

Herausforderungen für die Klassifikation historischer Buchillustrationen

Überlegungen am Beispiel retrodigitalisierter Kinder- und Jugendsachbücher des 19. Jahrhunderts

Helm, Wiebke

wiebke.helm@uni-leipzig.de
Universität Leipzig, Deutschland

Im, Chanjong

imchan@uni-hildesheim.de
Universität Hildesheim, Deutschland

Mandl, Thomas

mandl@uni-hildesheim.de
Universität Hildesheim, Deutschland

Schmideler, Sebastian

sebastian.schmideler@uni-leipzig.de
Universität Leipzig, Deutschland

Zusammenfassung

Die maschinelle Verarbeitung einer großen Anzahl von Abbildungen eröffnet unter anderem für Forschungen im Bereich der Kunstgeschichte neue Chancen. Obwohl eine Vielzahl von Buchillustrationen bereits digitalisiert vorliegt, werden diese in den Digital Humanities bisher noch wenig beachtet. Das „Distant Viewing“-Verfahren soll es ermöglichen, nach Ähnlichkeiten von Bildern zu suchen oder Bildinhalte mittels Objekterkennung zu identifizieren. Der folgende Beitrag stellt grundlegende technische Herausforderungen ins Zentrum, die sich bei der automatischen Klassifikation von Illustrationen in einem Korpus von historischen Kinder- und Jugendbüchern des 19. Jahrhunderts ergeben haben. Neben der genauen Lokalisierung, also dem Erkennen der Lage von Illustrationen auf einer Buchseite erwies sich die korrekte Bestimmung einzelner Bildobjekte als schwierig. Für die Objekterkennung stehen zwar derzeit leistungsfähige Systeme zur Verfügung, doch wurden diese mit Fotografien natürlicher Objekte optimiert, wodurch sie sich nur bedingt für die Anwendung auf die im 19. Jahrhundert vorherrschenden Drucktechniken

oder die präzise Zuweisung von fiktionalen Bildinhalten eignen. Eine weitere Schwierigkeit ergab sich durch die unterschiedlichen Formate, in denen Illustrationen angelegt sind. Tafelbilder oder Abbildungen im Text können gerahmt sein oder fließende Übergänge aufweisen und Ornamente oder Textbausteine enthalten. Darüber hinaus können auch in Initialen und ornamentalen Schmuckleisten weitere Bildmotive vorkommen. Nachfolgend sollen die Herangehensweisen an die genannten Herausforderungen umrissen, Testergebnisse diskutiert und Lösungsansätze vorgeschlagen werden.

Einleitung

Die Digital Humanities entwickeln sich kontinuierlich weiter und verarbeiten längst nicht mehr nur ausschließlich Texte (Kohle 2013). Doch widmen sich nur wenige Studien der massenhaften Analyse von Abbildungen, was unter anderem auf die innerhalb der (kunstgeschichtlichen) Forschungscommunity bestehenden Vorbehalte und Kontroversen, inwieweit digitale Unterstützung das Selbstverständnis der Disziplinen verändert, zurückzuführen ist.



Abbildung aus Körner, Friedrich (Hrsg.): *Das Buch der Welt: Wanderungen nach Nord und Süd, Ost und West, zu den Wohnstätten der Gesittung und den Bewohnern der Wildniss. Bd. I: Die alte Welt*, Leipzig: Spamer 1855: 101; urn:nbn:de:gbv:084-18938

Die automatische Erkennung von Objekten auf Abbildungen bietet ein nicht zu unterschätzendes Potential für künftige Forschungsvorhaben, kann doch mithilfe von Software ein weitaus größerer Korpus von Bildern untersucht werden als es dem menschlichen Betrachter allein möglich ist. Beispielsweise können so Analysen zur Häufigkeit von Objekten oder quantitative Erkenntnisse zur Motiv-Geschichte einer bestimmten Illustration in größerem Umfang durchgeführt werden.

