

ONiT Explorer – Multimodal analysis of historical images

Mag. Michela Vignoli (Austrian Institute of Technology)

Vienna, 04.11.2025

This presentation was created in collaboration with Dr. Doris Gruber
(Austrian Academy of Sciences)

CONTENT

1. The ONiT Project

- Introduction & Research Questions
- Corpus
- Approach

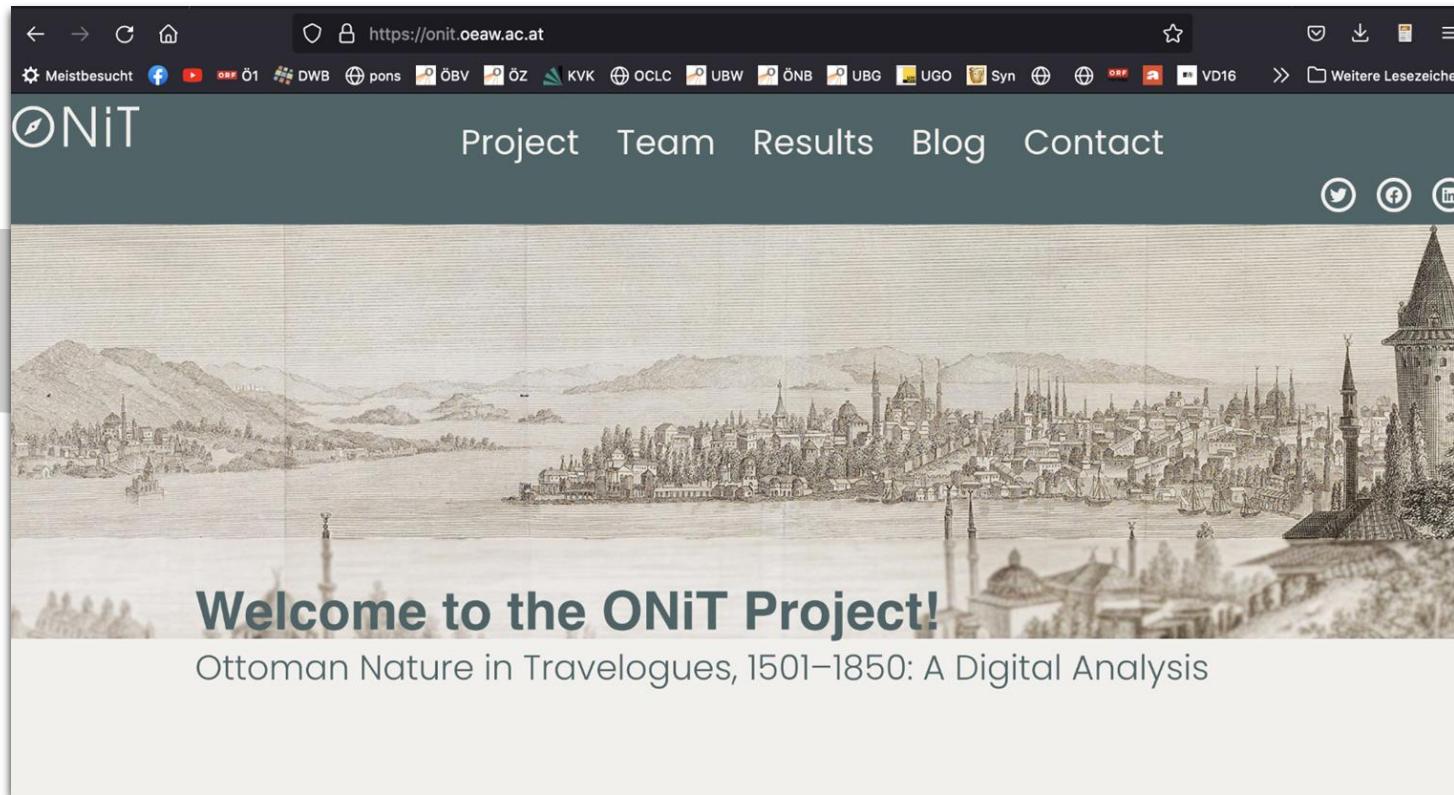
2. Multimodal Image Retrieval

- State of the Art
- Methodology
- ONiT Explorer Demo
- Introduction Vision-Language Models

3. Results and Conclusion

- Results
- Conclusions

THE ONiT-PROJECT: INTRODUCTION



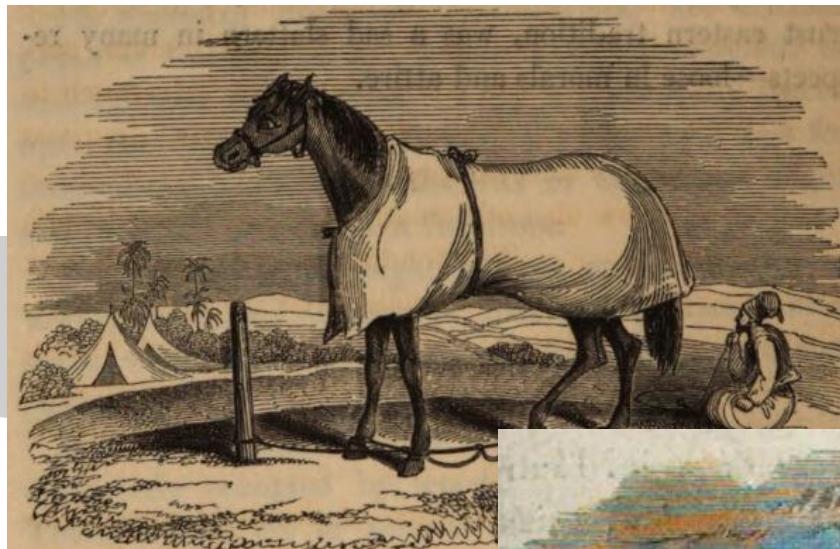
<https://onit.oeaw.ac.at>

FWF
Der Wissenschaftsfonds.



Project duration: 2022–2025
Team: [Institute for Habsburg and Balkan Studies \(IHB\)](#) of the [Austrian Academy of Sciences \(ÖAW\)](#), the [Austrian Institute of Technology \(AIT\)](#), the [Austrian National Library \(ÖNB\)](#) in Vienna, and the [Marmara University](#) Istanbul

THE ONIT-PROJECT: INTRODUCTION



Aim: Analysis of „nature“ representations
(i.e. flora, fauna, landscapes, maps) in travelogues
on the Ottoman Empire



THE ONiT-PROJECT: RESEARCH QUESTIONS



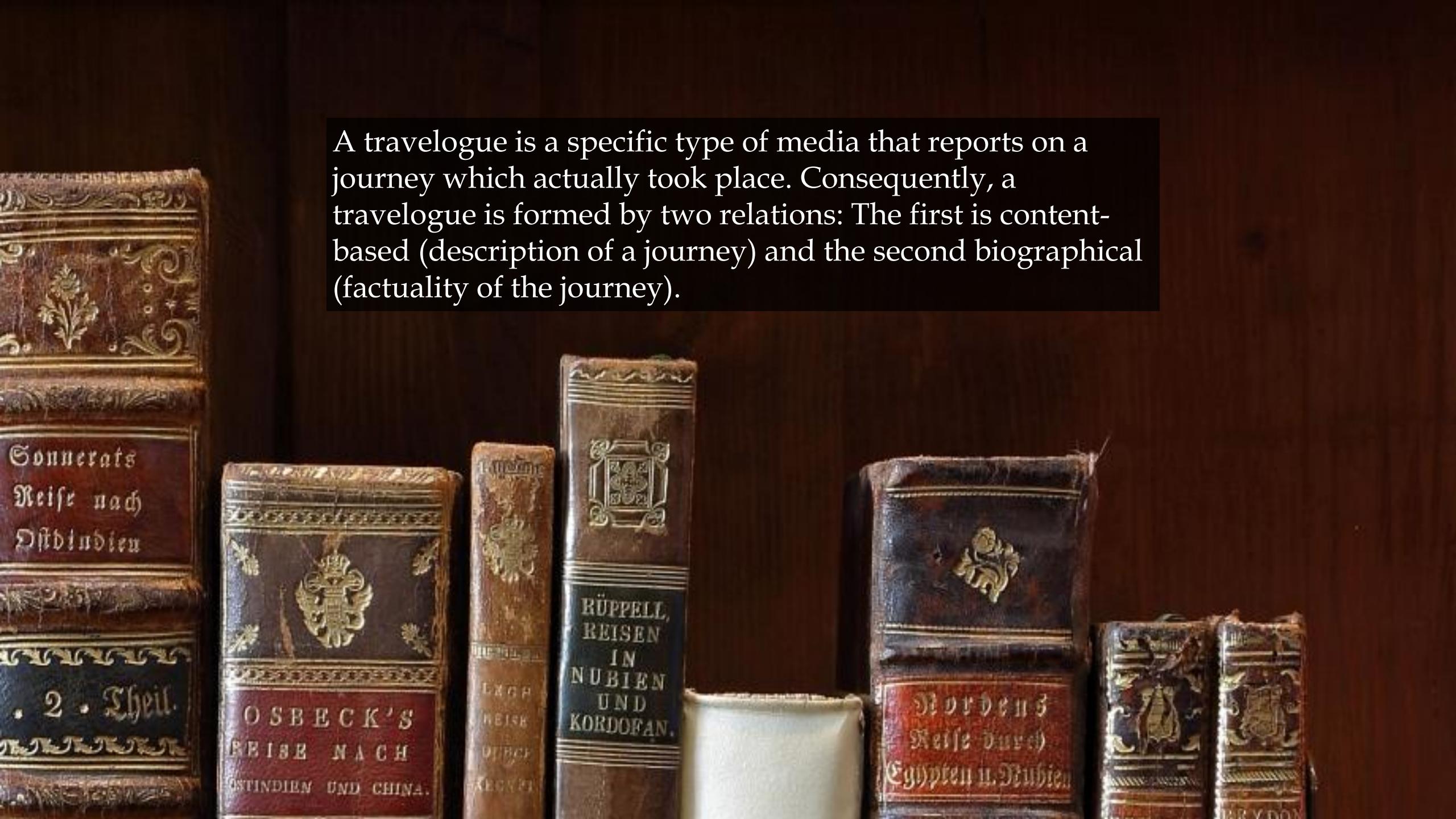
- What role did "**nature**" play in the reports?
- Which and, if so, why did **differences** occur in diachronic and synchronic perspectives?
- How do the **texts and images** on "nature" relate to each other?
- New light on transnational **environmental and natural history**

THE ONiT-PROJECT: CORPUS

- ⌚ Ottoman Nature in Travelogues (ONiT)
- ⌚ Western Travelogues to Ottoman Empire
 - ⌚ Travel reports from 16th-19th century in four languages
 - ⌚ > 2,500 early modern book prints (1,700 digitized)
 - ⌚ > 22,000 images extracted (woodcuts, engravings, etc.)



Image Source: <http://data.onb.ac.at/rep/108B5BC1>, p. 7

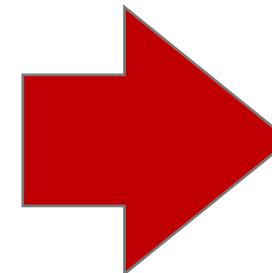


A travelogue is a specific type of media that reports on a journey which actually took place. Consequently, a travelogue is formed by two relations: The first is content-based (description of a journey) and the second biographical (factuality of the journey).

THE ONiT-PROJECT: CORPUS

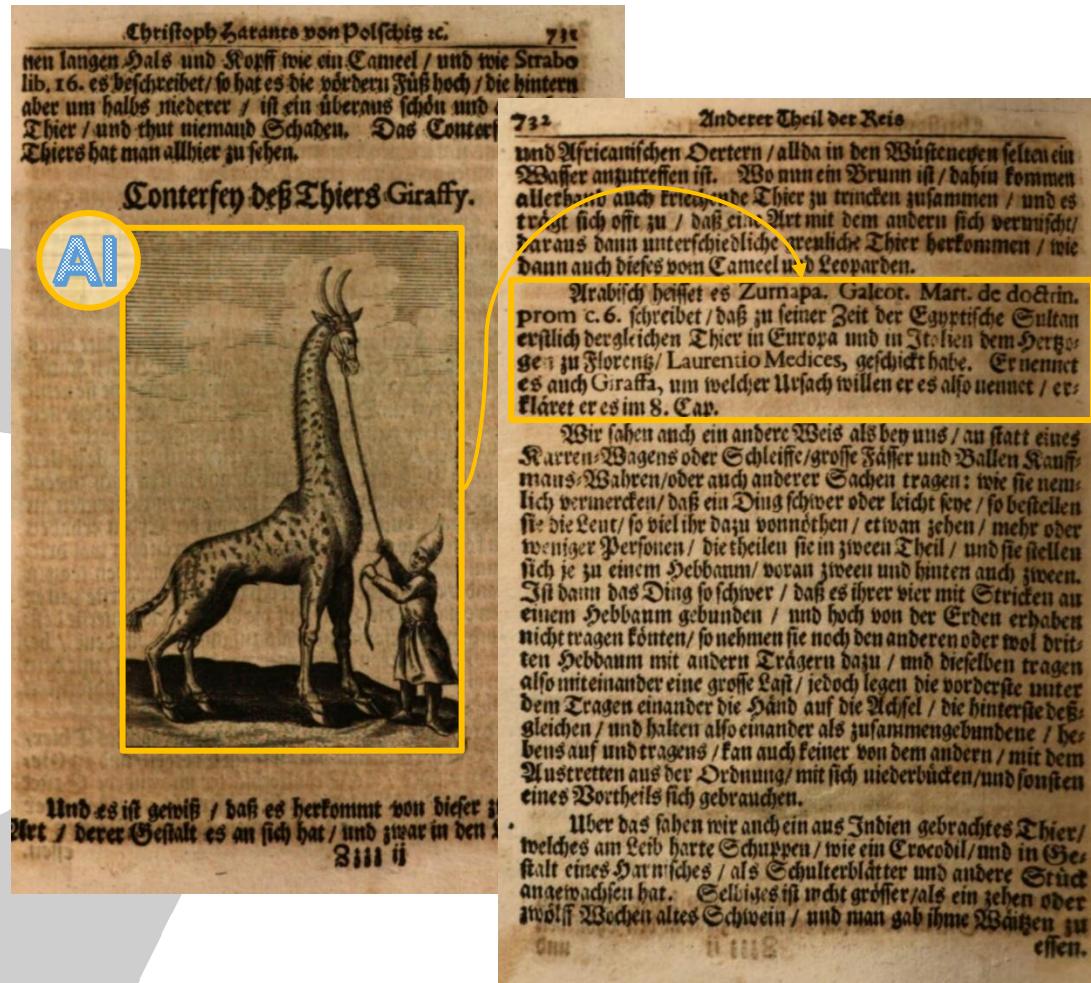
Printing Year	DE	EN	FR	LA	Total	Marker
1501–1599	53	0	30	54	137	ProjectOnit*16
1600–1699	114	12	120	55	301	ProjectOnit*17
1700–1799	201	60	360	11	632	ProjectOnit*18
1800–1850	393	321	795	6	1,515	ProjectOnit*19
1501–1850	761	393	1,305	126	2,585	ProjectOnit*

ProjectONiT
ProjectONiTDE*
ProjectONiTFR*
ProjectONiTITL*
ProjectONiTTE*



Corpus accessible
via Opac of ÖNB
<https://search.onb.ac.at/>

THE ONiT-PROJECT: APPROACH



Interdisciplinary workflow for image/ text extraction and analysis:

- image/text represent **specific coding of information**
- at first analyzed **separately with specific tools**, some of which use machine learning
- tension between the **numerical modelling and interpretive methods** to analyse narratives in historical sources

CONTENT

1. The ONiT Project

- Introduction & Research Questions
- Corpus
- Approach

2. Multimodal Image Retrieval

- State of the Art
- Methodology
- ONiT Explorer Demo
- Introduction Vision-Language Models

3. Results and Conclusion

- Results
- Conclusions

STATE OF THE ART

- **text-based searchability** of most humanities image search engines builds on manually-created metadata (*Bildindex, REALonline*)
- **creation of metadata:** time and resource-intense, often inconsistent or insufficient, esp. regarding historical images
- **image-based search:** most classification and object detection **algorithms** cannot be used out of the box for historical material (trained on modern photographic content)
- **training of new algorithms or fine tuning of existing models** for the analysis of early modern prints, e.g. [Visual Geometry Group \(VGG\)](#), University of Oxford; [Sphaera Project](#), MPIWG Berlin

STATE OF THE ART

[Learning Transferable Visual Models From Natural Language Supervision](#), Radford et al. 2021

- Contrastive Language-Image Pre-training (CLIP):
- multimodal model
- high accuracy in zero-shot classification
(predicting unseen categories across heterogeneous datasets)

JOURNAL ARTICLE

A multimodal turn in Digital Humanities. Using contrastive machine learning models to explore, enrich, and analyze digital visual historical collections 

Thomas Smits , Melvin Wevers

Digital Scholarship in the Humanities, Volume 38, Issue 3, September 2023,
Pages 1267–1280, <https://doi.org/10.1093/llc/fqad008>

