The Illustrated Page: Analyzing Illustrations of Historical Children's Books Using Citizen Science

Andrew Piper[1], Jiaming Jiang[2], and Robert Budac[3]

[1] Department of Languages, Literatures, and Cultures, McGill University, Montreal, Canada
[2] College of Humanities, EPFL, Lausanne, Switzerland
[3] Digital Humanities, University of Alberta, Edmonton, Canada

Abstract

This paper presents the first large-scale, systematic study of historical children's book illustrations through a combination of citizen science and computational analysis. Using a corpus of 27,901 digitized illustrations from 2,827 books from the Internet Archive's Children's Library, we developed a structured annotation workflow deployed on Zooniverse to collect over 400,000 annotations from 902 volunteers. Tasks included identifying depicted characters, objects, settings, and emotional tone. We assess inter-annotator reliability across task types and derive consensus labels to explore three central questions: who and what is most commonly visualized, which entities co-occur, and how visual depictions change over time. Findings reveal dominant portrayals of patriarchal figures and animals, the centrality of nature, and gendered patterns in emotional framing. Temporal analysis shows a surprising visual stability over 140 years. This work demonstrates the value of human-in-the-loop annotation for visual cultural heritage and provides a new resource for studying the visual language of childhood in print.

Keywords: children's literature, illustrated books, history of print, image understanding, citizen science, distant viewing

1 Introduction

Illustrated children's books offer a rich archive for understanding cultural history, combining visual and textual modes of storytelling that reflect societal values across time. Scholars have long emphasized that picture books are complex narrative artifacts in which words and images interact to produce meaning [24; 25]. This multimodal structure makes them powerful instruments of cultural transmission, encoding assumptions about childhood, gender, identity, morality, and emotion within their language and visual design [12; 15]. Historical studies have traced how these books foreground ideological currents and pedagogical aims, while psychological research underscores their formative role in early cognitive and emotional development [16; 23; 35]. Yet despite this longstanding interdisciplinary recognition, large-scale analysis of historical children's book illustrations remains limited especially with respect to their visual content.

Illustrations in children's books have long been read not merely as decorative accompaniments to text, but as cultural documents that encode shifting ideas of childhood and social order. Historians of childhood have shown how visual culture participates in defining what it means to be a child, often reinforcing classed, gendered, and racialized norms about innocence, obedience, play, and agency [1; 8; 11; 13; 30]. Studies of children's book illustration similarly emphasize the visual construction of gender roles, family relations, and moral conduct, revealing how images work

Andrew Piper, Jiaming Jiang, and Robert Budac. "The Illustrated Page: Analyzing Illustrations of Historical Children's Books Using Citizen Science." In: *Computational Humanities Research 2025*, ed. by Taylor Arnold, Margherita Fantoli, and Ruben Ros. Vol. 3. Anthology of Computers and the Humanities. 2025, 280–294. https://doi.org/10.63744/XVr0QDckSvkj.

pedagogically to shape expectations of behavior and identity [6; 7; 30]. At the same time, feminist and queer studies scholarship has highlighted the ways illustrations reproduce or contest cultural hierarchies, offering insights into how visual storytelling both reflects and regulates broader social imaginaries [14; 26; 33].

To support deeper computational research into the representation of childhood, this paper introduces a new dataset of ca. 28,000 annotated children's book illustrations created through an on-going citizen science project called "Picturing Children's Stories."[1] The project aims to enlist the public to help build more transparent, human-centered AI models for understanding human storytelling. The dataset is derived from the Internet Archive's Children's Library, which consists of 2,827 digitized books published between the years 1728 and 1932.[2] The first phase of annotations focuses on identifying core visual features associated with character age, gender, setting, the physical environment (objects and structures), and an assessment of emotional intensity to identify the degree of agency and activation connected to different character types and settings. Doing so allows us initial insights into the changing presence over time of different character types, social relationships, physical environments, and the emotional associations surrounding these interactions. The full dataset can be found here.[3]

2 Background

Recent advances in multimodal large language models (MLLMs) have opened new possibilities for the automated analysis of cultural heritage images. To support this work, researchers have developed specialized datasets such as SemArt [10], ArtPedia [32], ArtCap [18], and DEArt [29], which pair artworks with expert descriptions, iconographic labels, or object annotations. These resources have enabled the training of captioning models that incorporate art-historical knowledge [4] and the evaluation of zero-shot classification and metadata prediction by general-purpose LLMs [28; 34]. Arnold and Tilton [3] propose a general framework for using MLLMs to generate textual captions and embedding-based similarity scores for large image corpora. Their approach supports explainable recommendations and cluster-based exploration of visual heritage collections, demonstrating how automated captioning can enhance searchability and interpretability without the need for pre-annotated metadata. Further work has also explored the interpretibity of historical newspaper illustrations [9], photographs [2; 21], maps [19], as well as graphical aspects of scientific publishing [17; 27], providing a broad range of approaches to studying visual cultural heritage materials.

Compared to canonical artworks or photographs, historical book illustrations present distinct challenges for image understanding tasks [5; 20]. Unlike paintings, which are self-contained visual artifacts, or photographs which are more closely aligned with contemporary training data, book illustrations are embedded within larger narrative or instructional texts, such that their meaning often relies on sparser visual representations. Many illustrations are also low-resolution, monochromatic, or stylistically constrained by period-specific reproduction methods such as woodcuts or engravings. Additionally, some are subject to conditions of historical decay such as "foxing," where mold or oxidation commences prior to digitization, or quirks of historical production such as "show-through," when ink from the opposite side of a page shows through, and "off-setting," when ink from another page bleeds onto the back of the page. As a result, accurately interpreting these images requires models to recognize subtle, often noisy visual cues and historical iconography that may be underrepresented in modern training data or art historical data.

