Digital Historical Forensics: A Computational Approach to Wartime Media Cultures

LIN DU 杜琳 (NUS)

QTM SPEAKER SERIES, EMORY UNIVERSITY

FEBRUARY 24, 2025

Digital Historical Forensics

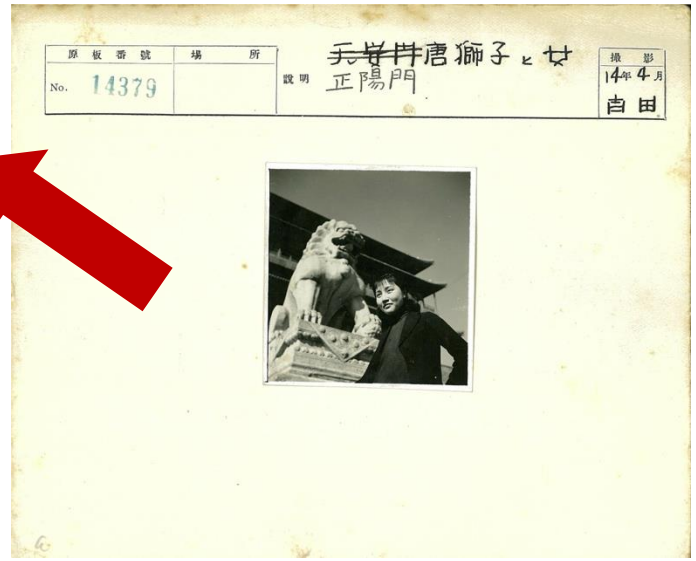
- This method blends AI technologies—including machine learning and computer vision—with traditional humanities methods, such as close reading and contextual analysis.
- It bridges empirical precision with interpretive inquiry.


華表と女
天安門前
吉田


支北
現地綜合 THE NORTH CHINA


原板番號
No. 14382
場所
華表と女
天安門前
撮影
吉田
P


原板番號
No. 14381
場所
華表と女
天安門前
撮影
吉田


原板番號
No. 14379
場所
唐獅子と女
正陽門
撮影
吉田

Digital Historical Forensics

- **This scalable, data-assisted approach integrates AI with traditional humanities methods, such as close reading and contextual analysis, to enhance media studies.**
- **Uncovers meaningful evidence in visual materials, transforming photographs into tools for accessing their historical indexical reality.**
- **Sees beyond static photographs and understand the dynamic historical forces at play.**

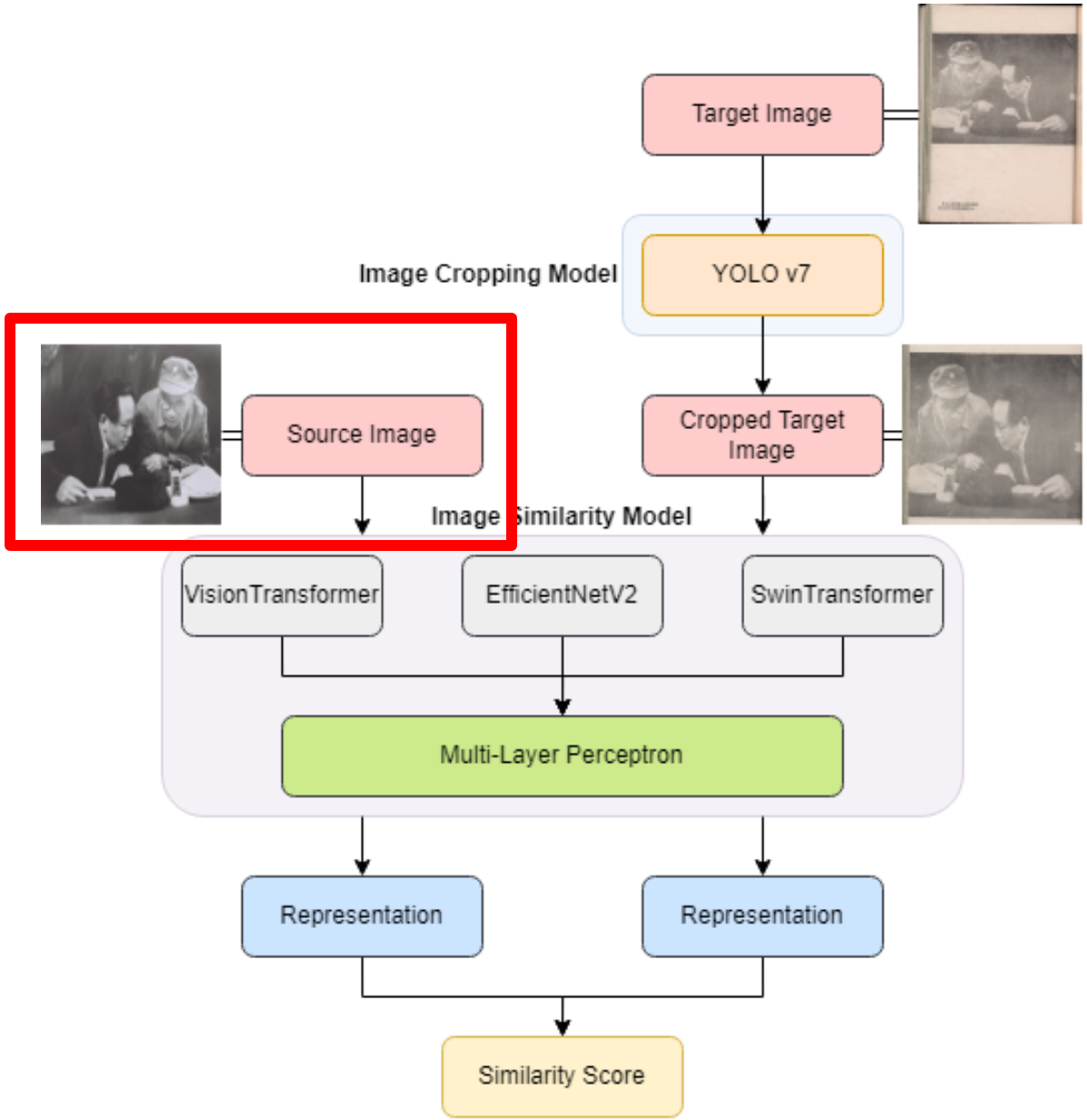
Challenges and Opportunities in Analyzing Print Media Images

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

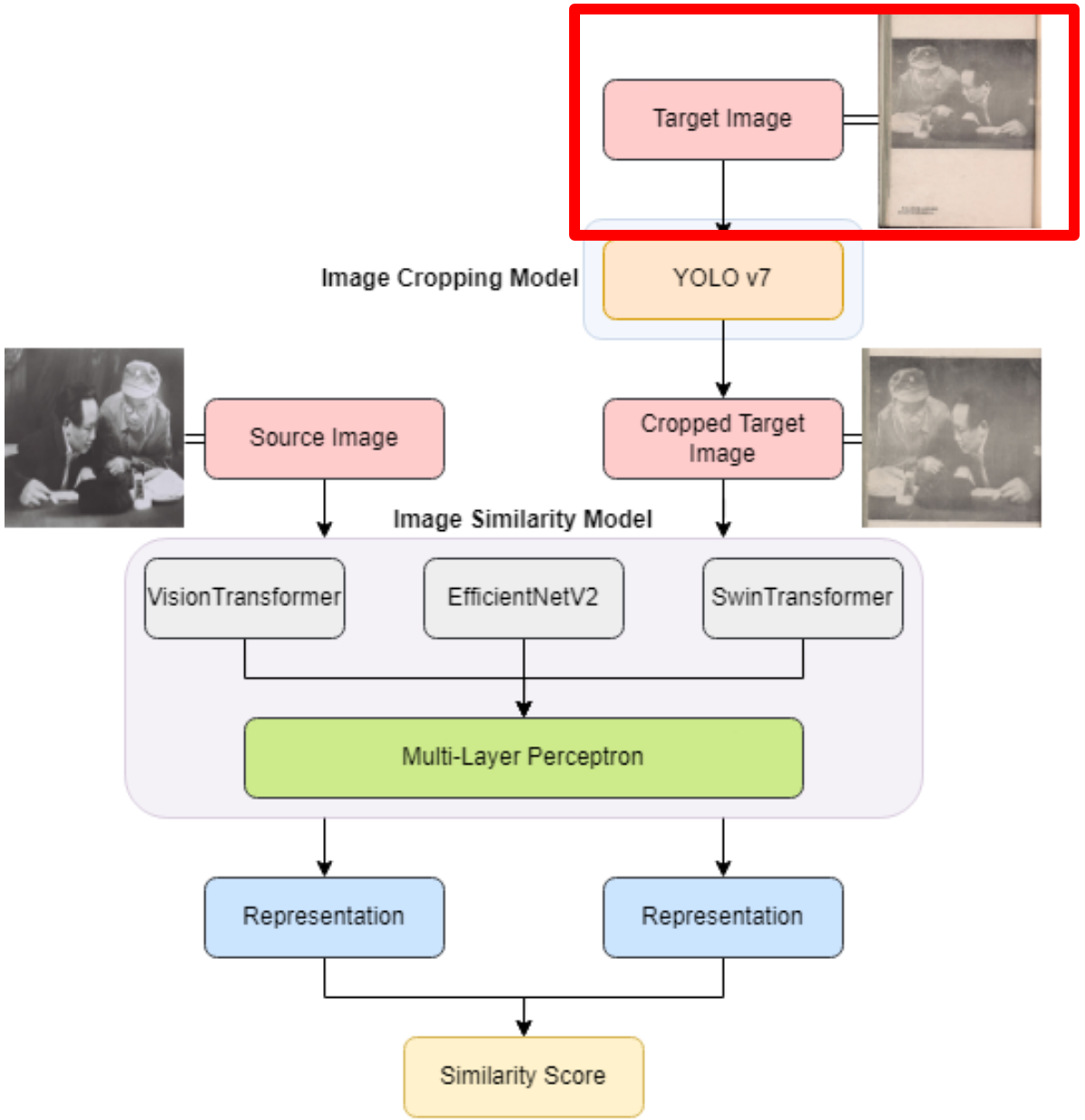
Challenges to humanities scholars tracing image circulation

