Project 1: Detecting Published vs. Unpublished Images in Wartime Archives

Challenges and Opportunities in Analyzing Print Media Images

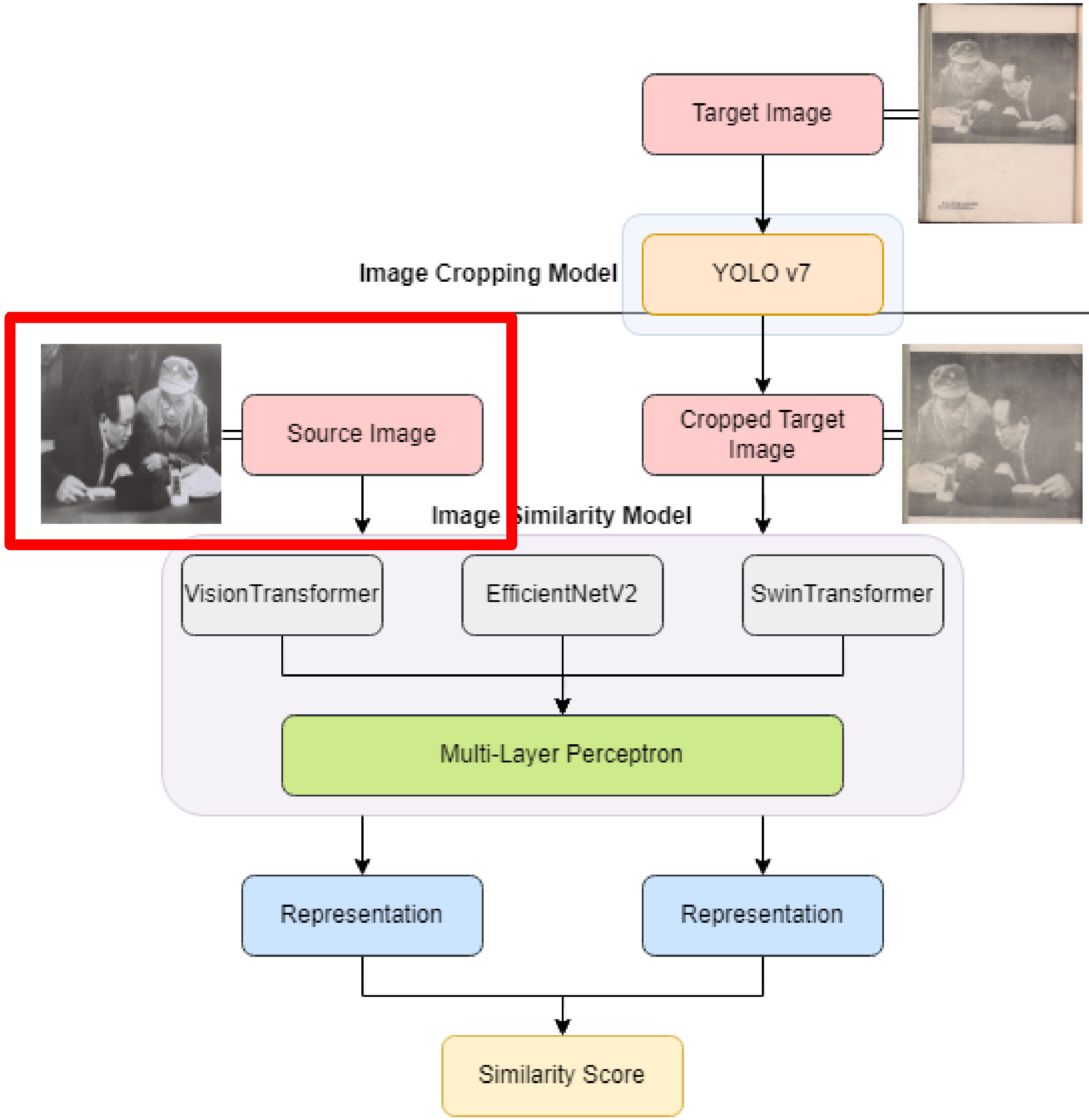
Dissemination of tons of photographs in 20th-century print media

Challenges to humanities scholars tracing image circulation

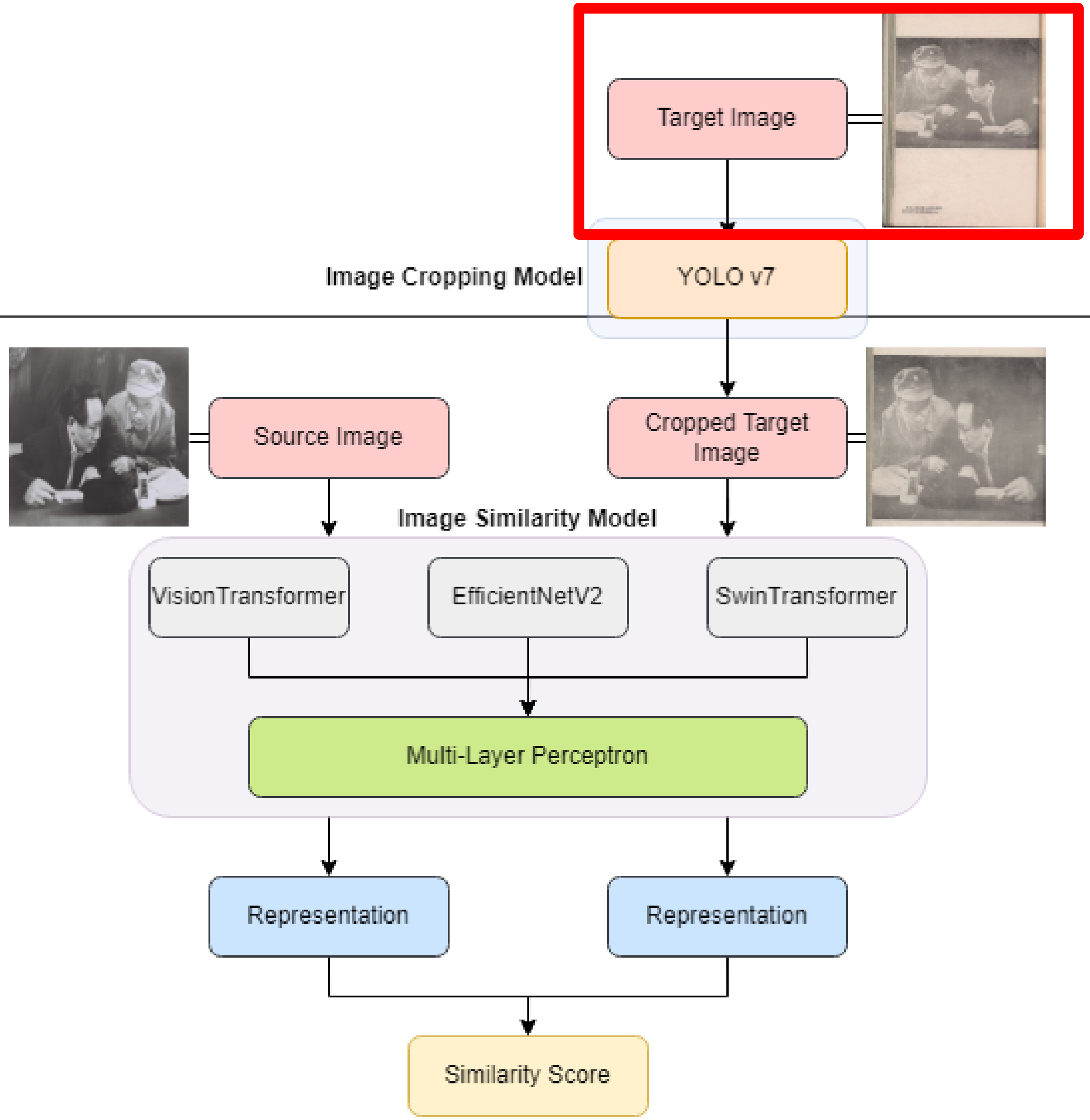
The combination of the traditional method—contextual analysis–in media studies and computer vision technologies