[1] https://www.zooniverse.org/projects/citizenreaders/picturing-childrens-stories
[2] https://archive.org/details/iacl
[3] https://doi.org/10.5683/SP3/KTSY9B

3 Data

3.1 Overview

Illustrations were sourced from a collection of 2,827 digitized books available through the Internet Archive's Children's Library. Over 91% of books included in the collection are English-language books, with the next most represented language being French at just under 5%. Based on an analysis of available regional information, we estimate that books originate equally from Britain and North America. To extract candidate images for annotation, we used PyMuPDF[4] to statistically analyze gray-scaled page images and identify illustration-containing pages. This resulted in 31,051 candidate page images. Figure 1 shows the distribution of publication dates of illustrations.

To support the large-scale human annotation of page images, we developed a citizen science project using the platform Zooniverse.org. Figure 6 in the Appendix shows the project homepage. The project interface included a number of features to support the reliable classification of material: an "About" page to inform participants of the purpose of the research and how the data will be handled; a tutorial which guides participants through the task structure; pop-up help boxes that provide further examples/instructions; and finally talk pages where participants can pose questions to researchers, where we had a dedicated team of three moderators.

Over a two-week period between May 20 and June 2, 2025, a total of 902 volunteers contributed a total of 457,756 annotations. Engagement was varied with 7.3% of users completing only one annotated image, while the top 25% of contributors accounted for 80% of all annotation activity. In all, 27,901 illustrations were positively identified by a consensus of annotators (i.e. majority vote).

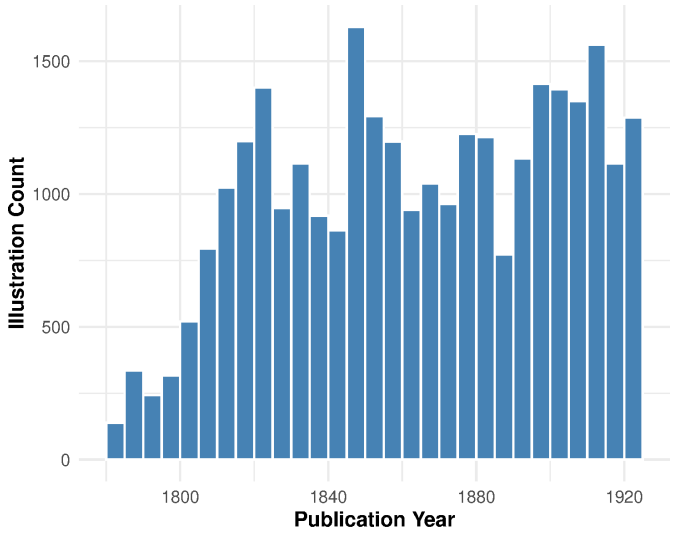

Figure 1: Distribution of publication dates of illustrations in the Internet Archive Children's Library.

3.2 Task Structure

Participants were presented with digitized pages from historical children's books and asked to complete a structured sequence of annotation tasks based on the visual and textual content of each page (see Figure 2 for an example of the task interface). We required a retirement rate of three

[4] https://github.com/pymupdf/PyMuPDF

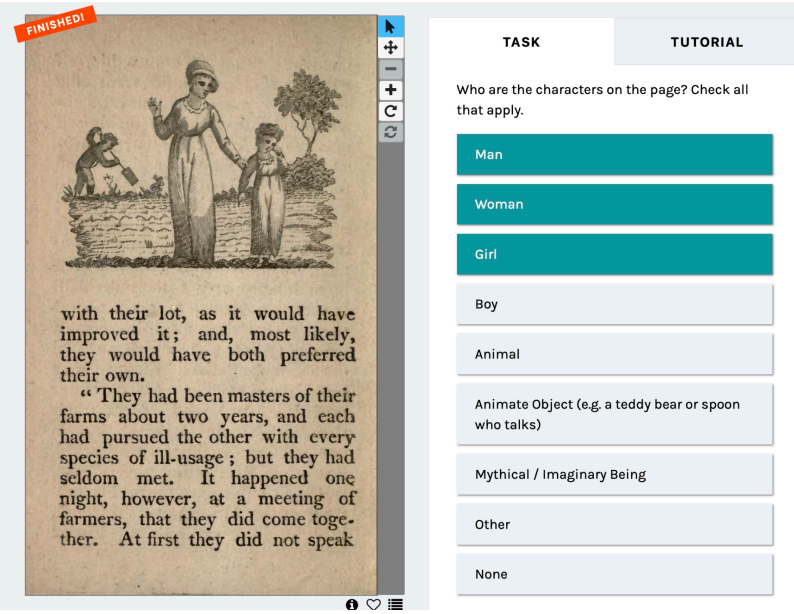

Figure 2: Example screenshot of one stage of the task workflow.

annotators per image. The workflow began with a screening question asking whether the page contained an illustration. We defined an illustration as:

> *An illustration is a realistic image of a scene or person from a story. An illustration is NOT: decorative graphics, a logo, book stamp, or photograph.*

Based on the majority agreement of annotators, we found that roughly 10% (ca. 3,150) of all extracted images were labeled as non-illustrations according to our criteria. This is a valuable step in the workflow given that the distinction between visual elements that are and are not considered illustrations can be very subtle.

If the participant answered "Yes" to this question, they were then guided through a series of visual classification tasks concerning the illustration as shown in Table 1. These included identifying the age, gender, and species of characters (T1); whether the illustration was set indoors or outdoors (T2); the objects and structures depicted (T3); and the emotional intensity of the scene (T4). Each task was presented with an optional tutorial and contextual help, ensuring consistent interpretation across annotators.

