



DRESDEN LEIPZIG

CENTER FOR SCALABLE DATA ANALYTICS
AND ARTIFICIAL INTELLIGENCE

Topic : Aktuelle OCR-Methoden für Handschriften
und Etiketten im deutschsprachigen Kontext

Speaker : Dr. Thomas Burghardt

Funded by



Federal Ministry
of Research, Technology
and Space



SACHSEN Diese Maßnahme wird gefördert durch die Bundesregierung
aufgrund eines Beschlusses des Deutschen Bundestages.
Diese Maßnahme wird mitfinanziert durch Steuermittel auf
der Grundlage des von den Abgeordneten des Sächsischen
Landtags beschlossenen Haushaltes.



Dresden
University of
Technology



UNIVERSITÄT
LEIPZIG

Einführung

Kurzvorstellung



- Dr. Thomas Burghardt
- Wissenschaftlicher Mitarbeiter am ScaDS.AI, Teil der zentralen Einrichtung Data Science Center (DSC) der Universität Leipzig
- Leitung Reallabor (Living Lab) in Leipzig, Lehre & Forschung in den Bereichen Datenwissenschaften und Künstliche Intelligenz



Living Lab ScaDS.AI

Living Lab ScaDS.AI Leipzig

Das Living Lab ScaDS.AI Leipzig ist eine offene Innovationsplattform. Es bietet Unternehmen, Forschenden und Studierenden die Möglichkeit, gemeinsam an der Entwicklung und Erprobung von KI- und Datenwissenschaftslösungen zu arbeiten. Wir fördern den Austausch von Ideen, die Zusammenarbeit und die praktische Anwendung von Technologien in realen Szenarien.

Demonstrators Development



Co-designing scientific demonstrators

Events / Exhibitions



Exhibiting the latest achievements in the field of AI

Trainings



Teaching required skills to enable active engagement of the public in the field of AI

School Programs



Delivering AI related lectures to pupils and familiarizing them with the benefits and challenges of latest AI technologies

Living Lab ScaDS.AI

Public Lectures



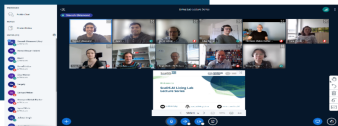
Presenting scientific topics in simplified language

Hands-on Sessions



Providing state-of-the-art achievements with an accessible and tangible approach

Online Lecture Series



Monthly online lectures; providing first hand insights in an easy and understandable language

Hackathon



Opportunity to learn and practice the state-of-the-art in machine learning in a friendly collaborative environment

Motivation

Motivation - um was geht es?

Möglichkeiten der Automatisierten Informationsgewinnung aus handschriftlichen Texten und industriellen Produkt Etiketten mittels klassischen OCR-Verfahren¹ und Bild-Text Modellen (engl. VLMs) niedrigschwellig aufzeigen.

Die Handschrift- und Etikettenerkennung sind nützlich für die Digitalisierung visueller Inhalte und der automatisierten Gewinnung strukturierter Informationen (z.B. Titel, Autor, Datum) aus unstrukturierten visuellen Datenquellen. Anwendungsgebiete:

- Personalakten und historische Archivadokumente
- Prüfprotokolle und Wartungsberichte
- Produktions- und Qualitätskontrollformulare
- Industrielle Etiketten mit handschriftlichen Anmerkungen
- Sitzungsprotokolle und Randbemerkungen

¹OCR (Optical Character Recognition) ist eine Technologie, die Text in Bildern, Scans oder Fotos automatisch erkennt und in maschinenlesbaren Text umwandelt (also von Pixel nach Text (z.B. Unicode (UTF-8))). Dadurch können beispielsweise Dokumente durchsucht, analysiert oder von Sprachmodellen weiterverarbeitet werden.

Motivation

Motivation - Beispiele

1. Unterschriebene Formulare und Anträge

Antrag auf Genehmigung

Antragsteller/in

Name, Vorname: Müller, Anna

Adresse: Beispielstraße 12
12345 Musterstadt

Antrag vom: 15.04.2024

Antragsgegenstand:

Genehmigung zur Veranstaltung
am 01.06.2024 im Stadtpark

Begründung:

Die Veranstaltung dient einem
gemeinnützigen Zweck und
ist öffentlich zugänglich.

Unterschrift: A. Müller

Datum: 15.04.2024

2. Prüfprotokolle und Wartungsberichte

WARTUNGSBERICHT

Anlage / Maschine: Presse 23

Datum: 17.05.2024

Wartungsart: Inspektion

Durchgeführt von: T. Schneider

Prüfpunkt	Status	Bemerkungen
Ölstand prüfen	<u>i.O.</u>	<u>nachgefüllt</u>
Feder prüfen	<u>i.O.</u>	<u>-</u>
Schraubenverbindungen kontrollieren	<u>i.O.</u>	<u>-</u>
Sicherheitseinrichtungen prüfen	<u>n.i.O.</u>	<u>Schalter defekt, austauschen</u>
Funktionstest	<u>i.O.</u>	<u>-</u>

Weitere Bemerkungen:

Nächste Wartung in 6 Monaten.

Unterschrift: T. Schneider

3. Industrielle Etiketten mit handschriftlichen Anmerkungen

Artikel-Nr.: **A12345**

Bezeichnung: **Hydraulikpumpe**

Serien-Nr.: **SN 987654**

Produktionsdatum: **12.03.2024**

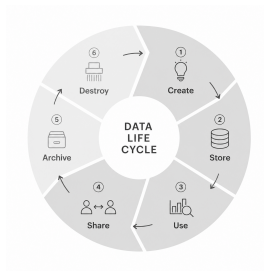
Barcode: 4 260123 456789

*OK
geprüft
13.03.
M.K.

Lager
Halle 3
Regal 12*

Struktur des Vortrages (Agenda)

Datenlebenszyklus



Der Data Life Cycle (DLC) beschreibt den Lebenszyklus von Daten über ihre gesamte Existenz hinweg – von der Entstehung (Create) bis hin zur Archivierung und endgültigen Löschung (Destroy).

Der Vortrag ist entlang dieses Datenlebenszyklus strukturiert, um die verschiedenen Aspekte der Datenverwaltung und -nutzung im Bereich der Erkennung von Handschriften zu beleuchten.

