Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240434

Unveiling the Potential of ChatGPT and YOLOv7 for Evaluating Children's Emotions Using Their Artistic Expressions

Uzair SHAH[a], Sulaiman KHAN[a], Mahmood ALZUBAIDI[a], Marco AGUS[a] and Mowafa HOUSEH[a,1]
[a] *College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar*

Abstract. Recent advancements in large language models (LLMs) have sparked considerable interest in their potential applications across various healthcare domains. One promising prospect is leveraging these generative models to accurately predict children's emotions by combining computer vision and natural language processing techniques. However, understanding children's emotional states based on their artistic expressions is equally crucial. To address this challenge, this paper presents a pipelined architecture comprising YOLOv7 and the powerful GPT-3.5 Turbo language model, where YOLOv7 is employed for object detection using art therapy imaging annotations, while GPT-3.5 interprets the sketches. After rigorously evaluating the proposed framework through a series of comprehensive experiments, we observed that our model achieved high confidence scores for both object detection and emotion interpretation. The robust performance of the proposed framework not only aids in explaining children's art but also provides valuable insights for parents and therapists. This capability enables them to better understand children's emotional states based on their artistic expressions, ultimately facilitating improved support and care.

Keywords. Emotion sensing, Kids drawing, LLM, ChatGPT, YOLOv7

1. Introduction

The evolving pressures of modern society increasingly cast a spotlight on mental health concerns, affecting not only adults but also children. The ability of children to regulate and express their emotions is a critical facet of their development, influencing their emotional competence, self-efficacy, and ultimately, their well-being [1,2]. Various strategies are employed by children to navigate their emotions, from seeking comfort to engaging in distraction techniques, which become more sophisticated with age [3]. Among these strategies, drawing stands out as a powerful means for children to articulate their feelings, serving both as an outlet for emotional expression and a window into their internal states [4,5]. The exploration of children's drawings as a medium to understand their emotional and social needs has a rich history. For instance, Steyn and Moen [6] highlighted the prevalence of themes such as violence and loss in children's representations of sadness, underscoring the potential of art to reveal significant insights

[1] Corresponding Author: Mowafa Househ; E-mail: mhouseh@hbku.edu.qa.

into the emotional landscapes of young minds [6]. However, the manual analysis of these drawings is both time-consuming and resource intensive.

Advancements in deep learning models, particularly in object detection and image processing, have transformed the technological landscape of emotion recognition and analysis. These innovations pave the way for more efficient, automated methods of interpreting the visual narratives presented in children's artwork. The emergence of LLMs has further expanded the possibilities for understanding and interacting with human-like text, presenting new avenues for analyzing and deriving meaning from children's drawings [7,8].

Despite these advancements, a gap remains in the literature regarding the integrated use of deep learning techniques and LLMs for object detection and emotional interpretation in children's art. Our research introduces a pioneering approach that utilizes YOLOv7 for object detection and GPT-3.5 Turbo for interpreting sketches. While our framework aims to estimate children's emotional states based on their artistic expressions, it is important to acknowledge that these interpretations are approximations and do not directly reflect the actual emotional states of the children at the time of drawing. This limitation underscores the necessity of integrating clinical insights and further research to enhance the accuracy and sensitivity of our emotional assessments.

This work introduces three key advancements in the interpretation of children's emotional expressions through art. We commence with the detailed annotation of children's drawings by a healthcare therapist to pinpoint emotional and stress-related themes, establishing a foundational dataset for analysis. This is complemented by our development of a cutting-edge architecture that marries YOLOv7's object detection capabilities with GPT-3.5 Turbo's prowess in interpreting artistic expressions, creating a holistic approach to understanding the emotional narratives within children's art. Finally, we validate our model's effectiveness by employing diverse metrics, ensuring its proficiency in both identifying and interpreting the nuanced emotional content depicted in drawings. Together, these contributions signify a significant leap forward in leveraging technology to decode the emotional well-being of children through their creative outputs.

