IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 10, 2021, accepted September 16, 2021, date of publication September 20, 2021,
date of current version September 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3113906

RPTK1: A New Thangka Data Set for Object Detection of Thangka Images

YUHONG CHEN, ZHEN FAN, AND XIAOJING LIU
Department of Computer Technology and Applications, Qinghai University, Xining 810016, China
Corresponding author: Xiaojing Liu (645020710@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61862053, and in part by Qinghai Provincial Department of Science and Technology Applied Basic Research Project under Grant 2021-ZJ-717.

ABSTRACT Data set is the basis of machine learning, a good data set can promote the development of various applications. Machine learning has been deeply involved in the protection and inheritance of cultural resources. However, there are few data sets about Thangka, and the types and quantity of Thangka images are relatively few. Therefore, we first establish a Thangka data set called RPTK1 (Religious Portrait Thangka Version 1), which contains 3,338 Thangka images, more than any other Thangka data set. The objects in the data set basically cover the common Thangka religious portraits, tools and headdresses, and are marked in the professional language of Buddhism. Then, on the basis of the RPTK1 data set, in order to achieve better detection of small Thangka objects (Thangka religious tools), we propose an improved Single Shot MultiBox Detector (SSD) method, called Single Shot MultiBox Detector with Improvement Feature Fusion And Loss Function (FALSSD). Finally, in order to verify the effectiveness of the FALSSD method, we conduct experiments on the RPTK1 data set. The experimental results show that the mAP of our method in the RPTK1 data set reaches 83.85%. Compared with the other ten state-of-the-art methods, the performance of our model is better.

INDEX TERMS Thangka data set, object detection, SSD method, feature fusion.

I. INTRODUCTION

In recent years, machine learning has developed rapidly, from natural scenes to paintings, comics, etc., but the premise is that machine learning needs data sets to support. The research field of machine learning has developed from natural scenes such as Pascal VOC [1], COCO [2] and ImageNet [3] to unnatural scenes such as Webcaricature [4], Danbooru2020 [5] and painting [6], and has achieved interesting and considerable effects. We focus on the study of Thangka painting scenes. Thangka painting is purely handmade. They are religious paintings on canvas or papers. Most of the contents of Thangka are about Tibetan Buddhism, the astronomical calendar and mythology [7]. No matter what kind of application scenario, all need a good data set to support.

At present, there are many scholars studying Thangka image understanding. The research of Thangka image mainly focuses on Thangka image restoration [8], [9], Thangka image retrieval [10], [11], classification [12], [13] and segmentation [14], [15]. However, there are some problems in their work, such as the small Tangka data set and the imbalance of classification; The annotation of image object detection data set is not standardized, and it is only expressed with simple words, and there is no detailed classification for different types of Thangka religious portraits [16]. The low accuracy of small object detection in Thangka image leads to less and less understanding of Thangka, and the understanding of Thangka image is relatively backward.

Based on this, we propose a new work: Thangka data set-RPTK1. RPTK1 is used for the detection of religious portraits, tools and headdresses in Thangka images. The collected data set contains a total of 3338 images of Thangka religious portraits, with a total of 57 labels of Buddhist professional terms. The religious portraits, tools and headdresses in each image are labeled by the labeling tool LabelImg [17], and the labels are named by the Buddhist professional names of each kind of religious portraits, tools and headdresses in Thangka images. Thangka images in the RPTK1 data set are rich in content, numerous composition elements and

The associate editor coordinating the review of this manuscript and approving it for publication was Ting Wang.

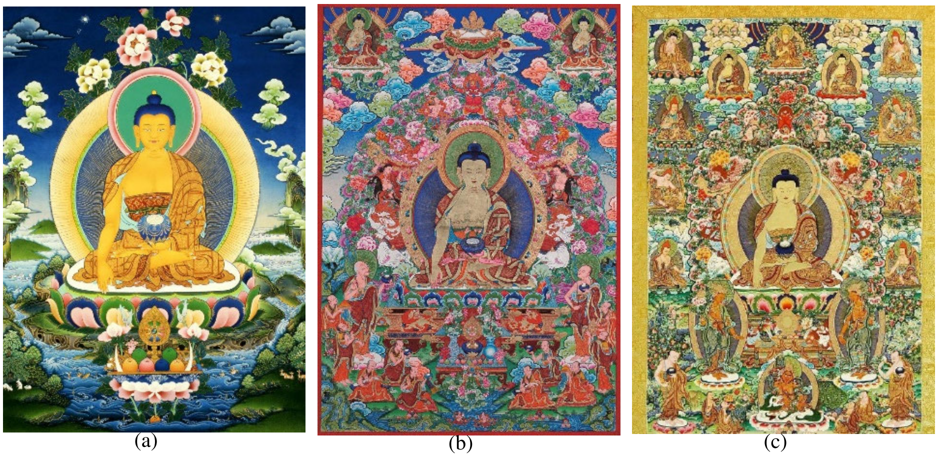


FIGURE 1. Same type of Thangka images with different complexity (a) Simple shakyamuni Buddha.(b)Complex shakyamuni Buddha.(c)The more complex Shakyamuni.

various forms of images. The main image of the same type of Thangka has different forms of expression. In Thangka, there is not only the main image of Thangka but also other complex elements of the Buddha image and background. As shown in Figure 1 below, Thangka images of Shakyamuni Buddha with different complexity are shown. It is not difficult to see from the figure that, compared with general images, Thangka images have rich content, complex background and extremely many composition elements. Figure 1(c) not only contains Lord and Buddha but also other Buddhas and other detailed parts. The classes are quite different and the forms are more abundant.

On this data set, we can perform object detection tasks. We try and propose a new SSD-based object detection method. Aiming at the poor detection effect of small objects in the RPTK1 data set, we propose a new feature fusion method, and a new positioning loss function to improve the accuracy of final detection. Compared with the other ten state-of-the-art methods, the performance of our model is better. Our new data set can be used for image object detection, image restoration, 3D modeling, etc., in order to inspire researchers and promote the scientific research of Thangka in a new way. Attract more Thangka researchers and ordinary people to protect and rescue Thangka and make it easier to appreciate Thangka.

Our main contributions are summarized as follows:

- We construct a data set RPTK1, which contains 3338 Thangka images. There are 3 types of headdresses, 18 types of religious tools, and 36 types of religious portraits among 57 classes. The main objects of Thangka images are manually marked, and the labels are more in line with the professional terms of Buddhism.
- We propose a new method for object detection in Thangka images based on SSD model. Feature fusion is used to solve the problem of small object detection in Thangka images. In addition, we propose a new loss function to improve the accuracy of location positioning.
- We also compare the most state-of-art object detection methods with our method on the data set RPTK1. Our experiments show that our FALSSD model has a better detection effect on the religious portraits and headdresses in Thangka images, and It also has an impact on smaller objects such as religious tools.

II. RELATED WORK

A. THANGKA

Thangka is a religious scroll painting on colored silk with pigments or other materials. It is a form of painting with local characteristics [18], and it is also one of the important intangible cultural heritage in China. The content of Thangka is rich, including buildings, medicine, astronomical calendar, legendary stories, etc. [19]. In addition, Thangka has a wide range of contents, bright colors, clear layers, unclear foreground and background, complex compositions and rigorous structures. First, according to the original writing method of religious portrait, Thangka painting has strict requirements on facial features, head, chest, waist and other parts. Thangka painters usually need many years of practice to draw a good Thangka [20], it is difficult to preserve Thangka paintings. Due to the influence of humidity, water stains, mildew and dirt, the number of images of well-maintained Thangka paintings is decreasing.

At present, people pay more and more attention to Thangka. Huaming *et al.* [21] supervise Thangka headdress through self-encoding, and then input the labeled training samples after the convolution pool operation process to train the soft maximum classifier. Finally, the performance of the classifier is tested by using the test set samples, and the method has good classification performance in Thangka headdress classification; In order to avoid errors being ignored in the training process of imbalanced Thangka image data set, Zeng *et al.* [12] use Resnet to improve the accuracy and convergence speed of imbalanced Thangka image classification; In addition, He *et al.* [22]

propose an improved Densely Connected Convolutional Networks(DenseNet) model to solve the problem of less Thangka images. Compared with DenseNet, the performance is improved by 1.1%. This method has better results in Thangka image classification; Ma *et al.* [16] provide a Thangka data set called Chomo Yarlung Tibet version 1(CYTKv1), which contains 1700 Thangka images, and they annotate Thangka images by hand, and finally, they release the detection baseline on this Thangka data set. In the third part, we elaborate some advantages of RPTK1 data set over CYTKv1 data set.