Die Verbreitung von visuellem Material erlebte im 19. Jahrhundert einen rasanten Anstieg. Die Zunahme von Wissen vermittelnden Illustrationen in Kinder- und

Jugendsachbüchern jener Epoche bildet diesen Prozess, der die Popularisierung von neuen Erkenntnissen und Entdeckungen beförderte, eindrücklich ab (vgl. Ries 1992, Schmideler 2014). Das Bildmaterial dieser Publikationen bietet daher eine ideale Voraussetzung für die Erforschung von Illustrationen mit Hilfe digitaler Methoden.



Abbildung aus Voltz, Johann Michael (Ill.)
(1815): *Bilderbogenbuch*, Nürnberg: Campe;
[urn:nbn:de:gbv:084-12012713339](http://urn.nbn.de/gbv:084-12012713339)

In den vergangenen Jahren wurden verschiedene Werkzeuge zur Klassifikation von Objekten in Abbildungen entwickelt und optimiert. Vor allem Ansätze aus Deep Learning Prozessen trugen dazu bei, für die Bilderkennung relevante Merkmale von Bildmotiven in Abhängigkeit von der Aufgabenstellung selbstständig herauszufiltern. Hier sind es insbesondere die aus mehreren Schichten bestehenden neuronalen Netzwerke (Convolutional Neural Networks, CNNs) (Krig 2016), die bei vielen Benchmarks hervorragende Werte aufweisen. Auf CNNs basieren auch die im Projekt getesteten Werkzeuge Yolo und SDS, die per Download installiert und lokal genutzt werden können. Sie sind auf Fotografien trainiert, können zahlreiche Arten von Objekten erkennen und weisen die untersuchten Objekte durch eine Markierung, ein Rechteck, aus (z.B. Redmon et al. 2016).

Fragestellungen zu Illustrationen aus der DH-Forschung

Innerhalb der Digital Humanities beschäftigen sich, wie eingangs erwähnt, nur wenige Forschungsprojekte mit Fragestellungen zu Illustrationen. Aufsehen erregte die erfolgreiche Klassifikation von zahlreichen Gemälden nach Künstler und Genre, bei der neuronale Netzwerke eingesetzt wurden (Saleh / Elgammal 2015).

In einem weiteren Ansatz wurde versucht, der Stil und der Inhalt von Gemälden zu trennen, um den Stil eines Künstlers auf andere Inhalte übertragen zu können (Gatys et

al. 2015). Die Studie erlaubt allerdings nur eine subjektive Bewertung der Ergebnisse.

Tiefergehende, auf Algorithmen gesteuerte Analysen wie in der Studie von Yarlagadda et al. (2013) versuchen auf einer Illustration die Geste einer Person zu erkennen, um daraus Rückschlüsse auf den damit verbundenen Rechtsakt eines Herrschers ziehen zu können.

Ein von der Projektgruppe erstellter Korpus von illustrierten Kinder- und Jugendsachbüchern des 19. Jahrhunderts repräsentiert einen Ausschnitt von verschiedenen Reproduktionsverfahren, die im Untersuchungszeitraum zur Herstellung von Buchillustrationen zur Anwendung kamen. Der Fokus wurde auf einige wenige ausgewählte Druckverfahren - Holzstich, Kupferstich und Feder- bzw. Kreidelithographie - gelegt. Es wurde ein System zur automatischen Klassifizierung dieser Verfahren erstellt, das allerdings noch kein zufriedenstellendes Ergebnis liefern konnte (Im et al. 2018).

Herausforderungen und Lösungsansätze

Werkzeuge zur Bilderkennung sollten zur korrekten Identifikation von Objekten bei einer großen Anzahl von Illustrationen führen. Aber wie verhält sich beispielsweise auf Fotografien trainierte Software bei der Anwendung auf Illustrationen, die in einer davon abweichenden reprografischen Technik angefertigt wurden und/oder weniger bekannte beziehungsweise untypische Bildobjekte ausweisen?

