitations of this paper and points out the possible solutions.

2. RELATED WORK

There are many researches related to the multi-modal knowledge graph. They are usually the fusion of image, text, audio, video, and other modals. Liu et al. [6] proposed MMKG, a collection of three knowledge graphs containing both numerical features and images of all entities, as well as the arrangement of entities between KG pairs. Wilcke et al. [7] proposed a multimodal message-passing network that allows end-to-end learning not only from the structure of graphs, but also from their possibly diverse multimodal node features. Zhu et al. [8] proposed a base framework for large-scale multi-modal knowledge to handle a wide variety of visual queries without the need to train new classifiers for new tasks. Huang et al. [9] proposed a novel image-text sentiment analysis model that uses discriminative features and internal correlations between visual and semantic content to perform sentiment analysis using a hybrid fusion framework. Xu et al. [10] proposed a novel self-supervised ternary adversarial network model named TANSS to overcome the problem of zero-shot cross-modal retrieval tasks. Lu et al. [11] extend the popular BERT architecture to a multi-modal dual-stream model that processes both visual and textual inputs in separate streams, interacting through co-attentional transformation layers to process a number of vision-and-language tasks.

Although these multi-modal knowledge graph researches have achieved outstanding results in their respective fields, and also provide useful guidance for us to construct the multi-modal OBS knowledge graph, they cannot solve the problem of OBI information processing well. These studies are dedicated to replacing traditional feature extraction with feature learning methods, but the current research on OBI information processing cannot get rid of artificial feature extraction. Most of them are overly dependent on big data, but lack a self-idea function [12]. They have many mature and big corpora for research. Most of these corpora can be automatically aligned using machine learning methods. But they still lack the feature extraction and feature representation for OBI. The basic data of the OBI itself cannot reach the volume of big data, and large-scale annotation is not possible due to the scarcity of OBI experts. Making full use of the existing research results and academic literature is the most feasible approach.

The biggest difference from other multi-modal knowledge graph research is that OBI information processing is highly dependent on experts' decision. For example, there are many OBI variant characters, and they correspond to the same modern Chinese characters, but they are very different in glyph (Figure 2 (a)). Some individual OBI characters sometimes appear in a combined form (Figure 2 (b)). A large number of unrecognized OBI characters are clear in the glyph, but their meanings are uncertain (Figure 2 (c)). OBI does not have a fixed writing sequence, and multiple writing sequences usually appear on one piece of bone or shell (Figure 2 (d)). At present, dealing with these problems requires the confirmation of OBI experts.

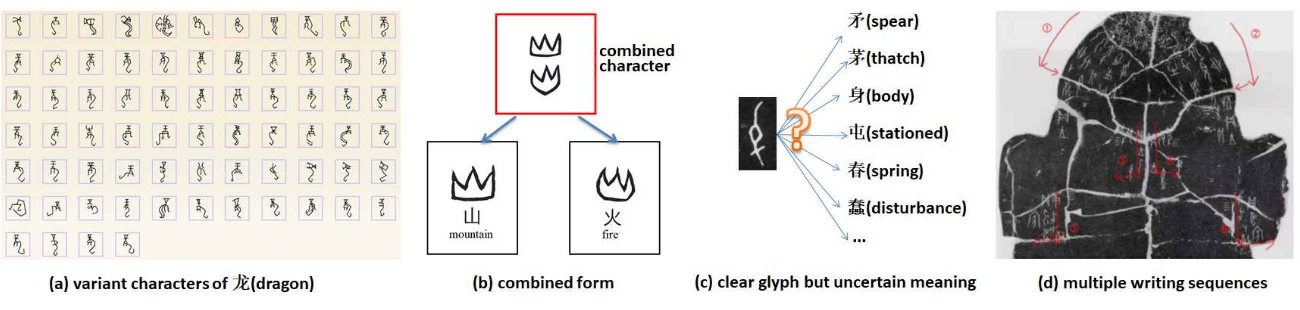

Figure 2. Some characteristic examples of OBI.

As far as the OBI interpretation is concerned, it is not the explanation and interpretation text of the characters on oracle bones or shells (for existing other multimodal corpus, images often have a paragraph of text to describe their content), but to interpret the OBI characters as modern Chinese characters. In this way, the known OBI characters are represented in modern Chinese characters in the text, and the unknown OBI characters still keep the original glyph. Even if the characters on the oracle bones or shells are recognized by image processing technology, the correct character order cannot be determined, as shown in Figure 2 (d). Therefore, it is necessary for OBI experts to determine the character order after interpretation, to form the interpretation text.

For OBI detection and recognition, there were also some researches. Xing et al. [3] summarized a guide for selecting a detection architecture for OBI that achieves the right speed/memory/accuracy balance for a given platform. However, it was only a comprehensive comparison experiment and did not provide a practical solution. Liu et al. [13] built a CNN model for recognizing OBI characters and achieved good results. However, the recognition objects were not the original OBI images, but the characters copied by experts. Unlike the actual rubbing features, they have been artificially denoised. Liu et al. [14] proposed an incomplete OBI recognition method based on convolutional neural networks. It could extract features from the data set of incomplete OBI with small samples and recognize incomplete OBI on different rubbings. However, the results are not satisfactory when the noise is large or the characters are severely incomplete. So far, the OBI detection and recognition are only based on image processing without considering other modalities such as OBI text.

3. OBI BIG DATA PLATFORM AND APPLICATION PYRAMID

After more than 20 years of research and accumulation, and nearly 5 years of development, we have built an OBI big data platform named YinqiWenyuan (http://jgw.aynu.edu.cn/). It is an OBI knowledge sharing platform that integrates font library, description library and literature library, and is open to the whole world for free. This is currently the world's largest and most complete OBI basic data service platform. With the continuous expansion of data in the platform and the continuous improvement of functions, more and more applications will serve the research of OBI information processing. Based on this platform, we proposed an application pyramid model framework, as shown in Figure 3.