Case Study: Jinchaji Pictorial series, a significant WWII-era photographic publication of the Chinese Communist Party (1942-1948)

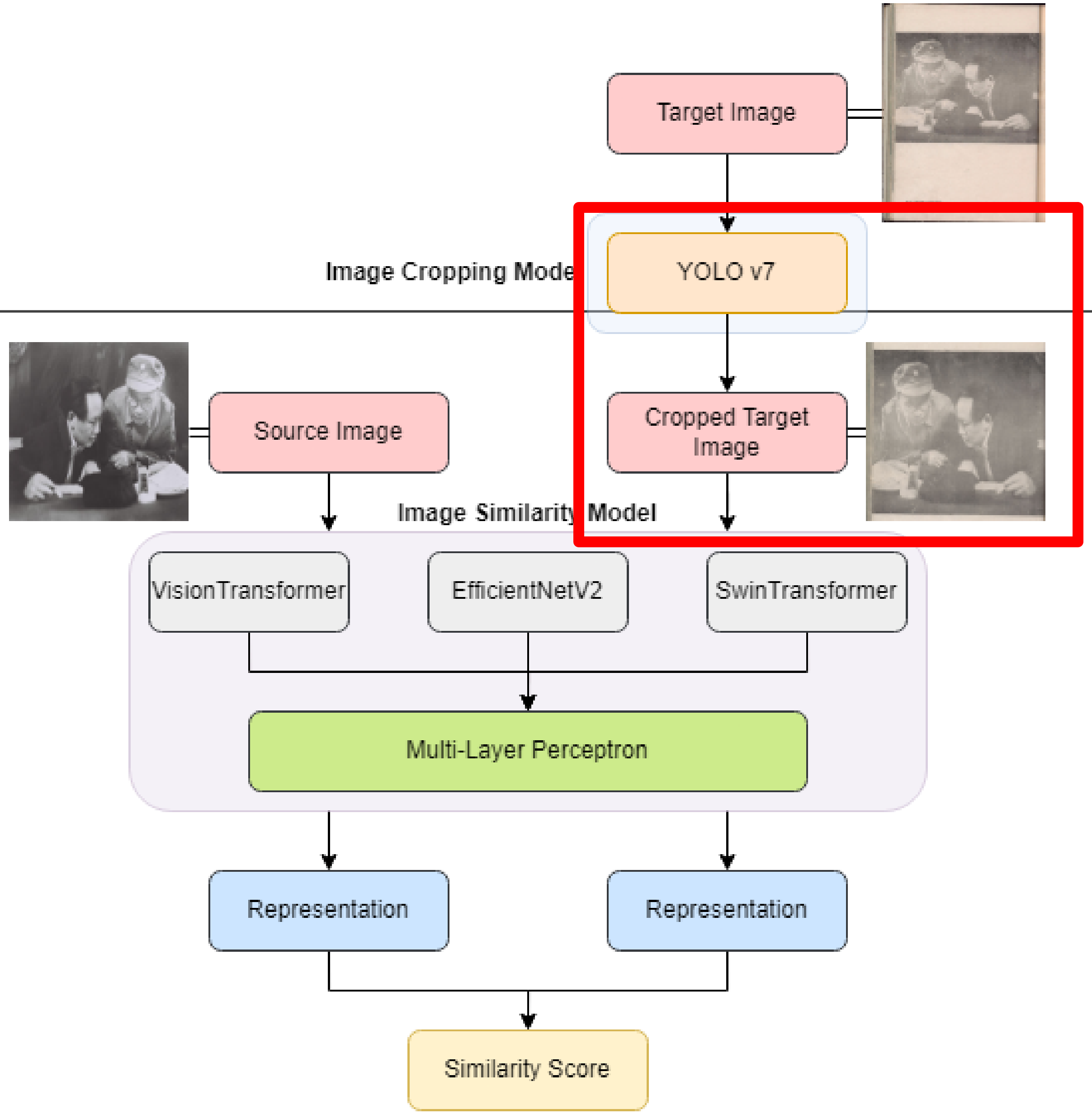
Computer Vision Pipeline for Image Retrieval



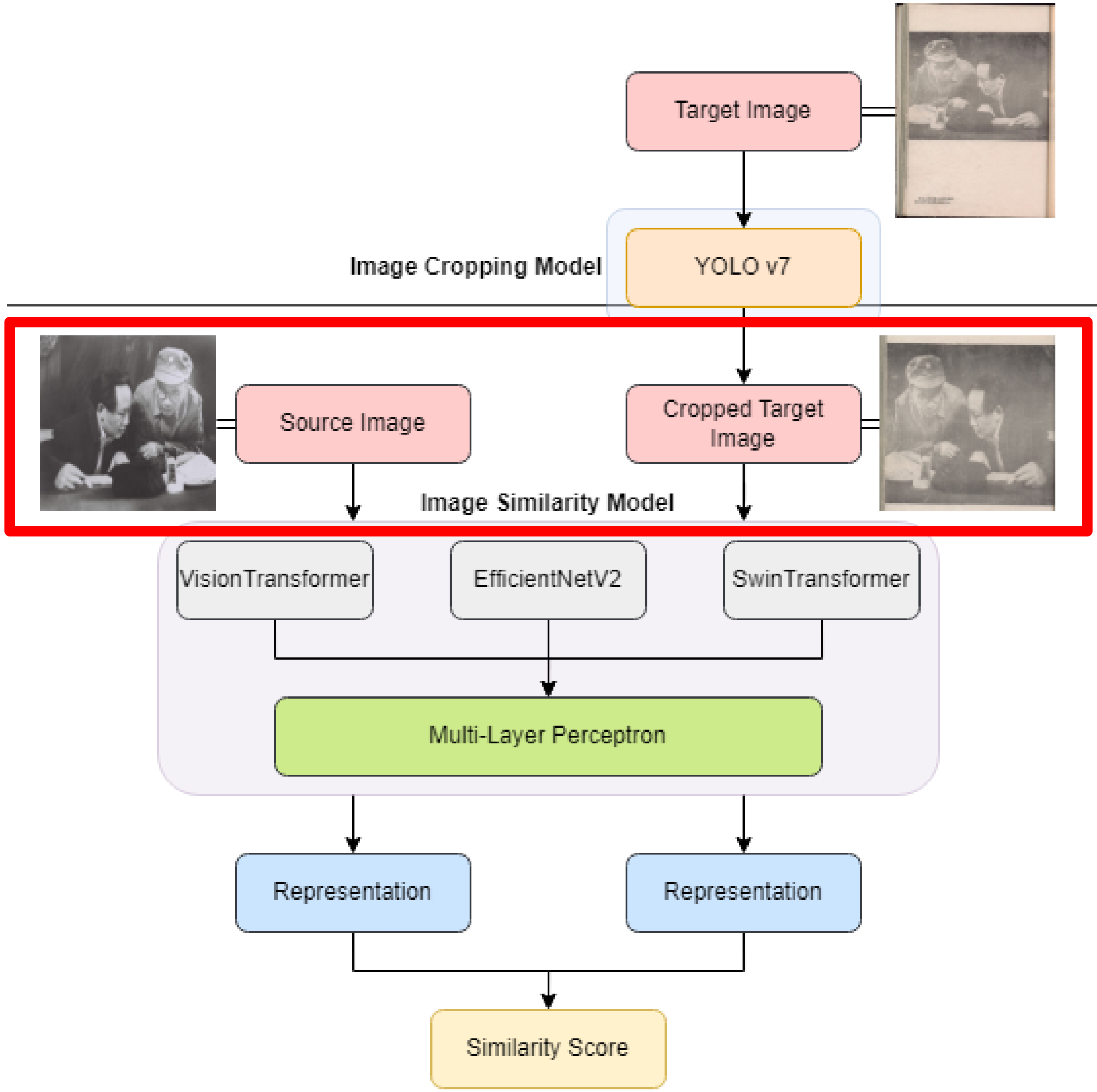
Computer Vision Pipeline for Image Retrieval



