Zwischen Ritual und Relief – wenn der Computer schielt

Analyse KI-gestützter Erschließung im Bildarchiv der ETH-Bibliothek

1. Das Bildarchiv der ETH-Bibliothek

Das Bildarchiv der ETH-Bibliothek in Zürich ist mit 3,6 Millionen Fotografien und anderen Bilddokumenten aus der Zeit zwischen 1860 und heute 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.[1]

Für die Erschließung und Vermittlung betreibt das Bildarchiv seit 2006 die webbasierte Bilddatenbank „E-Pics Bildarchiv“[2]. In E-Pics Bildarchiv sind rund eine Million Bilder öffentlich zugänglich. Der größte Teil der Bilder kann seit Einführung der Open-Data-Policy im März 2015 kostenfrei in verschiedenen Auflösungen heruntergeladen werden.[3]

2. Der Erstdurchlauf mit der Computer-Vision-Software „Clarifai“

Seit 2021 setzt das ETH-Bildarchiv Künstliche Intelligenz (KI)[4] ein, um die Bilderschließung zu automatisieren.[5] Ziel ist es, die intellektuelle Verschlagwortung mittels Schlagwörtern zu ergänzen und diese nicht zu ersetzen. Im ersten Durchlauf zwischen Februar 2021 und Mai 2022 wurden die über eine Million Bilder, die in Canto Cumulus verwaltet werden, mit der neu integrierten Computer-Vision-Software „Clarifai“ getaggt. Das heißt, Objekte wurden in den Bildern erkannt und mit Keywords klassifiziert. Für die vorwiegend dokumentarisch-wissenschaftlichen Fotografien wurde jeweils das Clarifai-Modell „General“ verwendet. Die anderen, spezifischeren Clarifai-Modelle „Wedding“, „Travel“, „Food“ und „Apparel“ wurden nicht benutzt. Bei Farbbildern wurden zusätzlich die maximal acht möglichen Hauptfarben („Colors“) extrahiert. Die Autotagging-Möglichkeiten der „Classification“ von sensiblen Inhalten oder die Gesichtserkennung „Face“ wurden nicht genutzt. Die Erkennungsrate,

1 Dieser Beitrag bezieht sich auf den Vortrag gleichen Titels der Autorin am 05.06.2024 auf der 112. BiblioCon 2024 in Hamburg.

2 Siehe https://ba.e-pics.ethz.ch, Stand: 19.07.2024. Das bisherige Digital Asset Management-System Canto Cumulus mit dem Webfrontend AWP wurde im August 2024 auf das DAM Sharedien/anura der Firma Advellence migriert.

3 Graf, Nicole: Bilder werden frei verfügbar. Open Data im Bildarchiv, in: Memoriav-Bulletin 22 (10), 2015, S. 28–30, https://doi.org/10.3929/ethz-b-000105881.

4 Im Weiteren wird der Begriff „KI“ teilweise synonym mit dem engeren Begriff „Computer Vision” verwendet.

5 Der vorliegende Aufsatz ist die inhaltliche Fortsetzung von: Graf, Nicole: Alles unter Kontrolle? KI im Einsatz im Bildarchiv der ETH-Bibliothek, in: o-bib. Das offene Bibliotheksjournal 10 (2), 2023, https://doi.org/10.5282/o-bib/5921.

das ist der Wert für die Wahrscheinlichkeit, dass die Keywords im Bild vorhanden sind, wurde auf 95 % gesetzt. Die berechneten Wahrscheinlichkeitswerte pro Bild und Keyword wurden nicht gespeichert. Die Speicherung der Daten – für jedes Keyword im Bild ein neues Feld – hätte eine größere technische Anpassung der Datenbank erfordert. Bei einem erneuten Durchlauf werden diese Werte nach Möglichkeit ebenfalls gespeichert werden.

Die initiale Berechnung der Keywords war – im Rahmen der zur Verfügung stehenden Infrastruktur – rechenintensiv und hätte 340 Tage im 24-Stunden-Betrieb erfordert. Da die Anwendung auf den Produktionsservern die Performance negativ beeinträchtigte, wurde die KI außerhalb der Bürozeiten betrieben.[6] Der finanzielle Aufwand ist hingegen gering. Ein Desiderat ist die Berechnung des CO2-Ausstoßes, der durch diese enorme Rechenleistung verursacht wurde.