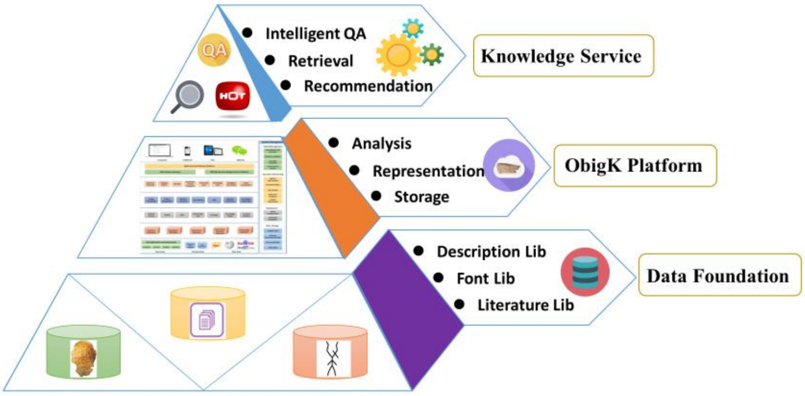

Figure 3. OBI information processing application pyramid model.

It can be seen from Figure 3 that the model is classified from the three levels of digitization, datafication and intelligence of OBI basic data. The bottom is data layer and it is the three foundations of the OBI big data platform: description library, literature library, and font library; the middle is representation layer which sorting out digital resources, optimizing storage, analysising and knowledge representation, forming an OBI big knowledge platform called ObigK[2]; the top is the application layer, which is based on the functions of ObigK and provides users with intelligent knowledge services, such as literature full-text search, knowledge recommendation, and intelligent question answering. The foundation of intelligent knowledge services is the OBS knowledge graph.

The innovations and contributions of the pyramid model are as follows:

1. Deploy OBI description, literature, and font library on the same platform to meet the different needs of OBS research. The three are independent and integrated, and establish a close relation between each other, which greatly facilitates the OBI researchers to access the literature and confirm the description.

2. Using fragmented annotations of OBI to semantically label the OBI character images in the literature, so as to solve the problem that the current literature retrieval cannot be based on the form of OBI character images.

3. Through the design and implementation of the OBI font library, the liding characters in the OBI interpretations are replaced, and the font library can be used to represent the unrecognized OBI characters, to realize the full-text literature retrieval based on the OBI characters.

4. By constructing the OBS knowledge graph, the relations between entities are established, and it can provide OBI experts with literature research and OBI interpretation based on knowledge reasoning clues. At the same time, it can provide OBI knowledge intelligent question answering service for OBI enthusiasts.

4. OBS MULTI-MODAL KNOWLEDGE GRAPH CONSTRUCTION

The key problem of constructing the OBS knowledge graph is to discover knowledge entities from various heterogeneous and multi-modal data sources and establish the correlation between entities. The construction process is shown in Figure 4.

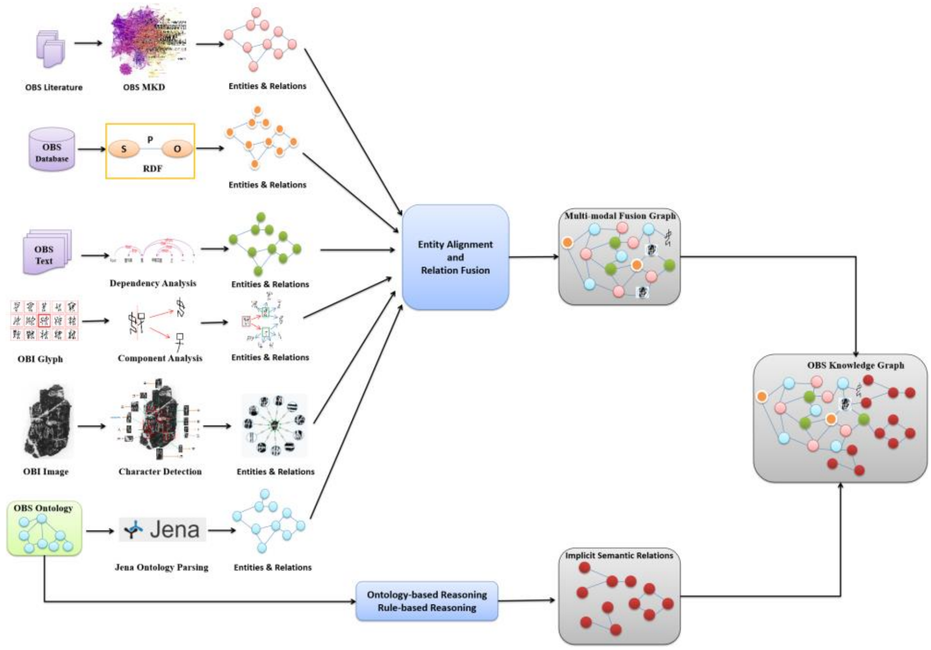

Figure 4. OBS knowledge graph construction process.

It can be seen from Figure 4 that entity extraction and relations extraction in the knowledge graph are implemented separately based on different types of data sources. Then the entities are aligned, the relations are merged, and the potential semantic relations are discovered based on knowledge inference. Finally, a large knowledge graph is formed.

4.1 Entities and relations extraction based on Mapping Knowledge Domains

Because of the ancient characteristics of OBI, the study of OBS has to rely on a large amount of literature and a series of knowledge correlation analysis based on the literature. For example, the relationship between scholars and literature, scholars and their cooperative relationship, research institutions and their cooperative relations, and the relations between citation and citation. These problems involve bibliometric and analytical techniques. Therefore, Mapping Knowledge Domains (MKD) [15] is also one of the data sources.

There are many existing MKD construction methods, such as co-citation analysis, co-word analysis, cluster analysis, social network analysis, and comprehensive analysis method integrating other literature features. Among them, the co-word analysis method determines the relationship between subjects in the subject field represented by the text and analyzes the scientific development of the field by analyzing the form of the co-occurrence of key words in the same text content. We choose co-word analysis to describe the relations between knowledge entities. And cosine function method is selected to calculate the co-word relations.

4.2 Entities and relations extraction based on relational database

During the study process of the OBS, many relational databases have been built, such as the OBI dictionary, OBI description database, OBS literature database, and OBI grammatical database. The data objects are actually analyzed and designed and the database schema is defined when the databases are built. Database schema has become an important method of knowledge acquisition in ontology development. Relations, relational attributes, atomic data types of attributes, attribute constraints, primary/foreign keys, etc. in the database schema provide descriptions of concepts or classes and their relations for knowledge base construction. Therefore, entity discovery and relation extraction can be realized based on database. RDB2RDF can convert data stored in relational databases into RDF triples. We used R2RML to realize the conversion from relational databases to RDF.