Published: 15 March 2023

 PDF  Split View  Cite  Permissions  Share ▾

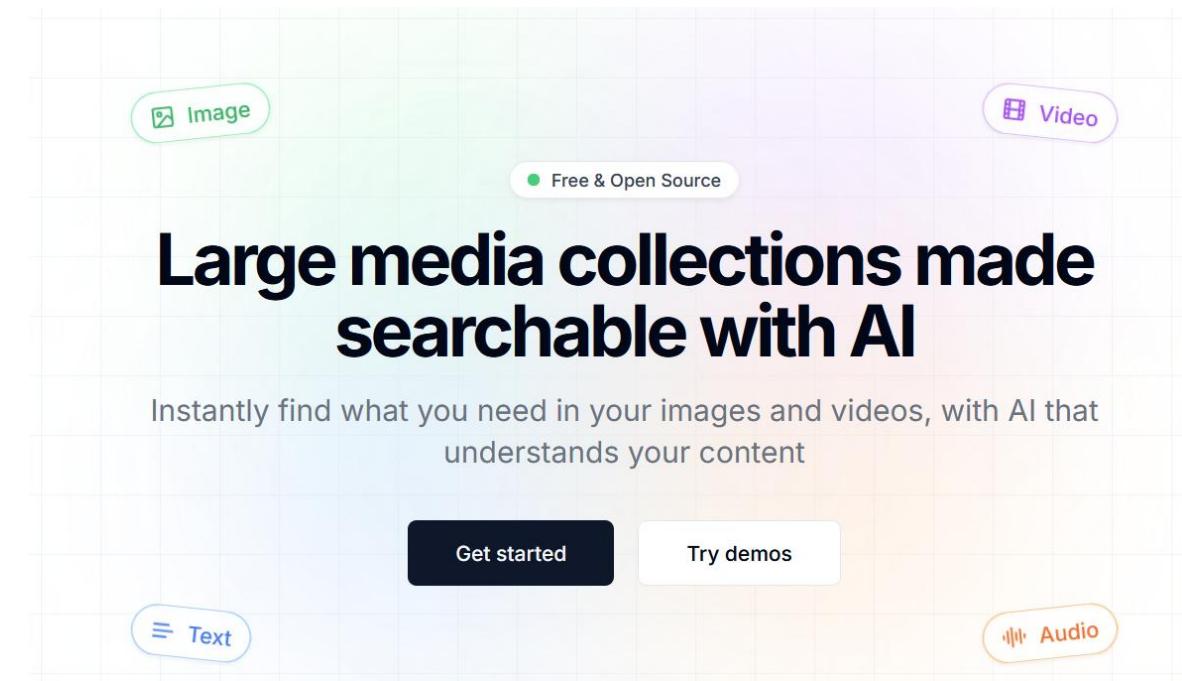
Abstract

Until recently, most research in the Digital Humanities (DH) was monomodal, meaning that the object of analysis was either textual or visual. Seeking to integrate multimodality theory into the DH, this

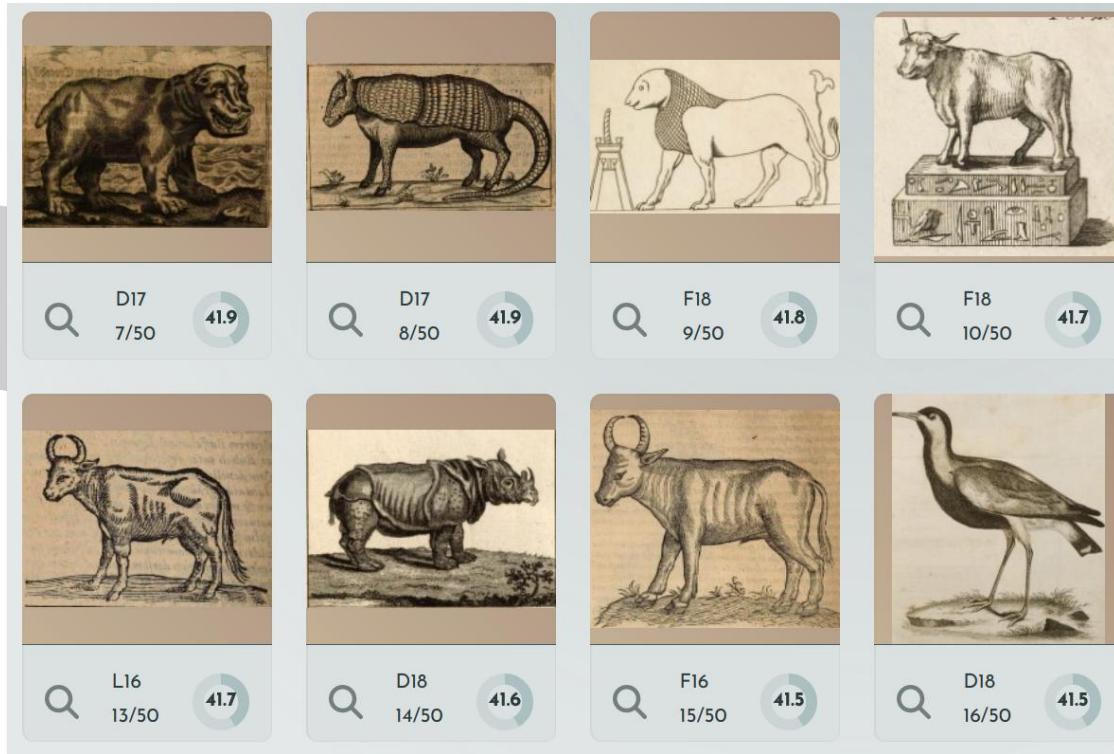
STATE OF THE ART

WISE is an open-source multimodal AI search tool developed and maintained by the VGG.

- Vision-language: CLIP, Google's SigLIP 2 (via OpenCLIP library)
- Audio-language: Microsoft CLAP and others
- Face recognition: InsightFace
- Audio transcription: WhisperX
- GitHub: <https://gitlab.com/vgg/wise/wise>



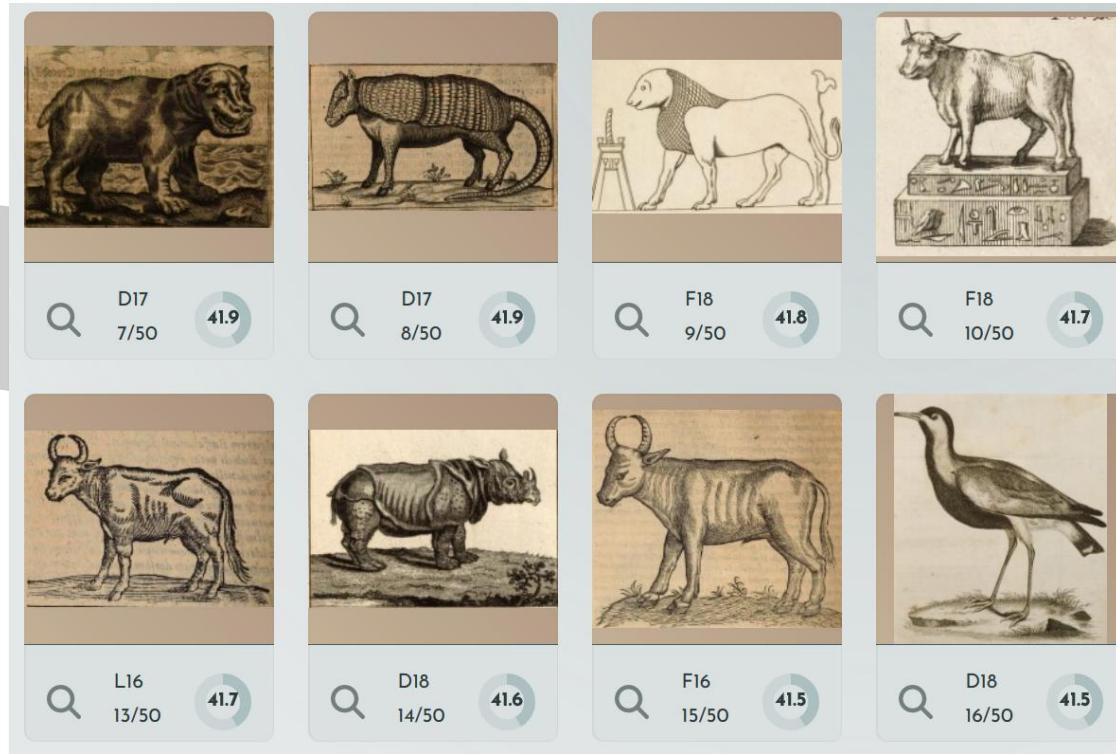
ONiT: METHODOLOGY



Focus on pictorial representations of Ottoman *nature* (flora, fauna, landscapes and maps):

- **Step 1: extraction**
- Automated with pre-trained tool
- **Step 2: curation**
- Semi-automatic
- **Step 3: classification**
- Manual annotation
- Fine-tuning and retrieval with CLIP model

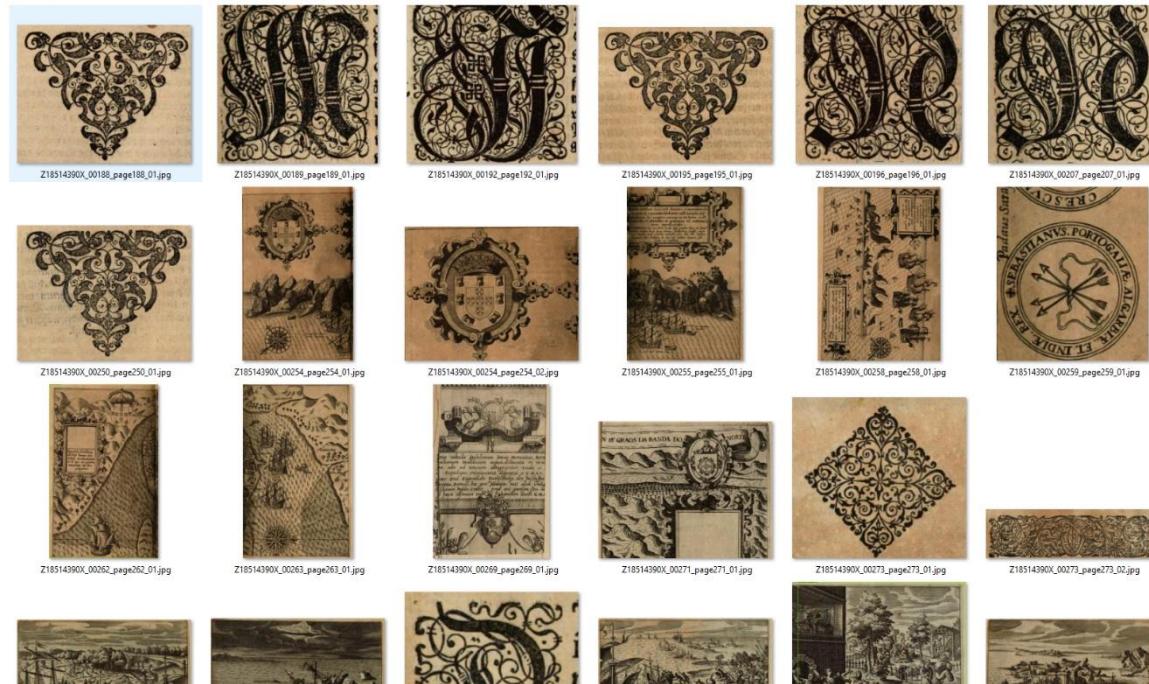
ONiT: METHODOLOGY



Step 1: extraction

- **tool:** *Illustration Detector / VGG*
- **result:** 22,300 pictorial elements extracted from over 1,500 digitised books (single editions)

ONiT: METHODOLOGY



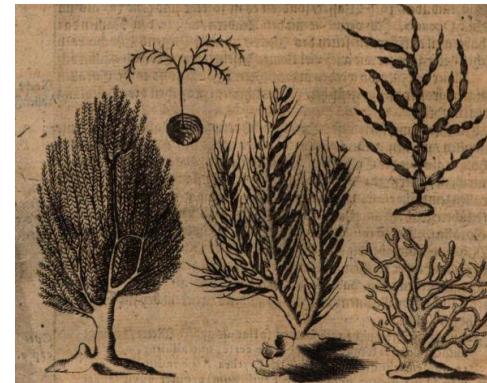
Step 2: curation

- **exclusion of irrelevant material** (error detections, library stamps, ornaments, initials, images not featuring nature)
- **image similarity** to support manual selection

ONiT: METHODOLOGY



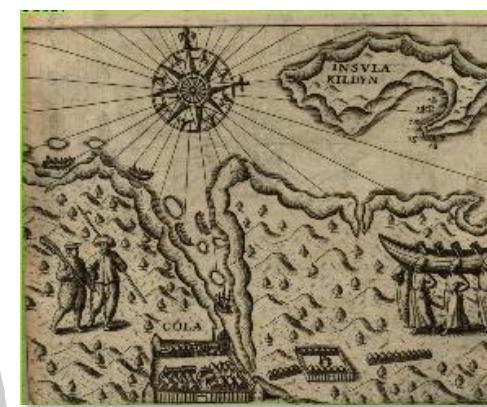
landscapes 25H



plants;
vegetation 25G



animals 25F



maps; atlases
25A

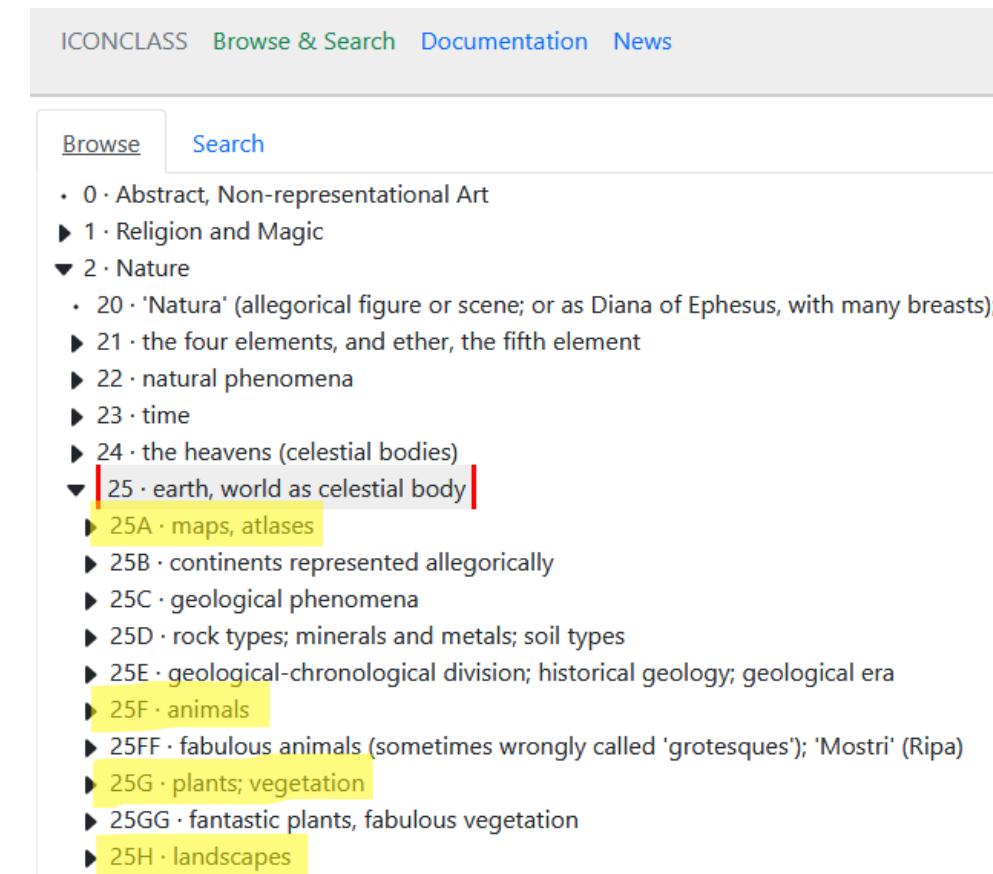
Step 3: classification

- annotations with *Iconclass*
- focus on four main categories
- more detailed annotations of animal classes

