

Ästhetik des Lesens: Ornament- Erkennung in historischen KJL- Illustrationen mithilfe von Machine Learning

Katharina U.

B.Sc. Informatik

Seminararbeit im Projektseminar “Digitale Analyse von historischen Kinder- und
Jugendbüchern”, Prof. Burghardt, Computational Humanities

Universität Leipzig, Sommersemester 2024

Gliederung

1. Einleitung
 1. Kontextualisierung
 2. Fragestellung und Related Work
 2. Experimentdesign
 1. Bildkorpus
 2. Vorstellung der Methodik
 3. Ergebnisse und Diskussion
 1. Erkenntnisse zu Prompting
 2. Auswertung der Annotationen
 4. Fazit
 5. Anhang
-

1. Einleitung

1.1 Kontextualisierung

Die vorliegende Arbeit findet im Kontext des Seminars "Digitale Analyse von historischen Kinder- und Jugendbüchern" (Universität Leipzig, Sommersemester 2024) im Rahmen des Forschungsprojekts „„Buchkindheiten“ – Innovationspotenziale skalierbarer Bildanalyseverfahren für Wissenskulturen des Kinder- und Jugendbuchs im „langen 19. Jahrhundert““ (Prof. Dr. Burghardt, Dr. Riegler, Dr. Schmideler) statt. Wie auch das Seminar bewegt sich diese Arbeit an der Schnittstelle zwischen historischer Kinderbuchforschung und den Methoden der Digital Humanities, hier vor allem Bildanalyseverfahren, welche sich aus den Möglichkeiten von Machine Learning ergeben.

Die Forschung zu historischer Kinder- und Jugendliteratur (KJL) beschäftigt sich zum einen mit Fragen, welche spezifisch die Gattung KJL oder Subgattungen wie etwa Märchen, Abenteuerliteratur oder Lehrbücher betreffen¹. Beispielsweise wird in Bezug auf das lange 19. Jahrhundert der Zusammenhang zu Ideen der Aufklärung und Romantik beleuchtet.² Zum anderen existieren zahlreiche fachübergreifende Fragestellungen wie bspw. der Zusammenhang zu historischen Konzepten von Kindheit und Spiel³, Fragen zu Aufbau und Funktionsweise des KJL-Verlagswesens (bspw.

1 Helm, W. (05.06.2024). *Historische Kinder- und Jugendliteratur und ihre digitale Analyse*. Präsentation im Rahmen des Seminars "Digitale Analyse von historischen Kinder- und Jugendbüchern", Universität Leipzig, Sommersemester 2024.

2 Wild, R. (2008). Aufklärung. In R. Wild (Ed.), *Geschichte der deutschen Kinder- und Jugendliteratur* (pp. 43ff). Stuttgart, Weimar: Metzler.

sowie

Ewers, H.-H. (2008). Romantik. In R. Wild (Ed.), *Geschichte der deutschen Kinder- und Jugendliteratur* (pp. 96ff). Stuttgart, Weimar: Metzler.

Aufteilung nach Konfession⁴, Wiederverwendung von Illustrationen⁵) oder auch zu Geschlechterrollen in historischer KJL.

Das Einbeziehen von computergestützten Methoden eröffnet neue Forschungsfelder im Diskurs der KJL Forschung, der bisher eher auf traditionelle philologische Methoden fokussiert war⁶. Dieser Einfluss ist vor allem der Zusammenarbeit mit den Digital Humanities zu verdanken, welche durch den Einsatz von Technik neue Forschungsfragen stellen und beantworten konnte.⁷ Aber auch die fortschreitende Digitalisierung von KJL-Beständen in Bibliotheken seit ca. 2006 macht diese Forschung erst möglich.

Neben den digitalen Methoden zur Erforschung des Texts (bspw. Sentiment Analysis in KJL⁸) oder von Text-Bild-Bezügen (z.B. Untersuchung von Bildunterschriften) entstanden auch verschiedene digitale Ansätze, welche sich auf die massenhafte, computergestützte Analyse von Bildern konzentrieren, bspw. um darin Muster zu erkennen ("Distant Viewing")⁹. Hierbei finden sich verschiedene Möglichkeiten der Bildanalyse: Segmentation, Object Detection, Object classification, Image classification, Face Recognition, oder Emotion Detection¹⁰. Auch auf historische Illustrationen, wie sie sich etwa in der KJL-Forschung finden, lässt sich dieser Ansatz gewinnbringend übertragen¹¹, wenn er auch noch etwas zaghaft genutzt wird¹².

3 Schmideler, S. (17.04.2024). PROJEKTSEMINAR „BUCHKINDHEITEN DIGITAL“ - Einführung in die KJL des 18. und 19. Jhs." Präsentation im Rahmen des Seminars "Digitale Analyse von historischen Kinder- und Jugendbüchern", Universität Leipzig, Sommersemester 2024.

Sowie

Burghardt, M. (03.04.2024). *Digitale Analyse von historischen Kinder- und Jugendbüchern*. Präsentation im Rahmen des Seminars "Digitale Analyse von historischen Kinder- und Jugendbüchern", Universität Leipzig, Sommersemester 2024.

4 Lauer, G. (08.05.2024). *Kinderbuch-Forschung digital*. Präsentation im Rahmen des Seminars "Digitale Analyse von historischen Kinder- und Jugendbüchern", Universität Leipzig, Sommersemester 2024.