4.3 Entities and relations extraction based on glyph structure

OBI is a key form of the development of Chinese characters and is called the earliest Chinese character. Its shape and structure gradually shifted from without analyzable components to synthesis character, and a large number of pictophonetic characters have appeared. At the same time, OBI characters have already possessed the rules for the construction of Chinese characters such as pictographic, referring, associating, borrowing, transcription, and phonogram. Therefore, there are semantic relations between the combined characters and their components, which play an important role in the OBI interpretation. So establishing the connection between the glyphs and their radicals is also part of our task of extracting entities and relations.

4.4 Entities and relations extraction based on text

OBS text refers to the text that involves OBS knowledge in addition to documents, databases and interpretations, including study notes and web pages. Extracting relations from unstructured text is extremely challenging. At present, entity and relation extraction based on deep learning has surpassed the traditional feature-based and kernel function-based methods and produced a number of important results. However, the existing studies on the OBS rely heavily on expert knowledge and lack of feasible annotated corpus or data set, so the advantages of deep learning cannot be highlighted. Therefore, we used dependency syntactic analysis to finish the entities and relations extraction, and used the language technology platform (LTP [16]) as a dependency syntactic analysis tool. Combined with OBS domain knowledge, it can extract entities and relationships from text [17].

4.5 Entities and relations extraction based on OBI images

The most basic work of OBS is to understand the inscriptions on bones or shells. At present, the images of OBI include photographs, rubbings and facsimiles, of which rubbings are the main forms. Researchers need to know what kind of description a certain rubbings came from, what types of OBI characters were on the bones or shells, what types of OBI characters had been interpreted, and what glyphs (variant characters) were in the bones or shells. Therefore, it is very important to establish the relations between the OBI characters and the bones or shells, between the bones or shells and the description, and between the OBI characters and its variant characters. The prerequisite is to detect and recognize the OBI characters from the bones or shells.

YOLOv3 [18] was used to realize automatic detection of OBI characters, with an accuracy rate of 77.6% [3], which reduced manual work. On this basis, the individual OBI characters were separated, so as to obtain the relations between the OBI characters and the bones or shells. Then find out the relevant variant characters of the recognized OBI characters and establish the variant character relations.

4.6 Entities and relations extraction based on ontology

Ontology can provide conceptual model and logical basis for knowledge graph. Since ontologies describe concepts, attributes, and their relations, they are already relational when they are created. When constructing knowledge graph, entities and their relations can be obtained directly from ontology. At present, we have built the OBS document ontology, OBS content ontology and OBS common ontology [19], among which the OBS document ontology is the resource ontology established based on OBS research papers and monographs. The content ontology of OBS is a knowledge base which describes the social life of Shang dynasty and its mutual relations and is interpreted by oracle experts and historians. The OBS common ontology describes the basic knowledge of OBS, including the discovery history, archaeological records, characters and grammar. We used Jena to parse the ontology to obtain entities and relation.

4.7 Multi-modal knowledge graph fusion

After obtaining a large number of entities and relationships from multi-modal data sources, knowledge fusion of these entities and relationships will be carried out, which requires consideration of two key issues: entity alignment and relation fusion [17].

Entity alignment is a process to determine whether two entities in the same or different data sets point to the same object in the real world [20]. The inscriptions of OBI have many kinds of equivalent entities. Such as the same entity may have different names, and the same oracle bone or shell may have different description numbers. If the entity pairs are recorded in interpretation text as "=" or are marked in ontology as "isSame" relation, or have corresponding fields in the database, then they will be directly judged as equivalent entities when they appear in the same or different data sets. Due to the strong specialization of OBS, many equivalent entities in the knowledge graph need to be determined using the OBS domain knowledge.

The key to relation fusion is to determine whether two entities express the same relation, or whether they have an inclusion relation, etc. Our relation fusion of the OBS knowledge graph mainly considers the equivalence class and the "subClassOf" relation.

After knowledge fusion, we got a huge knowledge graph. However, these entities and relations are all explicit elements directly obtained from multi-modal datasources, which cannot better meet the OBS knowledge reasoning needs. Therefore, it is an important task to extend the OBS knowledge graph by mining the potential entities and relations hidden behind the explicit ones. We realized the mining of implicit semantic relations by using both ontology-based reasoning and rule-based reasoning and thus extend the OBS knowledge graph. Among them, ontology-based reasoning is to use the defined relations and axioms in the OBI ontology for reasoning, and also to make full use of the transitivity and reflexivity of the relations. Rule-based reasoning requires writing rules under the guidance of OBI experts to compensate for reasoning that cannot be done directly by using OBI ontology.

The potential semantic relations obtained after knowledge reasoning and the newly discovered entities can further expand the scale of the OBS knowledge graph. The current OBS knowledge graph includes OBI experts and scholars, description, research institutions, publishers, oracle bones and shells, museums, OBI characters, OBI common sense, and other knowledge entities and relations. It includes 148305 entities and 434032 relations, as shown in Figure 5.

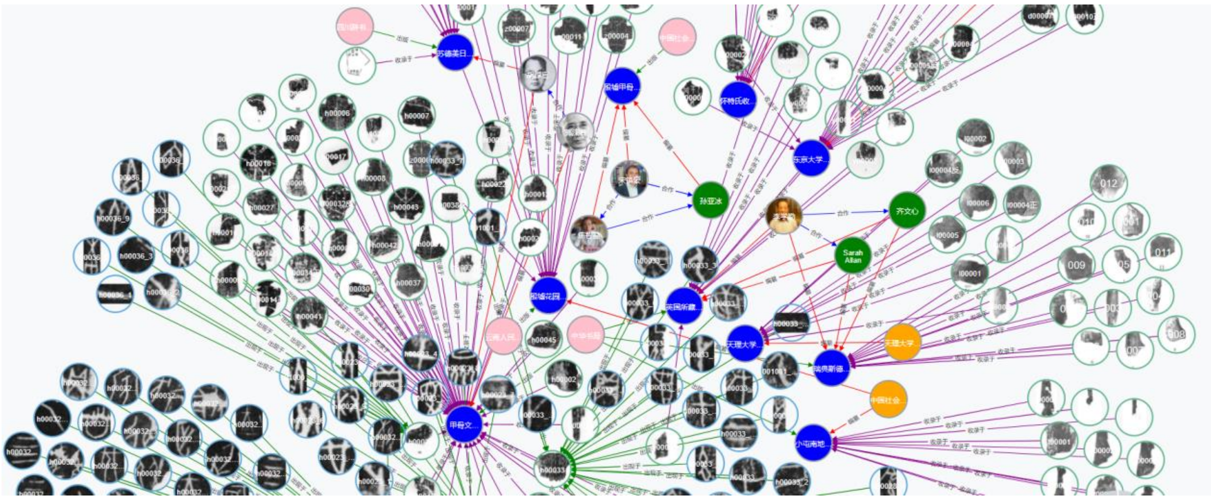

Figure 5. OBS knowledge graph fragment.

