

Alles unter Kontrolle?

KI im Einsatz im Bildarchiv der ETH-Bibliothek

Nicole Graf, ETH-Bibliothek Zürich, Bildarchiv

Zusammenfassung

Das Bildarchiv der ETH-Bibliothek in Zürich mit 3,6 Millionen Fotografien aus der Zeit zwischen 1860 und heute ist eines der größten historischen Bildarchive der Schweiz. Seit 2021 setzt das Bildarchiv nebst der herkömmlichen intellektuellen Erschließung auch künstliche Intelligenz für die inhaltliche Bilderschließung sowie für die Übersetzung von Metadaten ein. Die Ergebnisse sind gut. Die Nutzenden profitieren von zusätzlichen Recherchemöglichkeiten. Die automatisch generierten Tags (Autotags) sind näher an der Alltagssprache und gehen auch mehr ins Detail. Das Bildarchiv nutzt das Autotagging komplementär: Es ersetzt die intellektuelle Arbeit der Beschlagwortung nicht, sondern ergänzt sie. Während die Rechendauer beträchtlich ist, ist der finanzielle Aufwand vernachlässigbar. Angesichts der über 16 Monate dauernden Initialbearbeitung stellte sich die Frage, ob sich der Rechenaufwand zum Erzeugen der Autotags überhaupt lohnt. Eine Onlineumfrage brachte Erkenntnisse zur Akzeptanz der Autotags einerseits und den Recherchetechniken der Nutzenden andererseits.

Summary

With 3.6 million photographs – from the period between 1860 and today, the Image Archive of the ETH Library in Zurich is one of the largest historical image archives in Switzerland. Since 2021, the Image Archive uses artificial intelligence for content-based image indexing, in addition to intellectual indexing, and for the translation of metadata. The results are good. Users benefit from additional research options. The automatically generated tags are closer to everyday language and go into more detail. The Image Archive uses autotagging in a complementary way: It does not replace the intellectual work with the subject headings but complements it. While the computing time is considerable, the financial effort is negligible. In view of the initial processing, which took over 16 months, the question arose as to whether the computational effort required to generate the autotags was worthwhile at all. An online survey provided insights into the acceptance of the autotags on the one hand and the research techniques of the users on the other.

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5921>

Autorenidentifikation:

Graf, Nicole: ORCID: <https://orcid.org/0000-0003-2230-6679>

Schlagwörter: Bildarchiv; Automatische Erschließung; Computerunterstützte Inhaltsererschließung; Inhaltliche Erschließung; Formalerschließung; Bibliothekskongress; Künstliche Intelligenz; Bilderkennung

Dieses Werk steht unter der [Lizenz Creative Commons Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/).

1. Einführung zum Bildarchiv

Das Bildarchiv der ETH-Bibliothek in Zürich mit 3,6 Millionen Fotografien und anderen Bilddokumenten aus der Zeit zwischen 1860 und heute ist eines der größten historischen Bildarchive der Schweiz. Thematische Sammelschwerpunkte sind Bildbestände mit unmittelbarem Bezug zur Eidgenössischen Technischen Hochschule (ETH) Zürich wie Architektur und Bauwissenschaften, Ingenieurwissenschaften, Naturwissenschaften, Informatik oder Erd- und Umweltwissenschaften. Bildmaterial aus Organisationseinheiten der ETH Zürich, von Privatpersonen oder Institutionen mit direktem Bezug zur ETH Zürich werden ebenso wie Bildbestände und -archive externer Stellen (Privatpersonen, Organisationen, Stiftungen, Firmen) übernommen. Nebst der Bestandsbildung sind Erschließung, Digitalisierung, Vermittlung und Archivierung die klassischen Aufgabenfelder des Bildarchivs der ETH-Bibliothek.

Für die Erschließung und Vermittlung betreibt das Bildarchiv seit 2006 die webbasierte Bilddatenbank «E-Pics Bildarchiv» (<https://ba.e-pics.ethz.ch>), diese ist Teil von E-Pics, der Plattform der ETH Zürich für Fotografien und Bilddokumente (<https://www.e-pics.ethz.ch/>). Dahinter steht das Digital Asset Management System Canto Cumulus mit dem Webfrontend AWP. Auf E-Pics Bildarchiv sind rund 800 000 Bilder öffentlich zugänglich. Der größte Teil dieser veröffentlichten Bilder kann seit der Einführung der Open-Data-Policy¹ im März 2015 kostenfrei in verschiedenen Auflösungen heruntergeladen werden. Das Bildarchiv lizenziert diejenigen Bilder, deren Nutzungsrechte vollumfänglich bei der ETH-Bibliothek liegen, mit Creative Commons BY-SA 4.0. Diese Bilder können unter Angabe des korrekten Bildnachweises für wissenschaftliche, private, nicht-kommerzielle und kommerzielle Zwecke frei verwendet und müssen bei Veränderungen unter den gleichen Bedingungen weitergegeben werden.

2. Herkömmliche Datengenerierung

Die herkömmliche Datengenerierung in E-Pics Bildarchiv, das Inventarisieren und Erschließen von Bildern, kann in mehrere Arbeitsschritte zerlegt werden, die unterschiedliche Arten von Metadaten betreffen.

