WuMKG: a Chinese painting and calligraphy multimodal knowledge graph

Jing Wan[1*], Hao Zhang[1], Jun Zou[1], Ao Zou[1], Yubin Chen[1], Qingyang Zeng[1], Xinrong Li[1] and Qiya Wang[1]

Abstract

Chinese Painting and Calligraphy (ChP&C) holds significant cultural value, representing integral aspects of both Chinese culture and global art. A considerable amount of ChP&C works are dispersed worldwide. With the emergence of digital humanities, a vast collection of cultural artifacts data is now available online. However, the online databases of these artifacts remain decentralized and diverse, posing significant challenges to their effective organization and utilization. Addressing this, our paper focuses on the Wu Men School of Painting and proposes a framework for constructing a multimodal knowledge graph for the ChP&C domain. We construct the domain ontology by analyzing the ChP&C knowledge schema. Then, we acquire knowledge from diverse data sources, including textual and visual information. To enhance data collection around collecting historical context and subject matter, we propose methods for seal extraction and subject extraction specific to ChP&C. We validate the effectiveness of these methods on the constructed dataset. Finally, we construct the Wu Men Multimodal Knowledge Graph (WuMKG) and implement applications such as cross-modal retrieval, knowledge-based question-answering and visualization.

Keywords Multimodal knowledge graph, Chinese painting and calligraphy, Digital humanities

Introduction

Chinese **P**ainting and **C**alligraphy (ChP&C) is pivotal in traditional Chinese culture, significantly contributing to world art history. Beyond its aesthetic appeal, ChP&C embodies cultural richness and historical narratives, reflecting the profound heritage of China. ChP&C has provided valuable sources for historians and archeologists to study the history and humanity of the corresponding eras [1]. The Wu Men School of Painting, representative of the mid-Ming Dynasty in China, is one of the most extensive, accomplished, and influential painting schools in Chinese art history.

ChP&C collections are preserved in numerous museums around the world. In 2009, Chinese Calligraphy was officially recognized and included in the Representative List of the Intangible Cultural Heritage of Humanity by UNESCO's Intergovernmental Committee for Safeguarding the Intangible Cultural Heritage, promoting the safeguarding and transmission of ChP&C. The development of digital humanities has made extensive information on ChP&C freely accessible online. However, the invaluable ChP&C data are under-utilized, stored in different databases, and dispersed on official websites, which hampers the comprehensive understanding and research of ChP&C and limits the potential for academic collaboration and cultural exchange.

In recent years, Knowledge Graph (KG) has emerged as a novel method for organizing digital resource knowledge, providing new semantic insights. Additionally, Multi-modal Knowledge Graph (MKG) integrates various media types, including images, audio, and videos, which enhances the connection between the graph and the real world. MKG has become a key topic in recent research. IMGpedia, Visual Sem, and Richpedia [2–4] are typical multi-modal knowledge graphs.

*Correspondence:
Jing Wan
wanj@mail.buct.edu.cn
[1] College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China

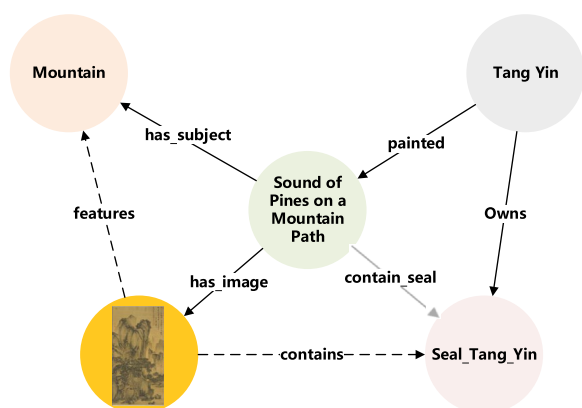Wan *et al. Heritage Science*   (2024) 12:159

Page 2 of 18

Fig. 1 Sound of pines on a mountain path[山路松声]: a Chinese painting entity in a multimodal knowledge graph. The dashed line denotes the relation extracted from the image
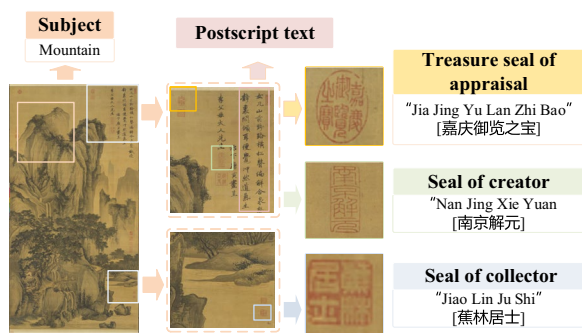

Fig. 2 An example of "Sound of pines on a mountain path". It features subjects, postscript and three types of seals

Digitizing cultural heritage facilitates reliable preservation and storage of cultural content in digital repositories [5, 6]. The advancement of the knowledge graph provides innovative approaches for knowledge organization and intelligent applications in the domain of cultural heritage. A common method for constructing a knowledge graph involves linking relevant knowledge to specific topics. As Fig. 1 illustrates, this can be achieved by developing the graph around a specific topic and extending it to relative knowledge. Currently, there is no specific ontology tailored for ChP&C. Consequently, utilizing a knowledge graph for researching and preserving ChP&C remains a challenge.

The research value of ChP&C is manifested in its rich and diverse visual information. As shown in Fig. 2, it includes subjects, postscript, and seals. The subject represents the primary elements of the painting. The postscript, closely related to the content of the painting, typically includes critiques and narrative records.

Seals are applied by either the artist or collectors. These are primarily divided into three types: the treasure seal of appraisal, the creator's seal, and the collector's seal. They explicitly reveal the painting's provenance and collecting history [7]. The primary challenge remains in accurately extracting and analyzing visual elements. Analyzing traditional Chinese paintings for automated detection and recognition is challenging due to their stylized and abstract approach, which lacks realistic details and emphasizes conveying the essence of subjects. Cultural interpretations of objects and symbols in these paintings require background knowledge, posing difficulties for algorithms and viewers unfamiliar with these contexts. Moreover, variability in conditions, preservation status, and medium further complicate target detection algorithms, which are impacted by factors such as fading colors, paper deterioration, and ink spread. Additionally, there is a scarcity of publicly available, high-quality datasets specifically tailored for traditional Chinese paintings.

To address these challenges, we present a framework for constructing a multi-modal knowledge graph in the cultural heritage domain. We construct the **Wu** **M**en **M**ultimodal **K**nowledge **G**raph (**WuMKG**) and develop associated applications. We design the ChP&C ontology based on CIDOC CRM [8] and Shlnames [9], which expands classes, attributes, and vocabulary. We refine descriptions of persons and artworks using multi-source heterogeneous data. Based on diverse data sources, we design methods to extract textual and visual knowledge including persons, ChP&C artworks, seals, and subjects. Then, we integrate and create associations between images and text to construct a multi-modal knowledge graph, which consists of 104,374 entities and 418,450 triples related to the Wu Men School. To facilitate the public's understanding and appreciation of ChP&C, we have developed a web-based system based on the WuMKG. This system implements features including cross-modal retrieval, a question-answering system (Q &A system), and KG visualization. These capabilities greatly enhance the knowledge discovery services for ChP&C and further advance its protection, inheritance, development, and utilization.

Our contributions can be summarized as follows:

- We propose a framework for constructing a multi-modal knowledge graph in the ChP&C domain. We analyze the knowledge schema in this domain and design the ontology based on CRM and Shlnames. Moreover, we introduce a method for knowledge acquisition, which includes extracting information from textual and visual data of ChP&C.
- We design distinct methods for seal extraction and Chinese painting subject detection. We have con-

Wan *et al. Heritage Science*   (2024) 12:159

Page 3 of 18

structed datasets specifically for seal extraction and subject detection, which consist of 8363 and 5182 samples respectively. Empirical experiments validate their effectiveness.

- We construct the WuMKG, which includes approximately 1200 artists and 12,000 works of painting and calligraphy. The knowledge graph and related datasets are released in OpenKG with the SPARQL endpoint provided. Additionally, we develop a web-based system based on the WuMKG.

Related work

In this section, we review related work on metadata and ontologies within the cultural heritage domain, investigate multimodal knowledge graphs, and explore their applications in this area. Additionally, we review image processing research related to ChP&C.

Cultural heritage metadata and ontology

In the realm of cultural heritage preservation, metadata and ontology play pivotal roles in facilitating the organization, retrieval, and understanding of historical artifacts and information. Metadata serves as the backbone for cataloging and describing cultural heritage items, capturing vital details such as provenance, condition, and historical significance. Ontology, on the other hand, provides a formal representation of concepts, entities, and relationships within a given domain, enabling a more structured and semantic understanding of cultural heritage data.

The early developments in metadata standards for cultural heritage occurred in the United States and Europe. The Getty Vocabulary Program [10] has categories for describing artistic works and data value standards for art, architecture, and material culture information. In addition, there are other standards like the Dublin Core Metadata Element Set [11], which describes general resources on the web, VRA core [12] categories for visual materials, and Information Object Definitions for exchanging cultural heritage content [13].

The Categories for the Description of Works of Art (CDWA) provides a framework to ensure consistent and comprehensive descriptions, including titles, creators, dates, materials, and styles, as well as the historical and cultural context of the works. [14]. Bobasheva et al. developed a pipeline using deep learning to analyze and correct metadata. Their method identifies contextual relationships between concepts in metadata by analyzing semantic models and concept store similarities [15]. The International Image Interoperability Framework (IIIF) is an innovative metadata solution that offers a standardized method for describing and delivering images on the web, facilitating the development of digital cultural heritage [16]. Xue et al. implemented the representation of the Chinese Seal Stamping Catalogs using IIIF and Serverless [17].

