Library Progress International
Vol.44 No.3, Jul-Dec 2024: P.5773-5788

Print version ISSN 0970 1052
Online version ISSN 2320 317X

Original Article

Available online at www.bpasjournals.com

Construction And Experimental Analysis Of House-Tree-Person Image Data Set

Min Wan

Department of Engineering Technology Management, International College, Krirk University, Bangkok 10220, Thailand ,wanmin@yibinu.edu.cn

This work was supported in part by the Yibin University Pre-research Project (2021YY07) and First-class Undergraduate Courses in Sichuan Province (YLKC02524) .

How to cite this article: Min Wan (2024) Construction And Experimental Analysis Of House-Tree-Person Image Data Set. *Library Progress International*, 44(3), 5773-5788.

Abstract

House-tree-person test (HTP) is a projected mental health test based on images. The image data set of the HTP is the basis of intelligent detection and analysis of HTP images. Although HTP is widely used, there are few general data sets. To this end, this study collected images in three ways: Internet download, book scanning, and reality drawing, obtaining thousands of digital images of the HTP. Then, the house, tree and person elements in each image were segmented to form four image data sets of the integral, house, tree and person. Therefore, according to the rules, the elements and features of the house, tree and person are individually calibrated, and the experimental data are divided. Finally, the YOLO target- detection model was used for the experiment, and the detection precision of the four data sets under different quantities was analyzed and compared. The experimental results showed that the average detection precision were equal to 0.828, 0.897, 0.700, and 0.734, respectively. The universality and stability of the HTP image data set were verified under objective indices.

Objective: House-tree-person test (HTP) is a projected mental health test based on images. The image data set of the HTP is the basis of intelligent detection and analysis of HTP images. Although HTP is widely used, there are few general data sets.

Theoretical framework: To this end, this study collected images in three ways: Internet download, book scanning, and reality drawing, obtaining thousands of digital images of the HTP. Then, the house, tree and person elements in each image were segmented to form four image data sets of the integral, house, tree and person.

Method: Therefore, according to the rules, the elements and features of the house, tree and person are individually calibrated, and the experimental data are divided. Finally, the YOLO target- detection model was used for the experiment, and the detection precision of the four data sets under different quantities was analyzed and compared.

Results and Conclusion: The experimental results showed that the average detection precision were equal to 0.828, 0.897, 0.700, and 0.734, respectively. The universality and stability of the HTP image data set were verified under objective indices.

Keywords：House-tree-person Test(HTP), YOLO, Image Data Set, Evaluation Indicators

1 Introduction

In 1948, Buck proposed the house-tree-person test (HTP) by drawing the elements of house, tree and person on three sheets of paper (BUCK,1948). Later, it was improved by Robert Burns to present the three elements of "house, tree and person" on a sheet of paper (Wang et al.,1998; Lee,2019). HTP has the characteristics of initiative, construction and non-verbal to avoid the distortion of the response content in the verbal process. Further, we understand the personality characteristics of the test more specifically and capture indescribable psychological conflict. Furthermore, the test can be repeated even if repeated testing does not result in practical effects, and it is easy to follow up. Thus, HTP has been widely used in practice and extensively studied in theory

(Zubala et al.,2021). The HTP conducts psychological analysis and diagnosis based on the contents, styles and details of the house, tree, person and other paintings drawn by the subjects. It is often combined with other psychological tests, clinical interviews, observations and other information to conduct a comprehensive assessment of the subjects' mental states, personality characteristics, emotions and cognitive functions. It is also an important means of mediating, treating and correcting the relationship between subjects and their surroundings. However, owing to the differences in the expression techniques of factors, the number of detection factors is large, and professionals need to invest a lot of manpower and time in the analysis and diagnosis of HTP (Amini et al.,2013; Sheng t al.,2019; Guo et al.,2022). Therefore, there is an urgent need to introduce artificial intelligence technology to reduce input costs, improve the objectivity, precision and accuracy of results, and deepen research on tracking and observation.

The image data set of HTP is the basis of the training, testing model and algorithm of artificial intelligence detection technology in HTP. Without relevant data sets, experiments cannot be conducted, let alone effective results analysis. Although a large number of published studies have shown that HTP images have been widely used in psychological analysis and diagnosis for different ages, regions, and environments, there are few public or downloadable universal HTP data sets. However, in the HTP, there are no specific instructions or requirements for the limitations of the house, tree, and person depicted by the subjects. The psychological analysis and diagnosis of HTP rely on the integration of various elements such as parts, proportions, structure, and positioning within the image. Therefore, the assessment criteria for HTP involve examination of the whole and its constituent parts as detection factors. These factors served as key indicators of the evaluation process. In fact, according to their own opinions or experiences from the environment, the subjects have different information about the size and type of house, the species and age of the tree, and the gender and behavior of the person (Liu et al.,2020). Therefore, the different performances of the three elements of house, tree and person form a variety of test images. Found two prominent phenomena in a single document or book in HTP (Cho et al.,2021; Chen et al.,2019): first, single and rare images are very scarce and limited for the establishment of data sets; second, the presence of noise and low resolution in the image will cause serious interference and influence artificial intelligence detection technology (Deshpande & Rao,2017; Greenwald et al.,2022). Therefore, it is necessary to collect and process existing HTP images using multiple channels (Zhao et al.,2020). First HTP images can be obtained from the Internet, related books and psychological consultations, and the images can be converted into digital images by downloading and scanning. At the same time, the operation manual of the HTP points out that the three elements and features of the house, tree and person are important analysis factors (Lin et al.,2022). For the confusion problem that is easily concentrated in the same image, the image segmentation principle and method are used to segment the elements to form an independent image data set of elements. In addition, supervised model training for target detection (Zou et al.,2023) is conducive to tracking and analysis. The target detection objects are labeled individually through annotation tools and principles (Saeed et al.,2020), and the annotation file of the HTP is generated. Finally, objective evaluation indices were used to analyze and compare the experimental results of target detection, to verify its effectiveness. Through these processes, an image data set of the HTP with diverse elements and features can be effectively constructed. Therefore, in addition to filling the gap in the current research field, these studies also aimed to explore the image data set.

Universally shared data sets must be represented by diversity, precision and robustness during detection (Wang et al.,2022). The data sets of universal exploration can compare results with each other, exchange experiences and techniques, and promote the application of artificial intelligence technology not only in the field of psychology and neurology (Guo et al.,2021; Schwartz, 2018), but also the practical application and in-depth development in the field of HTP. Therefore, it is particularly urgent and important to construct a high-quality image data set for HTP. First, the quality and quantity of the image data sets directly affect the performance and precision of the model. Whether the data set can provide a diverse and representative sample to determine whether the model can be generalized to a variety of scenes and complex images. Large and diverse data sets require longer training and testing times, often improving model performance, and sometimes overfitting can occur. Therefore, only a properly scaled dataset can enhance the accuracy and reliability of the model, while simultaneously minimizing the risks of overfitting and underfitting.

You Only Look Once (YOLO) is renowned for its rapid and efficient object detection algorithm, which features strong real time capabilities, a singular forward pass, and global information fusion advantages (Wang et al.,2023). In the domain of image detection, this algorithm processes images swiftly and efficiently (Jiang et al.,2022), achieving instantaneous and accurate detection results. Its distinctive characteristic lies in employing a singular forward pass, enabling the algorithm to promptly capture elements and features within the image, thereby mitigating the likelihood of false positives and negatives. Moreover, YOLO exhibits pronounced robustness, ensuring stable operation across diverse environments and further enhancing the reliability of detection. Consequently, the integration of YOLO with HTP image detection contributes to increase precision and real-time capabilities of HTP image detection. Simultaneously, it provides robust support for broader implementation of artificial intelligence technologies in various scenarios.

