

Fine-Grained Class Detection via Hierarchical Pattern Learning for Psychoanalysis Drawing Interpretation

ABSTRACT

Early identification of emotional distress, trauma, or abuse in children remains a critical challenge, especially among populations with limited verbal expression. The House-Tree-Person (HTP) test, a widely used projective drawing assessment, offers insight into a child's psychological state through symbolic illustrations. However, its traditional interpretation is subjective, labor-intensive, and difficult to scale. In this study, I propose a deep learning-based system for automated interpretation of HTP drawings. The system is developed with a convolutional neural network (CNN) to detect and classify the three core elements (house, tree, and person) while accounting for variations in children's drawing styles, such as disproportionate features or missing components. By transforming hand-drawn, unstructured sketches into structured visual data, the system enhances consistency, reduces human bias, and increases the accessibility of psychological screening tools. This approach enables scalable and cost-effective deployment of HTP testing in schools, foster care, and clinical settings. I hope the proposed system can be a promising solution for early mental health assessment in children.

1. Introduction

In recent years, concerns about children's mental health have risen with increased attention toward identifying early signs of trauma, neglect, or abuse. According to the National Children's Alliance, an estimated 558,899 children (unique incidents) were victims of abuse and neglect in the U.S. in 2022. Research reveals that traumatic experiences disproportionately affect minority youth. Reports highlight tragic cases where Black children, particularly those adopted transracially, are often linked to systemic issues like racism and bias. They endure severe abuse as a result of inadequate monitoring and a lack of psychological support, with behavioral warning signs repeatedly overlooked until a tragic outcome exposed years of emotional and physical mistreatment. These events emphasize the urgent need for more accessible, consistent, and scalable methods for evaluating children's emotional well-being, especially in populations where verbal expression is limited or suppressed.

One existing tool that can help assess children's inner emotional and psychological states is the House-Tree-Person (HTP) Test (Guo et al. 2023). In this projective drawing assessment, a child is asked to draw a house, a tree, and a person. Each of these elements reflects different aspects of the child's inner world. Through detailed analysis, psychologists can identify potential signs of anxiety, trauma, abuse, or emotional detachment (Zhang et al. 2024). The HTP test is especially valuable in assessing young children or those who may have experienced traumatic events that are difficult to verbalize.

Despite its utility, the traditional HTP test faces major limitations. Its interpretation is highly subjective, often varying between practitioners based on their clinical background, cultural understanding, or personal judgment. This introduces significant error-proneness, reducing the reliability and consistency of results. Moreover, performing and analyzing the HTP test manually is time-intensive and costly, making it inaccessible in many schools, foster care, or community health settings. Without standardization or automation, the test cannot be scaled effectively, and it limits its potential as a widespread screening tool for children at risk.

To address these issues, I propose a deep learning-based object detection system specifically designed for the HTP test. Unlike general-purpose object detection models, this system is purpose-built to analyze children's HTP drawings. It is trained on domain-specific data to detect, classify, and localize the three symbolic components- house, tree, and person- with high accuracy and sensitivity to child-drawn visual features. The system utilizes a convolutional neural network (CNN) to extract drawing-specific visual features, followed by a detection module that outputs bounding boxes and class labels assigned to HTP categories. Importantly, the model can account for variations in children's drawing styles, such as disproportionate shapes or missing elements, which are clinically relevant. This enables the transformation of freeform, subjective drawings into structured, analyzable data that can assist in psychological interpretation. By automating and standardizing the evaluation process, this system reduces human bias, lowers cost, and significantly increases accessibility, making HTP testing a scalable and effective screening tool for identifying emotional distress in children.