ONiT: METHODOLOGY

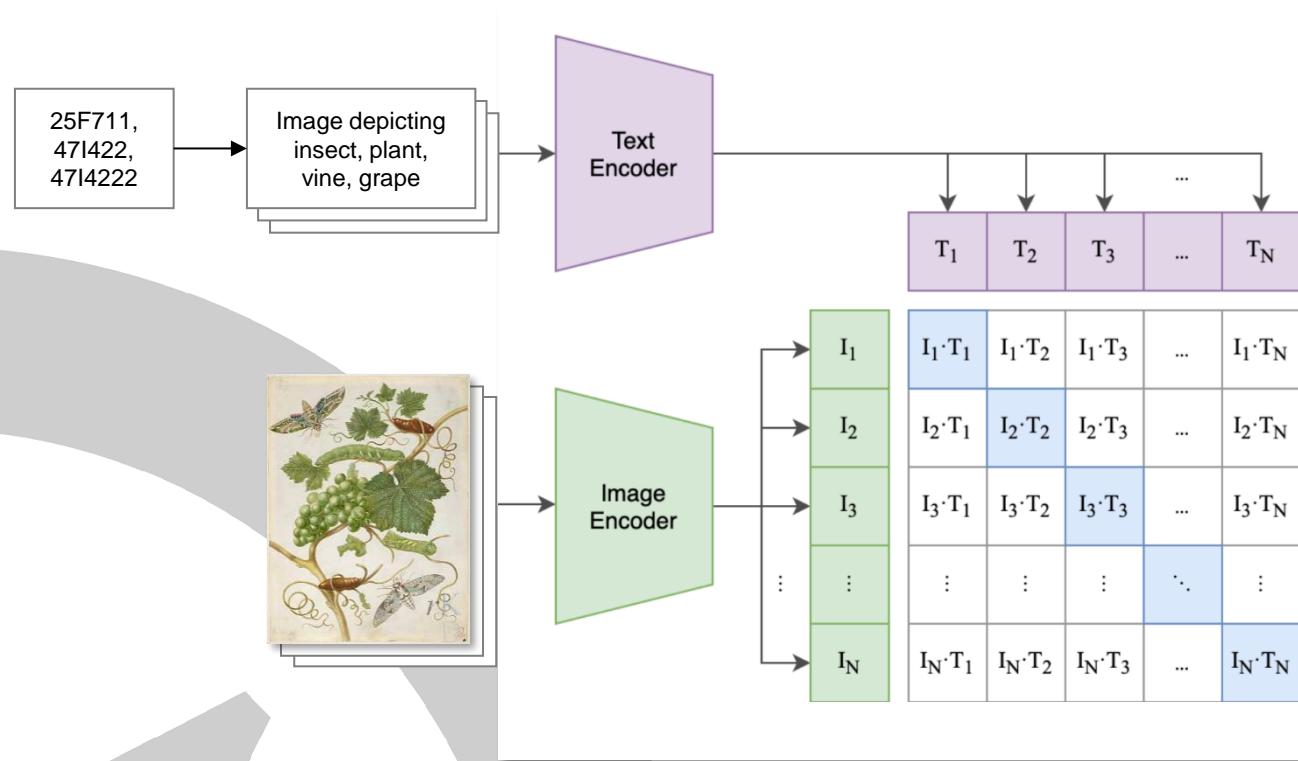
ICONCLASS

- Image classification vocabulary
- Coded with numbers and letters (e.g. 25F animals)
→ language agnostic
- Tree structure
- <https://iconclass.org/>



The screenshot shows the ICONCLASS website's browse interface. At the top, there is a navigation bar with links for 'ICONCLASS', 'Browse & Search', 'Documentation', and 'News'. Below the navigation bar, there are two tabs: 'Browse' (which is underlined) and 'Search'. The main content area displays a hierarchical tree structure. The root node is '0 · Abstract, Non-representational Art'. Below it is '1 · Religion and Magic', which is collapsed. The next level down is '2 · Nature', which is also collapsed. Under 'Nature', there are several nodes: '20 · 'Natura' (allegorical figure or scene; or as Diana of Ephesus, with many breasts);', '21 · the four elements, and ether, the fifth element', '22 · natural phenomena', '23 · time', '24 · the heavens (celestial bodies)', '25 · earth, world as celestial body' (this node is expanded and highlighted with a red border), '25A · maps, atlases' (highlighted with a yellow box), '25B · continents represented allegorically', '25C · geological phenomena', '25D · rock types; minerals and metals; soil types', '25E · geological-chronological division; historical geology; geological era', '25F · animals' (highlighted with a yellow box), '25FF · fabulous animals (sometimes wrongly called 'grotesques'); 'Mostri' (Ripa)', '25G · plants; vegetation', '25GG · fantastic plants, fabulous vegetation', and '25H · landscapes' (highlighted with a yellow box).

ONiT: MACHINE LEARNING TRAINING

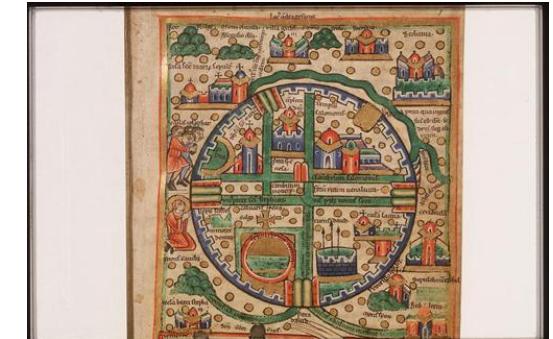
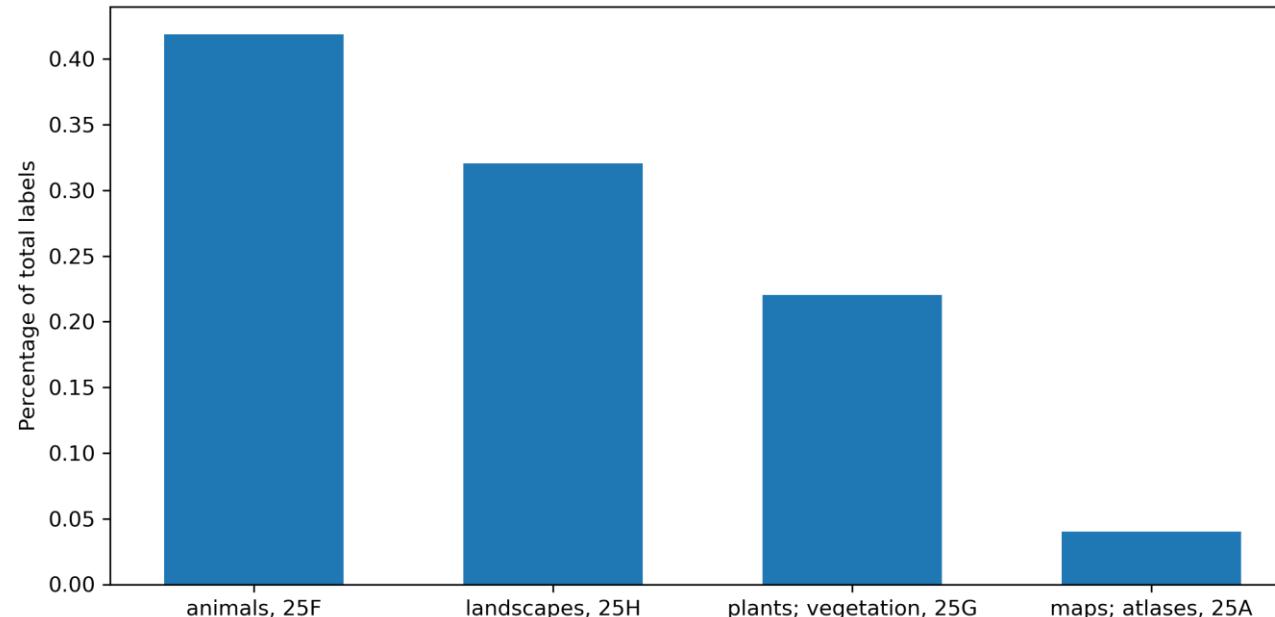


- **Model:** [openCLIP](#) ViT-B-32, pretrained on [LAION 400m](#) dataset
- Data split: 80/10/10 train/val/test
- Automatic mixed precision
- Params: 20 epochs; LR: 5e-5; batch: 128
- [WiSE-FT method](#) for better robustness (combining weights of pre-trained and fine-tuned models)

Image adapted from: Radford et al. 2021, <https://arxiv.org/abs/2103.00020>, p. 2

Nature Depictions in ICONCLASS AI Test Set

Label Frequencies ICONCLASS AI Test Set: Nature Subset



Source: Vignoli, Michela et al. (2023) *Impact of AI: Gamechanger for Image Classification in Historical Research?* In Konferenzbeiträge der Digital History 2023, Berlin. <<https://doi.org/10.5281/zenodo.8322398>>.

DEMO: ONiT EXPLORER

The screenshot shows the ONiT Explorer interface. At the top left is the ONiT logo and the text "Similarity Explorer". Below it is a search bar labeled "Search by Text:" with a placeholder "Search..." and a "Search" button. To the right is a "Number of Results:" dropdown set to "50" with "Search" and "Clear" buttons. In the center is a detailed botanical illustration of a plant with many leaves, labeled "Z17216440X (p. 92) E19". To the left of the image is a search bar placeholder "Search...". To the right is a "Select Corpora:" section with a grid of checkboxes for corpora A11 through L19. Below this are six smaller thumbnail images of botanical illustrations, each with a search icon, a count (e.g., 1/50, 2/50, 3/50), and a score (e.g., 99.0, 98.0, 96.5). At the bottom left is the copyright notice "© 2024 ONiT | Ottoman Nature in Travelogues" and the URL "onit.oeaw.ac.at". At the bottom right is the funding information "Funded by the Austrian Science Fund (FWF: P 35245)".

First version now online:

<https://labs.onb.ac.at/de/tool/onit-explorer/>



- **Web application for prompt-driven exploration and curation of historical image collections**
- Incorporates fine-tuned CLIP model
- **Exploration modes:** text prompt (**text to image**); retrieved or uploaded images (**image to image**)
- Advantage: **no need for metadata on the images**

CONTENT

1. The ONiT Project

- Introduction & Research Questions
- Corpus
- Approach

2. Multimodal Image Retrieval

- State of the Art
- Methodology
- ONiT Explorer Demo
- **Introduction Vision-Language Models**

3. Results and Conclusion

- Results
- Conclusions

INTRODUCTION: VISION-LANGUAGE MODELS

- Trained to match text descriptions with images.
- E.g. recognizes similarity of the sentence ‘a dog jumps over a fence’ with a specific image – even without the image having been manually labelled beforehand.

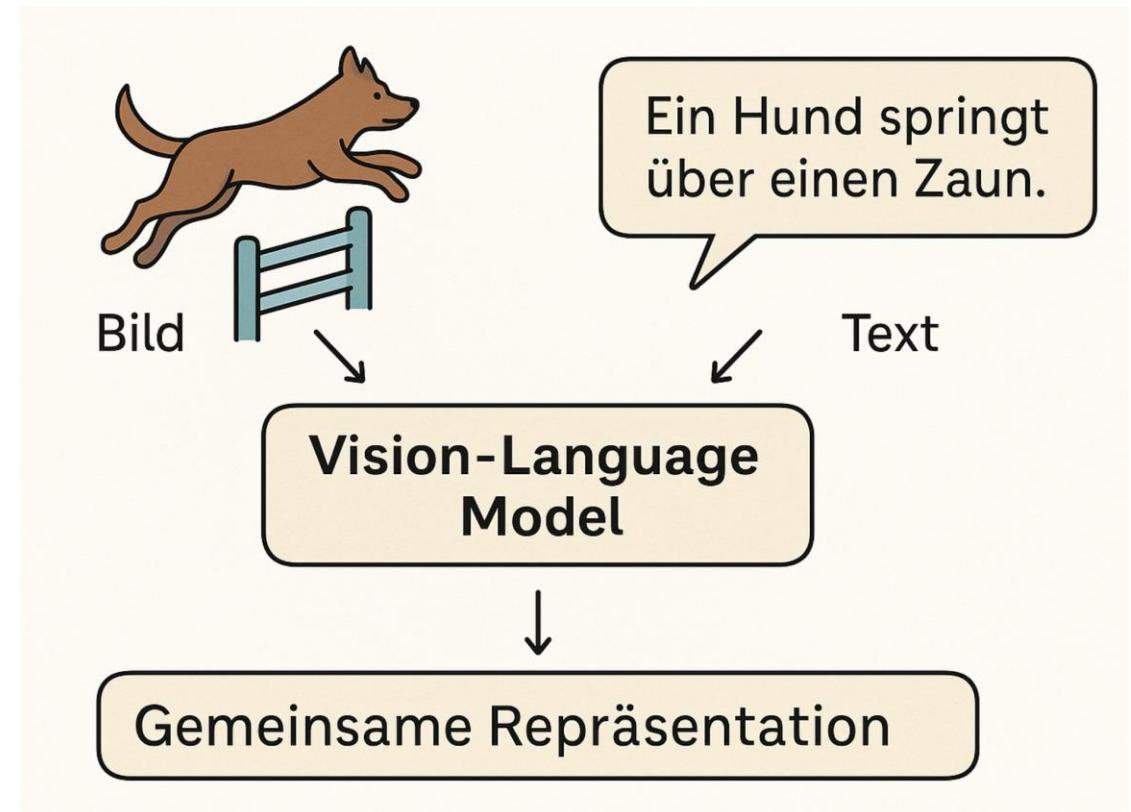
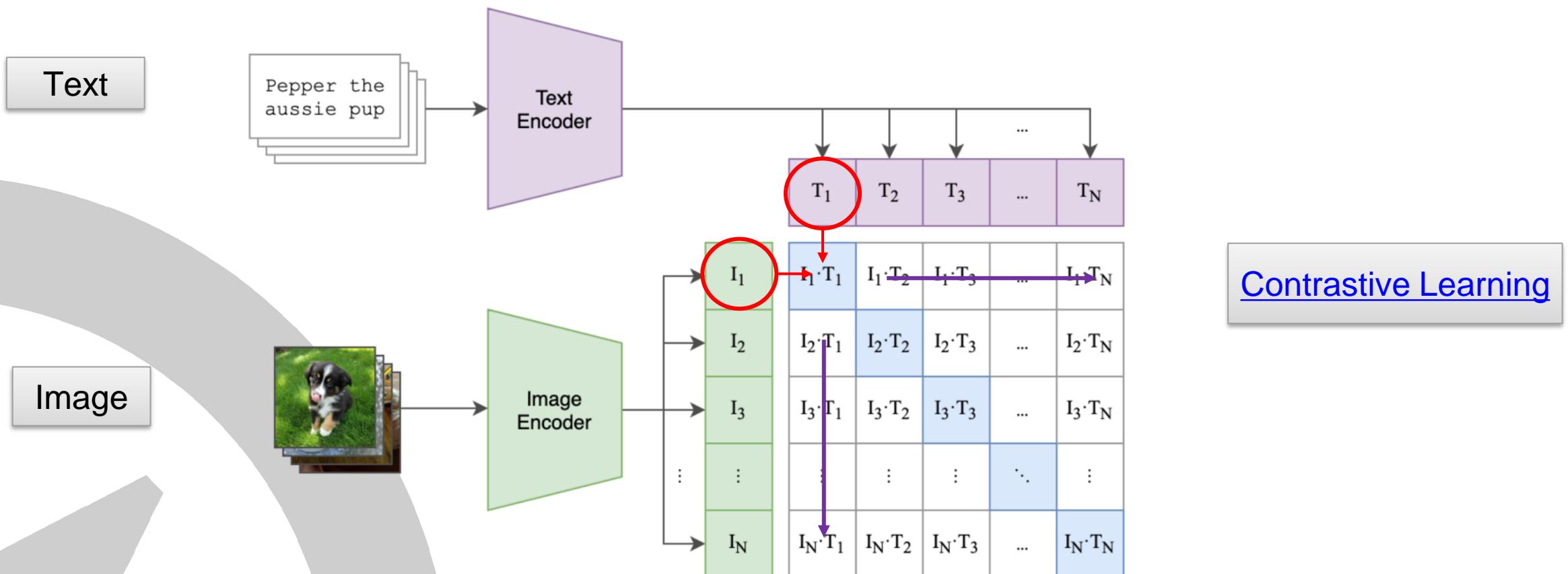
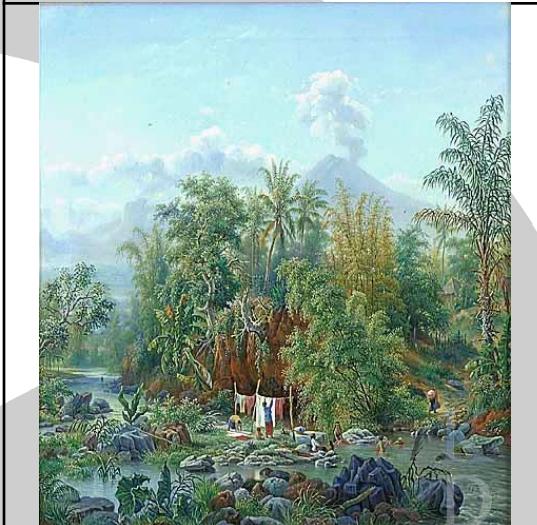


Image generated with ChatGPT-4. Prompt: Eine Grafik, die grundsätzliche Funktion von Vision-Language Models beschreibt.

INTRODUCTION: VISION-LANGUAGE MODELS



Training Input: Image-Caption Pairs

	<p>Image depicting back, company, elegant company, guest, receiving a person, visiting, couple, dog, castle, French garden, architectonic garden, formal garden, nourishment, nutrition, drinking, pouring, lute, theorbo, musical instrument, musician</p>		<p>Image depicting adult, man, wine, October, Ottobre, barrel, wood(en), tree, pouring</p>
	<p>Image depicting river, landscape, sub-tropical landscape, tropical landscape, washing (clothes)</p>		<p>Image depicting insect, plant, vine, grape</p>

Source: Vignoli, Michela et al. (2023) *Impact of AI: Gamechanger for Image Classification in Historical Research?* In Konferenzbeiträge der Digital History 2023, Berlin. <<https://doi.org/10.5281/zenodo.8322398>>.