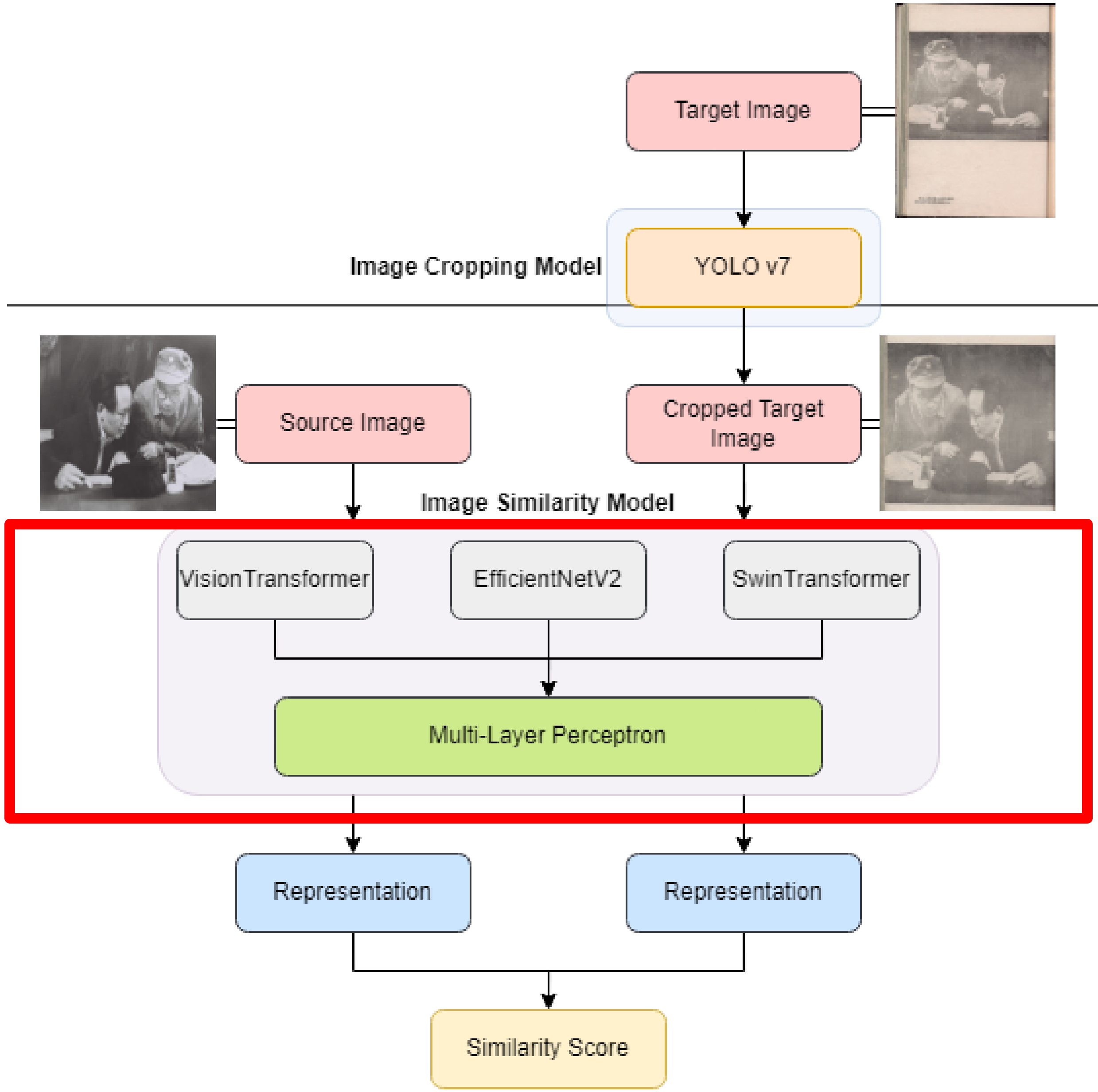
Computer Vision Pipeline for Image Retrieval



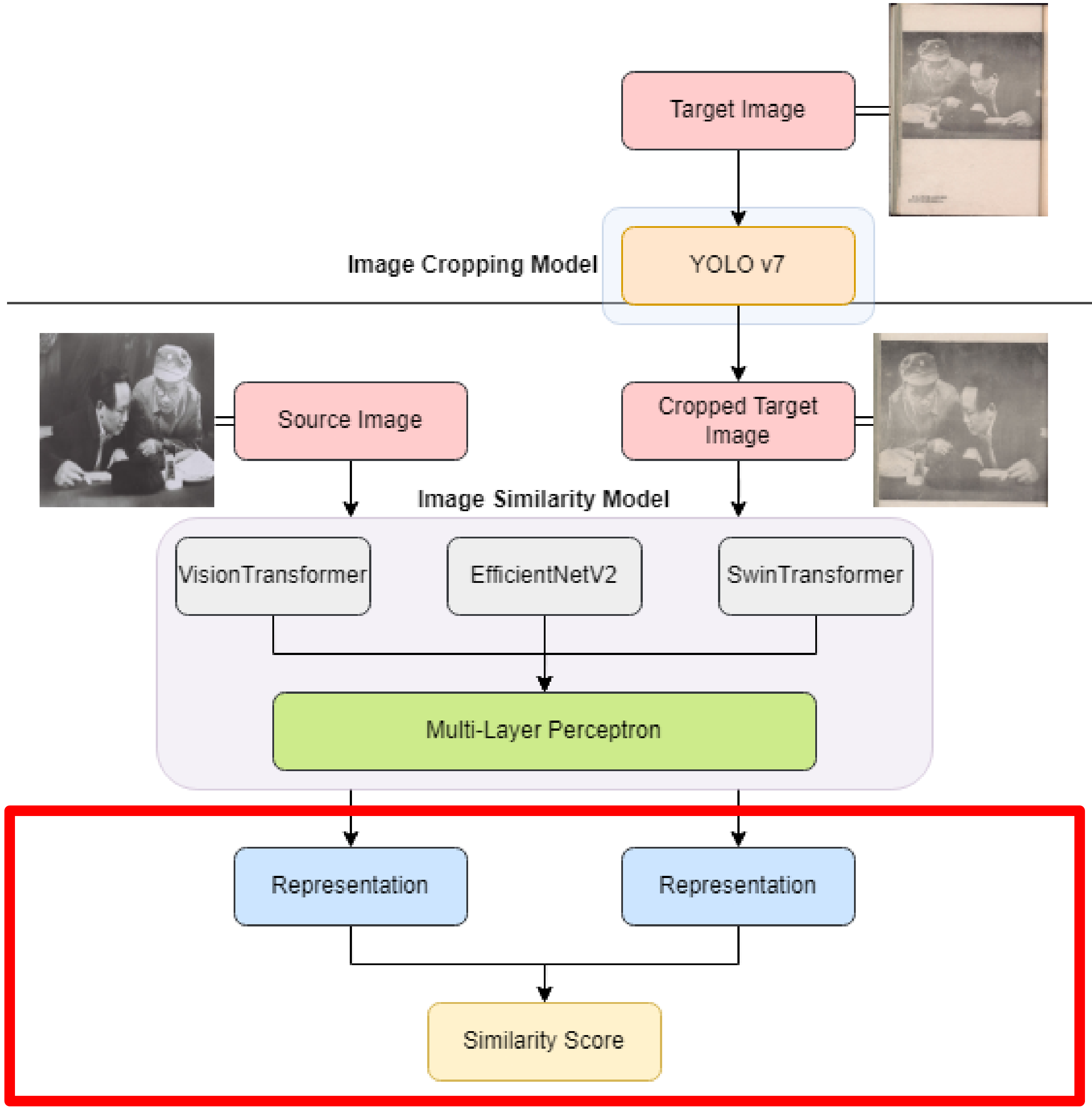
Computer Vision Pipeline for Image Retrieval