Struktur des Vortrages

Datenlebenszyklus

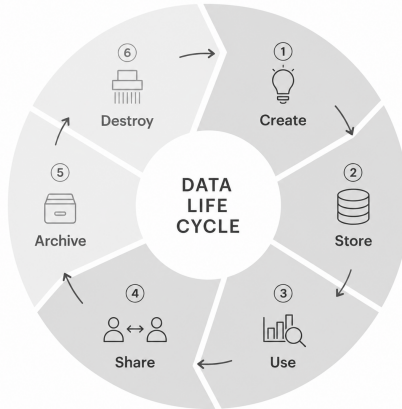
- **Planung & Erstellung (Generation):** Festlegung, welche Daten benötigt und wie sie erhoben werden (z. B. durch Messungen, manuelle Eingabe oder Systemerfassung).
- **Speicherung & Verwaltung (Storage & Management):** Ablage der Daten in Datenbanken, Data Warehouses oder Cloud-Systemen. Sicherheitsmaßnahmen und Qualitätsprüfungen sind hierbei essenziell.
- **Verarbeitung & Analyse (Processing & Usage):** Aufbereitung der Daten und Nutzung für Geschäftsentscheidungen, maschinelles Lernen oder Berichte.
- **Teilen der Daten:** Zugänglich machen von Daten, Einhaltung von FAIR-Prinzipien²
- Archivierung (Archiving): Sichere und langfristige Aufbewahrung von Daten, die nicht mehr aktiv genutzt werden, aber aus Einhaltung der Vorschriften (Compliance) - oder historischen Gründen erhalten bleiben müssen.
- Löschung (Destruction): Endgültige und datenschutzkonforme Vernichtung oder Löschung der Daten, um Speicherplatz freizugeben und rechtliche Vorgaben (z. B. DSGVO) zu erfüllen.

²FAIR steht für Findable (Auffindbar), Accessible (Zugänglich), Interoperable (Interoperabel) und Reusable (Wiederverwendbar)

Daten erzeugen

Daten erzeugen

Datenlebenszyklus



Daten erzeugen

1. Deutsche Handschriften

- 775 handgeschriebene Wikipedia-Abschriften (je ca. 0,5 DIN A4, etwa 15 Zeilen und rund 220 Wörter pro Seite)
- 775 Schreibende aus dem deutschsprachigen Raum sind anonym (Crowd Sourcing Ansatz)
- Bachelorarbeit von Ahmad Alzin: 42.000 Wörter und Textzeilen als Bildmaterial mit Grundwahrheit (Ground Truth) extrahiert
- Einsatz eines proprietären Labeling Tools, um die Grundwahrheit zu extrahieren, d. h. die genaue Transkription der handgeschriebenen Texte zu erhalten
- Zeitliche Aufwand: ca. 100 Stunden für die Erstellung der Grundwahrheit, abhängig von der Handschrift und der Komplexität der Texte
- Labeling Tool erkennt die Wörter und Sätze automatisch, aber die manuelle Überprüfung und Korrektur ist notwendig, um die Genauigkeit der Grundwahrheit sicherzustellen (nächste Folie)
- Ziel: Bewertung der Leistungsfähigkeit von modernen OCR-Ansätzen (trOCR Transformer Modell) bei der Informationsgewinnung aus handgeschriebenen Dokumenten

1. Deutsche Handschriften aus ScaDS.AI Datensatz

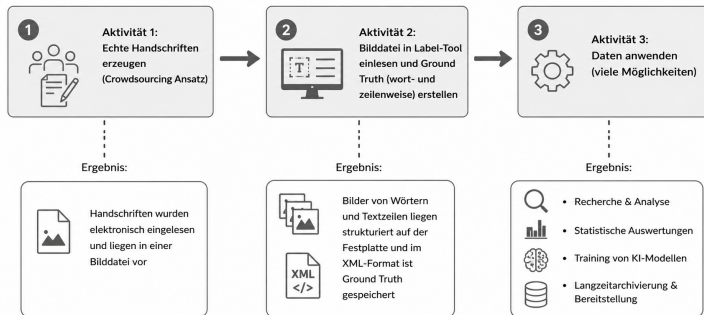
figure3: manuscript 027

figure4: manuscript 056

figure5: manuscript 066

Daten erzeugen

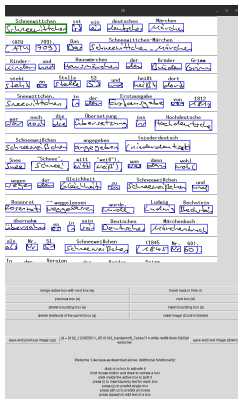
1. Deutsche Handschriften (Ablaufdiagramm)



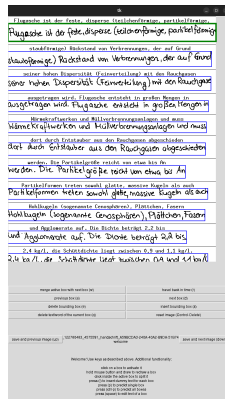
Ablaufdiagramm

Daten erzeugen

1. Deutsche Handschriften (Label Tool)



Annotation wortweise



Annotation zeilenweise

Daten erzeugen

1. Deutsche Handschriften (Export Label Tool)

Ein leichter, auch leichter, leichterschiff, Bording,

Fachmedienmarke mit mehreren Fachzeitschriften.

Wascheisich

der

Schneewittchen

Flugasche ist der feste, disperse (teilchenförmige, partikel förmige)

Exportierte Wörter und Zeilen

Daten erzeugen

1. Deutsche Handschriften (Export Label Tool)

```
<form id="0102">
  <handwritten-part>
    <line id="0102-000" x="15" y="25" width="618" height="32" text="Schneewittchen ist ein deutsches Märchen">
      <word id="0102-000-00" text="Schneewittchen" x="0" y="2" width="212" height="30"/>
      <word id="0102-000-01" text="ist" x="233" y="3" width="31" height="24"/>
      <word id="0102-000-02" text="ein" x="293" y="5" width="37" height="20"/>
      <word id="0102-000-03" text="deutsches" x="355" y="2" width="120" height="26"/>
      <word id="0102-000-04" text="Märchen" x="498" y="0" width="120" height="29"/>
    </line>
    <line id="0102-001" x="15" y="65" width="646" height="33" text="(ATU 709). Das Schneewittchen-Märchen">
      <word id="0102-001-00" text="(ATU" x="0" y="8" width="85" height="25"/>
      <word id="0102-001-01" text="709)." x="107" y="7" width="82" height="23"/>
      <word id="0102-001-02" text="Das" x="216" y="5" width="50" height="21"/>
      <word id="0102-001-03" text="Schneewittchen-Märchen" x="294" y="0" width="352" height="29"/>
    </line>
    <line id="0102-002" x="18" y="107" width="712" height="28" text="Kinder- und Hausmärchen der Brüder Grimm">
      <word id="0102-002-00" text="Kinder-" x="0" y="6" width="105" height="21"/>
      <word id="0102-002-01" text="und" x="117" y="6" width="55" height="17"/>
      <word id="0102-002-02" text="Hausmärchen" x="205" y="0" width="187" height="23"/>
      <word id="0102-002-03" text="der" x="420" y="2" width="53" height="19"/>
      <word id="0102-002-04" text="Brüder" x="507" y="3" width="103" height="23"/>
      <word id="0102-002-05" text="Grimm" x="626" y="3" width="86" height="25"/>
    </line>
  </handwritten-part>
</form>
```