This workflow was designed to balance interpretive nuance with structured metadata collection to facilitate both future analysis and multi-modal large-language model training. Entity types and their co-occurrence can help us understand the overall visual presence of entities over time, depicted social structures, orientations towards the natural and built environment, along with the emotional association of these interactions. We chose to focus on "emotional intensity" (also known as "arousal" in the psychological literature) for two reasons. The first is this is a well-established concept in psychology [31] and computational text analysis [22]. Second, we found this emotional dimension the most relevant for contextualizing the emotional agency of characters, which can begin to give us insights into the ideological portrayal of entities in children's books.

3.3 Inter-Annotator Agreement

To assess the reliability of the citizen-science annotations, we computed inter-annotator agreement separately for each task type. Tasks varied in structure: some required a single categorical response

Task	Category	Question	Possible Answers
T0	Illustration	Is there an illustration on this page?	*Yes, No*
T1	Characters	Who are the characters on the page? Check all that apply.	*Man, Woman, Girl, Boy, Animal, Animate Object (e.g., a teddy bear or spoon who talks), Mythical / Imaginary Being, Other, None*
T2	Setting	Where is the scene set?	*Indoors, Outdoors, Both / Ambiguous*
T3	Objects	What are the main objects in the scene? Check all that apply.	*Nature (trees, flowers, lakes), Architecture (buildings, walls, furniture), Vehicles (cars, boats, planes), Toys, Reading and Art (books, musical instruments, paintings), Household items and tools (pots, pans, hammers, sticks), Other, None*
T4	Intensity	How emotionally intense is this scene? What is the energy like?	*1 - Very calm (still, contemplative, restful), 2 - Mildly calm, 3 - Neutral, 4 - Mildly intense, 5 - Very intense (frantic, tense, active)*

Table 1: Annotation task structure: order, categories, questions, and response options.

(T0, T2), others allowed for multiple selections (T1,T3), and one required ordinal judgments (T4). Table 2 summarizes the different agreement measures across tasks.

For the single-answer tasks (T0, T2), we measured the percentage of majority versus full agreement as an intuitive baseline of how much disagreement there was. We did not use a traditional IAA measure such as Krippendorff's alpha due to the sparsity of annotator coverage across items—most participants annotated only a small subset of the dataset—violating the assumptions of that metric. As we can see in Table 2, agreement was high for both single-answer tasks, with 85% full agreement on the presence of an illustration and 93% full agreement on indoor/outdoor setting. The former is particularly surprising given the number of potential edge cases with decorative headpieces or borders.

For multi-answer tasks (T1 and T3), we calculated agreement using pairwise Jaccard similarity: the size of the intersection divided by the size of the union of label sets across annotators. Each item's agreement score was computed as the mean Jaccard similarity across all annotator pairs. Results showed moderate to high consistency. For T1 (character types), the mean Jaccard similarity across items was 0.73 (median = 0.78), with 25% of items showing perfect agreement (score = 1.0). For T3 (object types), the average was lower at 0.22 (median = 0.56), suggesting greater subjectivity and label diversity in this task.

To put these scores into context, we provide two illustrative scenarios involving hypothetical annotators assigning character labels and their attendant Jaccard agreement scores:

- **Example 1: High Agreement**
  - Annotator A: {Man, Woman, Girl}, Annotator B: {Man, Woman, Girl, Boy}
  - Intersection: {Man, Woman, Girl} = 3, Union: {Man, Woman, Girl, Boy} = 4
  - Jaccard similarity: $\frac{3}{4} = 0.75$
- **Example 2: Moderate Agreement**

- Annotator A: {Man, Woman}, Annotator B: {Man, Woman, Girl, Boy}
- Intersection: {Man, Woman} = 2, Union: {Man, Woman, Girl, Boy} = 4
- Jaccard similarity: $\frac{2}{4} = 0.5$

The ordinal rating task (T4), which asked annotators to assess the emotional intensity of each scene on a 1–5 scale, was evaluated using full agreement, majority agreement, and average pairwise distance. 20% exhibited full agreement and 69% achieved at least two-thirds majority agreement. We also calculated the mean pairwise distance in ordinal ratings per item (i.e., average absolute difference across annotator pairs). The mean pairwise difference was 0.98, with a median of 0.67, indicating that while exact agreement was uncommon, ratings tended to cluster within ±1 point on the scale.

Together, these results indicate a good degree of inter-annotator reliability across task types, with especially strong consensus on binary and character-labeling tasks, and moderate but interpretable variation on tasks involving emotional or perceptual judgment.

Task	Name	Type	Full Agree	Majority	Jaccard	Ordinal Distance
T0	Illustration	Single (2)	85.4%	99.9%	–	–
T2	Setting	Single (3)	75.3%	95.5%	–	–
T1	Characters	Multi (9)	52.5%	72.9%	0.73 / 0.78	–
T3	Objects	Multi (8)	33.7%	60.8%	0.22 / 0.56	–
T4	Intensity	Ordinal (5)	15.3%	68.9%	–	0.98 / 0.67

Table 2: Summary of inter-annotator agreement for each annotation task where there are a minimum of three annotators per image. Type includes answer type plus total number of classes considered. In addition to percentage of full / majority agreement, we report mean / median Jaccard similarity for multi-answer tasks (T1, T3) and mean and median pairwise distance in rating scores for the ordinal task (T4) (1.0 = annotators were on average 1 point apart in their ordinal ratings).

3.4 Consensus Table

To derive consensus labels from multiple annotators per image, we implemented a task-specific aggregation pipeline. We first filtered out all illustrations with fewer than three annotations to ensure reliability. For single-label tasks (T0, T2, T4), we computed the most frequent response per image. If any label was selected by a majority (>50%) of annotators, it was marked as a ”majority” consensus. In the case of n=3 annotators, if all annotators disagreed the item was labeled ”disagreement” and all unique labels were retained. In cases of n > 3 annotators where no majority existed but some overlap was present, we labeled the outcome as ”partial disagreement”, retaining any label selected by two or more annotators. For multi-label tasks we followed the same logic but given the possibility of multiple annotations per task, ”partial disagreement” was also possible for n=3 annotators.