Transformers (Deep Neural Networks)

The scaled dot-product attention mechanism is used in the multi-head attention blocks

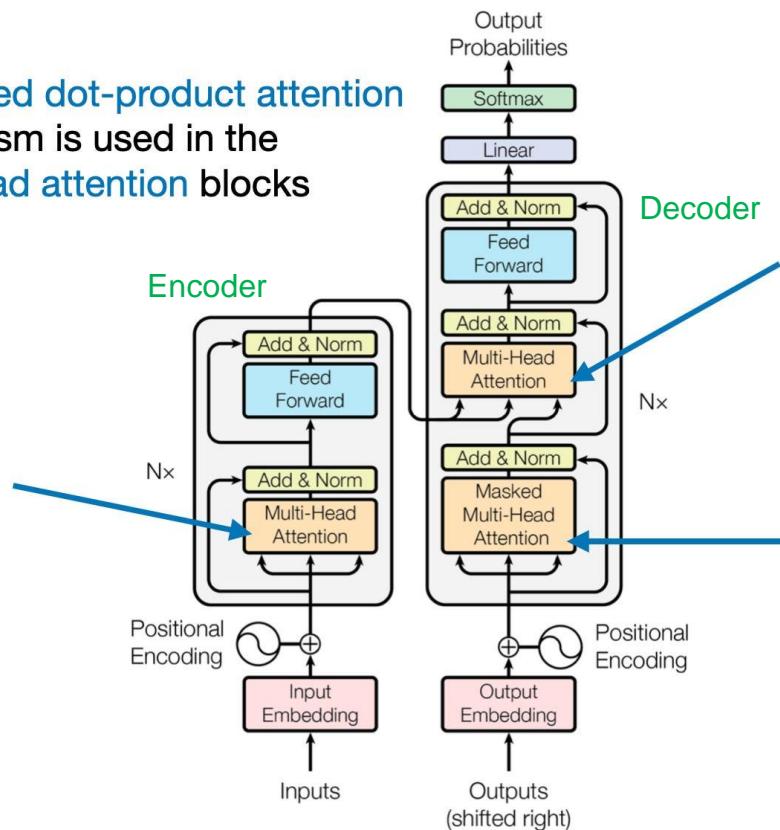
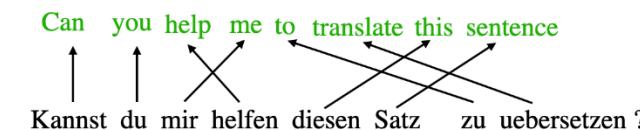
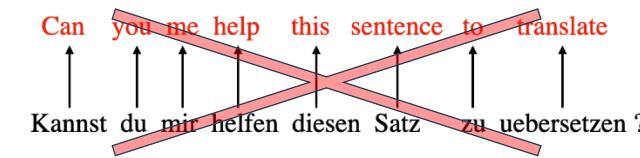


Image Source: <https://magazine.sebastianraschka.com/p/understanding-and-coding-self-attention>
(adapted from Vaswani et al. 2017, p. 3)

Attention Is All You Need

Vaswani et al. 2017

- **Self-attention:** give access to all sequence elements at each time step
- determine which words are most important in a specific context:



Transformers (Deep Neural Networks)

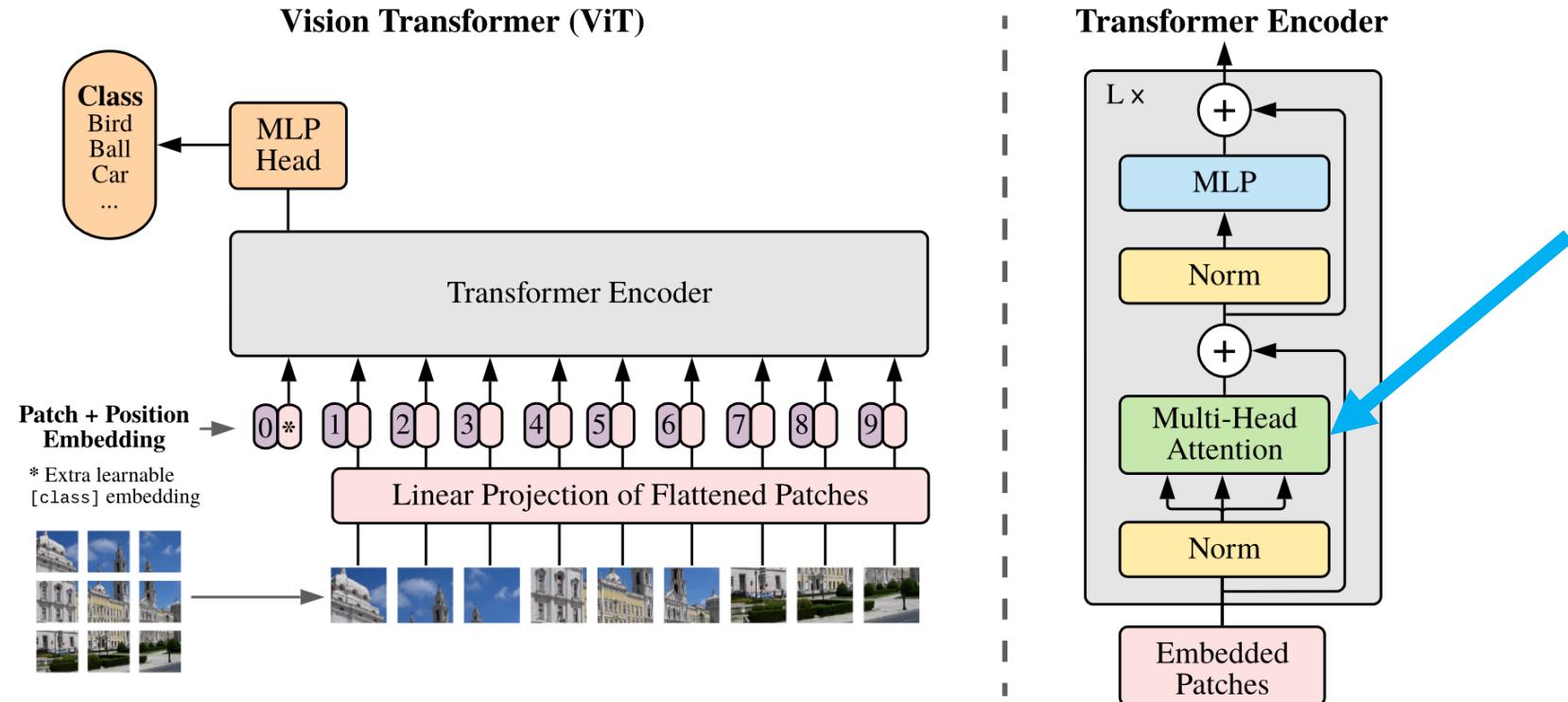
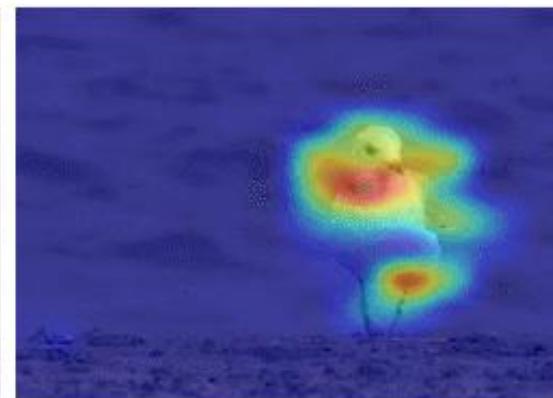


Image Source: Dosovitskiy et al. 2020, <https://arxiv.org/abs/2010.11929>, p. 3

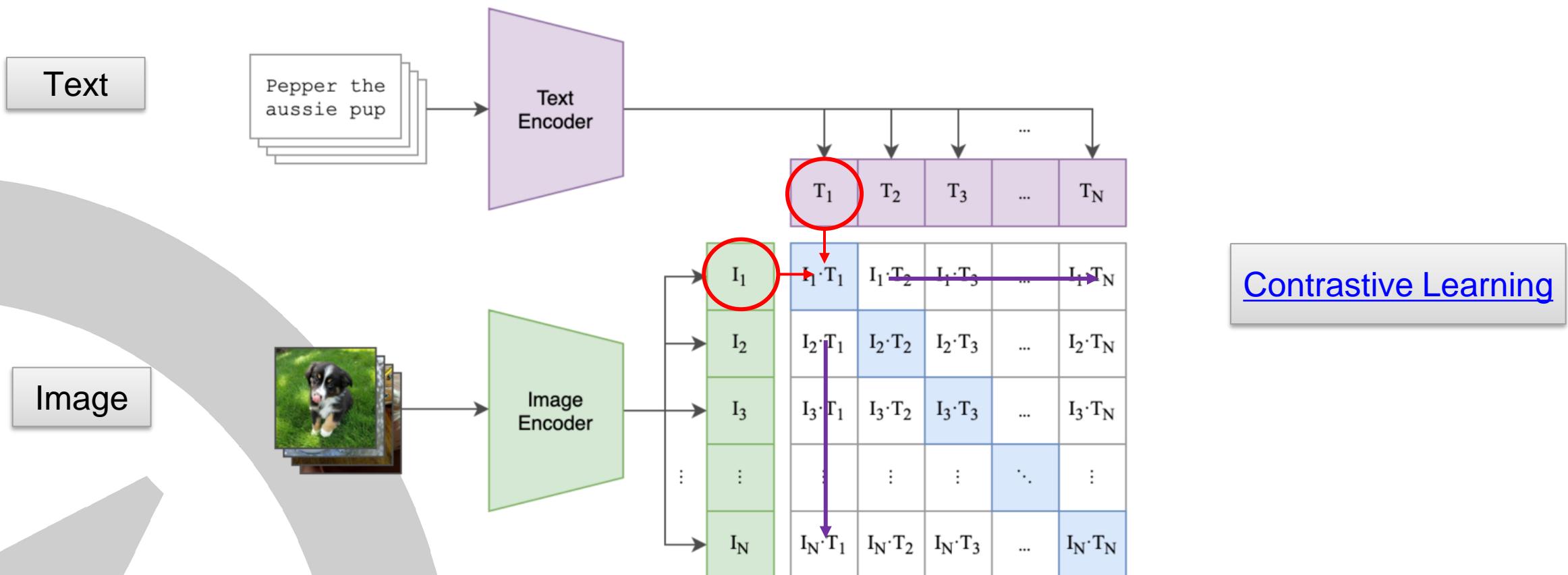
Transformers (Deep Neural Networks)



- **Attention** helps the network to focus on the necessary features/object

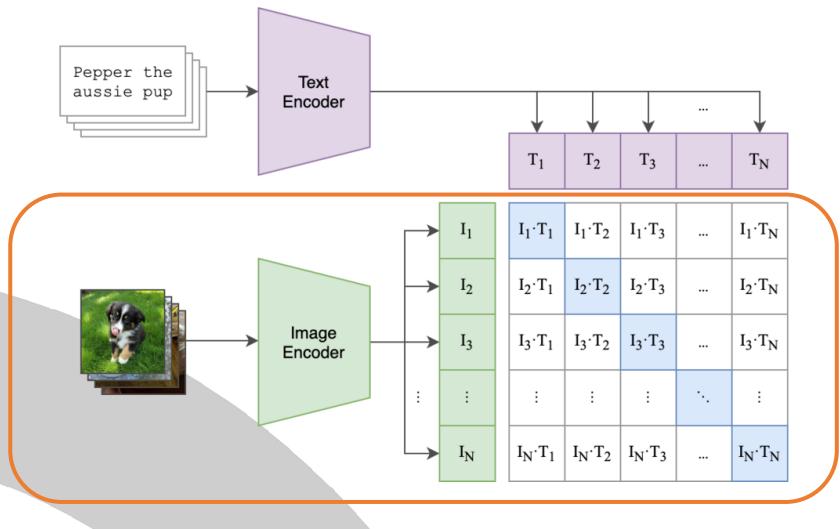
Image Source: <https://theaisummer.com/xai/>

INTRODUCTION: VISION-LANGUAGE MODELS

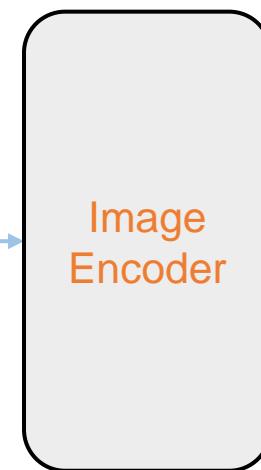
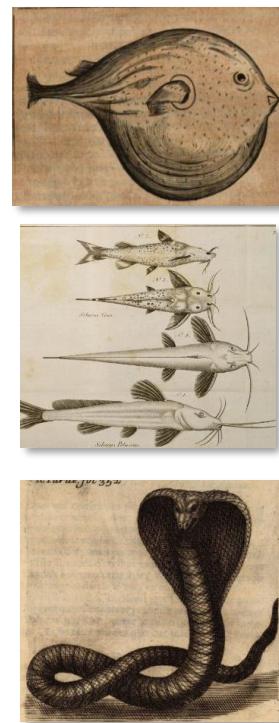


Source: Radford et al. 2021, <https://arxiv.org/abs/2103.00020>, p. 2)

CONCEPT: MULTIMODAL EMBEDDING



Image

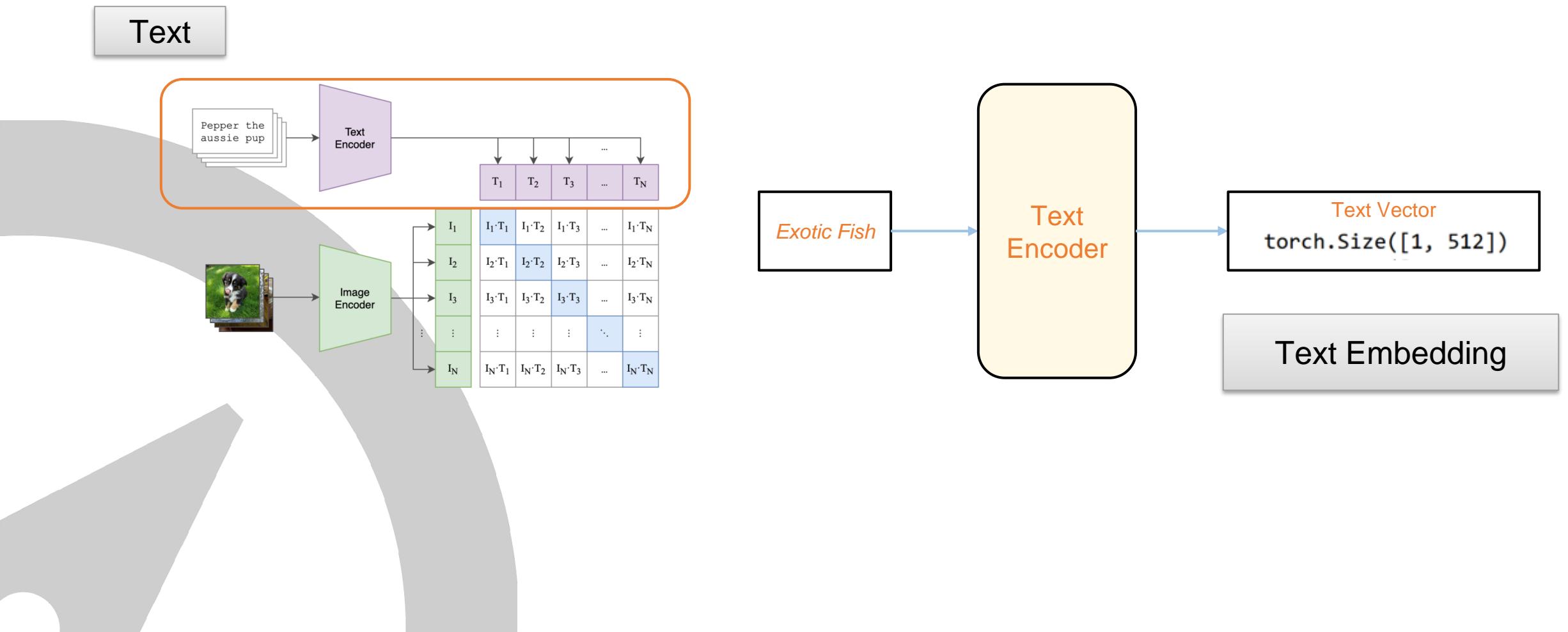


$$\begin{matrix} & 1 & 2 & \cdots & n \\ 1 & a_{11} & a_{12} & \cdots & a_{1n} \\ 2 & a_{21} & a_{22} & \cdots & a_{2n} \\ 3 & a_{31} & a_{32} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m & a_{m1} & a_{m2} & \cdots & a_{mn} \end{matrix}$$