5. APPLICATIONS BASED ON MULTI-MODAL KNOWLEDGE GRAPH

The OBS multi-modal knowledge graph integrates multi-source heterogeneous OBI basic data and establishes the association between knowledge entities. It forms a huge knowledge network by describing the attributes of these entities and the relations between them. At the same time, OBI experts directly participate in the process of constructing the knowledge graph. So it can be used for OBI information processing applications and provides background knowledge and knowledge reasoning ability. With the support of the knowledge graph, some challenging applications and research such as OBI detection and recognition, computer aided oracle-bones joint and knowledge question answering assistant can find solutions.

5.1 Multi-modal knowledge graph for detection and recognition

The automatic detection and recognition of OBI are very challenging tasks, which are mainly due to the noise of OBI images, incomplete information due to oracle bone or shell fragments, and large differences in multiple glyphs of the same OBI character. In the OBS knowledge graph, several subgraphs can be obtained by finding the association path between cross- modal entities. These subgraphs constitute some slots of OBI, and they work together to build the "OBI Profile", as shown in Figure 6. This can make up for the shortcomings of OBI detection and recognition using a single image processing method.

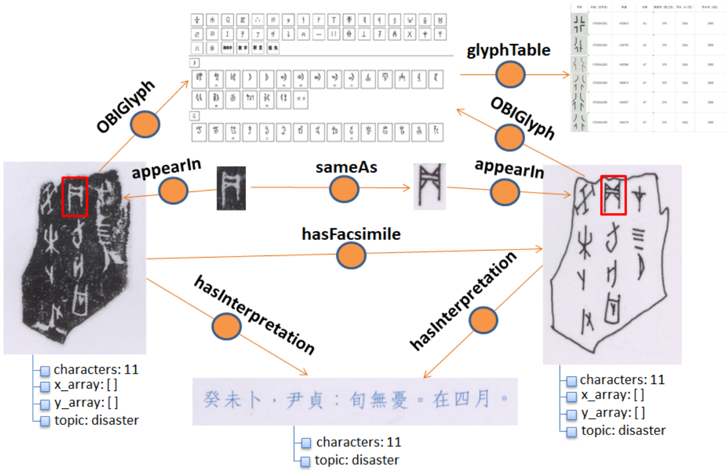

Figure 6. A sample of multi-modal knowledge graph for OBI detection and recognition.

Generally, the objects of OBI detection and recognition are rubbings. However, only using rubbings is not enough to accomplish this task, so it is necessary to comprehensively use all kinds of multi-modal data, including images (rubbings and facsimiles are both images, but not every rubbing has a corresponding facsimile), texts, databases, font libraries, etc., as shown in Figure 6. Since the facsimile is based on rubbing, the coordinate information of the OBI characters on the facsimile is consistent with the rubbing. They are stored separately in x_array and y_array. Although the shape of the characters on the facsimile has changed a lot, it clearly shows the outline of each OBI character, which can effectively remove the noise on the rubbing. As shown in Figure 6, the characters on the right side of the rubbings are so vague that they are almost unrecognizable, but they can be clearly distinguished from the facsimile. At the same time, the interpretation text can provide the semantic information which has been interpreted by OBI experts. So we can determine the performance of the detection model faster, and network parameters can be adjusted in time. In particular, it can be seen from the interpretation text that it is about consulting disaster. Therefore, the recognition of OBI characters should be limited to this topic. We have constructed an OBI semantic dictionary, which records the semantic categories and related topics of each OBI character. If the recognized OBI characters do not match the topics stored in the knowledge graph, they will be regarded as incorrect recognition.

5.2 OBI detection and recognition flow

Based on the OBS multi-modal knowledge graph, we extract the features of different morphological data of images and texts respectively and then map them to a unified semantic space, to realize knowledge fusion and semantic representation. This method can effectively compensate for the defect of single-mode representation. The detection and recognition process based on the OBS multi-modal knowledge graph is shown in Figure 7.

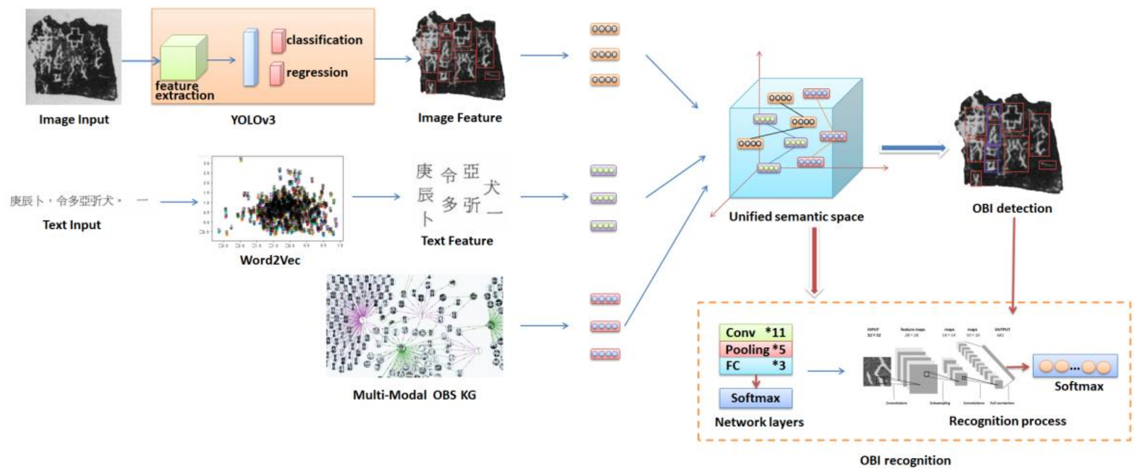

Figure 7. Multi-modal knowledge graph-based OBI detection and recognition flow.