Exportiertes Ground Truth (XML)³

³Export im proprietären IAM-Datensatzformat

Daten erzeugen

1. Deutsche Handschriften (Export Label Tool)

page_id	line_id	xml_file	line_file	line_idx	x	y	width	height	text
103	0103-001	0103.xml	0103-001.png	0	107	110	524	46	Nong Khiaw (Lao: ist ein Ort in der Provinz Luang Prabang
103	0103-002	0103.xml	0103-002.png	1	106	145	511	39	im Norden von Laos. Er liegt ca. 140 km nordöstlich der
103	0103-003	0103.xml	0103-003.png	2	103	173	512	44	Provinzhauptstadt Luang Prabang am Ufer des Flusses
103	0103-004	0103.xml	0103-004.png	3	107	204	488	43	Nam Ou. Nong Khiaw hat seit einigen Jahren für
103	0103-005	0103.xml	0103-005.png	4	108	233	460	35	den Tourismus an Bedeutung gewonnen und ist
103	0103-006	0103.xml	0103-006.png	5	109	257	519	50	zwischen Luang Prabang und der nördlichen Provinz
103	0103-007	0103.xml	0103-007.png	6	108	293	535	52	Phongsali eine beliebte Station für Individualreisende.
103	0103-008	0103.xml	0103-008.png	7	111	327	479	47	Ein Anziehungspunkt ist dabei der Bootsverkehr
103	0103-009	0103.xml	0103-009.png	8	112	359	433	47	auf dem Nam Ou, der für den einheimischen
103	0103-010	0103.xml	0103-010.png	9	107	384	488	49	Personen- und Warentransport nach wie vor eine
103	0103-011	0103.xml	0103-011.png	10	108	421	447	44	wichtige Rolle spielt. Durch den Bau eines
103	0103-012	0103.xml	0103-012.png	11	107	451	489	39	Staudamms ist der Bootsverkehr zwischen Nong
103	0103-013	0103.xml	0103-013.png	12	103	481	508	38	Khiaw und dem nördlich gelegenen Muang Khua
103	0103-014	0103.xml	0103-014.png	13	103	526	96	23	seit Ende

Exportiertes Ground Truth (CSV)

Daten erzeugen

1. Deutsche Handschriften

- Bachelorarbeit von Till Nestler: 77 ganzseitige handschriftliche Seiten als Bildmaterial mit Ground Truth extrahiert
- Ground Truth: nehme nicht das Wikipedia Original, sondern die tatsächliche Transkription der handgeschriebenen Seite, da sie die tatsächliche Handschrift repräsentiert, die von den Modellen erkannt werden soll.
- Einsatz eines Labeling Tools war nicht erforderlich, die 77 Texte wurden abgeschrieben (TXT-Datei)
- Zeitliche Aufwand: ca. 20 Stunden für die Erstellung der Grundwahrheit
- Ziel: Bewertung der Leistungsfähigkeit von Bild-Text-Modellen (VLM) bei der Informationsgewinnung aus handgeschriebenen Dokumenten

Daten erzeugen

1. Deutsche Handschriften (2 Beispiele)

Der Geschlechtseintrag dieses (von lateinisch dieses "ungleichartig, verschieden") bildet seit 2018 in Deutschland und seit 2019 in Österreich eine dritte Option neben "weiblich" und "männlich", die sich auf biologische Integrität/Einheit bezieht. Die Wahlmöglichkeit ist seit Ende 2021, inwieweit sich die Geschlechtsoption unabhängig von biologischen Eigenschaften allgemeiner auf eine nicht binäre Geschlechtsidentität beziehen kann. Vergleichbares gilt für Deutschland, gelten ebenfalls für Transsexuelle ohne Geschlechtseintrag in den Personalausweisen dieselben Regeln (siehe dritte Option im Transsexuellengesetz ab 2013). In Österreich gibt es bei unbekannten Geschlechtsgeschlechtseinführungswegen bei Indizgeschlechtlichkeit die Möglichkeit, den Geschlechtseintrag im Zentralen Personalausweisregister (ZPA) bewusst offen zu lassen, sodass es oder

Wikipedia Text Geschlechtseintrag

Eine Skala (von lateinisch scalae auch Skala, bezeichnet eine Aufeinanderfolge einer Anzahl von Teilstrichen auf einer Anzeigefläche. Die Teilung kann gleichmäßig oder, falls erforderlich, auch ungleichmäßig sein. Sie dient dazu, mit Hilfe einer Ableserate einen Wert anzuzeigen. Bei Werten von physikalischen Größen auf technischen Geräten werden die Striche in der Regel mit Zahlen und Einheit ergänzt; ansonsten auch Wertangaben oder Bewertungszeichen möglich. Eine Anzeige mittels einer Skala wird in der Messtechnik Skalenanzeige genannt. Die Alternative dazu ist die Ziffernanzeige. Der Begriff Skala wird in der messtechnischen Literatur benutzt und ist vom Deutschen Institut für Normung normiert.

Wikipedia Text Skala

Daten erzeugen

1. Deutsche Handschriften (2 Beispiele)

Şahmaran

(auch Schahmaran und Shohmaran) ist eine mythische Figur in Anatolien (besonders bei Tarsus und Marolin), dem Iran, dem Irak und Kurdistan. Sie gilt als Göttin der Weisheit und Beschützerin von Geheimnissen. Der Name besteht aus den Bestandteilen aus dem persischen "König" und (persisch "Schlangen") und bedeutet König (in) der Schlangen. wird oft als ein Wesen, halb weiße Frau, halb Schlange, dargestellt. Wenn sie stirbt, geht ihr Geist auf ihre Tochter über. In einer der Erzählungsvarianten verliebt sich in einen Mann namens Zamsad. In nachfolgenden Versionen heißt er oder Ghano.

Wikipedia Text Şahmaran

Vakuum ist in der technischen Praxis ein Raum mit weitgehender Abwesenheit von Materie. Im Vakuum gibt es keine festen Objekte oder Flüssigkeiten, nur extrem wenig Gas und damit auch einen extrem niedrigen Gasdruck. Ein technisches Vakuum wird erzeugt, indem mit der einer Pumpe (Gas-) Moleküle aus einem Behälter entfernt; der Druck darin sinkt (Gasdruck entsteht durch Stöße der Gas moleküle gegen die Behälterwand). Das Pumpen erzeugt einen Unterdruck also einen Druck, der geringer ist als der Umgebungsdruck. Sinkt der Druck im Behälter unter 300 mbar und entleert