2. Background Knowledge

2.1 House-Tree-Person Test

A House- Tree- Person test is a projective psychological assessment designed to diagnose aspects of an individual's personality through drawings (figure 1). Developed by John Buck in 1948, the so-called HTP test evaluates an individual's cognitive, emotional, and social functioning. The test is often used with children and individuals who have difficulty expressing themselves verbally. During the test, participants are asked to draw a house, a tree, and a person on a separate sheet of paper. They then analyze based on various elements such as size, placement, line, absence of content, and more. For example, a house with a smoking chimney might suggest warmth, comfort, or a sense of home and family, while the absence of it might suggest passivity or a lack of psychological warmth in home life. The HTP test is analyzed through quantitative and qualitative methods. Quantitative analysis scores specific drawing features using John N. Buck's standardized system to assess cognitive functioning. Qualitative analysis interprets the drawing and verbal responses to reveal emotional and personality traits along with the person's behavior during the test. Each element- the tree, person, and house- is assessed with indicators and context. As shown in Figure 1 (b), the tree is evaluated by its presence of branches, thickness of the trunk, depth or presence of roots, amount of leaves, and more.

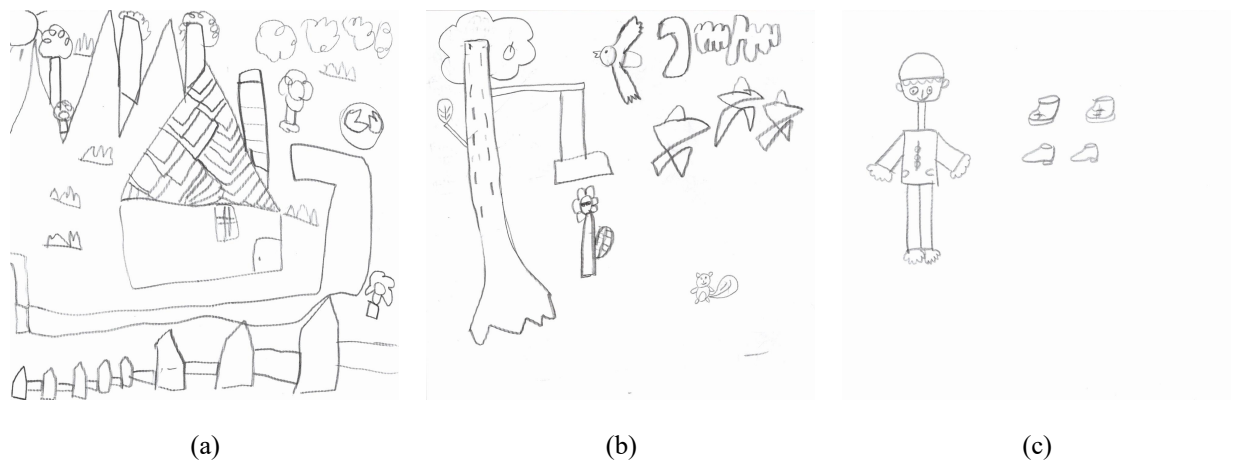


Figure 1. House-tree-person test

However, this test has several limitations. Its interpretation is highly subjective, relying on the psychologist's judgment and standards, and due to this fatal flaw, the results may be inconsistent. Especially during qualitative analysis, standards and interpretations can vary widely depending on the evaluator. It is also time-consuming and expensive, as it requires trained professionals to conduct the test with deep analysis. The test can take up to 150 minutes with full analysis and can range from \$200 to \$2,000,+ depending on location, provider, and what's included (Lacson 2024). This ultimately limits its accessibility for those who lack time, money, and for people without access to mental health services. To address these problems, I further explore this test by applying machine learning techniques to automate and standardize the analysis of HTP drawings, aiming to reduce bias, save time, cut costs, and elevate its accessibility.

2.2 Object Detection

Object detection is a computer vision technique used to identify and locate objects within an image by drawing bounding boxes around them and assigning labels (Zou et al. 2023). Unlike simple image classification, which only determines what is in the image, object detection also provides spatial information. This makes it essential for applications in autonomous driving, surveillance, and facial recognition.