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

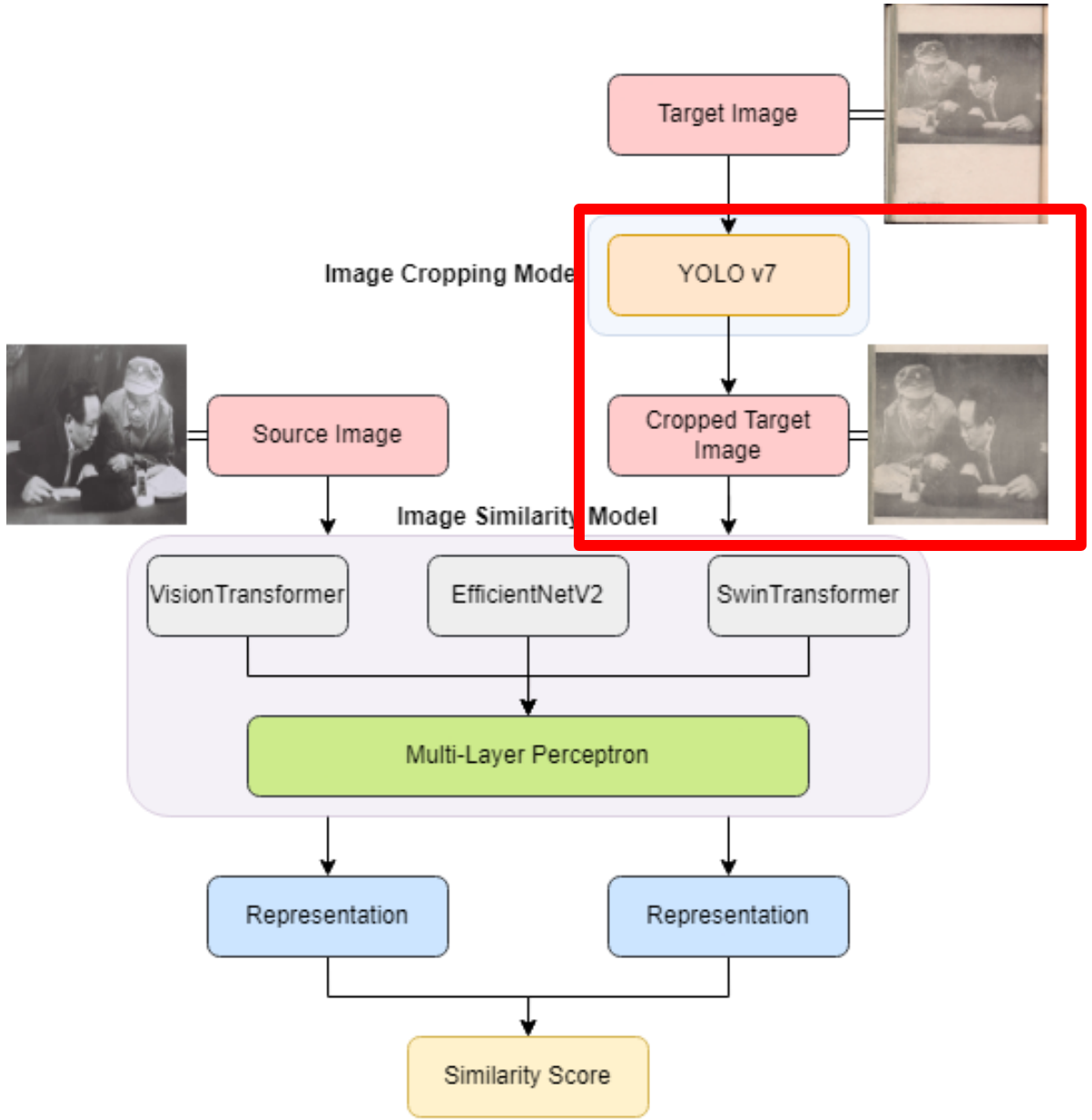
Case Study: Jinchaji Pictorial series, a significant WWII-era photographic publication of the Chinese Communist Party (1942-1948)



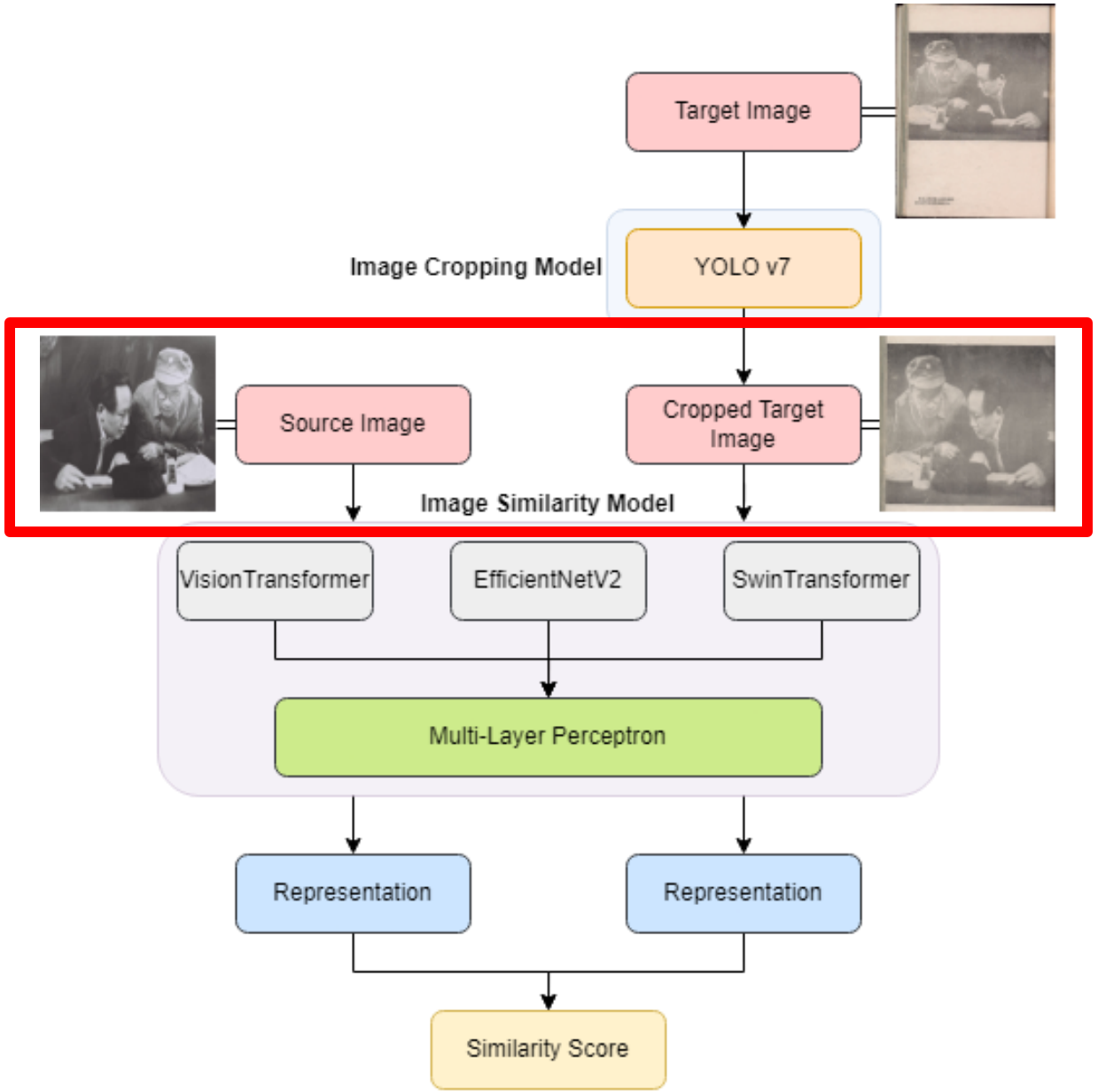
Computer Vision Pipeline for Image Retrieval