OBI detection and recognition use a deep learning framework. The function of the multi-modal knowledge graph is to provide a unified semantic representation. By integrating image features and text features, the semantic representation of the different modals can complement each other. Thus improve the accuracy and efficiency of OBI detection and recognition.

In the unified semantic space, OBI detection can be described as follows:

$$o_d{*} = \text{argmax}(o_d)P\left(o_d \mid I, I_f\right) \tag{1}$$

Where o_d means the output of OBI detection and it appears as bounding boxes around the OBI characters; *argmax* refers to the arguments, at which the output o_d is as large as possible; I means the input OBI rubbing and I_f means the input OBI facsimile.

YOLOv3 is a good choice for OBI detection. Our previous research has shown that YOLOv3 performed well in the test datasets, with higher precision, faster speed, less overlap, and sensitivity to small targets and suppress noise [3]. Compared with existing researches, we used text information to supervise and guide further network training. OBI experts can't label every OBI rubbing, but the accuracy of the labeling cannot be separated from the experts' manual proofreading. Our approach was to let students label the OBI characters on the rubbings of the training dataset first, and then obtained a detection model after training. This model was used to perform OBI detection and screen out the results with high detection accuracy. The results were submitted to OBI experts and after proofread manually by experts they would be used as new labeled samples, and then the detection model would be retrained. This method greatly reduced the manual proofreading work of experts. In the screening process, the interpretation text information can also help us quickly discard poorly performing detection results.

The facsimiles can be used to assist detection. The facsimiles of the OBI rubbings can be obtained by querying the knowledge graph, and they will be used for network training. Because the facsimiles can make up for the insufficiency of OBI rubbings in noise reduction and can achieve higher accuracy [13]. The characters on the facsimiles are extracted and imitated from oracle bones or shells after research and confirmation of OBI experts. Although they are not real OBI characters, their outlines and locations are the same as the original OBI characters. It is enough for OBI detection.

However, some OBI rubbings do not have corresponding facsimiles and they are difficult to detect the OBI characters. In this case, we use a deep learning network to extract clear OBI characters from rubbings and retain their original position information, and then perform OBI detection. This is actually a task of extracting characters on OBI rubbings. Therefore, we construct a deep neural network model incorporating semantic segmentation network and generative network to reduce the influence of oracle bone or shell background and inscriptions on the character extraction results. The hybrid model is used because there are several challenges in extracting and generating characters from OBI rubbings.

1. The surface of the OBI rubbing contains a large amount of irregular noise, which is densely distributed on the image surface of the rubbings and not only interferes with the recognition of character features, but also tends to increase the risk of overfitting in the character extraction.

2. There are various styles of crack interference on the surface of OBI rubbings, which have different scales and shapes, and they present very similar shapes to OBI characters on the rubbings, which can be easily mistaken as OBI characters. This seriously interferes with the accuracy of OBI character detection.

3. The position information and geometric priors of OBI characters in rubbings are unknown. This poses a great obstacle to the discrimination of character features and the spatial integrity of constrained characters.

Our model tries the best to overcome these problems. The model uses Generative Adversarial Networks (GAN) as the basic skeleton, and treats the OBI character extraction problem as an image-to-image conversion task; at the same time, in order to take advantage of the strong ability of the semantic segmentation network to distinguish the rubbings image background from the OBI characters, the model integrates the semantic segmentation network into the generator network, and embeds the segmentation network into the the encoder network to eliminate the influence of the rubbing background noise, so that establish a more accurate mapping relationship between the rubbing image and the corresponding OBI character image. Our approach is as follows:

1. To reduce the interference of background noise and cracks in the OBI rubbing images, the generation network of the model includes an embedding learning stream to achieve discriminable feature representation learning of the background and OBI characters in the feature embedding space.

2. In order to adapt to the variation of OBI characters' size in the rubbing images and generate clear and complete OBI character images, we construct a character generation stream in the generative network using residual block and multi-scale feature channel connections.

3. To ensure that the interference of the rubbings noise and cracks can be reduced while the integrity of the OBI characters in terms of spatial structure can be guaranteed, the generation network uses a Spatial Attention Model (SAM) to fuse the results of the two branches.

4. The model uses a combination of global discriminator network and local discriminator network to discriminate the generated OBI character images consistently.

The model can realize the automatic OBI character extraction with good results. It achieved 88.07% and 78.28% on Mean Intersection over Union (MioU) and Intersection over Union (IoU), respectively. The generated result is shown in Figure 8.

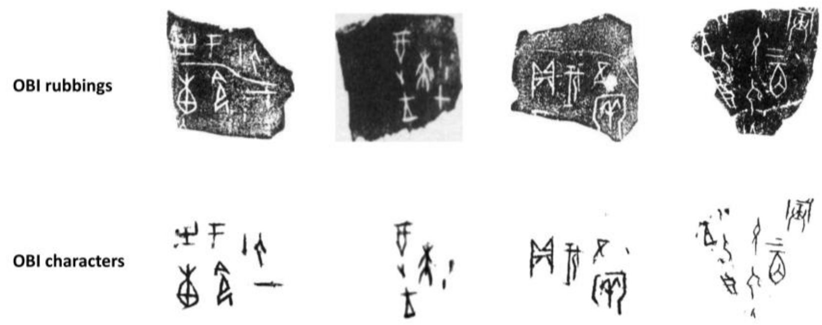

Figure 8. Deep learning based OBI character extraction.

OBI recognition is based on the results of OBI detection and it can be described as follows:

$$o_r* = \text{argmax}(o_r)P\left(o_r \mid I_d, T\right) \tag{2}$$

Where o_r means the output of recognition and it is the modern Chinese character corresponding to the input OBI character rubbing. I_d means the input OBI character rubbing vector, T means the input OBI interpretation text vector. In our algorithm, the OBI interpretation text vector uses the CBOW model of Word2vec [21].

OBI interpretation texts are unique character sequences in the OBS field, and the existing trained Word2vec corpus cannot be applied. In our research, we need to train the Word2vec model from scratch based on the OBI interpretation texts. What needs to be pointed out, in particular, is that usually, an OBI character is an independent semantic unit. Therefore, what we construct are character vectors.

Since the CBOW architecture predicts the current word based on the context, we choose it to generate OBI character vectors. The corpus now includes 4831 interpretation sentences. Separated by character, a total of 884 different characters were obtained. The window size is set to 2 and we took the embedding size as 200. The OBI character vector is shown in Figure 9.

