From Tradition to Technology: Leveraging Thangka Art Datasets for Cultural and Academic Advancements

Zhuo Li[1,2], Chunyan Peng[1,2†], Jing Zhang[1,2]

[1]School of Computer Science, Qinghai Normal University, Xining 810016, China

[2]The State Key Laboratory of Tibetan Intelligence, Qinghai Normal University, Xining 810016, China

Keywords: Thangka; Cultural heritage preservation; Data annotation; Dataset analysis; Image recognition

Citation: Li Z., Peng C.Y., Zhang J.: From Tradition to Technology: Leveraging Thangka Art Datasets for Cultural and Academic Advancements. Data Intelligence, Vol. XX, Art. No.: 2025??XX, pp. 1-22, 2025. DOI: https://doi.org/10.3724/2096-7004.di.xxx

ABSTRACT

Thangka is a distinctive and intricate form of Tibetan Buddhist religious art, combining religious beliefs, philosophical ideas, and fine painting techniques, and carrying profound cultural and religious significance. As a vital medium for Tibetan Buddhist art, Thangka plays a key role not only in religious rituals and practices but also in recording history and preserving culture. With the rapid advancement of artificial intelligence technologies, the digitalization of Thangka art has gained increasing attention, especially in the field of computer vision, where its potential for applications in object detection and image segmentation is becoming more evident. However, existing Thangka datasets face significant limitations in terms of scale, annotation accuracy, diversity, and category coverage, restricting their effectiveness in complex visual tasks. To address these challenges, this study introduces a high-quality, comprehensive Thangka dataset consisting of 6,134 high-resolution images, covering 37 primary elements such as Buddha statues, Bodhisattvas, ritual implements, headdresses, and mounts, with 17,628 precise annotations. The dataset supports both VOC and YOLO formats, making it suitable for various computer vision tasks. Compared to existing datasets, it offers substantial improvements in category diversity, annotation precision, and adaptability to complex visual scenarios. Extensive experiments were conducted using several state-of-the-art object detection models, including YOLOv5, YOLOv7, YOLOv8, YOLOv10, YOLOv11, and Faster R-CNN. The results demonstrate the dataset's robustness and reliability, with excellent performance in object detection tasks. Further analyses of bounding box scale distribution, spatial distribution, and size distribution provide deeper insights into the dataset's characteristics and model performance. This dataset not only establishes a solid foundation for digital cultural heritage preservation research but also provides valuable resources for applications in artistic analysis, automated

[†] Corresponding author: Chunyan Peng (E-mail: pcy@qhnu.edu.cn; ORCID: 0009-0001-8289-5092).

image classification, and fine-grained object detection in artificial intelligence. It opens up opportunities for future research in multi-modal learning, cross-domain adaptation, and AI-driven cultural heritage restoration, significantly contributing to both academic studies and digital humanities practices.

1. INTRODUCTION

Thangka, a traditional Tibetan scroll painting, has been practiced for centuries and often utilizes natural mineral pigments such as gold, silver, pearls, agate, coral, turquoise, chalchuite, cinnabar, and other precious minerals. These materials exhibit remarkable chemical stability, enabling Thangka artworks to retain their vivid colors for hundreds of years, thereby endowing them with immense artistic, historical, and scholarly value [1]. As a profound symbol of Tibetan Buddhist culture, Thangka reflects the spiritual beliefs of Tibet and embodies the artistic achievements and cultural heritage of the region. The creation of Thangka is a complex process, typically carried out by artists who undergo years of intensive training. The content of Thangka is diverse, including depictions of buildings, medicine, astronomical calendars, legendary stories, and more [2].

With the passage of time, Thangka art has evolved into various styles and schools, with distinct differences in content and aesthetics across regions and eras [3]. These differences highlight the diversity of Buddhist culture in Tibet and serve as historical records of the religious, social, and cultural shifts throughout different periods. Thangka is not merely a reflection of religious devotion but also a witness to the history, culture, and social transformations of the Tibetan people [4]. As such, the study of Thangka art holds great significance for understanding Tibetan culture, religious traditions, and artistic evolution.

In recent years, as globalization accelerates and cultural exchanges become more frequent, Thangka art, a valuable piece of cultural heritage, has gained increasing recognition. However, the preservation and transmission of Thangka art face significant challenges. Natural aging, environmental pollution, and improper storage methods have led to the gradual deterioration or even disappearance of numerous Thangka works [5]. Additionally, research on Thangka still largely depends on traditional methods of physical analysis and manual documentation, which are often inefficient and insufficient for comprehensively preserving and analyzing the rich details of this art form [6].

To tackle these issues, digital technology offers new opportunities for the preservation and study of Thangka art. Through high-resolution digital scanning and archiving technologies, Thangka works can be meticulously recorded and preserved, helping to prevent further degradation caused by natural or human factors [7]. Moreover, digital technology facilitates wider access and display of Thangkas through virtual exhibitions. With advancements in computer vision technology, researchers can now employ automated tools to identify and analyze the complex patterns and symbols found in Thangka art, thereby improving research efficiency and broadening the scope of scholarly inquiry [8].

While digital technology and computer vision hold immense potential for the preservation and study of Thangka art, the lack of high-quality, standardized Thangka image datasets presents a significant challenge. The absence of systematic and thoroughly annotated datasets restricts the widespread application of these technologies and impedes deeper research into the art of Thangka [9]. Therefore, the creation of a robust and well-curated Thangka image dataset is essential for advancing both the digital preservation and scholarly study of this culturally significant art form.

To address the aforementioned challenges, this paper proposes and constructs an image dataset encompassing a diverse range of Thangka art works. The dataset includes Thangka images from different regions, periods, and styles, with detailed annotations of the main elements within the images, including Buddhas, Bodhisattvas, protector deities, animals, and ritual implements. Using standardized annotation methods, this dataset provides critical resource support for the digital preservation of Thangka art and research in computer vision [19]. The main contributions of this paper are as follows:

1) We present a comprehensive Thangka image dataset, containing 6134 high-resolution Thangka images with detailed annotations of key elements.
2) We provide a detailed description of the dataset construction process, including image collection, selection, annotation, and image augmentation.
3) We demonstrate the effectiveness of this dataset for image classification and object detection tasks by utilizing well-established network models such as YOLOv5, YOLOv7, YOLOv8, YOLOv10, YOLOv11, and Faster R-CNN.

The remainder of this paper is organized as follows: Section 2 reviews related studies on digital preservation, computer vision applications, deep learning techniques, and image annotation methods in the context of Thangka art. Section 3 details the dataset construction process. Section 4 presents and analyzes the experimental results. Section 5 concludes the paper and discusses future research directions.

2. RELATED WORK

2.1 Current Status of Digital Preservation of Thangka Art

In recent years, digital preservation technologies have been widely adopted in the conservation of cultural heritage, with the primary goal of digitally preserving and restoring the original appearance of cultural artifacts [10]. Wang and Zhang [11] highlight that such technologies can be used not only to document Thangka paintings at risk of deterioration but also to reconstruct historical scenes and religious rituals through 3D modeling and virtual reality, thereby enhancing methods of research and exhibition. Moreover, digitized Thangka images can be disseminated via online platforms, enabling broader public access and academic engagement with this valuable form of art [12].