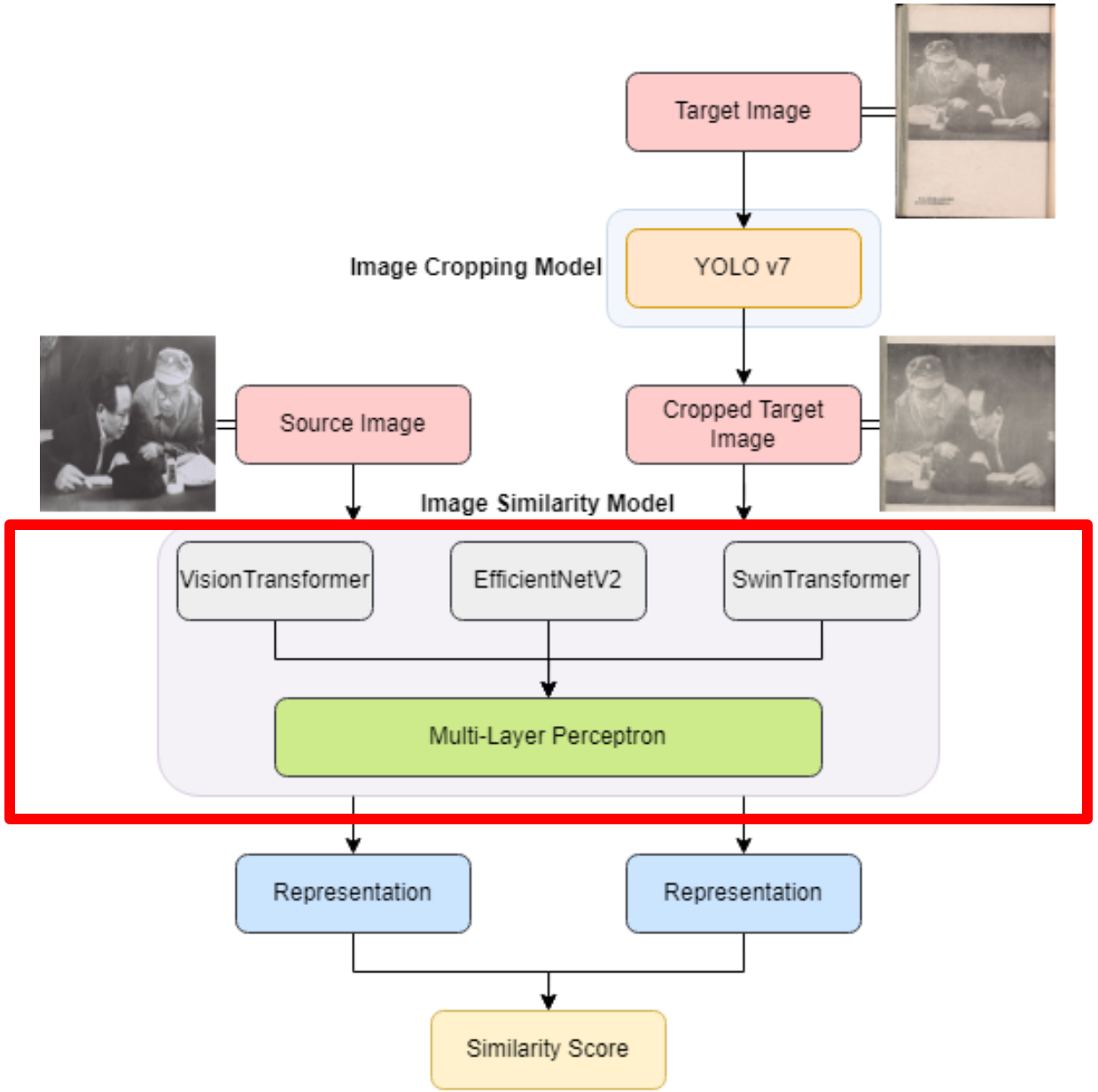
Computer Vision Pipeline for Image Retrieval



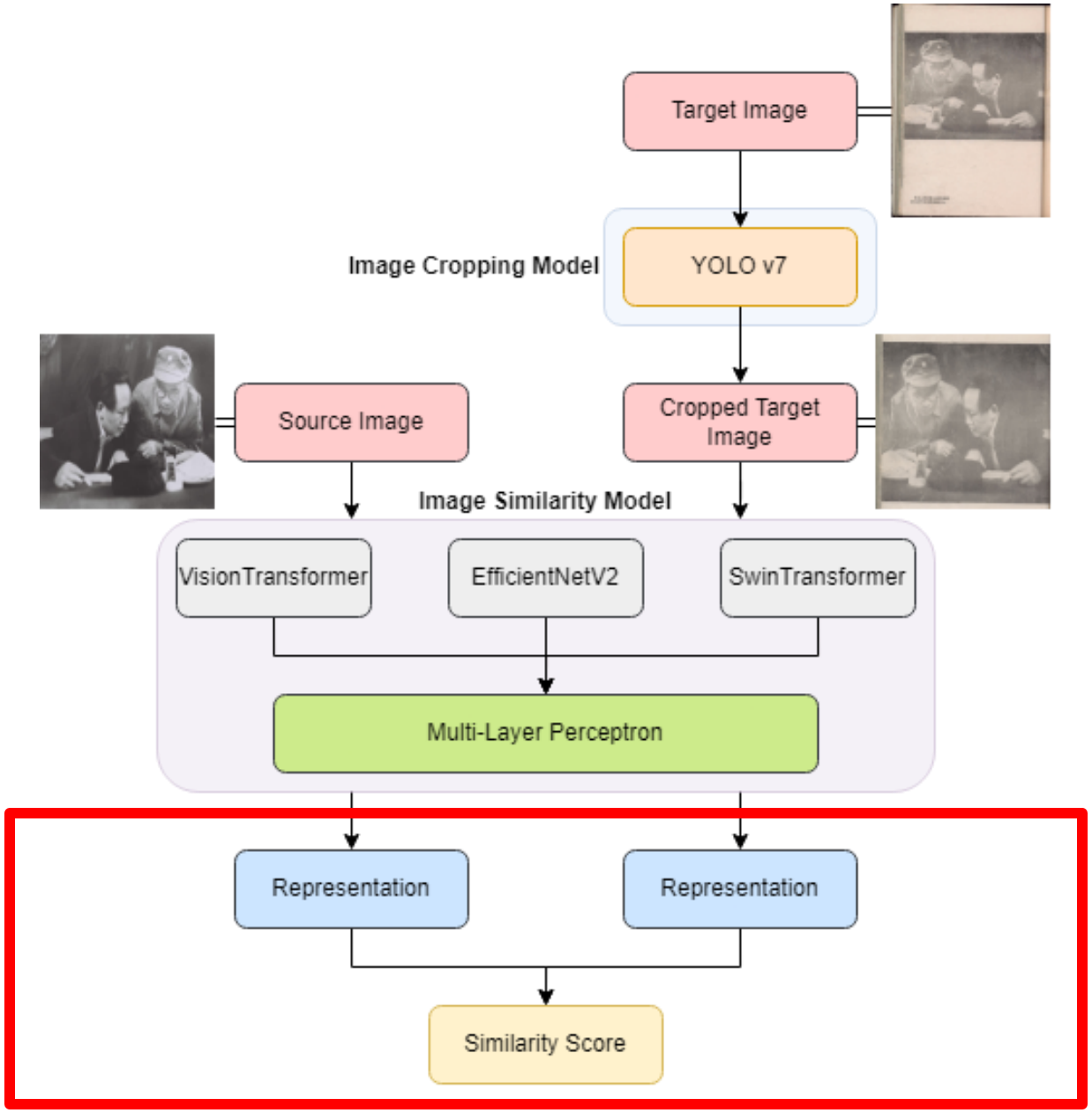
Computer Vision Pipeline for Image Retrieval