5 Siehe bspw.

Helm, W., Schmideler, S., Im, C., Mandl, T., Kollmann, S., & Müller, L. (2021). Wie sich die Bilder ähneln - Vom Zufallsfund zur systematischen Forschung im Bereich der automatisierten Bildähnlichkeitssuche. In M. Burghardt, L. Dieckmann, T. Steyer, P. Trilcke, N.-O. Walkowski, J. Weis, & U. Wuttke (Eds.), *Fabrikation von Erkenntnis: Experimente in den Digital Humanities* (Vol. 1). Melusina Press. <https://doi.org/10.17175/sb005>

6 Schmideler, S., & Helm, W. (Hrsg.). (2021). *BildWissen – KinderBuch: Historische Sachliteratur für Kinder und Jugendliche und ihre digitale Analyse* (S. 2). Stuttgart: Franz Steiner Verlag.

7 Blumesberger, S. (2021). Kinder- und Jugendliteraturforschung auf dem Weg zu Digitalisierung und Open Science. In S. Schmideler & W. Helm (Hrsg.), *BildWissen – KinderBuch: Historische Sachliteratur für Kinder und Jugendliche und ihre digitale Analyse* (S. 256 f.). Stuttgart: Franz Steiner Verlag.

8 Rebora, S., Lehmann, M., Heumann, A., Ding, W., & Lauer, G. (2023). Comparing ChatGPT to Human Raters and Sentiment Analysis Tools for German Children's Literature. In *Computational Humanities Research Conference (CHR 2023)*, December 6–8, 2023, Paris, France (S. 1–12). abgerufen unter: <https://ceur-ws.org/Vol-3558/paper3340.pdf>

9 Siehe bspw.: Arnold, T., & Tilton, L. (2019). *Distant viewing: Analyzing large visual corpora*. *Digital Scholarship in the Humanities*, 34(1), S. 176–192. <https://doi.org/10.1093/llc/fqy050>

10 Borst, J. (10.04.2024). *Digitale Analyse von historischen Kinder- und Jugendbüchern*. Präsentation im Rahmen des Seminars "Digitale Analyse von historischen Kinder- und Jugendbüchern," Universität Leipzig, Sommersemester 2024.

11 Siehe dazu bspw.: Wevers, M., & Smits, T. (2020). *The visual digital turn: Using neural networks to study historical images*. *Digital Scholarship in the Humanities*, 35(1), S. 194–207. <https://doi.org/10.1093/llc/fqy085>

12 Im, C., Helm, W., Mandl, T., & Schmideler, S. (2019). Herausforderungen für die Klassifikation historischer Buchillustrationen: Überlegungen am Beispiel retrdigitalisierter Kinder- und Jugendsachbücher des 19. Jahrhunderts. In 6. Jahrestagung „Digital Humanities im deutschsprachigen Raum (Dhd)“, 25.–29. März 2019, Frankfurt-Mainz (S. 300).

Vor diesem Hintergrund stellt sich die Frage, welche Chancen, aber auch welche Grenzen¹³ aktuelle computergestützte Methoden im Hinblick auf KJL-Forschungsfelder wie die Bildanalyse haben. Die vorliegende Arbeit soll ein Experiment sein, das diese Chancen und Grenzen ein Stück näher auslotet.

1.2 Fragestellung und Related Work

Dabei möchte ich mich besonders der Ornamentik in historischen KJL-Illustrationen widmen. Meine Arbeit bewegt sich also an den Schnittstellen zwischen historischer Kinderbuchforschung, Digital Humanities sowie Fragen der Kunst- bzw. Buchwissenschaft, welche sich mit ornamentalen Verzierungen beschäftigt. Ornamente dienen unter anderem als Schmuckelemente der Buchgestaltung, welche im 19. Jahrhundert und beginnenden 20. Jahrhundert von Verlagen häufig aus sogenannten Buchschmuckkatalogen bzw. Musterbüchern bezogen wurden¹⁴. Sie können dabei Funktionen wie z.B. Rahmung, Verzierung oder Gliederung einnehmen. Laut Wiebke Helm klafft hier in Bezug auf die Kinderbuchgestaltung noch eine große Forschungslücke.¹⁵

Das Ornament im engeren Sinne ist abzugrenzen von Randzeichnung, Arabeske, Vignette und Emblem.¹⁶ Die Abgrenzung zu diesen Schmuckelementen würde jedoch sowohl theoretisch als auch bzgl. der Auswertung der Qualität computergestützter Annotationen den Rahmen dieser Arbeit sprengen, weswegen ich mich im Folgenden einer weiteren Definition des Ornaments bediene. Diese Definition stellt vor allem die dekorative Funktion des Ornaments in den Mittelpunkt, also die Abgrenzung zwischen Hauptinhalt des Bildes und Verzierung:

“ornament: a generic term for the various kinds of decoration which compositors may use with type, e.g. arabesques, borders, flowers, head-pieces, rules, qq.v.”¹⁷

Diese Definition passe ich in meiner Arbeit an, um im Prompting für das Machine Learning Modell die Doppeldeutigkeit der Begriffe compositor und type zu vermeiden, welche außerhalb der Buchillustration weitere Bedeutungen haben. Ich verwende daher im Prompting die folgende Definition:

“Ornament is a generic term for the various kinds of decorative elements used in book illustration, e.g. arabesques, borders, flowers, head-pieces, rules, etc.”

¹³ Schmideler, S., & Helm, W. (Hrsg.). (2021). *BildWissen – KinderBuch: Historische Sachliteratur für Kinder und Jugendliche und ihre digitale Analyse* (S. 3). Stuttgart: Franz Steiner Verlag.