Mit dem ersten Durchlauf hat Clarifai eine alphabetische Liste von 4.573 Keywords von „Aasfresser" bis „Zypressenholz" erzeugt. Die Autotagging-Sprache wurde auf Deutsch eingestellt. Im Durchschnitt wurden 9,4 Keywords pro Bild vergeben. Bei der intellektuellen Beschlagwortung durch das Fachpersonal werden hingegen in der Regel rund fünf Schlagwörter pro Bild vergeben: Jedes Bild erhält zwingend ein Genre-Schlagwort, ein Ortsschlagwort und ein bis drei Sachschlagwörter. Der hierbei zum Tragen kommende, hierarchisch aufgebaute Schlagwortbaum enthält neun Hauptgenres mit sieben Untergenres, 11.212 Ortsschlagwörter, 2.418 Personenschlagwörter sowie 1.843 Sachschlagwörter.

3. Kritik an den Keywords

Die KI-generierten Keywords im Bildarchiv der ETH-Bibliothek sorgen für wachsende Kritik:

Kritikpunkt 1: Verwässerung der intellektuellen Arbeit

Mitarbeiter*innen des Bildarchivs kritisieren, dass die KI-generierten Keywords ihre intellektuelle Arbeit verwässern würden. Die Expertise und das Fachwissen des Fachpersonals würden durch die Algorithmen nicht ausreichend abgebildet. Zudem sei der Unterschied zwischen den automatisch erstellten Keywords und den vom Fachpersonal vergebenen Schlagwörtern im Web-Frontend nicht eindeutig gekennzeichnet. Dies kann zu Missverständnissen und Fehlinterpretationen führen.

Dieses Problem könnte durch eine klare Kennzeichnung und Abgrenzung zwischen KI-generierten Keywords und manuell vergebenen Schlagwörtern im Web-Frontend gelöst werden. Dies würde jedoch eine Anpassung der Software erfordern.

6 Das Autotagging mit Clarifai wurde vom 21. Februar 2021 bis 16. Mai 2022 jeweils ab 17 Uhr abends durchgeführt. Das nachfolgende Autotranslate mit DeepL wurde vom 15. März bis 23. Dezember 2022 jeweils ab 22 Uhr abends durchgeführt, da der Ressourcenaufwand so groß war, dass auch die Nutzer*innen im Web-Frontend dadurch beeinträchtigt wurden. Wenn beide Tools liefen, musste die Anzahl Datensätze für Clarifai von 10.000 auf 2.000 reduziert werden, damit der Clarifai-Batch um 22 Uhr beendet war und DeepL gestartet werden konnte. Siehe Graf: Kontrolle, 2023.

Kritikpunkt 2: Reproduktion von Stereotypen und Vorurteilen

Im Weiteren wird von Nutzer*innen kritisiert, dass Clarifai Stereotype und Vorurteile im Datensatz des Bildarchivs reproduziert hat. Die Clarifai-Algorithmen basieren auf dem großen ImageNet-Datensatz[7], der nachgewiesenermaßen verzerrte Darstellungen enthält.[8] Dies kann zu einer diskriminierenden und unfairen Beschlagwortung von Bildern führen. Problematische Beispiele sind etwa:

- „sexy" mit Abbildungen insbesondere von Frauen in Bikinis, Damenunterwäsche, Handschellen oder Plateauschuhen, aber auch von Bodybuildern usw. (45 Treffer);
- „hübsch" mit klassischen Porträtbildern vor allem von Frauen (176 Treffer) und „gutaussehend" als Äquivalent für Männer (1.059 Treffer);
- „Intelligenz" mit klassischen Porträtbildern von Frauen und Männern (355 Treffer).

Aufgrund dieser beiden Kritikpunkte und der Tatsache, dass das Bildarchiv mit über einer Million getaggten Bildern eine einzigartige Möglichkeit für eine vertiefte qualitative Analyse bietet, wurde entgegen der ursprünglichen Planung eine Qualitätskontrolle aller Bilder durchgeführt.

4. Qualitätskontrolle

Mangels eigener Groundtruth-Daten und um die Qualität der KI-gestützten Klassifikation zu verbessern, wurde der Datenpool des Bildarchivs mit über einer Million getaggten Bildern einer qualitativen Analyse unterzogen. Grundlage für die Qualitätskontrolle waren die Keywords aus dem ersten Durchlauf. Die visuelle Analyse der 4.573 Keywords erfolgte im Zeitraum von Juli 2023 bis Februar 2024. Der Aufwand betrug ca. 70 Arbeitsstunden. Das Autotagging neuer Bilder wurde in diesem Zeitraum ausgesetzt und danach wieder aufgenommen.