2. Methodology

Figure 1 represents the overall experimental setup followed to perform this research work. Kids drawings are accumulated from publicly available online resources and art therapist is engaged to annotate these drawings for emotions, objects, and perform interpretation. In this work, a Python script systematically extracted approximately 10,000 children's drawings from online sources. The images underwent meticulous annotation on the Roboflow platform, involving the identification and annotation of around 50 distinct objects. Out of the initial pool, 3,000 images were selected for model training and validation. To standardize the dataset, all images were resized to 640x640 pixels, suitable for YOLOv7. Augmentation techniques, such as flips, zooming, rotation (±15 degrees), and saturation adjustment (±25%), were applied, tripling the dataset size. After preprocessing, the dataset was partitioned for training (92%, 7,208 images), validation (4\%, 300 images), and testing (4%, another 300 images) to ensure a balanced distribution for evaluating the model's performance.

2.1. Experimental setup

YOLOv7, a state-of-the-art object detection algorithm, is chosen over alternatives like RCNN, FastRCNN, and Detector2 for its superior performance. The model is trained with a batch size of 8, using the default YOLOv7 optimizer over 100 epochs. Validation guides training, and the best weights are retained for enhanced performance during testing. To attribute emotions to objects, an emotion mapping dictionary is developed, aiding in associating emotions with identified objects in children's drawings. Positive, negative, and neutral objects are aggregated, and their sizes quantified. This comprehensive approach ensures a systematic evaluation of YOLOv7's efficacy in object detection, adding a nuanced layer to understand emotional expressions in visual content.

2.2. LLM Explainability

In collaboration with an art therapist, we crafted a comprehensive emotional description of objects in children's drawings to generate prompts for report generation. Using the YOLOv7 model, we identified objects, calculated average confidence scores, and aggregated areas covered. To create effective and visually appealing reports, we developed a prompt template incorporating object details, emotional descriptions, confidence scores, and spatial dimensions. This template also includes an overarching emotional analysis of the entire image.

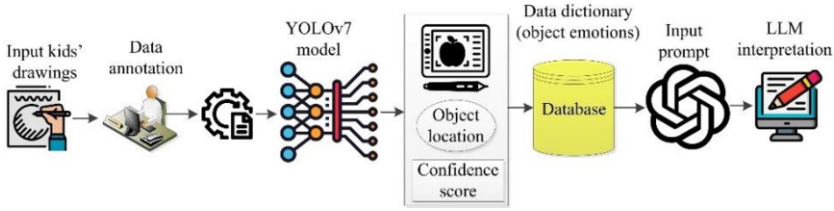
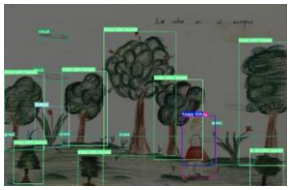
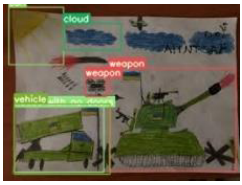


Figure 1. Proposed experimental setup.

3. Result

To consolidate the information obtained from the YOLOv7 model, encompassing object location, size, and confidence scores, we formulated a prompt template. This template was then utilized to convey the relevant details to ChatGPT, enabling the generation of meaningful reports for parents and teachers. Table 1, provides a comprehensive overview, including the input image, model annotated image, object information, and the corresponding generated reports. This integration of object detection, emotion assignment, and natural language processing ensures a holistic understanding of the emotional content within children's drawings. The crafted prompt was then input into the ChatGPT-3.5-turbo model for report generation. A representative sample of the prompt structure is provided, accounting for confidence levels, object names, and emotional interpretations. This method ensures a systematic and informed report generation process, combining human expertise with machine learning capabilities to provide insightful and nuanced interpretations of children's drawings.

Table 1. Experimental results generated by YOLOv7 and GPT-3.5 for original and annotated image.

Original drawing		
Annotated drawing		
YOLOv7 interpretation	'grass', 'clouds', 'flowers', 'trees with leaves', and 'happy kids'.	'weapons', a 'house with no doors', a 'vehicle', the 'sun', and a 'cloud'.
LLM interpretation	The drawing, rich with nature, shows trees with a 70% detection certainty dominating the scene, suggesting the child's affinity for nature. Grass at a 60% certainty and flowers at 75%, though less prominent, imply stability and bursts of happiness, respectively. A cloud, 80% likely, spans the sky, indicating complex emotions, while the image of happy children, 85% certain but less frequent, points to special joyful moments in the child's life.	The child's drawing subtly navigates a blend of emotions. A weapon, uncertain at 30% yet occupying 32.1% of the space, suggests danger and prompts resilience. The sun shines with 70% certainty and 7.49% presence, casting positive light. Freedom rolls in at 41% confidence, taking up 16.03%, embodied by a vehicle. A cloud, fairly certain at 77% yet only 3.62% of the image, reflects varied emotions. Lastly, a house without doors, at 27% certainty and 13.99% of the drawing, offers a sense of openness and positivity.