Im ersten Arbeitsschritt erfassen studentische Hilfskräfte die vorhandenen formalen Metadaten in Excel-Listen. Dies sind in der Regel Informationen, die mit den immer noch mehrheitlich analogen Originalbildern mitgeliefert werden, sei dies direkt auf dem Bild oder auf der Verpackung. Formale Metadaten sind etwa Titel und Beschreibung, Fotograf, Entstehungsdatum, physische Beschreibung mit Bildträger, Format, Größe und Farbe, Impressum bei Postkarten, Lizenz, Provenienz und Bildnachweis sowie interne Angaben etwa zum Standort oder der Digitalisierung. Das Metadatenfeld «Beschreibung» wird sparsam verwendet und primär für Kontextinformationen zum Bild und weniger zur generischen Beschreibung des Bildinhalts genutzt. Letzteres war in vordigitalen Zeiten in einem

1 ETH-Bibliothek: Open Data Zielsetzung, <<https://library.ethz.ch/archivieren-und-digitalisieren/open-data/zielsetzung.html>>, Stand: 31.01.2023.

Katalog wichtig, wenn auf der Katalogkarte keine Reproduktion montiert war, heute kann im digitalen Umfeld weitgehend darauf verzichtet werden.

Die im TIFF-Format gescannten Bilder enthalten technische Metadaten, die automatisch in die Bilddatenbank ausgelesen werden. In der Bilddatenbank werden die digitalen Dateien als Assets bezeichnet. Technische Metadaten sind etwa die Asset-Datengröße (in KB), das Asset-Erstellungs- und Änderungsdatum oder die Asset-Referenz im Vault (Speicherort).

Im nächsten Arbeitsschritt werden die formalen Metadaten aus den Excel-Listen mit den Masterfiles in der Datenbank zusammengeführt. Dies geschieht durch das Datenmatching via Bildcode. Der Bildcode ist gleichzeitig Bildsignatur und Name der Bilddatei. Anschließend werden die inhaltlichen Metadaten ergänzt, das fertig erschlossene Bild kann online gestellt werden. Die Beschlagwortung wird in der Regel durch Fachpersonal vorgenommen. Nebst den Schlagworten Ort, Person (für dargestellte Personen) und Sachschlagwort wird jedes Bild einem Bildgenre und einem Bildbestand zugeordnet.

Eine weniger klassische Form der Datengenerierung, die seit rund 15 Jahren in der digitalisierten Bibliothekswelt angewendet wird, ist das Crowdsourcing². Nach dem erfolgreichen Swissair-Projekt (2009–2013)³ können seit Dezember 2015 alle Nutzenden auf E-Pics Bildarchiv sämtliche Bilder per E-Mail kommentieren. Seit die «Neue Zürcher Zeitung» im Januar 2016 über diese Möglichkeit berichtete,⁴ erhält das Bildarchiv monatlich mehr als 1500 E-Mails. Dank der Hilfe der Freiwilligen können unter anderem Metadaten ergänzt, korrigiert und Bilder ohne Titel identifiziert werden. Eine weitere Form der Datengenerierung ist die Georeferenzierung von Bildern. Seit Februar 2018 werden Bilder aus dem Bildarchiv auf der kollaborativen Plattform sMapshot (<https://smAPSHOT.heig-vd.ch/owner/ethz>) ebenfalls durch Freiwillige georeferenziert. Bisher wurden rund 160 000 schrägaufgenommene Luftbilder der Schweiz, aber auch Bodenbilder, georeferenziert. Durch die Referenzierung berechnet der hinter sMapshot stehende Algorithmus die Geokoordinaten des Aufnahmestandpunkts, die Aufnahmehöhe sowie den geografischen Footprint des Bildinhaltes. Die Geodaten werden in die Bilddatenbank zurückgespeist und sind hier nicht nur abrufbar, sondern werden auch für wissenschaftliche Auswertungen nachgenutzt.

3. Bilderkennung mittels künstlicher Intelligenz

Die automatische, mittels künstlicher Intelligenz (KI) durchgeführte Bilderkennung resp. die Objekt- und Mustererkennung in Bildmaterialien ist in der Informatik seit Ende der 1950er-Jahre ein Thema. Erst seit rund zehn Jahren jedoch hat die Bilderkennung durch die Nutzung neuer Technologien, zum Beispiel Deep Learning mit Hilfe von *Convolutional Neural Networks*, signifikante Fortschritte gemacht.

2 Springer, Michelle et al.: For the Common Good: The Library of Congress Flickr Pilot Project, 30.10.2008, <https://www.loc.gov/rr/print/flickr_report_final.pdf>, Stand: 31.01.2023.

3 Graf, Nicole: Crowdsourcing. Die Erschließung des Fotoarchivs der Swissair im Bildarchiv der ETH-Bibliothek, Zürich, in: Rundbrief Fotografie, Vol. 23 (2016), No. 1, S. 24–32.

4 Kälin, Adi: Wer kennt die Berge, Orte und Fabriken?, in: Neue Zürcher Zeitung, 18. Januar 2016, <<https://www.nzz.ch/zuerich/wer-kennt-die-berge-orte-und-fabriken-1.18678913>>, Stand: 31.01.2023.