B. DATA SETS

Data sets play an important role in machine learning. Data sets also cover a wide range of fields and forms, such as finance [23], [24], transportation [25]–[27], medical treatment [28], human face [29], video [30], [31], etc.

A good point cloud data set is the key step of semantic segmentation to understand complex scenes in depth. Tong *et al.* [32] construct a new point cloud data set for semantic segmentation of large-scale scenes. This data set mainly includes six types of objects, such as ground, car, building, vegetation, bridge and pole, and has the advantages of more complete scenes and relatively uniform point density. Finally, an effective method is proposed, and the highest intersection (IOU) of poles, ground, buildings, cars, vegetation and bridges is 36.0%, 97.8%, 93.7%, 65.6%, 92.0% and 69.6%, respectively, for all benchmarks. In order to identify high-quality tea buds, Wang *et al.* [33] put forward a new deep neural network method to identify picking points of buds and establish an image data set of picking points of high-quality tea buds. Finally, the accuracy rate on this data set exceeds 90%. Insulators are an important part of transmission lines. Therefore, Yang *et al.* [34], in order to better detect insulators, construct an insulator data set while proposing the detection method. In this data set, the detection accuracy of the method they proposed for insulator defects is 98.38%, which is higher. In order to improve the accuracy and efficiency of casting defect identification, Duan *et al.* [35] construct their own casting defect detection data set through data preprocessing and labeling, and propose an improved YOLOV3 model, which is 26.1% higher than the original YOLOV3 model mAP. In the aircraft detection of remote sensing images, it is difficult to detect the aircraft due to the small object of the aircraft. Therefore, Wu *et al.* [36] experiment with the improved method on the remote sensing aircraft images constructed by themselves, which results in a lower false-positive rate and shorter training time. Schumann *et al.* [37] propose a new car radar data set, which contains measured values and point-by-point markings from the same car for more than four hours. The purpose of this data set is to develop a radar perception method based on machine learning for mobile road users. Maier *et al.* [38] propose the first publicly available colorectal data set. This data set comprises 30 laparoscopic videos and corresponding sensor data from medical devices in the operating room for three different types of laparoscopic surgery. The methods for medical device detection and segmentation can be fully tested.

C. SSD BASED METHODS

The SSD method is the representative of one-stage object detection network [39]. The biggest characteristic of one-stage method is that it can be completed in one step. It is a method to predict the classification and location of the object by directly inputting the image and extracting features from the network, to reduce the number of candidate regions. And all the boundary boxes can be predicted only once sent into the network. SSD directly returns the category and position of the object in the network, so the detection speed is faster. In the network with an input size of 300 × 300, NVIDIA TITAN can achieve 74.3% mAP and 59 FPS on VOC 2007 test set.

However, the problem is that the feature map extracted in the shallow layer is not strong enough, which leads to the low detection accuracy of SSD method for small objects. Therefore, the Deconvolutional Single Shot Detector(DSSD) method has improved the SSD method, which changing the backbone network of SSD from Vgg16 to Resnet. DSSD uses the deconvolution layer and skip connection to better characterize the shallow feature map, which can improve the accuracy of small object detection. Achieved 81.5% of mAP in the VOC2007 test, 80.0% of the mAP in the VOC2012 test, and 33.2% of the mAP on the COCO [40]; Considering that the SSD method independently uses deep and shallow layer feature maps for prediction, and its robustness to small objects is poor. Therefore, in Feature Fusion Single Shot Multibox Detector Motivation(FSSD) method, a feature fusion method is proposed to combine the different scales generated by Vgg16, the feature map is transformed into the feature map scale size of the conv4-3 layer by bilinear interpolation, and then all the feature maps are concatenating, and the obtained feature maps are re-subsampled to obtain different feature map scales, and then input into the perceptron network for prediction, and finally the detection accuracy on the Pascal VOC 2007 test set with an image size of 512 × 512 reaches 84.5% [41]; Deng *et al.* [42] combine inception and atrous convolution to propose a receptive filed block, which applies Receptive Field Block(RFB) to the head of the feature map of the SSD method, then predicts the processed feature map. The structure has a simple structure and excellent results, RFB Net300 is 80.5% and 83 FPS in mAP and FPS respectively.

Since the first several methods are too simple to make full use of the features of each layer, a new feature fusion method is designed in a Single-Shot Object Detector based on Multi-Level Feature Pyramid(M2Det) method to fuse the feature maps of SSD method. M2Det achieves 41% mAP on the COCO data set [43]; The Single Shot Refinement Neural Network for Object Detection(Refinedet) and the ideas of the previous several papers are different, mainly is to use the idea of the two-step method to first classify the priori box (binary classification, only distinguish the foreground and background) and regression, to solve the problem of extremely unbalanced samples, improve the accuracy; At the same time, FPN is used to fuse different feature maps to improve the robustness of small object detection, reaching 83.8% mAP on the VOC2007 data set with an input image size of 512 × 512 [44].

III. BUILD THE RPTK1 DATA SET

We collect a data set of Thangka religious portraits to study the detection of religious portraits, religious tools and headdresses in Thangka images. To easily access the data set, we have released the RPTK1 data set online using the Baidu cloud network disk.[1] In the construction of Thangka data set, the biggest challenge is that most elements in Thangka images are symbolic and contain Buddhist philosophy. Before labeling Thangka data set, we must have a certain understanding of this point and enough advanced semantic knowledge related to Buddhism. Ma *et al.* have also done the construction of Thangka data set. By comparison, our RPTK1 data set has the following advantages:

1.The annotation of our Thangka data set is based on communication with researchers. We summarize and annotate with the support of books such as "Introduction to Thangka Art" [45], "The World's Most Beautiful Thangka-Thangka Tools" [46], Bi *et al.* 's research on headdresses [47] and other relevant materials. Our RPTK1 data set contains more semantic knowledge related to Buddhism, and each label name has its corresponding name in Buddhism. As shown in Figure 2 below, the labeling of our RPTK1 data set is more professional, while the labeling of the CYTKv1 data set only uses simple words. In addition, our RPTK1 data set is more detailed in terms of classification, specific to each class of Buddha portraits, while CYTKv1 simply classifies all Buddha portraits into one class:

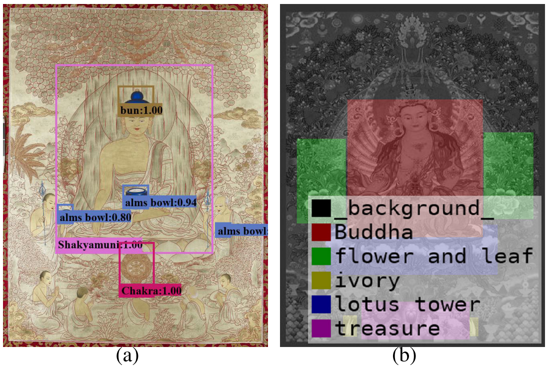


FIGURE 2. Schematic diagram of annotations of two Thangka data sets. From the comparison of the two pictures, we can see: first, the label of our RPTK1 data set is more professional. Second, our RPTK1 data set is more detailed in terms of classification, specific to each class of Buddha portraits, tools and headdresses.(a)annotations of RPTK1 data set.(b)annotations of CYTKv1 data set.