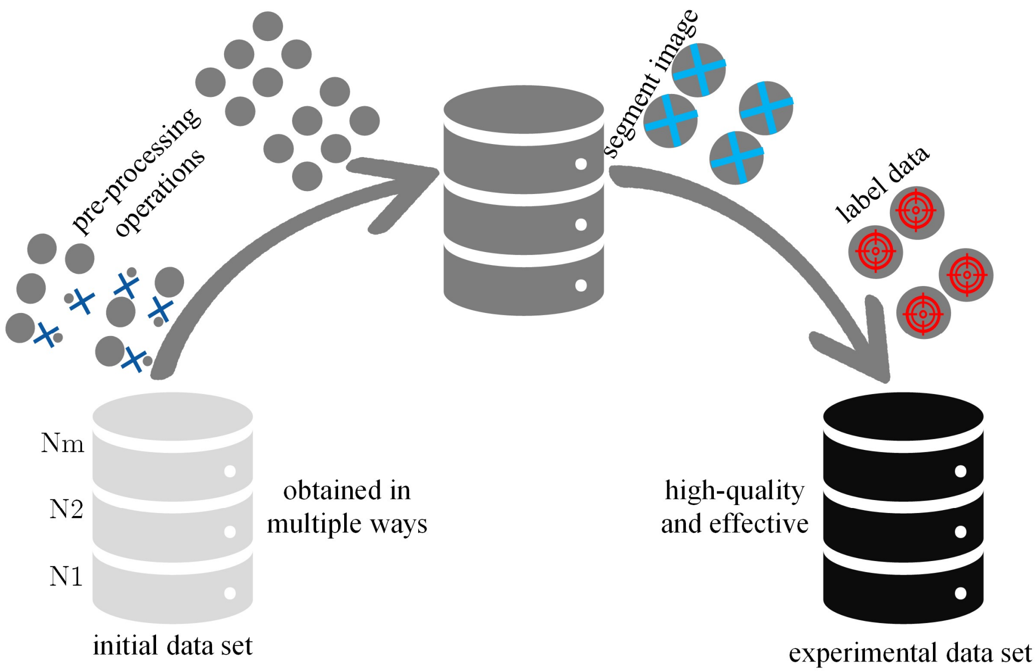
2 Proposed Methods

2.1 Data set construction process

The initial data set of the HTP images was obtained through multiple methods. Following pre-processing operations, including deletion and cleaning, the images within the data set were segmented. Subsequently, the factors and features of house, tree, and person were labeled to establish a high-quality and effective experimental data set of HTP images, as illustrated in Figure 1.

Figure 1

Flow chart of HTP data set construction



2.2 METHOD OF OBTAINING AND COMPOSITION OF DATA SET

To address the absence of a universal HTP data set, the solution involved integration and segmentation. Through the collection of images from various sources, the HTP image data set is formed by combining different numbers of sub-datasets. This method outlines the procedure for acquiring a sub-dataset. Specifically, the overall data set D consists of m sub-datasets, each sub-dataset *N* which comprises number is *j*, and every HTP image is *I*. Here, *D, m, N, j* ∈R. The formula for *N* and *D* are as follows:

$$\mathrm{N} = \{I_1, I_2, I_3, \cdots, I_j\} \tag{1}$$

$$\mathrm{D} = \{\mathrm{N}_1, \mathrm{N}_2, N_3, \cdots, \mathrm{N}_\mathrm{m}\} \tag{2}$$

Through an in-depth examination of current HTP images, the findings are primarily grouped into three categories: existing HTP images related to the author's theme in officially published books and literature, informal and scattered images on the Internet that align with HTP, and images collected by psychological counseling for HTP. After pre-processing operations such as denoising, deletion, and cleansing, three sub-data sets were constructed. Consequently, the total number of images encompassed by the data set used in this study was as follows:

$$\mathrm{D}_{total} = \{N_1, \mathrm{N}_2, \mathrm{N}_3\} = \sum_{m=1}^{3} N_m \tag{3}$$

2.3 Image segmentation in HTP

Image segmentation refers to the technology and process of dividing an image into several specific and unique regions and extracting the object of interest. It is segmented by detecting the edge of different regions. The edge usually represents the dividing line between different regions in the image. Often refers to the location of the change of the gray value of the image, such as the edge of the object, the texture, etc. Based on the coordinate transformation and cropping technology in image processing, the size and position of the rectangular box can be determined through the coordinates of the upper left and lower right corner. Then the required image area can be intercepted. So, the formula is as follows:

$$S = \{(x, y)|x_1 \leq x \leq x_2, y_1 \leq y \leq y_2\} \qquad (4)$$

where (x,y) represents the size of the segment in the image, x_1 and y_1 are the coordinates in the upper left corner of the rectangular box, and x_2 and y_2 represent the coordinates in the lower right corner of the rectangular box. Using this formula, the desired image area can be accurately captured.

In this study, the image data set of the HTP was designed, and the sub-data set of house, tree and person were constructed by image segmentation of houses, trees and persons. There is:

$$D_{HTP} = D_{htp} + D_h + D_t + D_p \qquad (5)$$

Where D_{HTP} is the image data set of HTP, D_{htp}, D_h, D_t, D_p are the segmented image data sets of the integral, house, tree and person, respectively.

2.4 Annotating of elements and features

Data annotation is the process of using AI image detection tools to label text or images in images as specific categories, targets, or attributes (Bouchakwa et al.,2020). There are *c* classes in the image that have *B* labels. Thus, one class is $Class_c$, and the label Number is Q_a. The formula for label number *B* is as follows:

$$\mathrm{B} = \sum_{q=1}^{Q1} \mathrm{Class}_1 + \cdots + \sum_{q=1}^{Qa} \mathrm{Class}_c \qquad (6)$$

In the HTP image, this paper designed to label the three elements and their features of "room", "tree" and "person". Specific areas where the three features of "feature 1", "feature 2" and "feature 3" are located. Therefore, the total number of data annotation labels in this study is:

$$\mathrm{B}_{total} = \sum_{j=1}^{D_{total}} \sum_{i=1}^{3} \sum_{q=1}^{3} Class_{iq}^{j} \qquad (7)$$

The first layer is to loop *j* through all the images; The second layer is a loop *i* traversing the three elements of each image: house, tree and person; The innermost loop *q* traverses the 3 features of each element.

1.5 Evaluation indicators of data set detection

In the target detection model, the evaluation indexes of data set detection mainly include the precision rate P, recall rate R, which expressed by the following formula:

$$P = \frac{TP}{TP+FP} \qquad (8)$$

$$R = \frac{TP}{TP+FN} \qquad (9)$$

where true positives (*TP*) indicate the number of samples correctly detected, false positives (*FP*) indicate the number of samples incorrectly detected, and false negatives (*FN*) indicate the number of samples not detected. Therefore, the average precision (*AP*) and mean average precision (*mAP*).

$$AP = \int_0^1 p(r)dr \qquad (10)$$

Where *p* stands for precision, *r* stands for recall, and *p* is an integral function that takes *r* as an argument.