Damit jedoch eine Software in Deep-Learning-Prozessen selbst lernen kann, muss ein Datenkorpus bereitstehen, der vorab händisch von Menschen in maschinenlesbarer Form ausgezeichnet wird. So können zum Beispiel Objekte innerhalb des Bildes mit Bounding Boxes⁵ umgeben und diese Regionen dann verschlagwortet werden. Die Bildinhalte werden in Konzepte (z.B. «Katze») aufgeteilt. Damit die Software trainiert werden kann, sind rund 500 bis 1000 repräsentative Bilder eines Konzepts notwendig. Für gute Trainingsdaten braucht es beispielsweise nicht nur Bilder von schwarzen Katzen auf grünem Gras, sondern unterschiedlichste Katzen in verschiedenen Settings und Körperhaltungen (siehe Bildbeispiele). Die kategorisierten Daten werden dann genutzt, um die Computer-Vision-Software resp. das künstliche Neuronale Netzwerk zu trainieren. Es sind große Mengen an Bildern mit möglichst vielen Konzepten notwendig. Die wichtigste Datenbank resp. der am meisten genutzte «vortrainierte» Datensatz ist ImageNet⁶ mit über 14 Millionen Bildern und über 20 000 Konzepten. Bei Deep-Learning-Modellen erzeugen die Algorithmen ohne die Programmierung durch den Menschen sowohl die Programmcodes als auch ein Regelwerk (Modell). Gibt man dem Algorithmus neue Daten hinzu, kann er diese nun bewerten und «weiter lernen».



Abbildung 1: Input «Katze». Foto: Josef Schmid⁷



Abbildung 2: Input «Schlafende Katze». Foto: Josef Schmid⁸

Dies bedeutet, in einfachen Worten und mit Beispielbildern des Bildarchivs illustriert, folgendes: Mit dem Trainingsbild einer Katze in einer klassischen Körperhaltung als Input würde der Algorithmus schließlich das Bild der «schlafenden Katze» als Konzept «Katze» erkennen und richtig kategorisieren können.

Da das Tagging der Trainingsdaten immer auch die Perspektive der taggenden Person auf ein Bild darstellt, kann es im Rahmen dieser Verfahren zu Gender- oder Kultur-Bias kommen, einem Phänomen, das sowohl von der kulturwissenschaftlichen Forschung als auch der Informatik erkannt und diskutiert wird.⁹

5 Begrenzungsrahmen, die als Bezugspunkte für die Objekterkennung dienen.

6 Englische Wikipedia: Artikel «ImageNet», <<https://en.wikipedia.org/wiki/ImageNet>>, Stand: 31.01.2023. Dave Gersghorn: The data that transformed AI research—and possibly the world, 26.07.2017, <<https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world>>, Stand: 31.01.2023.

7 Schmid, Josef: Katze, 2/1973. Bildnachweis: ETH-Bibliothek Zürich, Bildarchiv, <<http://doi.org/10.3932/ethz-a-001002487>>.

8 Schmid, Josef: Katze beim Schlafen, 2/1973. Bildnachweis: ETH-Bibliothek Zürich, Bildarchiv, <<http://doi.org/10.3932/ethz-a-001002485>>.

9 Xiang, Chloe: KI ist weder künstlich noch intelligent, 24.01.2023, <<https://www.vice.com/de/article/wxnaqz/ki-ist-weder-kuenstlich-noch-intelligent>>, Stand: 31.01.2023. Vgl. auch Meyer, Roland: Gesichtserkennung. Berlin 2021

4. Bildererkennung mit Clarifai (Autotagging)

Die Computer-Vision-Software «Clarifai»¹⁰ kann seit 2021 im Backend der Bilddatenbank Cumulus als Autotagging-Plugin integriert werden. Das Plugin wurde von CDS Gromke für Cumulus entwickelt.

Mittels automatischer Bildererkennung werden die Bildinhalte analysiert und als Ergebnis werden textuelle Autotags ausgegeben. Der Prozess wird «Autotagging» genannt. Im Autotagging Control Panel in Cumulus können folgende Einstellungen vorgenommen werden:

1. Modell («Autotag Model»): mögliche Modelle sind «General» sowie die spezifischeren «Wedding», «Travel», «Food» und «Apparel».
2. Farbe («Autotagging Color»): Auszug von bis zu acht Hauptfarben.
3. Klassifizierung («Autotagging Klassifizierung») von sensiblen Inhalten.
4. Gesichtserkennung («Autotagging Face»): Gesicht, Alter, Kultur, Geschlecht.

Das Bildarchiv nutzt das allgemeine Modell «General», das für historisches Bildmaterial geeignet ist und gute Ergebnisse liefert. Bei den Farbbildern wird zusätzlich die «Color»-Funktion aktiviert, die die Hauptfarben aus dem Bild herauszieht und im Webfrontend als zusätzlicher Filter «Farbenauswahl» bei der Recherche benutzt werden kann. Der Erkennungsprozess pro Bild dauert durchschnittlich 20 bis 30 Sekunden.

Die Autotags von Clarifai werden schließlich in einer alphabetischen Liste ausgegeben. Über den Gesamtbestand ergibt das im E-Pics Bildarchiv momentan eine Liste von rund 4600 automatisch generierten Tags, die bei «Aasfresser» beginnt und bis «Zypressenholz» geht. Die Original-Autotags sind in englischer Sprache, bei der automatischen Übersetzung ins Deutsche kommt gelegentlich die



Abbildung 3: Kategorien im Webfrontend von E-Pics Bildarchiv (Screenshot)

(Digitale Bildkulturen). Serie «Digital Warriors», in: Republik, 04.01.2022, <<https://www.republik.ch/2022/01/04/serie-digital-warriors>>, Stand: 31.01.2023. Kantayya, Shalini: Coded Bias – vorprogrammierte Diskriminierung (Dokumentarfilm), 2020, <<https://www.ajl.org/spotlight-documentary-coded-bias>>, Stand: 31.01.2023. Social Media Collective: Critical algorithm studies: a reading list, 2016, <<https://socialmediacollecive.org/reading-lists/critical-algorithm-studies>>, Stand: 31.01.2023. Webseite Algorithmic Justice League, <<https://www.ajl.org>>, Stand: 31.01.2023.