`torch.Size([22299, 512])`

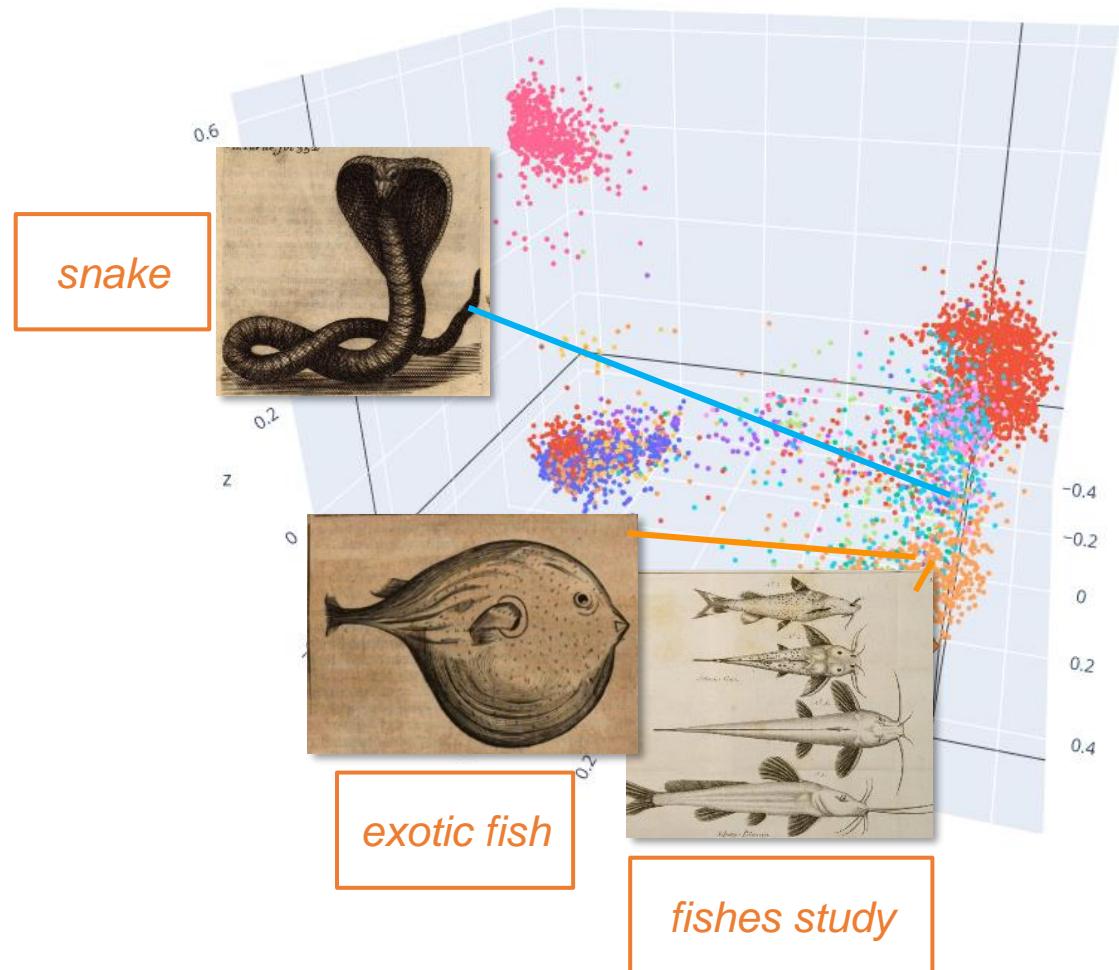
Image Embedding

CONCEPT: MULTIMODAL EMBEDDING



CONCEPT: SEMANTIC VECTOR SPACE

- Texts, images or other data are converted into vectors (series of numbers). These vectors are located in a common high-dimensional space.
- Semantically similar content is located closer together, while dissimilar content is further apart.



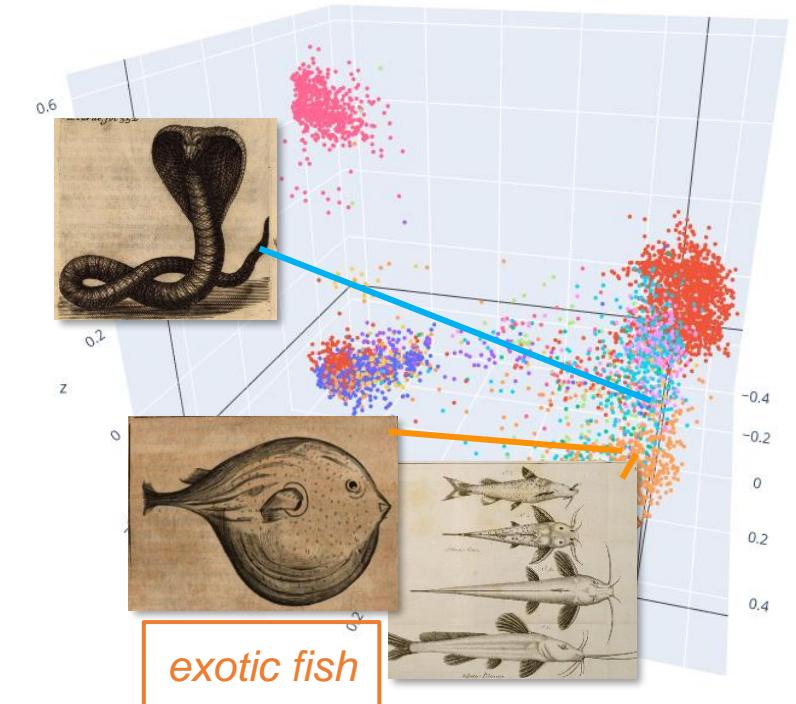
CONCEPT: SEMANTIC SEARCH

Semantic Search

- Based on vector representations (embeddings) of text, images, audio
- Calculates similarity in vector space (e.g. cosine similarity)
- Recognizes meaning and context, not just words

Advantages

- Takes synonyms, paraphrases and context into account
- Transferable to multimodal data (text ↔ image)



CONCEPT: SEMANTIC SEARCH

🔍 Similarity Metrics

- Cosine similarity: Measures angle between two vectors.
Very commonly used for text and multimodal embeddings (BERT, CLIP, etc.).
- Advantage: Independent of the absolute size of the vectors

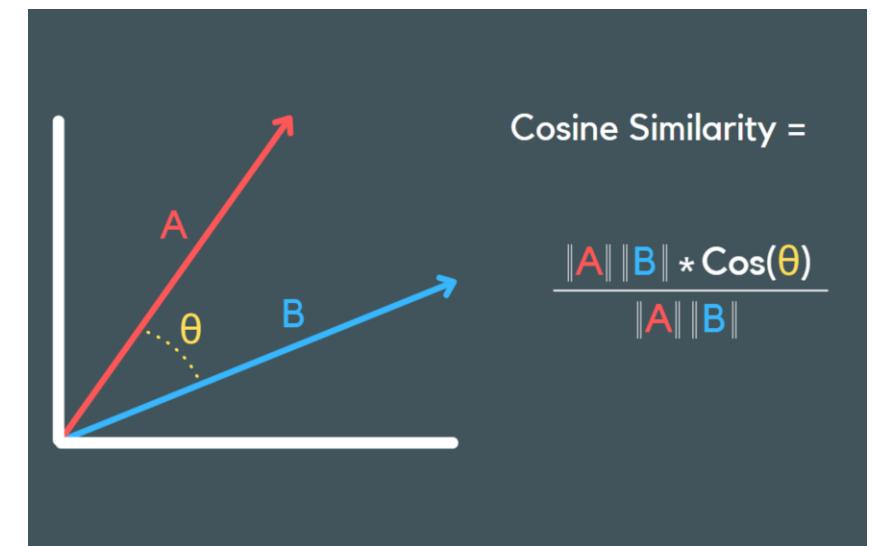


Image source: Shivanandan, [Cosine Similarity Explained](#), 2025

TUTORIAL: IMAGE RETRIEVAL WITH CLIP

→ https://github.com/ONiT-project/tutorial_image_retrieval_CLIP

The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** jupyter image_retrieval_with_CLIP_tutorial_DHRingVo Last Checkpoint: vor einer Stunde (autosaved)
- Toolbar:** File, Edit, View, Insert, Cell, Kernel, Navigate, Widgets, Help, Run, Cell, Code, etc.
- Header:** Not Trusted, CLIP-tutorial
- Section 1: 1 Image Retrieval with CLIP**
 - In [1]:** # Import packages
import numpy as np
import torch
from PIL import Image
import pandas as pd
import os
import open_clip
from tqdm.notebook import tqdm
import random
import matplotlib.pyplot as plt
- Section 1.1: Download Image Dataset**

This notebook uses the image data and annotations provided by the FWF-funded project [Ottoman Nature in Travelogues, 1501–1850: A Digital Analysis \(ONiT\)](#). The images that we will be using were extracted from travelogues about the Ottoman Empire in English, French, German and Latin language that were printed between 1501 and 1850 and survived in the holdings of the [Austrian National Library \(ÖNB\) in Vienna](#). The images have been scanned in the course of the [Austrian Books Online](#) project – a public private partnership between the ÖNB and Google Books.

 1. Download the ZIP and CSV files from the following URL: <https://1drv.ms/f/c/869f28ab041d44d9/ErTeV8fMekFCvtgxDOCxZoBCxCoxLJYUBOjam6rwBSdw?e=liK1F>
 2. Unzip the image data folder and save it in the Download folder
 3. Inspect the image data. Which types of images do you see?
- Section 1.2: Data Preprocessing**
 - 1.2.1: Load Data**

This cell loads the data from the downloaded CSV file. Filenames, paths, and other metadata of the subset of 786 images from the downloaded datasets are stored in lists. These will be used later to preview and load the images with the data loader and compute the image embeddings with CLIP.

CONTENT

1. The ONiT Project

- Introduction & Research Questions
- Corpus
- Approach

2. Multimodal Image Retrieval

- State of the Art
- Methodology
- ONiT Explorer Demo
- Machine Learning Training

3. Results and Conclusion

- Results
- Conclusions

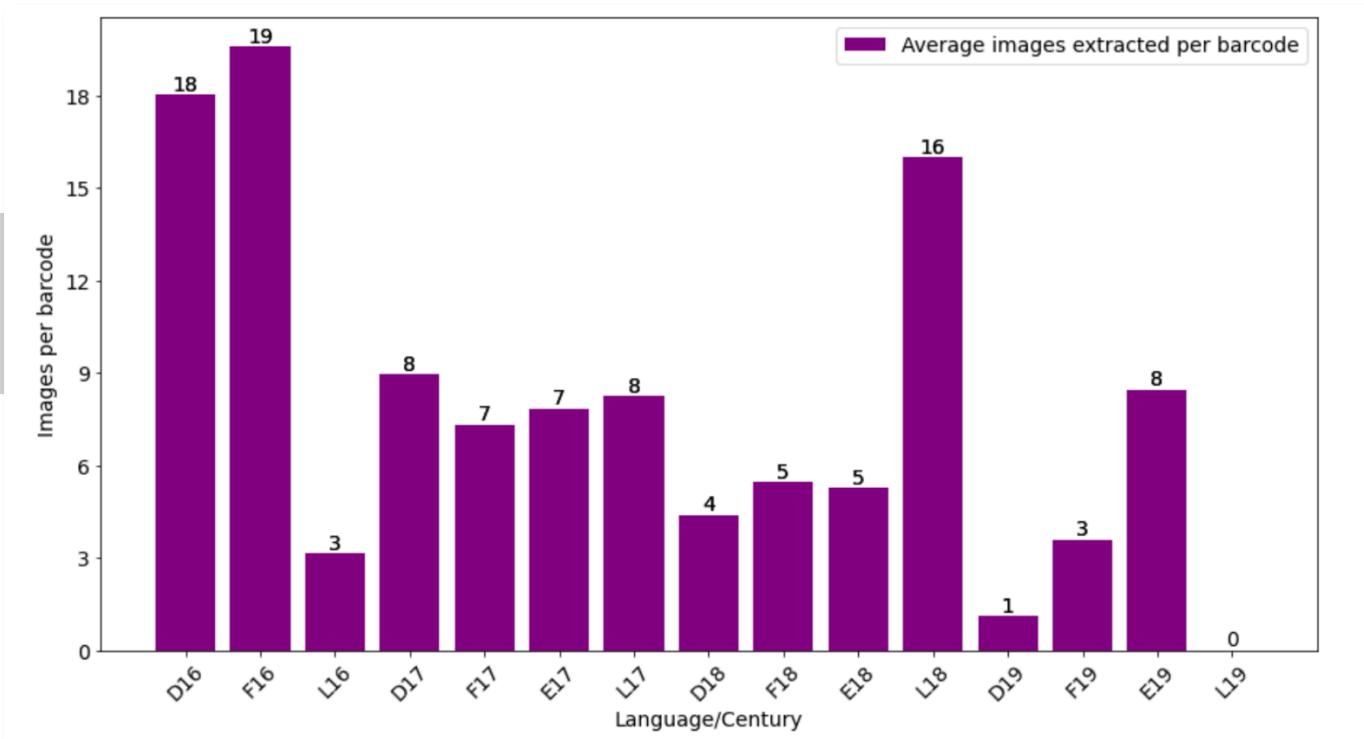
RESULTS: ONiT IMAGE DATASET



- **dataset with images and annotations**
- **more detailed annotations on animal species**
- final result: **8,700 images with 56 annotated classes**
in total 21,300 annotations

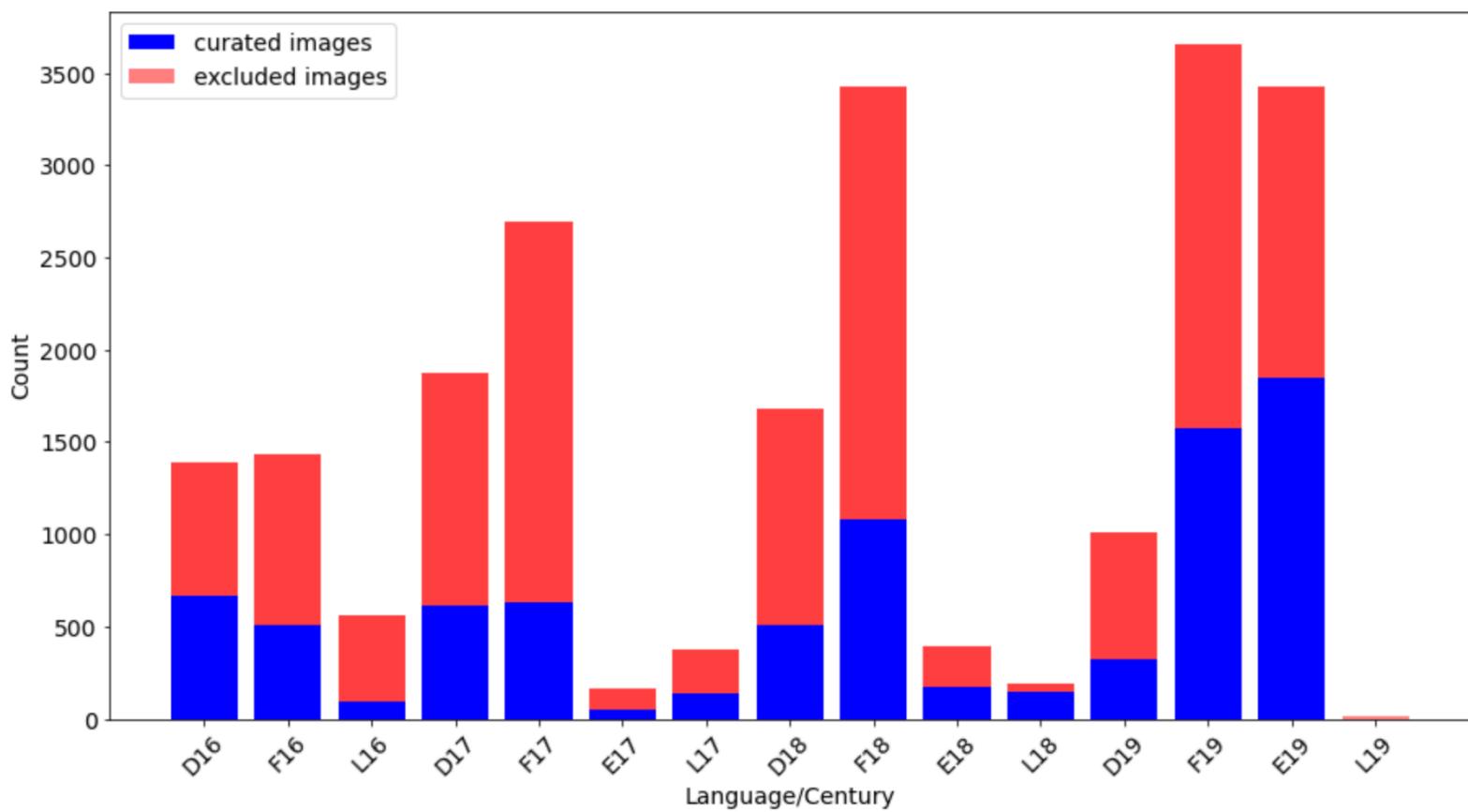
Buxbaum, Johann Christian, *Plantarum Minus Cognitarum Centuria: I. : complectens Plantas Circa Byzantium & In Oriente Observatas*, Ex Typographia Academiae: St. Petersburg 1728, <<http://data.onb.ac.at/ABO/%2BZ184478500>>, p. [61].

