Learning to Interpret Historical Maps by Exploiting Polygon Metadata

Fandel Lin

Ph.D. Dissertation Defense

April 8th, 2026

Committee Members:

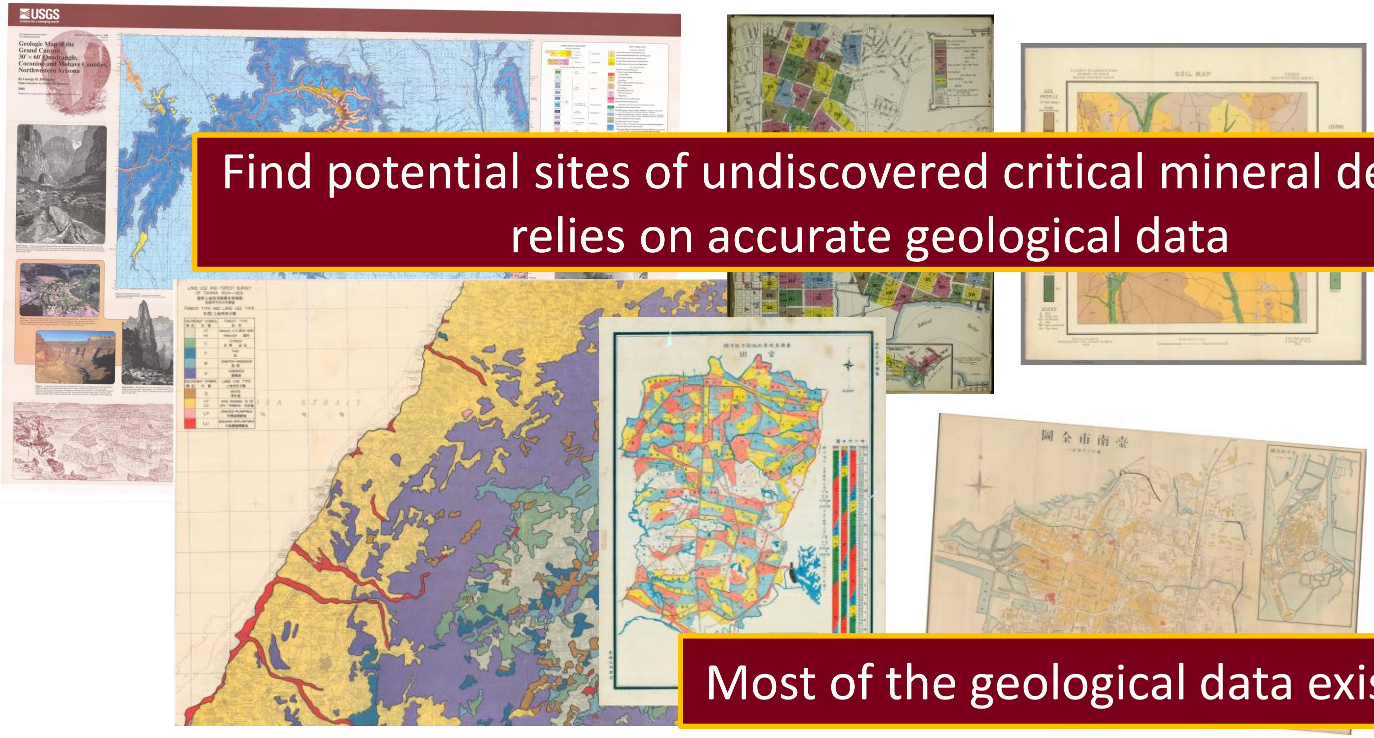
Craig A. Knoblock, Cyrus Shahabi, Yolanda Gil, Yao-Yi Chiang, and John P. Wilson

USC INFORMATION SCIENCES INSTITUTE

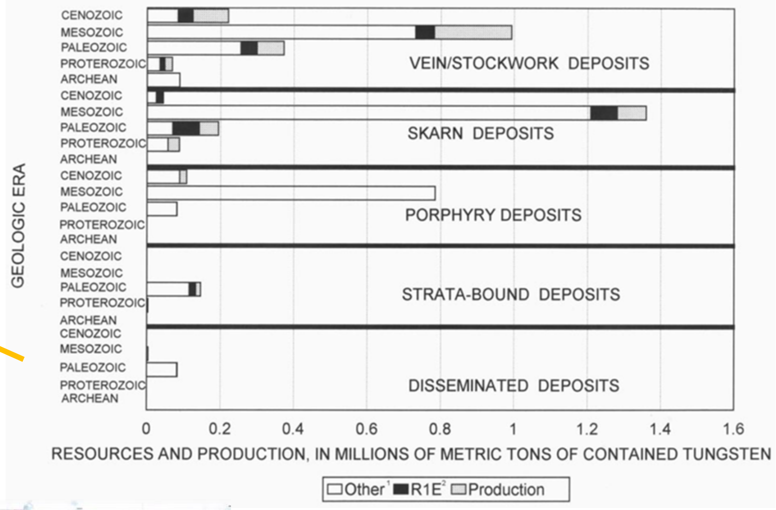
USC Viterbi School of Engineering

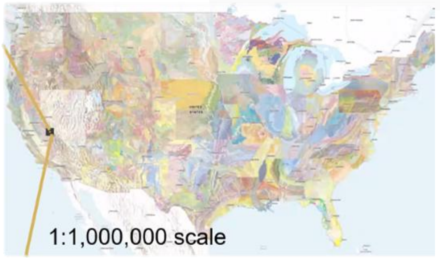
Motivation

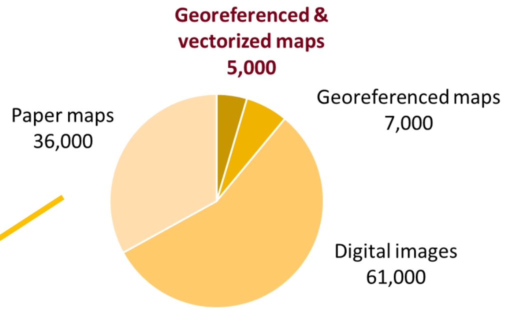
- Historical maps preserve long-term geographic, environmental, and geological information that is often unavailable from modern surveys
  - Enables a wide range of downstream analyses
  - Many exists only as scanned raster images



Find potential sites of undiscovered critical mineral deposits relies on accurate geological data





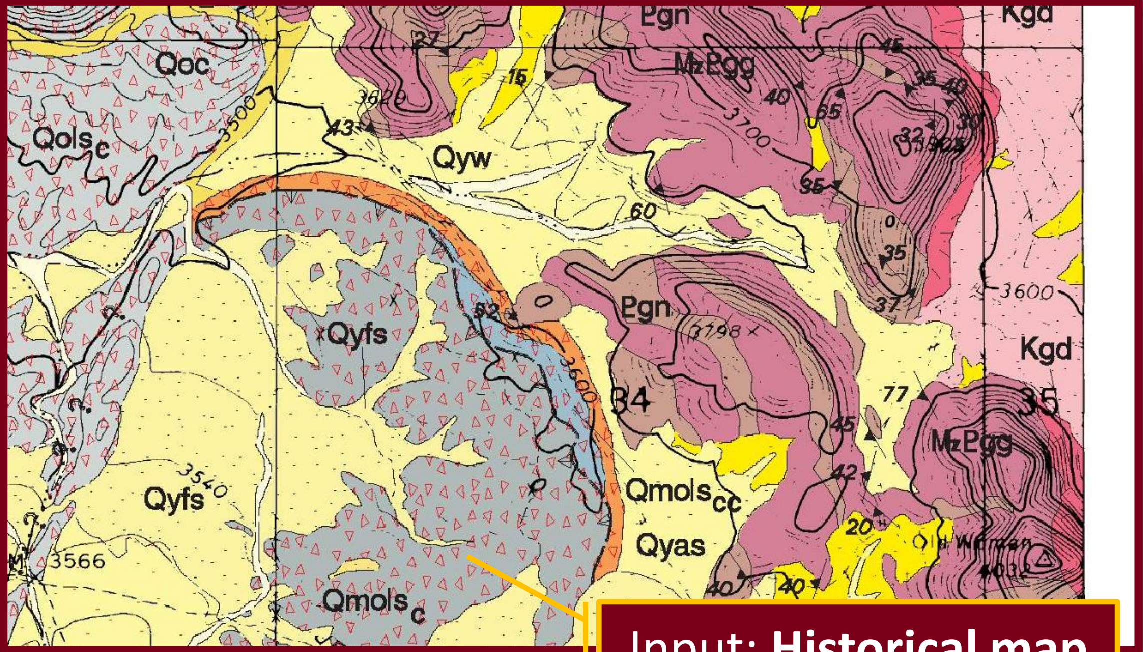


Most of the geological data exist as scanned images

Example: CMA

Problem Statement

- Given historical maps with identified map content area and polygon map keys
  - Automatically interpret and digitize polygonal features
  - Convert raster images into analysis-ready formats



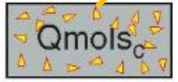
Input: **Historical map**

Output: **Digitized polygon**

We aim to digitize polygonal features from the map content area

Map content area

Polygon feature is identified by a **map key** from legend, specifying its color, pattern, and meaning



Moderately old landslide breccia, carbonate rocks (middle or early Pleistocene)—Rock avalanche breccia derived from metamorphosed Paleozoic carbonate strata

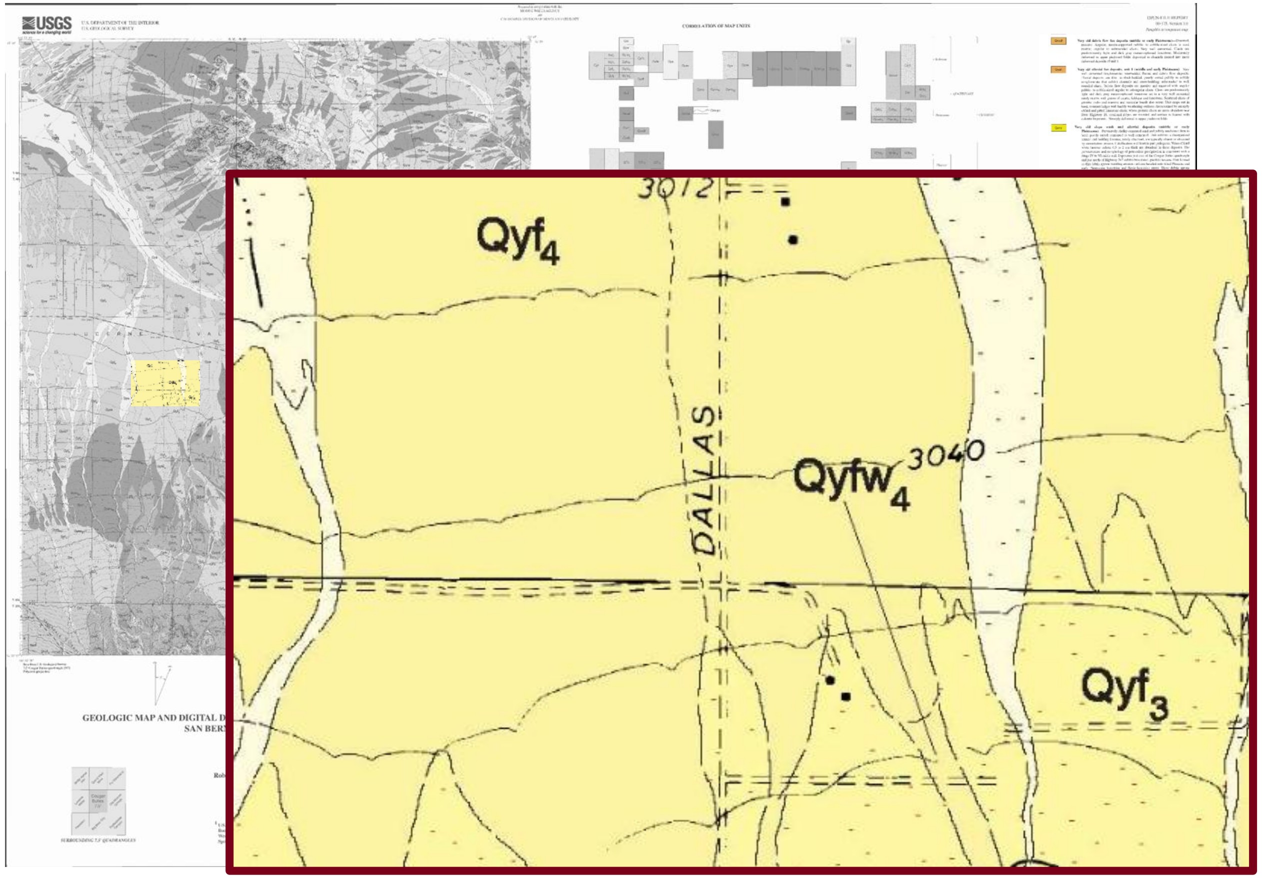
Input: **Map key**

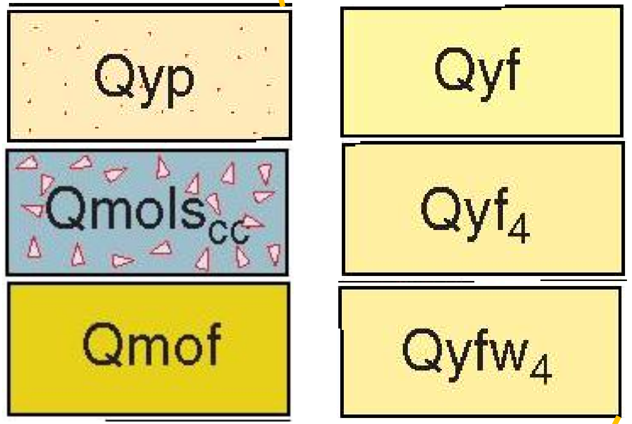
USC Viterbi School of Engineering

Challenges in Digitizing Polygonal Feature

1. Various styles of map keys
2. Same color, different text

A map has various colors or markings
to indicate the corresponding polygon features





They are all alluvial deposits. One is fan deposit;
the others are fan feeder wash deposit

Image Source: CA_CougarButtes, USGS

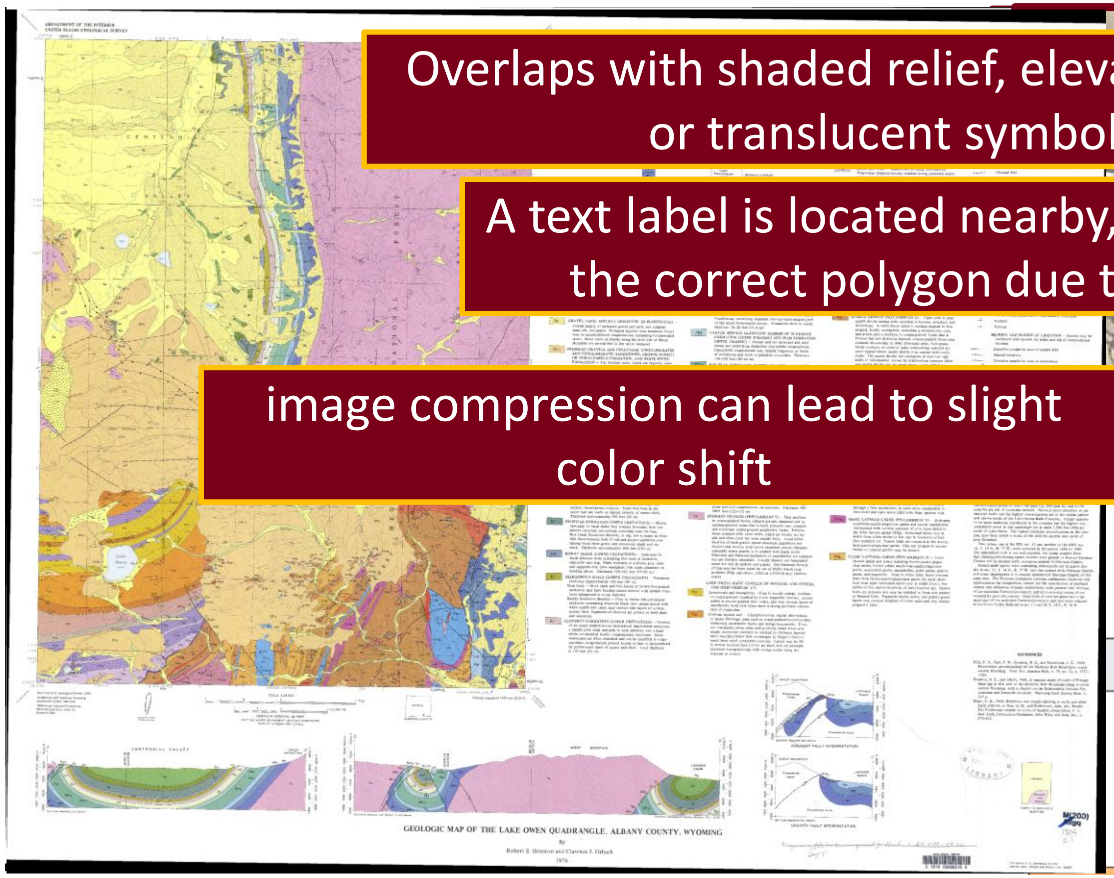
USC Information Sciences Institute

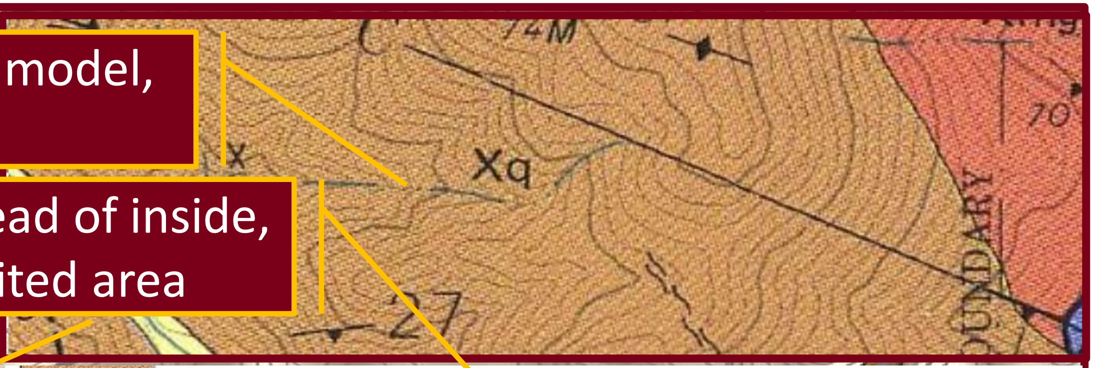
Rock info

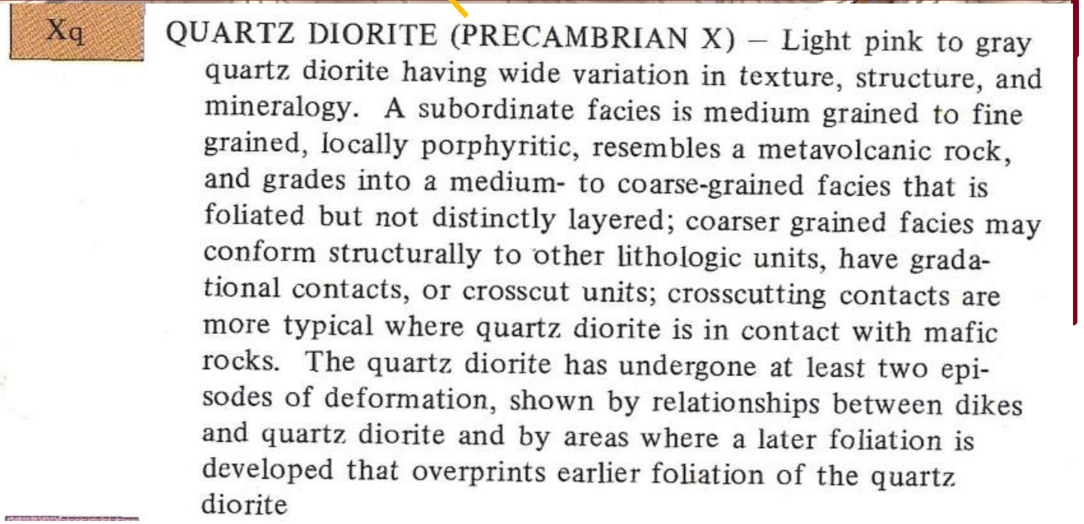
USC Viterbi School of Engineering

Challenges in Digitizing Polygonal Feature (cont.)