4. Discussion

In this work, we devised a comprehensive AI pipeline dedicated to recognizing emotions in children's drawings. Leveraging a combination of computer vision and natural language processing, we successfully crafted an end-to-end solution. Our methodology involved the initial training of a YOLOv7 object detection model, which adeptly identified various objects within the children's drawings [8].

The qualitative results, summarized in Table 1, demonstrate the efficacy of leveraging GPT-3.5 with YOLOv7 in generating concise and precise information based on the provided input prompt. Importantly, it is observed that YOLOv7 occasionally misidentifies objects, leading to conclusions with low confidence scores. These inaccuracies can impact the generation of emotionally accurate interpretations from the drawings. Given that objects in a child's drawing can evoke multiple emotions depending

on context, collaboration with art therapists proved invaluable [8]. Furthermore, it is crucial to note the inherent limitations of our methodology—specifically, the challenge of accurately interpreting emotions without direct knowledge of the children's emotional states when they created their drawings. This underscores the importance of viewing our results as initial indicators rather than definitive assessments of emotional states. For instance, a tree's emotional impact is influenced by factors such as its placement and the presence of leaves [8]. To address challenges like annotating trees with and without leaves as distinct entities, we sought insights from art therapists. In order to simplify model training and mitigate the complexities associated with various factors influencing emotional expression in children's drawings, we adopted a broader approach. This involved categorizing emotions, recognizing the intricate nature of emotional responses in children's artwork. Our work thus underscores the importance of addressing nuanced aspects in emotion recognition to enhance the accuracy of interpretations derived from children's drawings.

5. Conclusions

In this study, we presented a pipeline combining YOLOv7 for object detection and GPT-3.5 Turbo for interpreting children's drawings to discern their emotional themes. YOLOv7 detects objects within the artwork, and GPT-3.5 provides interpretations that align closely with those of a professional art therapist. Our extensive experimental evaluations demonstrate that the model achieves high confidence levels in both object detection and emotional interpretation. The efficacy of this system in decoding children's artistic expressions provides parents and therapists with deeper insights into the children's emotional states, thus enhancing the support and care for their developmental needs.

References

[1] Denham SA, Blair KA, DeMulder E, Levitas J, Sawyer K, Auerbach–Major S, et al. Preschool Emotional Competence: Pathway to Social Competence? Child Dev 2003;74:238–56. https://doi.org/https://doi.org/10.1111/1467-8624.00533.

[2] Nancy Eisenberg TLS, Morris AS. Regulation, Resiliency, and Quality of Social Functioning. Self Identity 2002;1:121–8. https://doi.org/10.1080/152988602317319294.

[3] Zimmer-Gembeck MJ, Skinner EA. Review: The development of coping across childhood and adolescence: An integrative review and critique of research. Int J Behav Dev 2011;35:1–17. https://doi.org/10.1177/0165025410384923.

[4] Drake JE. How Drawing to Distract Improves Mood in Children. Front Psychol 2021;12. https://doi.org/10.3389/fpsyg.2021.622927.

[5] Drake JE. How children can use drawing to regulate their emotions. Theory Pract 2023;62:181–92. https://doi.org/10.1080/00405841.2023.2202132.

[6] Steyn M, Moen M. Drawing sadness: what are young children telling us? Early Child Dev Care 2019;189:79–93. https://doi.org/10.1080/03004430.2017.1301936.

[7] Chung JJY, Kim W, Yoo KM, Lee H, Adar E, Chang M. TaleBrush: Sketching Stories with Generative Pretrained Language Models. Proc. 2022 CHI Conf. Hum. Factors Comput. Syst., New York, NY, USA: Association for Computing Machinery; 2022. https://doi.org/10.1145/3491102.3501819.

[8] Ali N, Abdrazaq A, Shah Z, Househ M. Artificial Intelligence-Based Mobile Application for Emotion Sensing for Children Through Art. MEDINFO 2021 One World, One Heal. Partnersh. Digit. Innov., IOS Press; 2022, p. 1130–1.