In the domain of cultural heritage ontology modeling, noteworthy contributions have been made by the International Committee for Documentation (CIDOC) CRM and the European Data Model (EDM) [8, 18]. The Getty Vocabulary Program has also been extended to accommodate Semantic Web representations. The Shanghai Library maintains the China Biographical Database Project (CBDB) ontology and also developed a person name standard ontology(Shlnames) to support the description of Chinese individual information [19].

Matusiak et al. presented a model of creating bilingual parallel records that combines translation with controlled vocabulary mapping, which addresses the linguistic needs of the target audience, connects the digitized objects to their respective cultures and contributes to richer descriptive records [20].

Ontologies go beyond metadata's capabilities in cultural heritage by structuring complex relationships and hierarchies, enabling advanced data integration and knowledge representation. This structured, semantic framework allows ontologies to model intricacies that metadata alone cannot, facilitating deeper and more systematic discovery and management of knowledge.

Although there are some ontology practices in the cultural heritage domain, research specifically related to ChP&C remains relatively limited, often concentrating on modeling the components of these art forms. Comprehensive modeling of ChP&C is not only crucial but also presents significant challenges. In our research, we thoroughly analyze the knowledge schema of ChP&C and design an ontology for it. Utilizing this ontology, we construct the WuMKG, which has demonstrated the effectiveness of the ontology.

Cultural heritage knowledge graph

Extensive research and practice have been conducted on constructing knowledge graphs in the cultural heritage domain.

The British Museum released open-linked data on collections of cultural relics based on CIDOC CRM and developed a semantic retrieval system [21]. The Museum of the Netherlands constructed open-linked data based on EDM ontology. The Europeana project integrates data from more than 4000 museums, libraries and archives in Europe, and its knowledge graph contains 3.7 billion triples [22]. There are also the Italian cultural heritage knowledge graph ArCo [23], the Finnish World War II knowledge graph WarSamp [24] and the Nuremberg urban cultural heritage knowledge graph [25], etc.

Wan *et al. Heritage Science*   (2024) 12:159

Page 4 of 18

Chinese researchers have also conducted research and exploration in recent years. Based on Dong Qichang's collection of image data and related digital resources, the Shanghai Museum sorted out his cultural context, built related data based on the CIDOC CRM, and displayed Dong Qichang's creative history in a visual form [26]. Related research also includes the knowledge graph of Dunhuang murals and the knowledge graph of Fan Qin's life [27].

In terms of multi-modal knowledge graphs, application exploration has also been carried out in the field of cultural heritage. Thomas et al. [28] combined expert knowledge and a knowledge graph containing images to propose a method for retrieving and recommending European silk fabrics. Puren et al. [29] proposed a knowledge representation method for silk text images. Fan et al. [30] constructed a multi-modal knowledge graph of Chinese intangible cultural heritage and used visual features to select representative images for entities.

Diverging from prior works, we introduce a novel framework specifically tailored for the construction of a multimodal knowledge graph for ChP&C. We construct a domain ontology by analyzing the knowledge schema. Additionally, we collect heterogeneous data from museums worldwide, extracting relevant ChP&C knowledge. Notably, addressing the unique characteristics of ChP&C, we design methods for seal extraction and subject extraction to extract knowledge from images.

Image processing on ChP&C

Chinese paintings, known for their diverse styles, have been a subject of exploration for researchers in the realm of artificial intelligence (AI). This section delves into the existing literature and highlights noteworthy contributions in multidimensional feature extraction, classification, and sentiment recognition tasks related to Chinese paintings.

Liong et al. [31] constructed a comprehensive traditional Chinese painting dataset featuring six distinct classes. Leveraging deep learning methodologies, particularly convolutional neural networks (CNNs), they pursued classification objectives. Furthermore, the researchers employed a series of instance segmentation techniques to discern discriminant characteristics in Chinese paintings. Zhang et al. [32] focused on the development of a model capable of learning aesthetic features from both global attribute maps and localized patches. The derived features, coupled with manually crafted features rooted in the expertise of art professionals, culminated in creating a comprehensive computational model for the beauty evaluation of Chinese ink paintings. Recognizing the challenge of CNNs underperforming in classification tasks with limited sample sizes,

Li et al. [33] introduced a novel solution. They proposed a multi-instance learning algorithm, ALSTM-MIL, which employs a Long and Short-Term Memory neural network with an attention mechanism. It transformed Chinese painting images into multi-instance bags. The application of ALSTM facilitated the extraction of encoded features, enabling semantic classification of Chinese paintings.

Research related to Chinese paintings primarily focuses on feature extraction and classification. However, the presence of multiple distinct objects within a painting indicates that it possesses diverse subjects. To leverage this distinctive feature, we employ an object detection approach for subject detection, which enables the nuanced content representation in Chinese paintings. Differing from previous studies, we incorporate image processing technology into our framework. This involves extracting seals and subject objects from images of ChP&C artworks to facilitate knowledge acquisition, thereby providing a novel perspective for constructing multimodal knowledge graphs.

Methodology for WuMKG construction

In this section, we will introduce the framework for constructing the WuMKG. As illustrated in Fig. 3, the framework consists of four parts: data sources, ontology design, knowledge acquisition and applications.

Data sources We collect data from various sources, including the websites of museums, the website of the National Cultural Heritage Administration, specialized books, encyclopedias, and seal databases, to gather comprehensive data.

Ontology design We analyze the domain knowledge schema. We reuse the CIDOC CRM and Shlnames ontology, extending the ontology by adding new classes, relationships, and terms.

Knowledge acquisition Our aim is to extract structured data from various sources of textual and image data, transforming it into knowledge. Initially, we pre-process the collected data to facilitate the extraction task. We extract entities, relationships, attributes, and events from the text. Additionally, we extract seals and subjects from images to acquire historical and content-related data. Finally, we apply knowledge fusion and data mapping techniques to construct the WuMKG.

Applications   We release the SPARQL endpoint and build a web-based software platform based on the WuMKG, enabling multimodal retrieval, a question-and-answer system, and visualization applications.

Ontology of WuMKG

In this section, we will present the design methodology employed for the WuMKG ontology. To create a well-organized and domain-specific ontology, we
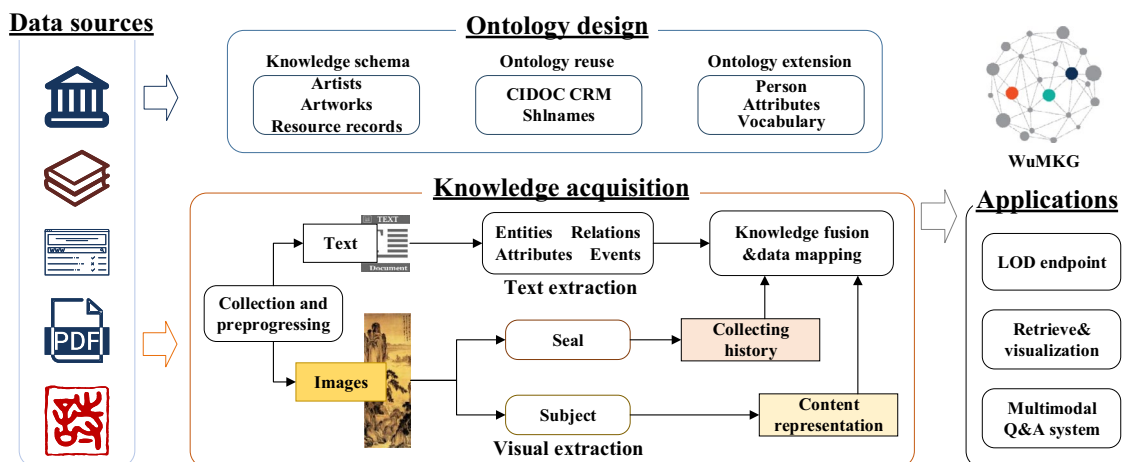
Wan *et al. Heritage Science*   (2024) 12:159

Page 5 of 18

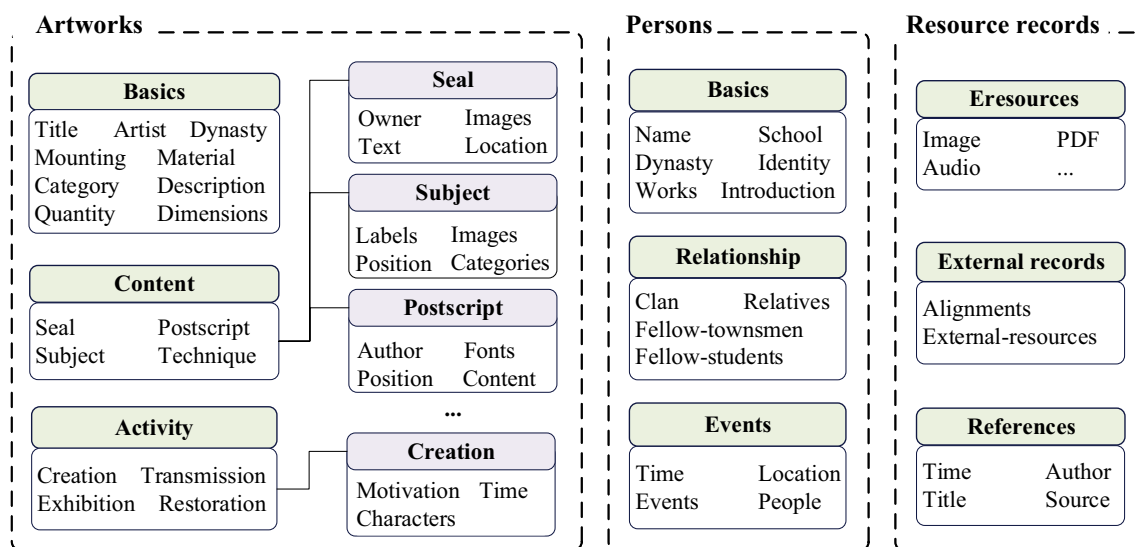**Fig. 3** The framework of WuMKG construction

Fig. 4 Knowledge schema of WuMKG

integrate guidance from art theory, knowledge schema design, ontology design, and expert review. Historical art insights and principles guide the data structuring, ensuring that the resulting ontology authentically reflects the distinctive attributes and deep cultural contexts of the art styles. We construct a knowledge schema after analyzing the ChP&C data from various resources. Afterward, we incorporate existing domain ontologies, such as CIDOC CRM and Shlnames, mapping objects, elements, and modifiers from the schema to the foundational classes, relationships, and properties to form the basic ontology for ChP&C. Subsequently, we expand the person ontology to encompass

various character relationships. The final work is validated by experts.

