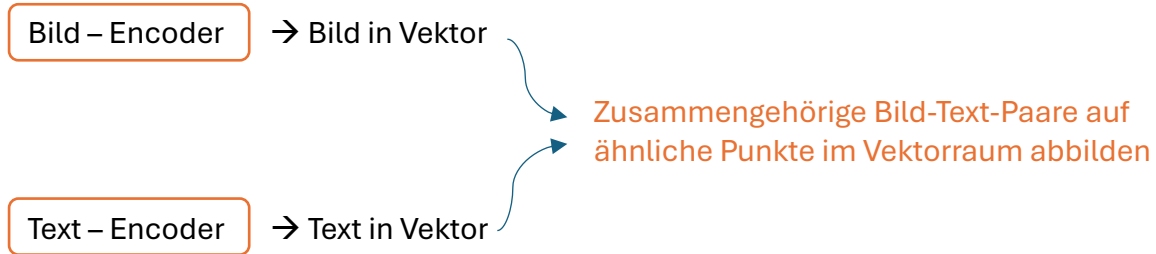


Bildbasiertes Empfehlungssystem mit FashionCLIP

CLIP → „Contrastive Language-Image Pretraining“ Natürliche Sprache mit Bildern verknüpfen



Nach Bild suchen, indem man Text eingibt

FashionCLIP → auf Kleidungsstücke spezialisiert

Nächste Schritte:

~~Bild – Text~~ → nur Bild – Encoder

1. Schuhbilder durch Bildencoder laufen lassen und in Dictionary speichern
2. Random Bilder aus Dictionary anzeigen und „likern“ lassen
3. Vektoren zu gelikten Bildern benutzen und nach nearest-neighbour suchen
4. Diese Bilder dem Nutzer anzeigen

Annoy → Nearest-Neighbour-Suche im Vektorraum (spotify nutzt diese)

Datenerhaltung

- Bild IDs + Embeddings (in Datei oder direkt in Annoy)
- Nutzer – Session (welche Bilder wurden bereits gezeigt, welche geliked)