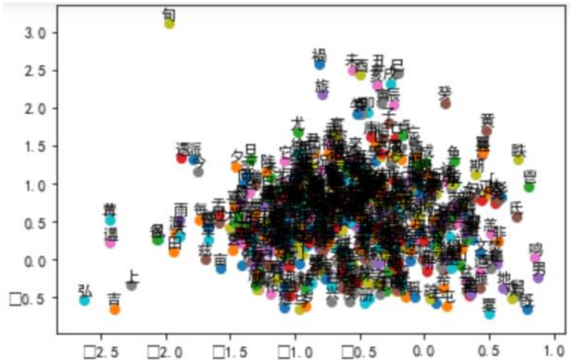

Figure 9. OBI character vector.

As can be seen from Figure 9, the distance between vectors indicates the degree of semantic similarity of the OBI characters. This helps us to predict the unclear or incomplete OBI characters.

The recognition model is expressed as follows:

$$L = \sum_{c \in C} \log p(c \mid Context(c)) \tag{3}$$

Where L is the objective function, it's a logarithmic likelihood function, c means the modern Chinese characters that need to be predicted, C is the collection of modern Chinese characters corresponding to the known OBI characters. $Context(c)$ is the context of c in interpretation text, which is composed of several characters before and after the character c.

The objective function of the CBOW model based on the hierarchical softmax is as follows:

$$\begin{aligned} L &= \sum_{c \in C} \log \prod_{j=2}^{l^c} \left\{ \left[\sigma\left(X_c^T \theta_{j-1}^c \right) \right]^{1-d_j^c} * \left[1 - \sigma\left(X_c^T \theta_{j-1}^c \right) \right]^{d_j^c} \right\} \\ &= \sum_{c \in C} \sum_{j=2}^{l^c} \left\{ (1-d_j^c) \cdot \log\left[\sigma\left(X_c^T \theta_{j-1}^c \right) \right] + d_j^c \cdot \log\left[1 - \sigma\left(X_c^T \theta_{j-1}^c \right) \right] \right\} \end{aligned} \tag{4}$$

Where X_c is the vector about c, which is shown in formula (3); l^c means the number of nodes contained in the path p^c, p^c means the path from the root node to leaf node c; $d_j^c \in \{0,1\}$ means the code corresponding to the j^{th} node on the path p^c, 1 represents the left direction and 0 represents the right direction; $\theta_j^c \in R^m$ is the vector corresponding to the j^{th} non-leaf node on the path p^c, here m is the dimension of the vector.

In the detection stage, since the image vectors and character vectors are in a unified semantic space, the accuracy of recognition can be improved by calculating the distance between the vectors. This is because in the case of character vectors and image vectors, the relative distances in the semantic space are the same, although their vector representations do not coincide. We used the cosine similarity to calculate the distance between the image vector and the character vector. It is expressed as follows:

$$S(I,T) = \frac{\sum_{i=1}^{n} I_i T_i}{\sqrt{\sum_{i=1}^{n} I_i} \sqrt{\sum_{i=1}^{n} T_i}} \tag{5}$$

Where I is the image vector of OBI, and T is the character vector of OBI. n means the vector dimensionality. The larger the value of $S(I,T)$, the farther the distance between the vectors and the smaller the similarity.

5.3 Computer aided oracle-bones joint

Oracle-bones joint is to splice and combine the broken oracle bones or shells together. Oracle bones or shells in the burial of thousands of years, due to the pressure of the stratum, water and moisture infiltration and excavation and other reasons, a lot of oracle bones or shells broken into several pieces. Some oracle bones or shells were artificially damaged after excavation. It brings many difficulties and inconvenience to OBS research. The OBI obtained through the oracle-bones joint are of high historical value. Therefore, it is one of the most important tasks in OBS research to recover the broken bone or shell pieces as much as possible.

After the OBIs are jointed, they can find out the interrelationships between the inscriptions, restore the original inscriptions, and become an important historical material for understanding the society of the Shang Dynasty. In the research of oracle-bone joint, it is necessary to comprehensively consider multiple data sources and data attributes, such as images, joint plates, interpretations, jointor, joint method, joint time, description, published articles about oracle-bone joint. It is also necessary to consider the information such as the shape of bones or shells, interpretation text, OBI characters, and stage divination.

It can be seen that to engage in oracle-bone joint research requires experts’ long-term scientific research accumulation and keen insight into OBI materials and literature, and many bits of joint clues are often hidden in massive OBI basic data and academic literature. Moreover, various data are often required to verify each other and complement each other in the joint process. If the material of the smaller rubbing cannot be determined, it can usually be supplemented with photos or 3D images. when the combination of candidate joints is correct, the interpretation is usually used for verification. These clues often have direct or indirect connections. Once the key points are found, they may be successfully jointed. Once the related information is stored and recorded in the multi-modal knowledge graph, valuable clues can be found through node path searching, which can help OBI experts to joint the OBI fragments.

In computer-assisted oracle-bones joint studies, it is also important to exclude incorrect joint candidates. On the one hand, the jointed oracle bones or shells are beneficial to restore the OBI interpretation text; on the other hand, the oracle bone inscriptions to be restored can also verify whether the joint candidates are correct. However, judging whether the candidate oracle bones to be conjugated is correct is a high-demand and high-standard work, and it is still impossible to get rid of the existing research dilemma by relying entirely on OBI experts. The knowledge graph can be advantageous in this respect. Based on the knowledge graph, and through retrieval, reasoning, consistency check, abnormal point analysis, and group mining can find abnormal information. For this reason, knowledge graph are widely used in anomaly detection, such as financial fraud issues. Similarly, this advantage can be transferred to oracle-bones joint research, and provide clues for OBI experts by finding "abnormal oracle-bones joint". An example of the oracle-bones joint is shown in Figure 10.

Figure 10. An example of oracle-bones joint.

As can be seen from Figure 10, this example was manually jointed. It has both rubbing and facsimile, which once again proves that the actual OBS research requires multiple modalities or forms of OBI data. Judging from the shape and outline of the oracle bone pieces, it matches relatively well and appears to be a correct example of oracle-bones joint. However, by joining the interpretations on the two fragments together, errors can be found in grammar and semantics, and thus it is judged to be an incorrect example of joint. This conclusion not only shows the importance of comprehensive utilization of multi-modal OBI data, but also proves that OBS multi-modal knowledge graph are effective in oracle-bones joint research.