1. Various styles of map keys
2. Same color, different text
3. Text labeled outside polygon
4. Color shift between key and content







Xq QUARTZ DIORITE (PRECAMBRIAN X) — Light pink to gray quartz diorite having wide variation in texture, structure, and mineralogy. A subordinate facies is medium grained to fine grained, locally porphyritic, resembles a metavolcanic rock, and grades into a medium- to coarse-grained facies that is foliated but not distinctly layered; coarser grained facies may conform structurally to other lithologic units, have gradational contacts, or crosscut units; crosscutting contacts are more typical where quartz diorite is in contact with mafic rocks. The quartz diorite has undergone at least two episodes of deformation, shown by relationships between dikes and quartz diorite and by areas where a later foliation is developed that overprints earlier foliation of the quartz diorite

Overlaps with shaded relief, elevation model, or translucent symbols

A text label is located nearby, instead of inside, the correct polygon due to limited area

image compression can lead to slight color shift

Challenges in Digitizing Polygonal Feature (cont.)

5. Significant color mismatch

Recoloring digitized polygons using corresponding keys' median color demonstrates the color shift/ mismatch in original map content

Scanning artifacts such as creases can lead to color mismatch

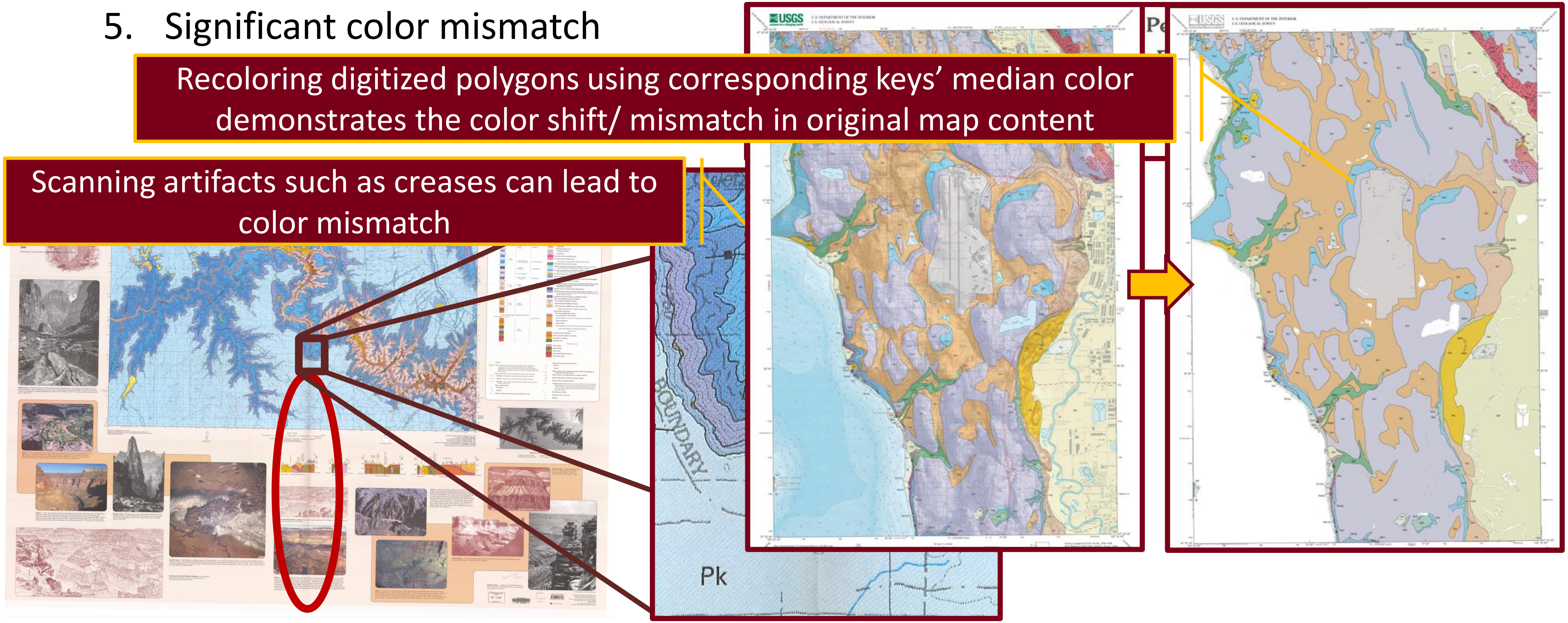
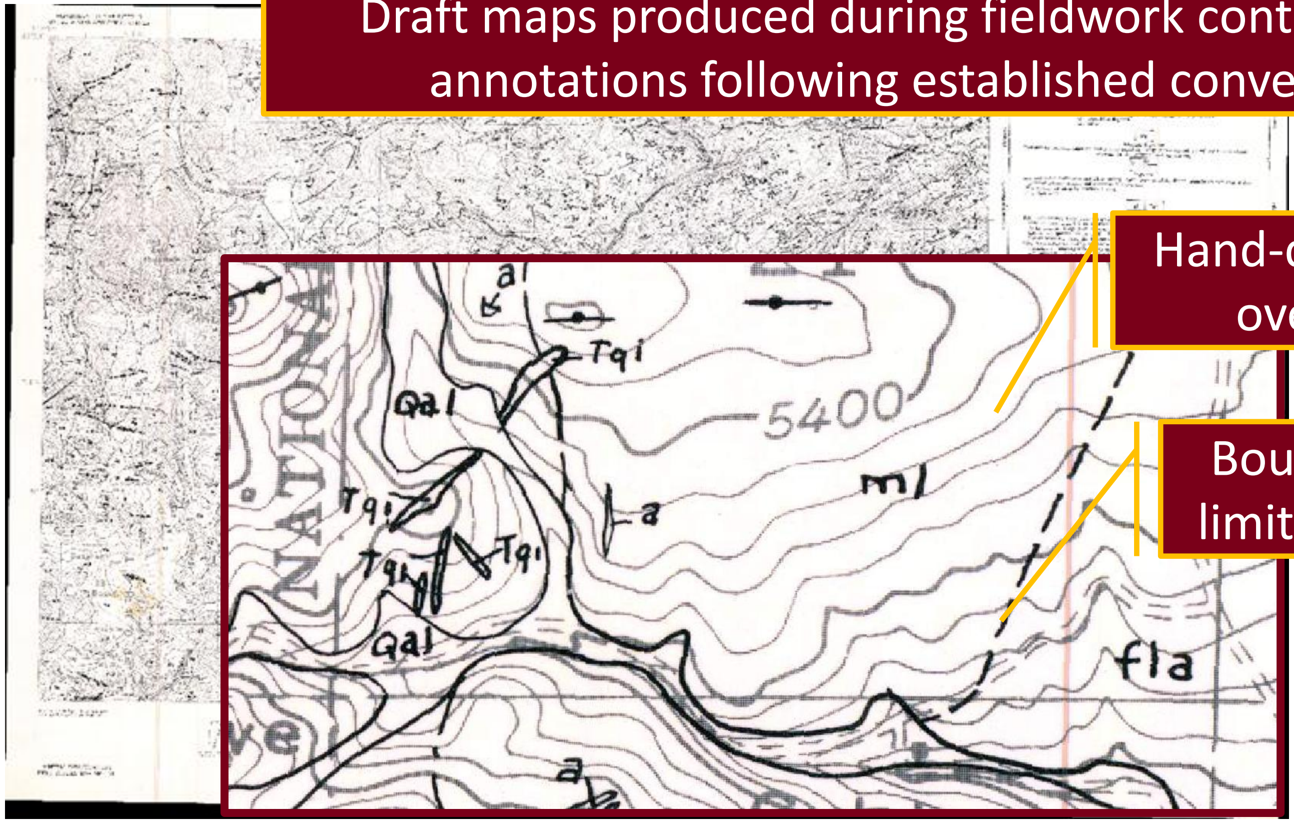


Image Source: AZ_GrandCanyon, USGS; WA_DesMoines, USGS

Challenges in Digitizing Polygonal Feature (cont.)

5. Significant color mismatch between key and content
6. Monochromatic or achromatic draft maps
7. Incomplete polygon boundaries

Draft maps produced during fieldwork contains hand annotations following established conventions



Hand-drawn boundary lines overlay on basemap

Boundary lines may no be fully closed, limiting segmentation-base approaches

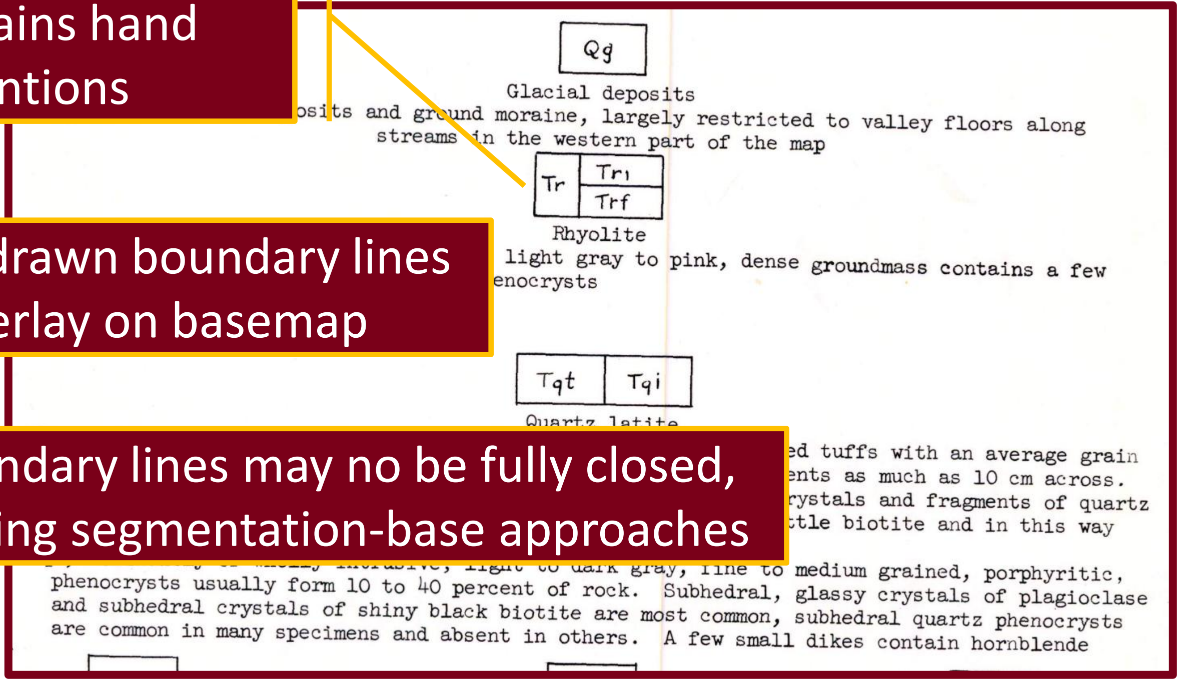


Image Source: USGS

Thesis Statement

- We can build a system that automatically learns to interpret historical maps by colorizing, recoloring, and digitizing polygonal features using polygon metadata

Metadata: *data that provides information about other data*

Any data that is on the map, describing or about the polygon feature, but cannot be the set of vertices that constitutes the polygon itself

Exploiting Polygon Metadata to Digitize Polygonal Features (*SIGSPATIAL 2023*)

Exploiting Polygon Metadata to Recolor Historical Maps (*SIGSPATIAL 2025*)

Exploiting Polygon Metadata to Colorize Draft Maps (*SIGSPATIAL 2025*)

Generalizing Polygon Digitization to Various Map Styles (*Under Review at KDD*)

Publication

- Exploiting Polygon Metadata to Digitize Polygonal Features
  - **Fandel Lin**, Craig A. Knoblock, Basel Shbita, Binh Vu, Zekun Li, and Yao-Yi Chiang. 2023. Exploiting Polygon Metadata to Understand Raster Maps: Accurate Polygonal Feature Extraction. In *The 31st ACM International Conference on Advances in Geographic Information Systems* (***SIGSPATIAL ’23***), November 13–16, 2023, Hamburg, Germany.
- Exploiting Polygon Metadata to Recolor Historical Maps
  - **Fandel Lin**, Craig A. Knoblock, Binh Vu, and Yao-Yi Chiang. 2025. Exploiting Polygon Metadata to Recolor Historical Maps. In *The 33rd ACM International Conference on Advances in Geographic Information Systems* (***SIGSPATIAL ’25***), November 3–6, 2025, Minneapolis, MN, USA.
- Exploiting Polygon Metadata to Colorize Draft Maps
  - **Fandel Lin**, Craig A. Knoblock, Binh Vu, Basel Shbita, and Yao-Yi Chiang. 2025. Exploiting Polygon Metadata to Colorize Draft Maps. In *The 33rd ACM International Conference on Advances in Geographic Information Systems* (***SIGSPATIAL ’25***), November 3–6, 2025, Minneapolis, MN, USA.
- Exploiting Polygon Metadata to Generalize Digitization across Styles
  - **Fandel Lin**, Zekun Li, Yao-Yi Chiang, and Craig A. Knoblock. 2026. Cross-domain Polygon Extraction from Historical Maps via Legend-guided Semantic Fusion. *Under Review at KDD*.

Outline

- Introduction and Overview
- Our Approach
  - Exploiting Polygon Metadata to Digitize Polygonal Features
  - Exploiting Polygon Metadata to Recolor Historical Maps
  - Exploiting Polygon Metadata to Colorize Draft Maps
  - Exploiting Polygon Metadata to Generalize Digitization across Styles
- Conclusion and Future Work

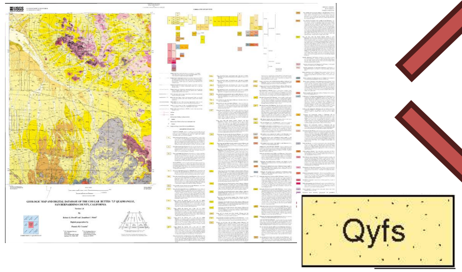
Outline

- Introduction and Overview
- Our Approach
  - **Exploiting Polygon Metadata to Digitize Polygonal Features**
  - Exploiting Polygon Metadata to Recolor Historical Maps
  - Exploiting Polygon Metadata to Colorize Draft Maps
  - Exploiting Polygon Metadata to Generalize Digitization across Styles
- Conclusion and Future Work

Our Approach to Digitizing Polygonal Features

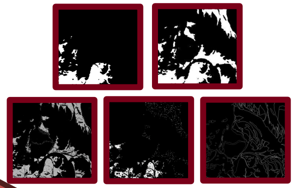
- Exploit **Polygon Metadata** in two ways

Input Map/ Map Key



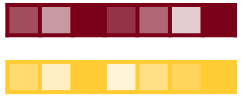
Polygon-feature Preprocessing

Generate representations indicating various map-understanding aspects



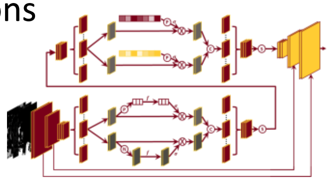
Auxiliary-information Embedding

Quantify the characteristics of (1) map and (2) map key



Polygon-recognition Model

Teach model to leverage different representations

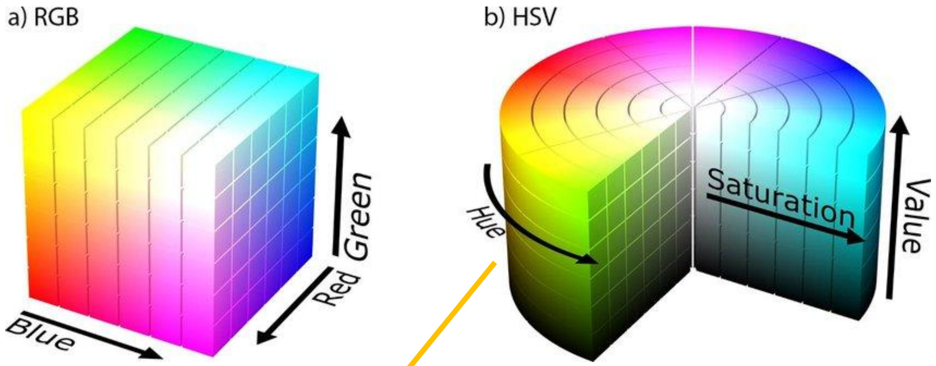


Output

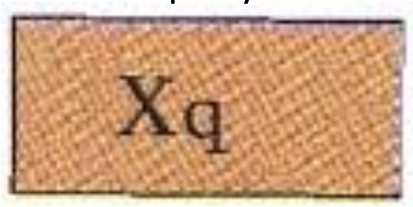
Overview

Polygon-feature Preprocessing

- Apply aspects that **human read a map**:
  - Find areas with similar colors
  - Distinguish overlapped symbols and textures
  - Use texts or boundaries

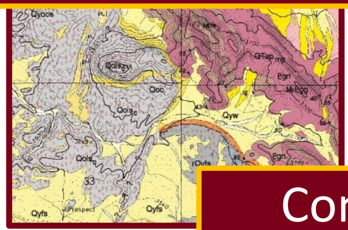


From map key:

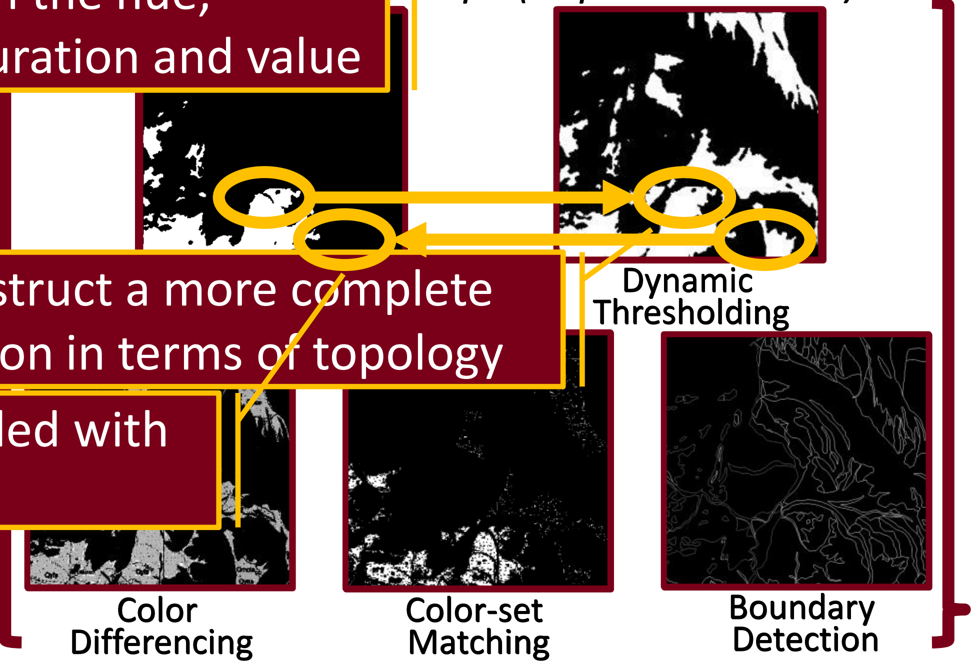


In map area

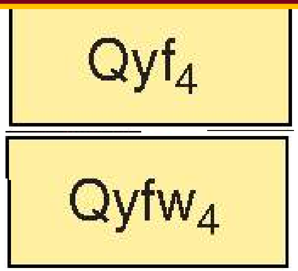
Put more emphasis on the hue, with higher tolerance for saturation and value

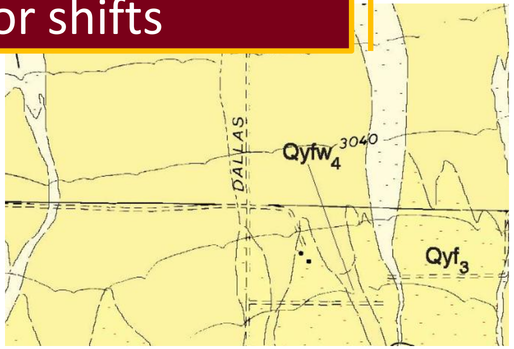


tmaps (Representations)