4 Analysis

To better understand how illustrations reflect and shape ideas about childhood, we analyze the citizen-science generated consensus table across three core dimensions: who appears most frequently, who is more likely to appear together, and how these patterns change over time. Together, these three perspectives—frequency, association, and change—offer preliminary insights into the

visual construction of childhood and its emotional worlds in historical print culture. For our measurements, we condition on labels from the consensus table under all conditions (majority, partial agreement, and disagreement). All visualizations of the data are included in Appendix B.

4.1 Jack and Jill Went Up the Page: Who and what is most often visualized?

We first tabulate the most commonly represented types of characters, objects, settings, and emotional intensities to establish baseline frequencies (Figure 4 in the Appendix). Our data indicates a few salient insights:

- **Patriarchy rules but animals aren't far behind.** Men are the most frequently depicted character in this collection and appear on almost half of all illustrated pages. They are 44% more likely to appear than women, while animals are 14% more likely to appear than women. Even though we are in the sphere of children and childhood, men and animals still predominate in the long nineteenth century.
- **Children show surprising gender balance.** While boys are statistically more likely to appear than girls, the effect is quite small with an 8% increased chance of representation.
- **Nature rules.** The central "space" of childhood storytelling in the long nineteenth century is the natural world as can be seen in the predominance of animals, outdoor settings and natural objects. Illustrations are 2.5x more likely to be set outdoors than inside.
- **Fantasy is a minor feature.** Despite the popularity of rhymes and stories featuring inanimate objects ("Hey diddle, diddle"), they are surprisingly infrequent in our data comprising just 3% of all entities.
- **Scenes skew calm, but cover the full spectrum of intensity.** Calm scenes occur 30% more often than intense scenes but the bulk of the data occurs around the neutral midpoint. While this may reflect annotator bias towards neutrality it also indicates important insights regarding how childhood is framed in printed books of the past, specifically, as an experience that can range across the full spectrum of emotional intensity.

4.2 Birds of a Feather: Who flocks together?

Next we examined co-occurrence patterns between entities (e.g., Girl–Woman), entities and objects (e.g., Boy–Toys), and entities and emotional intensities (e.g., Animal–Very Intense), using Fisher's exact tests (with Bonferroni correction) to identify statistically significant associations (Figure 5). Here we find:

- **Men and animals occupy their own (active) worlds.** Interestingly, both men and animals are not meaningfully associated with any other entities (despite a small positive association with women). This suggests they are more likely to appear alone than with other types of characters. At the same time, they are also the most strongly associated with the upper end of the emotional intensity scale, while women occupy more of the middle.
- **Fantasy is where the action happens.** Fantastic entities (animate objects and imaginary beings) are far more likely to appear together. Imaginary beings, but not animate objects, are strongly associated with high intensity events, while animate objects are not surprisingly associated with other toys and domestic items.

- **Boys and girls appear together.** Girls are the only entity that boys are more likely to appear with, while for girls they are also likely to be in the company of women. This suggests that childhood is being configured as a world unto itself, especially for boys. Given the co-occurrence of boys and girls we don't see differences in their object worlds. For emotional intensity, not surprisingly we do see girls less likely to be associated with intense actions than boys.

4.3 Hickory-Dickory-Decade: How has the visualization of children's stories changed over time?

Finally, we measure occurrences by decade to observe any meaningful shifts in depictions of childhood in our data (Figure 3 in the Appendix).

- **More historical stability than not.** With a few exceptions we see surprisingly consistent behavior across our 140-year timeline. Neutral emotional intensity, the balance between boys and girls, and the predominance of natural settings all persist over time.
- **Men return to their 18th-century levels after mid-century declines.** One notable exception to this rule is the decline and rise of men. We see a decline in the number of men depicted, which is replaced by children (both boys and girls) in the 1870s, followed by a stark increase in the number of men at the expense of representations of children. By the opening decades of the twentieth century men have achieved their pre-nineteenth-century levels. Women by contrast remain largely constant in their presence over time. Whether this is an artifact of this particular collection or represents broader historical trends requires further investigation.
- **Object worlds on the decline.** We see a notable decline of both natural objects and architectural features (the latter can include structural features like walls, fences or houses as well as interior items like desks and tables). We find that this decline is not due to rises in other types of objects but rather a decline in "background items," suggesting a potential shift in illustration style in which entities are foregrounded instead of object worlds.

5 Conclusion

This paper presents the first large-scale study of historical children's book illustrations using a combined computational and citizen science approach. By focusing on the visual dimension of children's literature—a rich but historically under-analyzed aspect of cultural transmission from a computational perspective—we open a new line of inquiry into how childhood, emotion, gender, and narrative settings have been visually represented over time. Our dataset, built through the contributions of nearly a thousand volunteers on 27,901 digitized images, demonstrates that citizen science can successfully support the annotation of historical multimodal data. The resulting corpus contains structured metadata with generally high inter-annotator agreement, showing that even subjective and interpretive tasks such as emotional tone or object recognition can yield consistent labels at scale.

Leveraging this dataset, we surface key insights about the visual language of childhood storytelling in the 19th century: the predominance of patriarchal figures, animals, and the natural world; the pairing of boys and girls in spaces of their own alongside gendered differences in the emotional framing of children's lives; the fall and rise of male adults; and finally the gradual decline of attention to architectural backgrounds in illustrative practices.