Computer Vision Pipeline for Image Retrieval

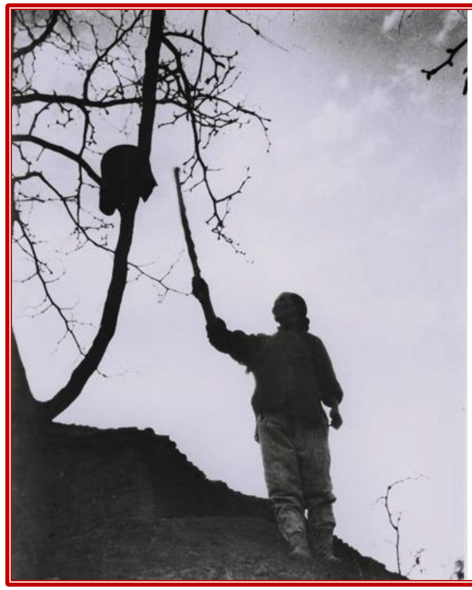


Computer Vision Pipeline for Image Retrieval

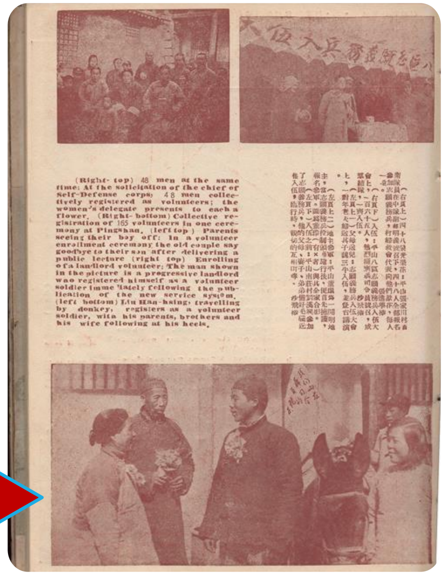


Two Key Ways Computer Vision Compares and Contextualizes Historical Photographs

1. Juxtaposing Original And Printed Images

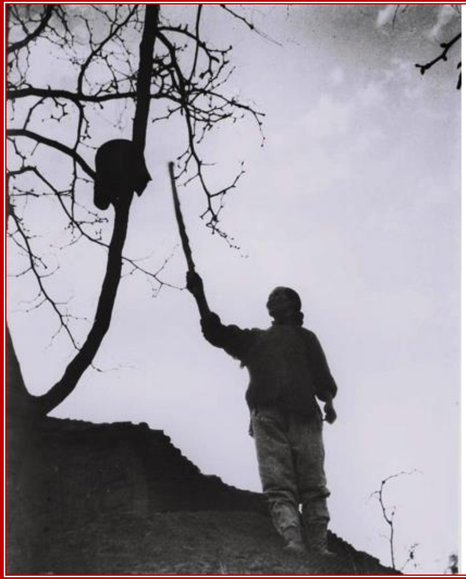


2. Tracking Photographic Circulation

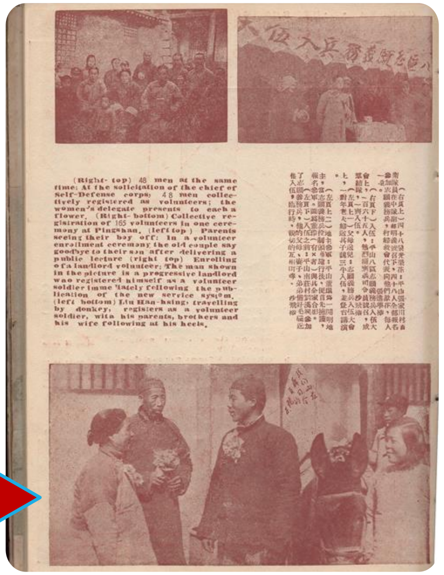


Two Key Ways Computer Vision Compares and Contextualizes Historical Photographs

1. Juxtaposing Original And Printed Images



2. Tracking Photographic Circulation



Case Study 1: Intentional Misinformation in Image Captioning

Case Study 1: Intentional Misinformation in Image Captioning

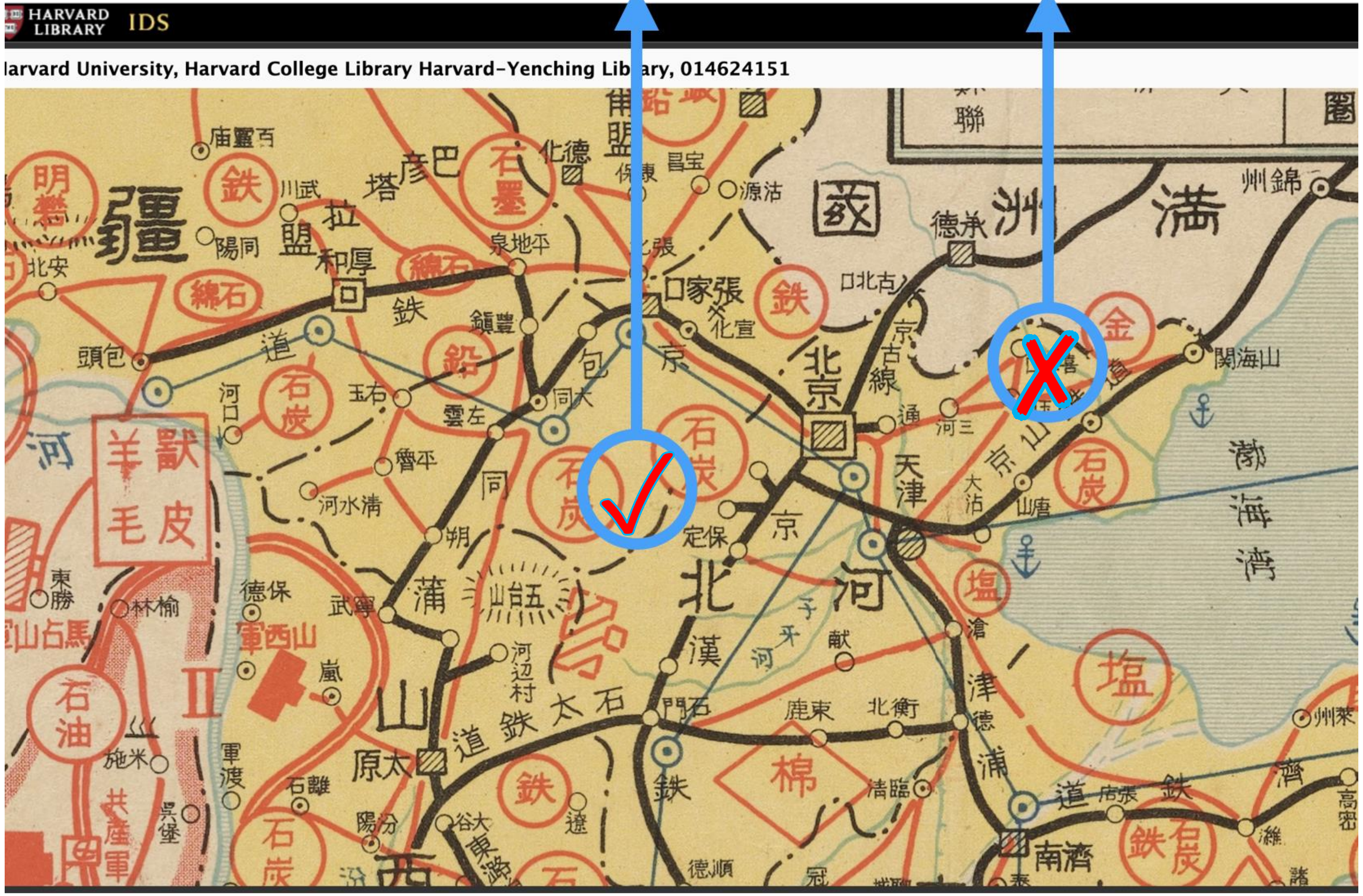
Fact Check

Scholar Si Sushi [Si 2016] has confirmed that this photographs was actually taken by photographer Sha Fei and the real shooting location was in Futuyu, Laiyuan County, Hebei Province, rather than Xifengkou in Qianxi County, Tangshan, Hebei Province.