2.2 Application of Computer Vision Technology in Thangka Research

The rapid development of computer vision technology has provided powerful tools for the study of Thangka art. Liu and Wang [13] employed image classification techniques to efficiently classify large volumes of Thangka images, distinguishing religious figures such as Buddhas, Bodhisattvas, and protector deities. Zhang and Wu [14] applied object detection methods to locate and annotate key elements within the images, improving the accuracy of content and structural analysis. Zhou and Zhang [15] achieved the separation of different elements in Thangka images, facilitating more detailed analysis. In addition, Tang and Pan proposed a model that integrates dynamic convolution with residual networks to improve the classification accuracy and robustness of Thangka images [30].

2.3 Research Progress in Deep Learning and Image Enhancement Technologies

With the rapid development of artificial intelligence, an increasing number of researchers have applied deep learning methods to the recognition and analysis of Thangka images. He and Li [13] utilized convolutional neural networks (CNNs) to achieve automatic recognition and classification of Thangka images, demonstrating strong feature extraction capabilities. Zhang et al. [8] incorporated attention mechanisms to enhance the model's focus on key regions, improving semantic understanding of the images. Wang and Fan [3, 19] employed image super-resolution techniques to reconstruct low-quality Thangka images, preserving more details and supporting digital restoration efforts. Recent studies have also proposed a semantic concept-guided multimodal optimization strategy (SCAMF-Net) to generate Thangka image descriptions with enhanced semantic expressiveness [31].

2.4 Applications of Thangka Image Annotation and Object Detection Models

Zhang and Wu [18] explored an image annotation process integrated with machine learning, effectively addressing the time-consuming nature of manual labeling and improving data consistency. The YOLO series of algorithms (e.g., YOLOv5, YOLOv8) have been widely applied to the recognition and localization of figures, deities, and ritual objects in Thangka images due to their efficiency and accuracy [14-15]. Models such as Faster R-CNN have also demonstrated excellent performance in high-precision detection tasks [13]. Additionally, Guo and Wang enhanced image-text matching in the context of Thangka art by improving the Transformer architecture with an adaptive pooling strategy [32]. Bai and Fan investigated and developed a Thangka image restoration method based on discrete codebooks and Transformer [33]. These studies have laid a solid foundation for the intelligent analysis and preservation of Thangka art.

3. MAIN METHODS

3.1 Semantic Introduction of Thangka

The main challenge in constructing a Thangka dataset lies in the fact that many elements within Thangka images possess strong symbolic meaning. These elements, such as deities, ritual implements, and headgear, have rich Buddhist semantics that require specialized knowledge for identification and understanding. Before annotating the Thangka dataset, one must have a certain level of understanding of Thangka and be familiar with advanced Buddhist-related semantic knowledge. Therefore, the primary task is to deeply study the semantic meanings of the various elements present in Thangka images.

Based on an in-depth analysis of works such as 《Study of Thangka Images》[20], 《The World's Most Beautiful Thangka—Ritual Implements in Thangka》[21], 《Thangka Art (Complete Illustrated)》[22], and Bi et al.'s research on Thangka headgear [23]. This article categorizes the collected Thangka images into six major categories: 1) Buddha Figures, 2) Bodhisattva Figures, 3) Vajra Figures, 4) Ritual Implements Figures, 5) Headwear Figures, 6) Mounts Figures and 37 subcategories. Table 1 below presents the main types found in Thangka.

Table 1. Description of Major Types in Thangka.

Category	Included Categories	Name	Sample Image	Feature
Buddha Figures	Shakyamuni Buddha Green Tara White Tara Medicine Buddha Laughing Buddha Amitayus	Shakyamuni Buddha		Shakyamuni Buddha is depicted as the historical founder of Buddhism, representing enlightenment and wisdom
Bodhisattva Figures	Four-Armed Avalokiteshvara Manjushri Samantabhadra Mahasthamaprapta Padmasambhava Sarasvati Ksitigarbha Bodhisattva Bodhisattva Thousand-Armed Avalokiteshvara	Manjushri Bodhisattva		Manjushri is the embodiment of wisdom, wielding a flaming sword to cut through ignorance.
Vajra Figures	Yellow Jambhala Hayagriva Yamantaka King Gesar Mahākāla	Hayagriva		Hayagriva is a protective deity in Tibetan Buddhism, symbolizing wrath and wisdom.

Table 1. *Continued.*

Category	Included Categories	Name	Sample Image	Feature
Ritual Implements Figures	Begging Bowl Medicine Bowl Vajra Trident Chintamani Staff Sword of Wisdom Kapala Pipa Bowl of Compassion Treasure-spitting Rat	Begging Bowl		The bowl represents simplicity, humility, and the renunciation of worldly possessions.
Headwear Figures	Crown Hair Knot Monk's Hat	Crown		The crown worn by deities symbolizes their supreme authority and spiritual power.
Mounts Figures	Six-tusked White Elephant Blue Lion Lotus Throne Horse	Six-tusked White Elephant		The six-tusked white elephant symbolizes strength, purity, and the perfection of wisdom.

In Table 1, the major types and their symbolic meanings in Thangka are presented. First, Shakyamuni Buddha, as the founder of Buddhism, symbolizes enlightenment and wisdom, reflecting the pursuit of true essence. Secondly, Manjushri Bodhisattva is the embodiment of wisdom, wielding a flaming sword that symbolizes the power to cut through ignorance. Hayagriva, as a protective deity, embodies the combination of wrath and wisdom, protecting followers. The Begging bowl represents simplicity and humility, reminding people to renounce worldly desires and seek inner tranquility. The hair crown symbolizes the supreme authority and spiritual power of deities, reflecting the noble status in the practice. Finally, the six-tusked white elephant symbolizes strength, purity, and the perfection of wisdom, signifying the importance of the path to enlightenment. These symbolic meanings not only enrich the artistic expression of Thangka but also provide profound insights into its significance in Tibetan Buddhist culture.

3.2 Data Collection

The creation of Thangka requires strict standards and a complex process, often involving the use of plant pigments and precious metals. This makes the collection of Thangka images particularly challenging. To date, datasets specifically focused on Thangka are not only limited in number but also generally incomplete. Therefore, this article adopts a combination of online and offline methods to collect Thangka images, aiming to gather them from the widest possible range. The offline collection channels mainly include the following:

On-site visits by lab personnel: 1) Our team conducted fieldwork in the Regong area of Jianzha County, Qinghai Province. In recent years, with government support, intangible cultural heritage protection projects, and art training, the inheritance and development of Regong Thangka have been effectively preserved. By consulting local artists and obtaining their permission, we were able to collect images. 2) Scanning Thangka images from relevant books: Thangka-related books such as 《The Thangka Collection of the Forbidden City》[24], 《Rebgong Nian Duhu Thangka Art》[25], and 《The Complete Collection of Tibetan Thangka》[26] were used to scan Thangka images.