Computer Vision Pipeline for Image Retrieval



Computer Vision Pipeline for Image Retrieval

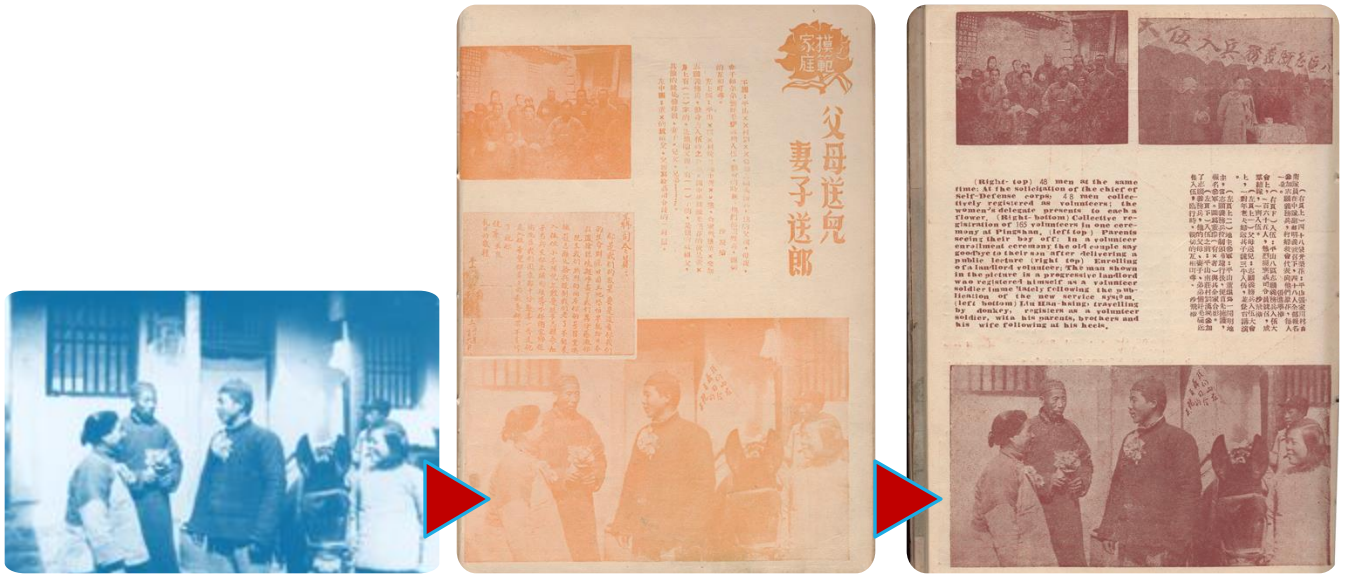


Computer Vision Pipeline for Image Retrieval

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

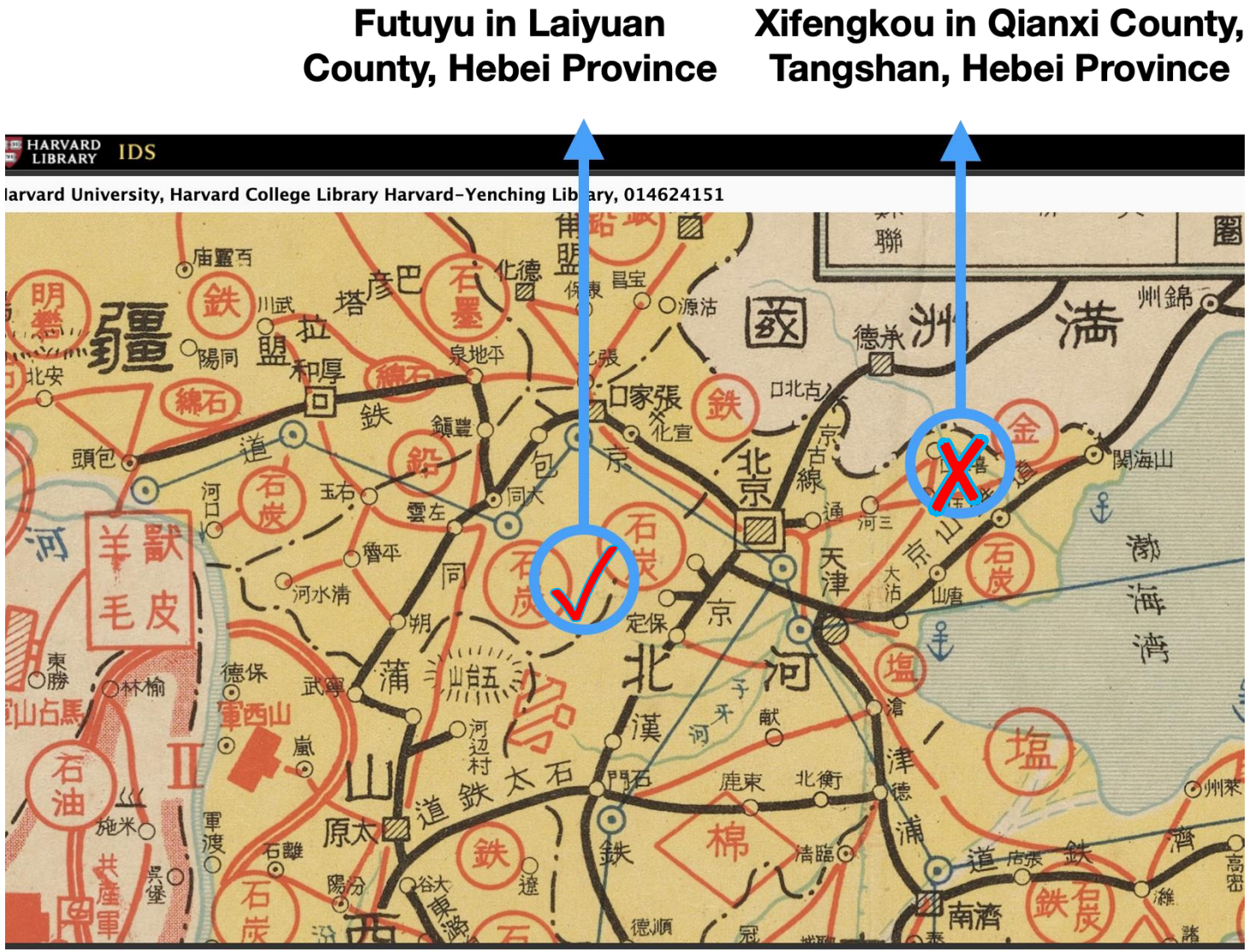
Case Study 1: Intentional Misinformation in Image Captioning

Case Study 1: Intentional Misinformation in Image Captioning

Fact Check

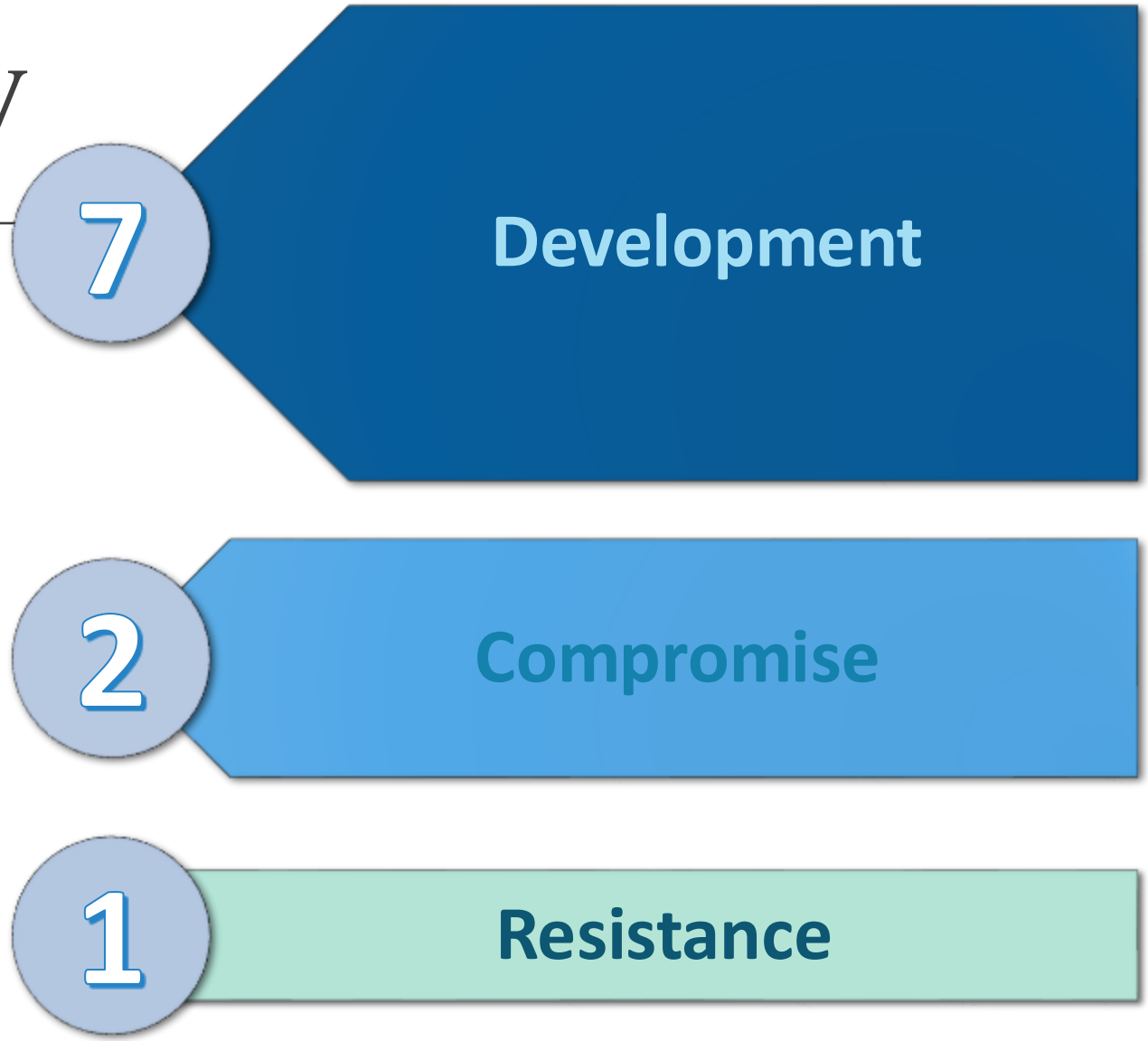
Scholar Si Sushi [Si 2016] has confirmed that this photographs was actually taken by photographer Sha Fei and the real shooting location was in Futuyu, Laiyuan County, Hebei Province, rather than Xifengkou in Qianxi County, Tangshan, Hebei Province.