Knowledge schema

We collect and analyze existing public data on the internet, as well as books and literature related to ChP&C, to determine the range and content of data for the ChP&C knowledge graph. After a thorough examination and consideration of current metadata sources related to cultural heritage, we construct a knowledge schema from the available data. We also seek insights from experts to ensure that the schema accurately reflects domain-specific knowledge and practices. As shown in Fig. 4,

Table 1 Examples of painting and calligraphy information

Type	Attr.	Desc.
Basics	Title	The sound of pines on the mountain road, Tang Yin, Ming Dynasty [明唐寅画山路松声]
	Author	Tang Yin[唐寅]
	Dynasty	Ming Dynasty[明]
	Material	Silk[本幅：绢]
	Dimensions	Width 194.5 cm, Height 102.8 cm[本幅：宽194.5cm，高102.8cm]
	Introduction	This piece was created as a farewell gift for Li Jing, the magistrate of Wu County, and was painted in the eleventh year of Zhengde. [本幅是为吴县知县李经赠别之作，画于正德十一年。…]
Content	Subject	Landscapes[山水], Flowers[花茸], Architecture[建筑], Vessel[器用], People[人物]
	Techniques	Freehand Brushwork[写意], Chaffing[皴法], Moss Dots[苔点]
	Postscripts	Author's Inscription[作者款识]: Tang Yin[唐寅], The main body[本幅], Regular Script[楷书]
	Seals	Collectors' Seals[收传印记]: Cangyan[苍岩], Jiaolin Collection[蕉林收藏]
		Appraisal Treasure Seals[鉴藏印记]: Xuantong Appreciation[宣统鉴赏], Jiaqing Appreciation[嘉庆鉴赏]…
Activity	Collecting History	Existing in the National Palace Museum, Taipei [现存于台北故宫博物院]

the painting and calligraphy knowledge schema mainly includes painting and calligraphy artworks, persons and resource records.

Artworks We summarize the knowledge of ChP&C into three dimensions: basics, content and activity. Taking the painting "The Sound of Pines on the Mountain Road" as an example, Table 1 shows its information. The basics of ChP&C, including the dynasty, materials, and dimensions, provide insights into their historical context and the nuances of their creation. These artworks primarily express the artist's ideals and vision through their visual impact.

The postscript denotes the calligraphic elements found within the artworks of ChP&C, encompassing aspects such as fonts and content.

The subjects of ChP&C are windows into the artist's emotions, thoughts, and aesthetic aspirations, while the employed techniques reveal their unique creative abilities and stylistic choices. Finally, the seals and other mark information in the artwork are content applied by the creator or collector, representing the process of creation and collecting history.

Persons Artists' creative styles and techniques are often influenced by their interpersonal relationships, including those with family members, teachers, and others. Additionally, significant life events can also lead to the evolution of an artist's creative style. To deeply study the creative characteristics of the characters, in addition to the basic information, the relationships and events of the characters were recorded as a reference for subsequent research. See Appendix A for examples of persons.

Resource records A large amount of knowledge in painting and calligraphy relies on records from multiple sources. Therefore, it is necessary to record information such as knowledge sources and bibliographies while introducing domain knowledge as auxiliary data for subsequent knowledge integration and research verification. The knowledge of ChP&C includes multimedia such as image resources and explanations of painting and calligraphy works, which should also be recorded as electronic resources. External resources store link relationships with other knowledge bases to enhance the connectivity between knowledge bases. See Appendix A for examples of resource records.

Ontology reuse

Reusing existing ontologies is an important principle in building ontologies, which helps to improve interoperability and avoid ambiguities and conflicts in expressions. We reuse CIDOC CRM and Shlnames ontology.

We use the CIDOC CRM as the basis of the ontology and map the concepts in painting and calligraphy to classes in the model based on semantics. The *E24_Physical_Human-Made_Thing* class in the model is the basic class for entities such as artworks, seals, and inscriptions, as shown in Fig. 5.

For resource records, classes such as *E73_Information_Object* and *E31_Document* are used to describe electronic resources, bibliographies, records, etc., and are related to entities such as works, persons, painting and calligraphy, and seals in the atlas.
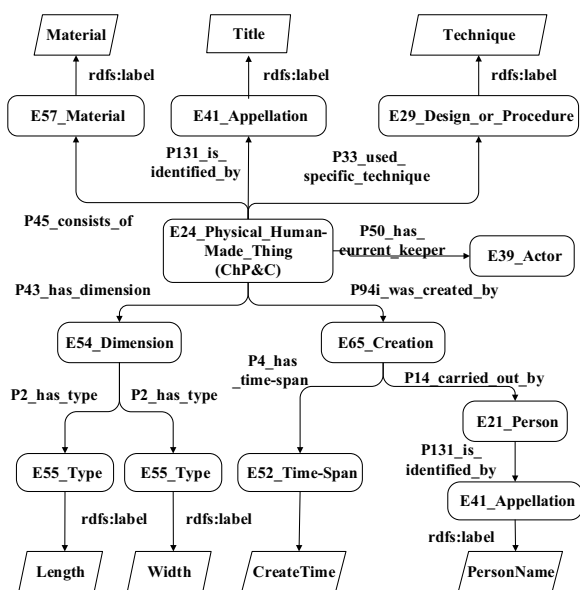
Wan *et al. Heritage Science* (2024) 12:159

Page 7 of 18

Fig. 5 Partial ChP&C ontology of WuMKG

Fig. 6 The ontology of person relationship

To describe person information, we reuse the Shlnames ontology. Shlnames is an ontology constructed by the Shanghai Library for describing basic information about Chinese personal names. For person entities, use *crm: E21_Person* class and *shl: Person* class for description. Use the *shl:courtesyName*, *shl: pseudonym*, *shl:livedAt*, and other attributes in the personal name specification library to describe the character's font size, place of residence, and other information.

Ontology extension

The current ontology struggles to effectively represent life events, social relationships, specific terms, and connections related to works. To address these issues, we have developed the person relationship ontology, which extends the classes, relationships, and attributes within the existing ontology. This expansion is accomplished through the utilization of the Ontology Web Language (OWL) and Simple Knowledge Organization System (SKOS) to precisely define terms [34]. This provides a more comprehensive and accurate relational framework for studying and analyzing artists.

As shown in Fig. 6, the detailed depiction of character relationships, encompassing kinship, master-disciple, and cooperative connections, is facilitated through the utilization of *Protégé* to expand the person class *E21_Person* within the CRM model. This expansion introduces two subclasses and delineates 38 specific types of person relationships. Additionally, the establishment of two gender classes, *EX_MALE* and *EX_FEMALE*, allows for a more nuanced representation of specific character relationships, such as sisters, brothers, husbands, wives, and more.
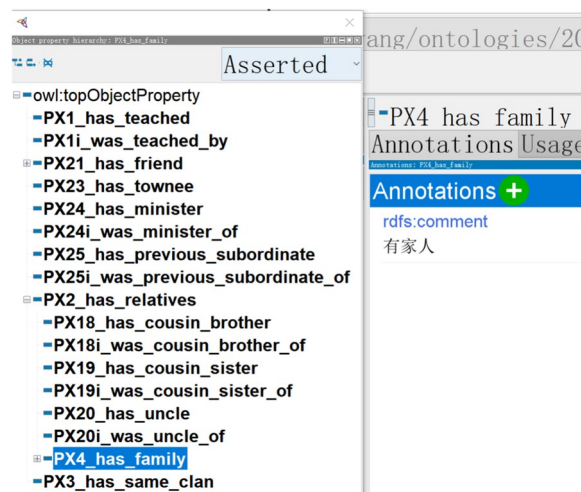
To ensure semantic consistency in the data mapping process of painting and calligraphy knowledge, we employ OWL and SKOS to construct a vocabulary of specialized terms. This vocabulary primarily encompasses terms related to painting and calligraphy types, materials, techniques, mounting, fonts, etc. The terms were systematically defined and categorized through expert validation and analysis of their expressions in data sources. The organization of these terms is based on their definitions and relationships. The terminology serves as a crucial tool for data mapping and maintaining consistency in painting and calligraphy knowledge.

Knowledge acquisition

In this section, we introduce the process of transforming textual and image data from various sources into knowledge within the WuMKG. Following the construction framework, knowledge acquisition is divided into three main components: collection &preprogressing, knowledge extraction, and knowledge fusion. Specifically, knowledge extraction includes text extraction, seal extraction and subject extraction.

Data collection

Our data primarily comes from professional books, museum websites, official websites, linked open data, and Baidu Baike [19, 35].