4.1 Zusammenführen von Dubletten

Zunächst wurden automatisch generierte, dublette Keywords zusammengeführt. Dubletten entstanden durch Groß- und Kleinschreibung eines Keywords (z.B. „baum" und „Baum"), durch Schreibweise mit scharfem S (ß) und doppeltem S (z.B. „Floss" und „floß") oder durch einfache Verdoppelung der Keywords (z.B. „Baum" und „Baum").

Die Bilder, die mit diesen dubletten Keywords beschlagwortet waren, waren inhaltlich nicht immer deckungsgleich, so dass die Tags auf Datensatzebene zusammengeführt wurden. Bei Keywords wie „Baum" (59.000 bzw. 66.000 Treffer) oder „Natur" (57.000 bzw. 700.000 Treffer) erfolgte die

7 Clarifai Marks 10th Anniversary Launching the 1st Full Stack Generative AI Platform, 20.11.2023, https://www.newswire.ca/news-releases/clarifai-marks-10th-anniversary-launching-the-1st-full-stack-generative-ai-platform-827037285.html, Stand: 19.07.2024.

8 Kantayya, Shalini: Coded Bias. Vorprogrammierte Diskriminierung, Dokumentarfilm, 86 min, Netflix, 05.04.2021, Online: https://www.netflix.com/title/81328723, Stand: 08.10.2024; weitere Informationen zum Film und den Hintergründen unter https://www.ajl.org/spotlight-documentary-coded-bias, Stand: 19.07.2024; Social Media Collective: Critical algorithm studies. A Reading List, 15.12.2016, https://socialmediacollective.org/reading-lists/critical-algorithm-studies, Stand: 19.07.2024; Webseite Algorithmic Justice League, https://www.ajl.org, Stand: 19.07.2024.

ressourcenintensive Zusammenführung der Datensätze jeweils nur nachts oder am Wochenende. Eine Blacklist der betroffenen Keywords wurde im System hinterlegt. Die ursprüngliche Liste enthielt 4.573 Keywords. Davon wurden 647 dublette Keywords (14 % der Gesamtzahl) gelöscht, so dass 3.926 Keywords (100 % bereinigte Gesamtzahl) übrigblieben.

4.2 Visuelle Analyse der Keywords

Die Hauptarbeit bestand in der visuellen Analyse der 3.926 verbleibenden Keywords. Dazu wurde jedes Keyword aufgerufen und die Treffer visuell geprüft, d.h. es wurde geprüft, ob die Bildinhalte korrekt getaggt wurden. Bei Keywords mit weniger als 100 Treffern (2.686 Keywords, 68 % der bereinigten Gesamtzahl) wurde exakt ausgezählt, wie viele Bilder korrekt bzw. falsch getaggt wurden. Bei Keywords mit mehr als 100 Treffern wurde, wenn eine Zählung zu aufwändig gewesen wäre, eine Schätzung vorgenommen.

Nach dieser Analyse wurde jeweils entschieden, ob ein Keyword behalten oder gelöscht wurde. In der Regel wurden Keywords aus einem oder mehreren der folgenden Gründe gelöscht:

- Quantitative Gründe: zu hoher Anteil an Falschtreffern, zu geringe Anzahl an Treffern.
- Qualitative Gründe: nicht-ETH-relevante Themen, „problematische“ Keywords (z.B. stereotypisierend).

Um einen ersten Eindruck zu vermitteln, um welche Art von Keywords es sich handelt, sind in Tabelle 1 die Keywords mit den meisten Treffern, also alle Keywords mit mehr als 100.000 Treffern, in absteigender Reihenfolge mit ihrem geschätzten prozentualen Anteil an korrekten Zuordnungen aufgelistet.