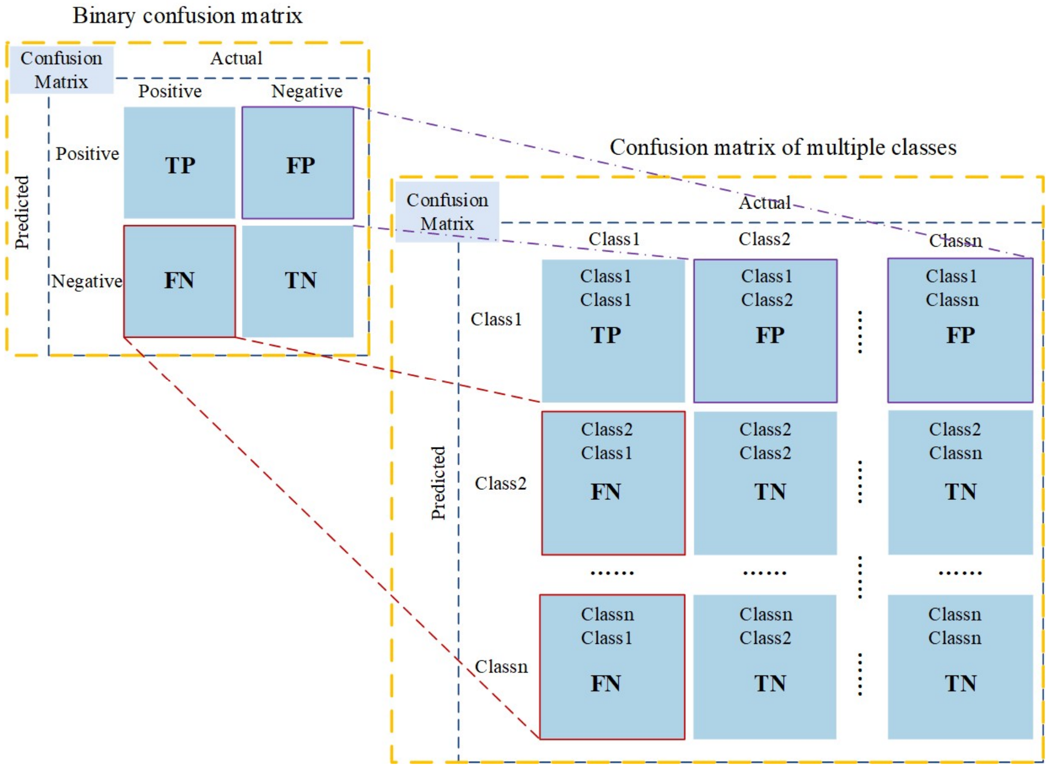
$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \qquad (11)$$

where *AP* reflects the prediction precision of each category *c*, and *mAP* is the average AP of all classes. The higher the *mAP* value, the better the detection effect of the object detection model on a given data set.

Another important metric was the confusion matrix. This shows the performance of the classifier for each category. The prediction classification of the model was compared with the actual classification of the example; thus, the precision or recall of the classification model was analyzed. Shown in Figure 2.

Figure 2

Confusion matrix of bi-classification and multi-classification



True Negatives (*TN*) occur when the true value is negative, and the model correctly predicts a negative outcome. Each row in the matrix represents the predicted class of the instance and each column represents the true class of the instance.

3 Experimental

3.1 Environment of data set experiment

Memory: 32G RAM; graphics card: GTX 1080ti 11G and RTX A5000 24G,Win11 64-bit; software: Python 3.9.7, Torch 1.10.1+cu113,Photoshop; labelImg for label calibration, YOLO v7.

3.2 Obtained Method and Pre-processing

Table 1

Overall image data set acquisition method of HTP

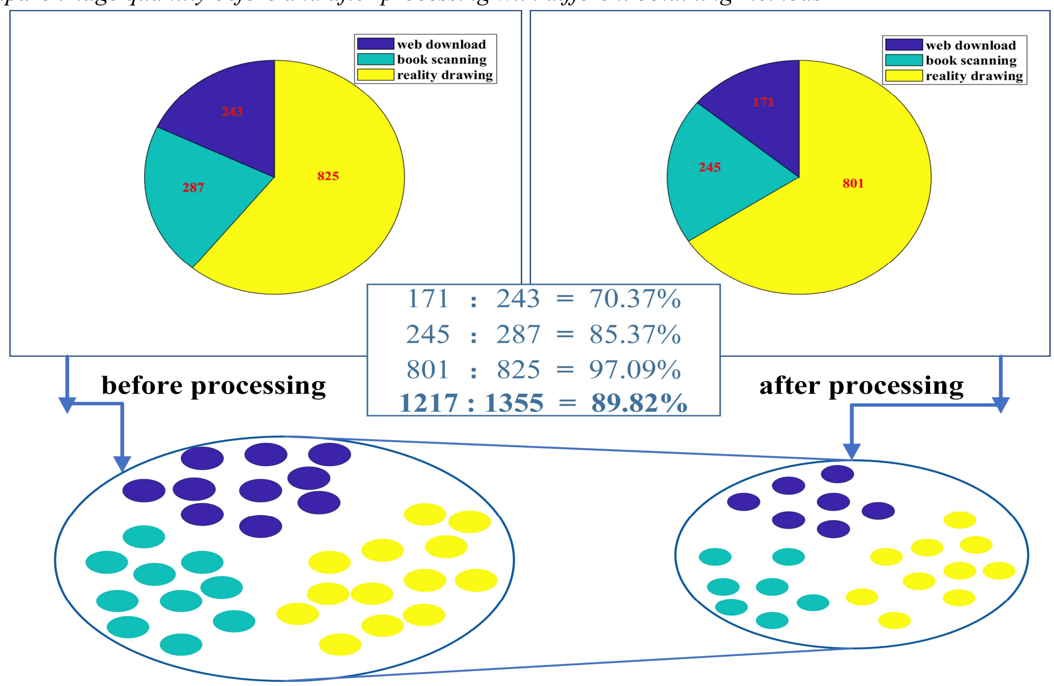
Acquisition mode	Specific information
Web download	Browser: Baidu; Search keywords: house-tree-person or house-tree-person drawing.
Book scanning	a. Pictorial psychological analysis of HTP in the field of management; b. Application manual of drawing psychology for middle school students; c. One Drawing, One World - Perspective of college students' painting psychological counseling cases; d. Theory and practice of painting psychoanalysis and therapy; e. Psychological drawing of college students.
Reality drawing	Organizer: Consultant, assistant; Instruction: Recording MP4, text description and on-site instruction; Collection: 825 people from 11 groups.

From Table 1, the Internet download is searched in the *Baidu* image library with "HTP or HTP drawings" as the keyword. The book scan covers five books related to the HTP. In reality painting, 825 participants from 11 groups were instructed to draw on the theme of the HTP in various ways.

Convert all collected images into digital images in *.jpg* or *.png* format. Subsequently, if the test does not conform to the HTP or the resolution is less than 10KB, it will be eliminated, duplicates will be deleted, and the work of auditing, naming, sorting and review will be carried out to constitute the experimental data set, as shown in Figure 3.

Figure 3

Compare image quantity before and after processing with different obtaining methods



Number of images obtained by different methods and the specific number of controls before and after processing. The experimental data sets after pre-treatment were all smaller than the initial data sets before processing. The proportion of control between the experimental data sets with different obtained methods and the initial data sets is significantly different. For example, the proportion of data downloaded from the Internet was the smallest, only 70.37%; The data of real scene drawing accounted for the largest proportion, as high as 97.09%, and the gap was 26.72%. The retention ratio for the three acquisition methods was 89.82%. Results analysis: Many irrelevant and low-resolution images appeared on the Internet download mode of keyword search, while a high repetition rate appeared in the book scan mode, and the participating groups performed well in understanding the guide language, showing high effectiveness.