¹⁴ Siehe etwa: Schneider-Henn, D. (1997). *Ornament und Dekoration: Vorlagenwerke und Motivsammlungen des 19. und 20. Jahrhunderts*. München: Callwey.

¹⁵ Email-Kommunikation mit Wiebke Helm (25.07.2024) im Rahmen des Seminars “Digitale Analyse von historischen Kinder- und Jugendbüchern” (Universität Leipzig, Sommersemester 2024)

¹⁶ Kreidt, U. (2008). Bilder in der Kinder- und Jugendliteratur. In O. Brunken, B. Hurrelmann, M. Michels-Kohlhage, & G. Wilkending (Hrsg.), *Handbuch zur Kinder- und Jugendliteratur. Von 1800 bis 1850* (S. 95ff, Spalten 134-136 und 139-144). Stuttgart: J.B. Metzler.

¹⁷ Glaister, G. A. (2001). *Encyclopedia of the Book* (S. 353). New Castle: Oak Knoll Press.

Die Inspiration für dieses Thema lieferte das Projekt „Entwicklung der Bildikonographie in Wissen vermittelnder Kinder- und Jugendliteratur und Schullehrbüchern des 19. Jahrhunderts. Ein Distant Viewing Ansatz“ (2017-2020)¹⁸. Hierbei konnte unter anderem die Ornament-Erkennung als Herausforderung in der automatisierten Bildanalyse identifiziert werden:

“Bei der automatischen Bilderkennung von Buchillustrationen ist der ornamentale Schmuck, der die Ränder der Buchseiten ziert oder die Abbildungen rahmt, ebenfalls zu berücksichtigen, da auf diese Weise nicht nur Aussagen zur Verbreitung von Bildmotiven erfasst werden. Auch für Fragen zur Stilistik von Bildern erweist sich dieses Element als relevant.”¹⁹

In der vorliegenden Arbeit will ich deshalb folgende Frage beantworten: Wie gut ist das Modell GPT 4o in der Lage, rahmende Ornamente in historischen KJL-Illustrationen zu erkennen und nach Inhalten zu klassifizieren?

Dabei weist meine Arbeit gewisse Parallelen zum Paper “Wie sich die Bilder ähneln - Vom Zufallsfund zur systematischen Forschung im Bereich der automatisierten Bildähnlichkeitssuche”²⁰ auf, welches ebenfalls mithilfe eines Distant Viewing Ansatzes Ähnlichkeiten in der Illustration von KJL untersucht. Es werden dabei jedoch Convolutional Neural Networks (CNNs – hier ResNet-101) betrachtet, ein Typ künstlicher neuronaler Netze, der besonders effektiv bei Aufgaben wie der Bildklassifikation, Objekterkennung und Bildsegmentierung ist. Dagegen sind CNNs nicht darauf ausgelegt, Text zu verstehen (z.B. Prompting) oder zu generieren. Ich teste in meinem Experiment stattdessen GPT 4o als Transformer-Modell (siehe 3.2 Vorstellung Methodik). Zudem fokussiere ich mich statt auf das Hauptmotiv auf Ähnlichkeiten in der Ornamentik. Um eine tatsächliche Wiederverwendung von Ornamenten (analog zu Wiederverwendung von Illustrationen, wie im o.g. Paper), bspw. aus Musterbüchern nachzuweisen (und ggf. in einem späteren Schritt festzustellen, zwischen welchen Verlagen, in welchen Zeiträumen etc die Wiederverwendung stattfand) bedürfte es jedoch einer noch genaueren Untersuchung der annotierten Bilder innerhalb der von mir festgelegten Labels (bspw. Suche nach Ähnlichkeiten innerhalb der Kategorie

18 Siehe <https://staatsbibliothek-berlin.de/die-staatsbibliothek/projekte/project-id-1112-2017>

19 Im, C., Helm, W., Mandl, T., & Schmideler, S. (2019). Herausforderungen für die Klassifikation historischer Buchillustrationen: Überlegungen am Beispiel retrodigitalisierter Kinder- und Jugendsachbücher des 19. Jahrhunderts. In 6. Jahrestagung „Digital Humanities im deutschsprachigen Raum (DHd)“, Frankfurt-Mainz, 25.-29. März 2019 (S. 302).

20 Helm, W., Schmideler, S., Im, C., Mandl, T., Kollmann, S., & Müller, L. (2021). Wie sich die Bilder ähneln - Vom Zufallsfund zur systematischen Forschung im Bereich der automatisierten Bildähnlichkeitssuche. In M. Burghardt, L. Dieckmann, T. Steyer, P. Trilcke, N.-O. Walkowski, J. Weis, & U. Wuttke (Hrsg.), *Fabrikation von Erkenntnis: Experimente in den Digital Humanities*. Melusina Press, Luxemburg. Online verfügbar unter <https://www.melusinapress.lu/read/melusina-8f8w-y749-wsdb-2/section/98bf793b-2978-42b1-8f62-494fa09f977c>

“Ornamente mit Pflanzenbestandteilen”). Meine Fragestellung stellt deshalb hier nur einen kleinen Schritt in eine ähnliche Richtung dar.

Auch die Arbeit “Herausforderungen für die Klassifikation historischer Buchillustrationen: Überlegungen am Beispiel retrodigitalisierter Kinder- und Jugendsachbücher des 19. Jahrhunderts”²¹ liefert interessante Einblicke in ein verwandtes Vorhaben, insbesondere in Bezug auf die Entwicklung eines Klassifikationssystems für historische Illustrationen, und welche Herausforderungen damit verbunden sind.