10 Clarifai wurde 2013 vom US-amerikanischen Informatiker Matthew Zeiler gegründet, der bei der ImageNet Challenge 2013 auf einem der ersten fünf Plätze landete.

Groß- und Kleinschreibung durcheinander und es kommt auch zu Doppelungen der Bezeichnungen.¹¹ Die Autotags sind im Webfrontend Teil der Kategorien. Sie heißen dort per default «Keywords»¹² und können, wie alle anderen Kategorien auch, komfortabel im Kategoriensuchschlitz durchsucht werden.

5. Clarifai: erste Ergebnisse

Die folgenden Beispiele diskutieren erste Ergebnisse mit Clarifai und setzen sie in Kontrast zur Beschlagwortung durch das Fachpersonal. Eine vertiefte Analyse steht allerdings noch aus.

Beispiel 1: «schlagzeuger»



Abbildung 4: Keyword «Schlagzeuger».
Foto: Comet Photo AG¹³

Beim ersten Bild (Abbildung 4) erzeugt das Autotagging eine Reihe an zusätzlichen Detailinformationen, die bei der intellektuellen Beschlagwortung weder vorgesehen noch als angemessen definiert sind. Und gerade dadurch entsteht der Mehrwert. Die Details ermöglichen neue Recherchemöglichkeiten mit oft überraschenden Resultaten. Der potenzielle Aufwand für das Fachpersonal mit der manuellen Beschlagwortung dieselbe Detailtiefe zu erzeugen, ist jedoch zu groß.

Inhaltlich betrachtet sind beim ersten Beispiel zehn von elf Autotags (s.u.) korrekt, es wurden keine widersprüchlichen Autotags ausgegeben. Das angeschnittene Schlagzeug und auch das Mikrofon wurden erkannt. Eine Gitarre lehnt am Verstärker. Der Schlagzeuger hingegen ist auf dem Bild nicht zu sehen. Persönlichkeiten wie Mick Jagger werden nicht ausgegeben. Mit der Einstellung «Autotagging Face» würde allenfalls auf generische Art Gesicht, Alter, Kultur oder Geschlecht erfasst werden. Die Verschlagwortung hingegen erfasst lediglich die Menschen im Kontext von Musik (Orchester, Gesang, Oper, Musical).

11 Die manuelle Nachbearbeitung der Liste ist nicht zweckmässig.

12 Die manuelle Neubenennung der Kategorie in «Autotags» ist nicht möglich. Der Begriff «Schlagwörter» musste gewählt werden, weil Clarifai die Keywords/Autotags in den Metadaten auf Einzelbildebene als «Schlagwörter» ausgibt.

13 Comet Photo AG: The Rolling Stones, Konzert im Hallenstadion, 14.04.1967, Bildnachweis: ETH-Bibliothek Zürich, Bildarchiv, <<http://doi.org/10.3932/ethz-a-000990703>>.

Die Metadaten zum Beispiel 1 «schlagzeuger»:

Comet Photo AG: The Rolling Stones, Konzert im Hallenstadion, 14.04.1967
(Com_L16-0280-1903, <http://doi.org/10.3932/ethz-a-000990703>)

Autotags: Erwachsener, gitarre, Instrument, Menschen, mikrofon, Musik, Musiker, Saiteninstrument, Schlaginstrument, schlagzeuger, trommel

Schlagwortketten:

- Menschen im Kontext / Bildung + Wissenschaft + Kultur / Musik (Orchester, Gesang, Oper, Musical)
- Schweiz / Zürich, Kanton / Zürich, Stadt

Beispiel 2: «schreiben»



Abbildung 5: Keyword «schreiben». Foto: Unbekannt¹⁴

Ein Beispiel für die erweiterte Bildrecherche-möglichkeiten veranschaulicht folgendes Bild. Konkret wurde ein «Teaserbild» für einen Wikipedia-Workshop gesucht. Auf gut Glück wurde «schreiben» eingegeben, ein gleichnamiger Autotag mit rund 1000 Treffern wurde gefunden. Mit den herkömmlichen Schlagwörtern hätte man ein vergleichbares Bild ohne Überlegungen zu möglichen Schreibkontexten nicht gefunden. Verschlagwortet war im vorliegenden Bild etwa das alte Bibliothekssystem ETHICS im Katalogsaal der ETH-Bibliothek. Der schreibende Mann war für die Beschlagwortung

nicht relevant. Die Autotags hingegen bringen weitere interessante Resultate hervor. Von der «Konzentration» – damit ist wohl die Konzentration des Schreibenden gemeint – zur Interpretation, dass das Setting in einem «Unternehmen» stattfindet. Die Technologie wird mit «Laptop» allerdings historisch nicht präzise erkannt, Laptops waren erst ab der 1990er-Jahre allgemein gebräuchlich. Hier ist das generischere Schlagwort «Geräte der Informationstechnik (Computer)» von Vorteil.