Overcome resolution issue and slight color shifts



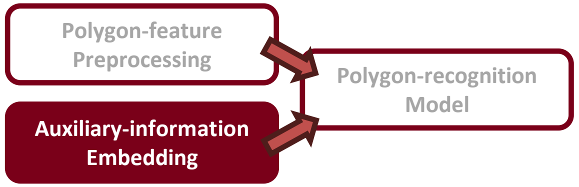


Construct a more complete polygon in terms of topology

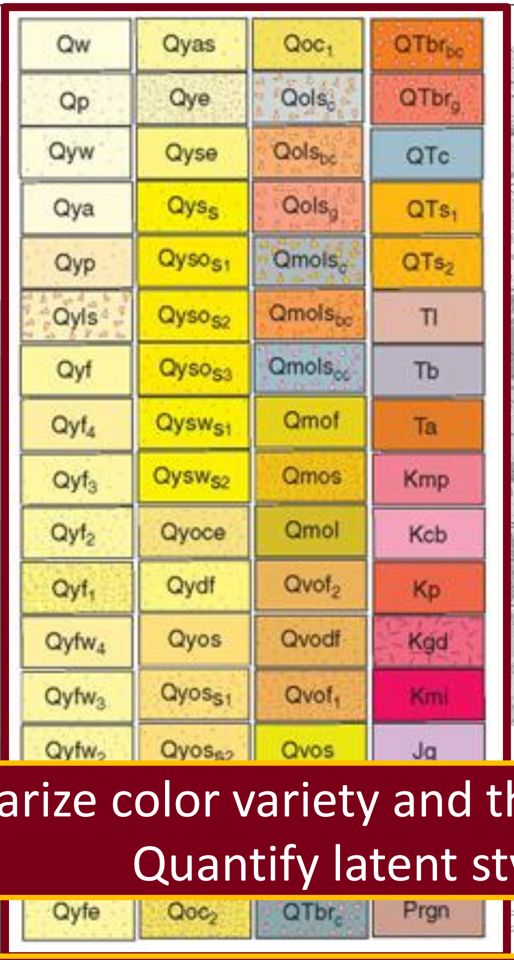
Identify polygon labeled with different texts

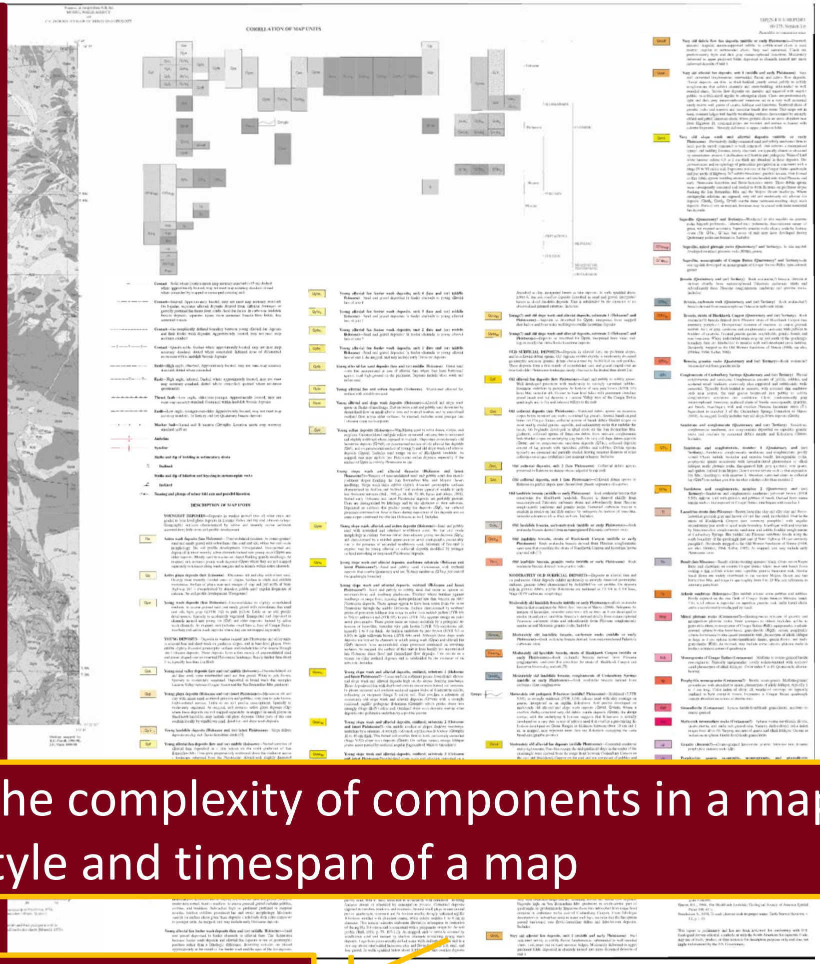
Image Source: Vencislav Popov, Practices and pitfalls in inferring neural representations. NeuroImage, 174.

Auxiliary-information Embedding



- The color of a map key indicates its thematic meaning (e.g., geologic time and rock type)
  - Quantify characteristics of:
    - Map
    - Map key

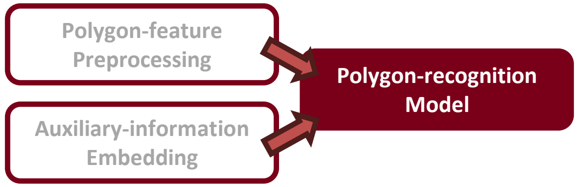




Summarize color variety and the complexity of components in a map.
Quantify latent style and timespan of a map

Summarize the color difference and variety of map keys.
Quantify whether a color is distinguishable

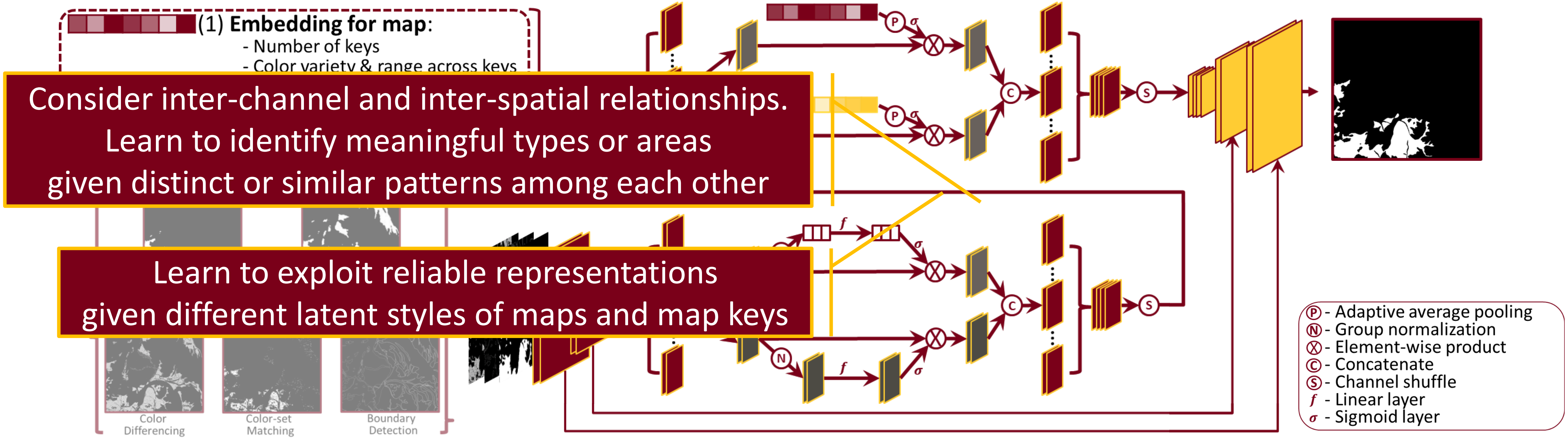
Polygon-recognition Model



- Leverage different representations to **adapt to various styles** of maps and map keys
  - Information from **representations** themselves
  - Information from characteristics of **map and map keys**

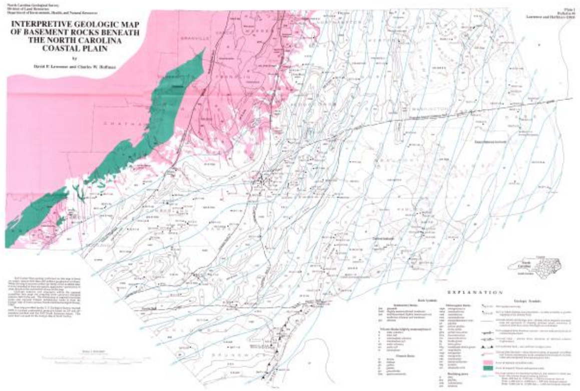
Consider inter-channel and inter-spatial relationships.
Learn to identify meaningful types or areas
given distinct or similar patterns among each other

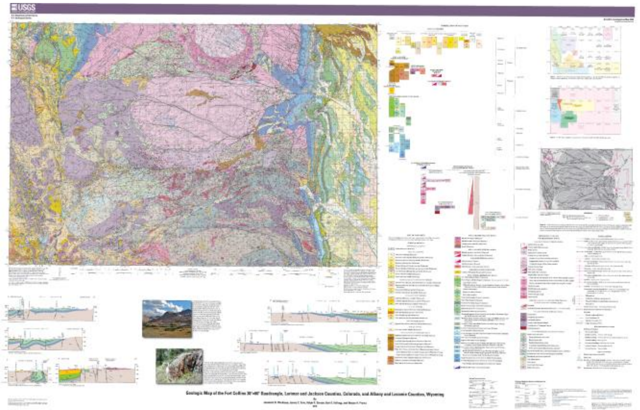
Learn to exploit reliable representations
given different latent styles of maps and map keys

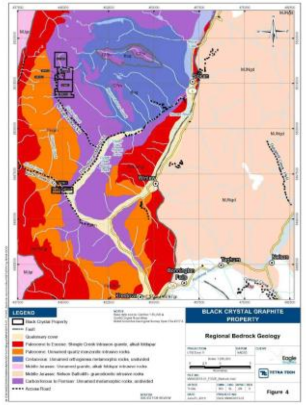


Evaluation Setting

- Use USGS geologic maps released in a competition
  - A public benchmark specifying training, validation, and testing sets with no overlap at key level





DARPA AI for Critical Mineral Assessment Competition

Attribute / Usage	Training	Testing
Number of raster maps	14	24
Number of map keys	536	849
Maximum number of map keys per map	100	103
Median number of map keys per map	31	31
Minimum number of map keys per map	13	1
Maximum size of a raster map (pixel, width x height)	17,572 x 15,950	13,200 x 18,450
Minimum size of a raster map (pixel, width x height)	6,479 x 13,614	7,200 x 11,113

Reference: https://criticalminerals.darpa.mil/Files/Map_Feature_Extraction_Challenge_Details.pdf

Overall Performance – Digitization

- Legend-Oriented Automated polygon digitization from Maps (LOAM)
  - Outperforming state-of-the-art methods

Method	F1 Score	
Our Approach - LOAM	**0.809**	
DARPA AI for CMA competition - Polygon		
Sub-task 1^{st}-place: team "ICM"	0.774	• U-Net with OCR and histogram equalization
Sub-task 2^{nd}-place: team "uncharted"	0.632	• Entropy analysis with spatial high-pass filter
Sub-task 3^{rd}-place: team "ISI-UMN"	0.629	• Color thresholding with pattern matching
Instance-segmentation method		
SAM with Color Matching	0.282	• Zero-shot generalization
Baseline		
Color Matching	0.046	• Competition baseline

Related Work – Digitization

- **Polygonal Feature Digitization**
  - Polygon Extraction from Geologic Maps (Luo 2023, Saxton 2024)
    - Apply U-Net with OCR and histogram equalization
    - Do not address the color shift and map styles
  - Polygon Extraction from Historical Maps (Arteaga 2013)
    - Apply a series of image-processing techniques to extract buildings from historical maps
    - Formulate as foreground detection with no map key
  - ML for Polygon Extraction from Maps (Wu 2023, Heitzler 2020, Jiao 2022, Wu 2022, Garcia 2021, Xia 2024)
    - Apply U-net, U-Net with a transformer, Segment Anything Model (SAM), fine-tuned SAM to extract buildings, roads, hydrological features, or archaeological features from historical maps
    - Require a tailored model for each type of polygon feature
    - U-net outperformed SAM-based methods given sufficient training data

Takeaways – Digitization

- LOAM exploits polygon metadata to digitize polygonal features from historical maps
  - Multiple representations using colors and patterns from the key
  - Adaptive digitization across diverse map styles
  - Handle arbitrary input keys in maps, with similar map style and unkown key during training

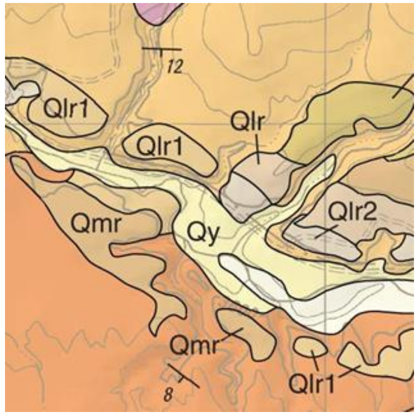
- Limitation
  - Extreme color shift

Outline

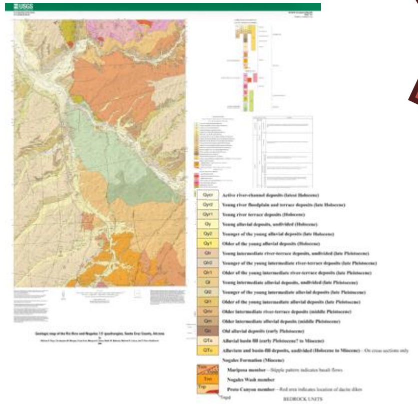
- Introduction and Overview
- Our Approach
  - Exploiting Polygon Metadata to Digitize Polygonal Features
  - **Exploiting Polygon Metadata to Recolor Historical Maps**
  - Exploiting Polygon Metadata to Colorize Draft Maps
  - Exploiting Polygon Metadata to Generalize Digitization across Styles
- Conclusion and Future Work

Our Approach to Recoloring Historical Maps

- Exploit **Polygon Metadata** in two ways

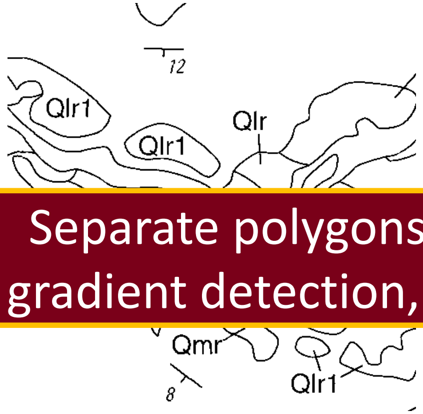


Input Map/ Map Key



Sketch Extraction

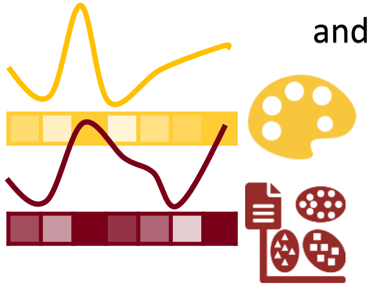
Preserve a structural backbone for regions with distinct colors



Separate polygons using color thresholding, gradient detection, and canny-edge detection

Color Style Extraction

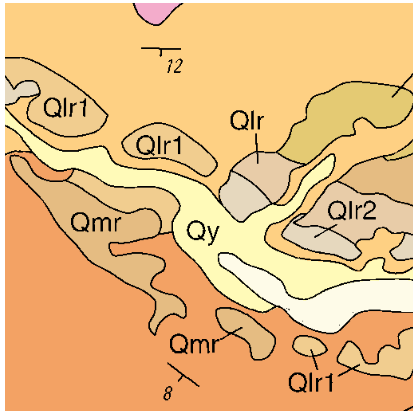
Quantify color histogram and palette from both map and map key



Multi-scale Recoloring Model

Teach model to align color histogram and adjust colors accordingly

Output Recolored Map



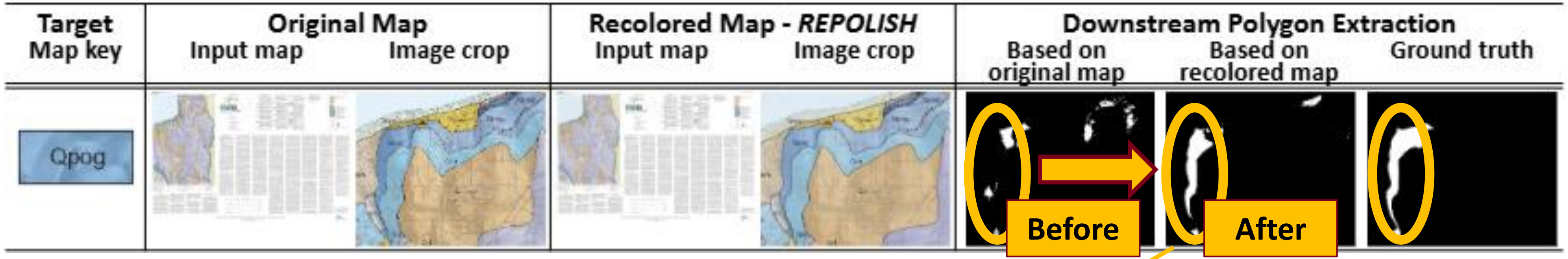
Overall Performance – Recoloring

- **RE**coloring via **P**olygon-**O**riented **L**earning with **I**nterpretative **S**pectra in **H**istorical maps (REPOLISH)
  - Outperforming state-of-the-art methods

Method	PSNR (↑)	SSIM (→1)
REPOLISH (Ours)	**21.652** ± 3.508	**0.930** ± 0.048
Reference-based Recoloring	20.015 ± 2.821	0.865 ± 0.061
Color-set-based Recoloring	16.451 ± 2.200	0.459 ± 0.043

- GAN with structure-oriented color styles
- Pixel-based color-set distance matching

- PSNR (↑) (peak signal-to-noise ratio): MSE-based color similarity
- SSIM (↑) (structural similarity index measure): Index combining luminance, contrast, and covariance



Improve downstream polygon digitization by 18% in precision

Related Work – Recoloring

- **Historical Maps Recoloring**
  - Synthetic Historical Map Generation (Li 2019, Li 2020, Lopez-Rauhut 2025)
    - Apply GAN to simulate particular map styles
    - Do not address the color inconsistency between map content and map keys
  - Line-art Recoloring (Ci 2018, Shi 2020, Liu 2022, Carrillo 2023, He 2025)
    - Propagate colors or transfer color styles from reference image
    - Exploit spontaneous markings with diffusion or adversarial models
    - Rely on explicit and strong alignment between the input image and guidance
  - Natural Image Recoloring (Chang 2015, Zhang 2017, Zhao 2021, Lin 2023, Qiu 2023)
    - Exploit text, palette, or region guidance with multimodal frameworks or optimization techniques
    - Aim to restore colors or improve perceptual quality
    - Unreliable for historical maps with various overlapping features

Takeaway – Recoloring

- REPOLISH exploits polygon metadata to recolor historical maps with color mismatch
  - Automated detection of color mismatches
  - Adaptive color re-alignment for polygonal features
  - Supporting polygon digitization using only in-map information

- Limitation
  - Still requiring some colors in the map content

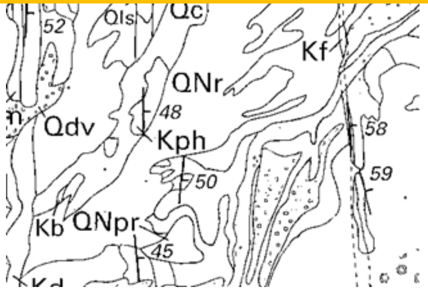
Outline

- Introduction and Overview
- Our Approach
  - Exploiting Polygon Metadata to Digitize Polygonal Features
  - Exploiting Polygon Metadata to Recolor Historical Maps
  - **Exploiting Polygon Metadata to Colorize Draft Maps**
  - Exploiting Polygon Metadata to Generalize Digitization across Styles
- Conclusion and Future Work

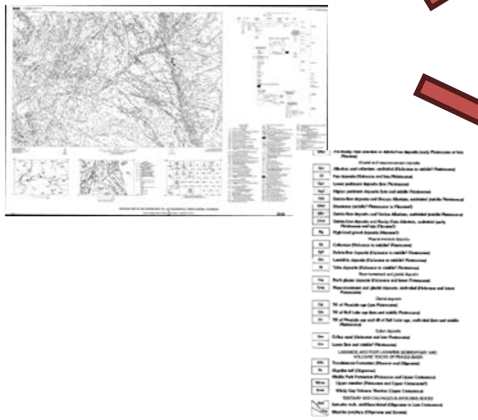
Our Approach to Colorizing Draft Map

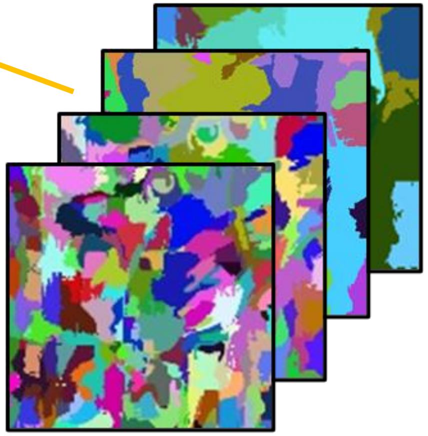
- Exploit **Polygon Metadata** in two ways