Von Bedeutung ist ebenfalls die Arbeit “Bildsuche: Erfahrungen zur Erkennung von Emblemen und zur automatischen Annotation von Segmenten”²², welche sich statt Ornamenten auf Embleme – also ebenfalls dekorative Elemente – fokussiert. Jedoch werden hier ebenfalls CNNs (YOLO) statt Transformers verwendet.

2. Experimentdesign

2.1 Bildkorpus

Der verwendete Bildkorpus besteht aus einem Teil des im Rahmen des Projektseminars zur Verfügung gestellten Korpus an Illustrationen aus KJL des langen 19. Jahrhunderts (Zeitraum 1801 bis 1914) aus dem Korpus Colibri, einer Sammlung von historischen Kinder- und Jugendbüchern aus den Beständen deutscher Forschungsbibliotheken.

Die Auswahl des Bildkorpus erfolgte anhand folgender Kriterien: Es muss, außer dem ggf. vorhandenen Ornament, eine Illustration mit einem Hauptmotiv vorhanden sein. Falls ein Ornament vorhanden ist, muss es sich um einen Rahmen handeln, welchen ich definiert habe als ein Ornament, das das Hauptmotiv an mindestens drei Seiten umgibt. Dabei müssen nicht alle Ornamentarten (bspw. Pflanzenteile, geometrische Formen) auf allen drei Seiten vertreten sein.

Ich habe die verwendeten Illustrationen so ausgewählt, dass mindestens 10 Illustrationen pro existierendem Label im Korpus enthalten sind, um die Ergebnisse der Annotationen für jedes Label aussagekräftig vergleichen zu können (Ornamente mit mehreren Labels werden für jedes Label mitgezählt). Eine Ausnahme bildet das Label “Tiere”, welches letztlich nur in vier Illustrationen annotiert werden konnte.

Teilweise habe ich jedoch auch mehr Illustrationen pro Label annotiert (bspw. 30 Illustrationen im Label “Pflanzenteile”), da besonders interessante Kombinationen an Ornamenten bzw. Herausforderungen enthalten waren, von denen ich gern wissen wollte, wie das Modell sie annotiert.

21 Im, C., Helm, W., Mandl, T., & Schmideler, S. (2019). Herausforderungen für die Klassifikation historischer Buchillustrationen: Überlegungen am Beispiel retrodigitalisierter Kinder- und Jugendsachbücher des 19. Jahrhunderts. In 6. Jahrestagung „Digital Humanities im deutschsprachigen Raum (DHd)“, Frankfurt-Mainz, 25.-29. März 2019 (S. 300-304).

22 Henrich, A., Bullin, M., & Engl, F. (2021). Bildsuche: Erfahrungen zur Erkennung von Emblemen und zur automatischen Annotation von Segmenten. In S. Schmideler & W. Helm (Hrsg.), *BildWissen – KinderBuch: Historische Sachliteratur für Kinder und Jugendliche und ihre digitale Analyse* (S. 115-135). Stuttgart: Franz Steiner Verlag.

Dabei habe ich jedoch allzu komplexe Mischformen und unklare Fälle weggelassen, wie z.B. das folgende Bild, bei dem aus meiner Sicht das Hauptmotiv schwer zu erkennen ist (sowohl die Blätter/Federn könnten als eine Art loser Rahmen interpretiert werden, aber auch das Sternenmuster rund um das Kreuzmotiv, zusätzlich könnte das Kreuzmotiv aber auch selbst als Teil eines Ornaments betrachtet werden, das den Text umrahmt).

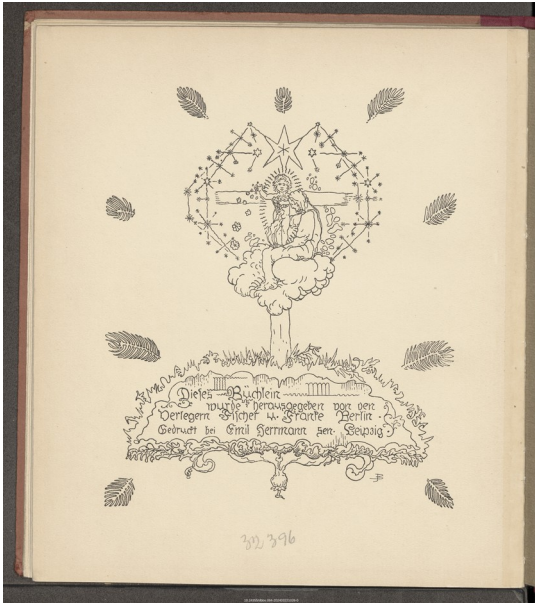


Abbildung 1. Originaldateiname: dbbs_derivate_00055179-images-00000024.png

Insgesamt habe ich 88 Illustrationen annotiert, von denen ich aufgrund der maximalen Dateigröße von GPT 4o später 85 automatisiert annotieren konnte.

2.2 Vorstellung der Methodik

Ich habe mich für die Verwendung von Category Sets statt eines Open Vocabulary Ansatzes entschieden, um eine einfachere Vergleichbarkeit zwischen meiner Annotation und den Ergebnissen des Modells sicherzustellen. Hierbei habe ich folgendes Label Set induktiv aus dem Bildkorpus des Projektseminars entwickelt:

Label Set 1: Inhalt des ornamentalen Rahmens

- keiner
- Pflanzenteile
- Tiere
- abstrakte geometrische Formen
- einfache Linie
- Banner/Bogen
- Buchstabe / Text
- sonstige Motive

Ein ursprünglich geplantes zweites Label Set (*Verhältnis zwischen Ornament und Bild*) habe ich verworfen, um die Komplexität der Aufgabe einzugrenzen – das Modell sollte sich nicht im ersten Schritt der Ornamenterkennung “verlieren”, sondern aussagekräftige Ergebnisse liefern, inwiefern die Inhalte des Ornaments erkannt werden können. Zudem wollte ich falsche Annotationen durch das Modell besser nachvollziehen können. Daher habe ich mich hier ausschließlich auf rahmende Ornamente beschränkt.