2.In our RPTK1 data set, the form of the main religious portrait is more abundant. There are two forms of Thangka Buddha in the same type of Thangka religious portraits: anger and compassion. As shown in Figure 3, Manjushri Bodhisattva has two forms, which are kind and quiet. We label Manjushri Bodhisattva uniformly for the transformation and the mighty transformation of wrath; The number of arms of Thangka religious portraits of the same type is different, and we also label them as the same type. For example, Mahakala has two arms, four arms and six arms. The image is marked as Mahakala. This is more abundant than Thangka religious portraits in the CYTKv1 data set:

3.Our RPTK1 data set has 3338 images, which is more than the CYTKv1 data set. The labels are more in line with the professional language of Buddhism. There are 57 labels related to Buddhist language, 3 labels related to headdress, 36 labels related to main religious portraits and 18 labels related to religious tools. As shown in Figure 4 below, it is the main statistical data diagram of the two data sets RPTK1 and CYTKv1:

The construction of Thangka image detection data set requires the following steps: first, collect Thangka images; second, preprocess Thangka images; third, label Thangka image data set. The following will be introduced separately:

A. THANGKA IMAGE COLLECTION AND PREPROCESSING

The main channels for collecting Thangka images are as follows: First, we use scanners to scan the illustrations in books related to Thangka, such as the "Forbidden City Thangka Picture Book" [48], "Regong Nianduhu Thangka Art" [49], "Tibet Thangka" [50], etc. The resolution of scanned images ranges from hundreds to thousands. Second, we obtain Thangka images through programs that automatically obtained from the web content, mainly for Baidu and Bing search engines, to obtain Buddha portraits images based on keywords such as "Thangka" and "Sakyamuni Buddha". Then we remove incomplete images (images related to Thangka images) and those that do not meet the requirements of Thangka image data set. Third, we consult relevant local Thangka technicians or well-known Thangka painters, we take and scan Thangka images, and we collect large-scale Thangka images in sections, then we conduct research on Thangka images stitching.

After collecting Thangka images, we simply process all Thangka images. First, We delete images that do not belong to or contain Thangka images; then, we use OpenCV to obtain the histogram data of Thangka images, and then normalize it, we use similarity detection to delete pictures with a similarity of less than 95%; finally, we need to convert all image suffixes into a unified jpg format. Figure 3 below is an example of the different types of Thangka images we have collected, and different Thangka religious portraits and tools have different meanings: As shown in Figure 3(a) above, Shakyamuni is the founder of Buddhism. After becoming a Buddha, Shakyamuni is honored as Buddha, meaning that those who fully understand the universe and the truth of life can save all lives. He usually holds a bowl in his left hand, which is often placed on his knee, symbolizing the wisdom that can directly realize

[1] https://pan.baidu.com/s/1LkhgHhLqY6xmLYbW6vz1lg

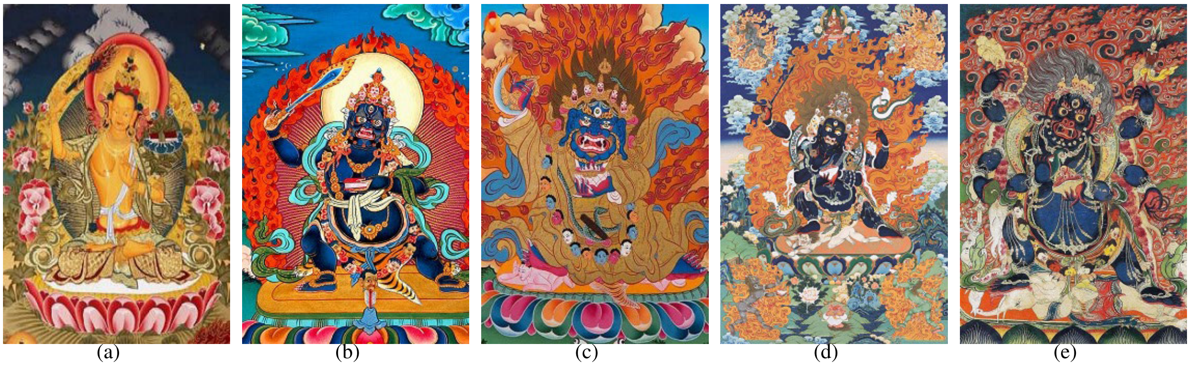


FIGURE 3. The abundant form of the main religious portraits.(a)Compassionate Manjushri.(b)Angry Manjushri.(c)Mahakala with two arms.(d)Mahakala with four arms.(e)Mahakala with six arms.

TABLE 1. The main statistical data diagram of RPTK1 and CYTKv1.

Name of data set	Images in data set	Total labels	Max labels of one image	Min labels of one image	Mean labels of one image	Total instances
RPTK1	3778	57	24	1	7	23665
CYTKv1	1778	79	16	1	5.5	11590

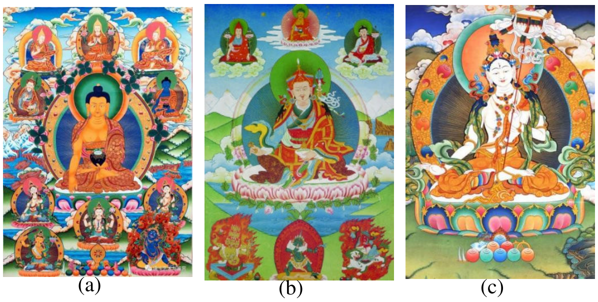


FIGURE 4. The different kinds of Thangka images. (a) Sakyamuni Buddha. (b) Padmasambhava. (c) Ushnisha Sita Tapatra.

emptiness. Figure 3(b) shows Master Padmasambhava, he is sitting in the center of the picture, with a skeleton stick wrapped around his left arm, a gabala bowl in his left hand and a diamond pestle in his right hand; the vajra is not only a simple weapon, but also a symbol of firmness, grandeur and majesty of Buddhist doctrine in a deeper spiritual sense; even the gabala bowl held by Master Peanut represents the cleansing of the mind and symbolizes the power of enlightenment that can transcend death. Figure 3(c) shows the Buddha Mother covered with a large white umbrella. Holding the big white umbrella in her left hand is her symbol. The meaning of the big white umbrella is to magnify the light and eliminate disasters and demons.

B. DATA ANNOTATION

After simple preprocessing, we then use the LabelImg labeling tool to label Thangka image data. The object in Thangka image is marked by the coordinates of upper left corner and lower right corner. The length and width of the marked rectangle should cover all the boundaries and contents of the object, and we name it by the name of the object in the Thangka image, then an .xml file will be automatically generated to save all coordinate information and label information. After marking, we construct this data set into Pascal VOC data set format. The annotations folder stores the .xml labels files obtained using the annotation tool; the ImageSets/Main folder stores the names of the training set, validation set, training validation set, and test set images. The pictures for subsequent code training are read from these files, we divide all the pictures into data set, the number of training set, validation set, and the test set is divided according to the ratio of 8:1:1; the original training and validation pictures are stored in the JPEGImages folder, and the verification pictures are randomly obtained from the training set. The main statistics of our RPTK1 data set are shown in Table 2 below: Figure 5 below shows the top 30 labels types in our RPTK1 data set, including religious portraits, tools and headdresses, in ascending order of the number of labels:

TABLE 2. Main statistics of RPTK1 data set.

Images in RPTK1	Total labels	Max labels of one image	Min labels of one image
3338	57	24	1
The most labels	Minimal labels	Total number of labels	Mean labels of one image
Crown	Dorije Legpa	23665	7

IV. THANGKA OBJECT DETECTION MODEL-FALSSD

Our method is an improved of SSD, mainly for the situation that SSD method fails to detect small objects and low-resolution objects. Due to the lack of semantic information of small objects, SSD uses low-level features to detect small objects, and the extracted features are insufficient, which leads to low accuracy of small object detection. We use