RESULTS: ONiT IMAGE DATASET



first larger scale data on pictorial representation of nature in travelogues on the Ottoman Empire

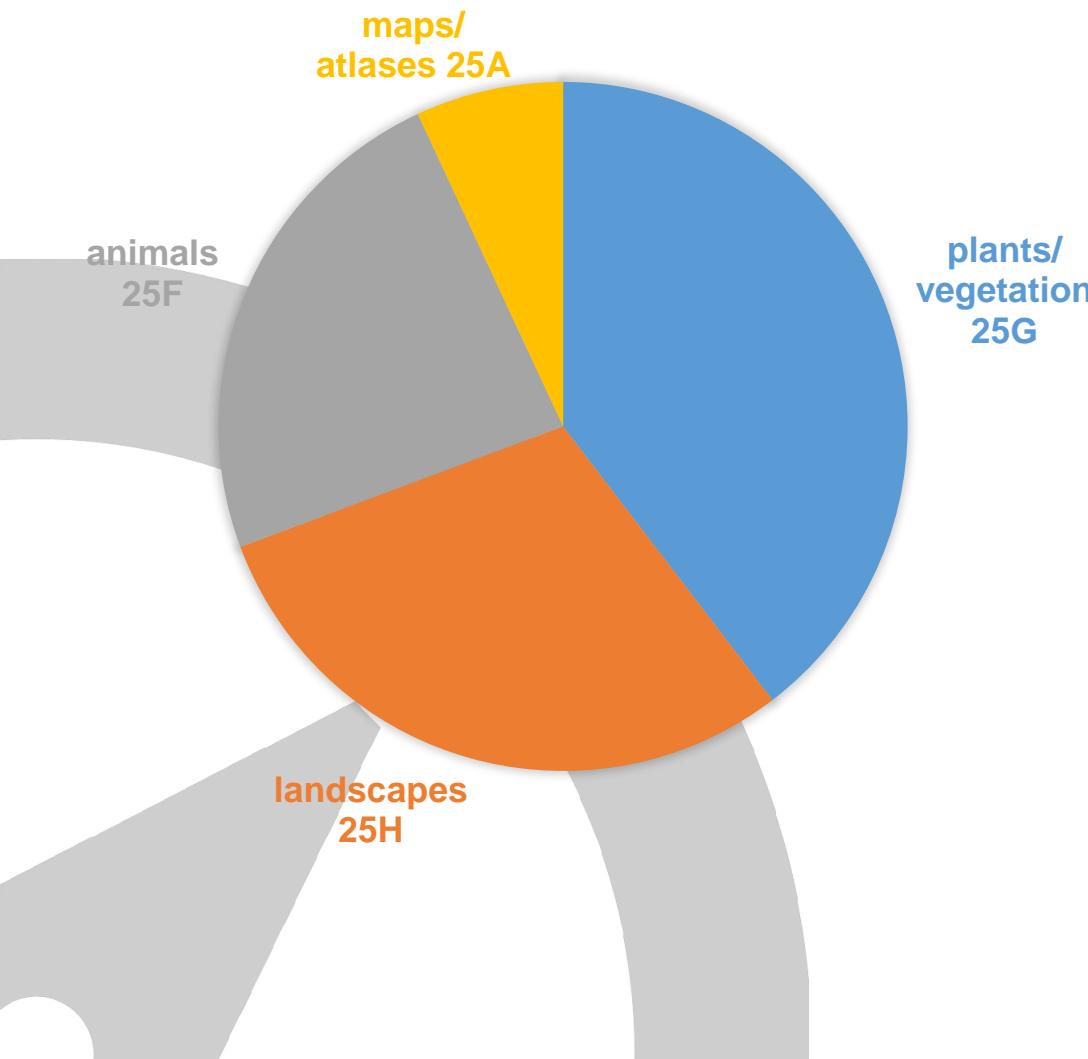
RESULTS: ONiT IMAGE DATASET



[dataset available](#) on our
GitHub repository

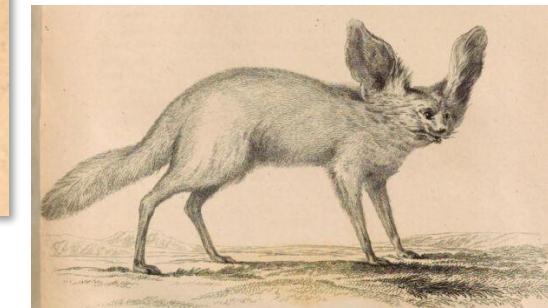
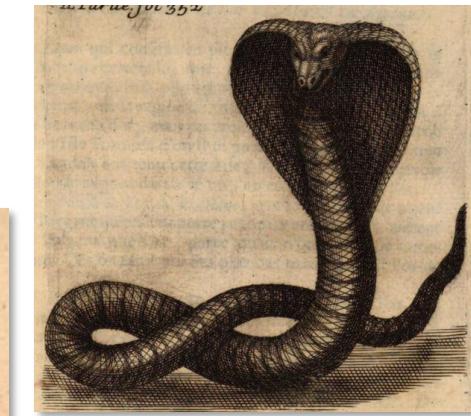
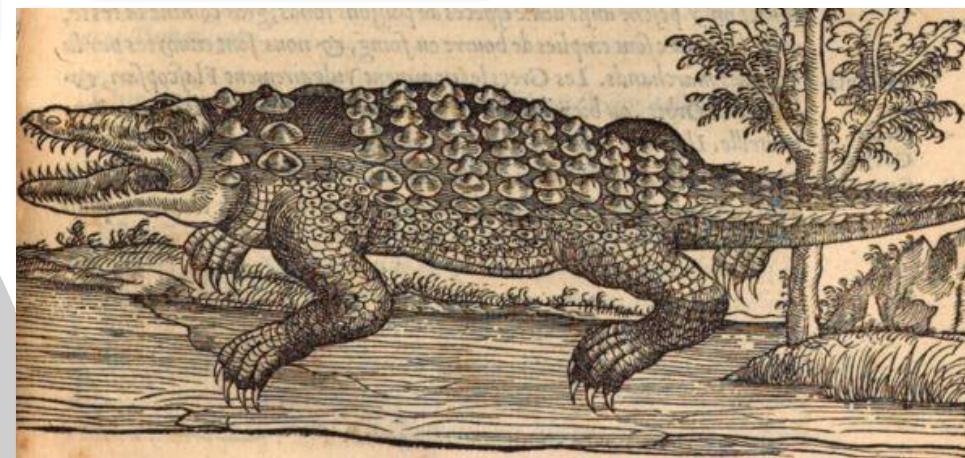
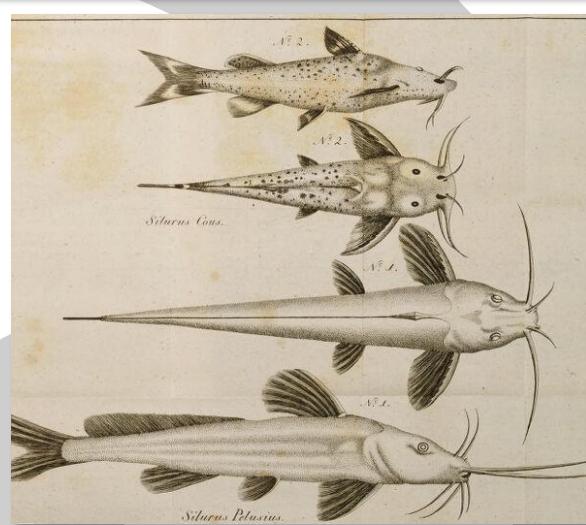
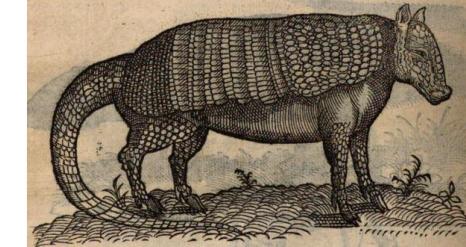
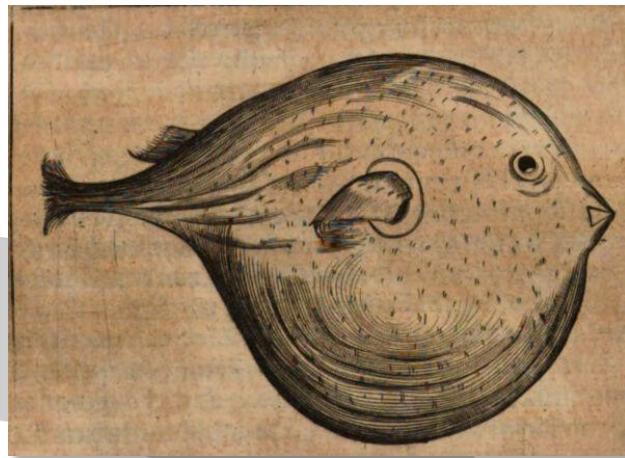
highest percentage of nature
representations in Latin
travelogues from eighteenth
century

RESULTS: ONiT IMAGE DATASET

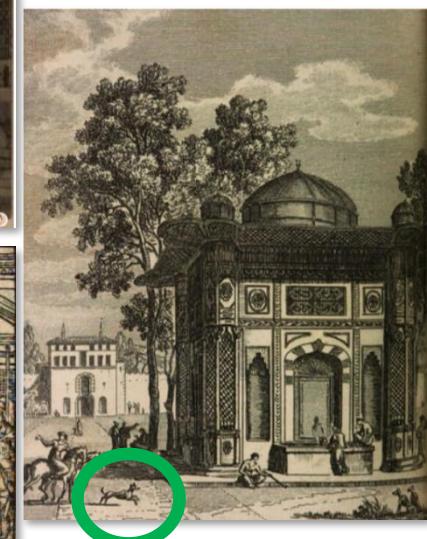


Class (ICONCLASS)	Percentage in corpus
horses and kindred animals 46C1314	29%
birds 25F3	26%
fabulous animals 25FF	10%
camels 25F24	7%
dogs 24B11	6%
snakes 25F42	6%
beasts of prey predatory animals lions 25F23L	6%
cattle 47121	5%
fishes 25F6	3%
hoofed animals goats 25F24	3%

RESULTS: DISTINCT REPRESENTATIONS



RESULTS: DOGS



RESULTS: MODEL EVALUATION

Class (ICONCLASS)	Total Examples	Recall@K (k=200)	Precision@ K (k=200)	Recall@K (k=1,000)	Precision@ K (k=1,000)	R- Precision
plants/vegetation 25G	5,708	0.02	0.71	0.1	0.62	0.51
landscapes 25H	4,234	0.05	0.98	0.22	0.95	0.7
animals 25F	3,506	0.05	0.92	0.17	0.69	0.46
maps/atlasses 25A	1,081	0.16	0.89	0.68	0.72	0.69

Pretrained CLIP Model						
plants/vegetation 25G	5,708	0.03	0.86	0.08	0.45	0.17
landscapes 25H	4,234	0.05	0.98	0.22	0.9	0.67
animals 25F	3,506	0.05	0.93	0.17	0.59	0.28
maps/atlasses 25A	1,081	0.17	0.94	0.71	0.77	0.75

Fine-tuned

Base model

RESULTS: MODEL EVALUATION

Class (ICONCLASS)	Total Examples	Recall@K (k=200)	Precision@ K (k=200)	Recall@K (k=1,000)	Precision@ K (k=1,000)	R- Precision
plants/vegetation 25G	5,708	0.02	0.71	0.1	0.62	0.51
landscapes 25H	4,234	0.05	0.98	0.22	0.95	0.7
animals 25F	3,506	0.05	0.92	0.17	0.69	0.46
maps/atlases 25A	1,081	0.16	0.89	0.68	0.72	0.69

Pretrained CLIP Model						
Class (ICONCLASS)	Total Examples	Recall@K (k=200)	Precision@ K (k=200)	Recall@K (k=1,000)	Precision@ K (k=1,000)	R- Precision
plants/vegetation 25G	5,708	0.03	0.86	0.08	0.45	0.17
landscapes 25H	4,234	0.05	0.98	0.22	0.9	0.67
animals 25F	3,506	0.05	0.93	0.17	0.59	0.28
maps/atlases 25A	1,081	0.17	0.94	0.71	0.77	0.75

Fine-tuned

Base model

RESULTS: MODEL EVALUATION

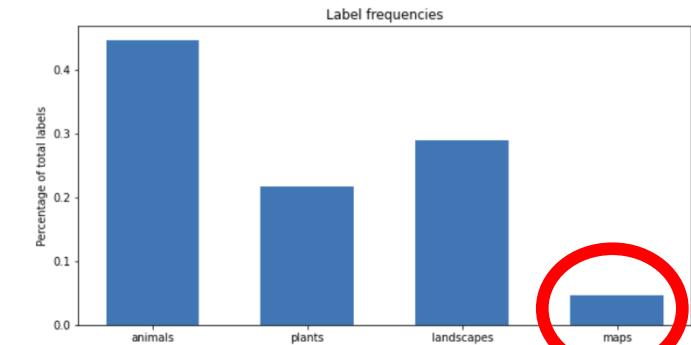
Class (ICONCLASS)	Total Examples	Recall@K (k=200)	Precision@ K (k=200)	Recall@K (k=1,000)	Precision@ K (k=1,000)	R- Precision
plants/vegetation 25G	5,708	0.02	0.71	0.1	0.62	0.51
landscapes 25H	4,234	0.05	0.98	0.22	0.95	0.7
animals 25F	3,506	0.05	0.92	0.17	0.69	0.46
maps/atlases 25A	1,081	0.16	0.89	0.68	0.72	0.69

Pretrained CLIP Model						
Class (ICONCLASS)	Total Examples	Recall@K (k=200)	Precision@ K (k=200)	Recall@K (k=1,000)	Precision@ K (k=1,000)	R- Precision
plants/vegetation 25G	5,708	0.03	0.86	0.08	0.45	0.17
landscapes 25H	4,234	0.05	0.98	0.22	0.9	0.67
animals 25F	3,506	0.05	0.93	0.17	0.59	0.28
maps/atlases 25A	1,081	0.17	0.94	0.71	0.77	0.75

Fine-tuned

Base model

RESULTS: MODEL EVALUATION



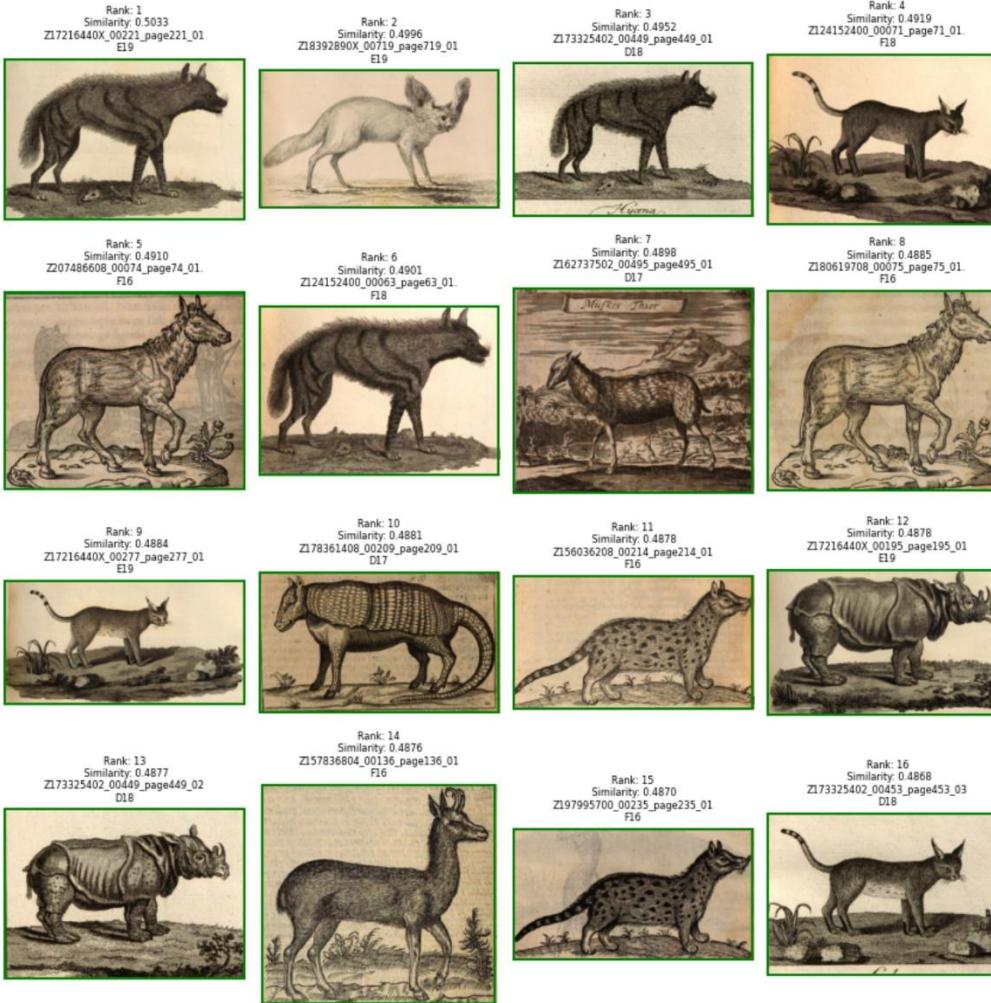
Class (ICONCLASS)	Total Examples	Recall@K (k=200)	Precision@ K (k=200)	Recall@K (k=1,000)	Precision@ K (k=1,000)	R- Precision
plants/vegetation 25G	5,708	0.02	0.71	0.1	0.62	0.51
landscapes 25H	4,234	0.05	0.98	0.22	0.95	0.7
animals 25F	3,506	0.05	0.92	0.17	0.69	0.46
maps/atlases 25A	1,081	0.16	0.89	0.68	0.72	0.69