The online collection channels include: 1) Web scraping technology: Images were scraped from browsers such as Baidu, Google, and Bing by inputting keywords like “Thangka” or the names of Buddhist figures. 2) Specific Thangka websites: Images were collected from publicly available resources on websites dedicated to Thangka, such as (http://www.datathangka.com/). 3) Visiting major museum websites: Thangka images were also collected by accessing the websites of various museums.

3.3 Data Filtering

Through both offline and online methods, a total of 9463 Thangka images were collected. However, some of these images were duplicates, unclear, or of poor quality. Therefore, we first used OpenCV to obtain the histogram data of the images and normalize them. Then, we calculated the image similarity and removed images with a similarity greater than 90%. Images that were of poor quality or where the content was clearly unclear were also removed. For images that were slightly unclear but still recognizable, image restoration techniques were applied to enhance their clarity. Table 2 shows the details of data preprocessing for the collected Thangka images, including duplicate removal, low-quality filtering, and restoration of slightly unclear images.

Table 2. data preprocessing details for the collected data.

Collect Images	Duplicate Images	Low-Quality Images	Restore Images	Remaining Images
9463	2149	1180	848	6134

3.4 Data Augmentation

During the annotation analysis, we observed significant disparities in the number of samples across different categories. Such class imbalance may cause bias during model training, which in turn affects overall recognition performance and generalization ability. To effectively address this issue, we systematically applied a series of data augmentation techniques to low-frequency categories with fewer than 100 annotated instances, aiming to increase the diversity and quantity of their samples and thereby improve the model’s recognition performance for these minority classes. The specific augmentation operations included random rotation (within a range of –30° to +30°), random cropping, horizontal flipping, image translation (with a maximum shift of 20% of the image width and height), and scaling (scaling ratios between 80% and 120%). Additionally, all images were normalized to stabilize the training process. Through these diverse data transformations, we not only expanded the dataset size but also effectively simulated various real-world variations of objects, significantly enhancing the model's robustness and generalization ability while reducing performance bias caused by class imbalance.

3.5 Data Annotation

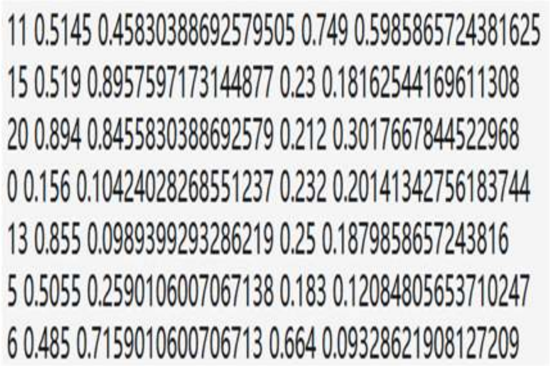
3.5.1 Annotation Process and Data Format

To train the Thangka object detection model, we conducted detailed manual annotation on the collected Thangka image dataset. The annotation work was jointly completed by two graduate students with backgrounds in Tibetan Buddhist art, and subsequently reviewed by an expert in the field to ensure the accuracy and consistency of category classification and object boundaries.

We used X-anylabeling as the annotation tool and saved the data in multiple formats to accommodate different computer vision frameworks and task requirements. Specifically, manual bounding box annotations were performed using the X-anylabelImg tool, and YOLO-format .txt files were exported, with each image having a corresponding .txt file containing the object category ID and normalized bounding box coordinates (center x, y, width, and height). Additionally, Pascal VOC-format .xml files were exported to ensure compatibility with various deep learning platforms.

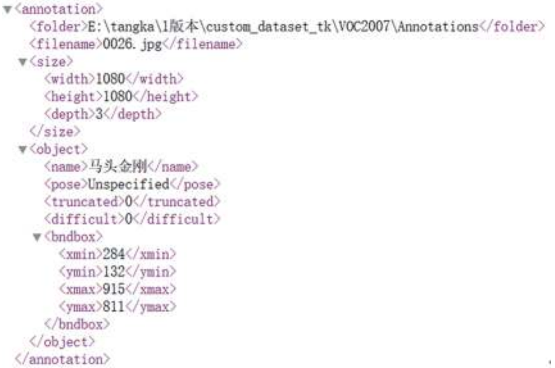
The dataset is divided into six major categories: 1) Buddha figures, 2) Bodhisattva figures, 3) Vajra figures, 4) Ritual implements, 5) Headwear, and 6) Mounts. Each major category is further subdivided into a total of 37 subclasses, covering typical target elements in Thangka art. Each category includes five attributes: category ID, name, annotation count, code, and annotation proportion. The category ID uniquely identifies each category; the name column lists specific category names such as Shakyamuni Buddha, Vajra, and Manjushri Bodhisattva; and the annotation count indicates the number of labeled samples for that category in the entire dataset. Below are example contents of annotation files in YOLO and VOC formats:

Figure 1(a) shows the YOLO format, where the first column represents the class ID, followed by four values indicating the normalized coordinates of the bounding box: the center point (x, y), width, and height. All coordinate values are normalized to the range [0, 1]. For example, the entry 11 0.5145 0.4583 0.749 0.5986 indicates an object of class 11, whose bounding box center is located at approximately 51.45% of the image width and 45.83% of the image height, with the box width and height accounting for about 74.9% and 59.86% of the image dimensions, respectively. Figure 1(b) shows the VOC format, which uses an XML structure to store annotations. This format includes the image filename, size information (width, height, and depth), object class name, and bounding box coordinates. The bounding box is defined by the pixel values of the top-left corner (x_{min}, y_{min}) and the bottom-right corner (x_{max}, y_{max}).



```
11 0.5145 0.45830388692579505 0.749 0.5985865724381625
15 0.519 0.895759717314877 0.23 0.18162544169611308
20 0.894 0.8455830388692579 0.212 0.3017667844522968
0 0.156 0.10424028268551237 0.232 0.20141342756183744
13 0.855 0.0989399293286219 0.25 0.1879858657243816
5 0.5055 0.2590106007067138 0.183 0.12084805653710247
6 0.485 0.7159010600706713 0.664 0.09328621908127209
```

Figure 1(a). YOLO Format Annotation.



```
▼<annotation>
   <folder>E:\tangka\1版本\custom_dataset_tk\VOC2007\Annotations</folder>
   <filename>0026.jpg</filename>
  ▼<size>
     <width>1080</width>
     <height>1080</height>
     <depth>3</depth>
   </size>
  ▼<object>
     <name>马头金刚</name>
     <pose>Unspecified</pose>
     <truncated>0</truncated>
     <difficult>0</difficult>
    ▼<bndbox>
       <xmin>284</xmin>
       <ymin>132</ymin>
       <xmax>915</xmax>
       <ymax>811</ymax>
     </bndbox>
   </object>
</annotation>
```

Figure 1(b). VOC format Annotation.