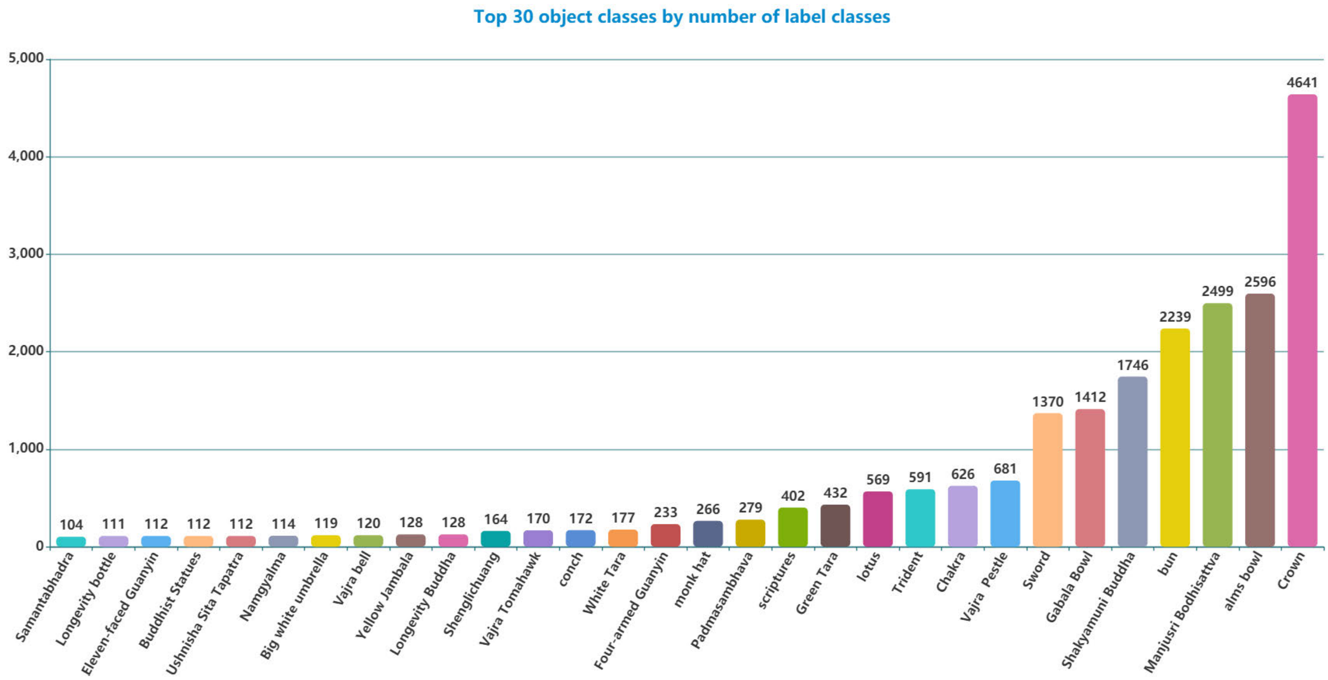


FIGURE 5. The number of the top 30 labels.

Resnet50 as the backbone network because when Vgg16 is used as the backbone network, only one shallow convolutional layer is used to extract features to predict small objects. However, due to the lack of relevant semantic information, the feature information extracted from shallow convolution layer can only be used for small objects. The effect of object detection is poor; Vgg16 lacks the ability to learn and express deep information; in the case of deepening the network, the gradient will disappear, resulting in slower learning rates. Therefore, we use Resnet50 as the backbone network to extract deeper image features, and use the deep residual network structure to solve the problem of gradient disappearance and extract deeper image features; In addition, the convolution layer after the third layer and the fourth layer in Resnet50 network structure is respectively upsampled to amplify the features, and then the convolution layer after the second layer and the two convolution layers after the upsampled are fused to improve the small object detection effect and object detection accuracy. The whole improved SSD object detection model is shown in Figure 6 below. With the deepening of the network depth, the gradient will decline, resulting in the decline of the accuracy of training set. In response to this phenomenon, we replace the backbone network in the original SSD network model of Vgg16 with Resnet50 to extract deeper image features and solve the problem of gradient disappearance. The Resnet50 network is deep and has many layers, so it consists of a 3 × 3 convolutional layer in the middle of two 1 × 1 convolutional layers, plus a zero-padding equivalent mapping. The 1 × 1 convolution layer at the beginning and the end is used to reduce and restore dimensions, to solve the problem of gradient disappearance and extract deeper image features, which makes the use of the backbone network of Resnet50 to greatly improve the detection accuracy. In order to improve the detection efficiency of the model, we upsample the layer 3 and layer 4 in Resnet50 network to amplify the features, and then add a convolution layer after the layer 2, and after that, we add a convolution layer after the upsampling. The convolution layer 5 is obtained by fusing the features of the two convolution layers. After the Batch Normalization(BN) layer, it passes through seven bottleneck layers, that is, the 1 × 1 convolution layer used in Resnet. Then, after sampling under the product layer and BN layer, three convolution layers are superimposed and input to the head network together. The predicted value is calculated according to the feature, and then the loss value is calculated with the real value of the object, and then input into the classifier. Finally, the final test result is obtained by non-maximum suppression.

A. FEATURE FUSION MODULE

The SSD model uses Vgg16 as the backbone network and shallow convolution conv4_3 for feature extraction to detect small objects. Because the features extracted by shallow convolution lack the relevant semantic information of small objects, the object detection effect is poor. Inspired by the DSSD model, we propose a feature fusion method to improve the drawback of poor detection of small objects. The feature fusion module we proposed is shown in the dotted line in Figure 7. Feature fusion is divided into the following three parts: Conv2 after Layer2, Conv3 upsampling after Layer3, and Conv4 upsampling after Layer4. Among them, the output of Conv3 upsampling is specified as twice the input and the

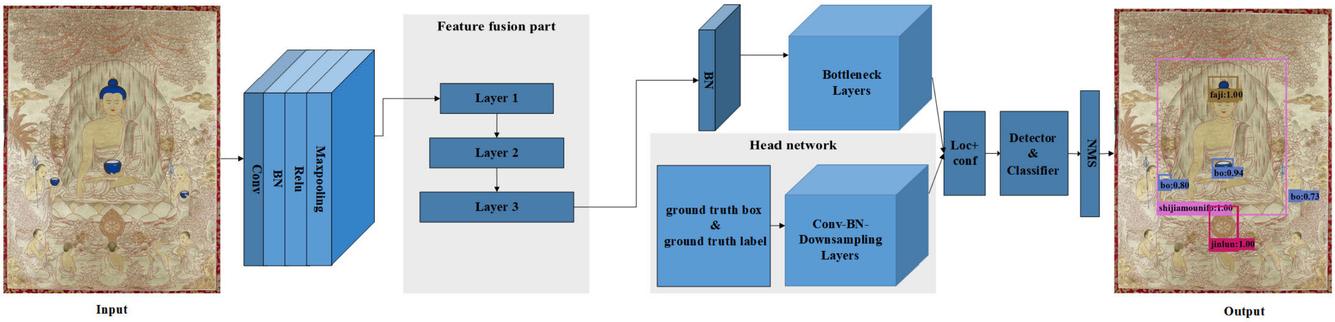


FIGURE 6. **Improved object detection model. The part that performs Conv-BN-Downsampling is the head network, whose input is calculated from the ground truth box and the ground truth label of the data set.**

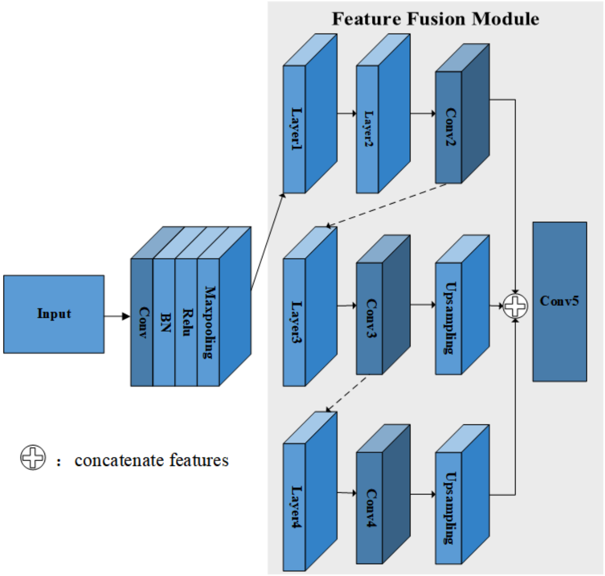


FIGURE 7. **Feature fusion module. The gray shaded part in the above figure is the specific structure of the feature fusion module.**

output of Conv4 upsampling is 38 × 38. We use the nearest neighbor interpolation algorithm, as shown in Figure 5 below, that is, 2D nearest neighbor up sampling for the input signal. We use the upsampling operation to amplify the extracted image features to increase the semantics of the low-level features, and then merge the three parts of the features to make the extracted features as diverse as possible. After the fusion of high-level features and low-level features, the model improves the low-level features with low semantics, multi defects in noise, and high-level features to perceive details at the same time. After combine the fusion layer of shallow local detail information and deep high semantic information, the fused features are used as the input of Conv5.