3.3 Image segmentation of HTP

The experimental data set is the whole image of the HTP, and it is necessary to train the detailed features of the elements in the feature detection of the house, tree and person. Therefore, in the overall image, according to the categories of house, tree and person, Photoshop was used to divide the image into house, tree and person. Specific data are presented in Table 2.

Table 2

Segmentation of HTP images

Serial number	Data set	Image quantity (page)	Main purpose
1	Integral graph	1217	Focus on the elements of house-tree-person in the overall image
2	house graph	1182	Focus on the feature details of the house in the image
3	Tree graph	1162	Focus on the feature details of the trees in the image
4	person graph	889	Focus on the details of the features of the people in the image

From Table 2, in the data set, the maximum number of integral was 1217, and the minimum number of person was 889. In the person figure, owing to some mismatches, unclear and complete situations, 309 matches and cartoons were eliminated according to the data requirements of the target detection model. The segmentation process is illustrated in Figure 4.

Figure 4

Image data set of HTP from left to right graphs: integral, house , tree and person

The four data sets in Table 2 are a comprehensive sample image data set of the HTP, which covers images of different themes and scenes and provides the basis for intelligent detection experiments.

3.4 Annotation of HTP images

Image annotation is to identify the target in the image, and label the location and category of the target in the area. In this paper, *LabelImg* software was used to select the house, tree and person elements in all sample data sets and the three features contained in them with rectangular envelope boxes, and the category sequence numbers were 0, 1 and 2. And the above location envelope information and category serial number and category information are written into the label text in turn. This process builds a training data set that conforms to the appropriate format for YOLO target detection. The quantity, serial number, category name and other information specified are shown in Table 3.

Table 3

Annotated information for each HTP image data set

Data set	Serial number 0	Serial number 1	Serial number 2	Label serial number and name
Integral	1283	1475	1612	0: house, 1: tree, 2: person
house	1252	1051	2032	0: roof, 1: window, 2: door
tree	2812	2666	516	0: crown, 1: trunk, 2: root
person	751	882	858	0: facial, 1: hand, 2: limb

From Table 3, in the entire graph data set, serial numbers 0, 1 and 2 represent house, tree and person respectively, with a total of 4370 annotations. The number of labels was 1283 for house, 1475 for tree and 1612 for person. In the house data set, serial numbers 0, 1 and 2 represent the roof, window and door, respectively, with 4335 annotations. The number of labels is 1252 for the roofs, 1051 for windows and 2032 for doors. In the tree data set, serial numbers 0, 1 and 2 correspond to the crown, trunk and root, respectively, with 5994 annotations. Crowns had the highest number of annotations 2812, trunks had 2666, and roots had 516. In the person data set, serial numbers 0, 1 and 2 represent facial, hand and limb, respectively, with a total of 2491 annotations. The number of facial labels was 751, hands had 882 and limbs had 858. The number of labels ranged from 516 to 2812, once again verifying the rich diversity of house-tree-person elements and their features.

3.5 Division of experimental data of data sets

The HTP image data set and four sub-data sets were divided into training, verification and test sets at a ratio of 8:1:1. The integral, the house, the tree and the person were tested by gradually increasing the amount of image data. 200, 500, 700 and 100% were selected successively for different numbers of experiments.

Experiments were conducted with the same value for each data set, ensuring that each data set has sufficient data for training, verification and testing in the experiment, which helps to ensure the fairness and comparability of the experiment. At the same time, the experiment of selecting all data in factor data sets with different amounts of data is helpful to verify the generalization ability of each data set with diverse analysis.

4 Result

In this section, the experimental verification, results analysis and comparison are carried out according to the experimental preparation of the HTP data set in Section 2.

4.1 Longitudinal analysis and comparison of data sets

HTP, HTP-h, HTP-t and HTP-p are used to represent four different categories of integral, house, tree and person, respectively, and HTP-class-quantity represents different selection quantities of images of different categories.

4.1.1 Analysis and comparison of 4 experiments of integral data set

From Table 2 and 3, 200, 500, 700 and 1217 images and their annotated information in the integral data set were selected for the experiments and the respective precision results of different value scales were obtained.

Table 4

Analysis and comparison of 4 experiments of the integral data set

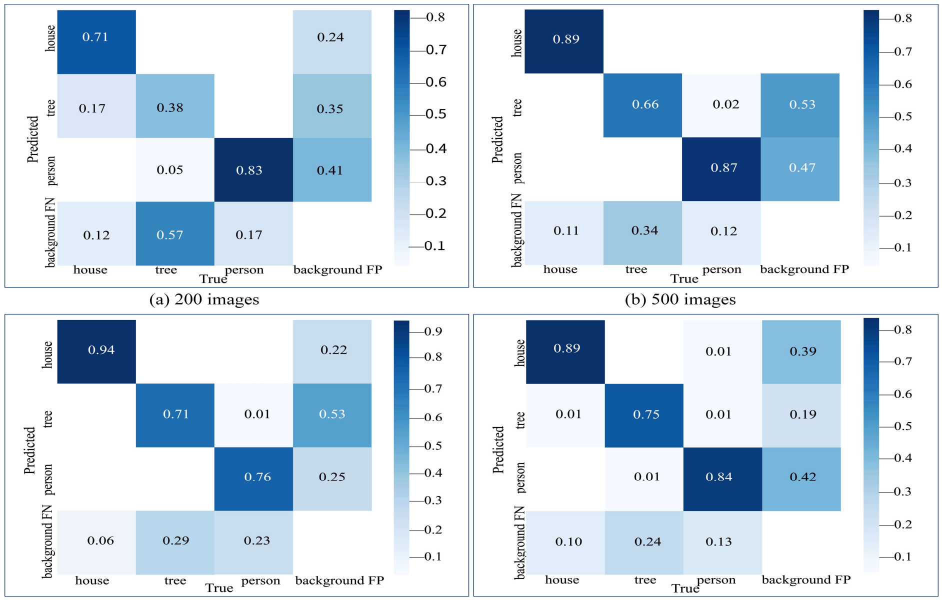
Serial number	House（mAP)	tree（mAP)	person（mAP)	Triad average（mAP)
HTP-200	0.855	0.618	0.874	0.782
HTP-500	0.975	0.667	0.849	0.830
HTP-700	0.958	0.658	0.849	0.822
HTP-1217	0.883	0.861	0.883	0.876
Four average	0.918	0.701	0.864	0.828

From Table 4:(a) Impact of data size and average precision: When the data is increased from 200 to 500, the average precision of the model is increased from 0.782 to 0.830. This indicates that when the data volume is small, increasing the data volume can significantly improve the generalization ability of the model. However, when the data were increased from 500 to 700, the average precision decreased slightly, from 0.830 to 0.822. This is because the new data introduce more noise or diversity, making it difficult for the model to learn a more accurate classification boundary. When the data were further increased to 1217, the average precision of the model was significantly improved, reaching 0.876. This shows that when the amount of data is sufficiently large, the model can learn more accurate and comprehensive feature representation from more samples, thus improving the classification precision. (b) Precision changes of different categories: "For house" and "tree" categories, the precision fluctuates with the increase of data volume, however the overall trend is rising. This is because the model can learn more about the features and patterns of the house using a larger amount of data. For the "person" category, the precision remains basically stable with the increase in data volume. This means that, in a small amount of data, the model is already able to identify the characteristics of a person.

The experiment also obtained confusion matrix of 200,500,700 and 1217, as shown in Figure 5.

