

ADS Projektarbeit

W.MA.WIN.21HS.b

EMILIA CHONIA

ALI MERT

BENJAMIN PETERS

Inhaltsverzeichnis

- ▶ Einführung
- ▶ Vorgehensweise
- ▶ Code-Durchlauf
- ▶ Diskussion & Ergebnisse
 - ▶ Jupyter
 - ▶ Generell
- ▶ Diskussion ethischer Aspekte
- ▶ Fazit
- ▶ Ende

Einleitung

Hintergrund

- ▶ Allgemeines Interesse für Schuhe
- ▶ Thematik Bildverarbeitung wurde im Unterricht behandelt, eine Umsetzung scheint gemäss Einschätzung Teammitglieder realisierbar
- ▶ Arbeiten mit CNN mit Bildern gut möglich

Problemstellung

Durch den digitalen Wandel ist es für Neueinsteiger im Detailhandelsbereich wichtig, sich über den Einsatz von ML Tools wie die automatisierte Klassifizierung von Bildern zu differenzieren oder in Zukunft zu skalieren.

Zielsetzung

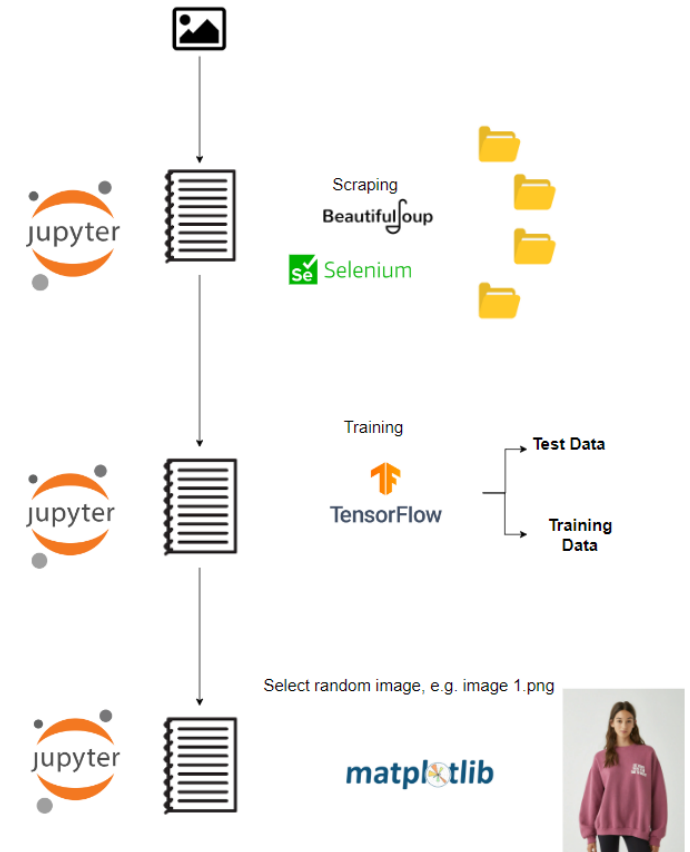
Ein Maschine Learning Modell entwickeln, welches Schuhbilder in vier verschiedene Kategorien klassifiziert und Vorhersagen mit neuen Bildern trifft.

Forschungsfrage

Wie kann man ein Machine Learning Modell von Grund auf aufsetzen, um vier verschiedene Kategorien von Schuhen zu scrapen und danach zu klassifizieren, damit zukünftig neue Bilder zur Klassifikation eingelesen werden können?

Vorgehensweise

- (1) Datengewinnung mittels Web Scraping (BeautifulSoup) und Web API (Selenium)
- (2) Datenspeicherung lokal (respektive OneDrive)
- (3) Datenaufruf aus Scraping
- (4) Bereinigung (Normalization & Resizing)
- (5) Classification normal
- (6) Classification CNN
- (7) Vorhersage / Prediction einbauen
- (8) Neues Bild einfügen – Testen
- (9) Versionierung und Publikation der Daten und des Modells über GitHub



Code - Durchlauf

Diskussion & Ergebnisse - Jupyter

Scraping

- WebScraping-Website (Zalando → Dosenbach → Koala → (ENG))
- Scraping von mehr als einer Seite - iterieren
- lokales Speichern der Bilder mit «write»

Classification

- Aufrufen & Labeln der Daten (mit Hilfe Array)
- Normalisierung & Resizing (Data Prep)
- CNN einbauen - höhere Accuracy als «altes» Model?
 - Accuracy zuerst bei ca. 80% dann bei über 90%
- Prediction einbauen, Problem Labelling

Diskussion & Ergebnisse - generell

- ▶ Themenfindung
→ genug schwierig aber im Rahmen & sinnvoll?
- ▶ Nutzung Docker
→ Sagemaker...
→ OneDrive
- ▶ Wo und wie bauen wir ein API ein?
→ Selenium (2. HT) – Bilderproblematik
- ▶ Arbeiten mit GitHub
→ Aufsetzung generell – Neuland
- ▶ Jedes Teammitglied Quereinsteiger, niemand geübt in der IT

Diskussion ethischer Aspekte, potentieller Risiken und Bias



Five Levels of Concern	Mögliche Use Cases und deren Einschätzung
Personal Privacy	<ul style="list-style-type: none"> Kommerzielle Nutzung der Koala / Dosenbach Bilder, z.B. im Rahmen einer neuen "Klassifikations-App" ist fraglich. Schutz von Endnutzern fraglich bei Bereitstellung von eigenen Bildern
Implicit Bias	<ul style="list-style-type: none"> Geltungsraum evtl. zu definieren, da Kategorien nicht weltweit gelten (Boots, Heels, Ballerinas, etc.)
Accountability	Man stelle sich die Anwendung des Modells in einer Marketplace-App vor, wo bereitgestellte Bilder von Verkäufern (z.B. Kleiderbörse) automatisch klassifiziert und bepreist werden sollten. In einem solchen Szenario müsste der Anbieter der Plattform die Verantwortung tragen für Fehlklassifikationen und damit verbundene Kosten
Human Agency	Grundsätzlich kann das Modell unterstützend für automatisierte Entscheidungen genutzt werden – mit menschlicher Überwachung
Innovation	Das Modell löst ein Klassifikationsproblem, was grundsätzlich in einem Effizienzgewinn und einem innovativen Service resultiert.

Fazit



- ▶ Zufriedenheit Ergebnisse – funktioniert
- ▶ Lernkurve in Python (*im Vergleich zu „nur“ Ausführen*)
- ▶ Einsatz verschiedener Techniken (*BeautifulSoup, Selenium, GitHub etc..*)
- ▶ Unterstützung Dozenten (*während und nach Vorlesungen*)



- ▶ Mangelnde Programmiererfahrung – Kostet viel Zeit und Nerven
- ▶ Aufwand vs. Bewertung (Prüfung zählt zu 70%)

“

Success is not final; failure is not
fatal: It is the courage to continue
that counts
unknown, unknown

”

Link zu Github

<https://github.com/echonia/ZHAW-ADS>