Data collection The principle of selecting data sources is based on the professionalism and reliability of the data. We collect open-access data from museum websites, National Cultural Heritage Administration websites, books and references, cultural relics dictionaries, domain
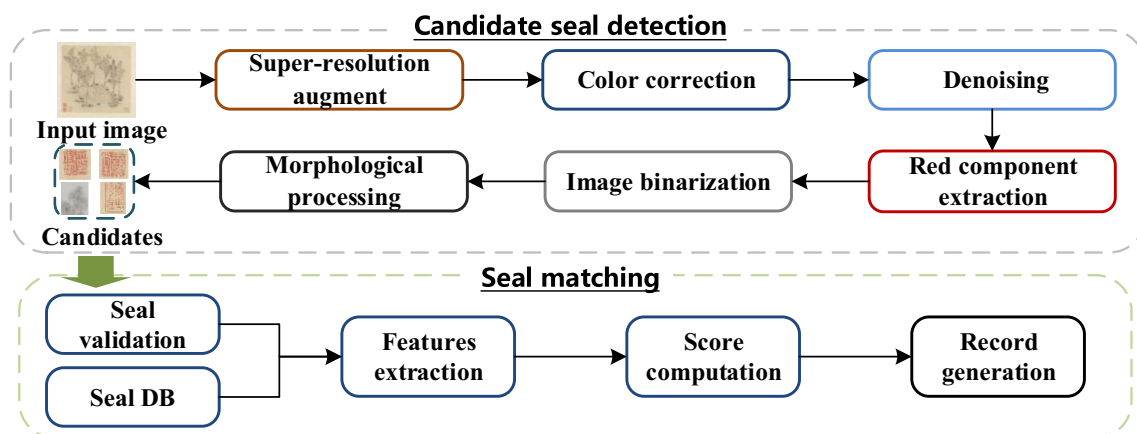
Fig. 7 The process of seal extraction

vocabularies, and the China Biographical Database Project (CBDB) [19] released by research institutions. The data encompasses structured (LOD), semi-structured (museum websites, etc.), and unstructured (PDF) formats. See Appendix B for details.

Data preprocessing We conduct data cleaning on semi-structured data, removing duplicates, and addressing issues such as erroneous characters. For PDF documents, we employed OCR for automated text recognition. Additionally, painting and calligraphy images underwent automatic processing to convert them into text.

Text extraction

The information about artists typically includes biographical overviews that provide unstructured summaries containing extensive details about their lives, works, and geographical locations. We extract relevant information from these entries such as the artists' names, important locations related to their lives, and significant artworks. These summaries are similar to general knowledge about individuals. To uncover this knowledge, we employ LeBERT [36] as our extraction model and devise an iterative annotation extraction strategy to assist in completing this task at a relatively low cost.

We used a pre-trained LeBERT model to extract entities from infoboxes and contents. Given a Chinese sentence with n characters $s_c = \{c_1, c_2, ..., c_n\}$, the process is expressed as follows:

$$\{e_1, e_2, ...e_n\} = \text{LeBERT}(s_c) \tag{1}$$

$$\{l_1, l_2, ...l_n\} = \text{Softmax}(\{e_1, e_2, ...e_n\}) \tag{2}$$

where e_i represents the embedding for c_i, and l_i represents the label for c_i.

Domain entity extraction requires professional annotation, which incurs significant costs. To minimize labeling costs, we adopted an **iterative annotation** approach. Firstly, we use the model pretrained on Weibo Dataset for rough data annotation, creating a coarse annotation. Subsequently, human reviewers filtered and corrected the coarse annotations to obtain a high-quality refined dataset. We then fine-tuned the model using the refined dataset and performed more information extraction.

We conducted experiments to validate the effectiveness of iterative annotation and examined the correlation between the quantity of finely annotated data and extraction performance.

Seal extraction

The seals on ChP&C serve as distinctive marks for both the original author and subsequent collectors. Extracting these seals enables the study of the origins and circulation history of ChP&C. Analyzing the provenance of ChP&C involves extracting and matching these seals. We propose a method for seal extraction which consists of two stages: **candidate seal detection** and **seal matching**, as shown in Fig. 7.

Candidate seal detection

Here, we detect and segment the potential positions of seals within paintings and calligraphy. The procedure involves super-resolution augment, color correction, image binarization, and morphological processing. These sequential steps collectively contribute to identifying candidate regions likely to contain seals.

Super-resolution augment Artworks can fade and flake due to aging, making details like seals difficult to identify. We adopt BSRGAN, a super-resolution technology [37], to enhance image details actively. Using
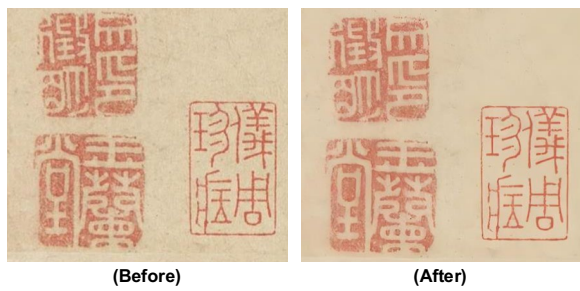
Wan *et al. Heritage Science*   (2024) 12:159

Page 9 of 18

(Before)   **(After)**
Fig. 8 Super-Resolution augment with BSRGAN: a visual comparison

(Red channel extracted)   **(Binarization)**
Fig. 9 Red channel extraction and binarization result

BSRGAN, we upscale each input image V_{in} to obtain augmented image V:

$$V = \text{BSRGAN}(V_{in}) \tag{3}$$

It aims to magnify the details of the seal area, facilitating improved identification in subsequent steps. The result is shown in Fig. 8.

Color correction Color deviations may arise between the captured image and the actual scene due to various factors, including lighting conditions and preservation conditions. For augmented image $V = (R, G, B)$, we employ white balance techniques to mitigate color bias [38]. The Gray World Algorithm serves as the primary method for implementing white balance. It determines the average color of the image to estimate the global color shift, pinpointing channel-specific deviations. It is used to calculate gain coefficients for each channel, aiding in the adjustment of white balance. For channel C, C_{avg} denotes the average values. k and k_c denote the identical grayscale value and gain coefficients of channel C. It is expressed as follows:

$$k = \frac{R_{avg} + G_{avg} + B_{avg}}{3} \tag{4}$$

$$k_c = \frac{k}{C_{avg}} \tag{5}$$

$$C_c = k_c \cdot C \tag{6}$$

where R_{avg}, G_{avg}, and B_{avg} represent the average values of the red, green, and blue channels, respectively. C_c corresponds to the component values of R, G, B channels in the image after white balance correction, and these components collectively form the corrected image denoted as V_c.

Denosing We apply a Gaussian kernel to reduce interference and enhance seal shape [39]. For image V_c, it adjusts pixel values with Gaussian-distributed weight coefficients to obtain the denoised image V_d. It is expressed as follows:
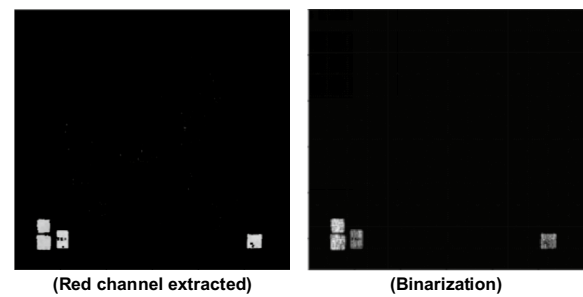

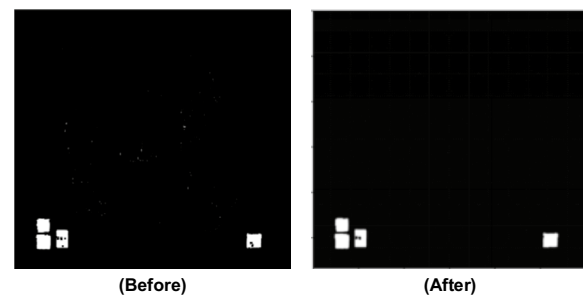**(Before)**   **(After)**
Fig. 10 Morphological processing before and after comparison

$$P_d(x, y) = \sum_{j=y-n}^{y+n} \sum_{i=x-n}^{x+n} w(i, j) \cdot P_c(i, j) \tag{7}$$

where $P_c(x, y)$ and $P_d(x, y)$ represent the pixel value at position (x, y) in the image V_c and the denoised image V_d, respectively. The weight coefficient $w(i, j)$ at each position is defined as:

$$w(i, j) = \frac{1}{2\pi\sigma^2} \cdot e^{-\frac{(i-n)^2 + (j-n)^2}{2\sigma^2}} \tag{8}$$

where σ represents the standard deviation of the Gaussian kernel, and n is the radius of the Gaussian kernel.

Image binarization The process involves keeping only the potentially seal-bearing red regions. The results are illustrated in Fig. 9. Converting the image to the HSV color space, which aligns with human color perception, enhances the extraction of the red component for improved identification of potential seal areas.

$$V_b = \text{Binarization}(V_d) \tag{9}$$

where V_b denotes the binarized image.

Morphological Processing The binarized image displays noise holes, impacting both the content within the seal and the smoothness of the seal's edges. This impedes the effectiveness of seal extraction. To address this issue, we utilize morphological processing,

Wan *et al. Heritage Science* (2024) 12:159

Page 10 of 18

Table 2 EfficientNet-B0 Architecture

Stage	Operator	Resolution	Layers
1	Conv3x3	$224 \times 224 \times 32$	1
2	MBConv1, k3x3	$112 \times 112 \times 16$	1
3	MBConv6, k3x3	$112 \times 112 \times 24$	2
4	MBConv6, k3x3	$56 \times 56 \times 40$	2
5	MBConv6, k3x3	$28 \times 28 \times 80$	3
6	MBConv6, k3x3	$14 \times 14 \times 112$	3
7	MBConv6, k3x3	$14 \times 14 \times 192$	4
8	MBConv6, k3x3	$7 \times 7 \times 320$	1
9	Conv1x1 & Avg Pooling, FC	$7 \times 7 \times 1280$	1

specifically employing dilation followed by erosion to eliminate gaps and connect fragmented regions. The results are illustrated in the Fig. 10.