Hierarchical segmentation serve as a soft constraint to encourage a flat-color style within each polygon



Input Map/ Map Key



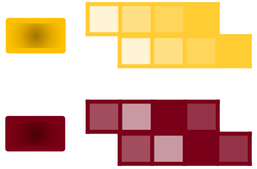


Region Map Generation

Preserve topology and determine similarity among over-segmented polygons

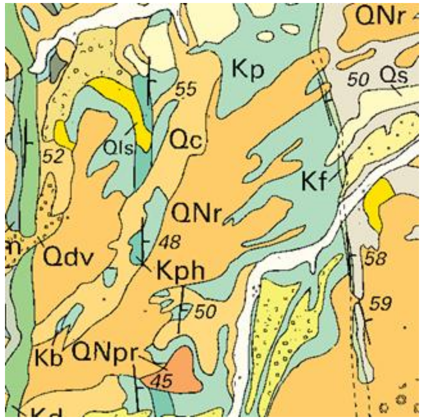
Semantic Tag Encoding

Quantify semantic identity and styles to determine potential colors



Conditioned Colorization Model

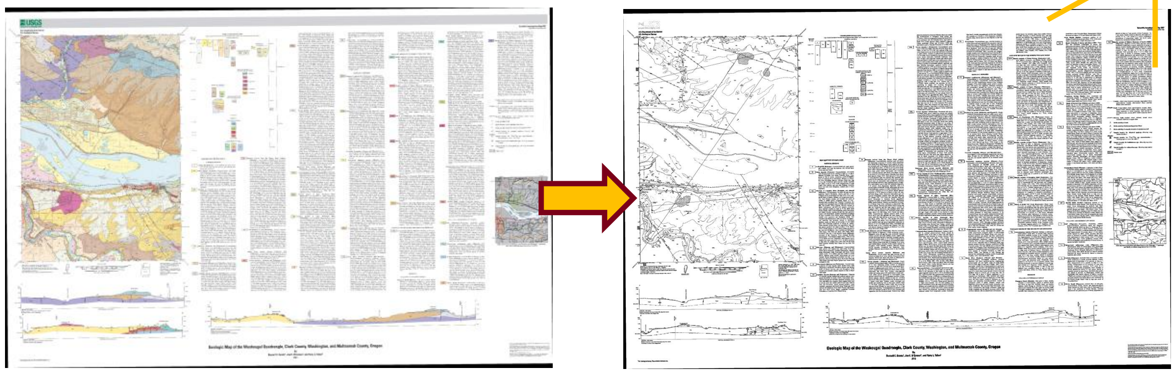
Teach model to align patterns and assign colors to nearby over-segmented polygons



Output Colorized Map

Overall Performance – Colorization

- **S**emantic-**H**armonic **A**chromatic **D**raft **I**nterpretation and **N**arration for **G**eological maps (SHADING)
  - Outperforming state-of-the-art methods



Generate synthetic data by sketch extraction and removing colors from the dataset

Evaluation is limited to maps with all keys having distinct text patterns

Method	PSNR (↑)	SSIM (→1)
SHADING (Ours)	19.339	0.798
Modified Tag2Pix	10.997	0.255
Pattern Filtering	16.721	0.356

- GAN with text-oriented color styles
- Pixel-based pattern-filtering assignment

- PSNR (↑) (peak signal-to-noise ratio): MSE-based color similarity
- SSIM (↑) (structural similarity index measure): Index combining luminance, contrast, and covariance

Outperform comparative methods by 15%

Related Work – Colorization

- **Draft Maps Colorization**
  - Synthetic Historical Map Generation (Li 2019, Li 2020, Lopez-Rauhut 2025)
    - Apply GAN to simulate particular map styles
    - Do not address the colorization guided by map keys
  - Line-art Colorization (Kim 2019, Liu 2022, Carrillo 2023, Wang 2023)
    - Apply image, region, or text as a guidance to assign colors to black-and-white line drawings
    - Do not address semantic interpretation of keys or flat-color consistency within each polygon
  - Natural Image Colorization (Wang 2022, Liang 2025)
    - Improve perceptual realism for grayscale images based on statistical priors or inputs
    - Apply image, text, palette, or explicit hints (human input) to guide the model for colorization
    - Unreliable for historical maps with various overlapping features

Takeaway – Colorization

- SHADING exploits polygon metadata to colorize draft thematic maps
  - Multi-level polygon over-segmentation
  - Adaptive color assignment for polygonal features
  - Supporting polygon digitization using only in-map information

- Limitation
  - Still requiring "in-domain" training data with similar cartographic styles

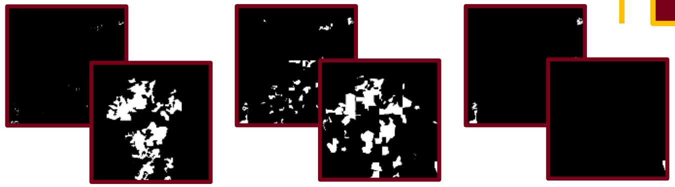
Outline

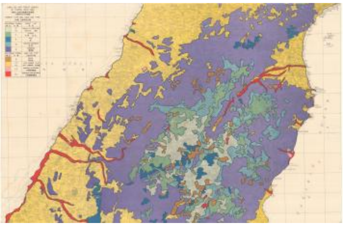
- Introduction and Overview
- Our Approach
  - Exploiting Polygon Metadata to Digitize Polygonal Features
  - Exploiting Polygon Metadata to Recolor Historical Maps
  - Exploiting Polygon Metadata to Colorize Draft Maps
  - **Exploiting Polygon Metadata to Generalize Digitization across Styles**
- Conclusion and Future Work

Our Approach to Generalize Digitization across Styles

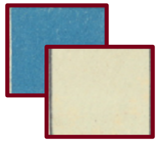
- Exploit **Polygon Metadata** in two ways

Leverage polygon-digitization results from methods with distinct emphases and potential failure patterns





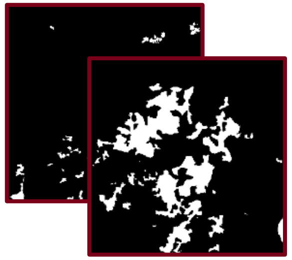
Input Map/ Map Key



Legend-oriented Polygon Extraction

Employ various methods to segment or extract polygon features per map keys

Test-time Adaptive Self Calibration

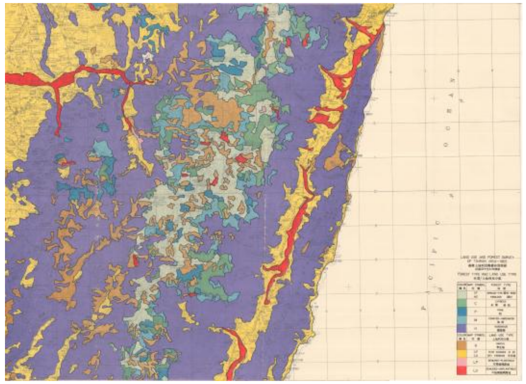


Learn entropy-regularized mixture weighting via uncertainty matching and contrastive representation alignment

Use polygon consensus for boundaries and pseudo labels; use map keys as soft anchors to support contrastive gating

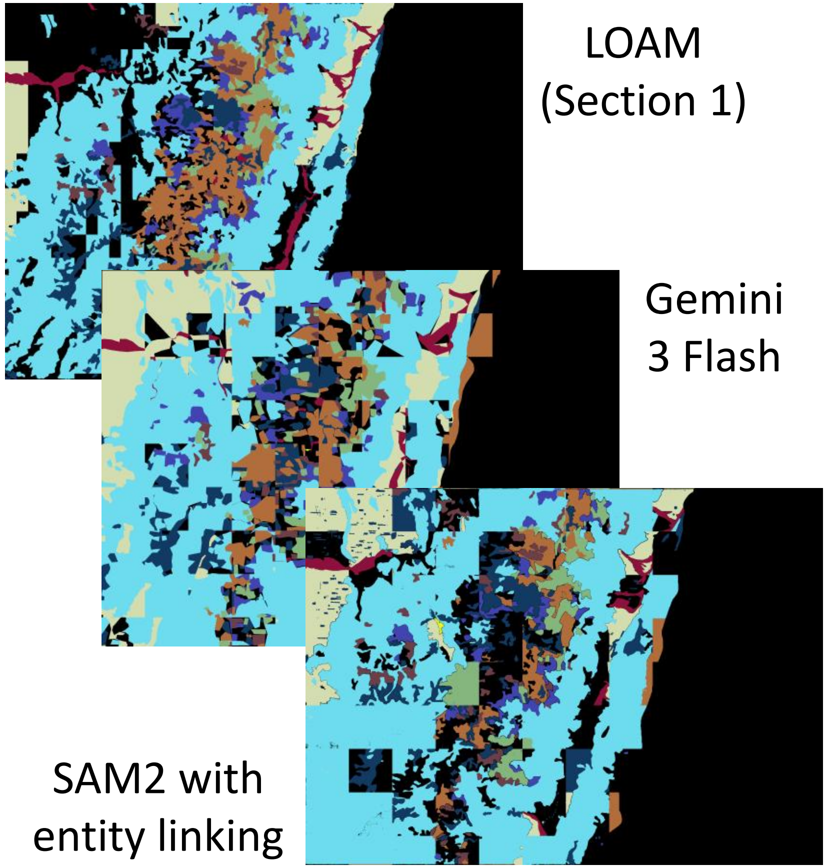
Legend-oriented Polygon Extraction

- Reconcile solutions for minimal polygon instances with potential linked map key



Input map

Map keys

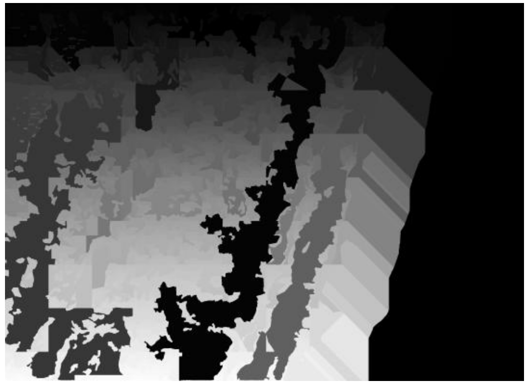


LOAM
(Section 1)

Gemini
3 Flash

SAM2 with
entity linking

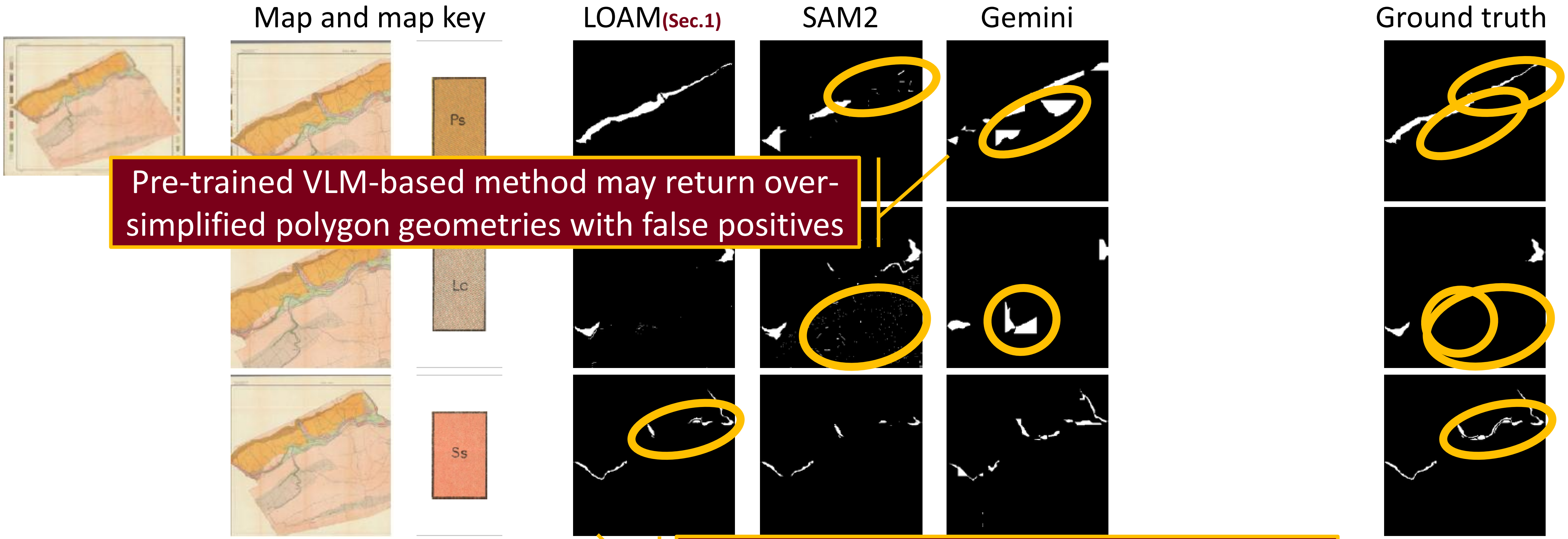
Reconciled polygons



Test-time Adaptive Self Calibration

- Different methods may succeed or fail for particular patterns or cartographic styles
  - Similar to LOAM’s different understanding aspects

Segmentation-based method with entity linking may derive over-segmented polygon instances

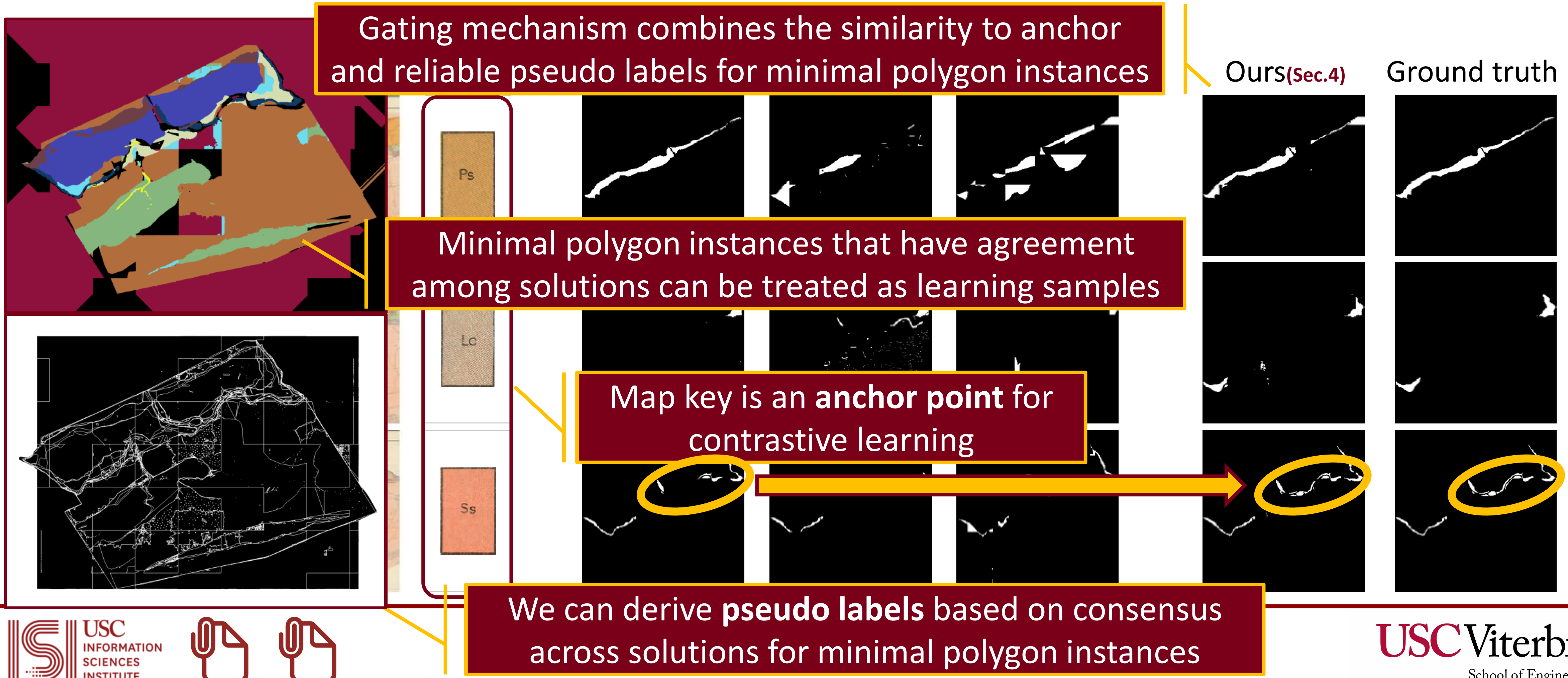


Pre-trained VLM-based method may return over-simplified polygon geometries with false positives

Method trained on USGS dataset may not work well on out-of-domain datasets

Test-time Adaptive Self Calibration (cont.)

- Reconcile aspects that **different machines read a map**
  - But at test time without explicit ground truth label



Dataset

- Select 5 datasets with diverse printing techniques and cartographic styles

Forest Type (FT) | Sanborn (SA) | Soil (SO) | Street Plan (SP) | Water Resource (WR)

Map

Snippet & Map legend

Annotation

Maps have the same or similar style within each map series

Only polygons having corresponding map keys in legend and existing in the map content area are annotated

Dataset (cont.)

- A total of three independent annotators manually label polygon ground truth
  - One Ph.D. candidate (myself) & two retired professors

	Forest Type (FT)	Sanborn (SA)	Soil (SO)	Street Plan (SP)	Water Resource (WR)
Data Source	Academia Sinica	Library of Congress	David Rumsey HMC	HMC of Tokyo, AS	Academia Sinica
Year	**1954 – 1955**	**1920 – 1960**	**1903 – 1908**	**1911 – 1950**	**1921 – 1995**
# of Map	6	6	6	4	4
# of Polygon key	38	27	37	7	16
# of Polygon feature	1,031	402	569	2,091	690
Printing technique	**Chromolithography**	**Lithography**	**Chromolithography**	**Hand-drawn, CMYK**	**Offset printing**
Language	Chinese, English	English	English	Japanese, English	Chinese, Japanese
Geographic coverage	Taiwan	United States	United States	Japan, Taiwan	Taiwan
Annotation time (min.)	1,925	840	1,660	1,405	2,655
Fleiss' Kappa	0.9758	0.9874	0.9950	0.9684	0.9828

More than 140 labor hours for 26 maps

Higher than 0.96 across datasets for inter-annotator agreement

* HMC refers to "Historical Map Collection"; AS refers to "Academia Sinica"
* One annotator has received formal training in GIS-related subjects at the university level
* All annotators are familiar with English, Chinese, and Japanese

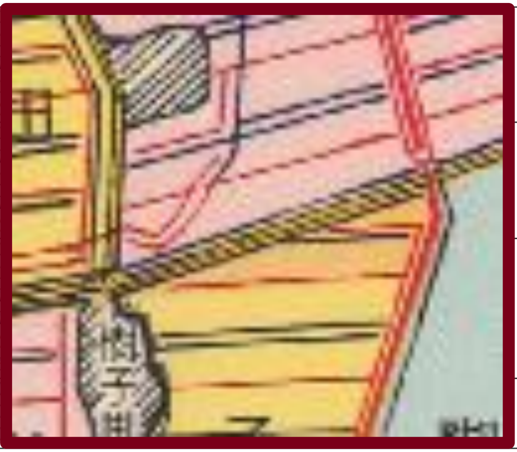
Overall Performance - Generalization

- **G**eneralization via **L**egend-guided **Y**oked **P**olygon extraction in **H**istorical maps (GLYPH)
  - Outperforming state-of-the-art methods (segmentation models, VLMs, etc.)