"The Anti-Japanese War Zone and Resource Transportation Network in China (支那抗日戰區及資源交通網要圖)," Yellow Region is Japanese-Army occupied area; This map was created by Japanese in 1941; Provided by Harvard University Library.

Futuyu in Laiyuan County, Hebei Province

Xifengkou in Qianxi County, Tangshan, Hebei Province



Summary

- The importance of context in analyzing historiography and image editing
- The fluidity and adaptability of photographs as historemes: photographs and accompanying captions can be manipulated or adapted to various contexts within the realm of propaganda.
- Our computer vision pipeline can be used in combination with contextual analysis, a traditional media studies approach, to compare images and map the publication and circulation history of photographs.
- Our paper: Du, Lin, Brandon Le, and Edouardo Honig. “Probing Historical Image Contexts: Enhancing Visual Archive Retrieval through Computer Vision.” *ACM Journal on Computing and Cultural Heritage* 16, no. 4: 84:1-84:17. https://doi.org/10.1145/3631129

Project 1: Goal

- Use supervised learning, leveraging local feature matching, to classify whether historical photographs from the Manchuria Railway archive (Kyoto University, 1930s-1940s) were published in *North China* Magazine or remained unpublished. This project aims to refine an existing image retrieval pipeline by reducing false positives in its candidate matches, using the Aspanformer model for local feature verification.
- The ultimate goal is to contribute to AI-assisted visual historiography by better understanding how editorial choices, aesthetics, and political agendas shaped wartime visual culture.
- We may turn this into a publishable research paper together.

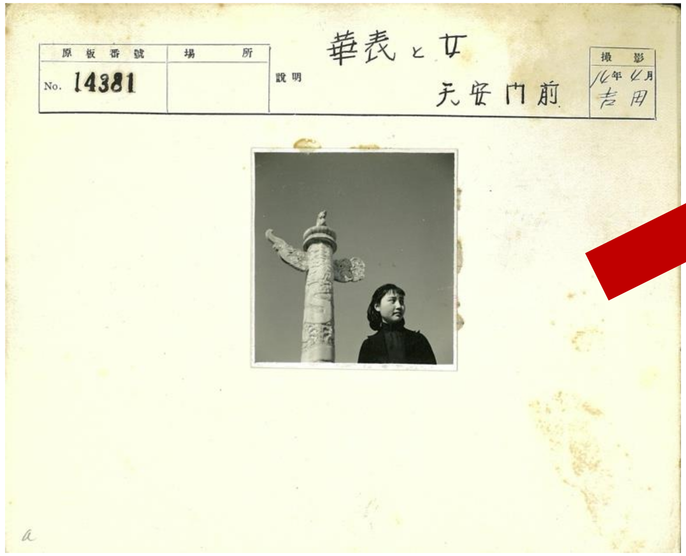
Dataset

- Archive Photos (A): ~40,000 digitized historical photographs. Each a ∈ A.
- Magazine scans (M): 2658 digitized pages from North China Magazine, from which individual images (m) are extracted. Each m ∈ M.
- Preprocessed Matches: For each archive photo a ∈ A, a list of the top-10 most visually similar magazine images m_1, ..., m_{10} ∈ M (identified using a self-supervised learning model) is provided. Some of these matches are correct (true positives), while others are incorrect (false positives).
- Labeled Subset (A_labeled): A small subset of archive photos (A_labeled ⊂ A) will be provided with ground truth labels y ∈ {published, unpublished} based on verifying the top-10 matches (published if ≥1 correct match, unpublished otherwise). We currently have 66 labeled examples and can label more if needed for training/evaluation.

所
說明
華表と女
天安门前
撮影
16年6月
吉田

北支
現地編輯
THE NORTH CHINA
6

原板番號
No. ~~14374~~ 14382
場所
說明
華表と女
天安門前
撮影
16年6月
吉田
P



原板番號
No. 14381
場所
說明
華表と女
天安门前
撮影
16年6月
吉田

原板番號
No. 14379
場所
說明
~~天安門~~唐獅子と女
正陽門
撮影
14年4月
吉田

Data Representation & Feature Extraction:

- Initial Representation and Matching: Each image (archive photo a or magazine image m) is represented by a pre-computed feature vector $\phi(\text{image}) \in R^{1000}$. These 1000-dimensional embeddings were generated using a self-supervised deep learning model (based on contrastive learning, detailed in previous work: https://doi.org/10.1145/3631129), capturing global visual similarity. The provided top-10 candidate matches $\{m_1, ..., m_{10}\}$ for each $a \in A$ were identified by finding the nearest neighbors in this embedding space using cosine similarity: $\text{sim}(\phi(a), \phi(m)) = (\phi(a) \cdot \phi(m)) / (||\phi(a)||\ ||\phi(m)||)$. These initial matches serve as the input for the refinement step.

Data Representation & Feature Extraction:

- Refinement Feature:

- Install Aspanformer (https://aspanformer.github.io/), a detector-free image matching model based on Transformers.

- Role in this Project:
  - Used to find reliable local feature matches (keypoints) between an archive photo (a) and its candidate magazine matches (m_j).
  - The number of these matches (k(a, m_j)) serves as a robust feature (f(a)) to classify if the archive photo was truly published, overcoming limitations of global similarity embeddings.

- Reference: https://aspanformer.github.io/ (Chen et al., ECCV 2022)

Data Representation & Feature Extraction:

- Refinement Feature: After installing Aspanformer, the core task is to use it to find local feature matches (keypoints) between an archive photo a and each of its top-10 candidate magazine images m_j. For each pair (a, m_j), Aspanformer outputs a set of matching keypoint pairs. Let $k(a, m_j)$ be the number of matching keypoints found.
- Classification Feature: For each archive photo a, we define a feature $f(a) = \max_{j=1...10} \{k(a, m_j)\}$. This feature $f(a)$ is a scalar value representing the strongest evidence of a match among the top 10 candidates based on local keypoints. Our dataset for classification can thus be represented as $\{(f(a_i), y_i)\}$ for $a_i \in A_{labeled}$, where y_i is the corresponding ground truth label.