3.5.2 Buddha Figures

Table 3 presents the annotation information for Buddha figures in Thangka, including category IDs, names, codes, label counts, and proportions for different Buddha representations. Shakyamuni Buddha, as the primary representative of this category, has 1580 annotations, accounting for 8.9% of the total, highlighting its central position in Thangka art. Additionally, Green Tara and White Tara have 583 and 587 annotations, respectively, with proportions of 3.3%, indicating their significant influence in Thangka. Medicine Buddha is represented with 335 annotations, making up 1.9%, while Laughing Buddha appears with 167 annotations at a proportion of 0.9%. Furthermore, Amitayus has 236 annotations, representing 1.3% of the total. These data not only reflect the rich diversity of Buddha figures but also emphasize their importance within the Thangka artistic tradition.

Table 3. Annotation Information for Buddha Figures.

Buddha Figures				
Category ID	**Name**	**Label Count**	**Code**	**Label Proportion**
A0	Shakyamuni Buddha	1580	1	8.9%
A1	Green Tara	583	21	3.3%
A2	White Tara	577	26	3.3%
A3	Medicine Buddha	335	14	1.9%
A4	Laughing Buddha	167	22	0.9%
A5	Amitayus	236	29	1.3%

3.5.3 Bodhisattva Figures

Table 4 presents the annotation statistics for Bodhisattva figures in Thangka. Manjushri Bodhisattva, with 1034 annotations and a proportion of 5.4%, holds a prominent position in this category. The Four-Armed Avalokiteshvara has 684 annotations, accounting for 3.9%, highlighting its importance in Thangka art. Samantabhadra Bodhisattva has 295 annotations, representing 1.6%. Mahasthamaprapta Bodhisattva and Padmasambhava appear with 147 (0.8%) and 358 (2.0%) annotations, respectively. Additionally, Sarasvati is annotated 283 times, comprising 1.6%. Finally, a total of 113 annotations were made for Ksitigarbha Bodhisattva, accounting for 0.6% of the labels. These annotation counts and proportions illustrate the breadth and frequency of different Bodhisattva figures in Thangka.

Table 4. Annotation information for the Bodhisattva Figures.

		Bodhisattva Figures		
Category ID	**Name**	**Label Count**	**Code**	**Label Proportion**
B0	Four-Armed Avalokiteshvara	684	12	3.9%
B1	Manjushri	1034	4	5.8%
B2	Samantabhadra	295	9	1.6%
B3	Mahasthamaprapta	147	28	0.8%
B4	Padmasambhava	358	13	2.0%
B5	Sarasvati	283	30	1.6%
B6	Ksitigarbha Bodhisattva	113	35	0.6%
B7	Thousand-Armed Avalokiteshvara	200	37	1.1%

3.5.4 Vajra Figures

Table 5 summarizes the annotation information for Vajrayana figures in Thangka. Among these figures, Hayagriva stands out with 1007 labels, accounting for 5.7% of the total annotations, highlighting its cultural and symbolic significance. Yellow Jambhala follows with 536 labels, making up 3.0%, while Yamantaka has 301 labels, contributing to 1.7%. Mahākāla is annotated 178 times, representing 1.0%, and Ge-sar rgal-po has the fewest annotations at 87 labels, accounting for 0.5%. This distribution underscores the prominence of Hayagriva among the Vajrayana figures.

Table 5. Annotation information for the Vajra Figures.

Vajrayana Figures				
Category ID	**Name**	**Label Count**	**Code**	**Label Proportion**
C0	Yellow Jambhala	536	16	3.0%
C1	Hayagriva	1007	8	5.7%
C2	Yamantaka	301	15	1.7%
C3	Mahākāla	178	36	1.0%
C4	King Gesar	87	32	0.5%
C5	Vajrapani Bodhisattva	113	34	0.6%

3.5.5 Headwear Figures

Table 6 summarizes the annotation information for the headwear category in Thangka. According to the data, the hair crown is the most common figure in the headwear category, with 1989 annotations, accounting for 11.2% of the total annotations. This is followed by the hair knot, which has 738 annotations, representing 4.1%. The monk's hat has 589 annotations, accounting for 3.3% of the total. These figures indicate that the hair crown occupies a prominent position in Thangka headwear imagery, while the hair knot and monk's hat appear less frequently.

Table 6. Annotation information for the Headwear Figures.

Headwear Figures				
Category ID	**Name**	**Label Count**	**Code**	**Label Proportion**
D0	Crown	1989	6	11.2%
D1	Hair Knot	738	2	4.1%
D2	Monk's Hat	589	18	3.3%

3.5.6 Ritual Implements Figures

Table 7 summarizes the annotation information for the ritual implements category in Thangka. According to the data, the Sword of Wisdom is the most common figure in the ritual implements category, with 591 annotations, accounting for 3.3% of the total annotations. This is followed by the Begging Bowl, which has 543 annotations, representing 3.0%. The Kapala has 452 annotations, accounting for 2.5% of the total. These figures indicate that the Sword of Wisdom occupies a prominent position in Thangka ritual implements imagery, while the Bowl and Kapala appear less frequently. From the label count and proportion, it can be seen that the dataset has a certain degree of imbalance in its category distribution.

This reflects that common ritual implements are well-annotated in the dataset, while some less common implements have relatively fewer samples.

Table 7. Annotation information for the Ritual Implements Figures.

Ritual Implements Figures				
Category ID	**Name**	**Label Count**	**Code**	**Label Proportion**
E0	Begging Bowl	543	3	3.0%
E1	Medicine Bowl	216	25	1.2%
E2	Vajra	400	23	2.2%
E3	Chintamani Staff	67	10	0.4%
E4	Sword of Wisdom	591	5	3.3%
E5	Kapala	452	20	2.5%
E6	Pipa	242	24	1.4%
E7	Bowl of Compassion	78	31	0.4%
E8	Treasure-Spitting Rat	410	17	2.3%
E9	Trident	460	19	2.6%

3.5.7 Mounts Figures

Table 8 summarizes the annotation information for the mounts category in Thangka. According to the data, the Lotus Throne is the most common figure in the mounts category, with 1490 annotations, accounting for 8.4% of the total annotations. This is followed by the Six-tusked White Elephant, which has 324 annotations, representing 1.8%. Next is the Blue Lion, with a total of 103 labels, accounting for 0.6% of the total. The Horse has only 27 annotations, accounting for 0.2% of the total.

Table 8. Annotation information for the Mounts Figures.

Mounts Figures				
Category ID	**Name**	**Label Count**	**Code**	**Label Proportion**
F0	Six-tusked White Elephant	324	11	1.8%
F1	Blue Lion	103	27	0.6%
F2	Horse	82	33	0.5%
F3	Lotus Throne	1490	7	8.4%

4. DATA ANALYSIS

4.1 Category Distribution Analysis

This section analyzes the distribution of label categories, as shown in Figure 2, which illustrates the number of labels for each category in the dataset. Crown and Shakyamuni Buddha have over 1,500 labels each, dominating the dataset and indicating their frequent appearance as common elements in Thangka imagery. The Lotus Throne category also has a relatively high label count, reaching around 1,500. Next are Hayagriva and Manjushri Bodhisattva, with over 1,000 labels each, reflecting their significance as major figures in Thangka. In contrast, other categories such as Green Tara, White Tara, and Bowls have label counts ranging between 500 and 1,000, indicating their moderate presence in the images. Further, categories like Samantabhadra and Yamantaka have label counts between 100 and 500, showing their lower frequency in the images. Finally, rare categories such as King Gesar and Chintamani Staff have fewer than 200 labels, emphasizing their infrequent appearance in the dataset.