Children's book illustrations have played a prominent role in shaping and manifesting cultural beliefs. Our findings can serve as an initial foundation for understanding the long history of the

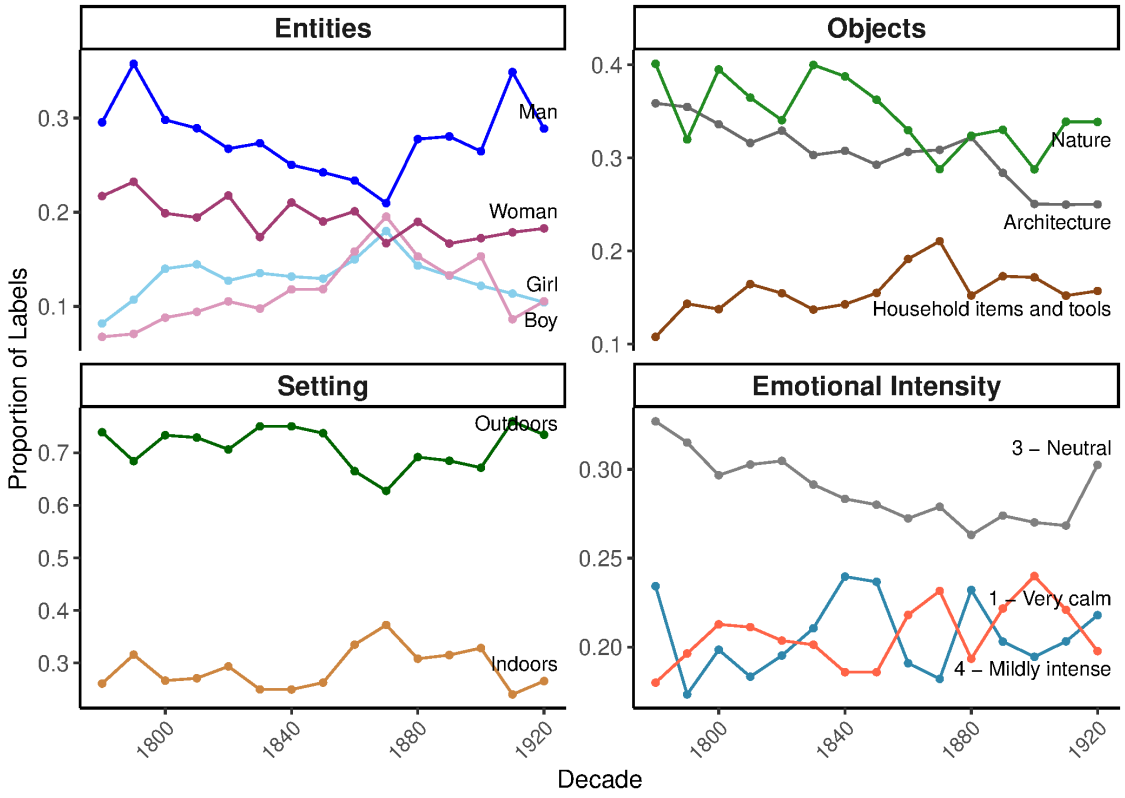

Figure 3: Temporal trends in the visual depiction of people, settings, objects, and emotional tone in 19th-century children's book illustrations. Each panel shows the changing proportion of the 3-4 most frequent labels in our four annotation categories. Values are normalized by decade to account for sample size variation.

visual representation of childhood via a genre that assumes major cultural significance around the world. At the same time, this data can be used to train and validate multi-modal large language models to apply to novel datasets to further shore up our understanding of the past.

Acknowledgements

We would like to acknowledge the generous funding of the Social Sciences and Humanities Research Council of Canada to support this research.

References

[1] Ariès, Philippe. *Centuries of Childhood: A Social History of Family Life*. Trans. by Robert Baldick. Translated from the French *L'Enfant et la vie familiale sous l'ancien régime*. New York: Alfred A. Knopf, 1962.

[2] Arnold, Taylor and Tilton, Lauren. "Automated Image Color Mapping for a Historic Photographic Collection". In: *Proceedings of the Computational Humanities Research Conference (CHR)*. Vol. 3834. Aarhus, Denmark: CEUR Workshop Proceedings, 2024, pp. 37–47. URL: `https://ceur-ws.org/Vol-3834/`.

[3] Arnold, Taylor and Tilton, Lauren. “Explainable Search and Discovery of Visual Cultural Heritage Collections with Multimodal Large Language Models”. In: *Proceedings of the Computational Humanities Research Conference*, ed. by Wouter Haverals, Marijn Koolen, and Laure Thompson. Vol. 3834. CEUR Workshop Proceedings. 2024, pp. 559–574.

[4] Bai, Zechen, Nakashima, Yuta, and Garcia, Noa. “Explain me the painting: Multi-topic knowledgeable art description generation”. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 5422–5432.

[5] Berne, Debbie. *The design of books: An explainer for authors, editors, agents, and other curious readers*. University of Chicago Press, 2024.

[6] Brown, Penny. “Capturing (and Captivating) Childhood: The Role of Illustrations in Eighteenth-Century Children’s Books in Britain and France.” In: *Journal for Eighteenth-Century Studies 31*, no. 3 (2008).

[7] Crain, Patricia. *Reading Children: Literacy, Property, and the Dilemmas of Childhood in Nineteenth-Century America*. University of Pennsylvania Press, 2016.

[8] Cunningham, Hugh. *Children and childhood in western society since 1500*. Routledge, 2020.

[9] Fyfe, Paul and Ge, Qian. “Image Analytics and the Nineteenth-Century Illustrated Newspaper”. In: *Journal of Cultural Analytics 3*, no. 1 (2018). DOI: 10.22148/16.036. URL: https://doi.org/10.22148/16.036.

[10] Garcia, Noa and Vogiatzis, George. “How to read paintings: semantic art understanding with multi-modal retrieval”. In: *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. 2018, pp. 0–0.