Task

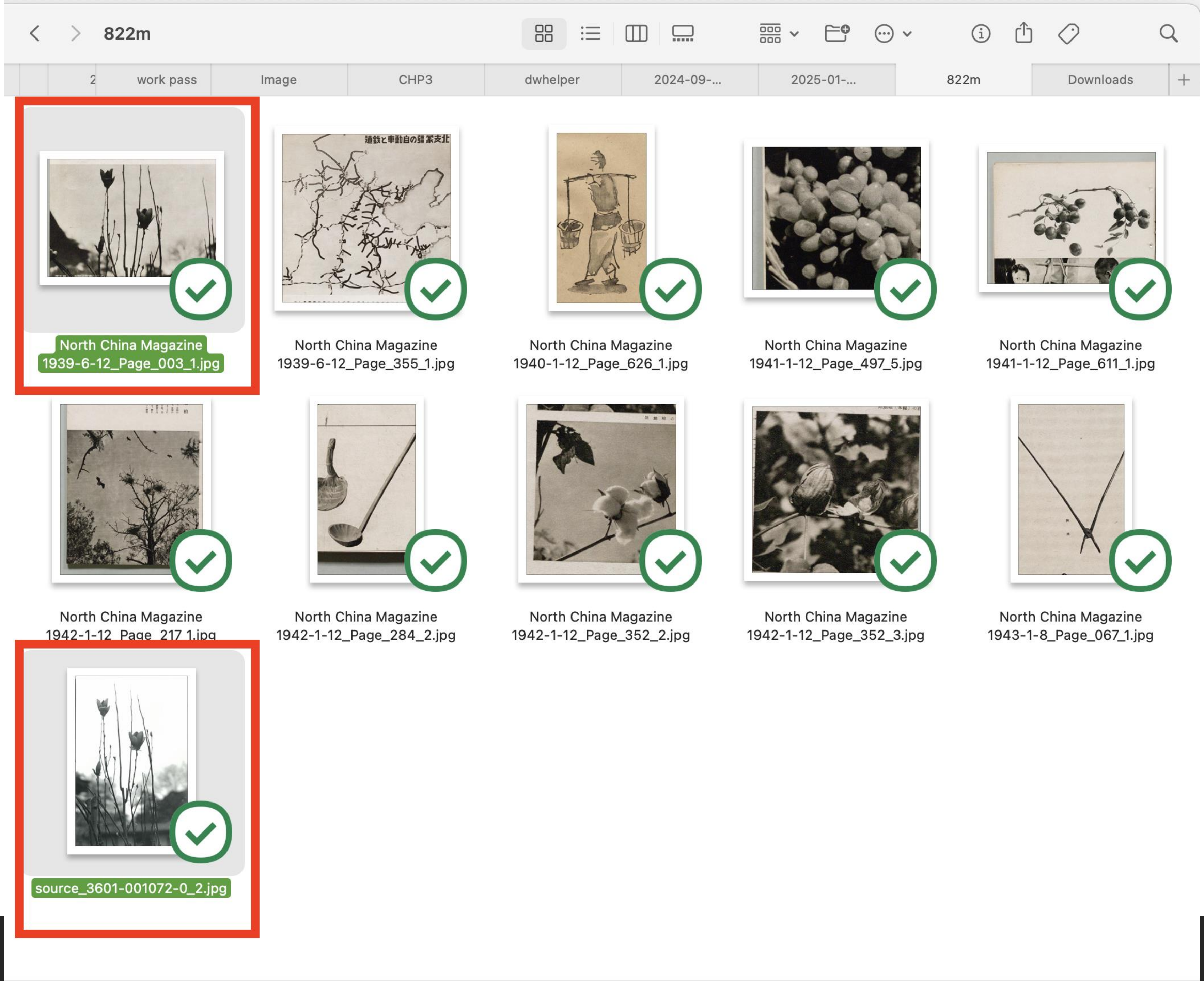
1. Implement Feature Extraction: Set up the Aspanformer environment. Run the model to compute the local keypoint counts $k(a_i, m_j)$ for all $a_i \in A_{labeled}$ and their respective top-10 matches m_j. Calculate the final feature $f(a_i) = \max_j\{k(a_i, m_j)\}$ for each labeled archive photo.
2. Analyze Feature Distribution: Examine the distribution of the feature $f(a)$ separately for the 'published' ($y=1$) and 'unpublished' ($y=0$) classes within the labeled set $A_{labeled}$. Visualize these distributions (e.g., using histograms or kernel density estimates) to assess the feature's discriminative power. Does $f(a)$ tend to be higher for published images?

Task

3. Develop Classification Model: Given the scalar nature of f(a), develop a supervised classification model.

- Thresholding Approach: Determine an optimal threshold τ such that an image a is classified as "published" if $f(a) > \tau$ and "unpublished" otherwise. Optimize τ using A_labeled (e.g., maximizing accuracy/F1-score via validation or ROC analysis).
- Alternative Classifiers: Optionally, train other simple classifiers (e.g., Logistic Regression, SVM) on the feature f(a).

4. Evaluate Performance: Evaluate the chosen classifier's performance using appropriate metrics (e.g., accuracy, precision, recall, F1-score) on a held-out test set or using cross-validation within A_labeled.


822m
2
work pass
Image
CHP3
dwhelper
2024-09-...
2025-01-...
822m
Downloads
North China Magazine 1939-6-12_Page_003_1.jpg
North China Magazine 1939-6-12_Page_355_1.jpg
North China Magazine 1940-1-12_Page_626_1.jpg
North China Magazine 1941-1-12_Page_497_5.jpg
North China Magazine 1941-1-12_Page_611_1.jpg
North China Magazine 1942-1-12_Page_217_1.jpg
North China Magazine 1942-1-12_Page_284_2.jpg
North China Magazine 1942-1-12_Page_352_2.jpg
North China Magazine 1942-1-12_Page_352_3.jpg
North China Magazine 1943-1-8_Page_067_1.jpg
source_3601-001072-0_2.jpg

Reference

1. Chen, Hongkai, Zixin Luo, Lei Zhou, Yurun Tian, Mingmin Zhen, Tian Fang, David Mckinnon, Yanghai Tsin, and Long Quan. 2022. "ASpanFormer: Detector-Free Image Matching with Adaptive Span Transformer." arXiv. https://doi.org/10.48550/arXiv.2208.14201.

2. Chen, Ting, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. "A Simple Framework for Contrastive Learning of Visual Representations." In *International Conference on Machine Learning*, 1597–1607. PMLR. http://proceedings.mlr.press/v119/chen20j.html.

3. Du, Lin, Brandon Le, and Edouardo Honig. 2024. "Probing Historical Image Contexts: Enhancing Visual Archive Retrieval through Computer Vision." *J. Comput. Cult. Herit.* 16 (4): 84:1-84:17. https://doi.org/10.1145/3631129.

4. Dubey, Shiv Ram. 2021. "A Decade Survey of Content Based Image Retrieval Using Deep Learning." *IEEE Transactions on Circuits and Systems for Video Technology* 32 (5): 2687–2704.

5. Ma, Jinping. 2020. "Visualizing North China Under Japanese Occupation: Digitized Photos of the North China Railway Archive." *The Digital Orientalist* (blog). November 27, 2020. https://digitalorientalist.com/2020/11/27/visualizing-north-china-under-japanese-occupation-digitized-photos-of-the-north-china-railway-archive/.

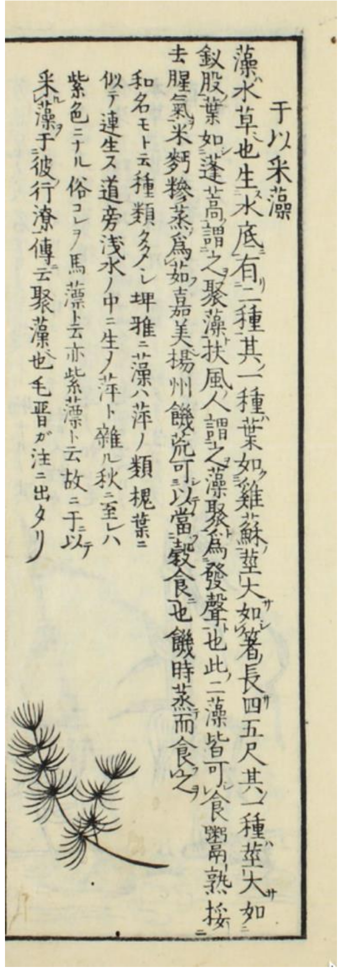
6. 華北交通アーカイブ作成委員会／ROIS-DS人文学オープンデータ共同利用センター. n.d. "華北交通アーカイブ：よみがえる膨大な白黒写真 - 国策鉄道会社が遺した戦時期広報用写真の研究データベース." Accessed March 30, 2025. https://codh.rois.ac.jp/north-china-railway/.