Bei einem Test an ca. 200 illustrierten historischen Kinderbüchern aus der retrodigitalisierten Kollektion des Sammlers Karl Hobrecker - digital bereitgestellt von der UB der TU Braunschweig - wurden alle als Bild erkannten Objekte mit dem Tool Yolo untersucht. Das Analyseergebnis machte deutlich, dass die Qualität des Werkzeuges zur Objekterkennung derzeit noch nicht ausreicht, um Bildmotive mit hoher Wahrscheinlichkeit für eine größere Kollektion zu verfolgen, da der Domain-Shift von Fotos zu Illustrationen zu einer Verschlechterung der Erkennungsleistung führt. Diese könnte durch umfangreiche Lerndaten verbessert werden, doch scheint dies keine tragfähige Lösung für Forschungsarbeiten im Rahmen der Digital Humanities zu sein, da selten einfache Objekte im Vordergrund stehen, sondern in der Regel Kompositionen oder komplexe Bildstrukturen sowie fiktionale Motive vorliegen. Somit kann die eindeutige Deklaration von Einzelmotiven lediglich der Klärung von Zwischenfragen im Verlauf eines Forschungsprozesses dienen. Auch aus Kostengründen können oftmals keine neuen Trainingsmengen für heterogene Fragestellungen erzeugt werden. Deshalb bietet sich das Nachtrainieren von bereits an Fotografien optimierten Systemen wie VGG-19 oder GoogleNet an.

Von Problemen bei der Erkennung von Objekten in Gemälden berichten beispielsweise auch Crowley und Zisserman (2016), welche ein CNN-Erkennungssystem auf Kunstwerke des 19. und 20. Jahrhunderts angewendet haben. Sehr kleine Objekte wie zum Beispiel winzige Tiere oder Flugzeuge im Bildhintergrund wurden kaum erkannt. Erst die Entwicklung eines zweistufigen Verfahrens führte zu verbesserten Ergebnissen.

In der Wissen vermittelnden Literatur - in Fachbüchern ebenso wie in populärwissenschaftlicher Literatur oder Kinder- und Jugendsachbüchern des 19. Jahrhunderts - werden Tiere teils in ungewöhnlichen Perspektiven dargestellt, die sich so nicht auf Fotos wiederfinden. Beispielsweise wird eine Fledermaus in der Draufsicht mit ausgestreckten Flügeln gezeigt. Auch werden Tiere selten in einem realistischen Größenverhältnis wiedergegeben. Das kann u.a. kompositionspragmatische Gründe haben. Somit ist das implizite Wissen eines vortrainierten Klassifikationssystems über Größen von Tieren in diesem Fall nicht weiterführend.

Die genutzte vortrainierte Version von Yolo eignet sich für die Erkennung von ca. 9.000 Klassen. Davon wurden in 1.891 Abbildungen aus 168 Titeln mit ca. 16.000 Seiten der digitalen Hobrecker-Kollektion insgesamt 4.600 Objekte aus 200 Klassen erkannt. Die häufigsten zeigt die nachfolgende Tabelle.

Objekt	Anzahl
Person	1.932
Organism	248
Artifact	188
Living thing	160
Entertainer	129
Animal	120
Worker	117
Bird	112
Horse	96

Tabelle: Frequenz der am häufigsten erkannten Bildobjekte mit Yolo

Es ist geplant, die Bilder zu einzelnen Klassen einer Inspektion zu unterziehen, um so einen Einblick in die Qualität der Klassifikation zu gewinnen.

In Kinderbüchern treten häufig anthropomorphisierte Tierfiguren auf, die den mit Fotografien trainierten Algorithmen nicht bekannt sind. Moderne Objekte wie Autos oder Baseball-Schläger werden hingegen gut erkannt, sie sind aber irrelevant für historische Illustrationen. Andererseits können Treffer in solchen Klassen als Fehler betrachtet werden, die bei der Verbesserung und Weiterentwicklung der Werkzeuge helfen können.