Die Metadaten zum Beispiel 2 «schreiben»:

Unbekannt: Zürich, ETH Zürich, Hauptgebäude (HG), Hauptbibliothek, Katalogsaal, ETHICS Online-Abfrage, 15.04.1986 (Ans_01693-005, <http://doi.org/10.3932/ethz-a-000014639>)

Autotags: büro, computer, Erwachsener, innerhalb, Konzentration, laptop, Mann, Menschen, Möbel, schreiben, Schreibtisch, Technologie, Unternehmen.

¹⁴ Unbekannt: Zürich, ETH Zürich, Hauptgebäude (HG), Hauptbibliothek, Katalogsaal, ETHICS Online-Abfrage, 15.04.1986, Bildnachweis: ETH-Bibliothek Zürich, Bildarchiv, <<http://doi.org/10.3932/ethz-a-000014639>>.

Schlagwortketten:

- Menschen im Kontext / Bildung + Wissenschaft + Kultur / Archiv + Bibliothek.
- ETH Zürich / Gebäude / ETH Zentrum / Hauptgebäude (HG) / ETH-Bibliothek / Katalogosaal / ETHICS.
- Objekte + Maschinen / Instrumente + Geräte / Geräte der Informationstechnik (Computer).
- Schweiz / Zürich, Kanton / Zürich, Stadt / Zürich-Oberstrass.

Beispiel 3: Luftbild



Abbildung 6: Luftbild. Foto: Walter Mittelholzer¹⁵

Beim letzten Beispiel handelt es sich um ein für das Bildarchiv typisches Bildgenre, ein schrägaufgenommenes Luftbild. Der Großteil der Autotags ist auch hier korrekt, wenn auch unspezifisch wie «Fluss» oder «Gewässer». Allerdings kommt hier wie auch häufig im Gesamtbestand die widersprüchliche Autotag-Kombination «Menschen» und «Keine Person» vor.¹⁶ Das abstrakte Konzept «Reise» würde bei einem Luftbild händisch nicht vergeben werden. Die weiße Fläche im Bild verleitet die KI irrtümlicherweise zu den Autotags «Schnee» und «Winter».

Die manuelle Beschlagwortung von Luftbildern fokussiert in der Regel nur auf das Ortsschlagwort, ausnahmsweise, falls auch im Titel eine Bezeichnung vorhanden ist, wird das präzise Sachschlagwort vergeben (hier: «Stauseen»).

Die Metadaten zum Beispiel 3 Luftbild

Mittelholzer, Walter: Flawil-Gossau, Stausee im Glatt-Tobel bei der Ruine Helfenberg, 17.08.1932 (LBS_MH01-006949, <http://doi.org/10.3932/ethz-a-000298816>)

Autotags: Baum, Berg, Fluss, Gewässer, keine Person, Landschaft, Menschen, Reise, Schnee, Winter.

Schlagwortketten:

- Natur / Gewässer / Brücken / Stauseen.
- Schweiz / St. Gallen, Kanton / Flawil.
- Schweiz / St. Gallen, Kanton / Gossau (SG).

15 Mittelholzer, Walter: Flawil-Gossau, Stausee im Glatt-Tobel bei der Ruine Helfenberg, 17.08.1932, Bildnachweis: ETH-Bibliothek Zürich, Bildarchiv/Stiftung Luftbild Schweiz, <<http://doi.org/10.3932/ethz-a-000298816>>.

16 Bei knapp der Hälfte kommt die widersprüchliche Kombination «Menschen» mit «Keine Person» vor.

Das Bildarchiv nimmt weder bei den Autotags auf Bildebene noch in der alphabetischen Liste der Keywords Korrekturen vor. Falsche oder widersprüchliche Autotags werden ebenso belassen wie Doppelungen durch Groß- und Kleinschreibung oder aufgrund des scharfen S, das als «ß» oder «ss» geschrieben wird. Bei den Doppelungen ist zudem auffällig, dass die Treffer unterschiedlich sind und sich wenig überschneiden.¹⁷ Der Großteil der Autotags sind Substantive, es kommen aber auch Verben und Adjektive vor. Es gibt sowohl konkrete (Abfalleimer, adlerfarm, afrikanischer Elefant usw.) wie auch abstrakte Konzepte, die einer Interpretation bedürfen (Ablehnen, Abstraktion (Philosophie), Abwesenheit usw.).

Der Einsatz der KI war von Anfang an nicht als Ersatz für die intellektuelle hierarchische Beschlagwortung konzipiert. Die vorliegenden Ergebnisse bestätigen diese Prämisse. Die KI ersetzt die intellektuelle Beschlagwortung nicht, sie ergänzt sie allenfalls. Bei der Beschlagwortung wird das zentrale Motiv erfasst, beim Taggen hingegen wird das ganze Bild bis in kleinste Details (vgl. das Mikrofon) «seziert». Der Schlagwortbaum ist teilweise kontextualisiert, d. h. es gibt Schlagwörter für einzelne Gebäude wie die Gebäude der ETH oder für Personen wie Albert Einstein. Die Autotags von Clarifai hingegen geben weder spezifische Personen noch Gebäude wie etwa den Eiffelturm aus. Eiffelturm-Bilder erhalten das generische «Turm» und allenfalls noch «Wahrzeichen». Im Bildarchiv werden die Bilder daher künftig sowohl manuell beschlagwortet als auch mittels KI erschlossen.