Tab. 1: Keywords mit mehr als 100.000 Treffern

Keywords	Anzahl Treffer ↓	korrekt in %
Menschen	366.167	74 %
Reise	337.731	89 %
Fahrzeug	235.355	85 %
Landschaft	225.722	100 %
Verkehrssystem	211.125	99 %
Erwachsener	173.716	95 %
Straße	168.166	95 %
Natur	159.738	99 %
Gewässer	149.538	100 %
Stadt	144.926	99 %
Gebäude	138.723	99 %
Mann	126.609	95 %

Keywords	Anzahl Treffer ↓	korrekt in %
Baum	125.659	8 %[9]
Behausung	123.164	100 %
Architektur	119.000	100 %
Berg	110.656	99 %

4.2.1 Visuelle Analyse: Was die KI kann

Erste Stichproben haben gezeigt, dass die Qualität der Ergebnisse der automatischen Erkennung sehr unterschiedlich ist. So gibt es Keywords mit nahezu 100 %-er Treffersicherheit, wie z.B. „Schlagzeuger" oder „Abenddämmerung", die eine zusätzliche inhaltliche Erschließung darstellen, da vergleichbare Schlagwörter im Bildarchiv nicht vorhanden sind. Die KI ist sehr mächtig bei sogenannten „ikonischen Bildern" wie von „Hunden" und „Katzen". „Ikonisch" bedeutet in diesem Zusammenhang eine eindeutige und unverwechselbare Darstellung einer Objektkategorie in einem Bild – zum Beispiel ein Bild, das verwendet werden könnte, um einem Kind eine bestimmte Kategorie wie „Fahrrad" oder „Leuchtturm" zu vermitteln.[10] Was passiert aber nun jenseits dieser ikonischen Bilder?

Am Beispiel der Keywords „Landschaft" und „Schlucht" kann gezeigt werden, wo die KI Ergebnisse liefert, die durch eine visuelle Prüfung bestätigt werden konnten. Das Keyword „Landschaft" liefert mit 225.722 Bildern sehr viele, auch sehr generische Treffer, von denen 42 % Luftbilder sind. Dagegen liefert das Keyword „Schlucht" mit 562 Treffern wenige, dafür aber präzisere Ergebnisse mit einem hohen Anteil von 93 % korrekt getaggten Bildern. Im Vergleich dazu liefert das intellektuell durch Fachpersonen vergebene Schlagwort „Schluchten" 799 Treffer.[11] Je spezifischer die Keywords sind, desto bessere Treffer werden erzielt: In der Sachgruppe Landschaft sind dies z.B. „Einschlagkrater" (45 Treffer, 100 % korrekt), „Geysir" (130, 100 % korrekt), „Ghetto" (17, 100 % korrekt), „Grube" (47, 100 % korrekt), „Lagune" (19, 100 % korrekt), „Moor" (26, 100 % korrekt), „Müllhalde" (60, 100 % korrekt) oder „Oase" (23, 100 % korrekt). Die aufgeführten Keywords wurden alle beibehalten, auch wenn einige nur wenige Treffer aufweisen. Gründe dafür sind der inhaltliche Mehrwert und die genaue Muster- und Objekterkennung.

4.2.2 Visuelle Analyse: Was die KI nicht kann

Ausgehend von der Beobachtung, dass die KI auch Keywords erzeugt, die schlicht falsch sind, wurde diese qualitative, visuelle Analyse des Gesamtbestandes vorgenommen. Bei den fehlerhaften Zuordnungen gibt es Beispiele mit klar erkennbaren Mustern, bei denen das jeweilige Keyword jedoch keinen Sinn ergibt. Ein solches Beispiel, das immer wieder für Heiterkeit sorgt, ist „Béchamelsauce".

9 Darin enthalten sind über 102.000 Luftbilder, Architekturfotografien, Landschaftsansichten und Ortsansichten mit Bäumen als peripheres „Beiwerk". Die 8 % korrekten Bilder beziehen sich auf Bilder mit erkennbaren Bäumen.

10 Vgl. dazu auch die ersten Auswertungen der Bildarchiv-Daten bei Graf: Kontrolle, 2023. Zum Konzept der ikonischen Bilder: Berg, Tamara L.; Berg, Alexander C.: Finding Iconic Images, in: 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Miami 2009, S. 1–8, https://doi.org/10.1109/cvprw.2009.5204174.

11 Schlagwörter sind im Plural angesetzt, Keywords im Singular. Dies ist ein Zufall, der nun bei gleichen Begriffen die Unterscheidung erleichtert.

Das Keyword wurde für 69 Bilder vergeben, und zwar ausschließlich für Bilder von schneebedeckten Bergen und Gipfeln. Das heißt, es enthält Bilder mit einem klar erkennbaren visuellen Muster. Das Keyword wurde gelöscht. Ein vergleichbares Schlagwort ist nicht vorhanden. Gelöschte Keywords wurden auf eine Blacklist gesetzt.