Pretrained CLIP Model						
plants/vegetation 25G	5,708	0.03	0.86	0.08	0.45	0.17
landscapes 25H	4,234	0.05	0.98	0.22	0.9	0.67
animals 25F	3,506	0.05	0.93	0.17	0.59	0.28
maps/atlases 25A	1,081	0.17	0.94	0.71	0.77	0.75

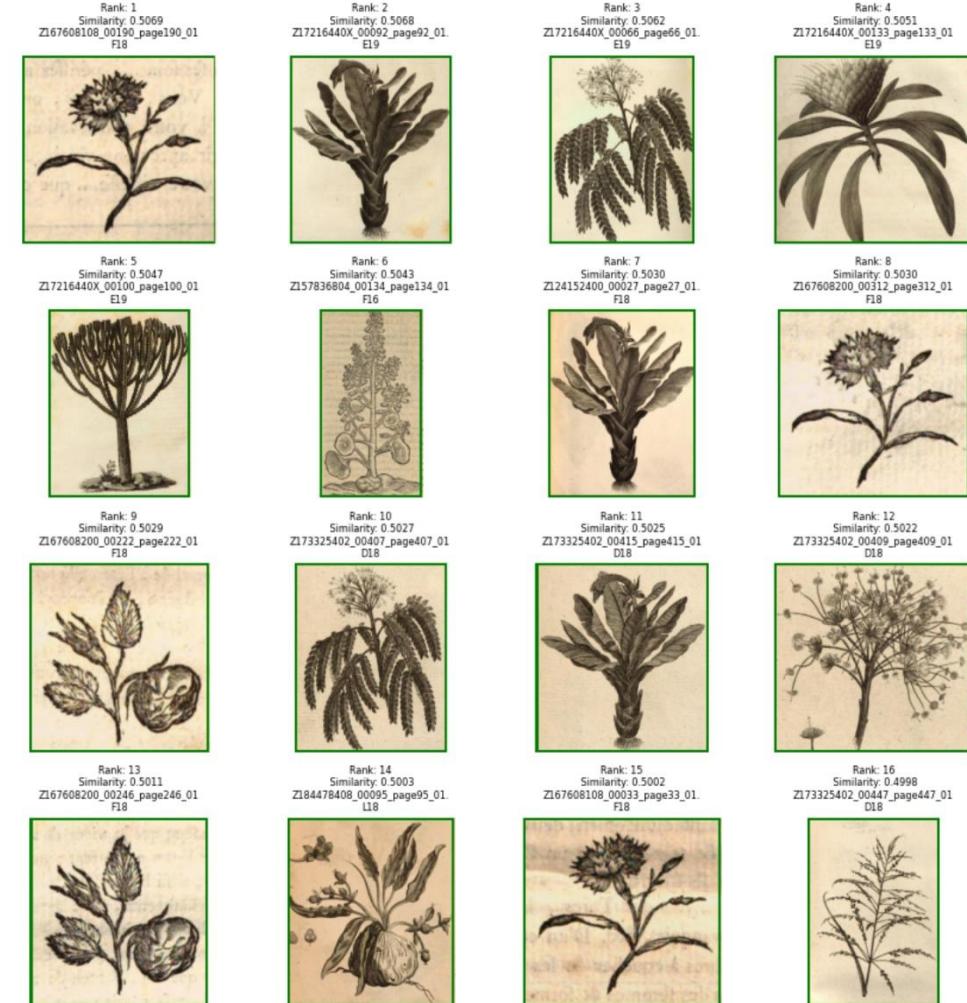


Highest Ranked Retrievals

animals



plants

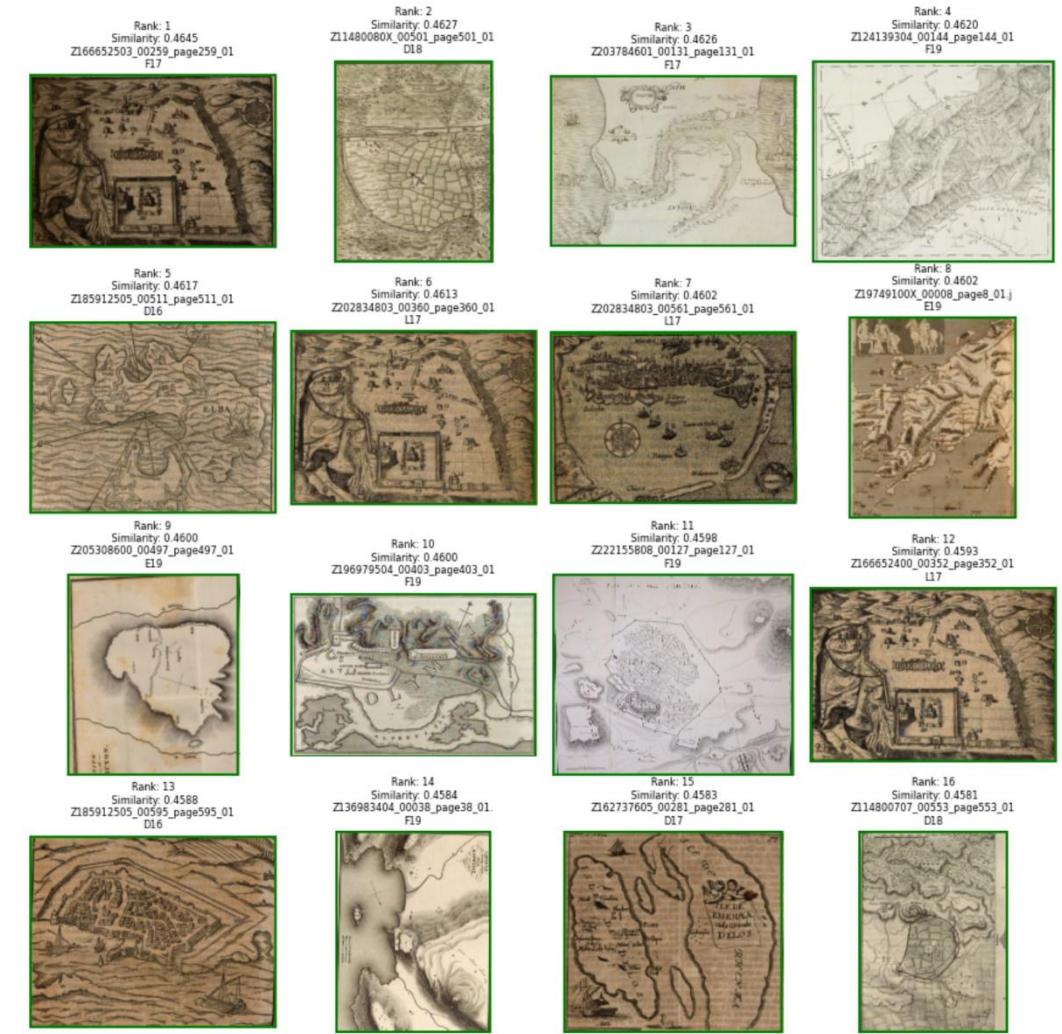


Highest Ranked Retrievals

landscapes



maps

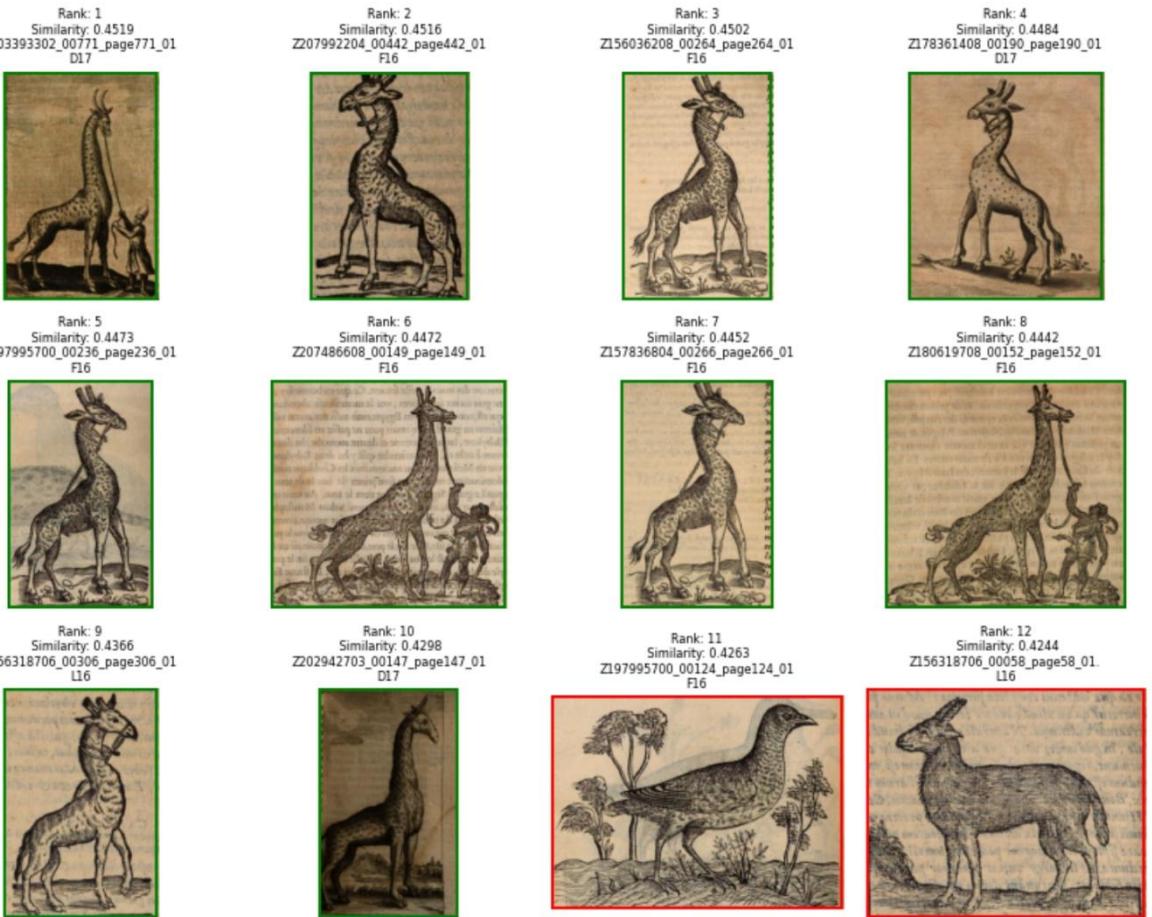


Highest Ranked Retrievals

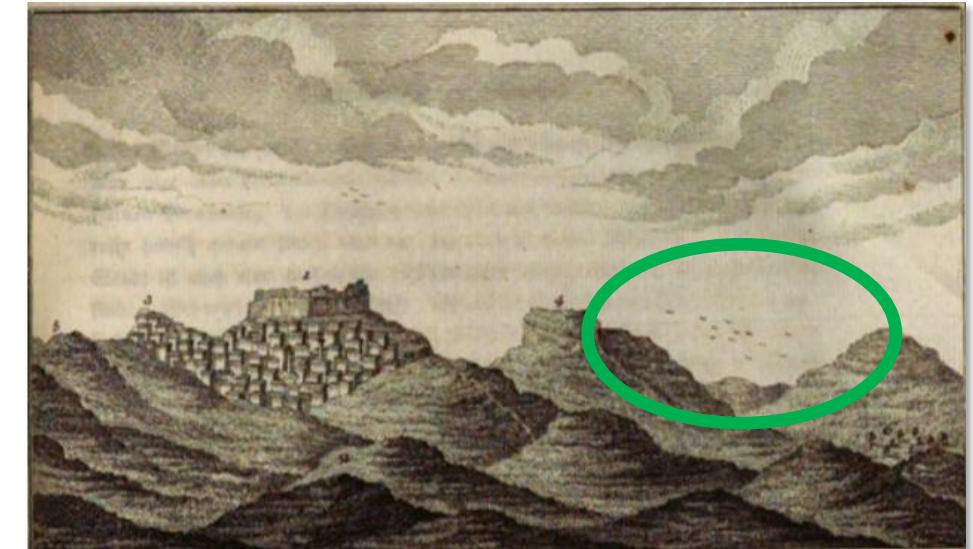
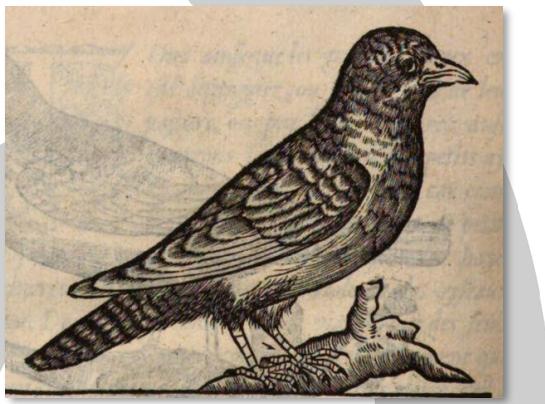
birds



giraffe

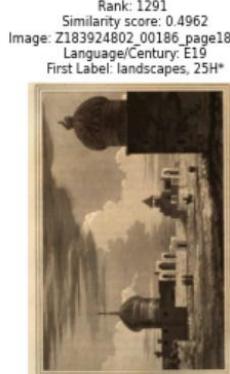
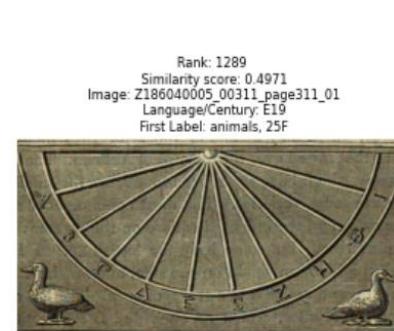
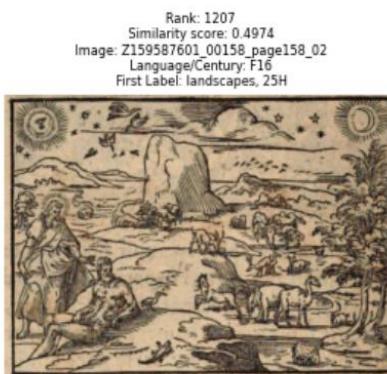
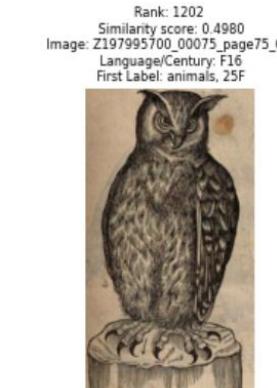
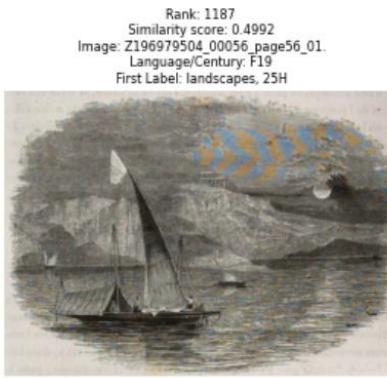
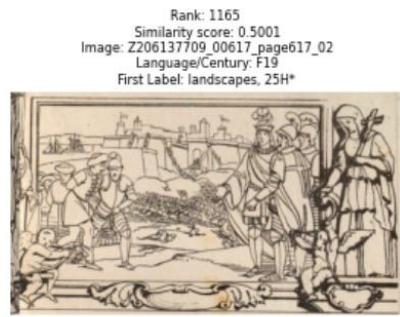
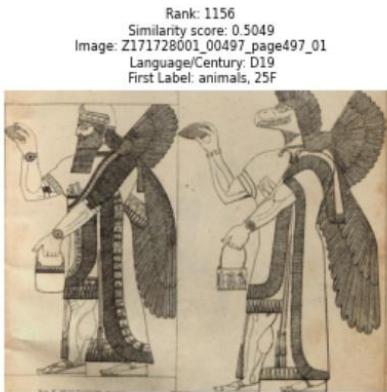