B. NEW LOSS FUNCTION

The loss function in SSD model consists of the weighted sum of the location loss function L_*loc* for classification and the confidence loss function L_*conf* for regression. As shown in Formula 1, the number of negative samples is much larger than the number of positive samples, so in this paper, we improve the optimization speed and the stability of training results by controlling the positive and negative sample ratio of SSD loss function [51]. The confidence loss function is softmax loss.

$$L(x, c, l, g) = \frac{1}{N}(L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \qquad (1)$$

where N is the number of positive samples of the prior box; c is the predicted value of category confidence; l is the predicted value of the position of the corresponding boundary box of the prior box; g is the position parameter of ground truth, and α represents the weight of both. Equation 2 is shown as the position loss function in SSD.

$$smooth_{L1}(x) = \begin{cases} 0.5x^2, & if\,(|x| < 1) \\ |x| - 0.5, & otherwise \end{cases} \qquad (2)$$

It can be seen from formula 2 and figure 8 that the loss function of Smooth L1 between [−1,1] is L2 loss, which makes up for the lack of smoothness of L1, and the other areas are L2 loss, which solves the problem of gradient explosion of the L2 loss function. Therefore, in order to make the location classification more accurate, in this article, we have improved the location loss function. On the basis of Smooth L1, we add the weight parameter gamma and use the parameter gamma to control the range in which the L1 loss is used (MAE loss function), in what range to use L2 loss (MSE loss function).

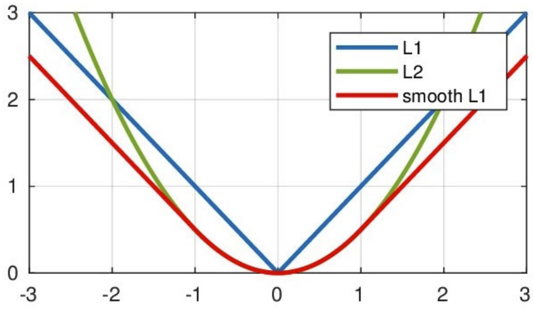


FIGURE 8. **Comparison of different loss function curves.**

The gamma parameter we used in the article is 1/9, and the specific formula is shown in 3:

$$smooth_{L1}(x) = \begin{cases} 0.5x^2/gamma, & if\,(|x| < 1) \\ |x| - 0.5 * gamma, & otherwise \end{cases} \quad (3)$$

C. EXPERIMENTAL ENVIRONMENT

We mainly conduct experiments on SSD method improvement, as well as experiments on other state-of-the-art methods. The training and testing process of all experiments are performed on NVIDIA GeForce GTX 1080Ti 12G GPU. We develop our model using the open-source deep learning framework PyTorch. We complete all the experiments on the Ubuntu Linux system equipped with NVIDIA GPU. We mainly use AP and mAP and FPS as the standard to evaluate the performance of our improved model. AP is the average accuracy, and the area enclosed by the pr curve and the coordinate axis is calculated using the integral method, as shown in formula 4. The pr curve refers to the curve obtained by taking the Recall value as the horizontal axis and the precision value as the vertical axis. Precious indicates how many of the predictions are correct, that is, the parameters that reflect the accuracy of the prediction, which is used to evaluate the correct rate of the prediction; while Recall is used to evaluate how many correct samples are predicted, that is, the total number of positive samples, how many successfully predicted.

$$AP = \int_0^1 p(r)dr \quad (4)$$

In the case of calculating the AP values for each class in the data set, the mAP sums and averages the AP values for each class. FPS is the number of frames transmitted per second, that is, the number of pictures that can be processed per second. All our experiment platforms are the same, all experiments are carried out on the same server, all use the same one GPU, we use a total of 101 Thangka images to calculate FPS, finally remove the speed of the first image. Therefore, the FPS of the remaining 100 images is averaged as the FPS of this model.

D. EXPERIMENTAL DETAILS

In all experiments, our Thangka image data set has 58 classes (plus background classes). We select 80% of Thangka images for training, about 10% images for verification, and finally randomly select 101 images from 10% for testing. For training, common data enhancement techniques, including zooming, moving, cutting, and flipping, are randomly activated, finally, 101 Thangka images out of order are used to test the FPS of the method. The whole experiment uses the Stochastic Gradient Descend(SGD) [52] method to optimize network parameters. According to the new real-time update model of Thangka training images, we use a warm-up learning rate. The initial learning rate is 1e-3 and the image size is 300 × 300, the batch size is 32; when the image size is 512 × 512, the batch size is 16. The maximum number of iterations is 120,000.

1) EXPERIMENTAL RESULTS

Our experiment is performed on the RPTK1 data set. Table 3 below is the result of our ablation experiment:

TABLE 3. Results of ablation experiment.

Resnet50(backbone)	Feature fusion	New loss function	mAP
			78.79%
✓			82.62%
✓	✓		83.47%
✓	✓	✓	83.85%

As can be seen from Table 3 above, the average accuracy of the original SSD model on the RPTK1 data set is 78.79%. After we replace the Vgg16 backbone network with Resnet50, the average accuracy increased by 3.83%, reaching 82.62%; then we use feature fusion, the average accuracy reached 83.47%. Finally, we change the backbone network to Resnet50 and adopt the feature fusion method and use the new loss function at the same time, the average accuracy on our RPTK1 data set reached 83.85%, which is the final result of our experiment.

2) BACKBONE NETWORK

In order to compare the effect of selecting Resnet50 for the backbone network, we conduct an ablation experiment under the condition of changing the backbone network. The details of the experimental results are shown in Figure 9, which includes the precision of each type of Thangka image object and the average precision of the whole data set. As shown in Figure 9 above, the average detection accuracy and detection rate of different backbone networks are shown. From the figure, we can know that compared with other backbone networks in the above figure, when Resnet50 is used as

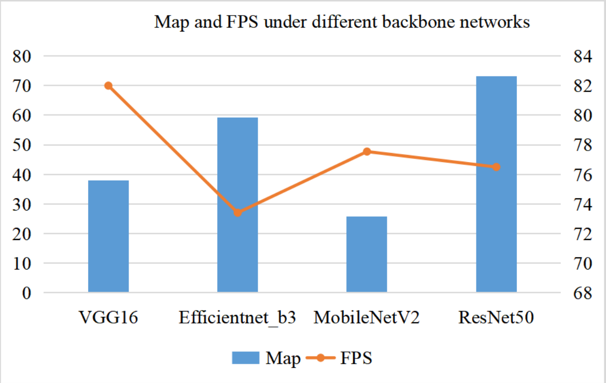


FIGURE 9. Average accuracy and FPS under different backbone networks.In the model inference process, the memory required by resnet50 reached 527.20MB, and Mobilenet v2 occupies the smallest memory of 280.23MB; but among the four backbone networks, Mobilenet v2 has the smallest parameter, only 10.97MB, and the largest is resnet50, which has reached 113.7MB.

TABLE 4. Fusion results of different modules in the backbone network Resnet50. CBAM is convolutional block attention module, SE block is Squeeze-and-Excitation networks.

CBAM	SE block	Feature fusion	Map	FPS
✓			82.55%	13.83
	✓		83.04%	29.91
		✓	83.47%	2

the backbone network, the detection accuracy is the highest. When we use Vgg16 as the backbone network, the detection rate is the highest. The proposed method of feature fusion for Resnet50 backbone network has higher detection accuracy.

3) FEATURE FUSION

With Resnet50 as the backbone network, we experiment on the improvement effect of feature fusion and add a convolutional block attention module, extrusion and excitation network, and feature fusion module to Resnet50 respectively for experiments. The experimental results are shown in Table 4 below. It can be seen from Table 3 above that when the attention module of the convolution block is added after the backbone network Resnet50, the average accuracy is only 82.55%, but after feature fusion, the average accuracy is the highest, reaching 83.47%. We put forward the method of feature fusion for the detection accuracy of ascension help is bigger, but because of the complexity of the network structure to improve, the detection rate of our proposed method is not high.