Wikipedia Text Vakuum

Daten erzeugen

1. Deutsche Handschriften (2 Beispiele)

Es Gottesdienst ist eine Zusammenkunft von Menschen mit einem Zweck,
mit Gott in Verbindung zu treten, mit ihm Gemeinschaft zu haben,
Orte zu besetzen, Lehrende zu empfangen oder eine aufsteigende religiöse
Offenheit zu erleben. Es kann in einer Kirche, Synagoge,
Gemeindehaus (Kloster, Synagoge, Kirche, Moschee, Tempel, Abtei, Klosterkirche etc.)
stattfinden, was auch mit privaten Hauskirchen, in Kirchen
(oder öffentlichen) öffentlichen Räumen oder im Haus Gottes.
Der Gottesdienst ist ein Teil des Lebens der Menschen.
Gottesdienst (Gottesdienst, "Gottesdienst") und Gottesdienst sind
religiöse Begegnungen, in denen sich die Menschen mit Gott
Gottesdienstleistungen verbinden, die gemeinsame Gottesdienste,
von den Menschen und dem

Wikipedia Text Gottesdienst

Bildschirmtext (kurz Btx oder BTX; in der Schweiz
Videotext) war ein interaktiver Onlinedienst. Es
kombinierte Funktionen des Telefons und des Fernstudiums
zu einem Kommunikationsmittel. Btx wurde in Österreich
im Juni 1984 eingeführt; in der Bundesrepublik Deutschland
ab dem 1. September 1983 bundesweit. Durch die
Konkurrenz des offenen Internets verlor Bildschirmtext
seine Bedeutung später wieder, inzwischen wurde der
Dienst in allen Ländern eingestellt, in Deutschland 2009.
Es gab Verwechslungen mit dem Fernseh-Teletext, was zu
auch beitrug, dass der Dienst in der Schweiz Videotext
(ohne "f" am Ende) hieß und damit eine Ähnlichkeit mit
dem in Deutschland verwendeten Teletext-Synonym
Videotext aufwies.

Wikipedia Text Bildschirmtext

1. Deutsche Handschriften (Referenztext bzw. Grundwahrheit) am Beispiel Gottesdienst

Ein Gottesdienst ist eine Zusammenkunft von Menschen mit dem Zweck, mit Gott in Verbindung zu treten, mit ihm Gemeinschaft zu haben, Opfer zu bringen, Sakramente zu empfangen oder eine auferlegte religiöse Pflicht zu erfüllen. Er kann in einer eigens vorgesehenen Räumlichkeit (Kirche, Synagoge, Moschee, Pagode, Tempel, Königreichssaal etc.) stattfinden, wie auch im privaten häuslichen Bereich, im städtischen (oder dörflichen) öffentlichen Raum oder in freier Natur. Das deutsche Wort Gottesdienst entspricht dem lateinischen Begriff cultus (Kultus, "Verehrung") und bezieht sich vornehmlich auf religiöse Feiern im Christentum, wird jedoch auch für andere Religionsgemeinschaften verwendet, die gemeinsame Gebete verrichten, wie im Judentum und dem

Daten erzeugen

2. Produkt Etiketten in der Industrie

- Masterarbeit von Jannes Nitzsche (an Universität Leipzig): **Multimodal Approaches for Information Extraction from Industrial Product Labels**
- Die Arbeit wurde bei Telekom MMS (Telekom MMS | Experience Beyond Digital) durchgeführt und finanziert im Rahmen des EU Projektes IPCEI-CIS – 8ra
- Industrielle Labels weisen eine hohe Vielfalt auf. Sie unterscheiden sich in Layout, Typografie, Inhaltsstruktur und visuellem Design.
- Komplexität entsteht durch reale Produktionsbedingungen. Labels werden auf unterschiedlichen Materialien wie Papier, Folie, Kunststoff oder Metall gedruckt, die verschiedene Reflexionen und Oberflächenstrukturen erzeugen
- Untersuchung von zwei Ansätzen zur Informationsgewinnung aus industriellen Produktetiketten: VLMs (End-to-end Extraction) und klassische OCR-Methoden (two-stage pipeline, z.B. PadleOCR).
- Insgesamt wurden über 1.000 synthetische Etiketten generiert (weil reale Daten Mangelware sind), um die Modelle zur Informationsextraktion zu bewerten.

Daten erzeugen

2. Produkt Etiketten (2 Beispiele)

Absender: Rockmann Feren-Böring-Straße 1457 81053 Egenauw		Empfänger: Gilbert Meyer Scheffstr. 18 83761 Schrobenhausen	
RECEIVED: 346047665323400037		Empfänger: 7	
Los Nummer: RYALCAPRR	Frachtnummer: 1852918	Referenz-Nr. Nummer: 5131602524	
Produktbeschreibung: chocolate kiss		Artikel-Nr.: ARAYE-CKE78	
Packungstyp: Envelope 6TX		Gewicht: 14.06 kg	
Referenz-Nr. Nummer: 1167		Datum: 2025-10-07	



33119603034



43510067028



38007346522



43510067028

Beispiel 1

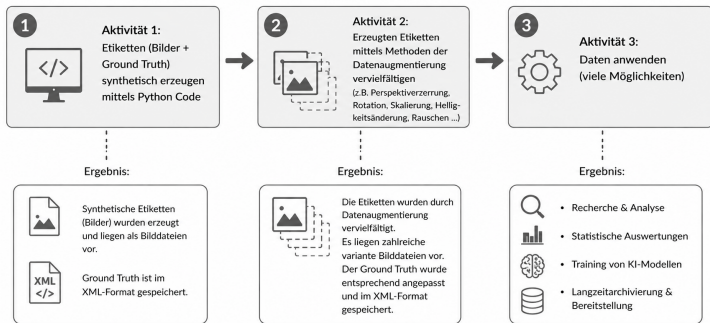
Absender: Aumann Brühl-Rosemann-Straße 8/9 63374 Gießen		Empfänger: Gabriella Henochel Kreuzweg 25 56673 Sankt Goar	
RECEIVED: 373590827238403120		Empfänger: DRIRPSS	
Los Nummer: RYALCAPRR		Frachtnummer: 1852918	
Produktbeschreibung: chocolate kiss		Artikel-Nr.: ARAYE-CKE78	
Packungstyp: Envelope 6TX		Gewicht: 14.06 kg	
Referenz-Nr. Nummer: 1167		Datum: 2025-10-07	



Beispiel 2

Daten erzeugen

2. Produkt Etiketten in der Industrie (Ablaufdiagramm)



Ablaufdiagramm

Daten erzeugen

2. Produkt Etiketten (Referenz bzw. Ground Truth im JSON-Format)

```
{
  "label_id": "label-0009",
  "image_file": "label_9.png",
  "objects": [
    {
      "type": "text",
      "value": "Sender: Beckmann\\nFerenc-Döring-Straße 1457\\n81033 Hagenow",
      "bbox": [
        6,
        12,
        301,
        99
      ]
    },
    {
      "type": "text",
      "value": "Recipient: Gilbert Mende\\nSchenkstr. 19\\n63761 Schrobhausen",
      "bbox": [
        686,
        17,
        251,
        90
      ]
    }
  ]
}
```