Dataset / Metric	FT			SA			SO			SP			WR		
Method - Tile Size	MMPQ ↑	F1@8 ↑	NBDR ↓	MMPQ ↑	F1@8 ↑	NBDR ↓	MMPQ ↑	F1@8 ↑	NBDR ↓	MMPQ ↑	F1@8 ↑	NBDR ↓	MMPQ ↑	F1@8 ↑	NBDR ↓
GLYPH-1024 (Ours)	0.90	0.90±0.16	**0.48±1.27**	0.69	0.87±0.16	0.12±0.16	0.81	0.91±0.18	0.36±1.12	0.30	0.59±0.21	0.79±0.82	0.65	0.82±0.13	0.15±0.16
LOAM-1024	0.71	0.83±0.25	0.51±2.26	0.38	0.57±0.28	0.39±0.55	0.76	0.88±0.26	0.38±1.01	0.13	0.32±0.27	1.33±1.53	0.33	0.89±0.08	0.08±0.06
SAM2-4096	0.33	0.53±0.32	2.97±9.89	0.47	0.61±0.35	0.45±0.50	0.41	0.64±0.29	0.72±1.57	0.06	0.39±0.29	1.20±1.12	0.18	[illegible]	[illegible]
SAM2-2048	**0.56**	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
SAM2-1024	0.54	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
SAM2-0512	0.54	0.60±0.26	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
SAM2-0256	0.53	0.60±0.25	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
SAM3-4096	0.00	0.00±0.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
SAM3-2048	0.02	0.01±0.05	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
SAM3-1024	0.02	0.02±0.06	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
SAM3-0512	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	**0.04±0.10**	[illegible]
SAM3-0256	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.01±0.02	[illegible]
Gemini-3-flash-4[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.81±0.50
Gemini-3-flash-2[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.29±0.14	[illegible]
Gemini-3-flash-1[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.47±0.20	0.51±0.43
Gemini-3-flash-0[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	**0.59±0.21**	**0.44±0.41**
Gemini-3-flash-0256	**0.74**	**0.79±0.26**	0.84±1.86	0.41	0.65±0.23	0.34±0.24	**0.54**	**0.67±0.26**	0.79±1.41	**0.10**	0.36±0.22	**0.90±0.89**	0.29	0.58±0.27	0.51±0.43
Gemini-3.1-pro-4096	0.09	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.01	0.06±0.07	1.21±0.64
Gemini-3.1-pro-2048	0.13	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Gemini-3.1-pro-1024	0.49	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Gemini-3.1-pro-0512	0.59	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Gemini-3.1-pro-0256	**0.65**	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Gemini-2.5-pro-4096	0.21	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Gemini-2.5-pro-2048	0.26	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.10	0.23±0.10	0.62±0.46
Gemini-2.5-pro-1024	0.34	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.16	0.35±0.16	0.48±0.35
Gemini-2.5-pro-0512	0.51	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.25	0.49±0.19	**0.34±0.29**
Gemini-2.5-pro-0256	**0.65**	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	**0.27**	**0.54±0.23**	0.55±0.44
GPT-4o-4096	0.11	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.00	0.00±0.00	N.A.
GPT-4o-2048	0.21	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.09	0.09±0.15	1.23±0.67
GPT-4o-1024	0.33	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.00	0.00±0.00	N.A.
GPT-4o-0512	0.34	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.16	0.24±0.18	0.59±0.40
GPT-4o-0256	**0.58**	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	**0.38**	**0.53±0.20**	**0.35±0.35**
GPT-4o-0128	0.17	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.20	0.38±0.25	0.54±0.61
GPT-5.2-pro	0.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.00	0.00±0.00	N.A.
Claude-sonnet-4.5	0.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.00	0.00±0.00	N.A.
Claude-opus-4.6	0.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.00	0.00±0.00	N.A.

Consistently better instance-based accuracy over LOAM, with emphasis on instance segmentation

Has dropped pixel-based accuracy with improved MMPQ for only one out-of-domain dataset



- MMPQ (↑): Instance-based accuracy
  - *Polygon are identified and grouped
- F1@8 (↑): pixel-based accuracy
  - *Polygon boundary are aligned
- NBDR (↓): geometry-based accuracy
  - *Estimated post-correction time

Dataset / Metric	WR		
Method - Tile Size	MMPQ ↑	F1@8 ↑	NBDR ↓
GLYPH-1024 (Ours)	0.65	0.82±0.13	0.15±0.16
LOAM-1024	0.33	0.89±0.08	0.08±0.06

Section 4 →

Section 1 →

- Comparative Methods:
  - * Segmentation model: SAM2, SAM3
  - * VLM: Gemini (3 Flash, 3.1 Pro, 2.5 Pro)
    GPT (4o, 5.2 Pro)
    Claude (Sonnet 4.5, Opus 4.6)

Statistical Analysis

- **G**eneralization via **L**egend-guided **Y**oked **P**olygon extraction in **H**istorical maps (GLYPH)
  - Conduct statistical analysis (one-way ANOVA and Fisher's LSD test) for solutions

Dataset / Metric	FT					SA					SO					SP					WR				
Method - Tile Size	MMPQ	P@8	R@8	F1@8	NBDR	MMPQ	P@8	R@8	F1@8	NBDR	MMPQ	P@8	R@8	F1@8	NBDR	MMPQ	P@8	R@8	F1@8	NBDR	MMPQ	P@8	R@8	F1@8	NBDR
GLYPH-1024 (Ours)	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
LOAM-1024	A	A	B	A	A	C	A	C	C	C	A	A	A	A	A	B	B	A	B	A	B	A	A	A	A
SAM2-4096	D	B	E	D	C	B	B	B	C	B	C	A	D	C	B	A	A	B	A	A	C	B	D	C	B
SAM2-2048	C	B	D	C	C	B	B	B	C	B	B	A	C	B	B	A	A	A	A	A	B	B	C	B	B
SAM2-1024	C	B	D	C	C	B	B	B	C	B	B	B	C	B	B	A	A	A	A	A	D	B	C	B	B
SAM2-0512	C	B	D	C	C	B	B	C	C	B	B	B	C	B	B	A	A	A	A	A	C	B	D	C	B
SAM2-0256	D	B	D	C	C	B	B	B	C	B	B	A	B	B	B	A	A	A	A	A	B	B	B	B	B
SAM3-4096	H	H	I	H	G	F	F	F	G	G	G	G	H	H	G	B	B	B	B	B	E	F	G	F	G
SAM3-2048	H	H	I	H	G	F	F	E	F	F	G	G	H	H	G	A	B	B	B	A	E	F	G	F	G
SAM3-1024	H	H	I	H	G	F	E	F	F	F	G	G	H	H	F	B	B	A	B	A	E	G	G	F	F
SAM3-0512	H	G	H	G	F	F	E	F	F	F	G	G	G	H	E	B	B	A	B	A	E	G	G	F	F
SAM3-0256	H	H	I	H	G	F	E	F	F	F	G	F	H	H	E	B	B	B	B	A	E	G	G	F	F
Gemini-3-flash-4096	E	E	E	E	C	D	C	D	D	C	E	D	E	E	C	B	B	B	B	A	E	E	F	E	D
Gemini-3-flash-2048	E	D	D	D	C	D	C	D	D	C	E	D	E	D	C	B	B	B	B	A	D	E	E	D	C
Gemini-3-flash-1024	C	C	B	C	B	B	B	A	B	B	D	C	C	C	B	A	A	B	A	A	C	D	C	C	C
Gemini-3-flash-0512	C	C	B	B	B	C	C	B	C	C	C	C	C	C	B	A	B	A	B	A	B	C	A	B	C
Gemini-3-flash-0256	B	B	A	B	B	B	C	A	B	C	C	C	B	B	C	A	B	A	B	A	B	C	A	B	C
Gemini-3.1-pro-4096	H	G	H	G	E	F	E	F	F	E	F	F	G	G	D	B	B	C	C	B	E	F	G	F	E
Gemini-3.1-pro-2048	G	E	F	F	C	E	C	E	E	D	F	E	F	F	C	C	B	D	C	B	E	F	F	E	D
Gemini-3.1-pro-1024	D	D	C	D	B	B	B	A	B	B	D	C	D	C	B	A	A	B	A	A	C	D	C	C	C
Gemini-3.1-pro-0512	C	C	C	C	B	C	B	C	C	B	D	C	D	C	B	A	A	B	A	A	B	C	B	B	B
Gemini-3.1-pro-0256	C	B	B	B	B	D	C	C	D	B	C	C	C	C	B	A	A	B	A	A	B	C	A	B	C
Gemini-2.5-pro-4096	G	G	G	G	E	F	F	F	F	E	G	F	G	G	D	B	B	C	C	B	E	F	F	F	D
Gemini-2.5-pro-2048	G	F	F	F	C	F	F	F	F	D	F	E	F	F	C	B	B	C	B	A	E	E	E	E	D
Gemini-2.5-pro-1024	E	E	E	E	C	D	D	C	D	C	E	D	E	D	C	B	B	B	B	A	D	E	D	D	C
Gemini-2.5-pro-0512	D	C	D	D	B	C	C	B	C	C	D	D	D	D	B	B	B	B	B	A	C	D	C	C	C
Gemini-2.5-pro-0256	C	C	C	C	B	C	C	A	C	B	D	C	C	C	B	A	B	A	B	A	C	C	B	B	C
GPT-4o-4096	H	G	H	H	F	F	G	G	G	G	G	F	H	H	F	C	C	D	C	N.A.	E	G	G	F	N.A.
GPT-4o-2048	G	G	H	G	E	F	F	G	G	F	G	F	H	H	D	C	C	D	C	B	E	F	G	F	E
GPT-4o-1024	G	F	F	F	D	F	D	F	F	E	G	F	G	G	D	B	C	C	C	B	E	G	G	F	N.A.
GPT-4o-0512	G	E	G	F	C	F	C	F	F	D	G	E	G	G	C	B	B	C	C	A	E	D	F	E	D
GPT-4o-0256	D	D	D	D	B	E	C	E	E	C	F	D	F	F	C	B	B	B	B	A	C	C	C	C	C
GPT-4o-0128	F	E	E	E	C	E	D	E	E	C	F	E	F	F	C	B	A	B	A	A	D	D	C	D	C
GPT-5.2-pro	H	H	I	H	N.A.	F	G	G	G	N.A.	G	G	H	H	N.A.	C	C	D	C	N.A.	E	G	G	F	N.A.
Claude-sonnet-4.5	H	H	I	H	N.A.	F	G	G	G	N.A.	G	G	H	H	N.A.	C	C	D	C	N.A.	E	G	G	F	N.A.
Claude-opus-4.6	H	H	I	H	N.A.	F	G	G	G	N.A.	G	G	H	H	N.A.	C	C	D	C	N.A.	E	G	G	F	N.A.
ANOVA p-value	e_2^{-242}	e_9^{-227}	$\to 0$	e_3^{-291}	e_1^{-194}	e_5^{-157}	e_8^{-152}	e_3^{-198}	e_6^{-190}	e_4^{-232}	e_1^{-252}	e_9^{-214}	$\to 0$	e_6^{-306}	e_4^{-171}	e_8^{-7}	e_1^{-9}	e_3^{-29}	e_4^{-14}	e_1^{-16}	e_3^{-76}	e_2^{-121}	e_9^{-145}	e_4^{-127}	e_4^{-143}
LSD (α=0.05)	0.08	0.11	0.09	0.09	0.77	0.10	0.12	0.11	0.11	0.51	0.08	0.11	0.08	0.09	0.71	0.17	0.22	0.25	0.22	1.47	0.11	0.13	0.12	0.12	0.65

- MMPQ: Instance-based accuracy
- P@8: pixel-based precision
- R@8: pixel-based recall
- F1@8: pixel-based accuracy
- NBDR: geometry-based accuracy

- Fisher's least significant difference (LSD) test

  *Cluster methods into groups (tiers) based on statistically significant difference

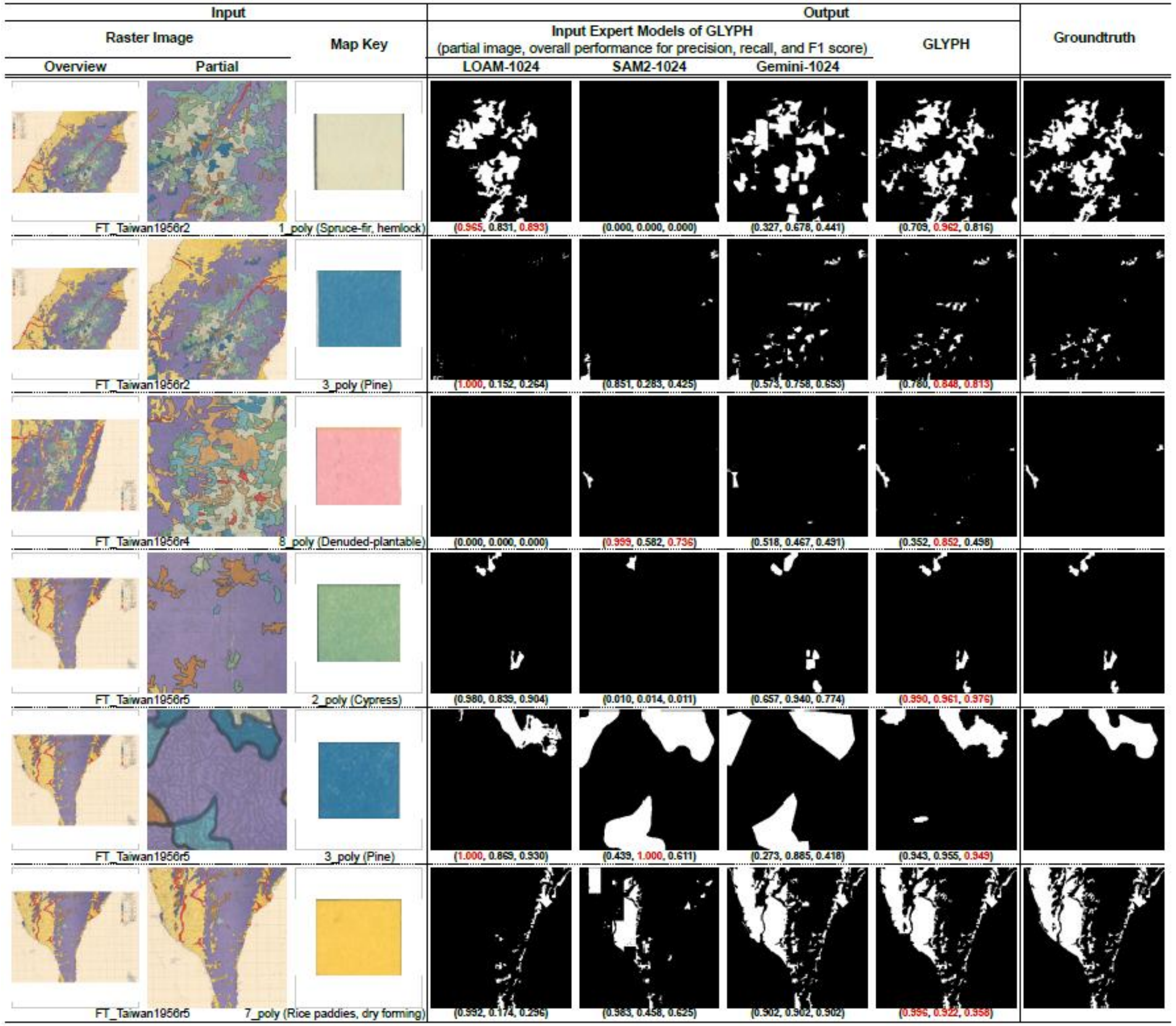
  *Methods of the same letter indicate they have statistically similar performance, "A" is the best tier

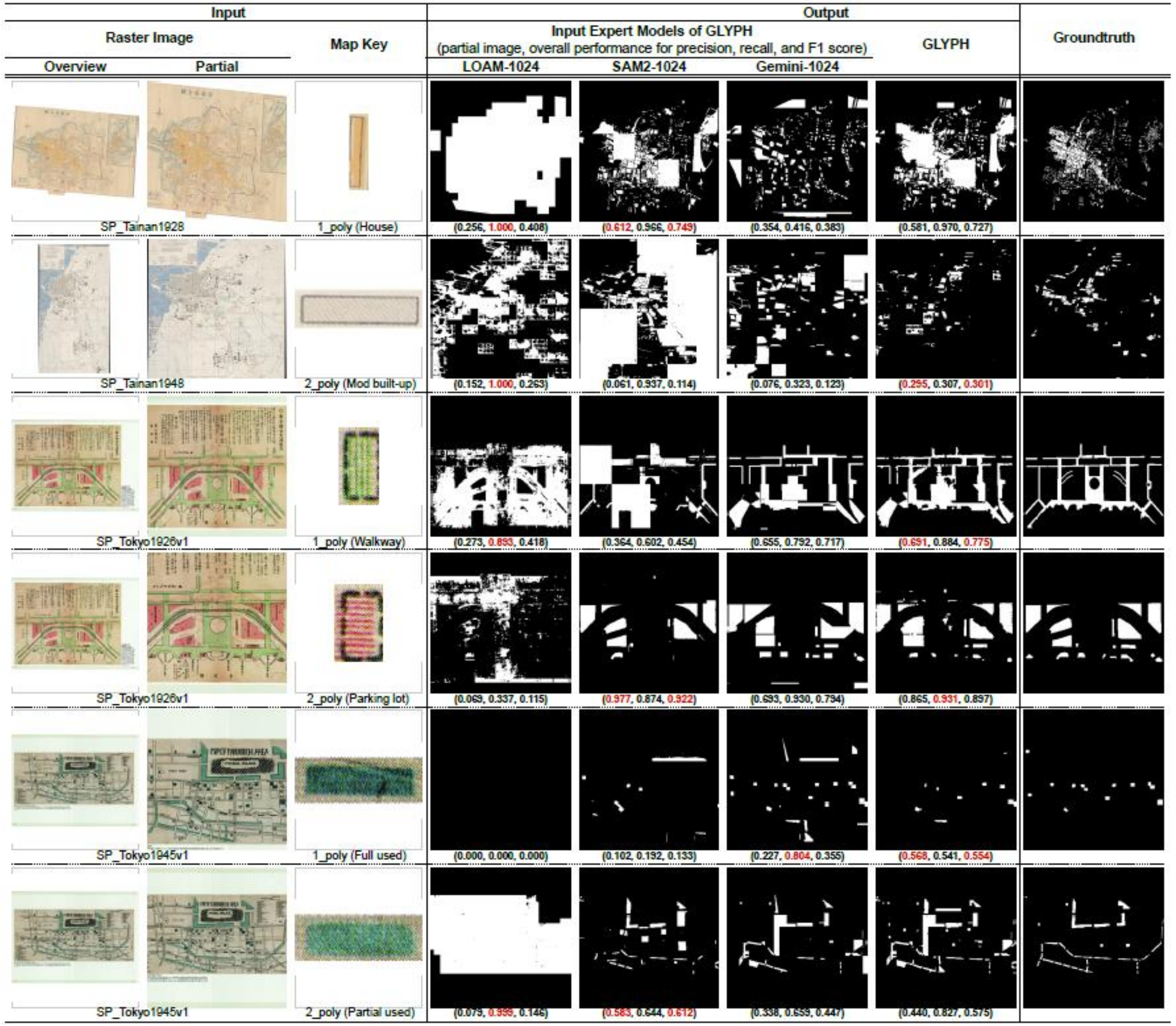
Methods' solutions have statistically significant difference for each metric, GLYPH is the only method to achieve the best tier ("A") across all metrics

Ablation

Pairwise

Qualitative Case Study





More!

Effectively and adaptively leverage solutions from different methods at test time for out-of-domain test maps (no training data)

Related Work - Generalization

- **Generalizing Polygon Digitization**
  - Synthetic Map Generation (Li 2019, Li 2020, Lopez-Rauhut 2025, Arzoumanidis 2025)
    - Apply GAN to simulate particular map styles
    - Do not explicitly address the map keys to support polygon digitization
  - Segmentation-based Extraction from Maps (Xia 2024, Xia 2025)
    - Apply SAM or SAM2 with YOLO to extract particular polygon features
    - Require a tailored model with at least a few training data for each type of polygon feature
  - Segmentation with Concepts from Images (Carion 2025)
    - SAM3 can directly incorporate positive and negative image exemplars for instance segmentation
    - Still seems to struggle with polygon digitization from historical maps
  - Visual-language Model (VLM) (Janowicz 2025, Li 2025, Xing 2025, Pyo 2026)
    - Have shown promising results on reasoning geospatial information
  - Test-time Adaptation (TTA) (Lei 2025, Wang 2025)
    - Combine TTA with MoE to identify and adapt domain shift with pseudo-label guidance

Takeaway - Generalization

- GLYPH exploits polygon metadata to generalize digitization across diverse map styles
  - Test-time adaptive self-calibration framework
  - Supporting polygon digitization for out-of-domain datasets
  - No annotation needed of maps in the targeted cartographic style
- Limitation
  - Limited to polygon features
  - Requires better estimation for post-correction manual effort

Outline

- Introduction and Overview
- Our Approach
  - Exploiting Polygon Metadata to Digitize Polygonal Features
  - Exploiting Polygon Metadata to Recolor Historical Maps
  - Exploiting Polygon Metadata to Colorize Draft Maps
  - Exploiting Polygon Metadata to Generalize Digitization across Styles
- **Conclusion and Future Work**

Conclusion

- Presented my dissertation, with 4 contributions surrounding polygon metadata
  - Digitize polygonal features from historical maps
    - A metadata-driven core for **in-domain polygon digitization** to handle arbitrary unseen polygon map keys with known cartographic styles
  - Recolor historical maps to support digitization
    - A **semantic restoration means** that regularize significantly noisy maps
  - Colorize draft maps to support digitization
    - A **semantic restoration means** that regularize incomplete draft maps
  - Generalize digitization across map styles
    - An extension to **out-of-domain polygon digitation** for unseen cartographic styles

Future Work

- Extend to other feature types
  - Point features
- Enhance entity linking
  - Polygon features
  - Text features
- Enhance geometric constraints
  - Contextual information
  - Polygon topology
- Incorporate various downstream tasks
  - Comprehensive estimation of applicability
  - Synthetic data generation

Thank You!