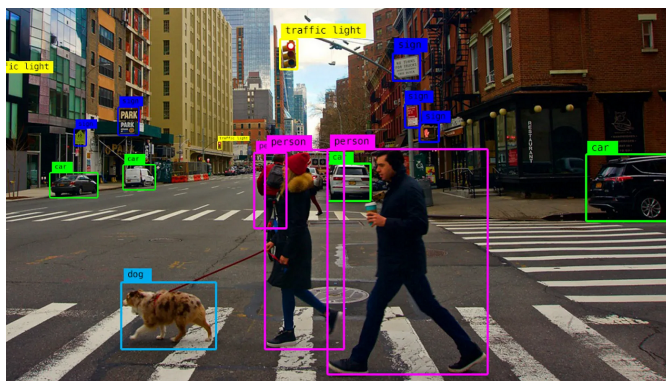


Figure 2. Object detection example (Yse 2024)

In this study, object detection is applied to automate the analysis of HTP (House-Tree-Person) drawings. Traditionally, HTP tests require manual interpretation by psychologists, which can be subjective, time-consuming, and costly. The system detects and labels key elements- house, tree, and person- within a child's drawing, providing bounding boxes and scores. It also labels the indicators of each element, such as identifying the presence of

windows, stability of the house, and analyzing the size of the door in a house. This allows for a more rapid and consistent analysis based on the spatial arrangement and presence of objects.

3. Psychoanalysis Drawing Interpretation Network

3.1 Training Methodology

Training an object detection model on House-Tree-Person (HTP) drawings poses challenges and hardships due to the grayscale input limitation of the drawings and the high domain variance of the input data. HTP drawings are typically done in pencil or a pen, resulting in grayscale images that lack the color and texture information commonly used by detection models to distinguish between classes. This leads to a loss of discriminative visual cues that are critical for the accurate detection of subtle features. Also, the drawings vary widely in style, proportion, and clarity due to differences in age, motor skills, and drawing ability of participants. This results in significant class variation, making it difficult for a model to detect effectively across different samples.

To address these challenges, this study adopts a curriculum-based training methodology, where the model is incrementally exposed to detection tasks. Starting from easier, more general classes and gradually moving toward more complex and fine-grained classes. This progressive approach helps the model stabilize early learning and adapt to increasing difficulty.

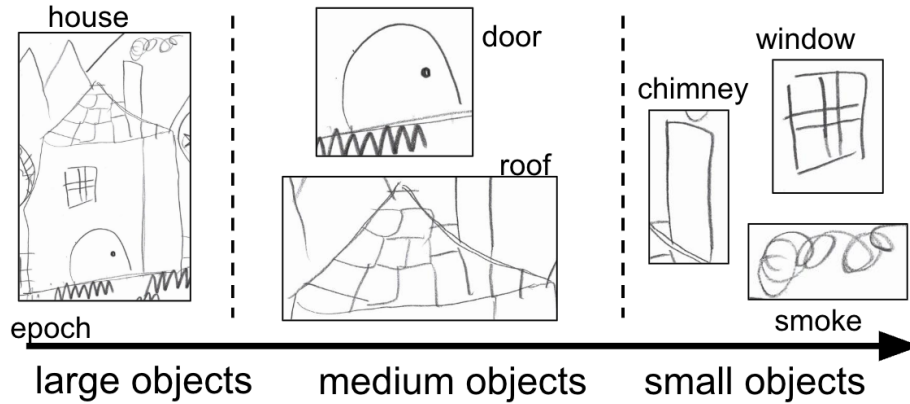


Figure 3. Gradual training curriculum

The training can be divided into three phases: coarse-level object detection, intermediate-level feature detection, and fine-grained object detection. The detection begins with large, easily recognizable object classes like the house, tree, and person. These objects are typically central, occupy more space on paper, and are more consistent across samples. This foundational phase enables the model to learn basic spatial context and structure within HTP drawings. Once the model satisfies the coarse classification, mid-size components such as windows, doors, and limbs are introduced. These features provide richer semantic detail that aligns with traditional psychological interpretation. Finally, the model is trained to detect small and subtle features such as a chimney, smoke, tree branches, facial features, and roots. These elements are crucial in nuanced psychological analysis but are also the most variable and difficult to identify, especially in low-resolution, grayscale images.