Referenz bzw. Ground Truth im JSON-Format

Daten erzeugen

2. Produkt Etiketten erzeugen (Ground Truth und Bild)

```
# =====  
# Main generation  
# =====  
  
def generate_labels() -> None:  
    """  
    Generate NUM_LABELS synthetic labels into OUTPUT_FOLDER and annotations into  
    OUTPUT_FOLDER/annotations. Uses Faker (de_DE) and randomized layout/codes.  
    """  
    ensure_clean_dirs()  
    faker = Faker('de_DE')  
  
    for i in range(NUM_LABELS):  
        # --- Build a data row ---  
        row: Dict[str, str] = {}  
        row.setdefault("order number", f"{random.randint(1000, 9999)}")  
        row.setdefault("weight", f"{random.uniform(0.1, 15.0):.2f} kg")  
        row.setdefault("quantity", str(random.randint(1, 30)))  
        row.setdefault(  
            "date",  
            (datetime.today() - timedelta(days=random.randint(0, 60))).strftime(  
                "%Y-%m-%d"  
            ),  
        )  
        row["sender"] = faker.company() + "\n" + faker.address()  
        row["addressee"] = faker.name() + "\n" + faker.address()  
        row["item_description"] = random_objects(k=1, with_spaces=True)[0]  
  
        # --- Pick base layout and scale its canvas + code sizes ---  
        layout_type = pick_layout()  
        cfg = LAYOUT_CONFIGS[layout_type]  
        scaled = get_scaled_canvas(cfg, layout_type)
```

Programmcode Ausschnitt zum Erzeugen von GT und Bild

Daten erzeugen

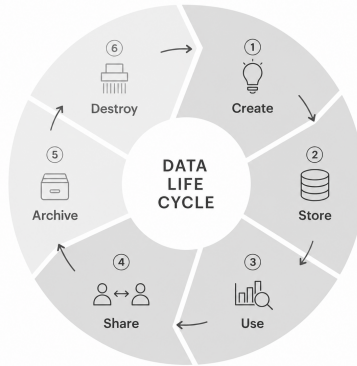
Zwischenfazit

- Daten erzeugen ist ein wichtiger Schritt im Datenlebenszyklus, um bspw. danach ML-Modelle trainieren und bewerten zu können
- Für die Bewertung von Methoden zur Informationsextraktion wurden handgeschriebene Wikipedia-Texte und synthetische Produktetiketten verwendet, um die Leistungsfähigkeit der OCR- und Bild-Text-Modelle zu bewerten
- Die Erstellung von Ground Truth ist zeitaufwendig, aber notwendig, um die Genauigkeit und Zuverlässigkeit der Methoden bewerten und optimieren zu können
- Eine hohe Vielfalt bei den Handschriften und Etiketten ist wichtig für das Training und die Bewertung der Modelle, da diese auf möglichst realistische Beispiele treffen sollen, um eine breite Anwendbarkeit zu gewährleisten.

Daten speichern

Daten speichern bzw. ablegen

Datenlebenszyklus



Daten speichern bzw. ablegen

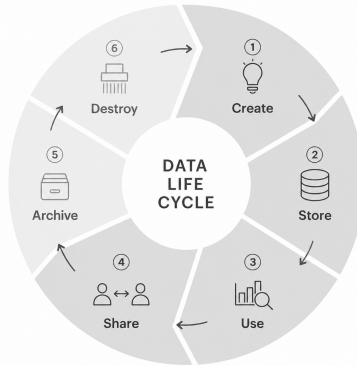
Data Science Cycle

- **Datenverfügbarkeit sicherstellen:** Daten werden zentral abgelegt und können für Analysen, Berichte, KI-Analysen und -trainings, Prozesse oder Entscheidungen genutzt werden.
- **Zusammenarbeit ermöglichen:** Mehrere Personen oder Systeme können auf dieselben Daten zugreifen und diese gemeinsam nutzen.
- **Rechtliche und organisatorische Anforderungen erfüllen:** Speicherung unterstützt die Einhaltung von Compliance-Vorgaben, Dokumentationspflichten und Datenschutzrichtlinien.

Daten anwenden

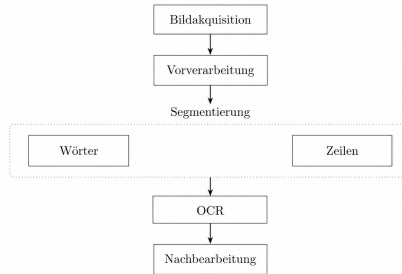
Daten anwenden

Datenlebenszyklus



Daten anwenden

OCR - Verarbeitungspipeline

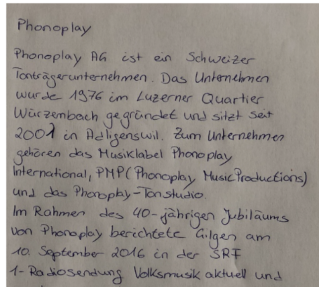


Verarbeitungspipeline für OCR-Methoden (Inferenz)⁴

⁴Inferenz (im Machine Learning) ist die Nutzungsphase eines trainierten Modells. Während das Modell beim Training Muster aus Daten lernt, wendet es dieses Wissen in der Inferenzphase an, um neue, unbekannte Daten zu analysieren und Vorhersagen oder Entscheidungen zu treffen

Daten anwenden

OCR - Vorverarbeitung (Binarisierung, Rauschunterdrückung, etc.)



Phonoplay
Phonoplay AG ist ein Schweizer
Tonträgerunternehmen. Das Unternehmen
wurde 1976 im Luzerner Quartier
Würzenbach gegründet und sitzt seit
2001 in Adligenswil. Zum Unternehmen
gehören das Musiklabel Phonoplay
International, PMP (Phonoplay Music Productions)
und das Phonoplay-Tonstudio.
Im Rahmen des 40-jährigen Jubiläums
von Phonoplay berichtete Gilgen am
10. September 2016 in der SRF
1-Radiosendung Volksmusik aktuell und

OCR-Vorverarbeitungsmethode Binarisierung

Daten anwenden

OCR - Segmentierung

für Hawaii und Zauberei und Magischer kongress der Vereinigten Staaten von Amerika geführt werden. Zum ersten Film erschien zudem das Begleitbuch mit dem Titel Fantastic Beasts and Where to Find Them The Original Screenplay, das das komplette Skript des Films enthält und die darin vorkommenden fantastischen Tieren und Figuren vorgestellt. Weitere Figuren werden in Mobile game Fantastic Beasts: Cases From the Wizarding World erwähnt, das von Warner Bros.



für Hawaii und Zauberei und Magischer kongress der Vereinigten Staaten von Amerika geführt werden. Zum ersten Film erschien zudem das Begleitbuch mit dem Titel Fantastic Beasts and Where to Find Them The Original Screenplay, das das komplette Skript des Films enthält und die darin vorkommenden fantastischen Tieren und Figuren vorgestellt. Weitere Figuren werden in Mobile game Fantastic Beasts: Cases From the Wizarding World erwähnt, das von Warner Bros.



für Hawaii und Zauberei und Magischer kongress der Vereinigten Staaten von Amerika geführt werden. Zum ersten Film erschien zudem das Begleitbuch mit dem Titel Fantastic Beasts and Where to Find Them The Original Screenplay, das das komplette Skript des Films enthält und die darin vorkommenden fantastischen Tieren und Figuren vorgestellt. Weitere Figuren werden in Mobile game Fantastic Beasts: Cases From the Wizarding World erwähnt, das von Warner Bros.



für Hawaii und Zauberei und Magischer kongress der Vereinigten Staaten von Amerika geführt werden. Zum ersten Film erschien zudem das Begleitbuch mit dem Titel Fantastic Beasts and Where to Find Them The Original Screenplay, das das komplette Skript des Films enthält und die darin vorkommenden fantastischen Tieren und Figuren vorgestellt. Weitere Figuren werden in Mobile game Fantastic Beasts: Cases From the Wizarding World erwähnt, das von Warner Bros.