Project 2: Quantifying Visual Similarity of Illustrations Across Cultures using Image Embeddings

藻

藻

左傳蘋蘩薀藻之菜杜預註薀聚也毛傳藻聚藻也
陸璣疏藻水草也生水底有二種其一種葉如雞蘇莖大如箸長四五尺其一種莖大如釵股葉如蓬蒿謂之聚藻二藻皆可食米麪糝蒸爲茹荆揚人饑荒以當穀食爾雅翼藻横陳於水如自澡濯若流水之中隨波衍漾莖葉條暢尤爲可喜故采藻於行潦也愚按葉生於莖一二寸兩兩對生即郭璞云馬藻陸璣所謂葉如雞蘇者是也葉細節節相生即傳云聚藻
于以采藻是也



于以采藻

藻水草也生水底有二種其一種葉如雞蘇莖大如箸長四五尺其一種莖大如釵股葉如蓬蒿謂之聚藻扶風人謂之藻聚爲發聲也此二藻皆可食蒸鬻熟挼去腥氣米麪糝蒸爲茹嘉美揚州饑荒可以當穀食也饑時蒸而食之

和名モト云種類多シ埤雅ニ藻ハ萍ノ類槐葉ニ似テ連生ス道旁浅水ノ中ニ生ノ萍ト雜ル秋ニ至レハ紫色ニナル俗コレヲ馬藻ト云亦紫藻ト云故ニ于以采藻于彼行潦傳云聚藻也毛晉ガ注ニ出タリ

Goal

- Employ unsupervised learning techniques, specifically image embeddings and similarity metrics, to quantitatively analyze and compare the visual representation styles of the same biological objects (plants and animals referenced in the Book of Songs / Shijing) as depicted in historical Chinese (Qing Dynasty) and Japanese (Tokugawa-to-Meiji era) illustrated commentaries. The aim is to explore patterns of visual continuity, divergence, or reinterpretation across these cultural contexts.
- We may turn this into a publishable research paper together too.

Datasets

- Chinese Illustrations (C): A set of n_C images, manually cropped from scanned pages of the Mao Shi Ming Wu Tu Shuo (毛诗名物图说, Qing Dynasty, Xu Ding). Each image $c \in C$ depicts a specific object mentioned in the Shijing.
- Japanese Illustrations (J): A set of n_J images, manually cropped from scanned pages of Mōshi Hinbutsu Zukō (毛诗品物图考, Okamoto Ryūho) and Rikushi Sōmoku Chōjū Chūgyo So Zukai (陸氏草木鳥獣虫魚疏図解, Tokugawa-to-Meiji era). Each image $j \in J$ depicts a specific object mentioned in the Shijing.
- Paired Data: The core dataset consists of N pairs $\{(c_i, j_i) \mid i = 1...N\}$, where $c_i \in C$ and $j_i \in J$ are illustrations of the same biological object (identified by its name in the Shijing).

Data Representation & Feature Engineering:

- Image Representation: Each illustration image (c or j) is treated as a data point. Given the nature of illustrations (potentially simpler line drawings compared to photographs), two main feature representations can be considered:
  - Deep Embeddings: Use a pre-trained Convolutional Neural Network (CNN), such as ResNet-18 (as mentioned in the slides), to extract a dense feature vector (embedding) $\phi(\text{image}) \in R^D$ for each illustration. D represents the embedding dimension (e.g., D=512 for the penultimate layer of ResNet-18). This captures higher-level visual semantics.
  - Traditional Local Features (Alternative): For simpler or line-art heavy illustrations, traditional computer vision features like SIFT (Scale-Invariant Feature Transform) could be explored. This would represent each image as a set of local feature descriptors, rather than a single vector. Comparing sets of features requires different similarity measures (e.g., matching SIFT keypoints and counting matches, similar to Project 1's use of Aspanformer, or using techniques like Bag-of-Visual-Words).
- Data Structure: If using CNN embeddings, we can conceptually think of two matrices: $X_C \in R^{(n_C \times D)}$ and $X_J \in R^{(n_J \times D)}$, although the primary analysis focuses on the N specified pairs.

Task & Methodology (Unsupervised Analysis)

- Feature Extraction: Compute the chosen feature representation (e.g., CNN embeddings $\phi(c_i)$ and $\phi(j_i)$) for all illustrations in the N pairs.
- Similarity Calculation: For each pair (c_i, j_i) representing the same object, calculate a visual similarity score. If using CNN embeddings $\phi(\text{image}) \in R^D$, the cosine similarity is a standard choice: $s_i = \text{sim}(\phi(c_i), \phi(j_i)) = (\phi(c_i) \cdot \phi(j_i)) / (||\phi(c_i)|| \, ||\phi(j_i)||)$ This results in a set of N scalar similarity scores $\{s_1, ..., s_N\}$.

Task & Methodology (Unsupervised Analysis)

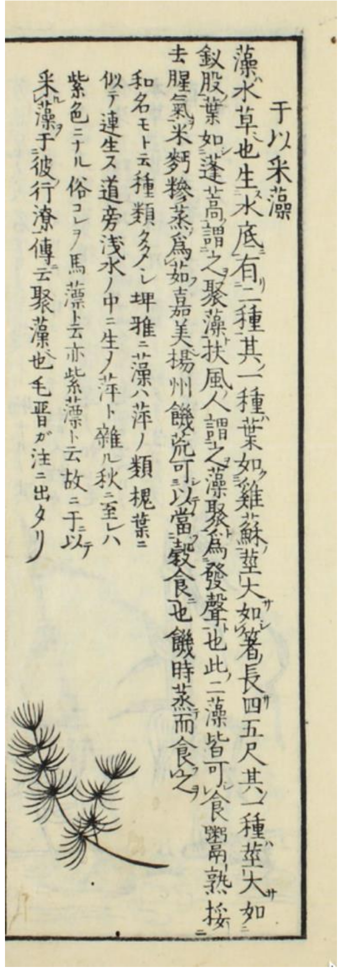
- Analysis of Similarity Scores:
  - Rank the object pairs (c_i, j_i) based on their similarity scores s_i from highest to lowest. Examine the illustrations at the extremes (most similar, least similar) to understand what drives the visual similarity/difference.
  - Analyze the overall distribution of the similarity scores $\{s_i\}$. Does the distribution suggest a general tendency towards high similarity (implying copying or strong influence) or low similarity (implying divergence or independent artistic traditions)?