This methodology enhances the model's capacity to handle domain variance while reducing the risk of misidentifying ambiguous objects early in training. The gradual introduction of complexity also mimics the interpretive approach of human psychologists, who begin with comprehensive impressions and narrow down to detailed analysis. By structuring the learning process this way, the model achieves better generalization and interpretability for automated HTP assessment.

3.2 Architecture and Loss Function

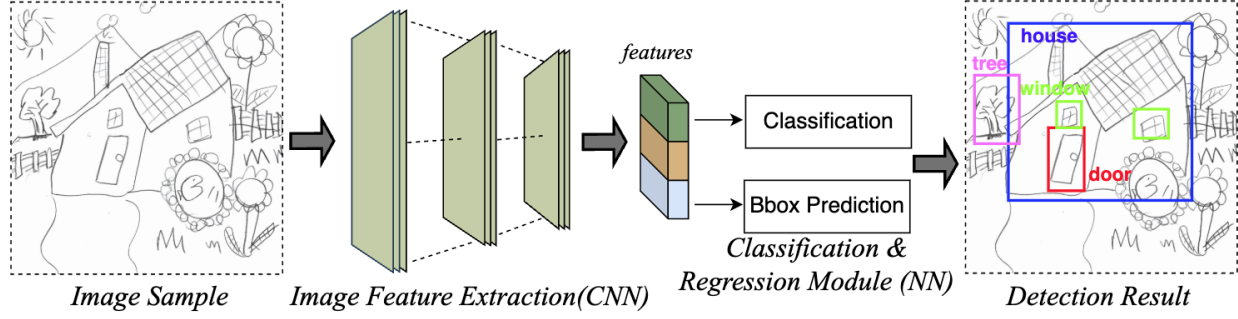


Figure 4. Network architecture

Figure 4 illustrates the architecture of the proposed network. The network takes an HTP image as input and generates a list of bounding boxes along with their corresponding class categories. To evaluate the network's performance, two loss functions are utilized: the bounding box regression loss function and the classification loss function.

The bounding box regression loss function enables the model to learn to draw boxes that closely match the location and size of the ground truth boxes. It compares the predicted box's center coordinates (x , y) with the ground truth to assess positional accuracy. Additionally, it compares the width (w) and height (h) to evaluate size differences. The loss decreases as the predicted box becomes more accurate in terms of position and size, guiding the model to improve its object localization.

Equation 1: Regression bounding box loss function

$$L_b = \sum_{i=1}^{S^2} \sum_{j=1}^M \mathbb{I}_{i,j}^{obj} (|x_{i,j} - \hat{x}_{i,j}|^2 + |y_{i,j} - \hat{y}_{i,j}|^2 + |\sqrt{w_{i,j}} - \sqrt{\hat{w}_{i,j}}|^2 + |\sqrt{h_{i,j}} - \sqrt{\hat{h}_{i,j}}|^2)$$

The classification loss function enables the model to correctly identify the object within the predicted box. It compares the predicted class probabilities with the true class label. The true label is marked as 1 if the object is present and 0 if it's not. The model gives a probability for each class, and the loss compares these predictions to the true labels. If the model is confident in the correct class (close to 1) and not confident in the wrong ones (close to 0), the loss is small. This helps the model learn to assign high scores to the right classes over time.

Equation 2: Classification loss function

$$L_c = \sum_{i=1}^{S^2} \mathbb{I}_i^{obj} \sum_{k=1}^K |p_k - \hat{p}_k|^2$$

4. Experimental Results

4.1 Dataset

To train and evaluate the system, I used 56,000 HTP (House-Tree-Person) samples collected from 7,000 children aged 7 to 13, with a balanced representation of genders. Each child submitted four types of HTP drawings- house, tree, male person, and female person- resulting in a comprehensive collection of child-drawn images. For each drawing, bounding box annotations were manually applied to mark key visual elements such as windows, doors, tree branches, arms, legs, and facial features. This results in 56,000 labeled samples specifically for object detection tasks.