4) LOSS FUNCTION

In this part of the loss function, we add the parameter gamma, so that we can control which error range uses MSE (L2 loss, mean square error) and which error range uses MAE (L1 loss, mean absolute error). This is different from the original smooth L1 loss function, and the final average accuracy is improved by 0.38%. The detailed experimental results are as follows: As can be seen from Figure 10 above, when the weight value of the loss function gamma is 1.0/9.0, the detection accuracy is most effective, that is, when the value of gamma is 1.0/9.0, we use L2 loss, and use L1 loss in the rest of the range. The location classification is more accurate, which is more conducive to improve the detection accuracy.

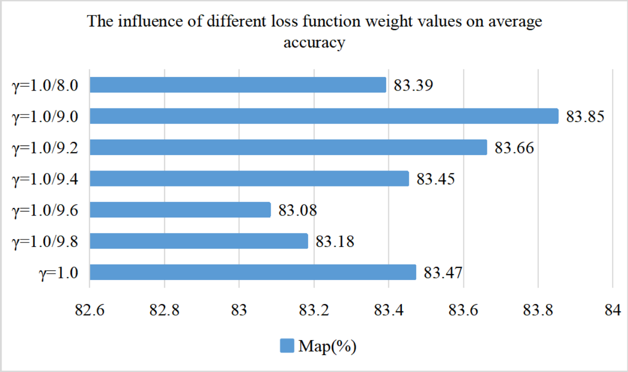


FIGURE 10. The influence of different weight values of loss function on average accuracy.

E. COMPARISON WITH OTHER STATE-OF-THE-ART METHODS

We also compare our proposed method with other state-of-the-art methods on the RPTK1 data set. We use different methods and different sizes for experiments. The results are as follows:

Table 5 above shows the backbone network, input size, mAP and FPS of various methods when the data sets are consistent. The FPS test uses a GPU. When the input size is 300 × 300, the average accuracy of our proposed method is the highest, which is 83.85%, but the detection speed is not significantly improved, which is not suitable for real-time tasks; When the backbone network of DSSD method is Resnet101, the average accuracy of Thangka data set is 83.4%, which is only 0.45% lower than the method proposed in this paper; When Mobilenet is used as the backbone network, the detection accuracy of SSD is only 73.14%. When the data set is the RPTK1 data set and the input size is 512 × 512, SSD uses Resnet50 as the backbone network,

TABLE 5. Experimental results compared with other state-of-the-art methods.

Methods	Backbone	Input Size	mAP	FPS
SSD-300	vgg16	300 × 300	78.79%	17.35
SSD-300	Resnet50	300 × 300	82.62%	18.79
SSD-300	efficientnet_b3	300 × 300	79.85%	18.59
Our proposed method	Resnet50	300 × 300	83.85%	19.62
SSD-300	MobileNetV2	320 × 320	73.14%	28.84
RefineDet	vgg16	320 × 320	75.46%	32.21
DSSD	Resnet101	320 × 320	83.40%	20.91
Faster-rcnn	Resnet50	320 × 320	74.17%	——
YoloV4	CSPDarkNet53	416 × 416	70.87%	13.67
YOLOV4-Tiny	CSPDarkNet53	416 × 416	62.5%	58.53
YOLOV3	darknet53	416 × 416	81.65%	15.63
SSD-512	vgg16	512 × 512	83.51%	16.86
SSD-512	Resnet50	512 × 512	85.04%	16
Efficientdet-d0	VGGNet-16	512 × 512	51.32%	7.77
centernet	Resnet50	512 × 512	29.28%	24.06
retinanet	Resnet50	600 × 600	77.16%	22.26
Our proposed method	Resnet50	512 × 512	84.53%	15.49

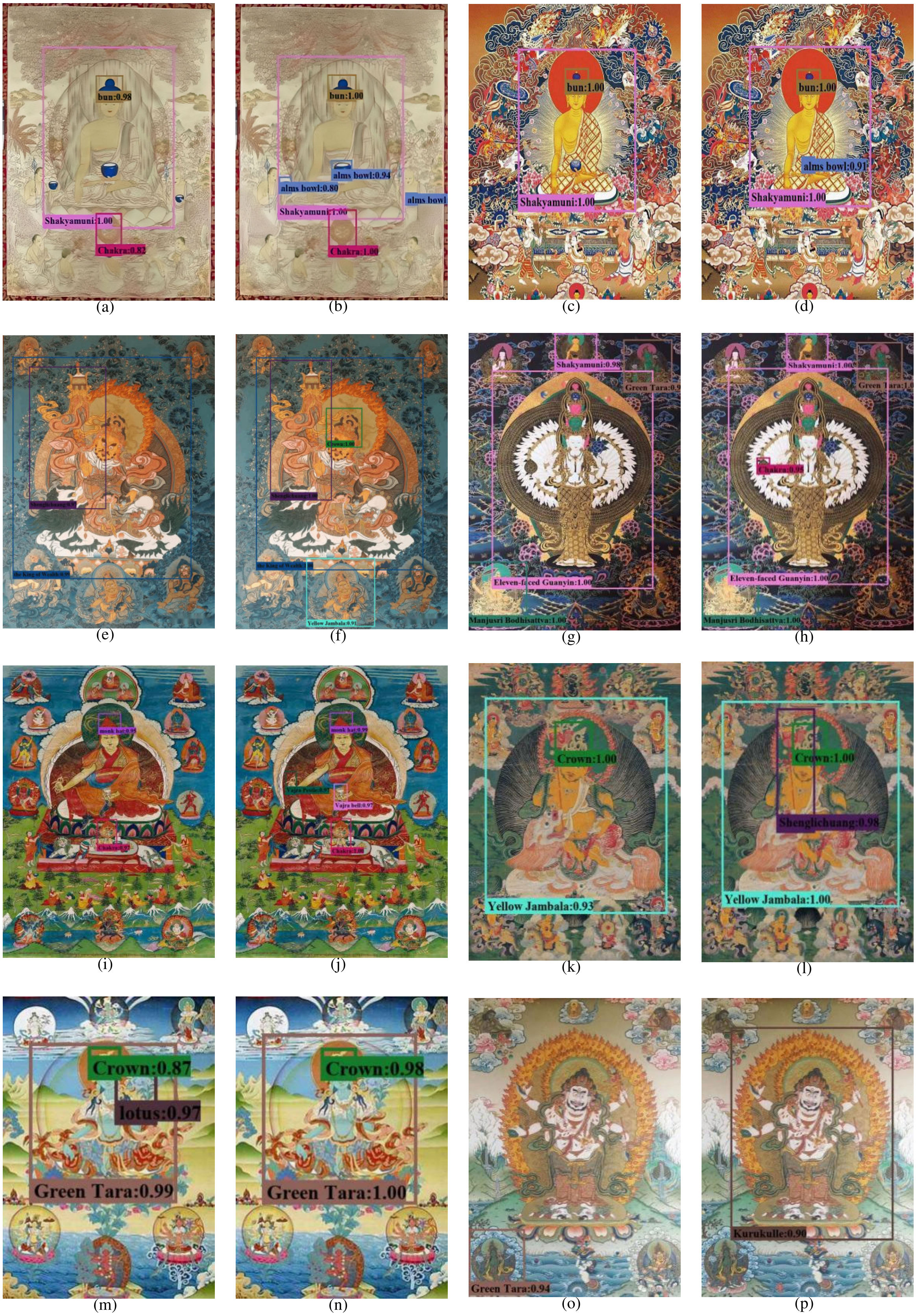


FIGURE 11. Experimental results. (a) SSD-Shakyamuni. (b) Ours-Shakyamuni. (c) SSD-Shakyamuni. (d) Ours-Shakyamuni. (e) SSD-King of Wealth. (f) Ours-King of Wealth. (g) SSD-Eleven faced Guanyin. (h) Ours-Eleven faced Guanyin. (i) SSD-The master. (j) Ours-The master. (k) SSD-Yellow Jambala. (l) Ours-Yellow Jambala. (m) SSD-Green Tara. (n) Ours-Green Tara. (o) SSD-Test results. (p) Ours-Detect errors.