6. Autotranslate mit DeepL

Ein lange bestehendes Desiderat wurde ebenfalls mit KI erfüllt. Seit Mitte März 2022 ist es möglich, die KI-basierte Übersetzungssoftware DeepLPro im Backend von Cumulus zu integrieren und die ausgewählten Metadatenfelder «Titel», «Beschreibung», «Enthalten in» (für Reportagetitel u. ä.) und «Kommentare» maschinell übersetzen zu lassen. Dazu mussten im Backend die entsprechenden Metadatenfelder neu angelegt (z.B. «Titel (Englisch)», ein spezifisches Datensatz-Ansichtenset eingerichtet und ein Trigger, um die automatische Übersetzung zu starten, definiert werden.

Ändern Nutzenden im Webfrontend die Sprache auf Englisch, werden die maschinell übersetzten Metadaten angezeigt und durchsucht. Die Inhalte von tabellarischen Metadatenfeldern (Farbe, Form, physische Beschreibung usw.), die Kategorien und auch die Feldbezeichnungen sind seit langem zweisprachig. Mit der Onlineschaltung der englischen Übersetzungen im August 2022 wurde das seit 2018 im Hintergrund geführte Feld «Originaltitel» erstmals auch online angezeigt. Im Originaltitel wird der Titel im Ursprungszustand, wie er beim Inventarisieren erfasst wurde, wiedergegeben. Die Metadaten des Bildarchivs sind sehr «volatil», mehrheitlich bedingt durch die unzähligen Kommentare der Freiwilligen. Das Feld Originaltitel wird in beiden Sprachversionen angezeigt, damit nach späteren Überarbeitungen zumindest teilweise nachvollzogen werden kann, welche Bestandteile des Titels geändert wurden. Bei missverständlichen oder unklaren maschinellen Übersetzungen kann also auf das Original zurückgegriffen werden.

¹⁷ «Menschen» wird bei 267 951 Bildern, «menschen» wiederum bei über 83 589 Bildern gefunden, das sind 351 540 Bilder. 350 473 Bilder enthalten jeweils nur eines der Keywords.

Beispiel in Deutsch mit englischer Übersetzung¹⁸

- Titel:
 - Aarburg, Brückenschlag zwischen Bifang-Kupferschmied und Ruttigerhof.
 - Aarburg, bridge building between Bifang coppersmith and Ruttigerhof.
- Beschreibung:
 - Fertigstellung des westlichen Endes der Brücke.
 - Completion of the western end of the bridge.
- Enthalten in:
 - [Pontoniere, 1914–915]. 3 Alben mit 198 Bildern (alle digitalisiert).
 - [Pontooniers, 1914–1915]. 3 albums with 198 pictures (all digitized).
- Kommentare:
 - Thomas Pfister: Aarburg, Brückenschlag zwischen Bifang-Kupferschmied und Ruttigerhof. Fertigstellung des westlichen Endes der Brücke.
 - Thomas Pfister: Aarburg, bridge between Bifang-Kupferschmied and Ruttigerhof. Completion of the western end of the bridge.

Im Moment ist DeepL eines der besten Übersetzungsprogramme auf dem Markt.¹⁹ Bei den Metadaten des Bildarchivs liefert DeepL größtenteils gute Übersetzungen. Luftbilder und Architekturaufnahmen, die einen beträchtlichen Anteil am Bestand ausmachen, sind meist nach dem relativ einfachen Schema «Ort, Ortsteil, Straße, Gebäude» aufgebaut.

Als problematisch stellten sich hingegen beispielsweise folgende Fälle heraus:

- Ortsnamen werden falsch übersetzt, z.B. Brunnen als «Fountain», Baden als «Bathing» oder Speicher als «Memory».
- Geografische Abkürzungen wie «v. S.» [von Süden], «v. N. O.» [von Nordosten] und schweizerische Begriffe wie ARA (Abwasserreinigungsanlage) werden nicht erkannt.

Eine weitere Hürde ist, dass jede Sprache einzeln übersetzt werden muss. Hierfür müssen diese zuerst identifiziert werden²⁰, denn die Metadaten enthalten keine fixen Felder für die Bezeichnung der Sprache des Katalogsats wie etwa im MARC-21-Format-Feld 008.

7. Aufwand, Kosten und Ertrag

Die maschinellen Erschließungsverfahren waren für die Initialbearbeitung der digitalen Daten wie folgt im Einsatz:

18 Unbekannt: Aarburg, Brückenschlag zwischen Bifang-Kupferschmied und Ruttigerhof, 1914–1915, <<http://doi.org/10.3932/ethz-a-000039406>>.

19 Wolfangel, Eva: Übersetzer DeepL: wirklich besser als Google?, in: Die Zeit, 16.01.2023, <<https://www.zeit.de/digital/2023-01/deep-uebersetzer-google-translate-ki/komplettansicht>>, Stand: 31.01.2023.

20 Ein Bruchteil des Bestandes (geschätzt ist es weniger als 1%) ist mit nicht-deutschsprachigen Metadaten versehen, etwa Englisch oder Französisch.