[11] Gittins, Diana. “The historical construction of childhood”. In: *An introduction to childhood studies 2* (2009), pp. 35–49.

[12] Grenby, Matthew Orville and Immel, Andrea. *The Cambridge companion to children’s literature*. Cambridge University Press, 2009.

[13] Jenkins, Henry. *The children’s culture reader*. NYU Press, 1998.

[14] Kincaid, James R. “Producing erotic children”. In: *Human, All Too Human*. Routledge, 2013, pp. 203–219.

[15] Kümmerling-Meibauer, Bettina. “Picturebooks”. In: *The Routledge Companion to Children’s Literature and Culture*. Routledge, 2023, pp. 95–105.

[16] Kümmerling-Meibauer, Bettina and Meibauer, Jörg. “Picturebooks and cognitive studies”. In: *The Routledge companion to picturebooks*. Routledge, 2017, pp. 391–400.

[17] Lang, Sarah, Liebl, Bernhard, and Burghardt, Manuel. “Toward a Computational Historiography of Alchemy: Challenges and Obstacles of Object Detection for Historical Illustrations of Mining, Metallurgy and Distillation in 16th–17th Century Print”. In: *Proceedings of the Computational Humanities Research Conference (CHR)*. Vol. 3558. Paris, France: CEUR Workshop Proceedings, 2023, pp. 29–48. URL: https://ceur-ws.org/Vol-3558/.

[18] Lu, Yue, Guo, Chao, Dai, Xingyuan, and Wang, Fei-Yue. “Artcap: A dataset for image captioning of fine art paintings”. In: *IEEE Transactions on Computational Social Systems 11*, no. 1 (2022), pp. 576–587.

[19] Mahowald, Jamie and Lee, Benjamin Charles Germain. “Integrating Visual and Textual Inputs for Searching Large-Scale Map Collections with CLIP”. In: *Proceedings of the Computational Humanities Research Conference (CHR)*. Vol. 3834. Aarhus, Denmark: CEUR Workshop Proceedings, 2024, pp. 528–547. URL: https://ceur-ws.org/Vol-3834/.

[20] Mak, Bonnie. *How the page matters*. University of Toronto Press, 2011.

[21] Maksimova, Erika, Meimer, Mari-Anna, Piirsalu, Mari, and Järv, Priit. “Viability of Zero-shot Classification and Search of Historical Photos”. In: *Proceedings of the Computational Humanities Research Conference*, ed. by Wouter Haverals, Marijn Koolen, and Laure Thompson. Vol. 3834. CEUR Workshop Proceedings. 2024, pp. 1242–1258.

[22] Mohammad, Saif. “Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words”. In: *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers)*. 2018, pp. 174–184.

[23] Montag, Jessica L, Jones, Michael N, and Smith, Linda B. “The words children hear: Picture books and the statistics for language learning”. In: *Psychological science 26*, no. 9 (2015), pp. 1489–1496.

[24] Nikolajeva, Maria. *The rhetoric of character in children’s literature*. Scarecrow Press, 2002.

[25] Nodelman, Perry. *Words about pictures: The narrative art of children’s picture books*. University of Georgia Press, 1988.

[26] Pifer, Ellen. *Demon or Doll: Images of the Child in Contemporary Writing and Culture*. Charlottesville: University Press of Virginia, 2000.

[27] Piper, Andrew, Wellmon, Chad, and Cheriet, Mohamed. “The page image: Towards a visual history of digital documents”. In: *Book History 23*, no. 1 (2020), pp. 365–397.

[28] Rei, Luis, Mladenic, Dunja, Dorozynski, Mareike, Rottensteiner, Franz, Schleider, Thomas, Troncy, Raphaël, Lozano, Jorge Sebastián, and Salvatella, Mar Gaitán. “Multimodal metadata assignment for cultural heritage artifacts”. In: *Multimedia Systems 29*, no. 2 (2023), pp. 847–869.

[29] Reshetnikov, Artem, Marinescu, Maria-Cristina, and Lopez, Joaquim More. “Deart: Dataset of european art”. In: *European conference on computer vision*. Springer. 2022, pp. 218–233.

[30] Roethler, Jacque. “Reading in color: Children’s book illustrations and identity formation for Black children in the United States”. In: *African American Review 32*, no. 1 (1998), pp. 95–105.

[31] Russell, James A. “A circumplex model of affect.” In: *Journal of personality and social psychology 39*, no. 6 (1980), p. 1161.

[32] Stefanini, Matteo, Cornia, Marcella, Baraldi, Lorenzo, Corsini, Massimiliano, and Cucchiara, Rita. “Artpedia: A new visual-semantic dataset with visual and contextual sentences in the artistic domain”. In: *Image Analysis and Processing–ICIAP 2019: 20th International Conference, Trento, Italy, September 9–13, 2019, Proceedings, Part II 20*. Springer. 2019, pp. 729–740.

[33] Thiel, Elizabeth. *The Fantasy of Family: Nineteenth-Century Children’s Literature and the Myth of the Domestic Ideal*. Routledge, 2013.

[34] Tojima, Tatsuya and Yoshida, Mitsuo. “Zero-shot Classification of Art with Large Language Models”. In: *IEEE Access* (2025).

[35] Wang, Zhenlin and Shao, Yihan. “Picture book reading improves children’s learning understanding”. In: *British Journal of Developmental Psychology 43*, no. 1 (2025), pp. 12–35.