Figure 5

Integral experiment of confusion matrix



(a) 200 images (b) 500 images

(c) 700 images (d) 1217 images

4.1.2 Analysis and comparison of 4 experiments of house dataset

From Table 2 and 3, 200, 500, 700 and 1182 images and their annotated information in the house data set were selected for the experiments, and the respective precision results of different value scales were obtained.

Table 5

Analysis and comparison of four experiments in house data set

Serial number	roof（mAP）	door（mAP）	windows（mAP）	Triad average（mAP）
HTP-h-200	0.741	0.860	0.860	0.821
HTP-h-500	0.937	0.938	0.955	0.943
HTP-h-700	0.933	0.926	0.931	0.930
HTP-h-1182	0.932	0.908	0.843	0.894
Four average	0.886	0.908	0.897	0.897

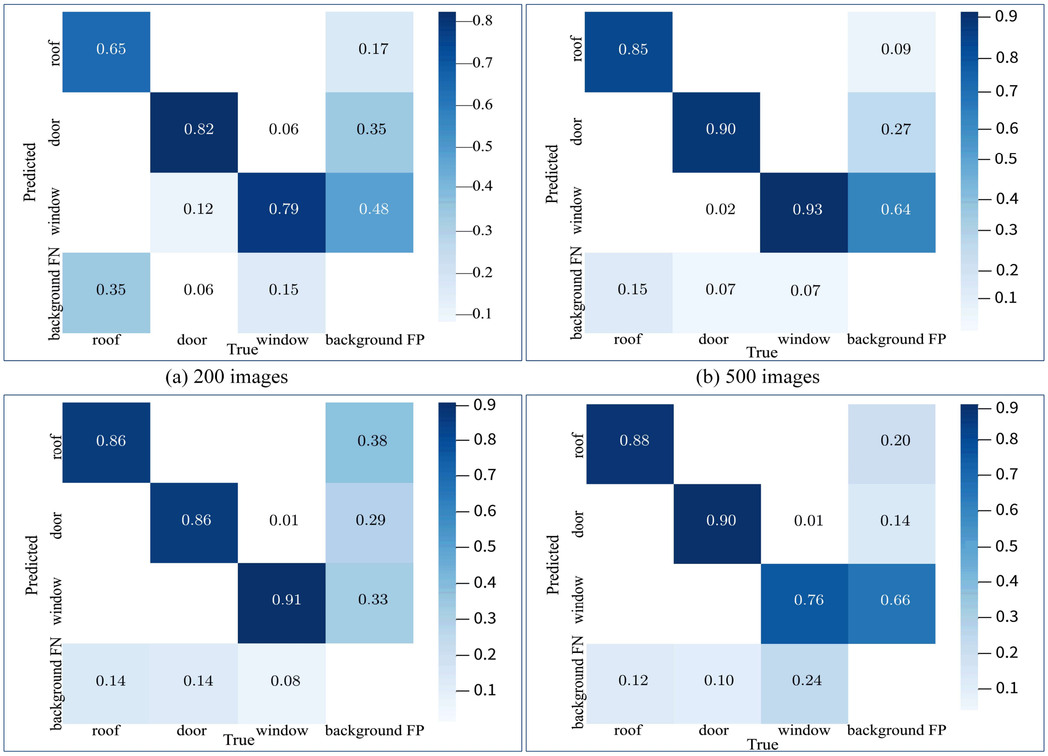
From Table 5: (a) Influence of data size and average precision: When the amount of data increased from 200 to 500, the average precision of the model increased from 0.821 to 0.943. This indicates that when the data volume is small, increasing the data volume can significantly improve the generalization ability of the model. However, when the data were increased from 500 to 700, the average precision decreased slightly, from 0.943 to 0.930. This is because the new data introduce more noise or diversity, making it difficult for the model to learn a more accurate classification boundary. When the data were further increased to 1182, the average precision of the model was significantly improved, reaching 0.894. This shows that when the amount of data is sufficiently large, the model can learn more accurate and comprehensive feature representations from more samples, thus improving the classification precision. (b) Precision changes of different categories: For the "roof" category, the precision fluctuates with the increase in data volume; however the overall trend is increasing. This is because the model can learn more about the characteristics and patterns of the roof using a larger amount of data. For the "door" category, there was a significant improvement in precision as the amount of data increased. In particular, for data volume of 1182 images, the precision reached 0.908. This shows that increasing the amount of data is very helpful for improving the recognition of the model on the "door" category. For the "windows" category, precision first increases and then decreases as the amount of data increases. This is because as the data grows, some specific types of window samples dominate, leading to a decline in the identification precision of other types of Windows.

Concurrently, the experiment also obtained confusion matrix of 200,500,700 and 1182, as shown in

Figure 6.

Figure 6

House experiment of confusion matrix



(a) 200 images (b) 500 images

(c) 700 images (d) 1182 images

4.1.3 Analysis and comparison of 4 experiments on tree data set

From Table 2 and 3, 200, 500, 700 and 1162 images and their annotated information in the tree data set were selected for the experiments, and the respective precision results of different value scales were obtained.

Table 6

Analysis and comparison of 4 experiments in tree data set

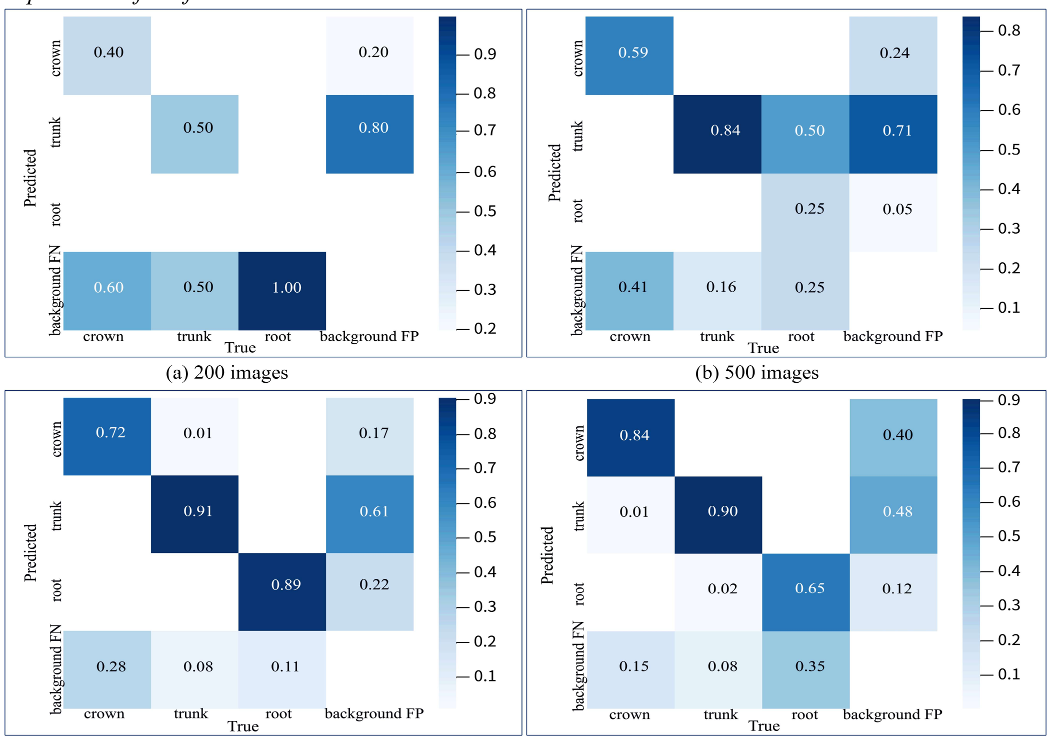
Serial number	crown（mAP）	trunk（mAP）	root（mAP）	Triad average（mAP）
HTP-t-200	0.717	0.502	0.000	0.406
HTP-t-500	0.747	0.754	0.504	0.669
HTP-t-700	0.843	0.866	0.807	0.839
HTP-t-1162	0.929	0.949	0.780	0.886
Four average	0.809	0.768	0.523	0.700

From Table 6:(a) Influence of data volume on average precision: With the increase in data volume from 200 to 1162, the average precision of the model steadily improves, and the average precision significantly increases from 0.406 to 0.886. This trend shows that the performance of the model continued to improve as more training samples were provided. This is because a larger amount of data contains more features and variations, which helps the model to learn and generalize better. (b) The variation of precision of various categories: For "crown" and "trunk" categories, the precision continues to improve with the increase of data volume, the former from 0.717 to 0.929, and the latter from 0.502 to 0.949. This shows that the model can capture more features and patterns about "crown" and "trunk" in a larger amount of data. For the "root" category, although the precision increased from 0.000 to 0.780, it is a smaller improvement compared to the other two categories. This is because the "root" category has greater diversity or complexity in some datasets, making it difficult for the model to identify precisely.