Finally, we perform segmentation and extraction on candidate regions V_{cand} by determining the coordinates of the maximum rectangular area within connected components and isolating them from the original image.

Seal matching

At this stage, we validate candidate seal regions to confirm their authenticity. Subsequently, corresponding records are generated through feature matching.

Seal validation We finetune a validation model to classify candidates into actual seals or not. We adopt EfficientNet-B0 as validation model [40]. EfficientNet-B0 architecture is shown in Table 2.

For each V_{cand}, the model assigns a score that indicates the likelihood of the image being an actual seal. When the score exceeds a predefined threshold t, V_{cand} is considered a genuine seal V_{real}. It is expressed as:

$$Score_{cand} = \text{EfficientNet}(V_{cand}) \tag{10}$$

$$V_{real} = V_{cand}, \text{if } Score_{cand} > t \tag{11}$$

Features extraction We separately extract seal features using the VGG16 model pre-trained on ImageNet and the SIFT (Scale Invariant Feature Transform) method [41, 42]. These features will be utilized to construct and match the seal database. Certainly, given the seal image V, the corresponding VGG16 feature K_v and SIFT feature K_s are expressed as follows:

$$K_v = VGG16(V) \tag{12}$$

$$K_s = SIFT(V) \tag{13}$$

Seal DB For matching of candidate seals, we construct a seal database (Seal DB). We obtain VGG16 and SIFT features of pre-collected seal images. For each seal image V_i and its corresponding description I_i, we obtain VGG16 features $K_{v,i}$ and SIFT feature $K_{s,i}$ through feature extraction. The seal collection D is defined as:

$$D = \{(V_i, I_i, K_{v,i}, K_{s,i})\}_{i=1}^n \tag{14}$$

Similarity computation To retrieve seals from the database that closely match the actual seal, we calculate similarity scores between the actual seal V_{real} and each seal in the Seal DB. Here, V_{real} serves as the query to Seal DB. The similarity score for VGG16 is computed as follows:

$$Score_{VGG}(Q_v, K_v) = \frac{Q_v \cdot K_v}{\|Q_v\| \cdot \|K_v\|} \tag{15}$$

where Q_v denotes the feature of seal extracted using VGG16 and K_v denotes the features of each seal in Seal DB.

The score for SIFT is computed as follows:

$$Score_{SIFT}(Q_s, K_s) = \arg \min_{K_i \in K_s} d(Q_s, K_i) \tag{16}$$

where Q_s denotes the feature of the actual seal V_{real} extracted using SIFT and K_s denotes the features of each seal in the library D. $d(X, Y)$ represents the Euclidean distance between the two features.

The total similarity score, denoted as $Score_{total}$, is calculated as a weighted sum of the VGG16 and SIFT scores, with α serving as the weighting factor. It is expressed as follows:

$$Score_{total} = Score_{VGG} + \alpha \times Score_{SIFT} \tag{17}$$

Upon successful matching, relationships are established between the input image and the matched seals. Collecting history records are generated based on the seals recorded in the calligraphy, seamlessly integrated into our knowledge graph.

Subject extraction

Chinese paintings commonly depict a variety of subjects, including people, landscapes, trees, flowers, animals, feathers, vessels, and buildings. Despite extensive documentation of these subjects in metadata, there can be significant variations in categorization accuracy and granularity. To address this, We propose a method that utilizes object detection based on EfficientDet to extract subject-related information in Chinese paintings and associate it with corresponding entity knowledge. This approach aims to provide a more detailed and systematic
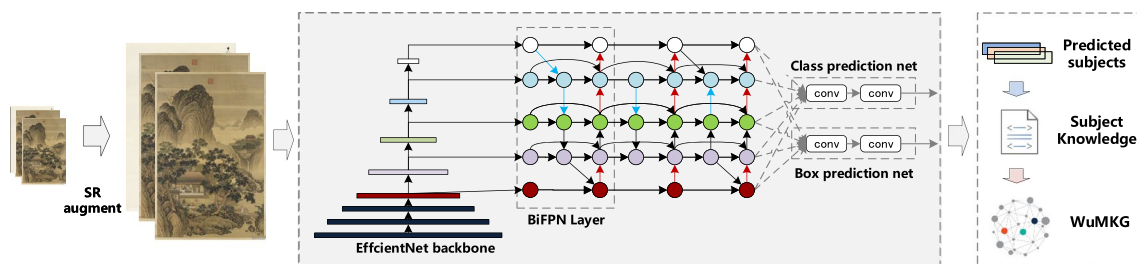
Fig. 11 The process of subject detection based on EfficientDet

classification of Chinese paintings. The process is illustrated in Fig. 11.

We use the super-resolution technology to enhance the image input V_{in}, enrich the image details, and improve the accuracy of small target detection.

$$V = \text{BSRGAN}(V_{in}) \tag{18}$$

We employ EfficientDet as our subject detection model. It consists of four components: EfficientNet backbone, BiFPN layer, class prediction network, and box prediction network. For image V, the model outputs predicted results S. It is expressed as follows:

$$S = \text{EfficientDet}(V) \tag{19}$$

$$S = \{(l_i, s_i)\}_{i=1}^{n} \tag{20}$$

where l_i and s_i denote the detected ith subject label and its corresponding confidence score, respectively. For label l_i, When the score s_i for an image V exceeds a predefined threshold, it indicates that the image contains the subject related to the label l. This implies that the artwork represented by V features the subject l. The identified subject from this detection is then converted into triples and subsequently integrated into the WuMKG.

Knowledge fusion and data mapping

In this section, we perform knowledge fusion, which includes entity alignment and conflict resolution to address duplication and conflicts in knowledge from different sources. By the established ontology, we convert the data into triples.

Entity alignment In light of the diverse representations of entities across various data sources, especially those representing individuals, calligraphy, and painting, alignment becomes essential to establish connections between entities referring to the same object [43]. Aligning person, calligraphy, and painting entities is crucial. We conducted both internal alignment and alignment with external knowledge graphs.

Internal aligment The alignment process relies on key attributes such as name, author and dynasty. We serialize entities and their respective attributes into sentences and calculate their similarities. The most similar entity pair is checked and merged into a single entity. We compute the similarity of attributes. Measures derived from the analysis of painting and calligraphy images. Notably, the domain vocabulary encompasses various aliases for individuals (e.g. Tang Yin [唐寅], courtesy name Bohu [字伯虎], and Liuru Jushi [号六如居士]). Therefore, aligning person entities primarily draws on the domain vocabulary, utilizing attributes like alias, dynasty, and residence for similarity calculations.

External aligment To achieve alignment with person entities in external graphs, we employ matching features such as person name, alias, dynasty, etc. The similarity with character entities in external knowledge graphs like DBpedia and CBDB is calculated, and the result with the highest similarity is used as the alignment. Subsequently, the owl: SameAs relationship is employed to link the aligned entitY pairs.

Conflict resolution Conflict resolution is used for merging and proofreading person attributes. Using the person details in *History of Wu Men Painting* as the references, the biography in CBDB is integrated, and the character data in the encyclopedia is used as a supplement. When inconsistent data is encountered, priority is given to information with higher frequency and reliability.

Data mapping Semi-structured data is converted into structured data automatically, and then data mapping is used to complete the structured data conversion into triples. As shown in Fig. 12, the software tool Karma is used to complete this work [44]. First, to generate a mapping model, we establish the mapping model between structured data samples and ontology classes, relationships, and attributes. Then, the model is executed to obtain knowledge graph data in the form of triples.

Evaluation of knowledge extraction

In this section, we conduct extensive experiments to evaluate the performance of the proposed methods and report detailed experimental results.
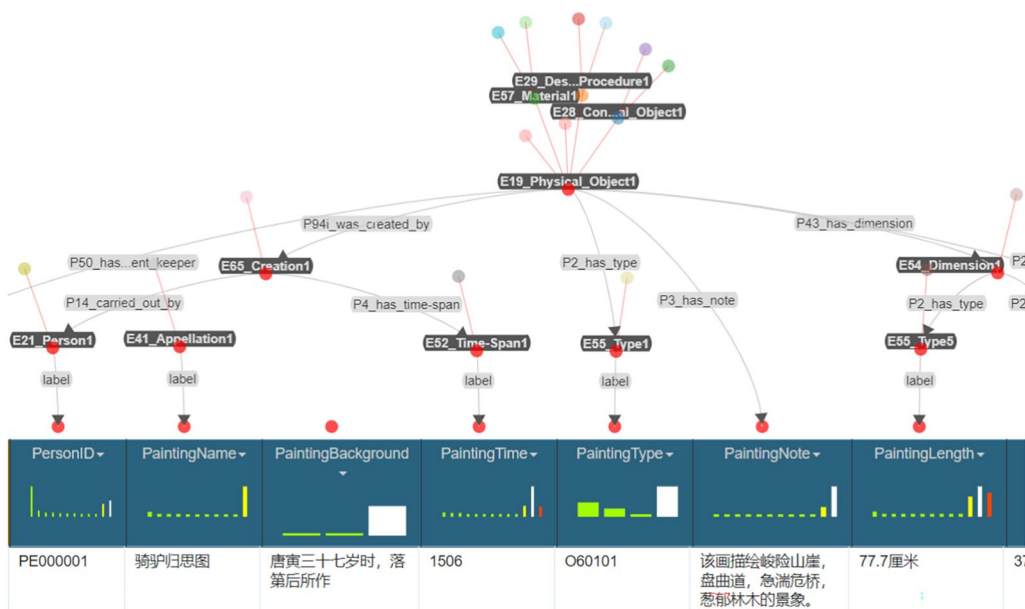
Fig. 12 Using Karma for data mapping

Table 3 Entity extraction result

ratio	PER			GPE			TITLE			total F1
	P	R	F1	P	R	F1	P	R	F1	
base	67.14	54.18	59.97	22.22	1.81	3.35	–	–	–	41.45
1x	88.77	94.76	91.67	54.39	65.99	59.63	17.83	12.43	14.65	75.15
2x	91.12	94.15	92.61	72.64	85.49	78.54	75.78	65.95	70.52	89.24
3x	91.05	95.04	93.00	81.22	87.30	84.15	79.62	67.57	73.10	90.63

Text extraction experiment

In this experiment, we assess the effectiveness of iterative labeling on the text extraction performance.