Acknowledgement

- Craig A. Knoblock (advisor)
- Cyrus Shahabi, Yolanda Gil, Yao-Yi Chiang, John P. Wilson (committee)
- Binh Vu, Basel Shbita, Minh Pham (USC ISI >> Meta/ IBM Research/ Google)
- Filip Ilievski, Min-Hsueh Chiu (USC ISI >> VU Amsterdam/ Berkeley Lab)
- Zekun Li, Yijun Lin, Jina Kim, Leeje Jang, Sofia Kirsanova, Theresa Chen, Min Namgung, Jiyoon Pyo, and many (UMN)
- Weiwei Duan, David Abbondanzio, Michael P. Gerlek, Steven N. Minton (Inferlink)
- T.K. Satish Kumar, Han Zhang, Ting-Rui Chiang, Shih-Yao Huang (USC)
- Hsun-Ping Hsieh and his team (NCKU)
- Jer-Yann Lin, Ding-Ying Guo (NUTN >> retired)
- Ta-Chien Chan and his team (Academia Sinica)
- Saty Raghavachary (USC) and Ning Wang (USC ICT)
- Graham W. Lederer (USGS)
- Colleagues, collaborators, and friends from USC, ISI, UMN, NCKU, and other places

Fellowships from Ministry of Education of Taiwan (**4-year**) and USC Viterbi School of Engineering (**3-year top off**).

Publication During PhD

- Cross-domain Polygon Extraction from Historical Maps via Legend-guided Semantic Fusion (*under review at KDD*)
  - **Fandel Lin**, Zekun Li, Yao-Yi Chiang, and Craig A. Knoblock
- Exploiting Polygon Metadata to Colorize Draft Maps (*SIGSPATIAL, 2025*) **Best Short Paper Award**
  - **Fandel Lin**, Craig A. Knoblock, Binh Vu, Basel Shbita, and Yao-Yi Chiang
- Exploiting Polygon Metadata to Recolor Historical Maps (*SIGSPATIAL, 2025*)
  - **Fandel Lin**, Craig A. Knoblock, Binh Vu, and Yao-Yi Chiang
- Exploiting Polygon Metadata to Understand Raster Maps - Accurate Polygonal Feature Extraction (*SIGSPATIAL, 2023*)
  - **Fandel Lin**, Craig A. Knoblock, Basel Shbita, Binh Vu, Zekun Li, and Yao-Yi Chiang

---

- An Integrated Approach to Multi-Agent Scheduling with Bounded Objectives in an Urban Setting (*ACM TSAS, 2025*)
  - **Fandel Lin**, Han Zhang, T. K. Satish Kumar, and Craig A. Knoblock
- A Hierarchical Voronoi Approach to Deploying New Charging Stations in an Existing Network (*SIGSPATIAL, 2024*) **GISCUP Finalist (Top 5)**
  - **Fandel Lin**, Craig A. Knoblock, and Binh Vu
- An Integrated Approach to Multi-Agent Scheduling with Bounded Objectives (*SIGSPATIAL, 2024*)
  - **Fandel Lin**, Han Zhang, T. K. Satish Kumar, and Craig A. Knoblock
- A Hierarchy-Aware Approach to Cross-Region Spatial-Temporal Inference of Unarchived Event in Urban Mobility Infrastructure (*DASFAA, 2024*)
  - **Fandel Lin** and Hsun-Ping Hsieh
- Indirect Cooperation in Distributed Stationary-Resource Searching with Predefined Destinations (*SIGSPATIAL, 2023*)
  - **Fandel Lin** and Craig A. Knoblock
- A Machine-Learning Approach to Recognizing Teaching Beliefs in Narrative Stories of Outstanding Professors (*AIED, 2023*)
  - **Fandel Lin**, Ding-Ying Guo, and Jer-Yann Lin
- Exploiting Network Structure in Multi-Criteria Distributed and Competitive Stationary-Resource Searching (*ACM TSAS, 2023*)
  - **Fandel Lin** and Hsun-Ping Hsieh

Publication During PhD (cont.)

- An Ensemble Approach to Text-Based Georeferencing of Historical Maps (*under review at KDD*)
  - Zekun Li, **Fandel Lin**, Yijun Lin, Yao-Yi Chiang, and Craig A Knoblock
- Enhancing POI Recommendation through Global Graph Disentanglement with POI Weighted Module (*ACM TIST, 2026*)
  - Pei-Xuan Li, Cheng-Ru Chou, Wei-Yun Liang, **Fandel Lin**, Hsun-Ping Hsieh
- Predicting Origin-Destination Traffic with Advanced Spatio-Temporal Networks (*Eng. Proc., 2026*)
  - Bo-Yan Zeng, Yen-An Chen, Shih-Hung Yang, **Fandel Lin**, Donna Hsu, and Hsun-Ping Hsieh
- MoVER: Modeling User Heterogeneity with Enriched Trajectory Representations for Human Mobility Prediction (*SIGSPATIAL, 2025*) **GISCUP 7th Place**
  - Yijun Lin, **Fandel Lin**, Jina Kim, and Yao-Yi Chiang
- DIGMAPPER: A Modular System for Automated Geologic Map Digitization (*SIGSPATIAL, 2025*)
  - Weiwei Duan, Yao-Yi Chiang, Theresa Chen, Michael P. Gerlek, Leeje Jang, Sofia Kirsanova, Craig A. Knoblock, **Fandel Lin**, Yijun Lin, Zekun Li, and Steven N. Minton
- A Domain-Independent Approach for Semantic Table Interpretation (*ISWC, 2025*)
  - Binh Vu, Craig A. Knoblock, and **Fandel Lin**
- FOG: Feature-Oriented Graph Neural Networks for Tabular Data (*PAKDD, 2025*)
  - Teng-Yuan Tsou, Pei-Xuan Li, **Fandel Lin**, and Hsun-Ping Hsieh
- Embedding Spatial and Semantic Contexts for Geo-Entity Typing in Smart City Applications (*WebAndTheCity @ WWW, 2025*)
  - Basel Shbita, Binh Vu, **Fandel Lin**, and Craig A. Knoblock
- Results of GRAMS+ at SemTab 2024 (*SemTab @ ISWC, 2024*)
  - Binh Vu, Craig A. Knoblock, and **Fandel Lin**
- Exploiting Distant Supervision to Learn Semantic Descriptions of Tables with Overlapping Data (*ISWC, 2024*)
  - Binh Vu, Craig A. Knoblock, Basel Shbita, and **Fandel Lin**
- Constructing a Knowledge Graph of Historical Mining Data (*GeoLD @ ESWC, 2024*)
  - Basel Shbita, Namrata Sharma, Binh Vu, **Fandel Lin**, and Craig A. Knoblock

1st Place, Map Feature Extraction Challenge, AI for Critical Mineral Assessment Competition (*DARPA, 2022*)
  - Weiwei Duan, Zekun Li, **Fandel Lin**, Yijun Lin, Tanisha Shrotriya, Craig A. Knoblock, and Yao-Yi Chiang

Appendix for Q&A

Appendix for Introduction and Overview

Preliminary: Polygon Metadata

- Metadata
  - We take the definition: *data that provides information about other data*

- Polygon Metadata
  - Any data that is on the map, describing or about the polygon feature, but cannot be the polygon (a set of line segments) itself
  - The map legend has a set of map keys that describe the visual representation (e.g., colors and text patterns) of the corresponding polygon features on the map

Motivating Example: Critical Mineral Assessment (CMA)

- **Critical minerals** are crucial to daily life and national security

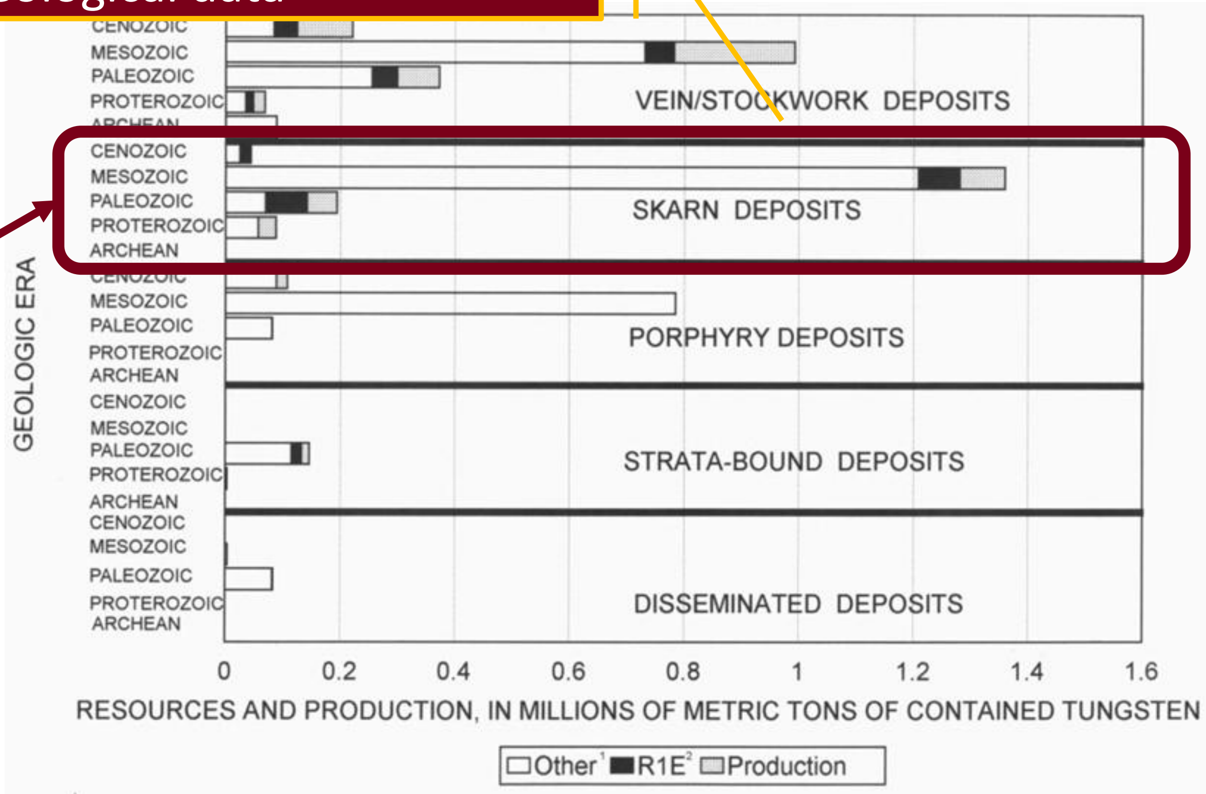
Find potential sites of undiscovered critical mineral deposits relies on accurate geological data

Scheelite
(ore)

Skarn Deposits
(rock)

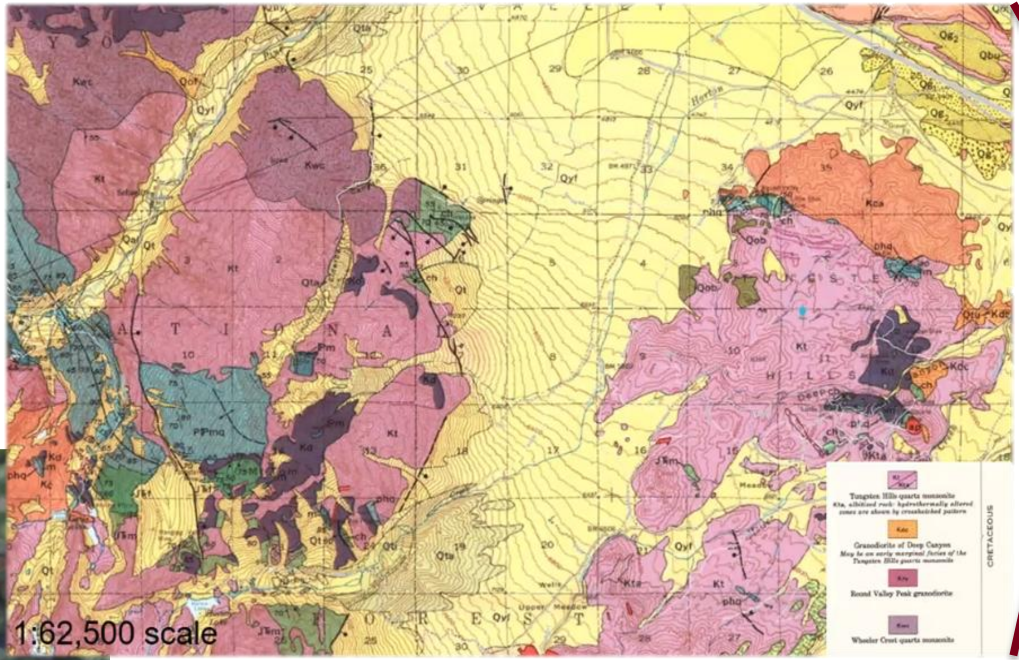
Tungsten
(element)

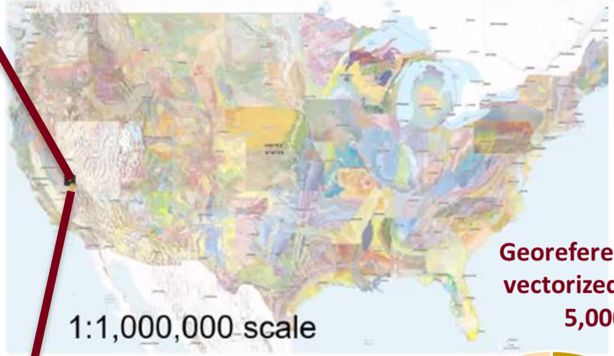
Tungsten Carbide
(product)

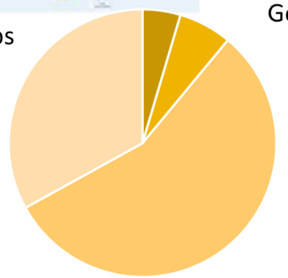


Motivation Example: CMA (cont.)

- United States Geological Survey (USGS) has about 100,000 geologic maps
  - They are the **only source** that records detailed information about geological features







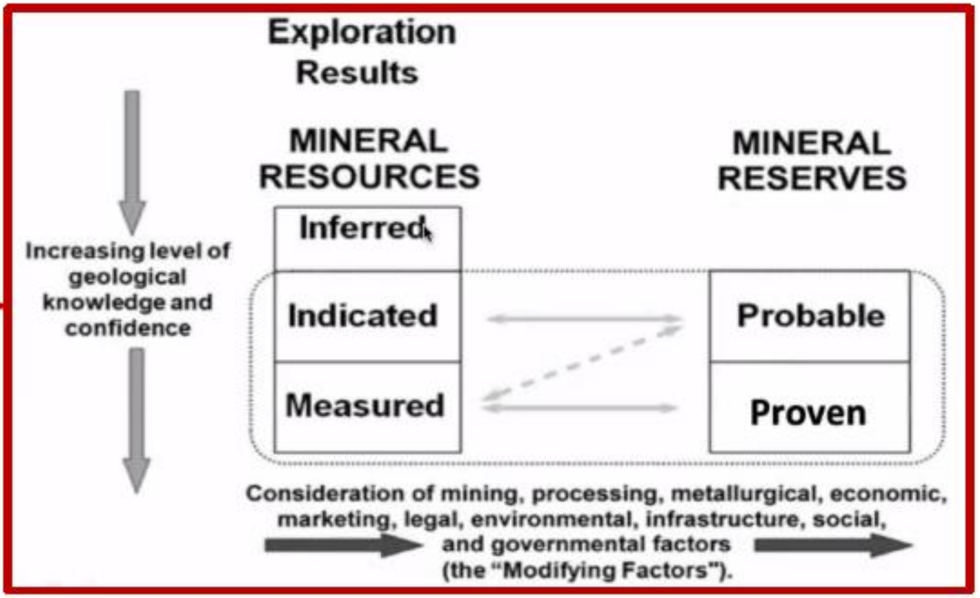
Most of them exist as scanned images

Image Source:
(1) Graham Lederer, USGS media;
(2) DARPA AI for Critical Mineral Assessment Competition. https://criticalminerals.darpa.mil/The-Competition

Tungsten Skarn

natural occurrence of minerals (geological formations)

naturally occurring solid material
usually mixtures of minerals and other materials
(bauxite = ore of aluminum; hematite = ore of iron)



COMMODITY	AS-REPORTED F(	AS-REPORTED A|	CONVERSION FA	CONVERTED FORM
Aluminum	Al		1.8895	Al2O3
Antimony	Sb		1.1971	Sb2O3
Barium	Ba		1.6995	BaSo4
Borates	B	Boron	3.2198	B2O3
Borates	H3BO3	Boric Acid	0.5629	B2O3
Cesium	Cs2O	Cesium Oxide, C	0.9432	Cs

material or primary agricultural product

Image Source: Sources are as described in page 2; notes were edited by the presenter and Basel Shbita.

Background: Features in Geologic Maps

- The geological features in USGS geologic maps have three representations
  - Each geological feature is identified by a **map key** specifying the feature's symbol and name

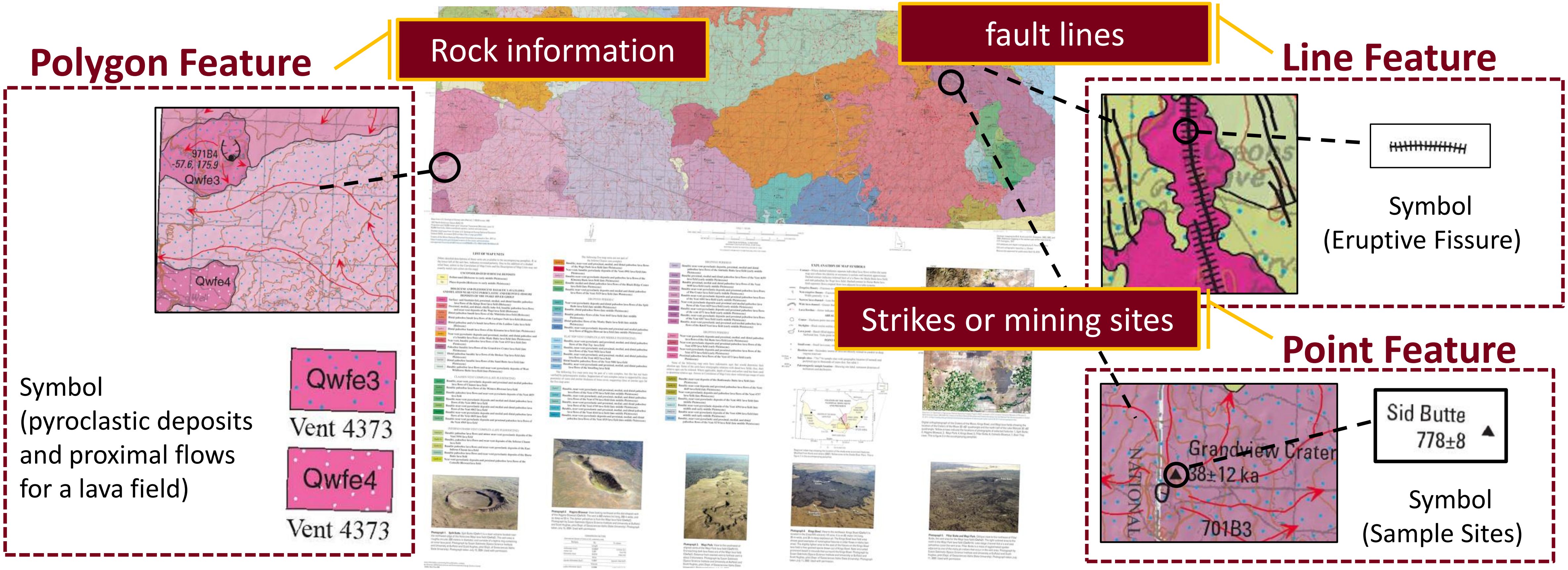
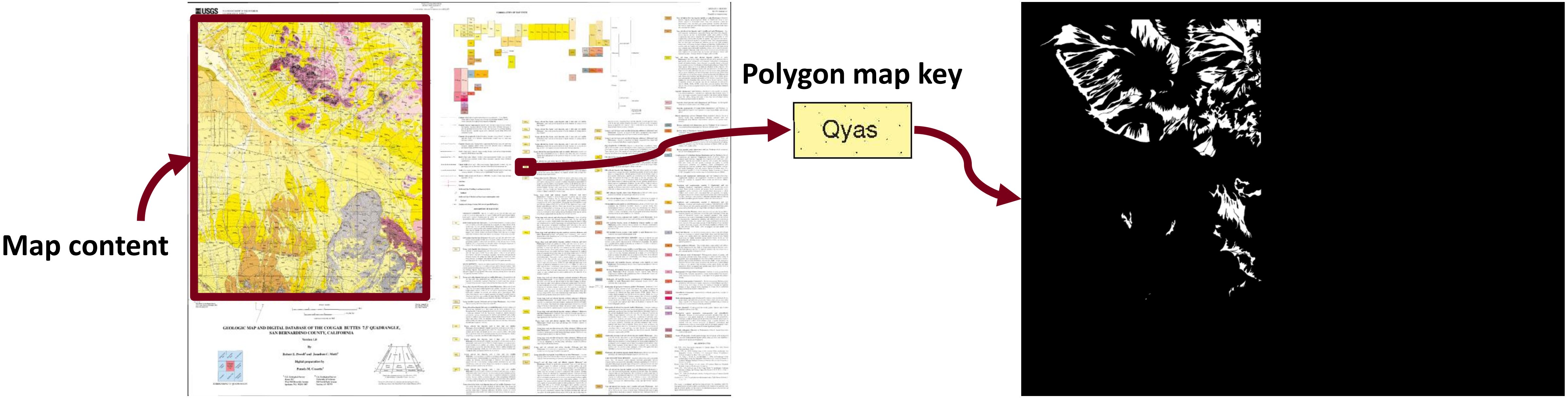