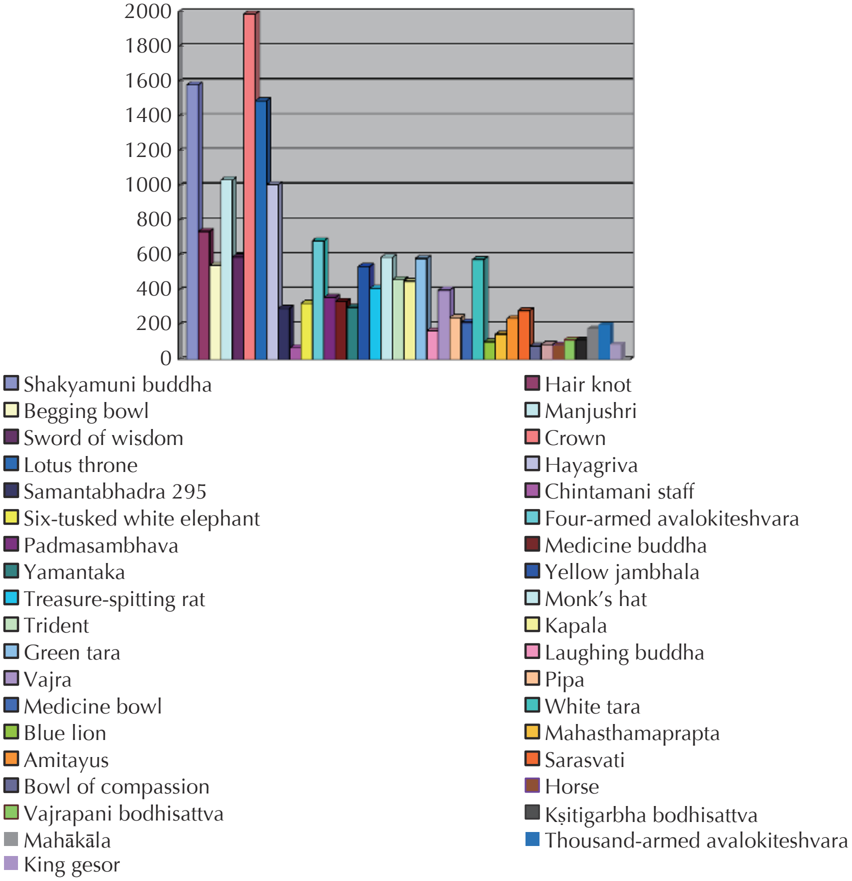


Figure 2. Resents the number of labels for each category in Thangka.

4.2 Model Performance Feedback Analysis

This dataset is designed for Thangka object detection. In the field of object detection, the main evaluation metrics are F1 score, precision, recall, and mean average precision (mAP). Therefore, to validate the quality, representativeness, and adaptability of the dataset, it is necessary to train the dataset using well-established network models and assess the feedback from the training results. In this paper, six current well-established network models—YOLOv5, YOLOv7, YOLOv8, YOLOv10, YOLOv11, and Faster-RCNN—were used to train the dataset, providing feedback on various performance metrics of the dataset. Table 9 shows the training performance of the dataset on these six network models.

Table 9. Model performance feedback.

Model name	F1	precision	recall	mAP@0.5
Yolov5	94%	100%	100%	97%
Yolov7	91%	100%	100%	94.8%
Yolov8	91%	100%	99%	95.4%
Yolov10	87%	100%	98%	93.5%
Yolo11	92%	100%	99%	95.8%
Faster-rcnn	80%	76.81%	82.82%	77.8%

According to the training results shown in Table 9, the Thangka dataset provides important feedback on the performance of different network models. Overall, the YOLO series models perform exceptionally well on this dataset, particularly YOLOv5, which achieves an F1 score of 94% and an mAP@0.5 of 97%, demonstrating very high precision and recall. This indicates that the Thangka dataset is of high quality and diversity, providing sufficient feature support for complex deep learning models, enabling them to accurately identify and classify categories. YOLOv7, YOLOv8, YOLOv10, and YOLOv11 also show similarly strong performance, with F1 scores of 91%, 91%, 87%, and 92%, and mAP values of 94.8%, 95.4%, 93.5%, and 95.8% respectively. This shows that the Thangka dataset has good adaptability and representativeness for YOLO models.

In contrast, Faster-RCNN performs relatively weaker on this dataset, with an F1 score of 80% and an mAP@0.5 of 77.8%. While its recall reaches 82.82%, indicating that it can detect most of the targets in the dataset, its precision is only 76.81%, suggesting a higher false positive rate. This may reflect that some categories in the Thangka dataset are more complex or imbalanced, causing Faster-RCNN to struggle in distinguishing these categories.

Although the Thangka dataset constructed in this study demonstrates high precision and recall across various mainstream object detection models-with some models achieving 100% precision and recall at an IoU threshold of 0.5-this does not imply that the detection task is simple or lacking in challenge.

These results may be attributed to several contributing factors. First, the dataset underwent meticulous expert annotation, with clearly defined categories and some objects possessing distinctive visual features, which facilitates accurate detection under standard IoU=0.5 conditions. Second, certain categories have a relatively small number of samples that are concentrated in images with clear and prominent features, which may make it easier for models to learn and recognize them. In addition, some images in the test set may share stylistic or compositional similarities with those in the training set, potentially leading to inflated performance due to reduced domain shift.

However, model performance still drops at higher IoU thresholds (e.g., mAP@0.75 and mAP@0.95), indicating that precise boundary localization remains a significant challenge. More importantly, the dataset contains a large number of fine-grained and semantically similar categories-such as various types of ritual implements and Buddha figures-along with densely distributed targets and common real-world complications such as occlusion, blur, and varying lighting conditions. These factors greatly increase the difficulty of detection and classification, particularly for small, overlapping, or boundary-blurred objects. The subsequent qualitative analysis results further reflect these complexities.

For example, Faster-RCNN performs notably worse than the YOLO series in distinguishing these semantically similar categories, with lower precision and mAP scores reflecting its limitations. These differences further validate the dataset's complexity and establish it not only as a valuable training resource, but also as a challenging benchmark for fine-grained object detection in the field of cultural heritage.

Therefore, while the high performance achieved under certain conditions is encouraging, it should not be interpreted as evidence of a simplistic task. Future work may focus on improving model robustness in detecting small objects, differentiating fine-grained categories, and achieving precise boundary localization under more stringent evaluation standards.