Meine Annotation der Illustrationen habe ich im ImageNet Format vorgenommen: Ich habe die Bilder aus meiner Auswahl also in Ordner mit den entsprechenden Labels sortiert und diese Annotation anschließend per Skript in eine Textdatei im ImageNet Format überführt. Anschließend habe ich mittels eines ausführlichen Prompts jeweils bis zu 10 Illustrationen durch GPT 4o annotieren und ebenfalls als ImageNet Textdatei ausgeben lassen. Ich habe die Bilder also in 9 Teilen bearbeitet. Die einzelnen Annotationsdateien habe ich anschließend ebenfalls durch Skripte zusammengeführt und mit meiner manuellen Annotationsdatei verglichen.

Meine anfängliche Auswahl von CLIP (Contrastive Language-Image Pre-Training) als Machine Learning Modell habe ich vor allem deshalb getroffen, weil es sich um ein Open Source Modell handelt und CLIP durch zero-shot classification und die Spezialisierung auf Image-Text Matching (Object Classification) gut geeignet erschien. Ich habe dazu ein Jupyter Notebook mit der Huggingface API (PyTorch library, Hugging Face Transformers library) genutzt, um ein vortrainiertes CLIP Modell für die Annotationen zu verwenden. Allerdings hat bei meinen Versuchen CLIP die Illustrationen häufig anhand des Hauptmotivs getaggt statt anhand des Ornaments – dies leider auch bei recht eindeutigen Illustrationen. Evtl. wäre daher für nennenswerte Ergebnisse bzgl. des Ornamentinhalts eine vorherige (manuelle oder maschinelle) Object Detection (Segmentation Annotation) notwendig gewesen, sodass sich CLIP anschließend nur auf das Ornament fokussieren kann. Dies hätte jedoch in Kombination mit dem eigentlichen Experiment – der Annotation der Ornamente – den Rahmen der Seminararbeit gesprengt. Auch wenn es sich um ein Sprachmodell handelt, das auf Textverarbeitung spezialisiert ist, bin ich daher zum Modell GPT 4o (Generative Pre-trained Transformer 4) gewechselt, sodass ich durch komplexeres Prompting bessere Ergebnisse bei der Ornamenterkennung erzielen konnte.

3. Ergebnisse und Diskussion

3.1 Erkenntnisse zu Prompting:

Den Prompt für die eigentliche Annotation durch GPT 4o habe ich schrittweise entwickelt. Ausgehend von Prompt 1 (siehe Anhang) habe ich festgestellt, dass teilweise durch gezielte Nachfrage Verbesserungen erzielt werden können:

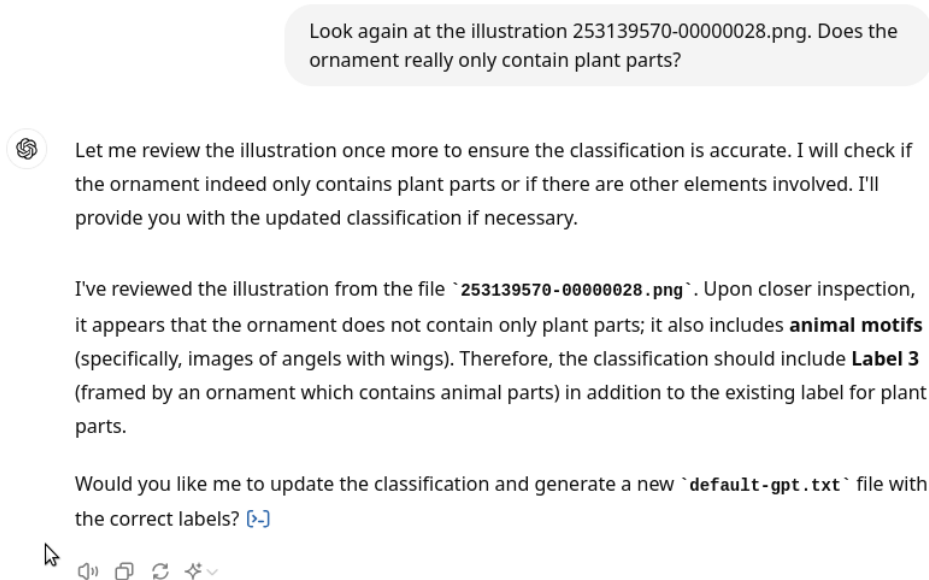


Abbildung 2.

Deshalb habe ich folgenden Satz hinzugefügt: "Please be as accurate as possible and inspect all ornaments closely." (Prompt 2, siehe Anhang). Dies führte nicht zu einer Verbesserung. Daraufhin habe ich versucht, nur eine Illustration auf einmal zu annotieren (Prompt 2a, siehe Anhang). Dies hat die Annotation deutlich verbessert: Statt nur dem Label "plants" wurden nun zusätzlich "animals" und "geometric shapes" zugeordnet.