A MLLM Prompt

```
You are given a children's book illustration. Analyze the image and answer the following
    5 questions
based on the visual content only. Respond using the exact JSON format provided.

Questions:
1. Is there an illustration on this page? Answer 'Yes' or 'No'.
2. Who are the characters on the page? Choose all that apply from:
   ['Man', 'Woman', 'Girl', 'Boy', 'Animal', 'Animate Object', 'Mythical / Imaginary
    Being', 'Other', 'None']
3. Where is the scene set? Choose one: 'Indoors', 'Outdoors', or 'Both / Ambiguous'
4. What are the main objects in the scene? Choose all that apply from:
   ['Nature', 'Architecture', 'Vehicles', 'Toys', 'Reading and Art', 'Household items and
    tools', 'Other', 'None']
5. How emotionally intense is the scene? Rate from 1 (very calm) to 5 (very intense).

Respond in this exact JSON format (no explanations or extra commentary):

{
  "illustration_present": "",
  "characters": [],
  "scene_setting": "",
  "objects": [],
  "emotional_intensity":
}
```

Listing 1: MLLM Image Annotation Prompt

B Supplementary Figures

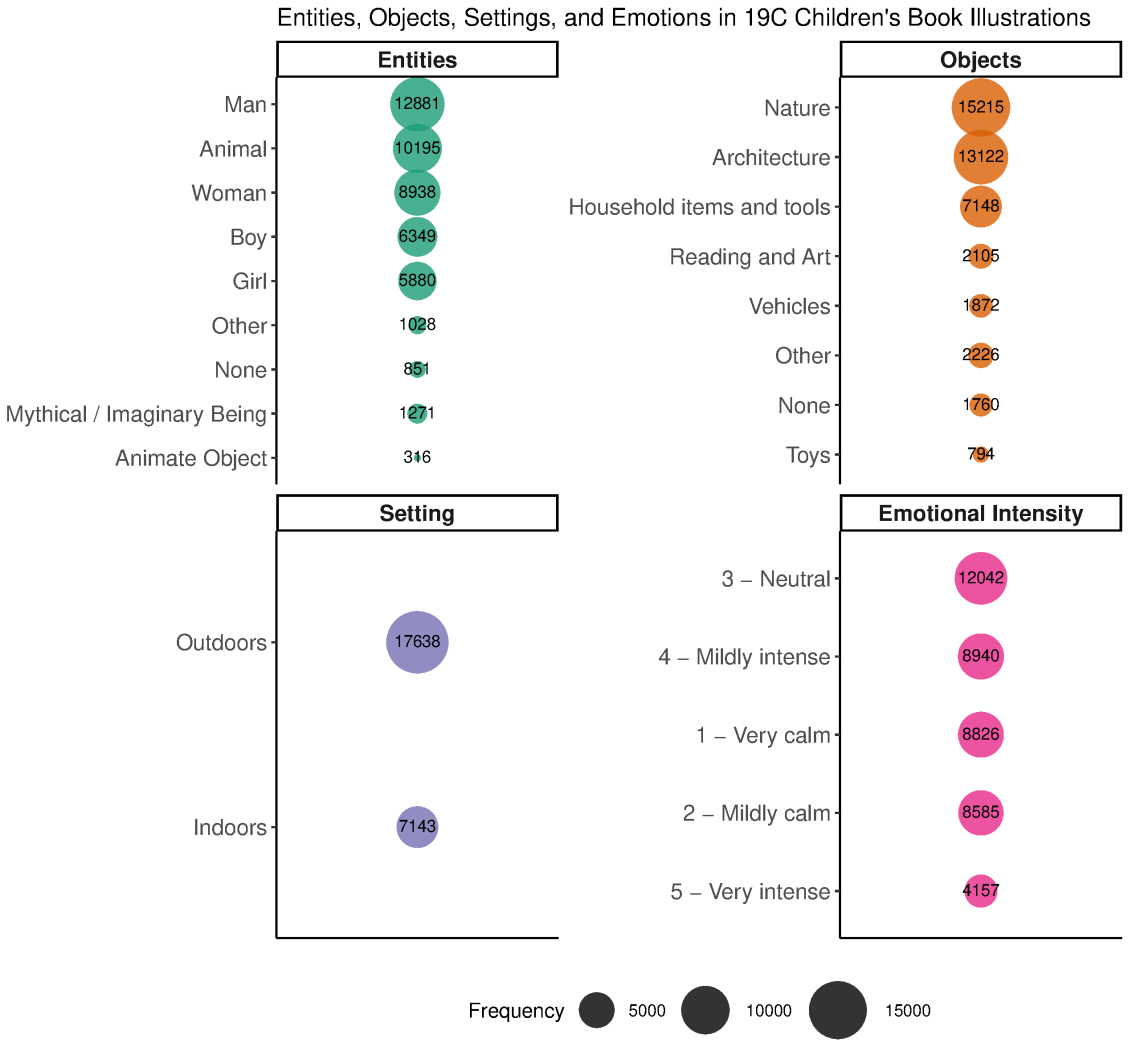

Figure 4: Frequency distributions of our different categories and sub-classes in the Internet Archive Children's Library.