4.3 Qualitative Analysis of Detection Results

As shown in Figure 3(a), the model successfully identifies multiple target categories in a representative Thangka image, including Shakyamuni Buddha, Manjushri Bodhisattva, and Four-Armed Avalokiteshvara, while also accurately localizing detailed elements such as hair bun and alms bowl. Even under conditions of complex image structure and dense object distribution, the model achieves high-confidence detection, demonstrating the effectiveness of the dataset constructed in this study in enhancing the model's fine-grained recognition capability.

On the other hand, Figure 3(b) presents failure cases in model prediction. Specifically, the model tends to produce misclassifications when dealing with targets that have high semantic similarity, such as between certain Buddha and Bodhisattva figures or among different types of ritual implements. In situations where the targets are heavily occluded or parts of the image are blurred, the model may fail to detect all objects correctly, resulting in missed detections. Moreover, in Thangka images with dense elements and complex compositions, the model often struggles to accurately localize every small object, leading to overlapping bounding boxes or incomplete recognition. These issues not only highlight the challenges inherent in understanding Thangka art but also reveal the current limitations of the model in small object detection, boundary sensitivity, and semantic discrimination. The primary causes of these failures include: (1) high visual similarity among certain object classes, making fine-grained classification difficult; (2) occlusions and image blur that reduce the visibility of critical features; (3) dense arrangements of small objects that strain the model's localization accuracy; and (4) insufficient sensitivity to object boundaries under complex backgrounds. These limitations suggest important directions for future improvements.

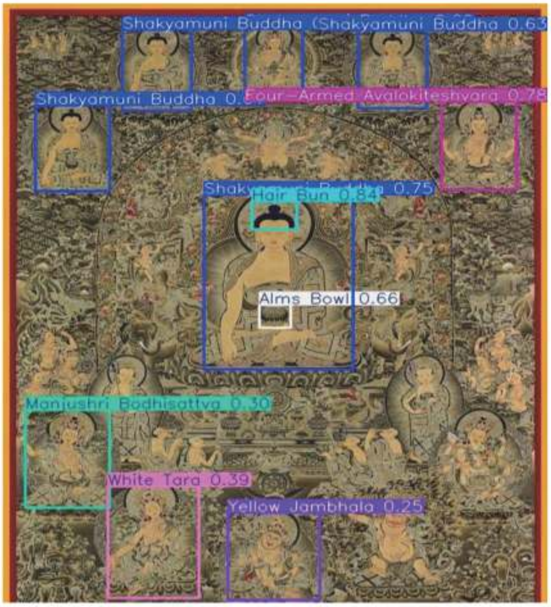

Shakyamuni Buddha (Shakyamuni Buddha 0.63
Shakyamuni Buddha 0.
Four-Armed Avalokiteshvara 0.78
Shak
Hair Bun 0.84
0.75
Alms Bowl 0.66
Manjushri Bodhisattva 0.30
White Tara 0.39
Yellow Jambhala 0.25

Figure 3(a). Good detection results.

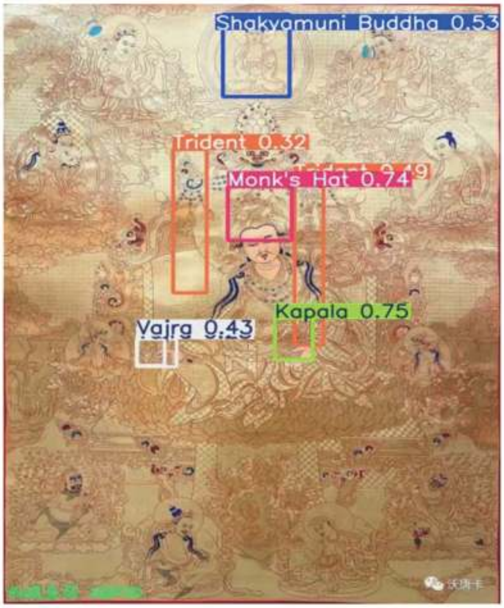


Figure 3(b). Poor detection results.

4.4 Evaluating the Effectiveness of the Thangka Dataset Through Comparative Analysis

4.4.1 Overview of Comparative Datasets

To evaluate the effectiveness of the Thangka dataset constructed in this study for object detection tasks, we selected three existing datasets (hereinafter referred to as Dataset A [27], Dataset B [28], and Dataset C [29]) as comparative benchmarks. These datasets are widely used in cultural heritage research and fine-grained object detection tasks, making them relatively representative. Although the specific names of the datasets are not explicitly disclosed in public sources, their data structures and annotation formats are compatible with the YOLO format, thus ensuring comparability. Table 10 summarizes the key properties of the three comparative datasets, including the number of images, number of categories, annotation formats, and annotation quality.

Table 10. Basic Attributes of the Comparative Datasets.

Attribute	Dataset A	Dataset B	Dataset C
Number of Images	3100	2954	2100
Number of Categories	35	15	15
Annotation Format	YOLO	YOLO	YOLO
Public Availability	No	Yes	No
Annotation Quality	Medium	High	Medium

4.4.2 Detection Performance Comparison Results

The evaluation metrics mainly include mAP@0.5 and mAP@0.5:0.95 (i.e., mean average precision across different IoU thresholds), as shown in Table 11:

Table 11. Detection Performance Comparison Results.

Dataset	Model	mAP@0.5	mAP@0.5:0.95
Dataset A	YOLOv5	72.5%	56.1%
Dataset B	YOLOv7	84.5%	56%
Dataset C	YOLOv7	88.5%	50.5%
Thangka Dataset	YOLOv5	97%	76.5%
Thangka Dataset	YOLOv7	94.8%	76.2%

The experimental results indicate that, compared to the other three datasets, the Thangka dataset constructed in this study achieves higher detection accuracy in terms of mAP, demonstrating its advantages in object detection tasks. This improvement is primarily attributed to finer annotation quality, a more diverse category coverage, and higher image resolution. Overall, the proposed Thangka dataset exhibits significant application potential in cultural heritage digitization and AI-driven Buddhist art analysis. Additionally, it provides valuable resources for future research in multimodal learning and cross-domain adaptation.

4.5 Scale Distribution Analysis of Target Bounding Boxes

In object detection tasks, understanding the spatial distribution of the targets in the dataset and the scale characteristics of the bounding boxes is critical for model training. To achieve this, we performed a statistical analysis of each target's bounding box attributes (x, y, width, height) in the dataset and visualized the correlations between these attributes and their independent distributions using joint distribution plots as shown in Figure 4.