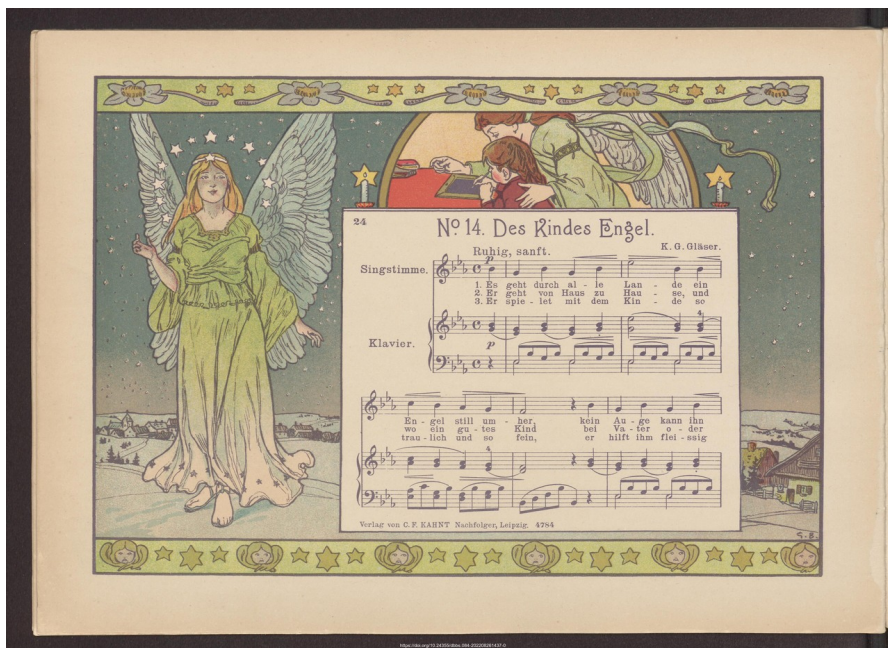


Abbildung 3. Originaldateiname: 253139570-00000028.png

Aber es blieb eine Frage offen:

Are angels animals?



No, angels are not considered animals. Angels are typically depicted as spiritual or supernatural beings in various religious traditions, rather than as living creatures that belong to the animal kingdom. In the context of labeling illustrations, angels would not be categorized under "animal parts." They would instead fall under other categories depending on the specific depiction and context.

< 2/2 > 🔊 📄 ↺ ⚙️

Abbildung 4.

Ich habe aus diesem Grund dem Prompt zusätzlich den folgenden Teil hinzugefügt: "Take only one illustration at a time and analyze it thoroughly as if I made this prompt only for this single illustration. Only after this, move to the next illustration, and do the same." (Prompt 3). Dies schien zu einer Verbesserung zu führen: Es wurden für die obige Illustration nun "plants", "geometric shapes", "banner/arch" sowie "other" ausgewählt. Engel werden nun – vermutlich aufgrund meiner Nachfrage – nicht mehr als Tiere gewertet; zusätzlich wird vermutlich der ornamentale Streifen als Banner eingeordnet. Da es sich gegenüber Prompt 2 um eine Verbesserung handelte, habe ich den Prompt 3 für die weiteren Annotationen verwendet.

Eine weitere Erkenntnis aus dem Prompting ist, dass bei vielen aufeinanderfolgenden Anfragen mit Bildern Vorsicht geboten ist: Teilweise bringt das Modell die Bilder durcheinander (Zusammenfügen der Ergebnisdateien aus zwei verschiedenen Prompts)

oder verwechselt die Dateinamen der geforderten Ausgabedatei mit denen vorheriger Prompts. In einem Fall hat GPT 4o auch die Bilddateien in der Ausgabedatei mit neuen, selbst festgelegten Dateinamen versehen. Hier war deshalb stets eine manuelle Kontrolle bzw. Bitten um Nachbesserung nötig. Teilweise wirkte es auf mich so, als würden die Ergebnisse besser, wenn man sich zwischendurch bedankt bzw. die Prompts durch "natürlichere" Interaktionen voneinander trennt. Bspw. wurden dem folgenden Bild im normalen Prompt 3 alle existierenden Labels zugeordnet. Nach einem "thanks" mit erneutem Prompt 3 erhielt es die Labels "plants", "other", "banner/arch" und "letter/word". Nach einem "thanks a lot" mit erneutem Prompt 3 erhielt es zusätzlich noch das Label "animals" (was sich auf das Holzpferd in den beiden Bogen beziehen könnte, oder aber auf das Hauptmotiv). Hier ist allerdings nicht ganz klar, ob die Änderungen der Labels durch das Bedanken oder durch die mehrfache Wiederholung zustande kamen.



Abbildung 5. Originaldateiname: PPN1771733616-00000009.png

3.2 Auswertung der Annotationen

Das Gesamtergebnis bezüglich der Annotationen des Ornamentinhalts sieht eher schlecht aus: vollständig identisch wurden nur 6 von 85 Illustrationen annotiert. Diese 6 Annotationen bezogen sich alle auf die Labels "none" oder "simple line".

Übereinstimmung mit meiner Annotation	Anzahl
100% (vollständig richtig)	6
100% (und zusätzl. falsche Annotationen durch GPT)	18

4o)	
67%	6
50%	15
40%	1
33%	1
0%	39

Abbildung 6.