"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.



Mao’s 721 Policy

A criticism against Mao for not fighting the Japanese army, a belief endorsed by the government of the Republic of China and the Kuomintang, as well as other groups opposing CCP.

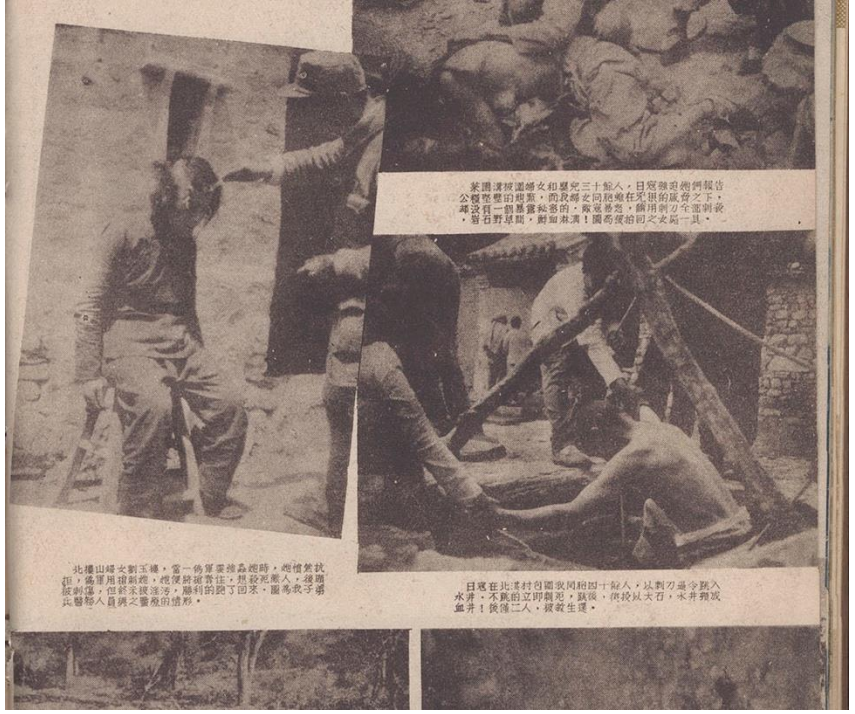


Mao’s 721 Policy

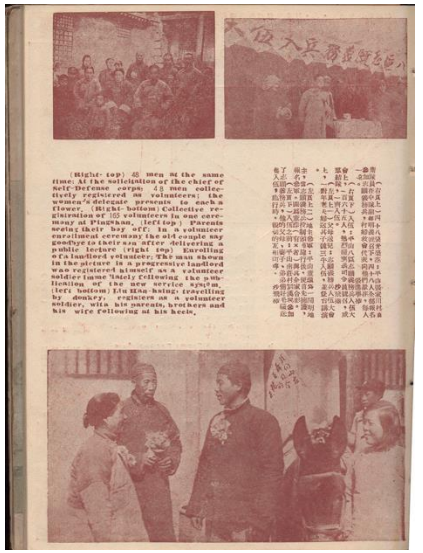
7： Development

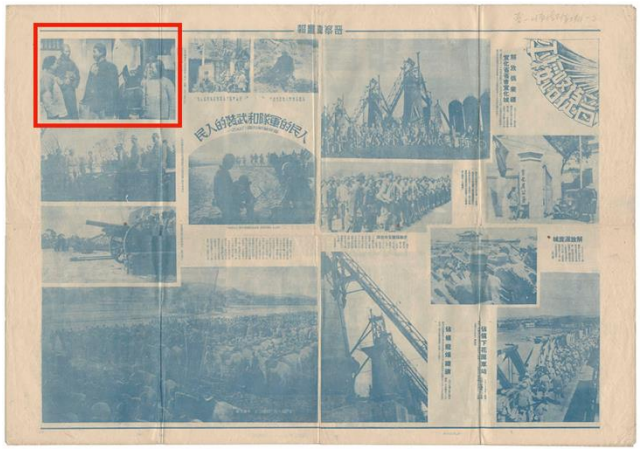
2： Compromise

1： Resistance



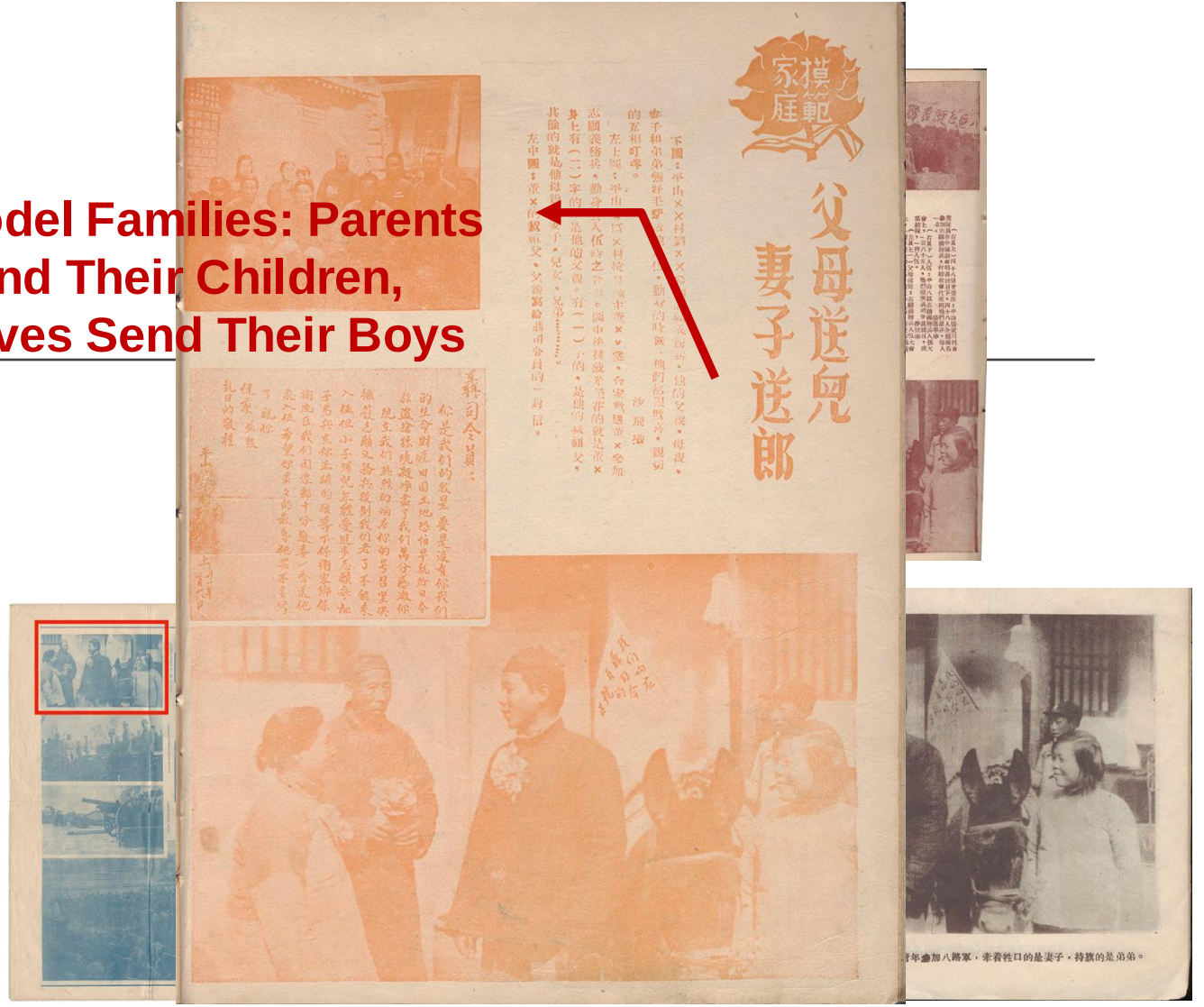
Case Study 2: Resource Scarcities and Image Manipulation