RESULTS: BIRDS

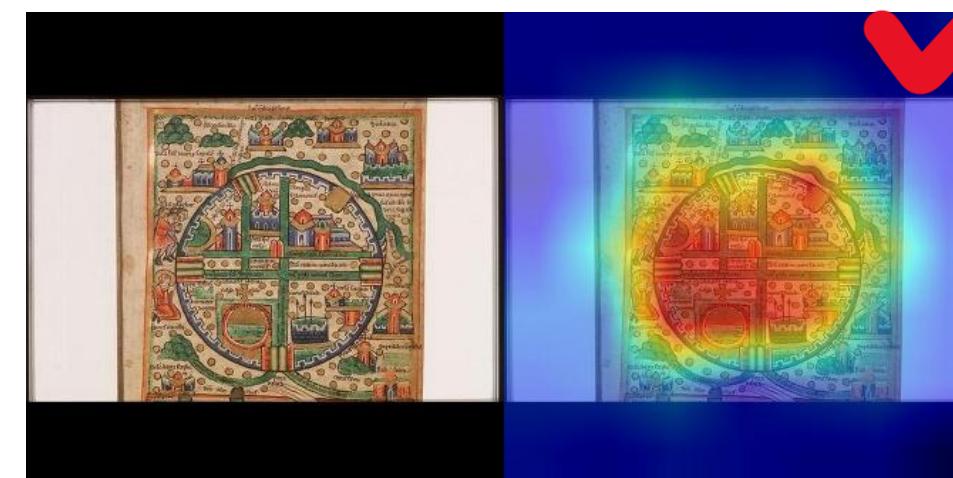
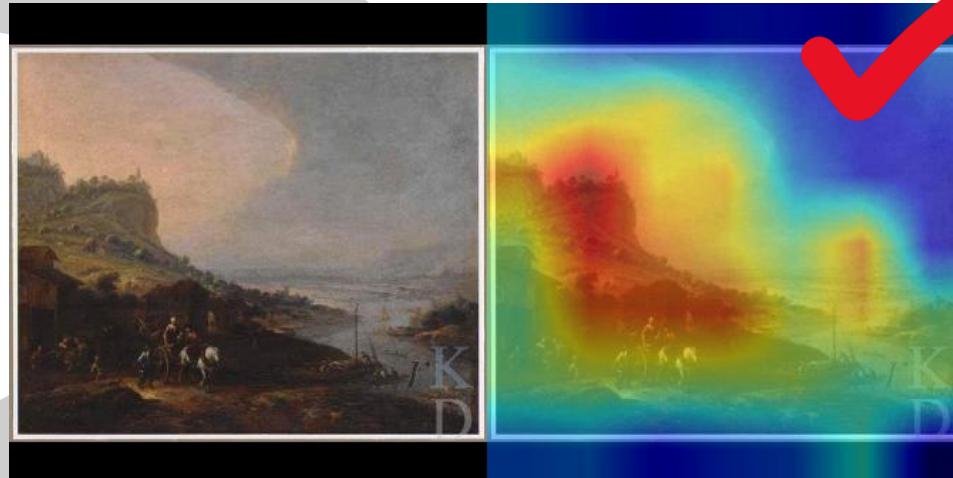
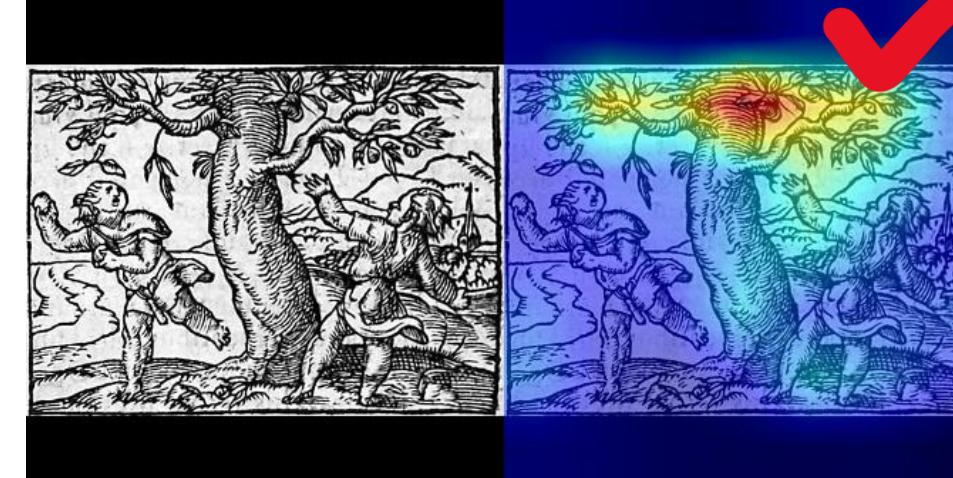


False Negatives: Small Size, Few Distinctive Features

birds



gScoreCAM Evaluation: What works well with CLIP



Source: Vignoli, Michela et al. (2023) *Impact of AI: Gamechanger for Image Classification in Historical Research?* In Konferenzbeiträge der Digital History 2023, Berlin. <<https://doi.org/10.5281/zenodo.8322398>>.

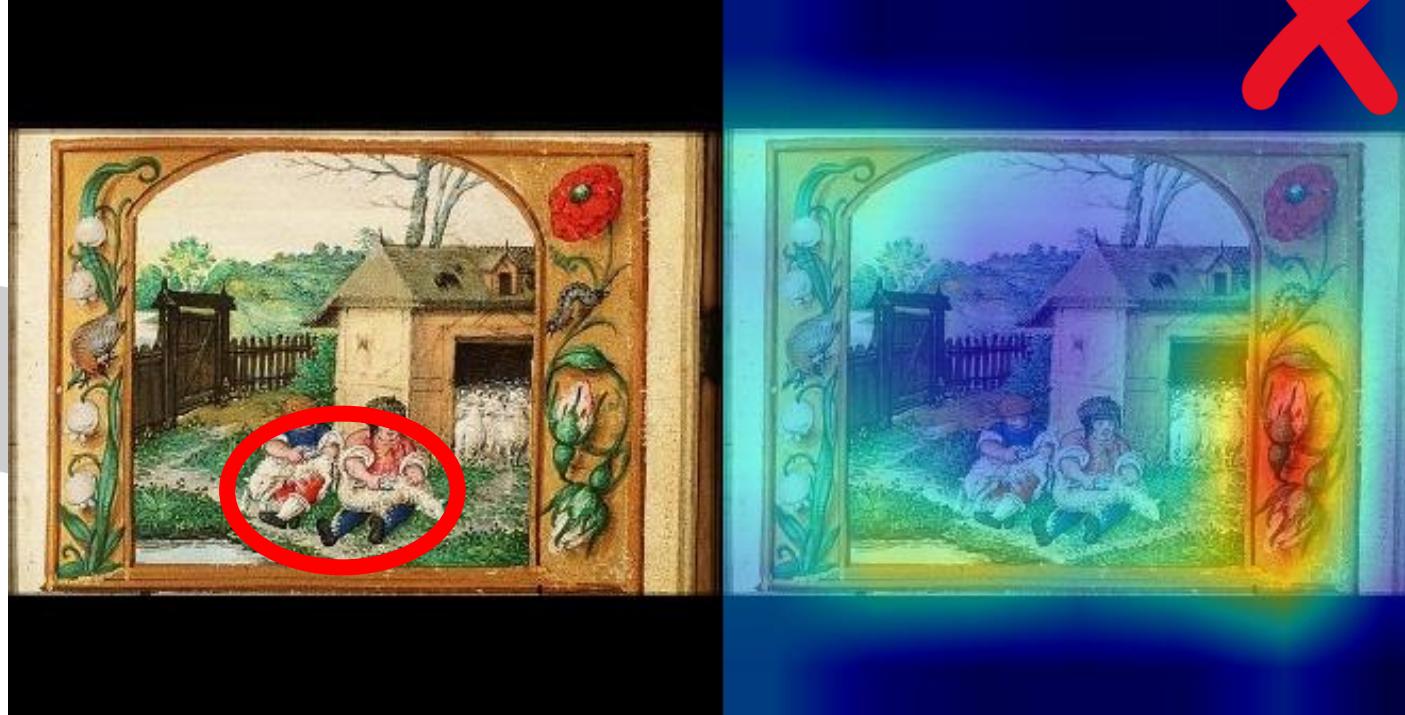
gScoreCAM Evaluation: Small Size, Few Distinctive Features



GS: “an image of animals”

Most probable class: landscapes

gScoreCAM Evaluation: Small Size, Few Distinctive Features



GS: “an image of animals”

Most probable class: maps

CONCLUSIONS

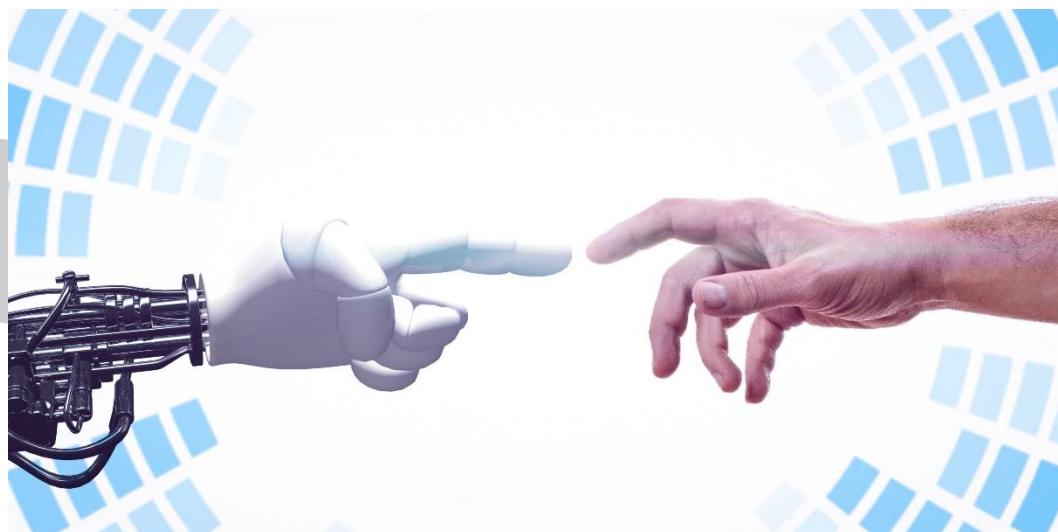
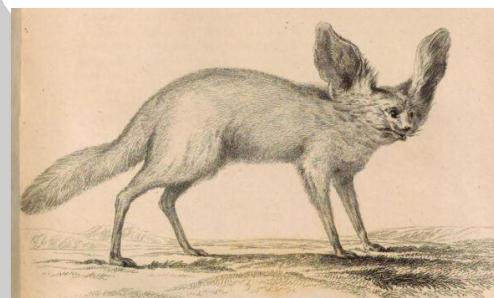
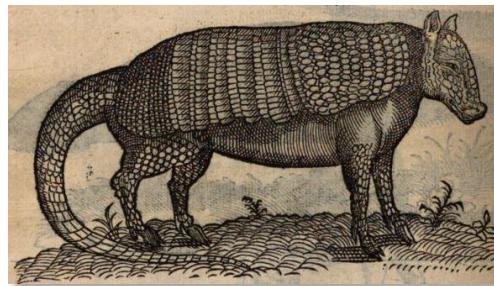
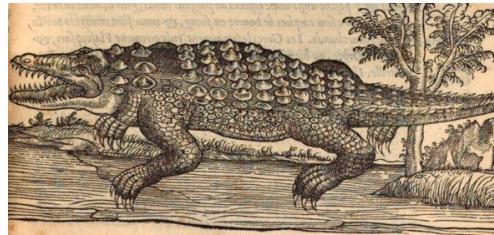


Image Source: <https://pxhere.com/en/photo/1638452> License: CC0

Artificial Intelligence (AI):

- How do quantitative methods support historical research questions?
- How do they impact the underlying epistemology from more traditional, historical methods?

CONCLUSIONS



Mixed results:

- Fine-tuned CLIP partially improves retrieval of nature representations in large, unlabeled collections
- However: model lacks robustness for some of the tested classes
- Fine-tuning time consuming and requires well balanced datasets
- In general: limited recall and precision

CONCLUSIONS



Image Source: <https://pxhere.com/en/photo/1638452> License: CC0

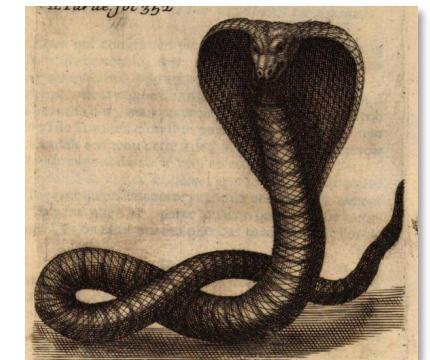
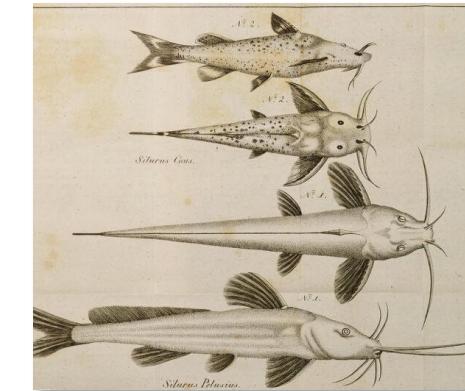
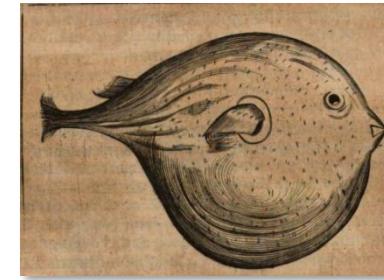
How do AI-methods support historical research?

- Retrieval of relevant elements in large corpora of images and texts
- But: retrieved content contains errors and remains partially incomplete
- Gaining detailed insight still requires a lot of manual work

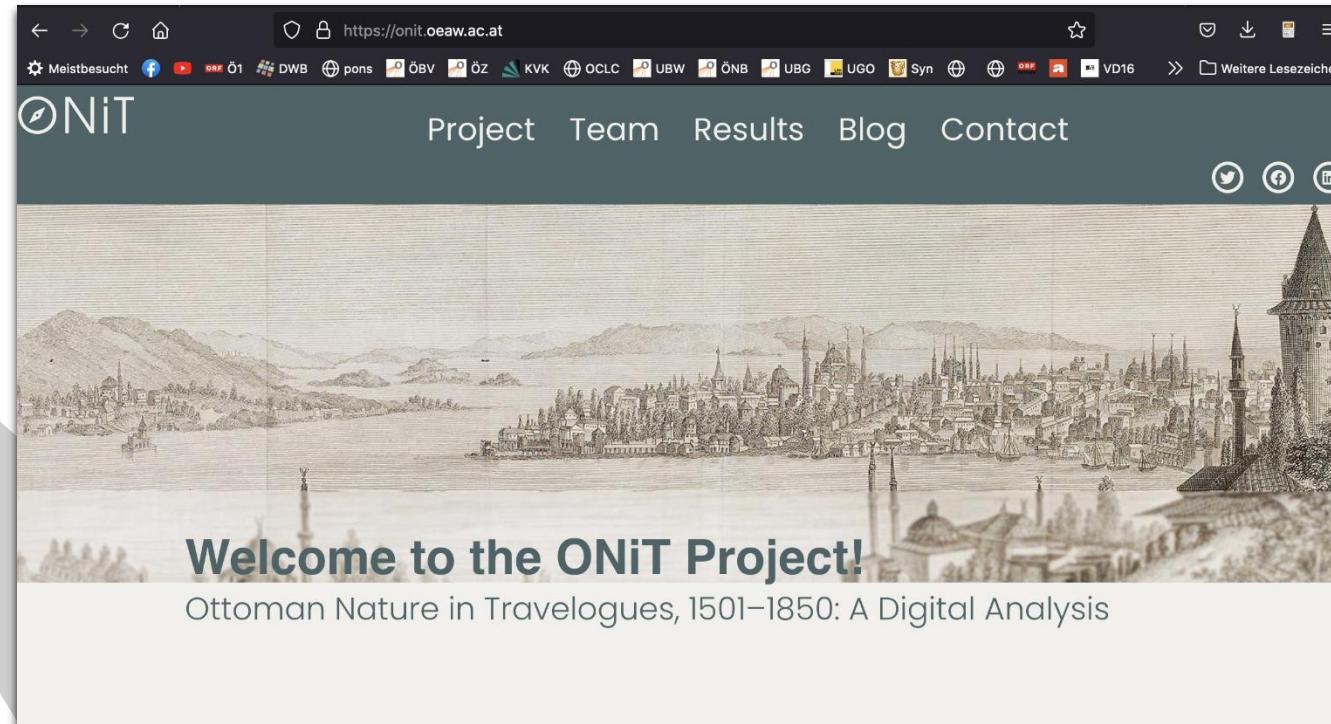
CONCLUSIONS

How does AI impact historical research and methods?

- Heuristics based on similarity ranking between natural language query and image vectors
→ not comprehensive and imprecise
- Bias: Concepts encoded through learning of formal visualisations and semantics from training data (*dataset bias, conceptual bias*) → images resembling CV training data
- Limitations: Good for exploring, but manual work and revisions are required to create comprehensive collections



THANK YOU!



<https://onit.oeaw.ac.at>