Die hohe Anzahl in der Kategorie “100% (und zusätzl. falsche Annotationen durch GPT 4o)” bezieht sich darauf, dass Teil 6 und 9 der annotierten Bilder alle Labels zugewiesen bekamen, unabhängig vom Bildinhalt. Dies ließ mich vermuten, dass ggf. die Qualität nach mehreren ähnlichen Prompts mit vielen Bildern sinkt, und es zu Vermischungen kommt (evtl. da die Bilder vorheriger Prompts noch im Speicher des Modells vorhanden sind, zusammen mit identischen Anfragen). Alternativ könnte dieses Verhalten als Phänomen analog zu “Langeweile” betrachtet werden; da in Machine Learning Modellen bereits andere menschenähnliche psychologische Mechanismen (etwa bessere Ergebnisse bei Ankündigung von Belohnungen) beobachtet wurden, wäre dies durchaus eine weitere Untersuchung wert. In diesem Fall wäre es für die Genauigkeit der Ergebnisse ggf. förderlich, gegenüber dem Modell ein “menschlicheres” Dialogverhalten an den Tag zu legen, anstatt immer dieselbe Anfrage mit unterschiedlichen Bildern zu schicken; dies würde allerdings auf Kosten der Vergleichbarkeit gehen. Eine Lösung könnte möglicherweise sein, zufällig erzeugten Smalltalk zwischen die einzelnen Prompts zu schieben.

Bei den teilweise korrekten Annotationen gab es zum Teil verständliche andere Interpretationen zum Ornament. Dies möchte ich anhand ein paar Beispielen veranschaulichen:

Bei der oben bereits gezeigten Abbildung 3 habe ich die Labels “plants” und “other” vergeben, während das Modell auf den Prompt 3 hin die Labels “plants”, “other”, “banner/arch” und “geometric shapes” ausgewählt hat. Es ist hier durchaus möglich, Sterne auch als geometrische Formen zu betrachten, und zumindest verständlich, dass der Ornamentstreifen als Banner gewertet wurde.

Im folgenden Bild wurde von mir dagegen das Label “none” vergeben, von GPT 4o “simple line”, “plants” und “other”.



Abbildung 7. Originaldateiname: 382992202-00000015.png

Hier wurde vermutlich die Definition aus dem Prompt nicht beachtet, dass ein Ornament eine Illustration rahmen soll (nicht einen Text). Ansonsten sind die Labels jedoch verständlich, wenn die gesamte Illustration als Ornament gelesen wird. Hier liegt die Herausforderung also eher darin, das Gesamtlayout der Seite zu verstehen. Ggf. könnten hier durch Anpassen des Prompts (mehr Ausführungen zum Seitenlayout bzw. zum Umgang mit Text) noch Verbesserungen erzielt werden.

Im folgenden, zugegebenermaßen recht komplexen Bild habe ich diese Labels vergeben: "plants", "banner/arch", "geometric shapes", "letter/word". GPT 4o hingegen hat nur "plants" und "banner/arch" zugeordnet.



Abbildung 8. Originaldateiname: BV010596711-00000001.png

Während sich tatsächlich darüber streiten lässt, ob die Textelemente im Ornament tatsächlich zu diesem dazugehören, z.B. weil sie keine in erster Linie dekorative Funktion erfüllen, überrascht es, dass die sehr präsenten abstrakten geometrischen Formen nicht als Ornament erkannt wurden. Für das menschliche Auge erfüllen diese recht eindeutig eine dekorative Funktion.

Die folgende Annotation ist vermutlich an der etwas untypischen Art eines Banners gescheitert: Während “plants” und “letter/word” korrekt gelabelt wurden, wurde “banner/arch” von GPT 4o nicht vergeben:



Abbildung 9. Originaldateiname: dbbs_mods_00068068-00000020.png

Das Banner ist vermutlich etwas untypisch in der Position (unten) sowie in der Form (nicht geschwungen); hier könnte ein vorheriges Einlesen typischer Ornamentbestandteile etwa aus den eingangs genannten Buchschmuckkatalogen interessant sein.

In manchen Fällen bleibt die Annotation des Modells aber auch vollständig mysteriös:



Abbildung 10. Originaldateiname: 578851466-00000001.png

In diesem Fall habe ich das Label "none" vergeben, GPT 4o jedoch "plant", "letter/word", "animal" sowie "banner/arch". Da nicht einmal die Illustration selbst Pflanzen oder Tiere enthält, ist die Zuordnung nicht zu erklären. Die Labels "banner/arch" und "letter/word" könnten aufgrund der Bibliothekssticker vergeben worden sein. Hier wäre eine Ergänzung des Promptings oder aber direkt eine automatisierte Korrektur des Bildes hilfreich.

Dies sollen ein paar Einblicke in konkrete Herausforderungen bei der automatisierten Ornamenterkennung mittels GPT 4o gewesen sein.

4. Fazit

Meine erwartete Herausforderung, dass besonders Mischformen der obigen Labels und doppelte Ornamente ein Problem darstellen könnten, hat sich so nicht bestätigt. Auch Verwechslungen mit Objekten im Bild schienen selten zu sein, wenn man von der Verwechslung des Gesamtmotivs mit einem Ornament einmal absieht. Dies konnte ich allerdings nur stichprobenartig überprüfen. Meine Erwartung, dass das Modell teilweise Bild und Ornament nicht klar voneinander trennen kann, wenn beide ohne Abstand ineinander übergehen, habe ich teils durch die Eingrenzung auf ornamentale Rahmen entkräftet – vor diesem Hintergrund habe ich dann allerdings keine Unterschiede anhand des Abstands zwischen Ornament und Hauptmotiv entdecken können.

Insgesamt ist festzuhalten, dass die Analyse einzelner Illustrationen zufriedenstellend ist, die Analyse vieler Bilder auf einmal mit standardisierter Labelzuordnung in vorgegebenem Dateiformat aber leider eher schlecht funktioniert. Es gibt zudem starke Qualitätsunterschiede zwischen den einzelnen Prompts bzw. unvorhersagbares Verhalten (z.B. eigenmächtiges Ändern des Dateinamens), welche eine manuelle Kontrolle nötig machen.