- Autotagging
 - Rechendauer pro Bild: 20 bis 30 Sekunden.
 - Initialbearbeitung von Februar 2021 bis Mai 2022.
- Color
 - Rechendauer pro Bild: ca. zwei Sekunden.
 - Initialbearbeitung von Mai 2022 bis Juli 2022.
- Autotranslate
 - Rechendauer pro Bild: ca. sechs Sekunden.
 - Initialbearbeitung von Mai 2022 bis Dezember 2022.

Da die Anwendung dieser Verfahren auf den Produktionsservern sehr ressourcenintensiv ist, so dass die Performance sowohl des Back- als auch des Webfrontends negativ davon betroffen ist, kann die Bildanalyse jeweils erst nach Feierabend, das Autotranslate sogar erst ab 22 Uhr resp. am Wochenende ausgeführt werden. Die Prozesse müssen jeweils manuell gestartet werden. Für die Initialbearbeitung aller Bilder mit den drei genannten Verfahren waren insgesamt 676 Tage (inkl. Wochenenden) reine Rechendauer notwendig.

Tabelle 1: Geschätzte Rechendauer für Autotagging und Autotranslate

	Autotagging	Color	Autotranslate
Anzahl Datensätze	980 142	595 760	1 067 094
Ø-Sekunden / Datensatz	20–30	2	6
Anzahl Zeichen			103 665 189
Anzahl Soll-Tage bei 24 Std.-Betrieb	340	14	74

Die Kosten werden beim Autotagging pro Aktion berechnet, das Modell «General» und die Aktion «Color» sind je eine Aktion. So kosten 500 Farbbilder beispielsweise USD 1,20. Rund eine Million mit einer Aktion getaggte Bilder kosten insgesamt USD 1.176.

Die Kosten für das Autotranslate werden nach Anzahl übersetzter Zeichen berechnet. Die rund eine Million bearbeiteten Bilder generierten rund 103 Millionen übersetzte Zeichen und kosteten EUR 2.073.

Tabelle 2: Kosten für Autotagging und Autotranslate

	Autotagging mit Color	Autotranslate
Anzahl Datensätze	USD 1,20 / 1000 Einheiten	EUR 500 / 25 Mio. Zeichen

Während die Rechendauer beträchtlich ist, ist der finanzielle Aufwand also vernachlässigbar. Angesichts der über 16 Monate dauernden Initialbearbeitung stellte sich die Frage, ob sich der doch

erhebliche zeitintensive Rechenaufwand – das allabendliche manuelle Starten des Rechenprozesses nicht zu vergessen – zum Erzeugen der Autotags überhaupt lohnt. Werden die mittels KI erzeugten Autotags von den Nutzenden genutzt und falls ja, wie?

Um Anhaltspunkte zu den Suchstrategien der Nutzenden im Webfrontend von E-Pics Bildarchiv zu gewinnen, wurde im Mai 2022 eine Online-Umfrage durchgeführt.²¹ Es wurden 1'206 Personen wie Nutzende, Crowdsourcer:innen oder Mitarbeitende von GLAM-Institutionen per Mail angeschrieben. Die Rücklaufquote betrug mit 199 ausgefüllten Fragebogen 16,5%. Wie zu erwarten war, nutzen beinahe alle befragten Personen die google-ähnliche Stichwortsuche im zentralen Suchschlitz. Die Hälfte der Befragten nutzt auch den zweiten und neueren Suchschlitz für die Kategorien. Jede/r Dritte nutzt die Schlagwörter regelmäßig und fast jede/r Vierte hat sie zumindest schon ausprobiert. Insgesamt wird die Schlagwortsuche gut akzeptiert und aktiv genutzt. Nur 16,6% der Antwortenden (?) nutzen die Schlagwörter nicht oder nicht mehr. Bereits 15% der Antwortenden (?) nutzen die neuen Keywords regelmäßig und je rund 25% der Antwortenden (?) hat sie bereits ausprobiert resp. entdeckte sie dank der Umfrage. Der Nutzen der inhaltlichen Erschließung wird von den meisten Nutzenden erkannt und geschätzt. Insgesamt werden Schlagwörter und Keywords als hilfreich angesehen. Die etwas provokante Aussage, dass es Schlagwörter und Keywords nicht bräuchte und die Stichwortsuche ausreichen würden, bejaht immerhin ein Drittel der Befragten. Ein Viertel der Befragten findet die Schlagwörter sehr hilfreich, da diese den Inhalt gut strukturieren. Die Frage, ob den Nutzenden der Unterschied zwischen Schlagwörtern/Autotags und Keywords klar sei, beantworteten 44,4% mit «Ja», 39,8% hingegen mit «Nein». Für 15,8% spielt diese Unterscheidung keine Rolle.

8. Fazit

Die Nutzenden bekommen dank der Autotags/Keywords eine zusätzliche Recherchemöglichkeit: die Tags sind näher an der Alltagssprache und gehen mehr ins Detail. Es sind neue und allenfalls auch unerwartete Suchergebnisse möglich.

Das Bildarchiv nutzt das Autotagging komplementär. Es ersetzt die intellektuelle Arbeit mit den Schlagwörtern nicht, sondern ergänzt sie. Falsche Tags werden nicht korrigiert, auch die Freiwilligen des Crowdsourcings wurden darauf hingewiesen, diese nicht zu korrigieren. Es ist vielmehr zu überlegen, ob die Bilder nun regelmässig neu getaggt werden müssen, da die KI sich ständig weiterentwickelt.