Image Source: ID_LakeWalcott, USGS; notes were edited by the presenter, Zekun Li, and Weiwei Duan

Problem: Overview

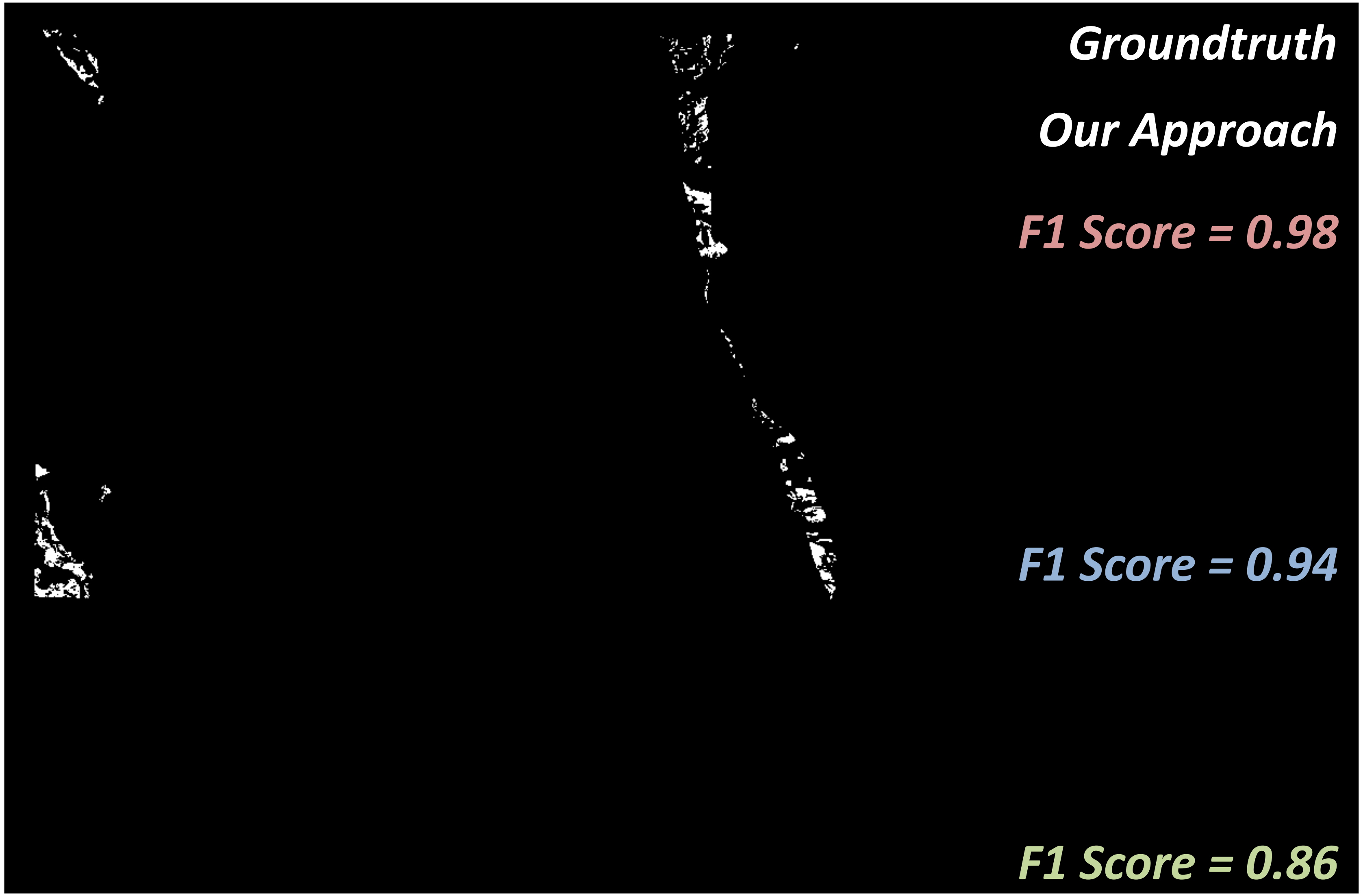
- Given historical maps with identified map content area and polygon map keys
  - Automatically interpret and digitize polygonal features
  - Convert raster images into analysis-ready formats



Historical map in raster format

Digitized polygon feature

Problem: Digitizing Polygonal Feature



XgdB

Boulder Creek Granodiorite (Early Proterozoic)

- mafic

PIPf

Pf Fountain Formation (Lower Permian and Pennsylvanian)

- quartz

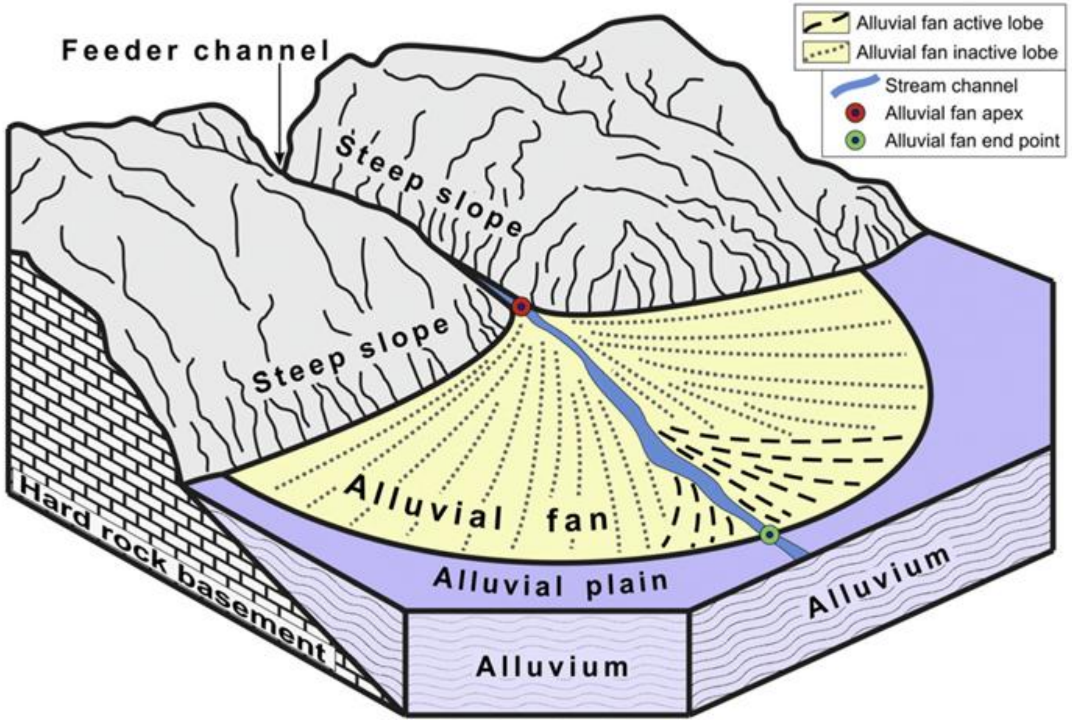
Kp

Pierre Shale (Upper Cretaceous)

- petroleum deposits

Challenge: Alluvial Deposit Category

- Polygon features of the same category are often in similar or similar color(s) in geologic maps
  - Alluvial fan feeder wash deposit
  - Alluvial fan deposit



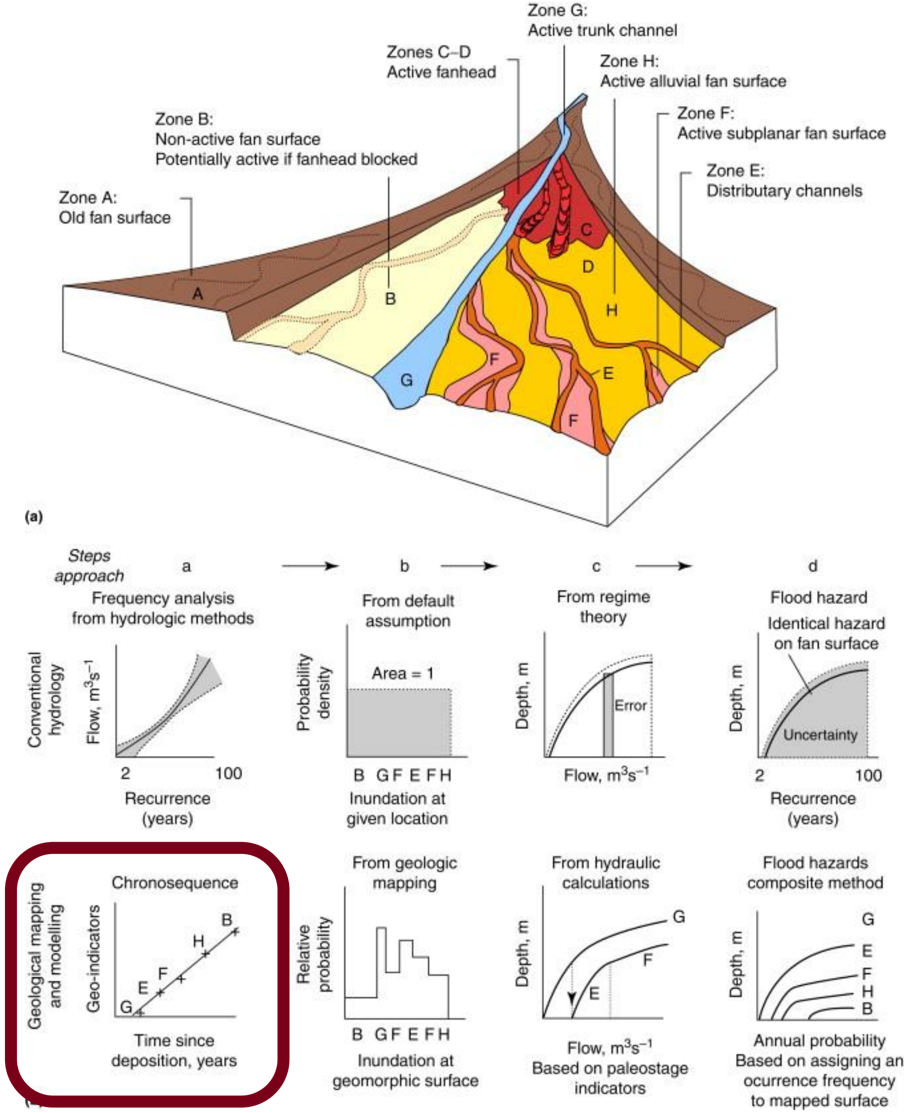
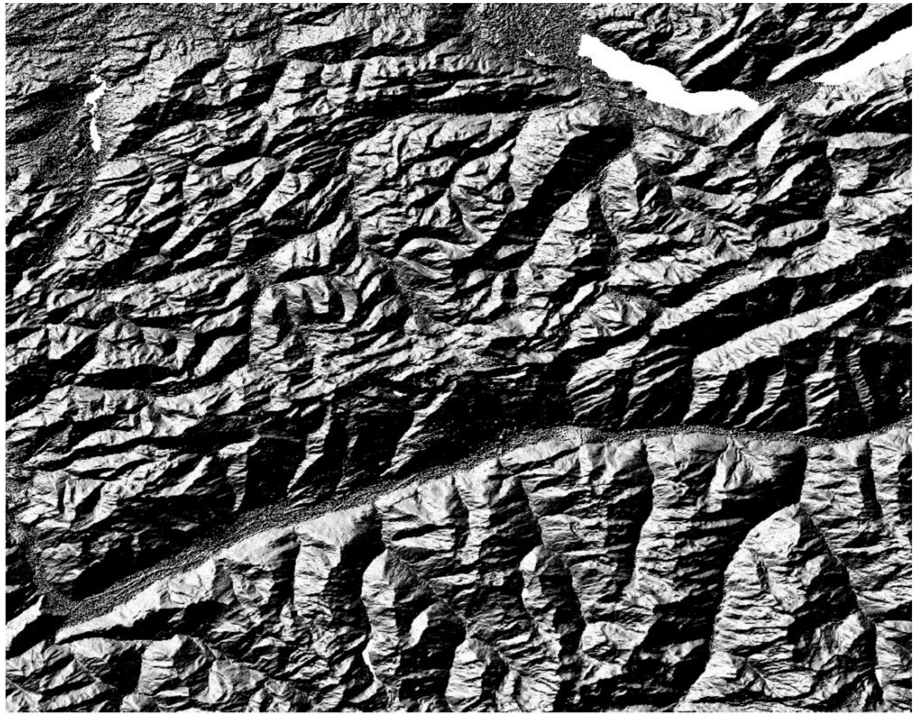
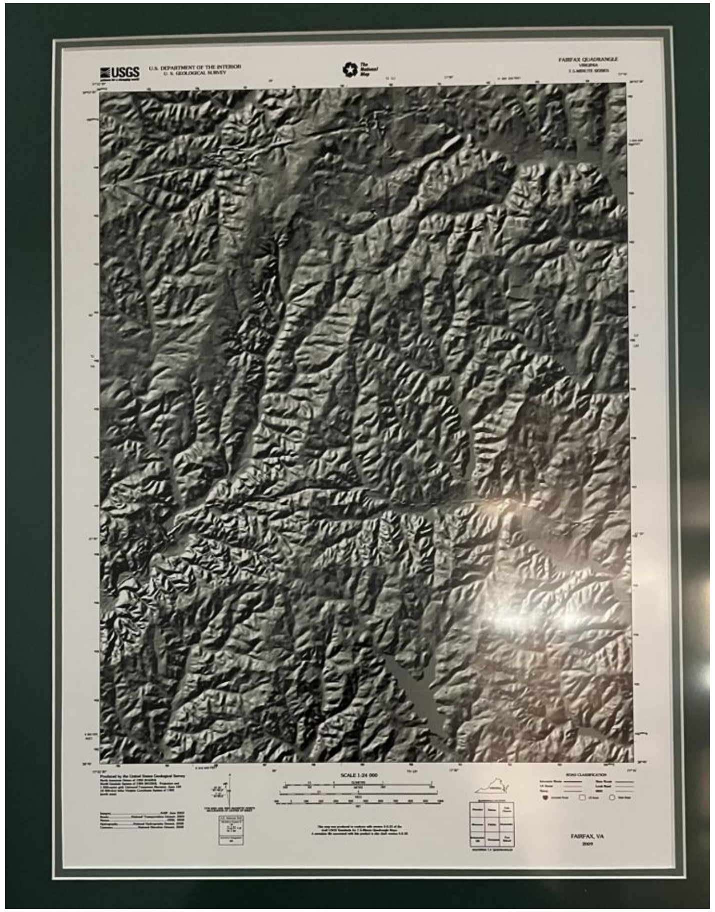


Image Source:
(1) Okanagan Alluvial Fan Hydrology: A Primer. Okanagan Water Stewardship Council, 2023;
(2) Geomorphology of Human Disturbances, Climate Change, and Natural Hazards. Treatise on Geomorphology, 2013.

Challenge: Shaded Relief Map

- Shaded relief displays the topography in a natural, aesthetic, and intuitive manner
  - Computer generated image (Already vectorized)
  - Cloud-free black and white aerial photo taken at a low sun angle
  - Data includes aspect, contour, elevation, hill shade, and slope

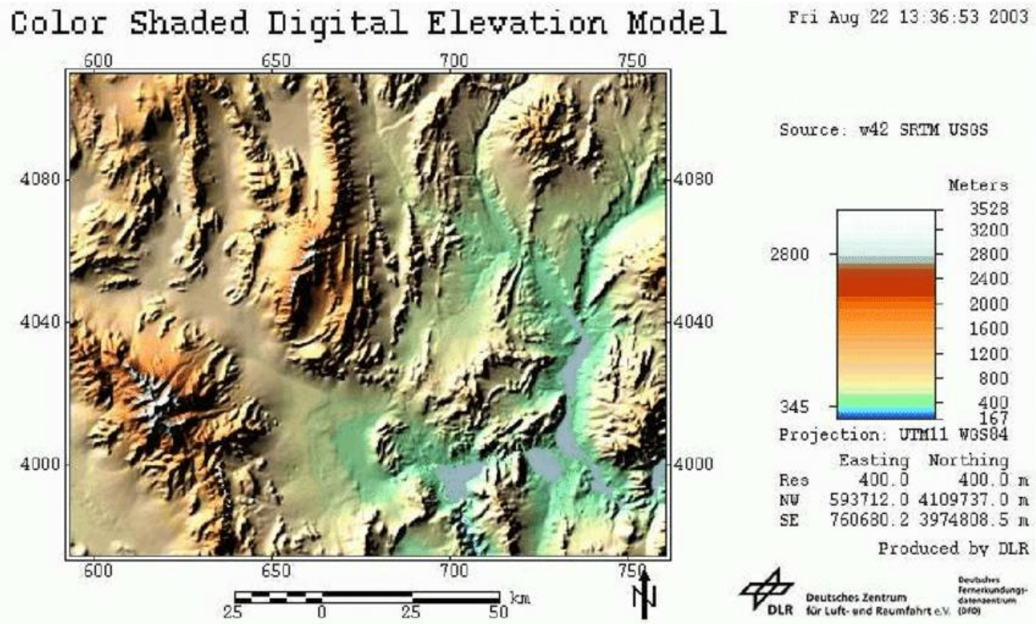




Reference: U.S. Geological Survey National Center.
Image Source: (1) https://blog.datawrapper.de/shaded-relief-with-gdal-python; (2) U.S. Geological Survey National Center.