Dies stellt ein Problem dar, da gerade für Distant Viewing eine Massenverarbeitung notwendig ist – andernfalls ergibt sich keine allzu große Erweiterung des Spektrums an möglichen Forschungsfragen.

Es sind allerdings auch mit dem bestehenden Modell einige Verbesserungsmöglichkeiten denkbar, vor allem durch die oben genannten Ideen für weiteres Anpassen des Prompts, sowie durch die Anreicherung des Prompts mit Katalogen (z.B. Buchschmuckkataloge, Frakturschriftbuchstaben).

5. Anhang

Prompt 1:

“Ornament is a generic term for the various kinds of decorative elements used in book illustration, e.g. arabesques, borders, flowers, head-pieces, rules, etc. Here, I want to focus only on ornaments that frame book illustrations, that is, they can be found on at least three sides of the illustration.

Please look at these images I am uploading. They are all book illustrations.

Then, please tell me if each book illustration contains a framing ornament, and label the illustration with one or more of the following categories. Please choose all labels that fit the illustration. Please be as accurate as possible and inspect all ornaments closely.

- Label 2: this illustration is not framed by an ornament
- label 6: this illustration is framed by an ornament which contains plant parts
- label 3: this illustration is framed by an ornament which contains animal parts
- label 8: this illustration is framed by an ornament which contains abstract geometric shapes
- label 5: this illustration is framed by a simple line
- label 7: this illustration is framed by a banner or arch
- Label 4: this illustration is framed by an ornament which contains letters or words

- label 1: this illustration is framed by an ornament which contains other motives not mentioned above.

Please output the result in the ImageNet format, containing one file called default-GPT-2.txt with the illustration filename and the label numbers assigned to it.”

Prompt 2:

Ornament is a generic term for the various kinds of decorative elements used in book illustration, e.g. arabesques, borders, flowers, head-pieces, rules, etc. Here, I want to focus only on ornaments that frame book illustrations, that is, they can be found on at least three sides of the illustration.

Please look at these images I am uploading. They are all book illustrations.

Then, please tell me if each book illustration contains a framing ornament, and label the illustration with one or more of the following categories. Please choose all labels that fit the illustration. Please be as accurate as possible and inspect all ornaments closely.

- Label 2: this illustration is not framed by an ornament
- label 6: this illustration is framed by an ornament which contains plant parts
- label 3: this illustration is framed by an ornament which contains animal parts
- label 8: this illustration is framed by an ornament which contains abstract geometric shapes
- label 5: this illustration is framed by a simple line
- label 7: this illustration is framed by a banner or arch
- Label 4: this illustration is framed by an ornament which contains letters or words
- label 1: this illustration is framed by an ornament which contains other motives not mentioned above.

Please output the result in the ImageNet format, containing one file called default-GPT-1_prompt2.txt with the illustration filename and the label numbers assigned to it.

Prompt 2a:

Ornament is a generic term for the various kinds of decorative elements used in book illustration, e.g. arabesques, borders, flowers, head-pieces, rules, etc. Here, I want to focus only on ornaments that frame book illustrations, that is, they can be found on at least three sides of the illustration.

Please look at the image I am uploading. It is a book illustration.

Then, please tell me if the book illustration contains a framing ornament, and label the illustration with one or more of the following categories. Please choose all labels that fit the illustration. Please be as accurate as possible and inspect all ornaments closely.

- Label 2: this illustration is not framed by an ornament
- label 6: this illustration is framed by an ornament which contains plant parts
- label 3: this illustration is framed by an ornament which contains animal parts
- label 8: this illustration is framed by an ornament which contains abstract geometric shapes
- label 5: this illustration is framed by a simple line
- label 7: this illustration is framed by a banner or arch
- Label 4: this illustration is framed by an ornament which contains letters or words
- label 1: this illustration is framed by an ornament which contains other motives not mentioned above.

Please output the result in the ImageNet format, containing one file called default-GPT-singlefile.txt with the illustration filename and the label number(s) assigned to it. Use only the illustration I upload with this prompt for this.

Prompt 3:

Ornament is a generic term for the various kinds of decorative elements used in book illustration, e.g. arabesques, borders, flowers, head-pieces, rules, etc. Here, I want to focus only on ornaments that frame book illustrations, that is, they can be found on at least three sides of the illustration.

Please look at these images I am uploading. They are all book illustrations.

Then, please tell me if each book illustration contains a framing ornament, and label the illustration with one or more of the following categories. Please choose all labels that fit the illustration. Please be as accurate as possible and inspect all ornaments closely. Take only one illustration at a time and analyze it thoroughly as if I made this prompt only for this single illustration. Only after this, move to the next illustration, and do the same.

- Label 2: this illustration is not framed by an ornament
- label 6: this illustration is framed by an ornament which contains plant parts
- label 3: this illustration is framed by an ornament which contains animal parts
- label 8: this illustration is framed by an ornament which contains abstract geometric shapes
- label 5: this illustration is framed by a simple line

- label 7: this illustration is framed by a banner or arch
- Label 4: this illustration is framed by an ornament which contains letters or words
- label 1: this illustration is framed by an ornament which contains other motives not mentioned above.

Please provide me with the result in the ImageNet format, containing one file called default-GPT-1_prompt3.txt with the illustration filename and the label numbers assigned to it.