Dataset We conduct experiments on the *Baidu Baike* dataset to assess the efficacy of iterative annotation. The dataset consists of the infobox and content text of artists in *Baidu Baike.* Following the BIO annotation scheme, entities were annotated with PER (Person), GPE (Location), and TITLE (Artwork name) tags. B and I denote the beginning and inside parts of each entity. The meticulously annotated data was partitioned into training and testing sets at a ratio of 3:7. Additionally, we randomly selected 33%(56 samples), 66%, and 100% of samples from the training set for training, denoted as 1x, 2x and 3x, respectively.

Implementation Details Initially, we employed LeB-ERT pretraining on the Weibo NER dataset, establishing it as our baseline. Subsequently, the model is fine-tuned using 33%, 66%, and 100% of the training set separately for 10 epochs, with a batch size of 32 and a learning rate of 1×10^{-4}.

Evaluation Metrics We use three evaluation metrics, precision (P), recall (R) and $F1$ value. The metrics are defined as follows:

$$P = \frac{TP}{TP + FP} \tag{21}$$

$$R = \frac{TP}{TP + FN} \tag{22}$$

$$F1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{23}$$

Results The entity extraction results are shown in Table 3. The model achieves excellent extraction performance on domain text even with small amounts of downstream task data for fine-tuning. Even for data
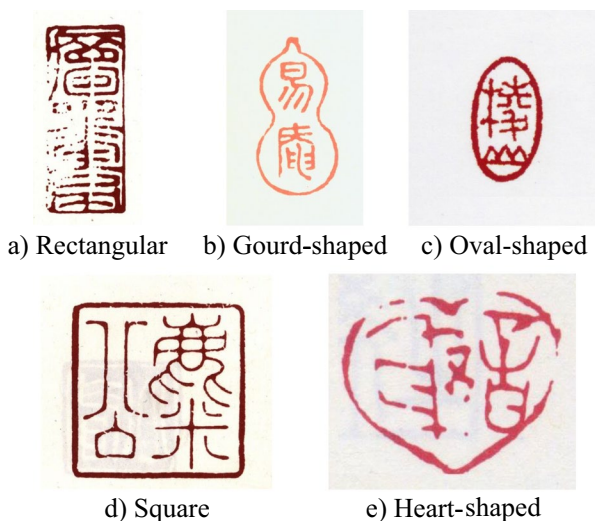
Wan *et al. Heritage Science*   (2024) 12:159

Page 13 of 18

a) Rectangular   b) Gourd-shaped   c) Oval-shaped

d) Square   e) Heart-shaped

Fig. 13 Examples of seals in various shapes

Table 4 Statistics for seal dataset

Datasets	Positive	Negative	Total
Train	3460	2396	5856
Dev	494	342	836
Test	988	683	1671

Table 5 Seal detection experimental results

Model	Param/M	P (%)	R (%)	F1 (%)
AlexNet	61.1	82.56	97.88	89.59
GoogLeNet	6.8	85.03	97.93	91.03
ResNet-50	25.6	99.29	99.80	99.54
Ours	5.3	99.09	99.69	99.28

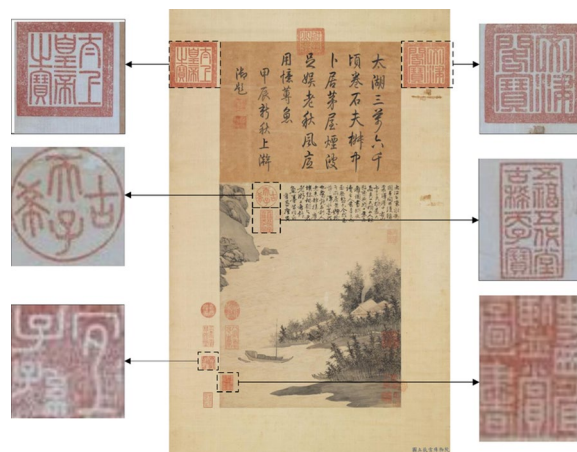**Fig. 14** An example of seal extraction

Table 6 Statistics for subject detection dataset annotations

Class	Instances	Class	Instances
Person	607	mountain	679
Fish	1015	beast	517
Insect	290	bird	692
Flower	1382		
Total	5182		

not previously labeled, such as TITLE, fine-tuning with a small amount of data still yields effective extraction results. The model's performance with 2x trainset nearly matches that achieved with 3x trainset. This indicates that our iterative annotation method can efficiently extract high-quality text at a significantly reduced cost. Additionally, it showcases the method's potential applicability in text extraction tasks across various domains.

Seal extraction experiment

We construct a seal dataset to evaluate the classification performance of our approach.

Dataset The dataset, as detailed in Table 4, comprises 8363 seal images, which are divided into training, development, and test sets with a ratio of 7:1:2. The seal images, sourced from the Shanghai Library's open data platform and manual annotations of candidate seals, include various shapes as depicted in Fig. 13. To enhance the diversity of negative samples, non-seal images were obtained through manual annotations and random cropping from larger images.

Baselines The models utilized in the experiments include AlexNet [45], GoogLeNet [46], ResNet-50 [47] and EfficientNet-B0 [40]. We train models for 300 epochs

with a batch size of 32 and a learning rate of 0.001. All model weights were initialized using pretrained checkpoints.

Evaluation metrics We use the same evaluation metrics as those used in the text extraction experiment.

Results The experimental results are shown in Table 5. It shows that the F1 score for ResNet-50 is 99.54%, while our method using EfficientNet-B0 achieves an F1 score of 99.28%. This demonstrates the effectiveness of our approach in performing seal validation. Furthermore, the total number of parameters in ResNet-50 is five times greater than that in EfficientNet-B0. Consequently, our method offers superior computational efficiency while still maintaining high recognition accuracy. Figure 14 is an example of seal detection.

Wan *et al. Heritage Science*   (2024) 12:159

Page 14 of 18

Subject extraction experiment

We construct a subject detection dataset and conduct a series of experiments to determine which model best suits our subject extraction task.

Dataset We developed a Chinese painting subject detection dataset using open and authorized digital resources from the museum, comprising 3261 annotated images. We annotated the images with 7 labels in COCO format, utilizing the Labelimg for annotation. The dataset's statistical details are presented in Table 6. Samples across all categories were randomly divided into training and test sets in an 8:2 ratio.

Baselines We conduct experiments using state-of-the-art object detection models, including DETR [48], YOLOS [49], YOLO v5 variants (s, m, l) [50], and EfficientDet [51] for comparison. DETR and YOLOS are sequence-to-sequence models, a distinctive architecture from traditional object detection networks.

Evaluation metrics We use two average precision metrics: mean average precision (mAP@0.5) and mean average precision with IoU thresholds ranging from 0.5 to 0.95 (mAP@0.5:0.95). These metrics are widely used to assess the performance of models in object detection tasks. Specifically, mAP@0.5 measures the average precision at an IoU threshold of 0.5, while mAP@0.5:0.95 measures the average precision across the IoU range from 0.5 to 0.95:

$$\text{mAP@0.5} = \frac{1}{N}\sum_{i=1}^{N}\text{AP}_{0.5}^{(i)} \tag{24}$$

$$\text{mAP@0.5:0.95} = \frac{1}{N}\sum_{i=1}^{N}\text{AP}_{0.5:0.95}^{(i)} \tag{25}$$

Results The comparative experimental results for painting and calligraphy subject detection are presented in Table 7. Our method using EfficientDet-D1 shows superior performance. It achieves mAP@0.5 of 74.3%, and a mAP@0.5:0.95 of 50.7%. These results surpass those of EfficientDet-D0 by margins of 0.5% and 0.7%, respectively. Notably, our method outperforms the YOLOv5 variants, with a particularly significant improvement over the YOLOv5l model by 2.6% in mAP@0.5 and 2.1% in mAP@0.5:0.95. DETR and YOLOS, featuring a sequence-to-sequence architecture, demonstrate lower performance in this specific task. An example of subject detection is shown in Fig. 15.