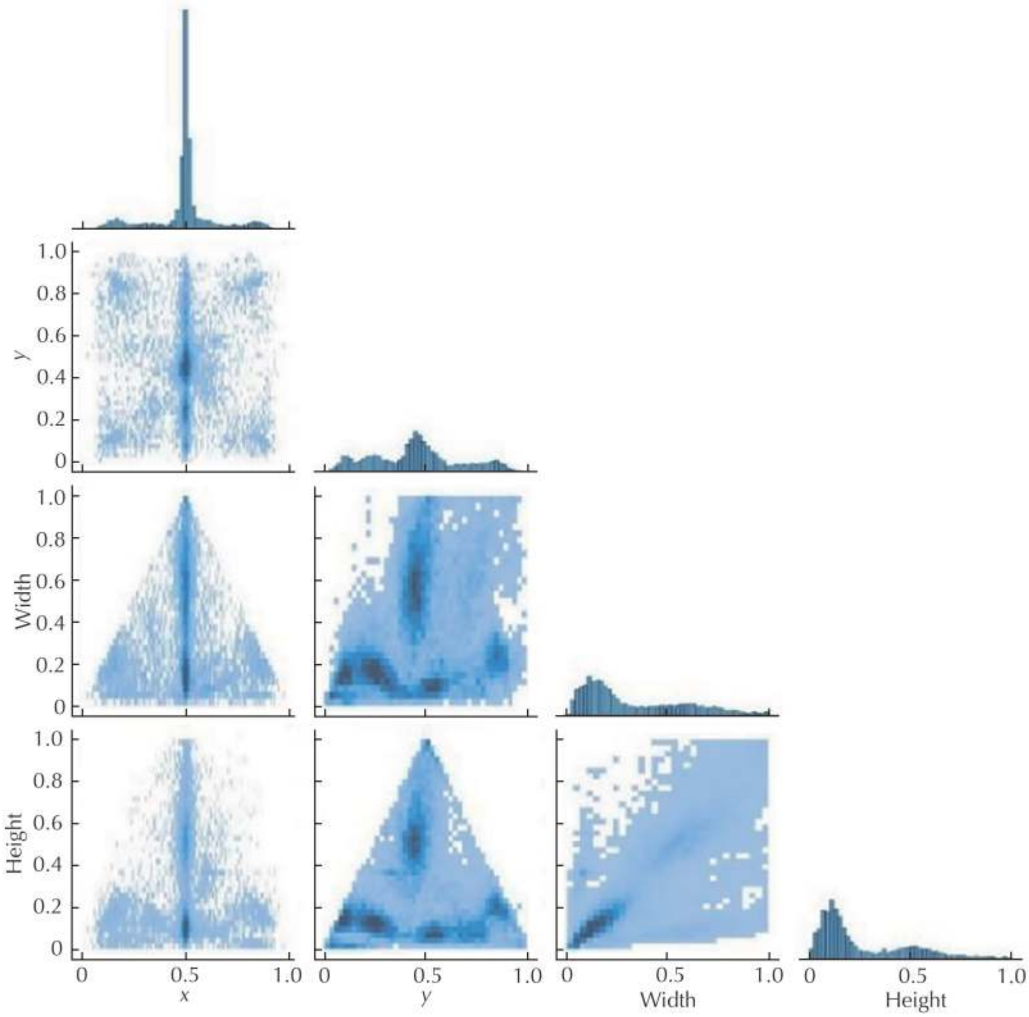


Figure 4. Joint Distribution Plot of Bounding Boxes.

4.5.1 Spatial Distribution of Bounding Boxes

The distribution of *x* and *y* coordinates: From the histograms, it can be seen that most of the target centers (*x*, *y* coordinates) in the dataset are concentrated in the center of the image, around (0.5, 0.5). This distribution indicates that the majority of objects in the dataset have been annotated near the center of the images. In the scatter plot, the distribution of *x* and *y* coordinates is concentrated in the center of the image, with fewer targets located near the edges.

4.5.2 Distribution of Bounding Box Sizes

Width and height distribution: From the histograms of width and height, it is clear that the majority of bounding boxes in the dataset are relatively small, with the width and height primarily concentrated between 0.2 and 0.3. This indicates that most of the targets in the images are relatively small in size.

Relationship between width and height: The joint distribution plot of width and height shows a positive correlation between the two. This suggests that most of the targets in the dataset have consistent aspect ratios, with no extreme distortions or particularly large objects present.

5. CONCLUSION AND FUTURE WORK

This paper establishes a systematically organized Thangka dataset, providing a digital resource for the study of the cultural, historical, and religious values of Thangka art. The dataset includes high-resolution images and classification labels, which can be used for image feature extraction, machine learning model training, and cultural preservation research in multiple fields. Through this data, scholars can better understand the uniqueness of Thangka art and promote the application of digital methods in cultural heritage preservation. This paper provides the theoretical foundation and technical support for the digital preservation and analysis of cultural heritage, laying a solid basis for future interdisciplinary research.

In future work, the Thangka dataset can be further expanded by incorporating a more diverse range of Thangka artworks and improving the annotation system to capture more detailed elements and iconographic features. Additionally, by integrating advanced artificial intelligence technologies such as deep learning, the potential for enhancing automated classification and recognition capabilities could be realized, further supporting the preservation and study of Thangka art. Interdisciplinary collaboration in the future could explore how this dataset can be applied in virtual exhibitions, educational platforms, and augmented reality experiences, enabling broader dissemination and appreciation of Thangka art worldwide. Moreover, the expansion and updating of the dataset will offer more innovative directions and opportunities for academic research, cultural heritage preservation, and digital applications.

DATA AVAILABILITY STATEMENT

All the data are available in the Science Data Bank repository (https: //www.scidb.cn/en/s/JJ7Nri) under an Attribution 4.0 International (CC BY 4.0).

AUTHOR CONTRIBUTIONS

Zhuo Li contributed to conceptualization, dataset construction, methodology, experiments, data curation, and writing the original draft. Jing Zhang assisted with data collection and literature review. Chunyan Peng provided supervision, project administration, funding acquisition, and writing review and editing. All authors have read and approved the final manuscript.

ACKNOWLEDGMENTS

Qinghai Province Key R & D and Achievement Transformation Project (No. 2022-GX-155); National Natural Science Foundation of China (Grant No: 62262056); National Natural Science Foundation of China (Grant No: 62441609).

REFERENCES

[1] R. Wen and F. Fan, "Quantifying pigment features of Thangka Five Buddhas using hyperspectral imaging," *Journal of Cultural Heritage*, vol. 70, pp. 120–133, 2024.

[2] Y. Chen, Z. Fan, and X. Liu, "RPTK1: A new Thangka dataset for object detection in Thangka images," *IEEE Access*, vol. 9, pp. 131696–131707, 2021.

[3] S. Wang and F. Fan, "Thangka Hyperspectral Image Super-Resolution Based on a Spatial-Spectral Integration Network," *Remote Sensing*, vol. 15, no. 14, p. 3603, 2023.

[4] S. Zhang, "The Portrait of Buddha: An Anthropological Approach Towards the Tibetan Religious Painting-The Thangka," *Contact Zone*, vol. 10, pp. 72–104, 2018.

[5] J. Sun and L. Huang, "Thangka Art Preservation and Cultural Industry Development from the Perspective of Social Change," *Sustainable Development*, vol. 14, no. 6, pp. 1386–1394, 2024.

[6] S. Xu, "Protection and Development of Intangible Cultural Heritage in the Digital Media Era-Taking Thangka Art APP Platform as an Example," M.S. thesis, Jiangxi Normal Univ., 2018.