OCR-Segmentierung in Wörter und/oder Zeilen

Daten anwenden

OCR - Methoden zur Erzeugung maschinenlesbaren Textes (kleine Auswahl)

- **Feature-basierte OCR** – Merkmalsextraktion mittels Konturen, Projektionen oder **manuell** definierten Zonenmerkmalen⁵
- **CNN + LSTM** – Kombination aus visueller Merkmalsextraktion und Sequenzmodellierung⁶
- **Transformer-basierte OCR** – Nutzung von Self-Attention für *In-Einem-Schritt-Alle-Bildelemente verstehen* (keine Segmentierung und anschließend OCR Zeichen für Zeichen, oder Wort für Wort, aber noch kein großes Sprachmodell vorhanden, sondern Transformer optimiert für OCR, außerdem nur 384x384 Pixel große Bilder)

⁵1980er Jahre bis Anfang 2000er Jahre, "Ein Zeichen (Zahl oder Buchstabe) wird nicht als Ganzes betrachtet, sondern durch mathematisch beschreibbare Merkmale charakterisiert (vertikale und horizontale Linien, Kreise, ...)."

⁶CNN lernt bzw. extrahiert automatisch relevante visuelle Merkmale, wie z.B. Kanten, ...

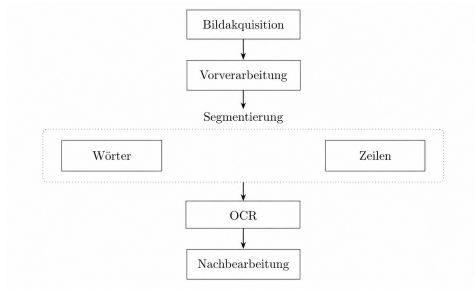
Neuere Bild-Text-Modelle zur Erzeugung maschinenlesbaren Textes

- **Vision-Language Models (VLM)** – Bild-Text-Modelle⁷: Das Modell sieht riesige Mengen von Bild-Text-Paaren, darunter auch Dokumente, Notizen und handschriftliche Texte. Dabei lernt es statistische Zusammenhänge zwischen visuellen Mustern und sprachlichen Symbolen. Ähnlich wie ein Sprachmodell Wortfolgen lernt, lernt ein VLM, dass bestimmte Pixelmuster häufig mit bestimmten Zeichen oder Wörtern korrespondieren.

⁷Verarbeitung von Bild- und Sprachinformationen, Text- und Kontextverständnis durch großes Sprachmodell, z.B. GPT-4o, Qwen-VL; Der Decoder-Teil bei GPT lernt, Text aus dem Bild zu extrahieren und die nächsten Wörter vorherzusagen, basierend auf der Bildrepräsentation und dem bereits generierten Text.

Daten anwenden

OCR - Verarbeitungspipeline



Verarbeitungspipeline für OCR-Methoden (Inferenz)

Daten anwenden

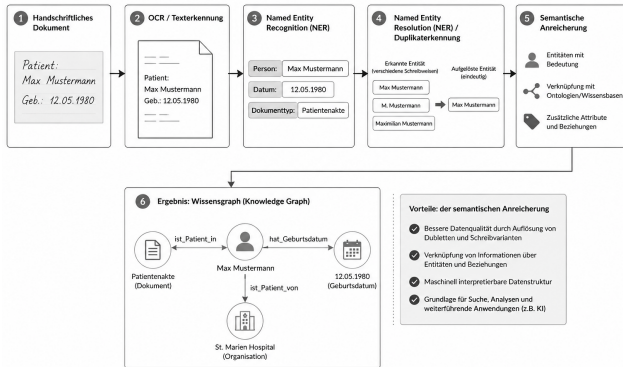
OCR - Nachbearbeitung (Fokus OCR oder trOCR)

- **Wörterbuchabgleich (Lexikon-Matching)** – Abgleich mit bekannten Wortlisten
- **Named Entity Recognition (NER)** – Erkennung von Personen-, Orts- oder Organisationsnamen (Entitätserkennung beginnend 1990er Jahre, NLP-Kontext, später auch in transformerbasierten Ansätzen wie BERT (Bidirectional Encoder Representations from Transformers, 2018) zu finden)
- **Duplikaterkennung** – Identifikation mehrfach erkannter Inhalte
- **Plausibilitätsprüfungen** – Validierung von Feldern anhand definierter Regeln, z.B. Formulare
- **Semantische Anreicherung** – Verknüpfung mit Metadaten oder Wissensbasen, z.B. Wissensgraphen, Ontologien

Daten anwenden

OCR - Nachbearbeitung - Vorgehen (Personen in Wissensgraph anreichern oder prüfen)

Beispiel: Semantische Anreicherung



OCR-Nachbearbeitung - Vorgehen

Daten anwenden

Ergebnis der 1. Bachelorarbeit

Modell	Trainingsdaten
Altes HTR-Modell	ScaDS.AI-Datensatz (37.406 Wörter, 400 Seiten)
TrOCR-Modell	

OCR-Trainingsdaten

Daten anwenden

Ergebnis der 1. Bachelorarbeit

Modell	CER (%)	Epochen	Batch-Größe
Altes HTR Modell	14.00%	40	16
TrOCR	8.17%	40	16

OCR-Bewertung

Daten anwenden

VLM-Benchmark für Handschriften (2. Bachelorarbeit)

OCRBench v2: An Improved Benchmark for Evaluating Large Multimodal Models on Visual Text Localization and Reasoning

Ling Fu¹, Zhenlin Kuang², Jiguan Song³, Mingxin Huang³, Biao Yang¹,
Yuhao Li¹, Jinghui Chen¹, Qiduo Luo¹, Xinyu Wang¹, Shao Lu¹, Zheng Li¹,
Guohui Tang¹, Bin Shao¹, Chaochao Liu¹, Qi Liu¹, Binghui Wu¹,
Han Feng¹, Han Lu¹, Cao Huang¹, Jiguo Wang¹, Wei Chen¹,
Guowen Jin¹, Yiding Liu¹, Xiang Bai¹