藻

藻

左傳蘋蘩薀藻之菜杜預註薀聚也毛傳藻聚藻也
陸璣疏藻水草也生水底有二種其一種葉如雞蘇
莖大如箸長四五尺其一種莖大如釵股葉如蓬蒿
謂之聚藻二藻皆可食米麪糝蒸爲茹荆揚人饑荒
以當穀食爾雅翼藻横陳於水如自澡濯若流水之中
隨波衍漾莖葉條暢尤爲可喜故采藻於行潦也愚
按葉生於莖一二寸兩兩對生即郭璞云馬藻陸璣
所謂葉如雞蘇者是也葉細節節相生即傳云聚藻

于以采藻是也



于以采藻

藻水草也生水底有二種其一種葉如雞蘇莖大如箸長四五尺其一種莖大如釵股葉如蓬蒿謂之聚藻扶風人謂之藻聚爲發聲也此二藻皆可食蒸熟挼去腥氣米麪糝蒸爲茹嘉美揚州饑荒可以當穀食也饑時蒸而食之

和名モト云種類多シ埤雅ニ藻ハ萍ノ類槐葉ニ似テ連生ス道旁浅水ノ中ニ生ノ萍ト雜ル秋ニ至レハ紫色ニナル俗コレヲ馬藻ト云赤紫藻ト云故ニ于以采藻于彼行潦傳ニ云聚藻也毛晋ガ注ニ出タリ

于以采藻 モ

傳藻聚藻也集傳生水底莖如釵股葉如蓬蒿

白茅包之 チカヤ

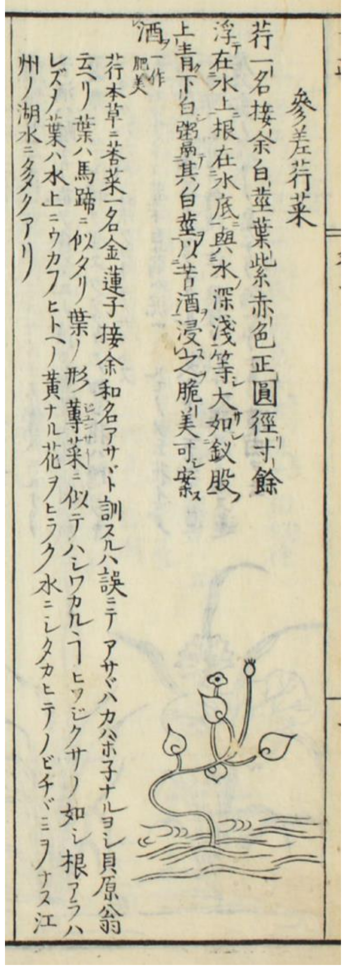
荇菜

荇

周南 關雎

爾雅釋草莕接余其葉苻郭璞註叢生水中葉圓在莖端長短隨水深淺江東食之亦呼莕陸璣草木蟲魚疏接余白莖葉紫赤色正圓徑寸餘浮在水上根在水底與水深淺等大如釵股上青下白鬻其白莖以苦酒浸之肥美可案酒蘇恭唐本草鳧葵即莕菜也生水中羅願爾雅翼荇菜今陂澤多有葉卷漸開雖圓而稍羨不若蓴之極圓也隨水平浮花則出水黄色六出今宛陵陂湖中瀰覆頃畝日出照之如金俗名金蓮子嚴粲詩緝參差訓不齊今池州人稱荇為莕公鬚蓋細荇亂生有若鬚然愚按荇似蓴菜而非蓴蓋蓴菜比荇而葉圓詩薄采其茆即蓴也陸氏德明云荇亦作莕接余也則荇與莕同俗呼為荇絲菜許氏說文謂之䓪楚辭謂之屏風云紫莖屏風文綠波皆指此也

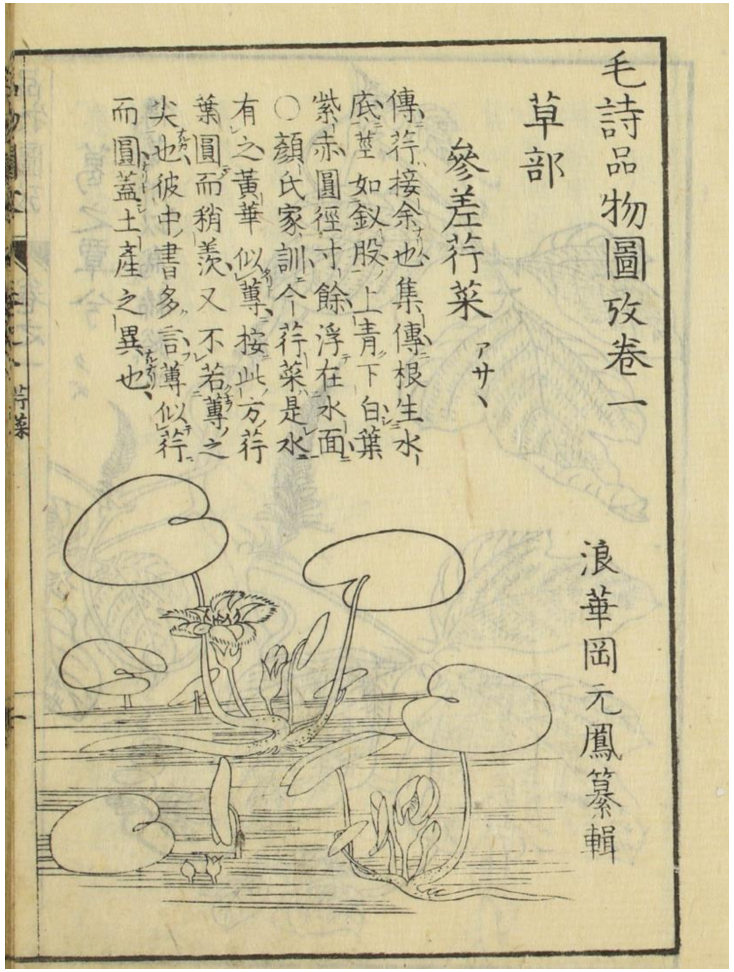
草上



參差荇菜

荇一名接余白莖葉紫赤色正圓徑寸餘浮在水上根在水底與水深淺等大如釵股上青下白鬻其白莖以苦酒浸之脆美可案酒（一作肥美）

荇本草ニ莕菜一名金蓮子接余和名アサザト訓スルハ誤ニテアサザハカハホネナルヨシ貝原翁云ヘリ葉ハ馬蹄ニ似タリ葉ノ形蓴菜ニ似テハシワカルヽコトヒツジクサノ如シ根アラハレズ葉ハ水上ニウカブヒトヘノ黄ナル花ヲヒラク水ニシタガヒテノビチヾミヲナス江州ノ湖水ニ多クアリ



毛詩品物圖攷巻一

草部

參差荇菜 アサヽ

傳荇接余也集傳根生水底莖如釵股上青下白葉紫赤圓徑寸餘浮在水面

○顏氏家訓今荇菜是水有之黄華似蓴按此方荇葉圓而稍羨又不若蓴之尖也彼中書多言蓴似荇而圓蓋土産之異也

浪華岡元鳳纂輯

荇菜

Reference

1. Du, Lin, Brandon Le, and Edouardo Honig. 2024. "Probing Historical Image Contexts: Enhancing Visual Archive Retrieval through Computer Vision." *J. Comput. Cult. Herit.* 16 (4): 84:1-84:17. https://doi.org/10.1145/3631129.

2. Karayev, Sergey, Matthew Trentacoste, Helen Han, Aseem Agarwala, Trevor Darrell, Aaron Hertzmann, and Holger Winnemoeller. 2014. "Recognizing Image Style." In *Proceedings of the British Machine Vision Conference 2014*, 122.1-122.11. https://doi.org/10.5244/C.28.122.

3. Lang, Sabine, and Björn Ommer. 2018a. "Attesting Similarity: Supporting the Organization and Study of Art Image Collections with Computer Vision." *Digital Scholarship in the Humanities* 33 (4): 845–56. https://doi.org/10.1093/llc/fqy006.

4. ———. 2018b. "Reconstructing Histories: Analyzing Exhibition Photographs with Computational Methods." *Arts* 7 (4): 64. https://doi.org/10.3390/arts7040064.

5. Resig, John. 2014. "Using Computer Vision to Increase the Research Potential of Photo Archives." *Journal of Digital Humanities* 3 (2): 3–2.

6. Rublee, E., V. Rabaud, K. Konolige, and G. Bradski. 2011. "ORB: An Efficient Alternative to SIFT or SURF. In: IEEE International Conference on Computer Vision."

7. "Ukiyo-e.Org." n.d. Accessed March 30, 2025. https://ukiyo-e.org/about.

If you’re interested in working with either project, please contact me at dulin525@gmail.com/WeChat: dulinlindu. I can share the datasets with you and provide Google Colab Pro Account.