[7] L. Wang and J. Zhang, "The Role of Digital Technology in Thangka Art Preservation," *Journal of Cultural Heritage Studies*, vol. 21, no. 3, pp. 203–215, 2019.

[8] J. Zhang, Y. Li, X. Wang, and Z. Liu, "Thangka Image Captioning Model with Salient Attention and Local Contextual Features," *Nature*, vol. 60, no. 2, pp. 123–135, 2024.

[9] Y. Ma, Y. Liu, Q. Xie, S. Xiong, L. Bai, and A. Hu, "A Tibetan Thangka data set and relative tasks," *Image and Vision Computing*, vol. 108, p. 104125, 2021.

[10] Z. Li and H. Zhang, "The Application of Machine Learning in Tibetan Thangka Art," *IEEE Trans. on Cultural Heritage Preservation*, vol. 17, no. 3, pp. 312–325, 2022.

[11] L. Xu and R. Chen, "Cultural Heritage and Digital Archives: The Case of Thangka," *Journal of Digital Heritage*, vol. 9, no. 2, pp. 123–135, 2020.

[12] F. Liu and Y. Wang, "Digital Archives of Thangka Art: Challenges and Opportunities," *Journal of Cultural Preservation*, vol. 6, no. 4, pp. 112–124, 2019.

[13] J. He and X. Li, "Thangka Art Recognition Using Convolutional Neural Networks," *Journal of Machine Vision*, vol. 18, no. 2, pp. 243–255, 2021.

[14] Q. Zhou and Y. Zhang, "Thangka Image Segmentation Based on Deep Learning," *Journal of Artificial Intelligence Research*, vol. 30, no. 5, pp. 456–470, 2022.

[15] Y. Chen and F. Wang, "The Digital Reconstruction of Tibetan Thangka Artworks," *Journal of Visual Arts and Digital Media*, vol. 11, no. 1, pp. 78–89, 2018.

[16] S. Liu and D. Zhao, "Preserving Tibetan Thangka through Digital Technology: A New Approach," *Journal of Cultural Heritage Preservation*, vol. 8, no. 3, pp. 45–57, 2019.

[17] X. Yang and W. Liu, "A Deep Learning Approach to Thangka Art Classification," *IEEE Trans. on Image Processing*, vol. 29, pp. 576–589, 2020.

[18] R. Zhang and T. Wu, "A Study on Thangka Image Data Annotation Using Machine Learning," *Journal of Digital Humanities*, vol. 7, no. 4, pp. 178–192, 2021.

[19] X. Chen, L. Ji, Z. Wang, et al., “Thangka Super-Resolution Diffusion Model Based on DCT-Upsample and High-Frequency Focused Attention,” 30 Aug. 2024.

[20] L. Zhang, *Study of Thangka Images*. Lhasa, Tibet: Tibet Publishing House, 2019.

[21] M. Li, *The World’s Most Beautiful Thangka-Ritual Implements in Thangka*. Xining, Qinghai: Qinghai People’s Publishing House, 2020.

[22] P. Wang, *Thangka Art (Complete Illustrated)*. Beijing, China: China Tibetan Studies Publishing House, 2018.

[23] H. Bi, W. Liu, and Y. Zhao, “Research on Thangka Headgear: A Study of Iconography and Artistic Style,” *Journal of Tibetan Art and Culture*, vol. 29, no. 3, pp. 145–160, 2021.

[24] Palace Museum, *Thangka Illustration of the Forbidden City*. Beijing, China: Forbidden City Publishing House, 2015.

[25] Nianduhu, *Nianduhu’s Thangka Art from Regong*. Xining, Qinghai: Qinghai People’s Publishing House, 2017.

[26] T. Jiansen, *Complete Collection of Tibetan Thangka*. Lhasa, Tibet: Tibet People’s Publishing House, 2018.

[27] X. Zhang, Y. Zhao, and Y. Zhao, “Automatic detection algorithm of Thangka elements based on circular smoothing YOLOv5-Ghost,” *Journal of Shanxi University (Nat. Sci. Ed.)*, vol. 46, no. 2, pp. 342–351, 2023.

[28] S. Guo, C. Peng, and X. Zhang, “Improved YOLOv6s algorithm for automatic detection of elements of Qinghai Farmer Painting and Thangka,” *Electronic Engineering and Informatics*, G. Izat Rashed (Ed.), doi:10.3233/ATDE240125, 2024.

[29] H. Li, X. Zhang, Y. Zhao, and C. Peng, “Improved YOLOX-Based Object Detection Algorithm for Thangka Murals,” *Computer Engineering and Applications*, vol. 60, no. 18, pp. 248–255, 2024.

[30] Z. Tang, C. Pan, and X. Zhang, “Thangka classification based on omni-dimensional dynamic convolution and ResNet,” in *Proc. 2024 8th Int. Conf. Graphics and Signal Processing (ICGSP)*, vol. 8, 2024.

[31] W. Hu, Q. Lang, W. Kang, and X. Shi, “Semantic concept-guided and multimodal feature-optimized Thangka image description method,” *Journal of Imaging*, vol. 9, no. 8, p. 162, 2023.

[32] K. Wang, T. Wang, X. Guo, K. Xu, and J. Wu, “Thangka image text matching based on adaptive pooling layer and improved Transformer,” *Applied Sciences*, vol. 14, p. 807, 2024.

[33] J. Bai, Y. Fan, and Z. Zhao, “Thangka image restoration with discrete codebook in collaboration with Transformer,” *Multimedia Systems*, vol. 30, p. 238, 2024.

AUTHOR BIOGRAPHY

Chunyan Peng, female, PhD, Professor, is a "High-end Innovation Talent" within the prestigious high-level talent cultivation program of Qinghai Province, affiliated with Qinghai Normal University. She also serves as a supervisor for master's degree candidates. Her current research endeavors are concentrated on the digital preservation of intangible cultural heritage, intelligent information processing, as well as teaching and research endeavors in the field of computer science. She has participated as a key researcher in one National Key R & D Project, one Ministry of Education project, one National Natural Science Foundation project, and one Qinghai Province Key R & D Project. Her research achievements have received one Third-Class Provincial Progress Award and one Third-Class Qinghai Province Science and Technology Paper Award. She has published 20 papers indexed by SCI/EI/ISTP, applied for 2 national invention patents, and obtained 18 software copyrights.

Zhuo Li, male, obtained a BS degree in Computer Science and Technology from Xinyang University in 2022. Currently, he is pursuing a Master's degree at the College of Computer Science, Qinghai Normal University, with a research focus on computer vision technology.

Zhang Jing, who graduated with a Bachelor's degree from Shandong Normal University in 2023, has not halted her academic journey. Currently, she is pursuing a Master's degree at the School of Computer Science, Qinghai Normal University, concentrating on research in the field of image vision. In this area, she is committed to exploring the application of deep learning techniques in image recognition and processing, as well as how to improve the precision and efficiency of image analysis through the use of computer vision technology.