Concurrently, the experiment also obtained confusion matrix of 200,500,700 and 1162, as shown in Figure 7.

Figure 7

Tree experiment of confusion matrix



(a) 200 images (b) 500 images

(c) 700 images (d) 1162 images

4.1.4 Analysis and comparison of 4 experiments on person data set

From Table 2 and 3, 200, 500, 700 and 889 images and their annotated information in the person data set were selected for the experiments, and the respective precision results of different value scales were obtained.

Table 7

Analysis and comparison of 4 experiments in the person data set

Serial number	facial（mAP）	hand（mAP）	limb（mAP）	Triad average（mAP）
HTP-p-200	0.893	0.750	0.429	0.691
HTP-p-500	0.932	0.798	0.673	0.801
HTP-p-700	0.899	0.706	0.609	0.738
HTP-p-889	0.845	0.825	0.451	0.707
Four average	0.892	0.770	0.541	0.734

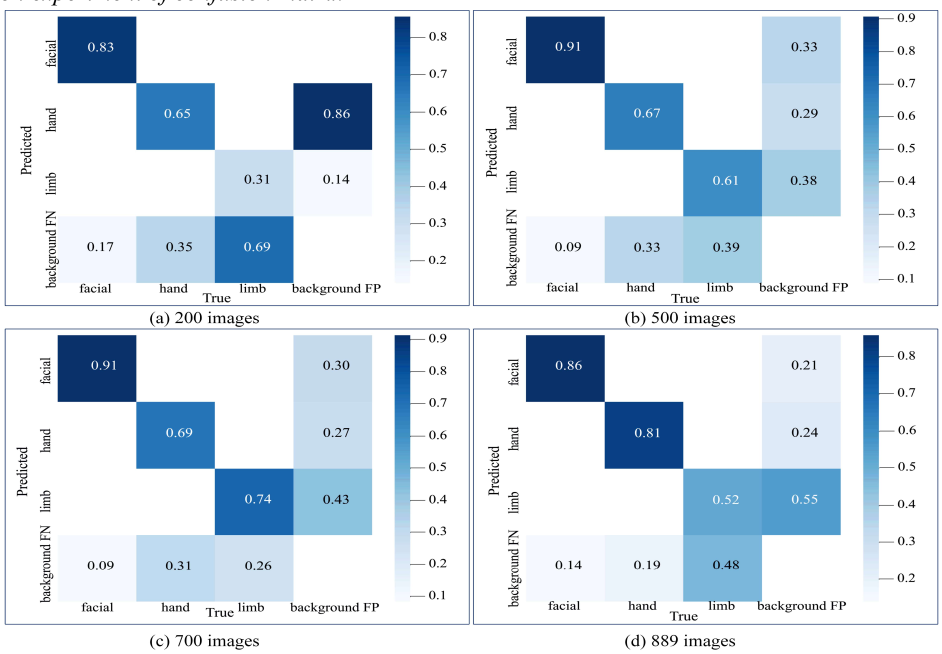
From Table 7: (a) Impact of data volume on model precision: With the increase in data volume from 200 to 889, the average precision was slightly improved, and the average precision of the model increased from 0.691 to 0.707. Although the improvement is small, it still shows that a larger amount of data helps the model to learn and generalize better. This indicates that the performance of the model improved slightly when more training samples were provided. (b) Variation in precision of various categories: For the "facial" category, the precision was higher at 500 sheets, which was 0.932. However, when the data volume was increased to 700 and 889 images, the precision decreased slightly. This is because larger data introduces more noise or diversity, leading to some fluctuations in the model's recognition of the "facial" category. For the "hand" category, the precision was highest at 0.825 when the amount of data was increased to 889 images. This is because larger datasets provide more information about the features and patterns of the "hand", which helps the model to learn and recognize better.

For the "limb" category, the precision was highest at 500 and 0.673, and also decreased and fluctuated when the data volume increased to 700 and 889. This is again because the larger amount of data provides more information about the features and patterns of the limb, making it difficult for the model to identify precisely.

The experiment also obtained confusion matrix of 200,500,700 and 889, as shown in Figure 8.

Figure 8

Person experiment of confusion matrix



(a) 200 images (b) 500 images

(c) 700 images (d) 889 images

4.2 Horizontal analysis and comparison of data sets

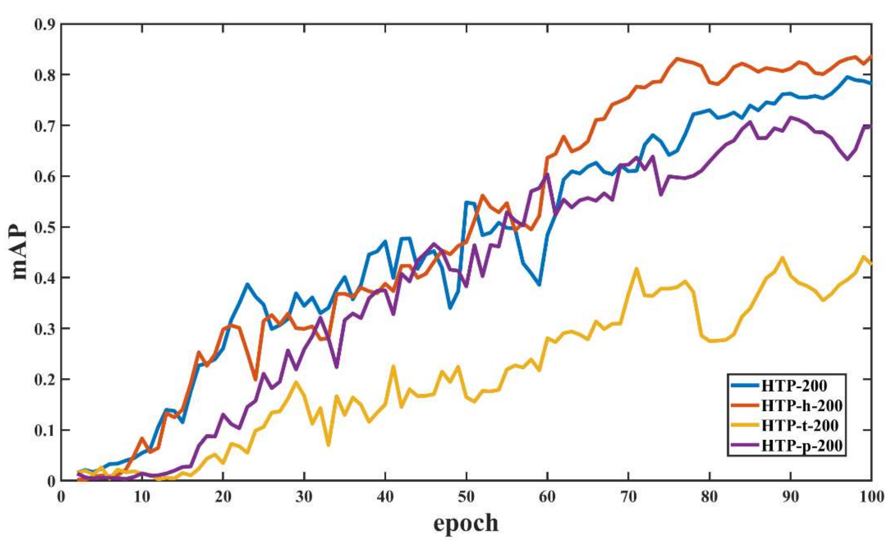
In the four data sets of integral, house, tree and person, different numbers of images were extracted for the experiment. The results in Table 4, 5, 6 and 7 are then compared and analyzed.

4.2.1 Analysis and comparison of 200 images

With the same amount, as the number of model training epochs increases, the precision changes as shown in Figure 9. The results of 200 in Table 4, 5, 6 and 7 were compared and analyzed.