Die Erkennung von Objekten in Buchillustrationen birgt aber über die bereits genannten noch weitere technische Herausforderungen. In den digitalisierten

Bibliotheksbeständen können Illustrationen durch Werkzeuge der Schrifterkennung identifiziert und ihre Position ausgezeichnet werden. Bei der Hobrecker-Kollektion wie auch bei einzelnen Buchdigitalisaten aus anderen Institutionen zeigte sich aber, dass diese Auszeichnung zu ungenau ist und häufig nicht die notwendige Qualität erreicht. Darum muss die Identifikation von Illustrationen einer Buchseite zunächst technisch gelöst werden. Dafür hat sich der Einsatz von CNNs bewährt. Kleinere Trainingsmengen sind in diesem Fall ausreichend, wie beispielsweise eine Data Challenge alter Handschriften belegt, aufgrund derer die Erkennung unterschiedlicher Inhaltsbereiche verbessert werden konnte (Mehri et al. 2017).

Eine weitere Herausforderung stellt bereits der Aufbau einer Trainingsmenge zur Erkennung von Abbildungen auf digitalisierten Buchseiten dar: So werden beispielsweise rechteckige Abbildungen mit und ohne Rahmen getrennt in zwei Klassen erfasst. Doch treten auch Abbildungen auf, die kein einfaches quadratisches Format aufweisen, wie die nachstehenden beiden Abbildungen belegen.



Abbildung aus Braun, Ferdinand: *Braun, Ferdinand: Der junge Mathematiker und Naturforscher : Einführung in die Geheimnisse der Zahl und Wunder der Rechenkunst ; eine Anleitung zu aufmerksamer Naturbetrachtung, begleitet von zahlreichen Aufgaben zur Uebung des Urtheils und der Anschauung*, Leipzig: Spamer 1876: 259; urn:nbn:de:gbv:084-14052311261

Eine weitere Schwierigkeit bei der Klassifikation von Bildern in retrodigitalisierten Kinder- und Jugendsachbüchern des 19. Jahrhunderts stellen doppelseitige oder faltbare Bildseiten dar. Für deren automatische Erfassung ist ein zusätzlicher Arbeitsschritt notwendig, der je zwei Druckseiten zusammenfasst und prüft, ob es sich um ein doppelseitiges Bild handelt, da

die Auszeichnung der digitalen Daten auch hierzu keine Anhaltspunkte liefert.

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. Eher selten kommen in der untersuchten Kollektion gestaltete Initialen vor. Hierbei handelt es sich um farbig abgehobene Buchstaben (Lombarden) oder um künstlerisch aufwändig gestaltete Initialen (z.B. Figureninitialen). Gerade letztgenannte Bildelemente sollten einbezogen werden, da sie beispielsweise in ABC-Büchern eine wichtige Funktion bei der Wissensvermittlung übernehmen.



Coverabbildung von Hoffmann, Franz (1873):
Land- und See-Bilder in Erzählungen für die reifere Jugend: Zwei Theile in einem Bande, Stuttgart:
 Schmidt und Spring; <https://nbn-resolving.org/urn:nbn:de:gbv:084-14110314057>



Personifizierter Buchstabe. Abbildung aus: *Das allergrösste Bilder-ABC*, Berlin: Winckelmann [1828]; <urn:nbn:de:gbv:084-13932>

Ausblick

Wichtig für den zukünftigen Einsatz von „Distant Viewing“-Verfahren ist die Entwicklung von Evaluierungskonzepten, die stärker auf die Arbeitsweise und Bedarfe der Digital Humanities ausgerichtet sind und sich nicht nur an den Bewertungsmaßstäben der Bildverarbeitung orientieren. Zudem müssen geeignete Benutzungsoberflächen entwickelt werden, die den Umgang mit Ergebnissen erleichtern und Möglichkeiten zur Interaktion anbieten. Auf diese Weise könnten fehlklassifizierte Bilder sehr schnell und einfach von kompetenten Bildbetrachtern identifiziert und deren Informationen zur Verbesserung bestehender Systeme beitragen. Somit sind die vorgestellten Fehlklassifikationen der automatischen Bilderkennung nicht ausschließlich als Negativkriterium zu bewerten, sondern vielmehr als eine Chance zur Weiterentwicklung des „Distant Viewing“-Verfahrens wahrzunehmen.