¹Huazhong University of Science and Technology ²South China University of Technology
³University of Adelaide ⁴HynDance

Abstract

Scoring the Optical Character Recognition (OCR) capabilities of Large Multimodal Models (LMMs) has witnessed growing interest. Existing benchmarks have highlighted the impressive performance of LMMs in text recognition; however, few abilities to contain challenging tasks, such as text localization, handwritten content extraction, and logical reasoning, remain underexplored. To bridge this gap, we introduce **OCRBench v2**, a large-scale bilingual text-centric benchmark with currently the most comprehensive set of tasks (i.e. more tasks than the previous multi-modal benchmark OCRI-Bench), the widest coverage of scenarios (14 diverse scenarios), and thorough evaluation metrics, with 10,000 human-verified question-answer pairs and a high proportion of difficult samples. Moreover, we construct a private test set with 1,500 manually generated images. The consistent evaluation results observed across both public and private test sets validate the OCRI-Bench v2's reliability. After carefully benchmarking state-of-the-art LMMs, we find that most LMMs score below 50/100 in total and suffer from five-type limitations, including less frequently encountered text recognition, fine-grained perception, layout perception, complex element parsing, and logical reasoning. The project website is at: https://github.com/zhongshen/ocrbench_v2

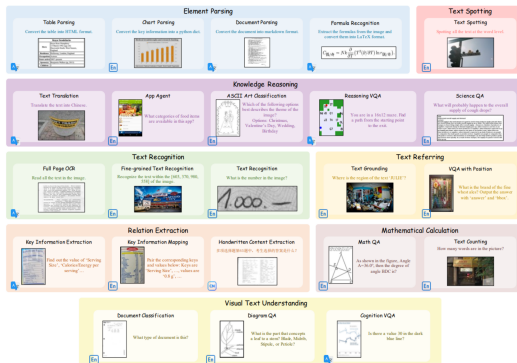
Aufsatz OCR-Benchmark v2⁸⁹

⁸<https://arxiv.org/pdf/2501.00321>

⁹OCRBench v2 wurde entwickelt (2025), weil OCRBench v1 (2023) zunehmend „gesättigt“ war: viele Modelle erreichten bereits sehr hohe Werte, die Aufgaben waren oft zu einfach, Handschriften, Layoutprobleme und schwierige Dokumente waren unterrepräsentiert. Deshalb erweitert OCRBench v2: von ca. 5 auf 23 Aufgaben, von ca. 14 auf 31 Szenarien, von 1.000 auf 10.000 Testfragen, deutlich mehr Handschriften, mehr Dokumentverständnis, mehr schwierige OCR-Situationen

Daten anwenden

VLM-Benchmark v2 für Handschriften (2. Bachelorarbeit)



Daten anwenden

Verwendete Datenquellen (2. Bachelorarbeit)

Quelle	Seiten
ScaDS-unlabeled	77
CVL-Database	10
Synthetisch	10
FH-SWF-Collage	3
Gesamt	100

Verwendete Datensätze

Daten anwenden

Ergebnisse der 2. Bachelorarbeit

Modell	Alle	ScaDS	CVL	Synth	FH-SWF
ChatGPT	0,027	0,025	0,021	0,006	0,191
Pixtral	0,104	0,110	0,075	0,030	0,286
Deepseek	0,355	0,384	0,205	0,170	0,719
Deepseek (det,)	0,393	0,453	0,157	0,144	0,448
MiniCPM	0,325	0,373	0,136	0,151	0,324
MiniCPM (cor,)	0,299	0,348	0,118	0,114	0,245
Qwen	0,040	0,042	0,042	0,007	0,068
Qwen (det,)	0,032	0,034	0,033	0,006	0,065
Llama	0,044	0,045	0,051	0,020	0,089
TrOCR (Alzin)	–	~0,06	–	–	–
TrOCR	0,434	0,374	0,605	0,530	0,624
TrOCR (cor,)	0,478	0,425	0,631	0,563	0,639
TrOCR-82	0,439	0,379	0,594	0,549	0,620
TrOCR-82 (cor,)	0,482	0,424	0,653	0,577	0,642
TrOCR-large	0,548	0,490	0,751	0,616	0,689
TrOCR-large (cor,)	0,582	0,530	0,783	0,633	0,698

Ergebnisse Erkennungsgüte

Daten anwenden

Verwendet Daten (Industrie Produkt Etiketten - Masterarbeit)

Variant	Labels	Images	Annotated Objects	Avg. Objects per Label
Synthetic Set	300	300	5272	17.57
Augmented Set	900	900	15816	17.57
Realistic Set	50	96	1718	17.90

Verwendeter Datensatz

Daten anwenden

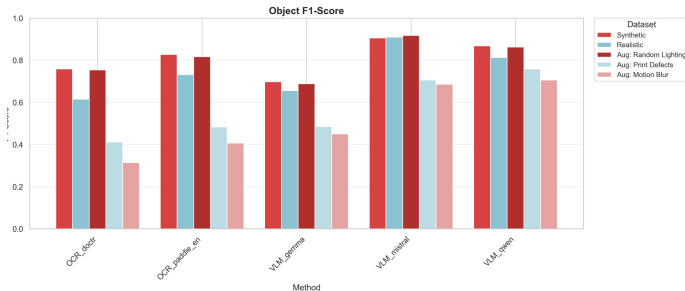
Verwendet Daten (Industrie Produkt Etiketten - Masterarbeit)



Verwendeter Datensatz

Daten anwenden

Industrie Produkt Etiketten - Ergebnisse der Masterarbeit

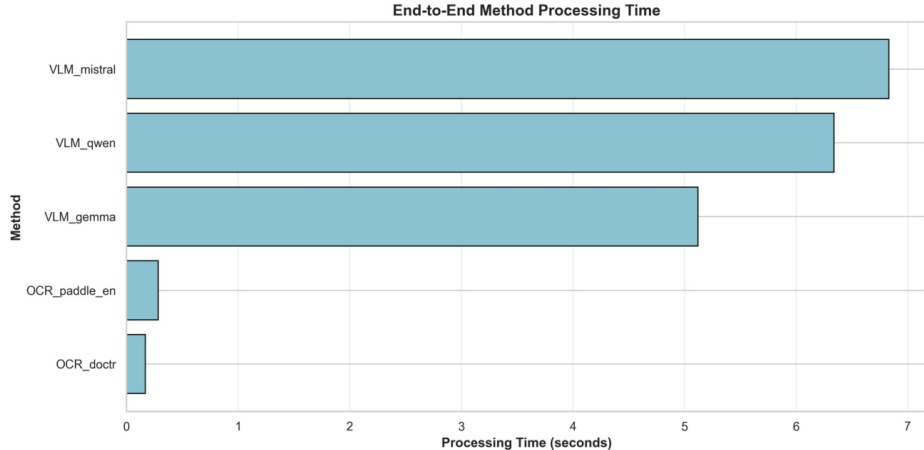


Ergebnisse Erkennungsgüte (F1-Score¹⁰)

¹⁰Genauigkeit von 90 % sind gute Werte, aber nicht perfekt. Nachbearbeitung: mit Methoden des Adressbuchabgleiches (z.B. Record Linkage mittels Ähnlichkeitsmaßen) können auch mit unvollständigen oder fehlerhaften Daten noch bessere Ergebnisse erzielen.