Figure 9

Comparison of detection process in 200 images on mAP



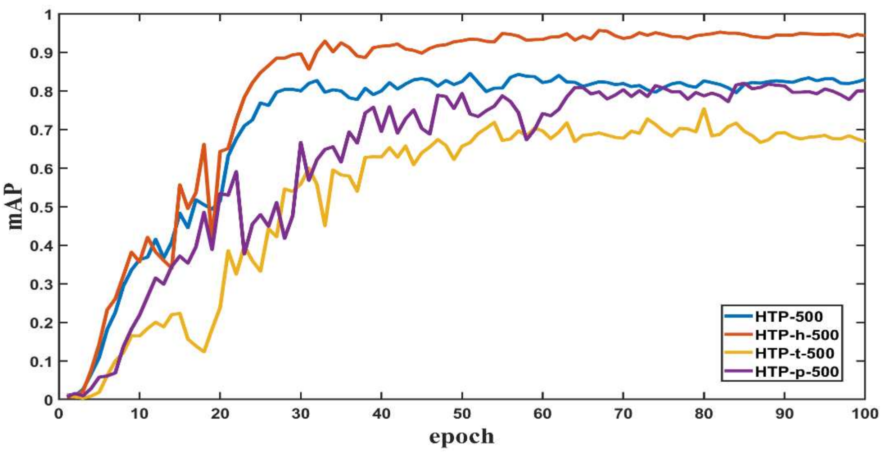
The analysis is as follows: (a) The fluctuation of the four curves is large, and the overall precision value is not high. The precision of the HTP-t-200 curve was lower than that of the other three curves. At epoch 100, HTP-h-200 was the highest (0.821). (b) The results show that HTP-h-200 has obvious features and is easier to extract and identify. Feature highlighting helps the model to learn and generalize better, thus improving the precision of classification and detection. (c) For the category of "detection 1", the precision is the highest in the HTP-p-200 data, which is 0.893, while in the HTP-h-200 and HTP-t-200 data, the precision is less than 0.75. For the "detection 2" category, the precision was the highest in the HTP-h-200 data set at 0.860. For HTP-t-200, the precision was the lowest at 0.502. For the "detection 3" category, the precision was higher for both the HP-200 and HP-h-200 data (0.874 and 0.860, respectively). For the HTP-t-200, the precision was 0. The low detection precision is due to the small number of complex features in these samples, which makes it difficult to identify the model accurately.

4.2.2 Analysis and comparison of 500 images

With the same amount, as the number of model training epochs increases, the precision changes as shown in Figure 10. The results of 500 in Table 4, 5, 6 and 7 were compared and analyzed.

Figure 10

Comparison of detection process in 500 images on mAP

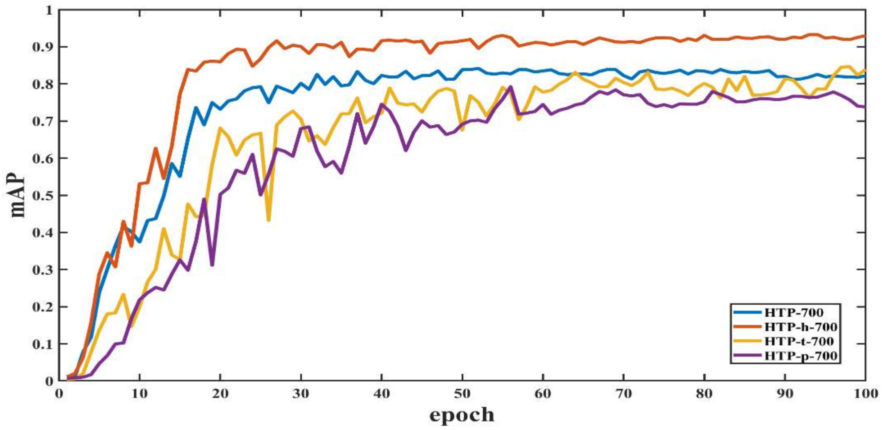


The analysis is as follows: (a) The four curves fluctuate greatly before 40 epochs, and then are relatively stable. The precision of the HTP-t-500 curve was lower than that of the other three curves. HTP-h-500 was obviously the highest at 0.943. (b) The average precision of the model was higher than 200, indicating that the model could handle these data better than 200. This is because in these number images, the features are significant, which helps the model learn and generalize better, thus improving the precision of classification and detection. (c) For the "detection1", the precision of HTP-500, HTP-h-500 and HTP-p-500 data was higher than 0.9, HTP-500 was the highest (0.975), and HTP-t-500 is the lowest (0.747). For the "detection2", the precision is highest in the HTP-h-500 data set at 0.938, much higher than the other three. For the "detection3", the precision of HTP-h-500 is the highest (0.955), and that of HTP-t-500 is the lowest (0.504), with a gap of 0.451. The results show that complex features make the model difficult to identify accurately, thus the precision is relatively low.

4.2.3 Analysis and comparison of 700 images

With the same amount, as the number of model training epochs increases, the precision changes as shown in Figure 11. The results of 700 in Table 4, 5, 6 and 7 were compared and analyzed.

Figure 11
Comparison of detection process in 700 images on mAP

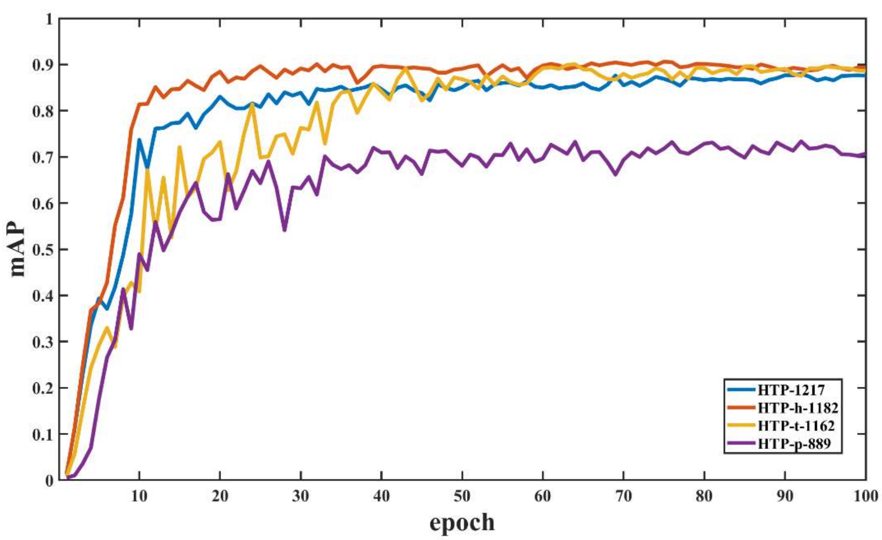


The analysis is as follows: (a) The four curves fluctuate greatly before 25 epoch, and in the later epoch only HTP-p-700 has a larger fluctuation, and the others relatively stable. The precision of the HTP-p-700 curve was lower than that of the other three curves. HTP-h-700 was the highest, with an average precision of 0.930. (b) This is because in this quantity, each feature can be better extracted and identified. For detection points with significant features, this helps the model to learn and generalize better, to improve the precision of classification and detection. (c) For the category of "detection1", the precision is the highest in the HTP-700 data set, which is 0.958. For the "detection 2", the precision is highest in the HTP-h-700 data set at 0.926. For the "detection 3", the precision is higher in HTP-h-700 at 0.931. In other data sets, low precision of different detection points is owing to the small number of samples or complex features of such data, which makes it difficult to precisely identify the model.

4.2.4 Analysis and comparison of all images

With the same amount, as the number of model training epochs increases, the precision changes as shown in Figure 12. The results in Table 4, 5, 6 and 7 were compared and analyzed.