4.2 Evaluation Metric

Confusion Matrix				
Total Observations (n)			Ground Truth Label	
			has disease Condition Positive (CP)	no disease Condition Negative (CN)
Predicted Label	test positive	Test Outcome Positive (TOP)	True Positive (TP)	False Positive (FP)
	test negative	Test Outcome Negative (TON)	False Negative (FN)	True Negative (TN)

Figure 5. Confusion matrix explanation (Cheyne 2025)

A confusion matrix is a tool used to evaluate the performance of a classification model by comparing the model's predictions with the actual labels. It displays the number of correct and incorrect predictions for each class: true positive, true negative, false negative, and false positive. From this matrix, important performance metrics such as accuracy, precision, recall, and F1-score can be calculated, providing insight into how well the model distinguishes between different categories.

Equation 3: Intersection over union

$$IOU = \frac{GT \cap PRED}{GT \cup PRED}$$

Intersection over union (IoU) is a common evaluation metric used in object detection tasks to measure how accurately a predicted bounding box overlaps with the ground-truth box (Rezatofighi et al. 2019). It is calculated by dividing the area of overlap between the predicted and actual boxes by the area of their union. A higher IoU value indicates a better prediction, and typically, a prediction is considered correct if the IoU exceeds a certain threshold. This metric helps determine not just whether an object was detected, but how precisely its location and size were identified.

4.3 Evaluation Result

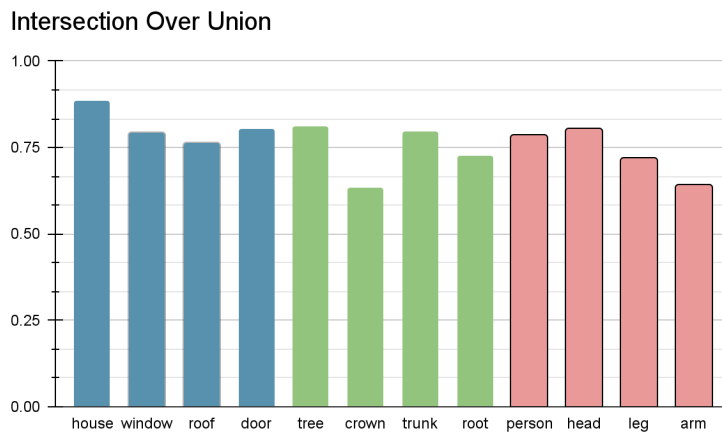


Figure 6. IoU evaluation for each category

As shown in Figure 6, the proposed system was evaluated using IoU scores across various HTP components to assess detection accuracy. Results demonstrated strong performance on major objects: House (0.885), tree (0.812), and person (0.787). Subcomponents also showed solid accuracy, including window (0.793), door (0.802), crown (0.635), arm (0.642), and leg (0.720), yielding slightly lower scores due to higher visual variability; their detection remains reliable. Components such as the crown had a lower accuracy because of their variance in range, while components like the head have a higher accuracy rate due to their consistency. These results validate the system's effectiveness in detecting both primary and detailed features of children's HTP drawings, enabling accurate and scalable psychological analysis.