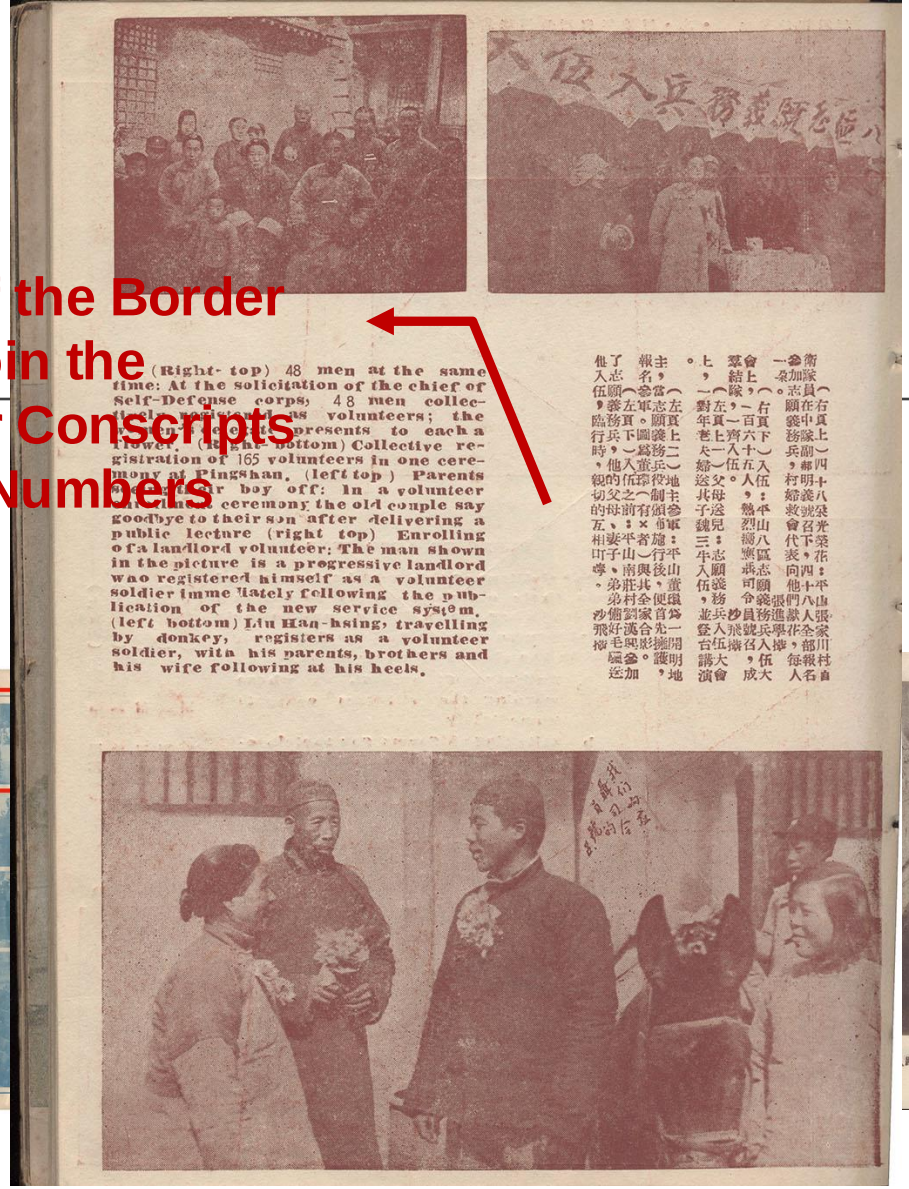
Case Study 3: Editorial Changes in Image Contextualization

Model Families: Parents Send Their Children, Wives Send Their Boys



Case Study 3: Editorial Changes in Image Contextualization

People of the Border Areas: Join the Volunteer Conscripts in Large Numbers



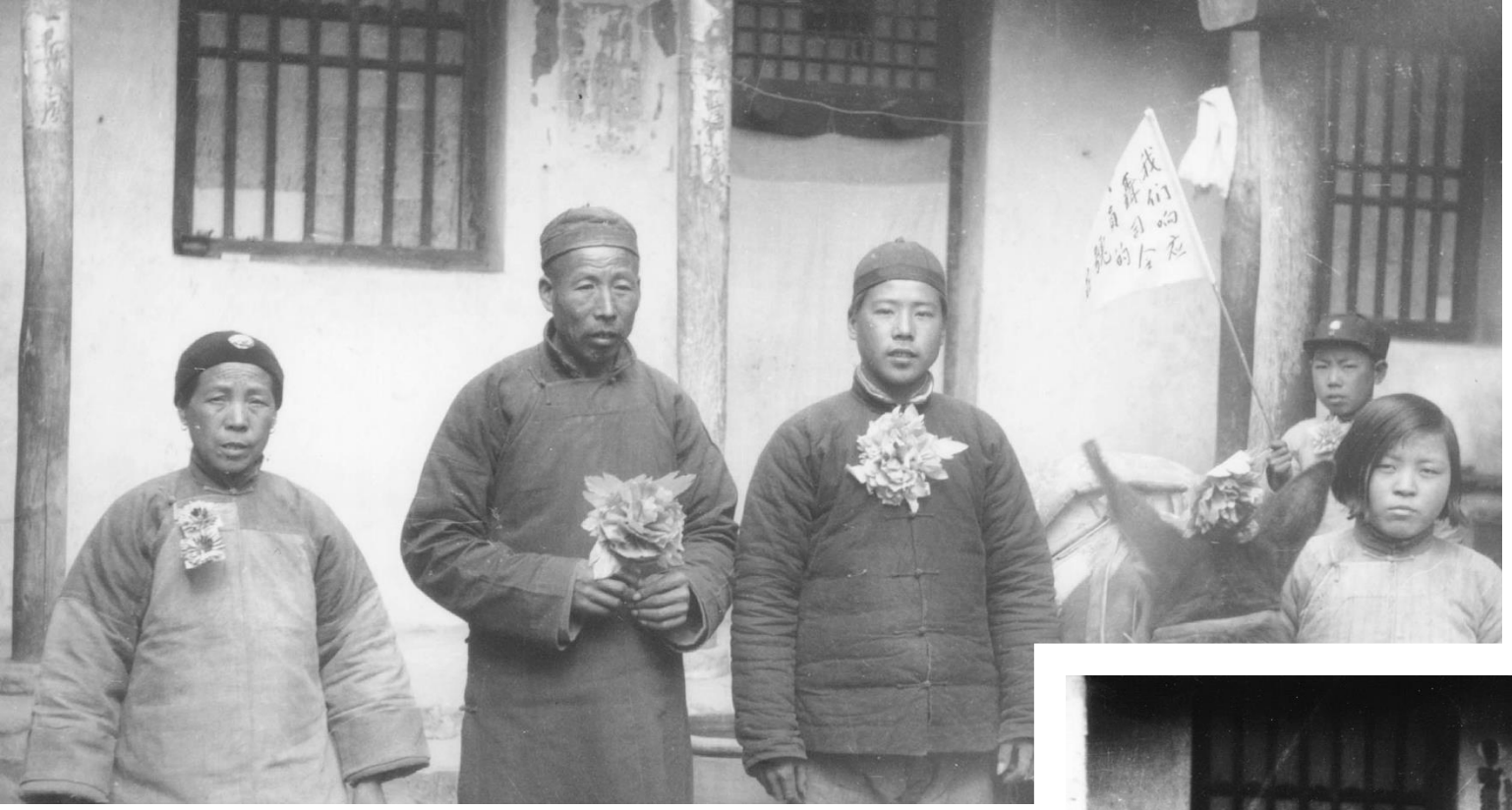
(Right- top) 48 men at the same time: At the solicitation of the chief of Self-Defense corps, 48 men collec- their patriotism volunteers; the [illegible] presents to each a (lower right- bottom) Collective re- gistration of 165 volunteers in one cere- mony at Pingshan. (left top) Parents [illegible] r boy off: In a volunteer [illegible] ceremony the old couple say goodbye to their son after delivering a public lecture (right top) Enrolling of a landlord volunteer: The man shown in the picture is a progressive landlord wao registered himself as a volunteer soldier imme diately following the pub- lication of the new service system. (left bottom) Liu Han-hsing, travelling by donkey, registers as a volunteer soldier, with his parents, brothers and his wife following at his heels.

Case Study 3: Editorial Changes in Image Contextualization

Case Study 3: Editorial Changes in Image Contextualization

Liu Hanxing's name was omitted, and he was referred to as a "youth".

Case Study 3: Editorial Changes in Image Contextualization



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

The Power of Computer Vision in Historical Analysis

Role of computer vision in identifying and tracking image modifications and circulations

Looking into the complex intersections of aesthetics, politics, and misinformation in image circulation

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

Project 1: Goal

- Use supervised learning to classify whether historical photographs from the Manchuria Railway archive were published in North China Magazine or remained unpublished.
- Why?
  - Many wartime photos were taken, but only some were published.
  - By identifying which were published or even repeatedly published, we can study how visual culture was shaped by editorial choices, aesthetic preferences, and political agendas.
- This is part of a broader project on AI-assisted visual historiography — and we may turn this into a publishable research paper together.

Dataset

- Kyoto University Archive: ~40,000 wartime photographs (1930s–1940s)
- North China Magazine: Digitized magazines from the same era
- Preprocessed Matches: Your instructor’s pipeline already matched each archive photo to its "top 10 most similar" magazine images using self-supervised learning (no human labels needed for matching).

Task

- Problem: The current pipeline has false positives (incorrect matches).
- Your Job: Improve accuracy by filtering bad matches using Aspanformer and supervised learning.
- How?
- Step 1: Label Training Data:
  - If a photo has ≥1 correct match in the magazine → label as published.
  - If all matches are incorrect → label as unpublished.
  - This becomes your labeled dataset (I already have 66 labeled matches, but we can label more if necessary).