5.4 OBI knowledge question answering assistant

The purpose of constructing the OBS knowledge graph is to establish connections between OBI knowledge entities and form a large-scale knowledge network. Its function is to sort out the knowledge chain according to a certain logic from the complicated and intertwined massive knowledge points, so as to realize knowledge retrieval and knowledge service. When the knowledge navigation path is formed in the knowledge network, the intelligent question answering (QA) function can be realized. Based on the OBS multi-modal knowledge graph, we have developed an intelligent QA assistant named "Wen Xiaoyuan", which can conduct less complicated OBI knowledge question answering. As shown in Figure 11.

Figure 11. OBI question answering assistant-Wen Xiaoyuan.

The realization method of intelligent QA based on the knowledge graph is to use NLP technology to parse the natural language question input by the user, extract the named entities and the relations in the question sentence, and match them to the entities and relations in the knowledge graph. Then generate and execute knowledge graph query statements, thereby returning possible answers. Here we used Neo4j Cypher query language to implement knowledge graph query. At the same time, we have adopted the following methods:

1. Commonly used word segmentation tools cannot achieve high accuracy in the OBS field, so we have built an OBI dictionary to manage OBS terms and common word collocations in a unified manner.

2. Usually the nouns in the question input by the user are the candidate elements of the entities, and the verbs are the candidate elements of the relations. But in the OBS knowledge graph, the relations may be expressed by nouns, or they may be multi-speech words. To improve the quality of QA, we maintain a relational vocabulary. This vocabulary is extracted from the relation set of the OBS knowledge graph, and their synonyms are considered.

6. EXPERIMENT AND ANALYSIS

The OBS multi-modal knowledge graph can comprehensively utilize the basic OBI data of different modalities, which can complement and verify each other, and it also contains the empirical knowledge of OBI experts. Therefore, it can solve some problems faced by OBI information processing research. The following only takes OBI detection and recognition as examples for experiments and analysis. The OBI detection and recognition we designed is a 2-stage task. That is, the first step is to complete the detection task, and then complete the recognition based on the detection results.

6.1 OBI detection

We have collected 9500 OBI rubbings from "*Collection of Oracle Bone Inscriptions*" and "*Supplement of Oracle Bone Inscriptions*" by a high-resolution scanner then labeled every character with the upper left and lower right coordinates [3]. Notably, this stage only focuses on the detection task, so the number of classes in this dataset is regarded as one. In other words, we just need the results detected to be OBI characters. The data has been published and will be updated on the website http://jgw.aynu.edu.cn.

The experimental results of OBI detection are shown in Table 1.

Table 1. OBI detection results.

Method	Precision (%)	Recall (%)	F-score (%)
SSD	74.8	75.8	75.3
Faster R-CNN	75.4	77.8	76.6
RFBnet	76.1	78.9	77.5
RefineDet	75.2	80.5	77.8
YOLOv3	77.6	78.4	78
Ours	81.3	87.6	84.3

It can be seen from Table 1 that our method based on the multi-modal knowledge graph has improved the accuracy and recall compared with single image processing. The recall has been greatly improved with the help of facsimiles corresponding to the rubbings, and the detection model found more correct bounding boxes. However, the improvement of the accuracy is relatively small, mainly because there are still incomplete or unclear characters on the oracle bones or shells, and these characters are also incomplete in the facsimiles, so they can not get more accurate box coordinate from corresponding facsimiles.

6.2 OBI recognition

The dataset of OBI recognition is the data of character images cut from the rubbings published in authoritative OBI descriptions. It was named OBIS163 [22], because there are 163 kinds of OBI characters in the dataset. It consists of a training dataset and a testing dataset. Rotation and deformation are used for data enhancement to equalize the data samples in each class. The training dataset uses this data enhancement method to make 250 images in each class, and the testing dataset uses this data enhancement method to make 50 images in each class.

Based on AlexNet, we have constructed a CNN network for OBI recognition named OBI-CNN [22]. OBI-CNN is a network structure for image modality alone, and our OBI recognition algorithm is a text modality incorporated on the structure of OBI-CNN by using OBS multi-modal knowledge graph.

The size of the input OBI images was 224×224, and ReLU was choosed as the activation function and it was added after the convolutional layer for data training. Partial normalization was added to replace the partial response normalization layer in the AlexNet network to improve network training and improve the generalization ability of the designed model.

We have also made some optimizations, such as add batch normalization layer, changing the square convolution kernel into the strip convolution kernel, and adjust the hyperparameters and optimize the scale transformation strategy.

OBI recognition is a multi-classification problem. Its cross-entropy loss can be expressed as follows:

$$loss = -\sum_{i}^{K} y_i \log(p_i) \tag{6}$$

Where K is the number of categories, in fact, it is the number of modern Chinese characters corresponding to the known OBI characters. In other words, in our research, a different modern Chinese character that can be recognized is a class; y is the label, that is, if the category is the i^{th} class, then $y_i = 1$, otherwise $y_i = 0$; p_i is the output of CNN, it means the probability that the category is the i^{th} class. The p_i is calculated based on softmax which is shown as follows:

$$p_i = S_i = \frac{e^{Z_i}}{\sum_{j=1}^{n} e^{Z_j}} \tag{7}$$

Where Z is the n-dimensional vector of the modern Chinese characters corresponding to the OBI characters to be recognized, and n is the number of the modern Chinese character categories. The softmax S_i means the probability that the recognized OBI character belongs to the i^{th} class.

In the experiments the GPU is Titan XP of NVIDIA, and the video memory is 16GB, with 4 pieces in total. The software environment includes CUDA 10.0 and CUDNN 10.0. In the recognition phase, we focus on accuracy and efficiency. The experimental results of OBI recognition are shown in Table 2.

Table 2. OBI recognition results.

Method	Precision (%)	Training time (hour)
DenseNet	28.12	21.16
ResNet18	30.83	17.11
VGG19	31.04	19.34
AlexNe	46.94	5.73
OBI-CNN	65.64	5.32
Ours	80.43	3.07

As shown in Table 2, our method has a great reduction in training time compared with other frameworks and also has an absolute advantage in accuracy. In particular, compared with OBI-CNN, our algorithm has better performance in terms of accuracy and training time after incorporating text modal information. Because we mapped the image vectors and the text vectors to a unified semantic space. The similarity of vectors was used to characterize the matching degree between images and texts. There were three reasons why the recognition accuracy has not reached a very high level. First, the OBI detection itself did not achieve high accuracy, resulting in low accuracy in subsequent recognition. Second, the effect of prediction for incomplete and unclear OBI characters depends on the scale of the character vectors and the prediction accuracy of CBOW. Last, the unknown OBI characters are not listed as other categories separately, so they could only be selected in the designated categories.