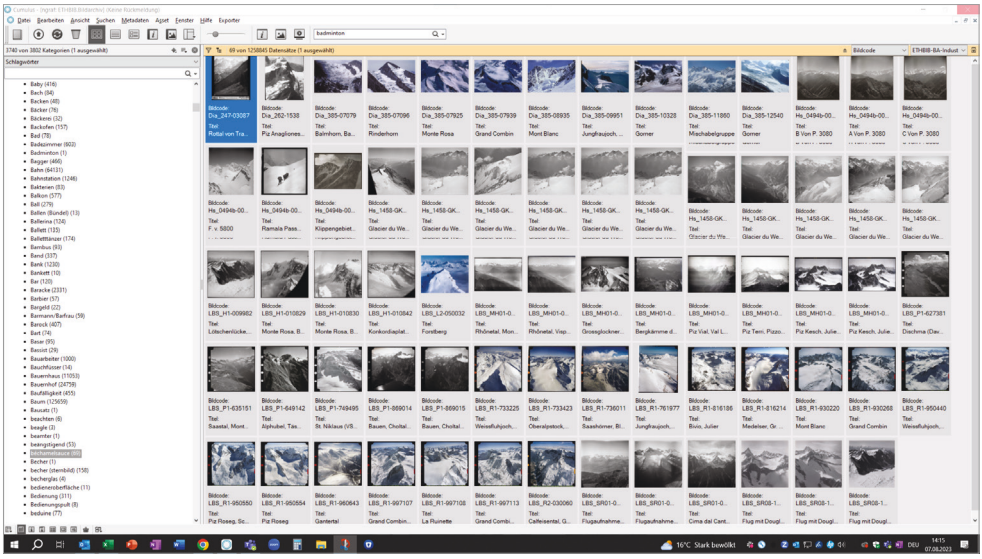

Abb. 1: Miniaturansicht der Treffer von „Béchamelsauce" im Datenbank-Backend Cumulus (Screenshot)

Eine weitere Kategorie von Fehlzuordnungen sind solche, die nur indirekt mit den Bildinhalten zusammenhängen. Dazu gehören beispielsweise 21 Architekturbilder des Kolosseums in Rom oder von Amphitheatern, die mit „Gladiator" getaggt wurden. Ein ähnliches Beispiel ist „Sopran" mit 86 Treffern, die Ausschnitte von Landkarten zeigen, die vermutlich als Musiknoten identifiziert wurden. Ähnliche Bilder wurden mit dem Tag „Notenschlüssel" (16 Treffer) gefunden. Bei allen drei Tags lag die Quote der korrekten Zuordnungen bei 0 %, die drei Tags wurden gelöscht. Vergleichbare Schlagwörter sind nicht vorhanden.

Eine weitere Kategorie mit problematischen Zuordnungen sind Keywords, die Vorurteile und Stereotype reproduzieren. Das Autotag „Ritual" zum Beispiel hat 21 Treffer, von denen weniger als die Hälfte als gelungen eingestuft werden können. Auffällig ist jedoch, dass die Treffer überwiegend Szenen von Menschenansammlungen im „Globalen Süden" zeigen, z.B. Wäscherinnen, Kinder beim Haarewaschen am Brunnen oder allgemeine Marktszenen. Szenen mit Ritualen aus dem „Globalen Norden", etwa Beerdigungen oder Gottesdienste, fehlen gänzlich. Zum Vergleich: Beim Schlagwort „Menschen im Kontext" gibt es das Unterschlagwort „Religion (religiöse Handlung)" mit 1.012 Treffern, das genau solche Szenen enthält. Das Autotag „Ritual" wurde entfernt.

Abb. 2: Miniaturansicht der Treffer von „Sopran“ im Datenbank-Frontend E-Pics Bildarchiv (Screenshot)

4.2.3 Visuelle Analyse: Was die KI kann und trotzdem gelöscht wurde

Es gibt verschiedene Keywords mit korrekten Bildzuordnungen, die dennoch gelöscht wurden. Dabei sind insbesondere folgende zwei Typen zu unterscheiden: Einerseits Tags mit einer zu geringen Treffermenge (≤ 10 Treffer) und/oder andererseits Tags mit nicht-ETH-relevanten Inhalten. Beispielsweise haben 764 Autotags (19,5 % der bereinigten Gesamtzahl) eine Quote an korrekten Zuordnungen von 100 %, davon sind aber 343 Tags mit 10 oder weniger Treffern (8,7 % der bereinigten Gesamtzahl). Zum Vergleich: Schlagwörter werden erst ab einer bestimmten Anzahl von Bildern, üblicherweise etwa 50, neu gesetzt.