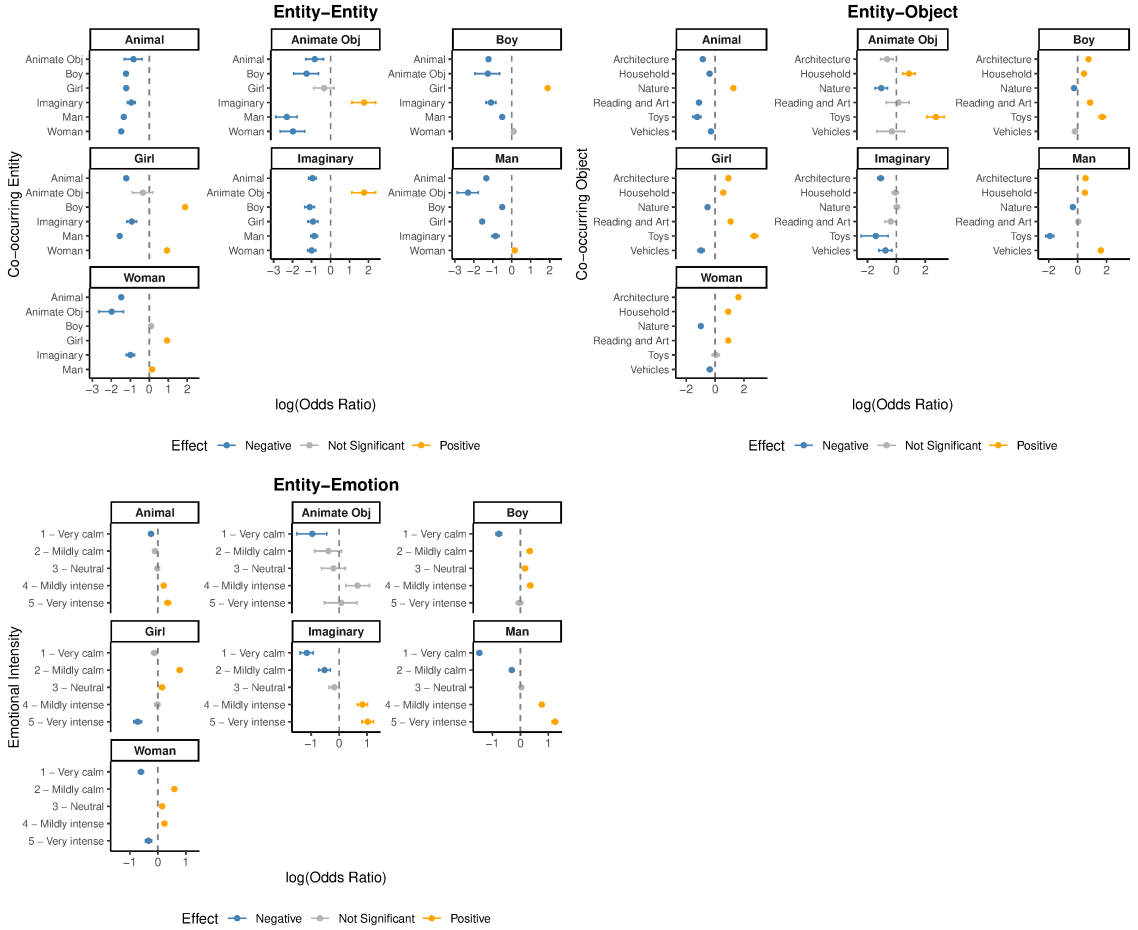

Figure 5: Co-occurrence analysis of illustrated entities in 19th-century children's books. We use Fisher's exact test (with Bonferroni correction) to estimate which types of figures tend to co-occur within the same illustration across three domains: (A) Entities; (B) Entities and Objects; and (C) Entities and Emotions. Each point shows the log odds ratio of co-occurrence with 95% confidence intervals. Colors indicate the direction and significance of the association (positive, negative, or non-significant), and facets group results by focal entity.

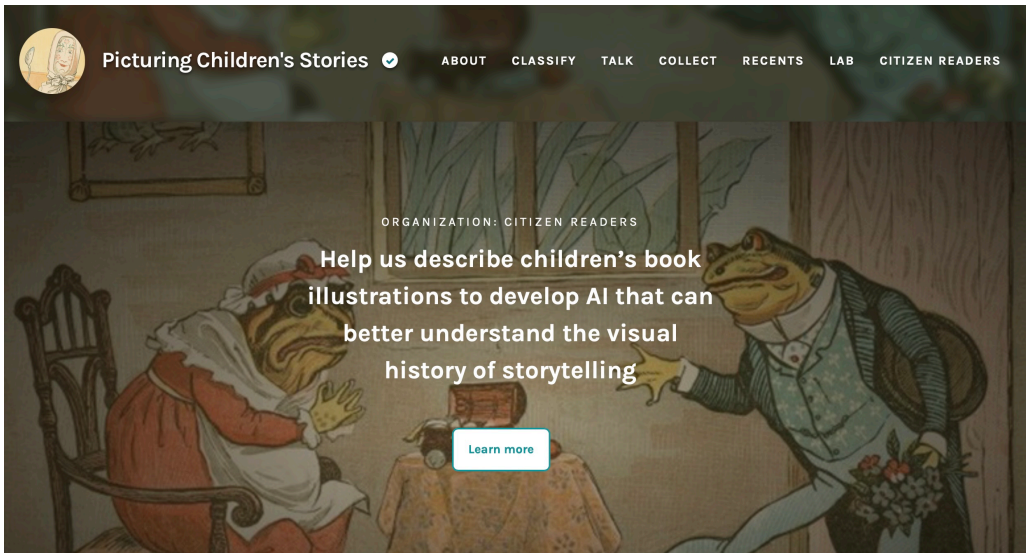

Figure 6: Screenshot of project homepage.

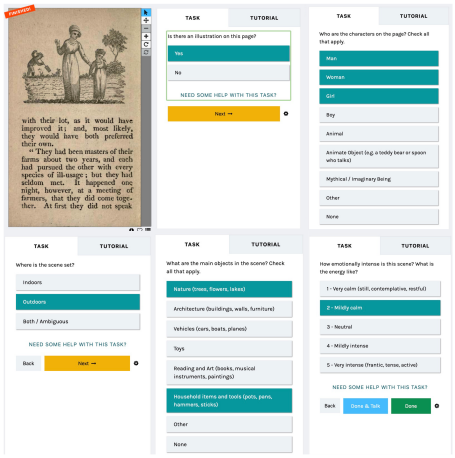

Figure 7: Screenshots of annotation workflow moving left to right beginning in the upper-left.