Task

- Step 2: Deploy Aspanformer to calculate local feature matches (keypoints) between archive photos and magazine images.
  - Install Aspanformer: Follow instructions at https://aspanformer.github.io/.
  - Save Results: For each archive photo, record the maximum number of matching points among its top 10 matches.
    - Matching points reflect how much two images overlap in local features.
    - Published photos (even if cropped/edited) will share many keypoints with their magazine versions.
    - Unpublished photos will have few/no matching keypoints.

Task

- Step 3: Determine the Threshold
  - Correct matches will likely have high matching points (e.g., >50).
  - False postives will have low matching points (e.g., <20).
  - Set a threshold (e.g., 40 matches) to separate the correct matches.

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

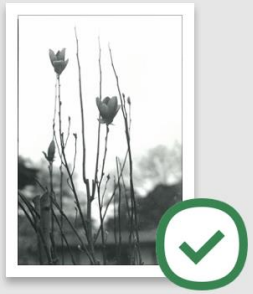

North China Magazine
1942-1-12_Page_217_1.jpg


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.

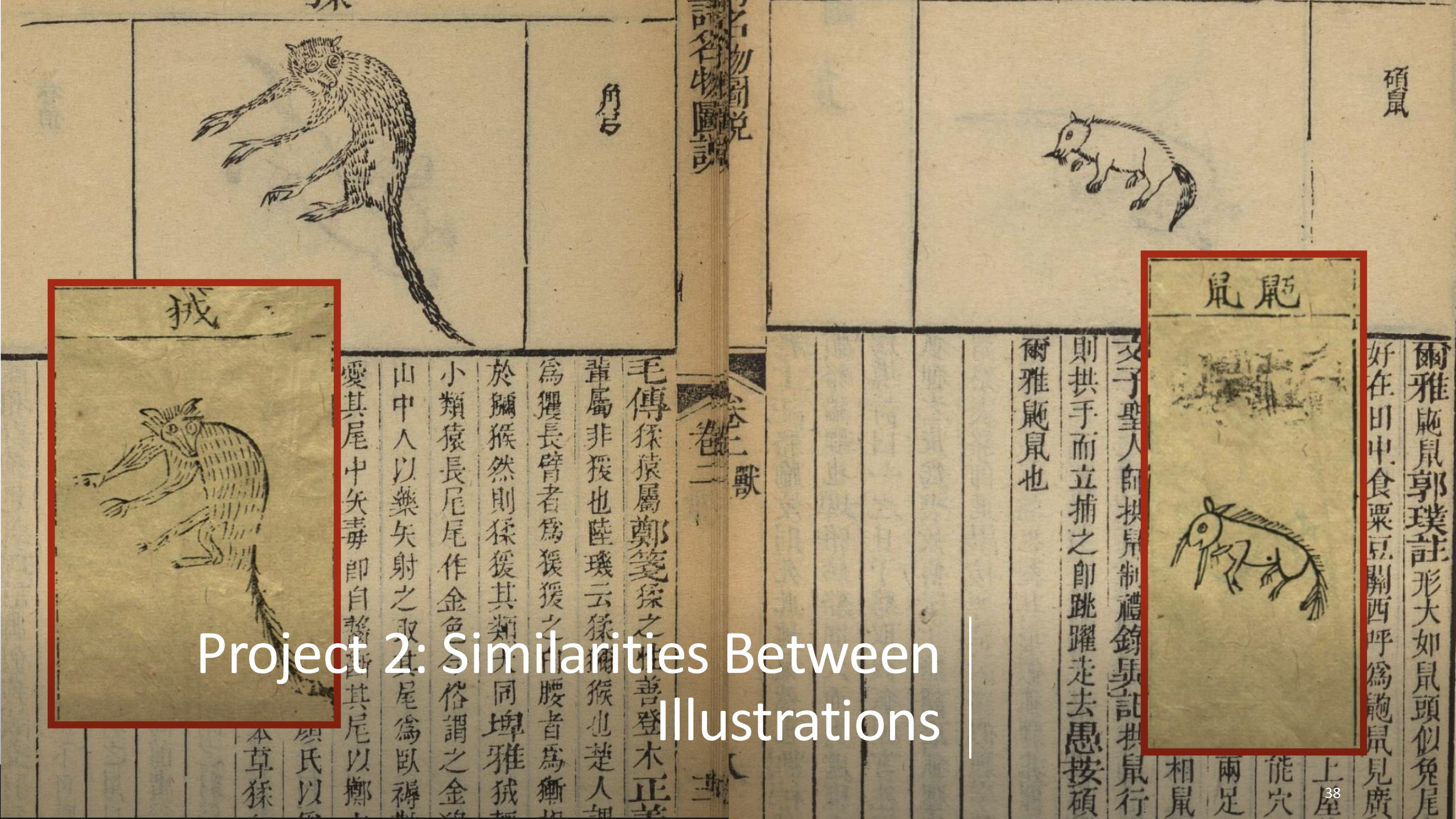
2. Chen, Ting, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. “A Simple Framework for Contrastive Learning of Visual Representations.” In *International Conference on Machine Learning*, 1597–1607. PMLR. http://proceedings.mlr.press/v119/chen20j.html.

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.

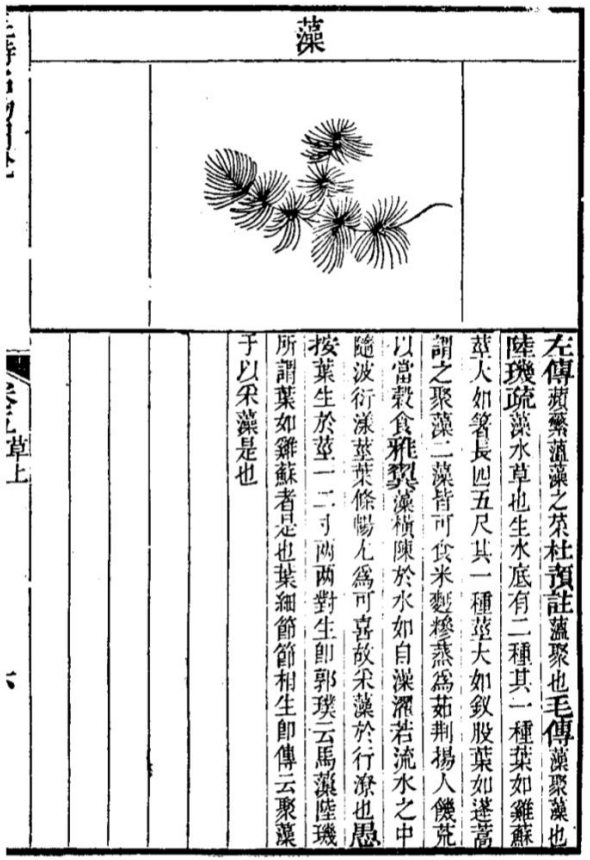
4. Dubey, Shiv Ram. 2021. “A Decade Survey of Content Based Image Retrieval Using Deep Learning.” *IEEE Transactions on Circuits and Systems for Video Technology* 32 (5): 2687–2704.

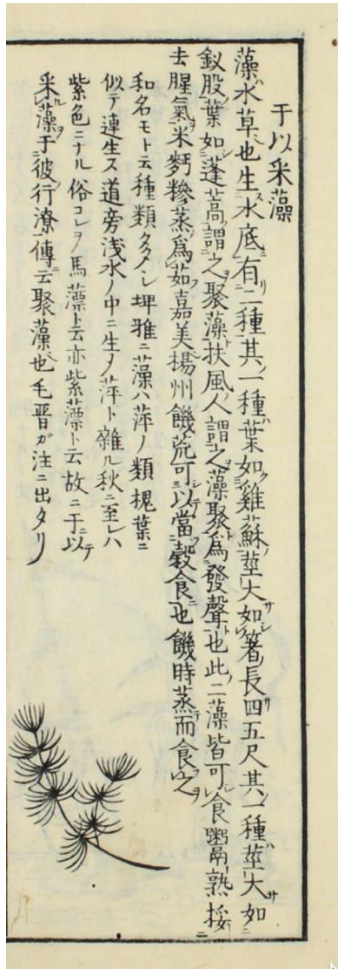
5. Ma, Jinping. 2020. “Visualizing North China Under Japanese Occupation: Digitized Photos of the North China Railway Archive.” *The Digital Orientalist* (blog). November 27, 2020. https://digitalorientalist.com/2020/11/27/visualizing-north-china-under-japanese-occupation-digitized-photos-of-the-north-china-railway-archive/.

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: Similarities Between Illustrations

藻


藻


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煮令熟接去腥氣米麪糝蒸爲如嘉美揚州饑窮可以當穀食也饑時蒸而食之

和名モトモ五種類多クシテ坪雅ニ藻ハ萍ノ類梶葉ニ似テ連生ス道旁浅水ノ中ニ生ノ萍ト雑ル秋ニ至レハ紫色ニナル俗コレヲ馬藻上云亦紫藻ト云故ニ于以采藻于彼行潦傳云聚藻也毛晉カ注ニ出タリ


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 チガヤ

Goal

- Explore how Book of Songs (Shijing 詩經) plant and animal imagery was visualized and interpreted differently in Qing China and Tokugawa-to-Meiji Japan through annotated illustrations.
- When Chinese and Japanese commentaries both include illustrations for the same object (e.g. 荇菜 xìng cài, 蘋 píng),
  - → Do they visualize the object in a similar way?
- Did Japanese illustrators copy Chinese visualizations? Or did they reinterpret them in unique ways?
- We may turn this into a publishable research paper together too.