„Dicht“ ist ein Beispiel für die erste Art: sechs korrekte Angaben mit Großstadt- und Dschungelansichten. Hier hätte es ein Vielfaches an Angaben geben müssen. Ein vergleichbares Schlagwort gibt es nicht. Das Keyword wurde gelöscht.

„Romanze“ oder „schön“ sind zwei Beispiele mit nicht-ETH-relevanten Inhalten und problematischer Zuordnung. Das Keyword „Romanze“ liefert 67 Treffer, darunter Bilder von Rosen, Pralinen, Edelsteinen oder zwei nahe beieinanderstehenden Menschen. „Schön“ liefert 2.839 Treffer, darunter (nicht nur idyllische) Postkartenbilder, (eher zufällig „ausgewählte“) Landschafts- und Ortsansichten, viele Schwarz-Weiß-Bilder. Als erkennbares Muster kämen allenfalls (Urlaubs-)Orte in Frage, und die Frage nach dem „Schönen“ wäre damit noch nicht beantwortet. Eine Schätzung der gelungenen Zuordnungen wurde nicht vorgenommen. Die Keywords wurden gelöscht.

4.2.4 Visuelle Analyse: Wo die KI unterstützt wurde

Schließlich gibt es noch eine weitere Kategorie von Keywords, die bearbeitet wurden: Keywords mit offensichtlichen Übersetzungsfehlern wurden umbenannt. Inhaltlich ähnliche Autotags wurden zusammengeführt.

Das Autotag „Erleichterung" spiegelt ein eindeutiges Objektmuster wider: Von den 435 Bildern sind 92 % Wandreliefs. „Relief" bedeutet im Englischen sowohl Relief als auch Erleichterung. In der Keywordliste wurde also der falsche Begriff für das Objektmuster Relief hinterlegt. Das Autotag wurde zu „Relief" korrigiert. Insgesamt gibt es 27 Keywords (0,7 % der bereinigten Gesamtzahl) mit eindeutigen Übersetzungsfehlern bzw. ohne Übersetzung, die korrigiert wurden. Weitere Beispiele sind „Wurf" (113 Treffer, 100 % korrekt), übersetzt von „litter", neu übersetzt mit „Müll", oder „Kämpfer" (445 Treffer, 90 % korrekt), übersetzt von „fighter", neu übersetzt mit „Kampfflugzeug".

4.2.5 Visuelle Analyse: Auswertungen

Über die Hälfte der Autotags wurden nach der Analyse der knapp 4.000 Tags gelöscht: das sind 2.179 der 3.926 Autotags (55 % der bereinigten Gesamtzahl). Nach der visuellen Analyse verbleiben also noch 1.747 Autotags (45 % der bereinigten Gesamtzahl) in der Bilddatenbank.

Schaut man sich die Spitzen bei der Quote der korrekten bzw. nicht korrekten Zuordnungen an, ergibt sich folgendes Bild:

- 0 % korrekte Zuordnung:
  - 1.358 Autotags (34,6 % der bereinigten Gesamtzahl),
  - davon wiederum mit ≤ 10 Treffern: 851 Autotags (62,6 % bzw. 21,7 % der bereinigten Gesamtzahl).
  - Alle Tags mit Trefferquote 0 % wurden gelöscht.

- 100 % korrekte Zuordnung:
  - 764 Autotags (19,5 % der bereinigten Gesamtzahl),
  - davon mit ≤ 10 Treffern 343 Autotags (44,9 % bzw. 8,7 % der bereinigten Gesamtzahl).
  - 289 Tags (37,8 % bzw. 7,3 % der bereinigten Gesamtzahl) mit Trefferquote 100 % wurden gelöscht.[12]

Bei kleinen Treffermengen sieht die Verteilung wie folgt aus:

- ≤ 10 Treffer
  - Insgesamt gab es 1.484 Autotags,
  - davon wurden 1.321 Autotags (89,0 % bzw. 33,6 % der bereinigten Gesamtzahl) gelöscht.