the average detection accuracy is 85.04%, which is 0.51% higher than the detection accuracy of our proposed method, but the detection rate is lower. The method we proposed is more suitable for use in images with an image size of 300 × 300, and it is also not suitable for real-time tasks. It can be seen from Table 5 above that our method performs better than most other methods. Figure 11 below is a comparison of the results of the SSD with Vgg16 as the backbone network and the method we proposed: It can be seen from Figure 11 that our proposed method has a certain effect on small object detection. For example, in Figure 11(a), the religious tools bowl is not detected, but our proposed method detects the religious tools bowl in Figure 11(b). To the small object of bowl, through the comparison between Figure 11(c)-Figure 11(l), SSD with Vgg16 as the backbone network and our proposed method, it can also be seen that our proposed method does not lose the accuracy of normal scale object. Under this premise, the accuracy of small object detection is improved. Figure 11(m)- (p) are examples of the poor detection effect of our proposed method. From the two pictures (n) (p), the proposed method has similar detection objects and background colors. The image detection effect is poor, and there are detection errors. In our analysis, the reason for the poor detection effect is mainly due to the high similarity of the shape of the detected object and the class of the detection error, and the similarity of texture features is also high [53], which results in low detection accuracy and detection errors. Perhaps in the future work, we can start from the texture of the image and constantly improve the accuracy of object detection.

V. CONCLUSION

In this work, we first collect a RPTK1 data set, which contains 3338 Thangka images of different types and contents, with a total of 57 labels of Buddhist professional terms. To further promote our research, we benchmark the Thangka image object detection data set. Finally, when the image size is 300 × 300 and the backbone network is feature fusion Resnet50, the average accuracy of our method reaches 83.85%; under the same conditions, when the image size is 512 × 512, the average accuracy of our method is 84.53 %. We also compare the average accuracy with other state-of-the-art methods, and our model has higher accuracy than other methods. Our experiments show that object detection on Thangka images with rich colors and complex content can also achieve better results. Our experimental results provide an effective way for people to inspect, understand and master Thangka culture, which is conducive to the protection and rescue of Thangka art. In this way, more people can get in touch with and understand the Thangka culture, improve the understanding of Thangka intangible cultural heritage, and accelerate the spread of Thangka intangible cultural heritage. Although our model improves the detection of small objects to a certain extent, the detection results of a small number of Thangka images with lower resolution are still not satisfactory.

In the follow-up research, we will continue this research, mainly to improve the accuracy of object detection on the Thangka data set and improve the detection rate of our method. In the follow-up work, we will continue to improve the feature fusion part of the experiment, because this method requires too much computing resources in the current situation. We hope that our research will encourage more researchers to devote themselves to the investigation of Thangka and even more intangible cultural heritage, and continue to conduct in-depth and comprehensive explorations on the digital protection of intangible cultural heritage.

ACKNOWLEDGMENT

The authors would like to thank Pei Yang, Zhengjun Du, and Sheng Yi for their guidance on experiment and writing.

REFERENCES

[1] Q. Kai, C. Jian, W. Linyuan, Z. Lei, and Y. Bin, "Evaluation results of the PASCAL VOC 2012 test set," in *Proc. Eur. Conf. Comput. Vis.*, vol. 88, no. 2, Jun. 2017, pp. 303–338.
[2] T. Y. Lin, M. Maire, S. Belongie, J. Hays, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.
[3] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, and M. Bernstein, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.
[4] J. Huo, W. Li, Y. Shi, Y. Gao, and H. Yin, "WebCaricature: A benchmark for caricature recognition," *CoRR*, vol. abs/1703.03230, pp. 1–12, Mar. 2017. [Online]. Available: http://arxiv.org/abs/1703.03230
[5] Anonymous, The Danbooru Community, and Gwern Branwen. *Danbooru2020: A Large-Scale Crowdsourced and Tagged Anime Illustration Dataset*. Accessed: Jan. 2021. [Online]. Available: https://www.gwern.net/Danbooru2020
[6] G. Folego, O. Gomes, and A. Rocha, "From impressionism to expressionism: Automatically identifying van Gogh's paintings," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2016, pp. 141–145.
[7] R. Xiao, "Tibetan Thangka, China's intangible cultural heritage," *China Trade Union Accounting*, vol. 2, p. 59, May 2017.
[8] X. Li, W. Wang, and W. Yang, "Improved local accumulate histogram-based Thangka image retrieval," in *Proc. Int. Conf. Image Anal. Signal Process. (IASP)*, 2010, pp. 318–321.
[9] W. Hu, F. Zen, J. Meng, and Y. Ye, "Digital restoration for damaged Thangka image," in *Proc. Int. Conf. Appl. Intell. Syst. Multi-Modal Inf. Anal.* Springer, 2019, pp. 857–865.
[10] Y. Li and X. Liu, "Sketch based Thangka image retrieval," in *Proc. IEEE 5th Adv. Inf. Technol., Electron. Autom. Control Conf. (IAEAC)*, Mar. 2021, pp. 2066–2070.
[11] W. Wang, J. Qian, and L. Yin, "High level semantic retrieval of Thangka image based on C-K relation net," in *Proc. 5th Int. Multi-Conf. Comput. Global Inf. Technol.*, Sep. 2010, pp. 77–81.
[12] F. Zeng, W. Hu, G. He, and C. Yue, "Imbalanced Thangka image classification research based on the ResNet network," *J. Phys., Conf. Ser.*, vol. 1748, Jan. 2021, Art. no. 042054.
[13] H. Liu, X. Wang, X. Bi, X. Wang, and J. Zhao, "A multi-feature SVM classification of Thangka headdress," in *Proc. 8th Int. Symp. Comput. Intell. Design (ISCID)*, Dec. 2015, pp. 160–163.
[14] W. Honghui and L. Xiaojing, "Summary on Thangka image segmentation," in *Proc. Int. Conf. Intell. Comput., Autom. Syst. (ICICAS)*, Dec. 2020, pp. 72–77.
[15] J. Meng, W. Hu, L. Jia, G. He, and P. Xue, "A semantic segmentation model for headdresses in Thangka image based on line drawing augmentation and spatial prior knowledge," *IEEE Sensors J.*, early access, Apr. 30, 2021, doi: 10.1109/JSEN.2021.3076765.
[16] Y. Ma, Y. Liu, Q. Xie, S. Xiong, L. Bai, and A. Hu, "A Tibetan Thangka data set and relative tasks," *Image Vis. Comput.*, vol. 108, Apr. 2021, Art. no. 104125.
[17] D. Tzutalin, "LabelImg," *GitHub Repository*, vol. 6, 2015.
[18] Z. Mecuo, "Inheritance and innovation of traditional crafts of regong Thangka," *Qinghai Technol.*, vol. 28, no. 1, pp. 51–54, 2021.

[19] T. L. An, "A brief talk on the artistic style of contemporary Thangka painting," *Cultural Relics World*, vol. 7, pp. 56–60, 2020.

[20] Gazangnima and Niangjiecairang, "Thoughts on the Dilemma in the inheritance, protection and development of regong Thangka," *Qinghai Technol.*, vol. 27, no. 5, pp. 19–22, 2020.

[21] L. Huaming, B. Xuehui, W. Xiuyou, and W. Weilan, "Automatic classification of Thangka headdresses based on convolutional depth neural networks," in *Proc. Int. Conf. Mach. Learn. Soft Comput.*, Jan. 2017, pp. 105–110.

[22] G. He, P. Xue, and J. Meng, "Few-shot Thangka image classification based on improved DenseNet," *J. Phys., Conf. Ser.*, vol. 1678, Nov. 2020, Art. no. 012087.

[23] K. Rydqvist and R. Guo, "Performance and development of a thin stock market: The Stockholm stock exchange 1912–2017," *Financial Hist. Rev.*, vol. 28, no. 1, pp. 26–44, Apr. 2021.

[24] R. A. Tunio, R. H. Jamali, A. A. Mirani, G. Das, M. A. Laghari, and J. Xiao, "The relationship between corporate social responsibility disclosures and financial performance: A mediating role of employee productivity," *Environ. Sci. Pollut. Res.*, vol. 28, no. 9, pp. 10661–10677, Mar. 2021.