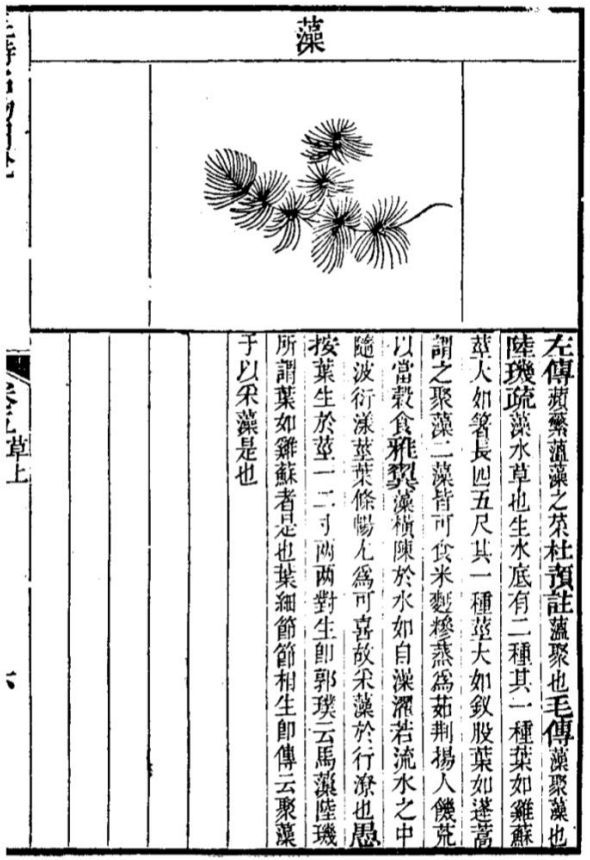
Tasks

- Goal: Use unsupervised learning to analyze how the same plant/animal from the Book of Songs (Shijing) was illustrated differently in Chinese (Qing Dynasty) and Japanese (Tokugawa-to-Meiji era) commentaries.
- Dataset:
  - Chinese: 毛诗名物图说 (Qing Dynasty, Xu Ding).
  - Japanese: 毛诗品物图考 (Okamoto Ryūho) and 陸氏草木鳥獸虫魚疏図解.

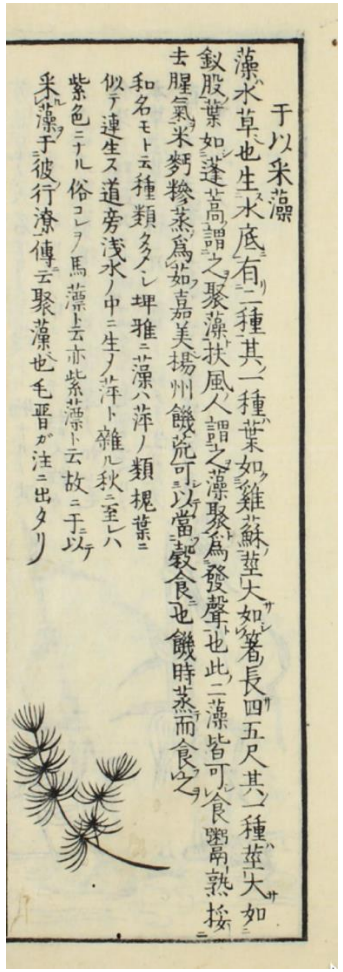
Task

- Step 1: Data Preparation: Manually crop illustrations from scanned book pages
- Step 2: Visual Comparison
- 1. Extract Image Features: Use a pretrained CNN (e.g., ResNet-18) to convert images into numerical vectors (embeddings).
  - Alternative: Use traditional features such as SIFT for simpler illustrations.
- 2. Calculate Similarity Scores: For each Chinese-Japanese pair, compute cosine similarity between their embeddings.
  - High score = Similar visuals; Low score = Different visuals.
- 3. Rank Pairs: Sort all object pairs by similarity scores (most to least similar).

藻


藻

左傳蘋蘩蕰藻之菜杜預註蕰聚也毛傳藻聚藻也
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荊揚人饑荒
以當穀食雅翟藻橫陳於水如自澡濯若流水之中
隨波行漾莖葉條暢九爲可喜故采藻於行潦也愚
按葉生於莖一二寸兩兩對生郭璞云馬藻陸璣
所謂葉如雞蘇者是也葉細節節相生郎傳云聚藻
于以采藻是也

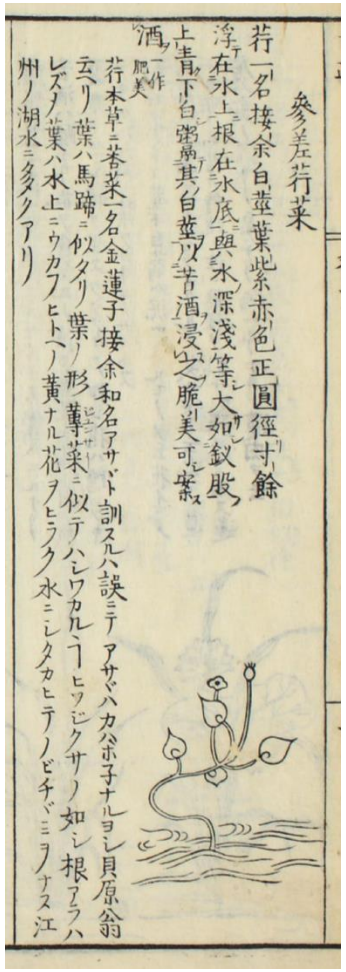

于以采藻
藻水草也生水底有二種其一種葉如雞蘇莖大如箸長四五尺其一種莖大如釵股葉如蓬蒿謂之聚藻扶風人謂之藻聚爲發聲也此二藻皆可食鄙俗熟挼去腥氣米麪糝蒸爲茹如嘉美揚州饑窮可以當穀食也饑時蒸而食之
和名モ五種類多クシテ坪雅ニ藻ハ萍ノ類梶葉ニ似テ連生ス道旁浅水ノ中ニ生ノ萍ト雜ル秋ニ至レハ紫色ニナル俗コレヲ馬藻上云亦紫藻ト云故ニ于以采藻于彼行潦傳云聚藻也毛晉ガ注ニ出タリ

荇菜


荇

周南關雎

爾雅釋草莕接余其葉荇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鄙人採其白莖以苦酒浸之脆美可案酒蘇莕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斷開雖圓而稍狹不若蓴之極圓也隨水平浮花則出水黃色六出今宛陵陂湖中頗覆項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爲莕公羹蓋細荇亂生有若髮然愚按荇似蓴菜而非蓴蓋蓴菜比荇而葉圓詩傳采其莖即蓴也陸氏德明云荇亦作莕接余也則荇與莕同俗呼爲荇絲菜許氏說文謂之攣楚辭謂之屏風云紫莖屏風文綠波皆指此也


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鄙人採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ト訓スルハ誤テアザハカホ子ナルヨシ貝原翁云リ葉ハ馬蹄ニ似タリ葉ノ形蓴菜ニ似テハレワカレテヒツジクサノ如シ根アラハレズ葉ハ水上ニウカフヒトヘノ黄ナル花ヲヒラク水ニヒタカヒテノビチヾミヲナス江州ノ湖水ニ多クアリ


毛詩品物圖攷卷一

草部

參差荇菜 アサ、

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黃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產之異也

浪華岡元鳳纂輯

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.

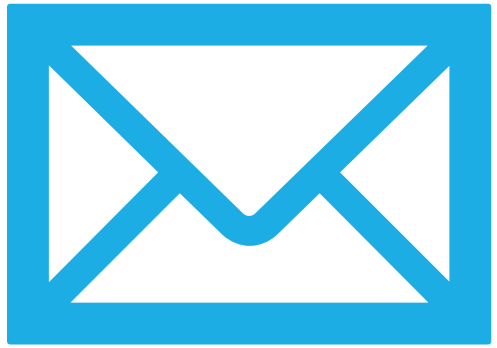
3. Lang, Sabine, and Björn Ommer. 2018a. “Attesting Similarity: Supporting the Organization and Study of Art Image Collections with Computer Vision.” *Digital Scholarship in the Humanities* 33 (4): 845–56. https://doi.org/10.1093/llc/fqy006.

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.

CONCLUSION: DIGITAL HISTORICAL FORENSICS

- Digital Historical Forensics, particularly through AI, bridge different visual media including pictorials and exhibitions with textual media to **track media circulation** and **enhance media studies**.
- Its importance in Japan and China studies lies in discovering how Chinese photojournalists and editors, despite limited material resources, effectively **made history** and competed with Japanese propaganda efforts.
- The methodology is **generalizable and extendable** across diverse cultural contexts and archives, facilitating cross-regional research.
- This approach enables **robust source criticism** at a new scale, applicable to both historical and modern media, and encourages the digitization of offline archives.
- Contact: dulin525@gmail.com