12 Davon enthielten allein 222 Keywords weniger als 10 Treffer.

- 11 bis 100 Treffer
  - Insgesamt gab es 1.199 Autotags,
  - davon wurden 536 Tags (44,7 % bzw. 13,6 % der bereinigten Gesamtzahl) gelöscht.

In Tabelle 2 finden sich aggregierte Auswertungen zu den Trefferquoten.

Tab. 2: Analyse nach Trefferquoten

Trefferquote	Anzahl Keywords	Keywords nach Analyse ok	Gelöscht in %
0–25 %	1.576	42	97 %
26–50 %	224	62	72 %
51–75 %	253	173	32 %
76–100 %	1.873	1.470	22 %
Summe	3.926	1.747	55 %

4.3 Auswertung der Keywords nach Sachgruppen

Neben der visuellen Auswertung wurden die Keywords „von Hand“ in Sachgruppen eingeteilt. Das Ergebnis sind 24 Sachgruppen. In Tabelle 3 sind die Sachgruppen, absteigend sortiert nach dem prozentualen Anteil der korrekten Keywords, aufgelistet.

Tab. 3: Einteilung in Sachgruppen, absteigend sortiert nach dem prozentualen Anteil der korrekten Keywords

Sachgruppe	Keywords	OK	in % ↓
Landschaft	91	72	79%
Architektur	319	216	68%
Verkehrsmittel	110	72	65%
Musik	21	12	57%
Tier	248	148	60%
Pflanze	171	98	57%
Gestein	25	14	56%
Wetter	38	21	55%
Chemie	33	17	52%
Mode	82	41	50%
Sport	57	28	49%
Objekt	718	341	47%
Natur	46	21	46%
Geografie	33	15	45%
Menschen	251	114	45%
Computer	42	16	38%

Sachgruppe	Keywords	OK	in % ↓
Wissenschaft	45	17	38%
Mythologie	22	8	36%
Körper	66	21	32%
Food	154	48	31%
Tätigkeit	76	23	30%
Konzept	1.241	373	30%
Universum	25	7	28%
Ethnie	12	2	17%

Im Folgenden werden fünf Sachgruppen näher betrachtet. Es werden jeweils die ersten 15 Keywords, die nicht gelöscht wurden, mit der Anzahl der Treffer sowie der Quote der korrekten Zuordnungen aufgeführt.

- Architektur: „Abtei“ (538 Treffer, 43 % korrekt), „Abwasserkanal“ (19, 100 %), „Allee“ (23, 65 %), „Altar“ (629, 100 %), „am Wasser“ (476, 100 %), „Amphitheater“ (113, 80 %), „Architektur“ (119.000, 100 %), „Auditorium“ (211, 100 %), „Aufzug“ (153, 100 %), „Ausgang“ (8, 100 %), „außerhalb“ (1.267, 95 %), „Autobahn“ (5.972, 33 %), „Bad“ (69, 87 %), „Badezimmer“ (238, 97 %), „Bahnstation“ (1.246, 80 %) usw.

- Pflanzen: „Agave“ (39, 85 %), „Ahorn“ (53, 100 %), „Aloe“ (12, 25 %), „Ananas“ (6, 83 %), „Apfel“ (48, 79 %), „Apfelbaum“ (11, 64 %), „Artischocke“ (2, 100 %), „Bambus“ (93, 97 %), „Baum“ (125.659, 8 %), „Beere“ (44, 91 %), „Birke“ (61, 90 %), „blühend“ (654, 100 %), „Blume“ (5.115, 100 %), „Blumenarrangement“ (159, 100 %), „Blumenstrauß“ (43, 100 %) usw.

- Gestein: „Aluminium“ (176, 56 %), „Amethyst“ (25, 60 %), „Edelstein“ (142, 99 %), „Felsbrocken“ (53, 94 %), „felsig“ (1.037, 96 %), „Granit“ (473, 99 %), „Kalkstein“ (172, 93 %), „Kies“ (1.378, 87 %), „Marmor“ (843, 95 %), „Megalith“ (22, 100 %), „Quarz“ (127, 88 %), „Sandstein (Geologie)“ (221, 95 %), „Stalaktit“ (133, 75 %), „Stein“ (13.973, 72 %)

- Konzept: „Abend“ (1.991, 100 %), „Abenteuer“ (4.394, 64 %), „Abfahrt“ (43, 100 %), „Abgabetermin“ (18, 72 %), „Abgenutzt“ (253, 47 %,) „Abreise“ (1.683, 100 %), „Abriss“ (2.593, 94 %), „Abschlag“ (39, 33 %), „Abstraktion (Philosophie)“ (121, 100 %), „Absturz“ (44, 11 %), „Abwechslung“ (15.686, 100 %), „Abwesenheit“ (335, 100 %), „Aggression“ (43, 100 %), „Aktion“ (3.167, 99 %), „aktiv“ (65, 100 %) usw.