Wie gezeigt werden konnte, ist die zeitliche Rechendauer sehr hoch. Für die Initialbearbeitung wären 428 Tage in einem 24-Stundenbetrieb notwendig gewesen. Die Prozesse müssen jeweils manuell angestoßen werden. Da die Anwendung dieser Verfahren auf den Produktionsservern sehr ressourcenintensiv ist und die Performance sowohl des Back- als auch des Webfrontends negativ davon betroffen ist, wird die KI nur außerhalb der Bürozeiten betrieben. Gemessen daran ist der finanzielle Aufwand gering.

21 Graf, Nicole: Bildersuche in E-Pics Bildarchiv. Online-Umfrage zu Suchstrategien von E-Pics-Nutzenden. Zürich: ETH-Bibliothek, Dezember 2022. <<https://doi.org/10.3929/ethz-b-000588615>>.

Es ist zu hoffen, dass künftig auch Modelle für historische Bilder trainiert werden. Der über eine Million umfassende Bestand des Bildarchivs wäre ein guter Anfang für die Verwandlung von Bibliotheksarchiven in KI-Trainingsdaten wie es die Königliche Bibliothek Schwedens gerade vormacht.²²

Die IFLA hat zwar im September 2020 ein «Statement on Libraries and Artificial Intelligence»²³ veröffentlicht, allerdings fehlt bisher noch eine Debatte in Bezug auf ethische Richtlinien für KI in den GLAM-Institutionen.

Literaturverzeichnis

- Dave Gersghorn: The data that transformed AI research – and possibly the world, 26.07.2017, <<https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world>>, Stand: 31.01.2023.
- Englische Wikipedia: Artikel «ImageNet», <<https://en.wikipedia.org/wiki/ImageNet>>, Stand: 31.01.2023.
- ETH-Bibliothek: Open Data Zielsetzung, <<https://library.ethz.ch/archivieren-und-digitalisieren/open-data/zielsetzung.html>>, Stand: 31.01.2023.
- Graf, Nicole: Bildersuche in E-Pics Bildarchiv. Online-Umfrage zu Suchstrategien von E-Pics-Nutzenden. Zürich: ETH-Bibliothek, Dezember 2022, <<https://doi.org/10.3929/ethz-b-000588615>>.
- Graf, Nicole: Crowdsourcing. Die Erschließung des Fotoarchivs der Swissair im Bildarchiv der ETH-Bibliothek, Zürich, in: Rundbrief Fotografie, Vol. 23 (2016), No. 1, S. 24–32.
- International Federation of Library Associations: Statement on Libraries and Artificial Intelligence, 17.09.2020, <https://www.ifla.org/wp-content/uploads/2019/05/assets/faife/ifla_statement_on_libraries_and_artificial_intelligence.pdf>, Stand: 31.01.2023.
- Kälän, Adi: Wer kennt die Berge, Orte und Fabriken?, in: Neue Zürcher Zeitung, 18.1.2016, <<https://www.nzz.ch/zuerich/wer-kennt-die-berge-orte-und-fabriken-1.18678913>>, Stand: 31.01.2023.
- Kantayya, Shalini: Coded Bias – vorprogrammierte Diskriminierung (Dokumentarfilm), 2020, <<https://www.ajl.org/spotlight-documentary-coded-bias>>, Stand: 31.01.2023.
- Meyer, Roland: Gesichtserkennung. Berlin 2021 (Digitale Bildkulturen).
- Salian, Isha: Booked for brilliance: Sweden's National Library turns page to AI to parse centuries of data, 23.01.2023, <https://blogs.nvidia.com/blog/2023/01/23/sweden-library-ai-open-source/?ncid=so-twit-236541#cid=gnl_so-twit_en-eu>, Stand: 31.01.2023.
- Serie «Digital Warriors», Republik, ab 04.01.2022, <<https://www.republik.ch/2022/01/04/serie-digital-warriors>>, Stand: 31.01.2023.

22 Salian, Isha: Booked for Brilliance: Sweden's National Library Turns Page to AI to Parse Centuries of Data, 23.01.2023, <https://blogs.nvidia.com/blog/2023/01/23/sweden-library-ai-open-source/?ncid=so-twit-236541#cid=gnl_so-twit_en-eu>, Stand: 31.01.2023.

23 International Federation of Library Associations: Statement on Libraries and Artificial Intelligence, 17.09.2020, <https://www.ifla.org/wp-content/uploads/2019/05/assets/faife/ifla_statement_on_libraries_and_artificial_intelligence.pdf>, Stand: 31.01.2023.

- Social Media Collective: Critical algorithm studies: a reading list, 2016, <https://socialmediacollective.org/reading-lists/critical-algorithm-studies/>, Stand: 31.01.2023.
- Springer, Michelle et al.: For the common good: The Library of Congress Flickr pilot project, 30.10.2008, <https://www.loc.gov/rr/print/flickr_report_final.pdf>, Stand: 31.01.2023.
- Wolfangel, Eva: Übersetzer DeepL: wirklich besser als Google?, in: Die Zeit, 16.01.2023, <<https://www.zeit.de/digital/2023-01/deepl-uebersetzer-google-translate-ki/komplettansicht>>, Stand: 31.01.2023.
- Xiang, Chloe: KI ist weder künstlich noch intelligent, 24.01.2023, <<https://www.vice.com/de/article/wxnaqz/ki-ist-weder-kuenstlich-noch-intelligent>>, Stand: 31.01.2023.