Danksagung

Die Autoren danken für die Förderung durch die Fritz Thyssen Stiftung für das Projekt „Entwicklung der Bildikonographie in Wissen vermittelnder Kinder- und Jugendliteratur und Schullehrbüchern des 19. Jahrhunderts: ein Distant Viewing Ansatz“. Ebenso wird der UB der TU Braunschweig für die Bereitstellung der Daten der Hobrecker-Sammlung gedankt.

Bibliographie

Crowley, Elliot / Zisserman, Andrew (2016): “The Art of Detection”, in: Computer Vision – ECCV 2016 Workshops. ECCV 2016. Lecture Notes in Computer Science, vol 9913. Springer, Cham. S. 721-737

Gatys, Leon A. / Ecker, Alexander S. / Bethge, Matthias (2015): *A neural algorithm of artistic style* arXiv preprint arXiv:1508.06576.

Heuwing, Ben / Mandl, Thomas / Womser-Hacker, Christa (2016): “Combining contextual interviews and

participative design to define requirements for text analysis of historical media”, in: Proceedings of ISIC, the Information Behaviour Conference, Zadar, Croatia, 20-23 September, 2016: Part 1. Information Research, 21 (4) <http://www.informationr.net/ir/21-4/isic/isic1606.html>

Im, Chanjong / Ghauri, Junaid / Rothman, John / Mandl, Thomas (2018): “Deep Learning Approaches to Classification of Production Technology for 19th Century Books”, in: Lernen. Wissen. Daten. Analysen (LWDA 2018) Workshop Fachgruppe Information Retrieval (FGIR 2018) August 22-24, Mannheim: 150-158. <http://ceur-ws.org/Vol-2191>

Kohle, Hubertus (2013): *Digitale Bildwissenschaft*. Glückstadt: Verlag Werner Hülsbusch.

Krig, Scott (2016): *Computer Vision Metrics*. Cham: Springer.

Mehri, Maroua / Heroux, Pierre / Gomez-Krämer, Petra / Mullot, Rémy (2017): “Texture feature benchmarking and evaluation for historical document image analysis”, in: International Journal on Document Analysis and Recognition (IJDAR) 20 (1): 325-364.

Pech, Klaus-Ulrich (2008): *Der ökonomisch-technische Prozess und die Entwicklung der Kinder- und Jugendbuchliteratur*, in: **Brunken, Otto / Brüggemann, Theodor / Hurrelmann, Bettina / Michels-Kohlhage, Maria / Wilkending, Gisela (Hrsg.): *Handbuch zur Kinder- und Jugendliteratur. Von 1850 bis 1900*.** Stuttgart, Weimar: J. B. Metzler 15-22.

Redmon, Joseph / Divvala, Santosh / Girshick, Ross / Farhadi, Ali (2016): “You only look once: Unified, real-time object detection”, in: Proceedings of the IEEE conference on computer vision and pattern recognition: 779-788.

Ries, Hans (1992): *Illustration und Illustratoren des Kinder- und Jugendbuchs im deutschsprachigen Raum 1871–1914*. Osnabrück: Wenner.

Saleh, Babak / Elgammal, Ahmed (2015): “Large-scale Classification of Fine-Art Paintings: Learning The Right Metric on The Right Feature”, in: Computer Science > Computer Vision and Pattern Recognition, Arbeitsberichte. <http://arxiv.org/abs/1505.00855>

Schmideler, Sebastian (2014): „Das bildende Bild, das unterhaltende Bild, das bewegte Bild – Zur Codalität und Medialität in der Wissen vermittelnden Kinder- und Jugendliteratur des 18. und 19. Jahrhunderts“, in: **Weinkauf, Gina u.a. (Hrsg.): *Kinder- und Jugendliteratur in Medienkontexten. Adaption – Hybridisierung – Intermedialität – Konvergenz*.** Frankfurt a. M.: Peter Lang 13-26.

Yarlagadda, Pradeep / Monroy, Antonio / Carque, Bernd / Ommer, Björn (2013): “Towards a Computer-based Understanding of Medieval Images”, in: **Bock, Hans Georg / Jäger, Willi / Winckler, Michael J. (eds.): *Scientific Computing and Cultural Heritage. Contributions in Computational Humanities*.** Berlin: Springer 89-97.