- Food: „Abendessen“ (250, 68 %), „Alkohol“ (43, 100 %), „Anbauen“ (1.513, 73 %), „aromatisch“ (11, 91 %), „Backen“ (48, 100 %), „Bier“ (130, 69 %), „Brot“ (92, 71 %), „Büfett“ (10, 100 %), „cremig“ (9, 100 %), „Croissant“ (6, 83 %), „Espresso“ (18, 33 %), „essbar“ (26, 96 %), „Fleisch“ (820, 98 %), „Frühstück“ (123, 98 %), „Gebäck“ (21, 95 %) usw.

Wie oben gezeigt, kann es vorkommen, dass die Trefferzahl unter 100 Bildern liegt, das Keyword aber trotzdem erhalten bleibt. In der Regel handelt es sich dabei um Tags, die einen Mehrwert bieten, sehr genaue Muster- und Objekterkennungen enthalten und/oder um Tags mit ETH-relevanten Inhalten. In Ausnahmefällen wurden auch Keywords mit einer geringen Quote korrekter Zuordnungen beibehalten, um an diesen Beispielen die Arbeitsweise von Clarifai zu veranschaulichen. So enthält „Abschlag" (für Golfspielen) auch Luftbilder von vermuteten Golfplätzen (es sind Dörfer inmitten einer parzellierten Landschaft) oder „Absturz" (Flugzeugabsturz) enthält auch Autofriedhöfe.

5. Fazit

Computer Vision bzw. Künstliche Intelligenz kann die intellektuelle Erschließung durch Fachpersonal ergänzen, indem große Datenmengen automatisch mit Keywords versehen werden. Wenn in kleinen Archiven aus Ressourcengründen keine intellektuelle Erschließung erfolgen kann und nur wenige Metadaten vorhanden sind, kann KI eine erste wichtige Erschließungshilfe leisten und die Auffindbarkeit von Bildern sicherstellen.

KI ersetzt noch nicht die intellektuelle Erschließung durch Fachpersonal. Wie gezeigt werden konnte, können die Computer-Vision-Modelle keine bildinhärenten Kontextualisierungen und Abstrahierungen vornehmen. KI kann zu falschen, problematischen oder irreführenden Verschlagwortungen führen, Vorurteile und Stereotype reproduzieren und allenfalls die Wahrnehmung der intellektuellen Arbeit des Fachpersonals verwässern. Wie die qualitative Analyse mit der Löschung von 55 % der Keywords zeigt, ist eine Kontrolle der von der KI generierten Tags weiterhin notwendig, um die Qualität der Bilderschließung zu gewährleisten. Ideal wäre die Einrichtung einer Feedbackschleife zur kontinuierlichen Verbesserung der KI-Ergebnisse – ebenso die Zusammenarbeit mit anderen Institutionen, um Erfahrungen und Wissen auszutauschen. KI wird in Zukunft eine noch wichtigere Rolle in der Bilderschließung spielen. Es ist wichtig, die Möglichkeiten und Grenzen von KI zu kennen, um sie optimal nutzen zu können.

Im Bildarchiv der ETH-Bibliothek wird 2024 das DAM-System Cumulus durch das System Sharedien abgelöst. Zwischen April und August 2024 fand die Datenmigration statt, d.h. es wurden keine Metadaten erfasst und keine Keywords erstellt. Im neuen System sind auch KI-Systeme integriert, die das Bildarchiv ebenfalls nutzen wird, sobald die klassischen Workflows neu aufgesetzt sind. Was dies für die bereits getaggten Bilder bedeutet, muss noch getestet werden.

Nicole Graf, Bibliothek der Eidgenössischen Technischen Hochschule Zürich,
ORCID: https://orcid.org/0000-0003-2230-6679

Zitierfähiger Link (DOI): https://doi.org/10.5282/o-bib/6090