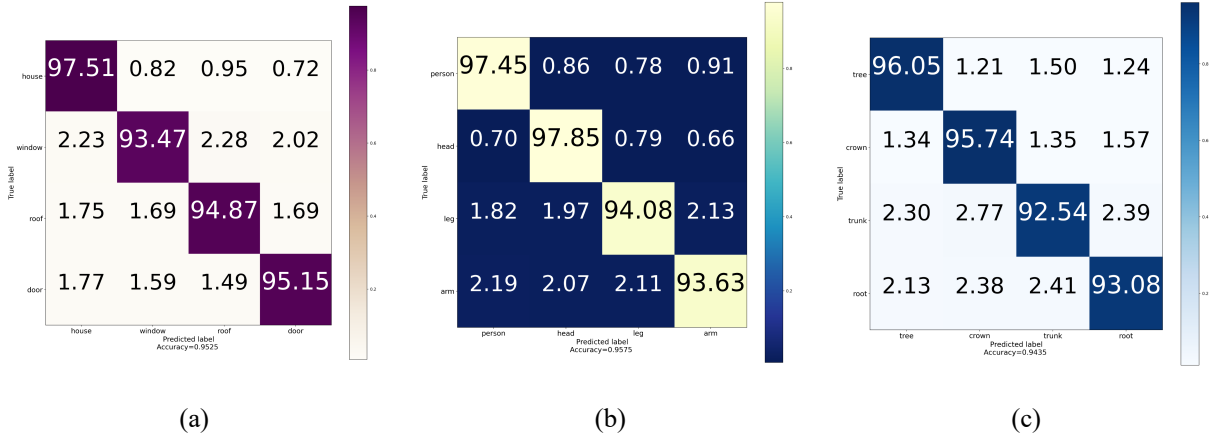


Figure 7. Confusion matrix evaluation
(a): house, (b): person, and (c): tree

Figure 7 presents a confusion matrix evaluation for three categories: (a) house, (b) person, and (c) tree. Figure 7 (a) shows the confusion matrix result of house categories which demonstrates strong classification performance with an average diagonal value of 95.25. All diagonal entries exceed 93 which indicates a high true positive rate and strong classification performance. Similarly, the confusion matrices for the person and tree categories yield an average value of 95.75 and 94.355, respectively. These consistently high diagonal values across all categories suggest that the model effectively distinguishes between the different object classes, with minimal misclassification.

5. Conclusion

In this research, I proposed a deep learning-based object detection system to automate and standardize the analysis of House-Tree-Person (HTP) drawings for psychological screening in children. By addressing the limitations of traditional HTP tests, such as subjectivity, time, cost, and limited accessibility, the model uses a curriculum-based training approach to detect both major components (house, tree, person) and finer details (windows, limbs, facial features) from grayscale child-drawn images. Trained on 56,000 annotated drawings, the system demonstrated strong potential for accurate and scalable analysis, with performance measured using metrics like IoU and confusion matrix. With the evaluation metrics, the system achieved average performance scores of 95.25%; house, 95.75%; person, and 94.35%; tree, which highlights its potential for reliable and consistent analysis across a wide range of drawing styles. These results confirm the feasibility of applying object detection techniques in a clinical psychological context and mark a critical advancement in digitizing and automating projective tests. This study advances the intersection of machine learning and mental health by creating a more accessible, objective framework for early psychological screening, especially for underserved populations. Ultimately, this research lays the groundwork for scalable, inclusive tools to detect emotional distress in children through their drawings.

References

AI Hub. (2025, Jun 23). "Dataset for AI-based child art psychology diagnosis": AI Hub.
<https://aihub.or.kr/aihubdata/data/view.do?dataSetSn=71399>

Cheyne, H. (2025, Jun 18). "Binary Classification Metrics": Biosymetrics.
<https://www.biosymetrics.com/blog/binary-classification-metrics>

Guo, H., Feng, B., Ma, Y., Zhang, X., Fan, H., Dong, Z., ... & Gong, Q. (2023). Analysis of the screening and predicting characteristics of the house-tree-person drawing test for mental disorders: A systematic review and meta-analysis. *Frontiers in Psychiatry*, 13, 1041770.

Khanam, R., & Hussain, M. (2024). What is YOLOv5: A deep look into the internal features of the popular object detector. *arXiv preprint arXiv:2407.20892*.

Lacson, N. (2024, Jul 15). "What is the concept of the house-tree-person?": Carepatron.
<https://www.carepatron.com/templates/house-tree-person-test>

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 658-666).

Yse, D. L. (2024, Oct 23). "Computer Vision | Object Detection using Python": Medium.
<https://lopezyse.medium.com/computer-vision-object-detection-with-python-14b241f97fd8>

Zhang, J., Yu, Y., Barra, V., Ruan, X., Chen, Y., & Cai, B. (2024). Feasibility study on using house-tree-person drawings for automatic analysis of depression. *Computer Methods in Biomechanics and Biomedical Engineering*, 27(9), 1129-1140.

Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. *Proceedings of the IEEE*, 111(3), 257-276.