Applications

Based on the WuMKG, a web platform is b dal retrieval, knowledge-based question and answer system (Q &A system), and visualization.
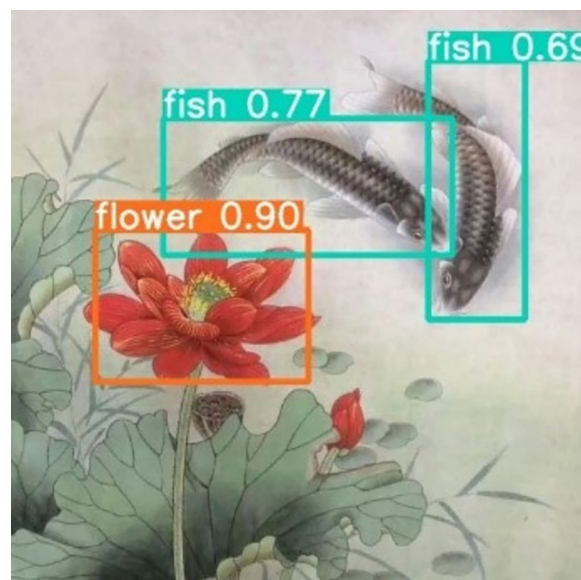

Fig. 15 An example of subject detection

Table 7 Subject detection experimental results

Model	mAP@0.5/%	mAP@0.5:0.95/%
DETR	56.5	37.9
YOLOS	56.2	36.8
YOLOv5-s	71.3	45.3
YOLOv5-m	67.5	46.8
YOLOv5-l	71.7	48.6
EfficientDet-D0	73.8	50.0
Ours	74.3	50.7

Multimodal retrieval It allows retrieval by text and image. Image retrieval is based on VGGNet16 to extract image features for entity matching. When retrieving seals, paintings and calligraphy, we use Faiss for similarity search to achieve faster search speed. The visualization application is developed based on Echarts, as shown in Fig. 16. Entities and relationships are displayed in the form of a force-directed diagram.

Q &A system we develop a multimodal Q &A system based on the WuMKG framework. The system consists of three key modules: intent classifier, SPARQL parser, and answer generator. The intent classifier identifies the user's input question and calculates its similarity to intent templates to determine the corresponding intent type, as illustrated in Table 8. We construct 17 types of intents, each including multiple templates, and use a Bayesian classifier as the classifier. Subsequently, the intent is forwarded to the parser to generate the appropriate query.

Wan *et al. Heritage Science*   (2024) 12:159

Page 15 of 18

Table 8 Category of intents

Intent num	Intent name	Examples
1	Author	Who is the artist of [Song Yin Gao Yin](松阴高隐)?
2	Length of Work	How long is the [Qingming Shanghe Tu](清明上河图)?
3	Width of Work	How wide is the [Five Horses painting](五马图)?
4	Painting material	What material of painting is [Youchun Tu](游春图)?
5	Count	How many paintings did [Tang Yin](唐寅) create?
6	Style	What Style of paintings did Tang Yin create?
7	Location of Collection	Where is [Wangchuan Shijin Tu](辋川十景图) preserved?
8	Description of Collection	What does [Kuan He Tu](款鹤图) depict?
9	Author's Birth and Death	When was [Qiu Ying](仇英) born?
10	Author's Alias	What is Tang Yin's alias?
11	Author's Dynasty	Which dynasty did [Dong Qichang](董其昌) belong to?
12	Author's Introduction	Provide an introduction to Qiu Ying.
13	Dimensions	What size is [Song Yin Gao Yin](松阴高隐)?
14	Subject	What is the Subject of Kuan He Tu?
15	Image	Show me Kuan He Tu.
16	Grade[a]	Which Grade does [Song Yin Gao Yin](松阴高隐) belong to?
17	Relationships	Who is Tang Yin's teacher/father?

[a] According to the classification standards for cultural relics collections announced by the Ministry of Culture of China, cultural relics are divided into precious and general cultural relics, with precious cultural relics further classified into first, second, and third grades

Fig. 16 Example of WuMKG visualization

The generator module then employs this query to extract values from WuMKG, transforming them into coherent and informative answer statements.

The knowledge question and answer uses a template matching method based on Bayesian classification, which supports 17 types of questions and also supports image-based question and answer, as shown in Fig. 17.
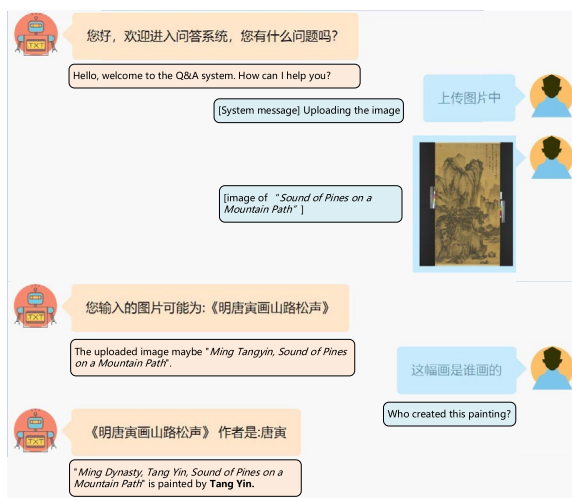
Fig. 17 Example of Q &A system

Table 9 Example of personal information

Type	Attr.	Desc.(Trans.)
Basics	Name	Tang Yin [唐寅]
	Dynasty	Ming Dynasty [明]
	Artworks	The sound of pines on the mountain road painted by Tang Yin in the Ming Dynasty [明唐寅画山路松声]…
Relations	Relatives	Brother: Tang Shen, Wife: Xu [兄弟：唐申, 妻子：徐氏]
	Friends	Zhang Ling [张灵]
	Clan	Tang Zhongmian [唐仲冕]
	Learn-from	Wang Ao [王鏊]
Activity	Time	The second year of Zhengde (1507) [正德二年(1507年)]
	Artist	Tang Yin [唐寅]
	Event	Tang Yin's construction of Tao Hua Temple and Mengmo Pavilion [唐寅筑桃花庵别业及梦墨亭]

Conclusions

In this paper, aimed at aggregating the diverse knowledge of ChP&C and extracting knowledge from text and image, we propose to build the large-scale multimodal knowledge graph WuMKG integrating heterogeneous text and images. We give a practical construction framework and KG-based applications. Specifically, we first design the ChP&C ontology based on CRM and Shl-names, which is extended to enhance the description of personal information. Thereafter, we extract textual and visual knowledge from massive multimodal data obtained from the internet and specialized books. We propose seal extraction and subject extraction methods that extract visual information and convert them into knowledge. We construct seal and subject datasets and conduct empirical experiments, demonstrating that our

proposed method is promising and effective. Furthermore, we ensure the correctness and rationality of knowledge by aligning entities and resolving conflicts through the calculation of entity similarities. Finally, we construct WuMKG, comprising 418,450 triples that encompass data on approximately 1,200 artists and 12,000 painting and calligraphy works. We implement various applications based on WuMKG, including multi-modal retrieval, Q &A system, and visualization.

Although we have made progress in utilizing MKG to explore ChP&C, there is still room for improvement. The decentralized nature of ChP&C artifacts presents challenges in data collection. Moreover, the application of the knowledge graph remains elementary. The work offers

Table 10 Example of resource records

Type	Attr.	Desc.(Trans.)
Electronic Resources	Images	File name: The sound of pine trees on the mountain road.jpg [文件名：山路松声.jpg]
External Records	Alignment	http://cbdb.library.sh.cn/entity/person/ia36ximanh5nzsxr [URI in CBDB]
References	Bibliography	Lin Lina, Chen Jianzhi, Zheng Shufang, "The Scroll of Tang Yin's Painting of Mountain Paths and the Sound of Pines," included in "Special Exhibition of Ming Four Masters - Tang Yin" (Taipei: National Palace Museum, July 2014), pages 50-53, 294. [林莉娜、陈建志、郑淑方，〈明唐寅画山路松声轴〉，收入《明四大家特展-唐寅》（台北：国立故宫博物院，2014.07），页50-53、294。]
	Source Record	Taipei National Palace Museum Collection Records [台北故宫博物院典藏记录]

Wan *et al. Heritage Science*   (2024) 12:159

Page 17 of 18

insights into the application of multi-modal knowledge graphs in the domain of ChP&C, providing a reference for similar studies in the broader field of cultural heritage. It is essential to extend our research to encompass a wider range of cultural relics, exploring diverse types and subjects. Such future endeavors will not only address the current limitations but also enrich our understanding and preservation of cultural heritage.

Appendix A Examples of knowledge schema

We present examples of knowledge schema showcasing person and resource records in Tables 9 and 10.

Appendix B Data source details

Painting and calligraphy: Our primary sources for painting and calligraphy data are the following:

1. The Book：*Catalogue of Authenticated Works of Ancient Chinese Painting and Calligraphy* (24 volumes) and *Illustrated Catalogue of Selected Works of Ancient Chinese Painting and Calligraphy* (10 volumes). These reference books collectively contain 20,117 works, making them invaluable resources for ancient painting and calligraphy in China. We extract text and image information from the pages using OCR.
2. Museum websites, including those of the Beijing Palace Museum, the National Palace Museum in Taipei, the Metropolitan Museum of Art in the United States, the British Museum, and over 30 other museums. These websites provide access to a wealth of visual and descriptive information on painting and calligraphy exhibits. These data are transformed into semi-structured and structured data.
3. Cultural Heritage Administration websites, specifically the Catalog of Precious Cultural Relics, serve as a valuable source of historical and cultural artifacts.
4. The Shanghai Library Knowledge Service Platform, which offers access to a diverse range of relevant data.
5. Various book and literature databases that supplement our research with additional textual information.

Person: For person data related to historical figures, we have drawn from the following sources:

1. The Book：*History of Wu Men Painting School.*
2. China Biographical Database Project (CBDB).

3. The Shanghai Library's Personal Name Standard Database.
4. Online references such as the Baidu Baike, Wikipedia, and the Getty vocabulary, contribute to our comprehensive dataset of historical figures.