Figure 12
Comparison of detection process in all images on mAP



The analysis is as follows: (a) The four curves fluctuate significantly before 30 epoch, and other relatively stable. Among them, the precision of curves HTP-1217, HTP-h-1182 and HTP-t-1162, their variation or size are almost the same after 40 epoch. However, that of HTP-p-889 was lower than that of the other three curves. (b)

For the four data sets, the data set size ranged from 889 to 1217. Despite the differences in data set sizes, the average precision of the models is relatively high, which indicates that the models perform well when dealing with these images. (c) For category "detection1", the precision is the highest in the HTP-h-1182 data set, which is 0.932. For the "detection2", the precision is highest in the HTP-t-1162 data set at 0.949. For the "detection3", the precision is highest in HTP-1217 at 0.883. The results show that the model is easier to learn and recognize when there are more samples in full image.

5 Discussion

The intelligent detection experiment images of HTP are mainly obtained in three ways: Internet download, book scanning and real scene drawing. The image data set of the HTP was constructed by pre-processing, segmentation and labeling the initial image. From the data sets of integral, house, tree and person, 200, 500, 700 and all number images were selected to carry out the detection experiment of different image numbers, and analyzed and compared from the vertical and horizontal experimental results. Bidirectional results analysis and comparison show that a small amount of data is prone to overfitting or insufficient generalization ability, whereas a large amount of data introduces noise and diversity, which affects the performance. Only a suitable data size can balance the learning and generalization abilities of the model. Therefore, different tasks and categories require different optimization strategies, analysis of data set distribution and features, and selection of appropriate sizes to achieve the best results. Through data analysis of the classification and quantity experimental results, the YOLO target detection model in artificial intelligence technology is used to verify the effectiveness of the constructed HTP image data set, which has practical application value and significance, and the average detection precision can be equal to or even higher than 0.801 in different scenarios of house, tree and person elements and their features.

6 Conclusion

For the first time, this paper constructs the image data set of HTP by collecting, sorting and processing the images of HTP, uses the YOLO model with fast detection speed and high precision conducts sufficient experiments and detailed analysis of the data set, and further explores the fitting ability of the HTP image data set in the detection model experiments. In this paper, we focus on the completion of the work: First, the existing HTP is introduced, and the methods of acquiring HTP images are systematically organized. The HTP images obtained from various sources were then compiled into an initial data set. second, the original data set underwent pre-processing, involving deletion, naming, cleaning, and segmentation. Four distinct data sets were then established: integral, house, tree, and person subsets. The elements of the house, tree, and person, along with their respective features, were meticulously annotated. This process contributes to the creation of experimental data sets characterized by enhanced realism in classification, increased annotations, and improved quality for the intelligent detection of HTP images. Third, diverse samples of varying data sizes were chosen using an incremental method for the constructed HTP image data set. The YOLO target detection model was employed for experimental validation, and a comprehensive analysis and comparison were conducted to assess the robustness and generalization of the constructed experimental data set in detecting elements and features within the HTP. Next, based on the HTP image data set constructed in this study, we combined other advanced algorithms to improve the precision of model detection and its potential in practical application. Simultaneously, we should fully explore more feature detection in HTP to promote the in-depth application of artificial intelligence technology in psychological analysis and diagnosis.

Reference

Amini D., Afrooz G. A., Daramadi P. Sharifi, Homan H. A. (2013) "Recognition of disorders and emotional problems of deaf children using house-tree-person and draw-a-person tests in comparison with normal children of Hamadan Province", Avicenna Journal of Clinical Medicine, 20(1), pp.49-58.

Bouchakwa M., Ayadi Y., Amous I. (2020)"A review on visual content-based and users' tags-based image annotation: methods and techniques", Multimedia Tools and applications, 79, pp.21679-21741.

BUCK JN (1948)"The H-T-P technique:a qualitative and quantitative scoring manual",Journal of Clinical Psychology, 4,pp.151-159.

Chen T., Wang P. G., Lin X., Pei H. C., Xing Y. L., Luo J., Wang Y. (2019) "Application of machine learning to predict adolescent dependent personality deviation in drawing test", Chinese Mental Health Journal, 33,10.

Cho S. J., , Cho G. W., Kim Y. W. (2021) "Development of Deep Learning-Based House-Tree-Person Test

Analysis Model", In Proceedings of the Korea Information Processing Society Conference, 558-561.

Deshpande M., Rao V. (2017) "Depression detection using emotion artificial intelligence", In 2017 international conference on intelligent sustainable systems (iciss), 12,pp.858-862.

Jiang P., Ergu D., Liu F., Cai Y. B. (2022)"A Review of Yolo algorithm developments", Procedia Computer Science, 199, pp.1066-1073.

Greenwald N. F., Miller G., Moen E., Kong A., Kagel A., Dougherty T., Valen D. V. (2022) "Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning", Nature biotechnology, 40(4), pp.555-565.

Guo Q., Yu G., Wang J., Qin Y., Zhang L. (2022) "Characteristics of House-Tree-Person Drawing Test in Junior High School Students with Depressive Symptoms", Clinical Child Psychology and Psychiatry, 13591045221129706.

Guo W., Yang H., Liu Z., Xu Y., Hu B. (2021) "Deep Neural Networks for Depression Recognition Based on 2D and 3D Facial Expressions Under Emotional Stimulus Tasks", Frontiers in Neuroscience, 15.

Lee E. J. (2019)"Correlations among depressive symptoms, personality, and Synthetic House-Tree-Person Drawings in South Korean adults", Psychologia. 61(4), pp.211-220.

Liu C., Chen H., Liu C. Y., Lin R. T., Chiou W. K. (2020) "Cooperative and individual mandala drawing have different effects on mindfulness, spirituality, and subjective well-being", Frontiers in Psychology, 11, 564430.

Lin Y., Zhang N., Qu Y., Li Y., Liu J., Song J. (2022) "The House-Tree-Person test is not valid for the prediction of mental health: An empirical study using deep neural networks", Acta Psychologica, 230, 103734.

Saeed S. M. U., Anwar S. M., Khalid H., Majid M., Bagci U. (2020) "EEG based classification of long-term stress using psychological labeling", Sensors, 20(7): 1886.

Schwartz K. M. (2018)"Making unformulated experience real through painting: Painting and psychoanalytic psychotherapy practice as two ways of making sense", Journal of Clinical Psychology, 74(2), pp.239-248.

Sheng L., Yang G., Pan Q., Xia C., Zhao L. (2019) "Synthetic house-tree-person drawing test: a new method for screening anxiety in cancer patients", Journal of Oncology, 2019.

Wang C. Y., Bochkovskiy A., Liao H. Y. M. (2023)"YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors", In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,pp.7464-7475.

Wang J., Wang J., Zhang H., Guo C., Wang Y., Lu B., Feng C., Pan F., Li T. (2022) "Construction and application of time-effect assessment database for experiments on war-traumatized animals", Emergency and critical care medicine, 2(4), pp.219-224.

Wang H. X., Ericsson K., Winblad B., Fratiglioni L. (1998)"The Human Figure Drawing test as a screen for dementia in the elderly: A community-based study", Archives of gerontology and Geriatrics, 27(1), pp.25-34.

Zhao B., Zhang X., Li H., Yang Z. (2020) "Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions", Knowledge-Based Systems, 199, 105971.

Zou Z., Chen K., Shi Z., Guo Y., Ye J. (2023) "Object detection in 20 years: A survey", Proceedings of the IEEE, 2023.

Zubala A., Kennell N., Hackett S. (2021) "Art therapy in the digital world: An integrative review of current practice and future directions", Frontiers in Psychology, 12, 595536.