Daten anwenden

Industrie Produkt Etiketten - Ergebnisse der Masterarbeit



Daten anwenden

Herausforderungen bei der Anwendung von OCR-Methoden und VLMs

- **Hohe Variabilität von Handschriften:** Individuelle Schreibstile, unterschiedliche Buchstabenformen und Schreibgewohnheiten erschweren die Generalisierung von OCR-Modellen.
- **Fragmentierte Zeichen:** erschweren Segmentierung und Klassifikation, z.B. Buchstaben mit Tintenlücken, unvollständige oder beschädigte Zeichen, durchgestrichene Wörter, Überlappungen oder sich berührende Zeichen.
- **Berührende und verschmolzene Zeichen:** erlauben Interpretationen
- **Auch unscharfe Strickonturen durch Tintenverlauf:** erschweren die Erkennung von Zeichen, insbesondere bei dünnen Linien oder ungleichmäßiger Tintenverteilung.

Daten anwenden

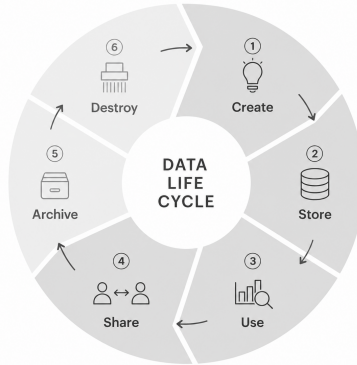
Herausforderungen bei der Anwendung von OCR-Methoden und VLMs

- **Schlechte Bildqualität:** Unschärfe, Bildrauschen, geringe Auflösung oder ungünstige Beleuchtung können die Erkennung erheblich beeinträchtigen.
- **Dokumentalterung und Artefakte:** Flecken, Verfärbungen, Knicke, Durchscheineffekte oder beschädigtes Papier erschweren die Segmentierung und Texterkennung.
- **Komplexe Layouts und Überlappungen:** Randnotizen, Tabellen, Stempel oder sich überschneidende Textbereiche können zu fehlerhaften Segmentierungen führen.
- **Uneinheitliche Schreibweise und Abkürzungen:** Historische Schreibweisen, Fachterminologie oder individuelle Abkürzungen reduzieren die Erkennungsgenauigkeit.
- **Zeichenverwechslungen:** Visuell ähnliche Zeichen (z. B. 0/O oder 1/I/l) werden aufgrund ihrer ähnlichen Form falsch erkannt.

Daten teilen

Daten teilen

Datenlebenszyklus



Daten teilen

Zenodo Datensätze und mehr

- Zenodo Datensatz **Industrial Label Dataset for Structured Information Extraction**
<https://doi.org/10.5281/zenodo.19480237>
- Zenodo Datensatz **ScaDS.AI German Full-Page Handwriting Dataset Authors/Creators**
<https://doi.org/10.5281/zenodo.18283705>
- Zenodo Datensatz **ScaDS.AI German Line- and Word-Level Handwriting Dataset**
<https://doi.org/10.5281/zenodo.18301532>
- Anwendung zur Erkundung der Handschrifterkennung (HTR Explorer) im Internet:
<https://livinglab.scadsai.uni-leipzig.de/htr-viewer/>
- Anwendung zur Erkundung der Industrie Produkt Etiketten (IPL Explorer) im Internet:
<https://livinglab.scadsai.uni-leipzig.de/ipl-viewer/>
- Blog-Beitrag zu allen vorgestellten Aktivitäten auf der Seite
<https://scads.ai/open-data-for-german-handwriting-recognition/>

Daten teilen

Zenodo Datensätze und mehr

- **Sicher** -- Ihre Forschungsergebnisse werden langfristig und sicher im Rechenzentrum des CERN gespeichert, solange CERN existiert.
- **Vertrauenswürdig** -- Zenodo wird von CERN und OpenAIRE entwickelt und betrieben, um sicherzustellen, dass alle an Open Science teilhaben können.
- **Zitierbar** -- Jeder Upload erhält einen Digital Object Identifier (DOI), sodass die Inhalte zitierbar und nachverfolgbar sind.
- **Offen oder geschlossen** -- Beispielsweise können anonymisierte Daten aus klinischen Studien über den Modus für eingeschränkten Zugriff gezielt nur mit medizinischem Fachpersonal geteilt werden.
- **Versionierung** -- Datensätze können mithilfe der integrierten Versionierungsfunktion unkompliziert aktualisiert werden.
- **GitHub-Integration** -- GitHub-Repositorien lassen sich einfach und dauerhaft in Zenodo archivieren.

Zusammenfassung

Zusammenfassung

Zusammenfassung

- Deutsche Handschriftproben sind eine wertvolle Ressource für die Forschung im Bereich der Handschrifterkennung.
- Multimodale Sprachmodelle (Vision-Language Models, VLMs) zeigen überraschend robuste Ergebnisse bei der Erkennung lateinischer Handschriften und Etiketten.
- Die Qualität der Daten (z. B. Scanqualität und Zuschnitt der Dokumente) hat einen wesentlichen Einfluss auf die Leistung von Handschrifterkennungssystemen.
- Die Wahl des Erkennungssystems (klassische OCR-Systeme oder End-to-End-Ansätze) hängt vom jeweiligen Anwendungskontext ab (z. B. Geschwindigkeit).

Vielen Dank für die Aufmerksamkeit!

- Zenodo Datensatz **Industrial Label Dataset for Structured Information Extraction**
<https://doi.org/10.5281/zenodo.19480237>
- Zenodo Datensatz **ScaDS.AI German Full-Page Handwriting Dataset Authors/Creators**
<https://doi.org/10.5281/zenodo.18283705>
- Zenodo Datensatz **ScaDS.AI German Line- and Word-Level Handwriting Dataset**
<https://doi.org/10.5281/zenodo.18301532>
- Anwendung zur Erkundung der Handschrifterkennung (HTR Explorer) auf
<https://livinglab.scadsai.uni-leipzig.de/htr-viewer/>
- Anwendung zur Erkundung der Industrie Produkt Etiketten (IPL Explorer) auf
<https://livinglab.scadsai.uni-leipzig.de/ip1-viewer/>
- Blog-Beitrag zu allen vorgestellten Aktivitäten auf
<https://scads.ai/open-data-for-german-handwriting-recognition/>