References
1. Zou Q, Cao Y, Li Q, Huang C, Wang S. Chronological classification of ancient paintings using appearance and shape features. Pattern Recogn Lett. 2014;49:146–54.
2. Ferrada S, Bustos B, Hogan A. IMGpedia: a linked dataset with content-based analysis of Wikimedia images. In: The Semantic Web–ISWC 2017: 16th International Semantic Web Conference, Vienna, Austria, October 21–25, 2017, Proceedings, Part II 16. Springer; 2017; p. 84–93.
3. Alberts H, Huang N, Deshpande Y, Liu Y, Cho K, Vania C, et al. VisualSem: a high-quality knowledge graph for vision and language. In: Proceedings of the 1st Workshop on Multilingual Representation Learning; 2021; p. 138–152.
4. Wang M, Wang H, Qi G, Zheng Q. Richpedia: a large-scale, comprehensive multi-modal knowledge graph. Big Data Res. 2020;22: 100159.
5. Champion E, Rahaman H. Survey of 3D digital heritage repositories and platforms. Virtual Archaeol Rev. 2020;11(23):1–15.
6. Davis E, Heravi B. Linked data and cultural heritage: a systematic review of participation, collaboration, and motivation. J Comput Cult Herit (JOCCH). 2021;14(2):1–18.
7. Lyu S, Yang X, Pan N, Hou M, Wu W, Peng M, et al. Spectral heat aging model to estimate the age of seals on painting and calligraphy. J Cult Herit. 2020;46:119–30.
8. Agathos M, Kalogeros E, Gergatsoulis M. CIDOC CRM. In: From Born-Physical to Born-Virtual: Augmenting Intelligence in Digital Libraries: 24th International Conference on Asian Digital Libraries, ICADL 2022, Hanoi, Vietnam, November 30–December 2, 2022, Proceedings. vol. 13636. Springer Nature; 2022p; p. 345.
9. Shanghai Library Person Names Ontology; 2022. https://data.library.sh.cn/ont/ontology/tree?g=http://ont.li.
10. Cobb J. The journey to linked open data: the Getty vocabularies. J Libr Metadata. 2015;15(3–4):142–56.

Wan *et al. Heritage Science*   (2024) 12:159

Page 18 of 18

11. Wan J, Zhou Y, Chen G, Yi J. Designing a multi-level metadata standard based on Dublin core for museum data. In: International Conference on Dublin Core and Metadata Applications; 2014; p. 31–36.
12. Lima VMA, Macambyra M. The VRA core in a digital library of artistic production. Vis Resour Assoc Bull. 2023;50(2).
13. Ciortan IM, Pintus R, Marchioro G, Daffara C, Gobbetti E, Giachetti A. A DICOM-inspired metadata architecture for managing multimodal acquisitions in Cultural Heritage. In: Digital Cultural Heritage: Final Conference of the Marie Skłodowska-Curie Initial Training Network for Digital Cultural Heritage, ITN-DCH 2017, Olimje, Slovenia, May 23–25, 2017, Revised Selected Papers. Springer; 2018; p. 37–49.
14. Yu-Yun L. A comparative study of the VRA Core, CDWA and Archaeodata. J Libr Inf Sci. 2005;31(2).
15. Bobasheva A, Gandon F, Precioso F. Learning and reasoning for cultural metadata quality. J Comput Cult Herit. 2022;15(3).
16. Freire N, Robson G, Howard JB, Manguinhas H, Isaac A. Metadata aggregation: assessing the application of IIIF and sitemaps within cultural heritage. In: International Conference on Theory and Practice of Digital Libraries. Springer; 2017; p. 220–232.
17. Xue S, Li Y, Ren L. Representing the Chinese Seal Stamping Catalogs Using IIIF & Serverless. In: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020. JCDL '20. New York, NY, USA: Association for Computing Machinery; 2020; p. 547–548.
18. Freire N, Meijers E, de Valk S, Raemy JA, Isaac A. Metadata aggregation via linked data: results of the Europeana Common Culture project. In: Research Conference on Metadata and Semantics Research. Springer; 2020; p. 383–394.
19. Tsui LH, Wang H. Harvesting big biographical data for Chinese history: the China Biographical Database (CBDB). J Chin History. 2020;4(2):505–11.
20. Matusiak KK, Meng L, Barczyk E, Shih CJ. Multilingual metadata for cultural heritage materials: the case of the Tse-Tsung Chow collection of Chinese scrolls and fan paintings. Electron Libr. 2015;33(1):136–51.
21. Hyvönen E. Digital humanities on the Semantic Web: Sampo model and portal series. Semantic Web. 2023;14(4):729–44.
22. Isaac A, Haslhofer B. Europeana linked open data-data. Europeana. eu. Semantic Web. 2013;4(3):291–7.
23. Carriero VA, Gangemi A, Mancinelli ML, Marinucci L, Nuzzolese AG, Presutti V, et al. ArCo: The Italian cultural heritage knowledge graph. In: The Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part II 18. Springer; 2019; p. 36–52.
24. Koho M, Ikkala E, Leskinen P, Tamper M, Tuominen J, Hyvönen E. WarSampo knowledge graph: Finland in the second world war as linked open data. Semantic Web. 2021;12(2):265–78.
25. Bruns O, Tietz T, Chaabane MB, Portz M, Xiong F, The Sack H, Graph Nuremberg Address Knowledge, In: The Semantic Web: ESWC,. Satellite Events: Virtual Event, June 6–10, 2021, Revised Selected Papers 18. Springer. 2021;2021:115–9.
26. Tong Yin ZB. Exploration and Practice of the Dong Qichang Digital Humanities Project. Chinese Museum. 2018;04(114-118).
27. Wang X, Tan X, Gui H, Song N. A semantic enrichment approach to linking and enhancing Dunhuang cultural heritage data. In: Information and Knowledge Organisation in Digital Humanities. Routledge; 2021; p. 87–105.
28. Schleider T, Troncy R, Ehrhart T, Dorozynski M, Rottensteiner F, Sebastián Lozano J, et al. Searching silk fabrics by images leveraging on knowledge graph and domain expert rules. In: Proceedings of the 3rd Workshop on Structuring and Understanding of Multimedia heritAge Contents; 2021; p. 41–49.
29. Puren M, Vernus P. Conceptual Modelling of the European Silk Heritage with the SILKNOW Data Model and Extension; 2022. Working paper or preprint.
30. Fan T, Wang H, Hodel T. CICHMKG: a large-scale and comprehensive Chinese intangible cultural heritage multimodal knowledge graph. Herit Sci. 2023;11(1):1–18.
31. Liong ST, Huang YC, Li S, Huang Z, Ma J, Gan YS. Automatic traditional Chinese painting classification: a benchmarking analysis. Comput Intell. 2020;36(3):1183–99.
32. Zhang J, Miao Y, Zhang J, Yu J. Inkthetics: a comprehensive computational model for aesthetic evaluation of Chinese ink paintings. IEEE Access. 2020;8:225857–71.
33. Li D, Zhang Y. Multi-instance learning algorithm based on LSTM for Chinese painting image classification. IEEE Access. 2020;8:179336–45.
34. Smith A. Simple Knowledge Organization System (SKOS). Ko Knowl Org. 2022;49(5):371–84.
35. Shanghai Library open data platform; 2021. https://data.library.sh.cn/index.
36. Liu W, Fu X, Zhang Y, Xiao W. Lexicon Enhanced Chinese Sequence Labeling Using BERT Adapter. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers); 2021; p. 5847–5858.
37. Zhang K, Liang J, Van Gool L, Timofte R. Designing a practical degradation model for deep blind image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2021; p. 4791–4800.
38. Kumar M, Bhandari AK. Contrast enhancement using novel white balancing parameter optimization for perceptually invisible images. IEEE Trans Image Process. 2020;29:7525–36.
39. Mafi M, Martin H, Cabrerizo M, Andrian J, Barreto A, Adjouadi M. A comprehensive survey on impulse and Gaussian denoising filters for digital images. Signal Process. 2019;157:236–60.
40. Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. PMLR; 2019; p. 6105–6114.
41. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014.
42. Zhou Z, Wu QJ, Wan S, Sun W, Sun X. Integrating SIFT and CNN feature matching for partial-duplicate image detection. IEEE Trans Emerg Topics Comput Intell. 2020;4(5):593–604.
43. Zhao X, Jia Y, Li A, Jiang R, Song Y. Multi-source knowledge fusion: a survey. World Wide Web. 2020;23:2567–92.
44. Yun H, He Y, Lin L, Wang X. Research on multi-source data integration based on ontology and karma modeling. Int J Intell Inf Technol (IJIIT). 2019;15(2):69–87.
45. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015; p. 1–9.
46. Davari N, Akbarizadeh G, Mashhour E. Corona detection and power equipment classification based on GoogleNet-AlexNet: an accurate and intelligent defect detection model based on deep learning for power distribution lines. IEEE Trans Power Delivery. 2021;37(4):2766–74.
47. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016; p. 770–778.
48. Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S. End-to-end object detection with transformers. In: European conference on computer vision. Springer; 2020; p. 213–229.
49. Fang Y, Liao B, Wang X, Fang J, Qi J, Wu R, et al. You only look at one sequence: rethinking transformer in vision through object detection. Adv Neural Inf Process Syst. 2021;34:26183–97.
50. Wu W, Liu H, Li L, Long Y, Wang X, Wang Z, et al. Application of local fully Convolutional Neural Network combined with YOLO v5 algorithm in small target detection of remote sensing image. PLoS ONE. 2021;16(10): e0259283.
51. Tan M, Pang R, Le QV. Efficientdet: Scalable and efficient object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2020; p. 10781–10790.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.