[25] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3213–3223.

[26] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu, "Traffic-sign detection and classification in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2110–2118.

[27] N. Xiang, Z. Cao, Y. Wang, and Q. Jia, "A real-time vehicle traffic light detection algorithm based on modified YOLOv3," in *Proc. IEEE 4th Int. Conf. Electron. Technol. (ICET)*, May 2021, pp. 844–850.

[28] A. Andreopoulos and J. K. Tsotsos, "Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI," *Med. Image Anal.*, vol. 12, no. 3, pp. 335–357, 2008.

[29] R. Rothe, R. Timofte, and L. V. Gool, "DEX: Deep EXpectation of apparent age from a single image," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Dec. 2015, pp. 10–15.

[30] S. Caelles, J. Pont-Tuset, F. Perazzi, A. Montes, K.-K. Maninis, and L. Van Gool, "The 2019 Davis challenge on VOS: Unsupervised multi-object segmentation," 2019, *arXiv:1905.00737*. [Online]. Available: http://arxiv.org/abs/1905.00737

[31] S. Caelles, A. Montes, K.-K. Maninis, Y. Chen, L. Van Gool, F. Perazzi, and J. Pont-Tuset, "The 2018 Davis challenge on video object segmentation," 2018, *arXiv:1803.00557*. [Online]. Available: http://arxiv.org/abs/1803.00557

[32] G. Tong, Y. Li, D. Chen, Q. Sun, W. Cao, and G. Xiang, "CSPC-dataset: New LiDAR point cloud dataset and benchmark for large-scale scene semantic segmentation," *IEEE Access*, vol. 8, pp. 87695–87718, 2020.

[33] S. Wang, Y. Liu, Y. Qing, C. Wang, T. Lan, and R. Yao, "Detection of insulator defects with improved ResNeSt and region proposal network," *IEEE Access*, vol. 8, pp. 184841–184850, 2020.

[34] H. Yang, L. Chen, M. Chen, Z. Ma, F. Deng, M. Li, and X. Li, "Tender tea shoots recognition and positioning for picking robot using improved YOLO-V3 model," *IEEE Access*, vol. 7, pp. 180998–181011, 2019.

[35] L. Duan, K. Yang, and L. Ruan, "Research on automatic recognition of casting defects based on deep learning," *IEEE Access*, vol. 9, pp. 12209–12216, 2021.

[36] Z.-Z. Wu, T. Weise, Y. Wang, and Y. Wang, "Convolutional neural network based weakly supervised learning for aircraft detection from remote sensing image," *IEEE Access*, vol. 8, pp. 158097–158106, 2020.

[37] O. Schumann, M. Hahn, N. Scheiner, F. Weishaupt, J. F. Tilly, J. Dickmann, and C. Wöhler, "RadarScenes: A real-world radar point cloud data set for automotive applications," *CoRR*, vol. abs/2104.02493, pp. 1–8, Apr. 2021. [Online]. Available: https://arxiv.org/abs/2104.02493

[38] L. Maier-Hein *et al.*, "Heidelberg colorectal data set for surgical data science in the sensor operating room," *Sci. Data*, vol. 8, no. 1, p. 101, Dec. 2021.

[39] B. Guo, J. Shi, L. Zhu, and Z. Yu, "High-speed railway clearance intrusion detection with improved SSD network," *Appl. Sci.*, vol. 9, no. 15, p. 2981, Jul. 2019. [Online]. Available: https://www.mdpi.com/2076-3417/9/15/2981

[40] C. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "DSSD: Deconvolutional single shot detector," *CoRR*, vol. abs/1701.06659, pp. 1–11, Jan. 2017. [Online]. Available: http://arxiv.org/abs/1701.06659

[41] Z. Li and F. Zhou, "FSSD: Feature fusion single shot multibox detector," *CoRR*, vol. abs/1712.00960, pp. 1–10, Dec. 2017. [Online]. Available: http://arxiv.org/abs/1712.00960

[42] L. Deng, M. Yang, T. Li, Y. He, and C. Wang, "RFBNet: Deep multimodal networks with residual fusion blocks for RGB-D semantic segmentation," *CoRR*, vol. abs/1907.00135, pp. 1–7, Jun. 2019. [Online]. Available: http://arxiv.org/abs/1907.00135

[43] Q. Zhao, T. Sheng, Y. Wang, Z. Tang, Y. Chen, L. Cai, and H. Ling, "M2Det: A single-shot object detector based on multi-level feature pyramid network," *CoRR*, vol. abs/1811.04533, pp. 1–8, Nov. 2018. [Online]. Available: http://arxiv.org/abs/1811.04533

[44] S. Zhang, L. Wen, Z. Lei, and S. Z. Li, "RefineDet++: Single-shot refinement neural network for object detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 2, pp. 674–687, Feb. 2021.

[45] K. Gesangyixi, *Introduction to Thangka Art*. Beijing, China: Cultural Relics Publishing House, 2015.

[46] Nuobuwangdian, *Religious Portrait Tools in Thangka (Full-Color Illustrated Collection)*, no. 2. Beijing, China: Forbidden City, 2009, pp. 1–279.

[47] H. Liu, X. Wang, X. Bi, X. Wang, and J. Zhao, "A multi-feature SVM classification of Thangka headdress," in *Proc. 8th Int. Symp. Comput. Intell. Des. (ISCID)*, vol. 2, no. 8, 2015, pp. 160–163, doi: 10.1109/ISCID.2015.25.

[48] The Palace Museum, *The Forbidden City Thangka Illustrated Book*. Beijing, China: Forbidden City Press, 2011.

[49] Jiumeijiancuo, *Regong New Year is All About Thangka Art*, China Acad. Art Press, 2016.

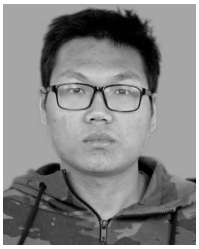
[50] *Tibetan Thangka (Fine)*, Tibet Auton. Region Cultural Relics Manage. Committee, Cultural Relics Publishing House, 2005.

[51] Y. Li, H. Huang, Q. Xie, L. Yao, and Q. Chen, "Research on a surface defect detection algorithm based on MobileNet-SSD," *Appl. Sci.*, vol. 8, no. 9, p. 1678, Sep. 2018. [Online]. Available: https://www.mdpi.com/2076-3417/8/9/1678

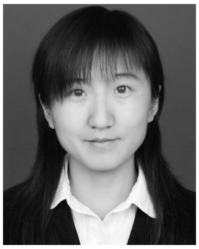
[52] J. M. Cherry *et al.*, "SGD: *Saccharomyces* genome database," *Nucleic Acids Res.*, vol. 26, no. 1, pp. 73–79, 1998.

[53] J. Miao, S. Xu, B. Zou, and Y. Qiao, "ResNet based on feature-inspired gating strategy," *Multimedia Tools Appl.*, vol. 5, pp. 1–18, Mar. 2021.

YUHONG CHEN was born in Qinghai, China. She received the bachelor's degree from Qinghai University, in 2019, where she is currently pursuing the master's degree in computer technology. She received a Computer Professional Qualification Certificate, which is a Software Designer Certificate, in 2018. She published an article.



ZHEN FAN was born in Hubei, China. He received the bachelor's degree from Hubei University of Technology, in 2020. He is currently pursuing the master's degree with Qinghai University. He majored in software engineering and devoted himself to the research of software development technology when he was an undergraduate. In the graduate stage, he turned to the study of mural digitization, and devoted himself to making mural content into animation to record the content of mural permanently.



XIAOJING LIU was born in Anhui, China. She is currently an Associate Professor with the Department of Computer Science, Qinghai University, the Director of the Institute of Information Visualization and Media Computing, the Director of the Chinese Society of Image and Graphics and Qinghai Youth Joint Committee, and the Deputy Secretary-General of Qinghai Computer Society. She received 32 awards, including eight provincial awards, presided over two National Natural Science Foundation projects, two provincial and ministerial projects, and published more than 30 teaching and research articles.