7. CONCLUSION

We have accumulated a large amount of basic OBI data and established a big data sharing platform. Based on this platform, we proposed an application pyramid model for OBI information processing. We also constructed an OBS knowledge graph based on the multi-modal data to better represent, share and reuse the OBS knowledge, aiming to break the bottleneck of OBS research. We conducted experiments with OBI detection and recognition as examples and achieved good results, thus verifying the advantages of OBS multi-modal knowledge graph. However, at present, it is not able to deal with the bones or shells which contain too many incomplete characters. Besides, for the unknown OBI characters, it is still unable to complete the recognition task well, even if some characters can be detected. Moreover, we have not yet implemented the knowledge recommendation system based on the OBS multi-modal knowledge graph. In our next research plan, we will expand the scale of the OBS corpus and consider better feature representation or feature learning methods for these problems, such as to mask the unrecognized characters and predict them by constructing the Bert corpus of OBI (we call it OBIBert). The prediction results of OBIBert will be provided to the experts as clues to help them find new approaches to OBI interpretation. And the knowledge recommendation system can also help OBI experts to obtain useful information for the interpretation of OBI.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (U1504612, U1804153), the Chinese Ministry of Education and National Language Commission Special Project for Research and Application of Oracle Bone Inscriptions and other ancient characters (YWZ-J023, YWZ-J010), the Development Projects of Henan Province Science and Technology, China (182102310039, 202102310562), the Major Commissioned Project of National Social Science Foundation, China (16@ZH017A3), and the Program for Ministry of Education Innovation Research Team, China (2017PT35).

REFERENCES

[1] Jing Xiong, Guoying Liu, Yongge Liu, Mengting Liu, "Oracle Bone Inscriptions information processing based on multi-modal knowledge graph, " Computers & Electrical Engineering 92, 107173 (2021).

[2] Jing Xiong, Qingju Jiao, Guoying Liu, Yongge Liu, "Oracle bone inscriptions big knowledge management and service platform," In International Conference of Pioneering Computer Scientists, Engineers and Educators, 124-136 (2019).

[3] Jici Xing, Guoying Liu, Jing Xiong, "Oracle bone inscription detection: a survey of oracle bone inscription detection based on deep learning algorithm," In Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing, 1-8 (2019).

[4] X. S. Shi, Y. J. Huang, Y. G. Liu, "Location method of rubbing text of oracle bone inscriptions based on threshold segmentation and morphology," Journal of Beijing Information Science and Technology University 29(6), 7-10, 2014.

[5] H. E. Ying, H. E. Xiaoju, G. Zhang, "Image segmentation of jiaguwen image based on wavelet transform and fcm," Journal of Tianjin University of Science and Technology 33(6), 62-66 (2018).

[6] Ye Liu, Hui Li, Alberto Garcia-Duran, Mathias Niepert, Daniel Onoro-Rubio, David S Rosenblum, "Mmkg: multi-modal knowledge graphs," In European Semantic Web Conference, 459-474 (2019).

[7] WX Wilcke, Peter Bloem, Victor de Boer, RH van t Veer, FAH van Harmelen, "End-to-end entity classification on multimodal knowledge graphs," arXiv preprint arXiv:2003.12383, (2020).

[8] Yuke Zhu, Ce Zhang, Christopher Ré, Li Fei-Fei, "Building a large-scale multimodal knowledge base system for answering visual queries," arXiv preprint arXiv:1507.05670, (2015).

[9] Feiran Huang, Xiaoming Zhang, Zhonghua Zhao, Jie Xu, Zhoujun Li, "Image-text sentiment analysis via deep multimodal attentive fusion," Knowledge-Based Systems 167, 26-37 (2019).

[10] Xing Xu, Huimin Lu, Jingkuan Song, Yang Yang, Heng Tao Shen, Xuelong Li, "Ternary adversarial networks with self-supervision for zero-shot cross-modal retrieval," IEEE transactions on cybernetics 50(6), 2400-2413 (2019).

[11] Jiasen Lu, Dhruv Batra, Devi Parikh, Stefan Lee, "Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks," arXiv preprint arXiv:1908.02265, (2019).

[12] Huimin Lu, Yujie Li, Min Chen, Hyoungseop Kim, Seiichi Serikawa. Brain intelligence: go beyond artificial intelligence," Mobile Networks and Applications, 23(2):368–375, 2018.

[13] Guoying Liu, Feng Gao, "Oracle-bone inscription recognition based on deep convolutional neural network," Journal of Computers 13(12), 1442-1450 (2018).

[14] Mengting Liu, Guoying Liu, Yongge Liu, Qingju Jiao, "Oracle-bone inscription recognition based on deep convolutional neural network," Journal of Image and Graphics 8(4), 114-119 (2020).

[15] Changjiang Qin, Hanqing Hou, "Mapping knowledge domain: A new field of information management and knowledge management," Journal of Academic Libraries 27(1), 30-37 (2009).

[16] Wanxiang Che, Zhenghua Li, Ting Liu, "LTP: A Chinese Language Technology Platform, " In Proceedings of the Coling, 13-16 (2010).

[17] Jing Xiong, Qingju Jiao, Yuntong Liu, "Oracle bone studies knowledge graph construction based on multi-source heterogeneous data," Journal of ZheJiang University(Science Edition) 47(2), 131-141 (2020).

[18] Joseph Redmon, Ali Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, (2018).

[19] Jing Xiong, Feng Gao, Qinxia Wu, "Research on semantic mining for large-scale Oracle bone inscriptions foundation data, " Data Analysis and Knowledge Discovery 31(2), 7-14 (2015).

[20] Yan Z, Guoliang L, Jianhua F, "A survey on entity alignment of knowledge base," Journal of Computer Research and Development 53(1), 165 (2016).

[21] T Mikolov, K Chen, G Corrado, J Dean, "Efficient estimation of word representations in vector space, " arXiv preprint arXiv:1301.3781, 2013.

[22] Mengting Liu, "Research on OBI Recognition Based on Deep Convolutional Neural Network," Master, Zhengzhou University, (2020).