Challenge: Digital Elevation Model

- Digital elevation model (DEM) / digital surface model (DSM) displays the terrain or overlaying objects
  - Computer generated image (Already vectorized)



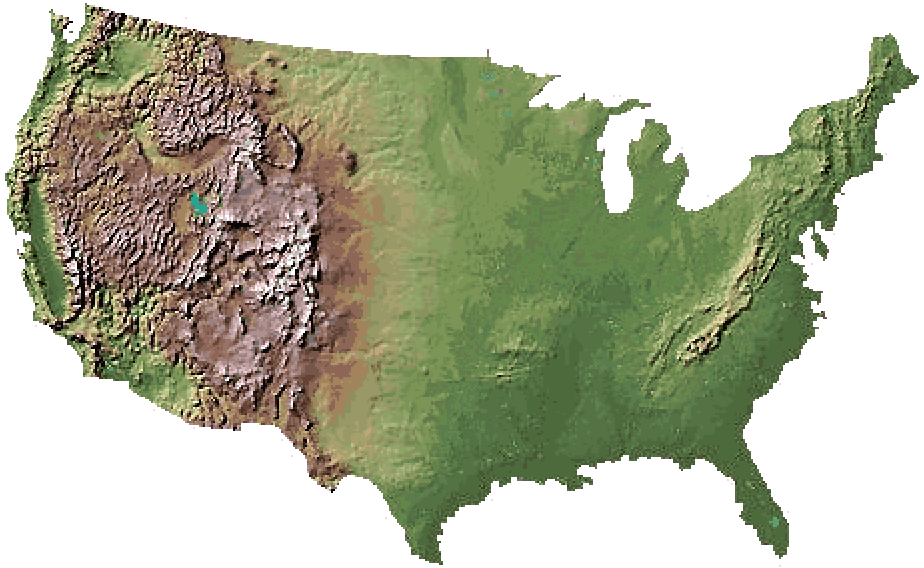
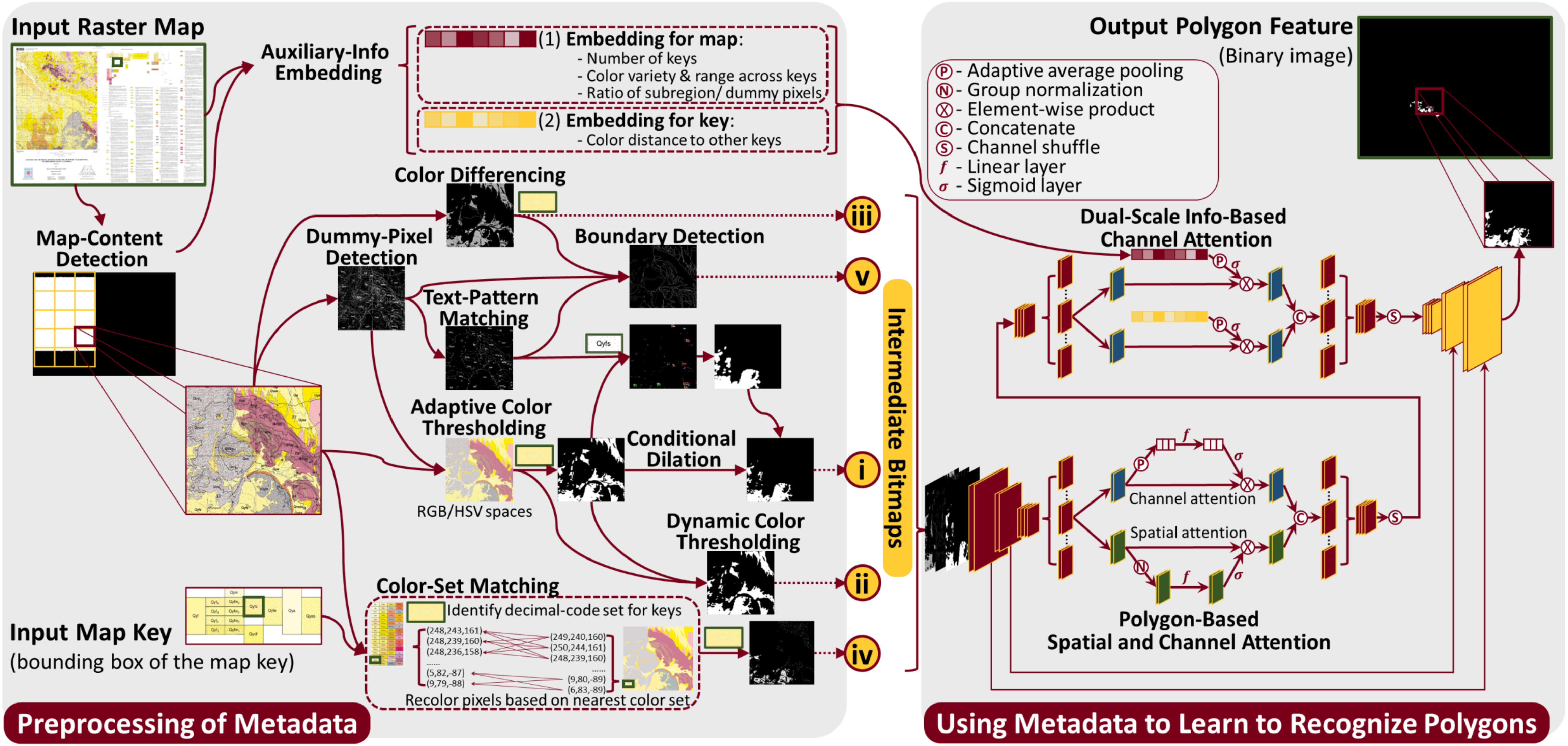


Image Source: (1) Radar Interferometry With Public Domain Tools; (2) USGS.

Appendix for Digitizing Polygonal Features

Methodology: Approach Overview

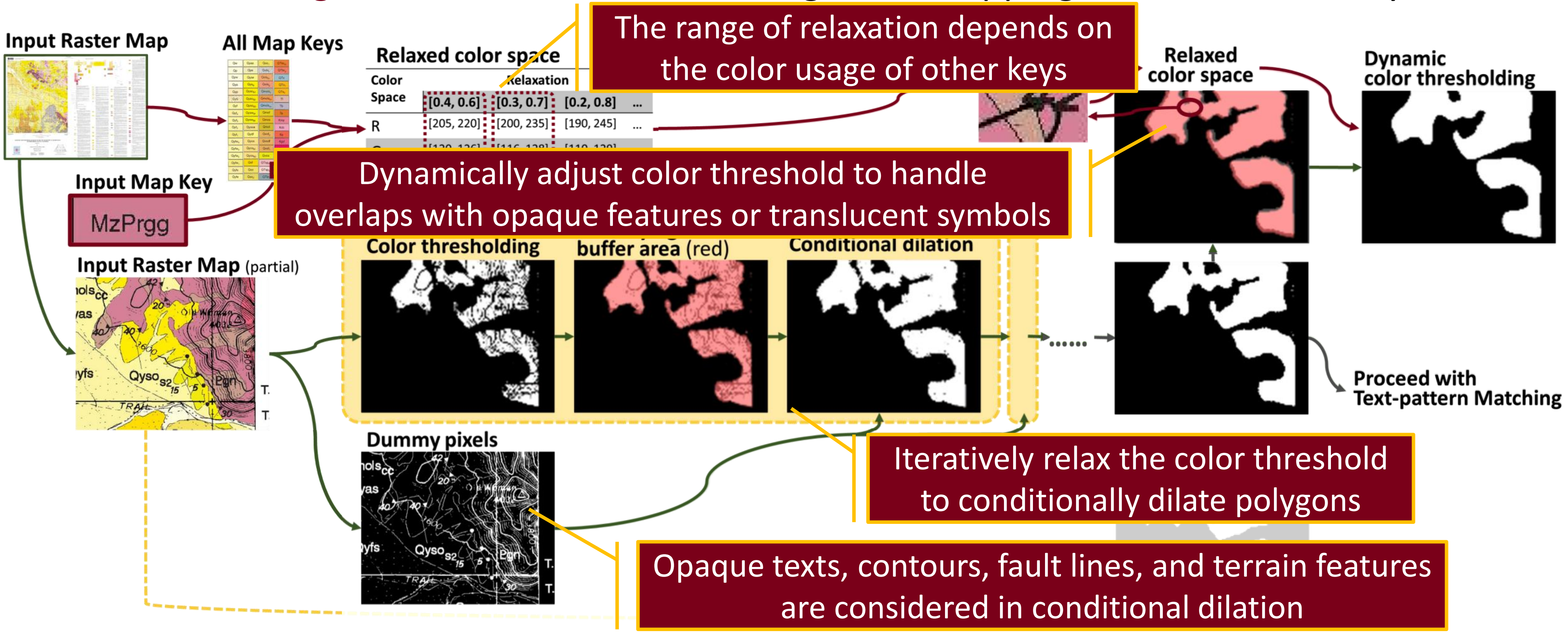


Methodology: Intuition for Polygon-Feature Preprocessing

How human read the historical maps	How our approach imitate
comprehending the colors, texts, or symbols used by map keys in the map legend	assuming that each pixel in raster maps can only belong to one map key, finding the areas that have the most-similar colors to our targeted map key
identifying the region of interest in the map	using boundaries to differentiate the polygon features if nearby polygons have the same colors but belong to other map keys
finding the areas with colors similar to the color of our targeted map key	directly finding the areas that have colors similar to the color of our targeted map key
distinguishing symbols and textures overlapped with or nearby the found areas	identifying the polygon feature based on the color of the map key, and including surrounding areas that have different colors due to translucent symbols or textures
using texts or boundaries to further differentiate the polygon features if multiple keys have the same colors	identifying the polygon feature based on the color of the map key, and excluding polygons that are labeled with texts different from the targeted one

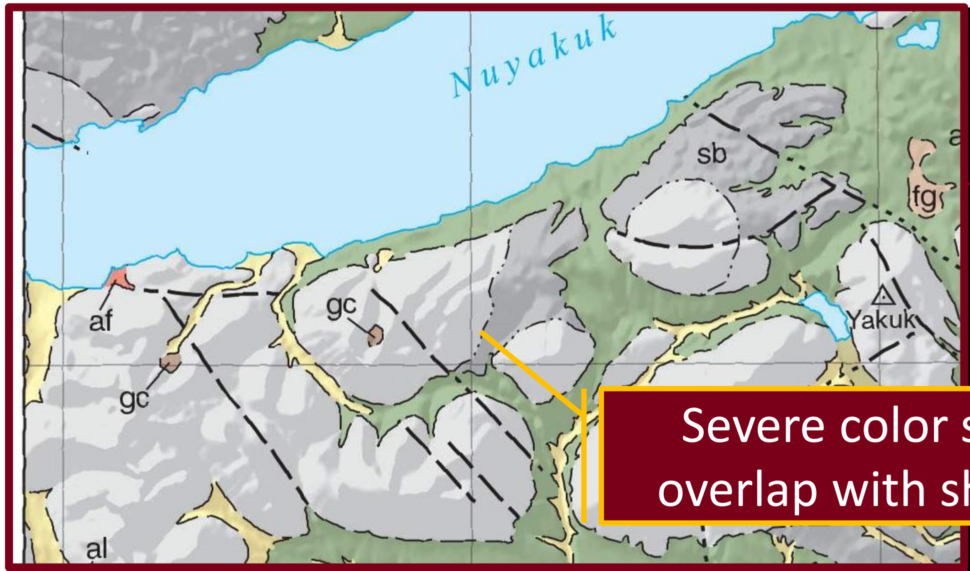
Methodology: Adaptive Color Thresholding

- Overcome slight color shift due to scanning or overlapping with translucent symbols

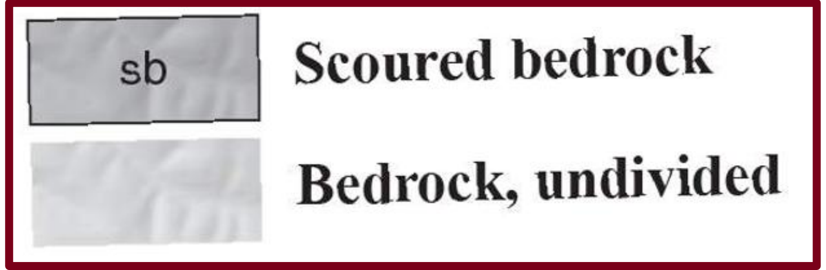


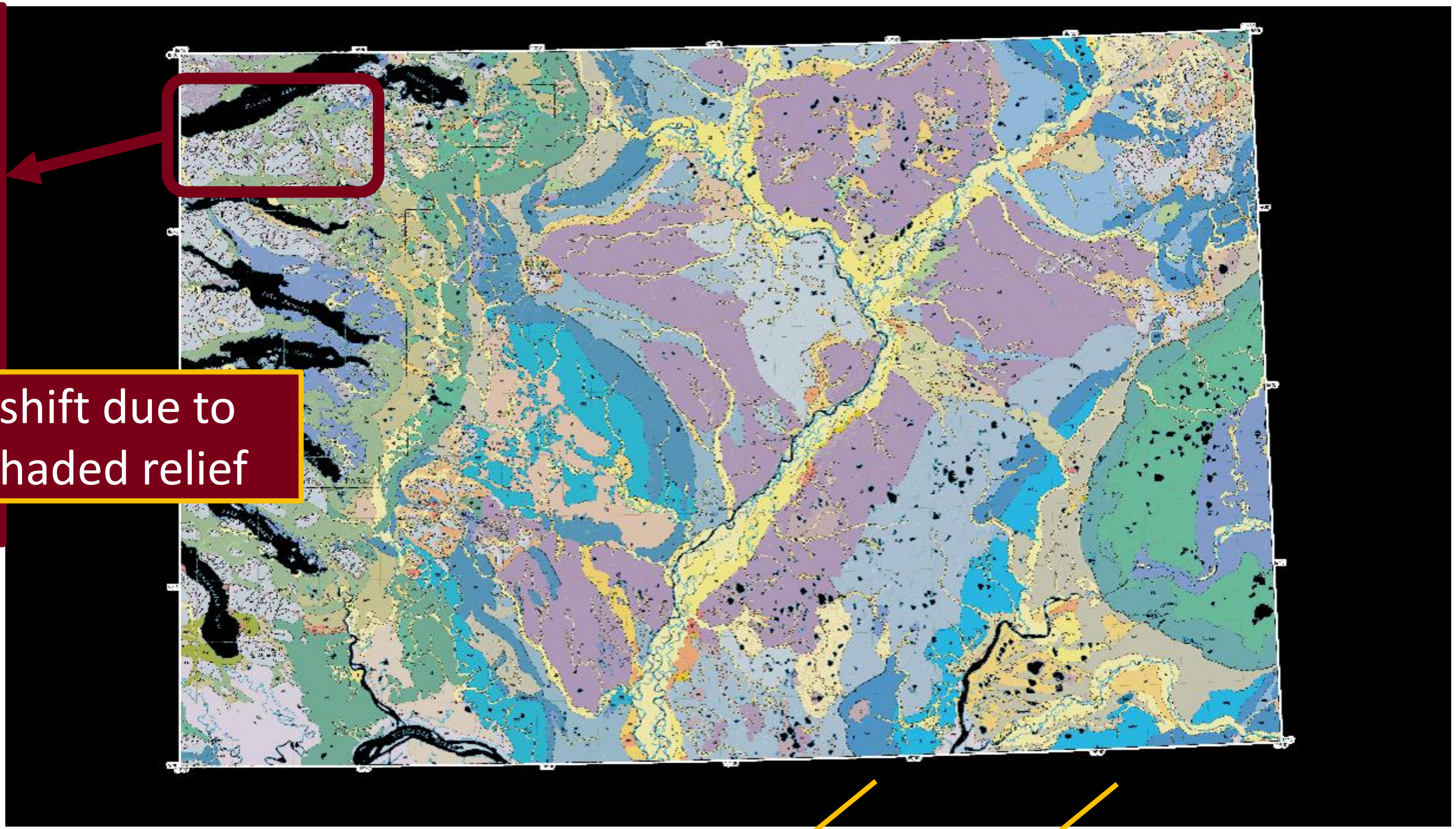
Color-set Matching

- Polygon topology
  - A pixel in the map content can only belong to one map key



Severe color shift due to overlap with shaded relief





The re-c map ke

The preliminary output of our color-set matching

Methodology: Adaptive Color Thresholding (cont.)

- We iteratively relax the tolerance in the RGB and HSV thresholds for each key while preventing overlapping with the thresholds of other keys

Current upper bound

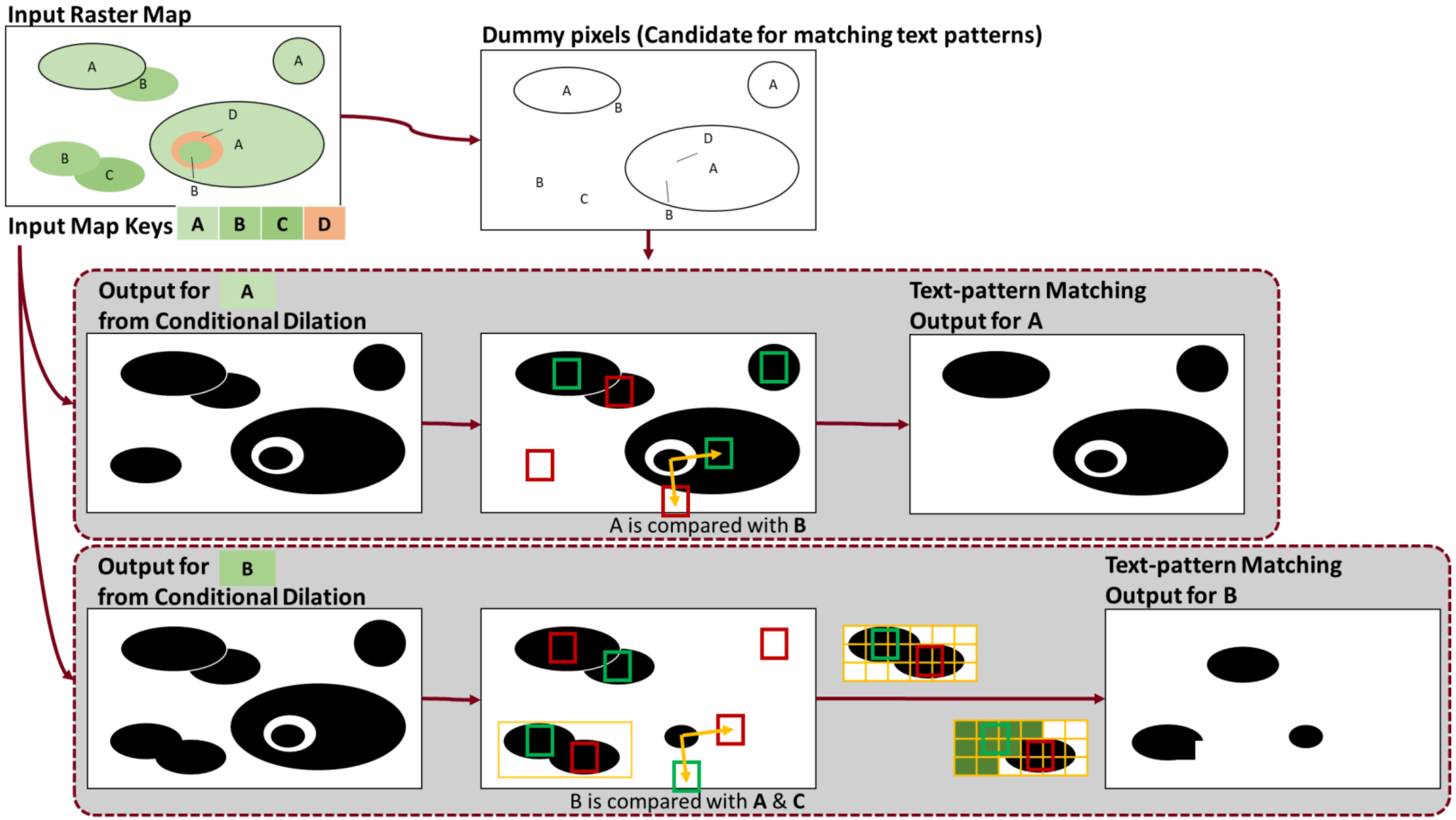
The minimum lower bound threshold for all keys that are larger than the current upper bound

$$h^{up}_{adapt} = h^{up}_{th} + min(\alpha, \beta(h'^{lw}_{th} - h^{up}_{th}))$$

$$h^{lw}_{adapt} = h^{lw}_{th} - min(\alpha, \beta(h^{lw}_{th} - h'^{up}_{th}))$$

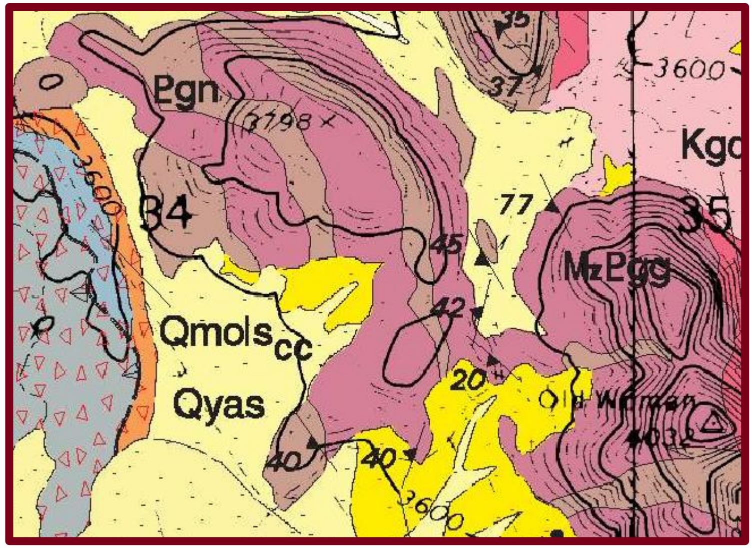
Methodology: Text-pattern Matching

- Exclude connected components close to a text pattern different from the target one
  - Inconsistency in text representation between map key and map content



From map key:

In map area:



Methodology: Color-set Matching

- Considering the difference between two color spaces provides a reference when all keys in a raster map have similar values in two of the three spaces

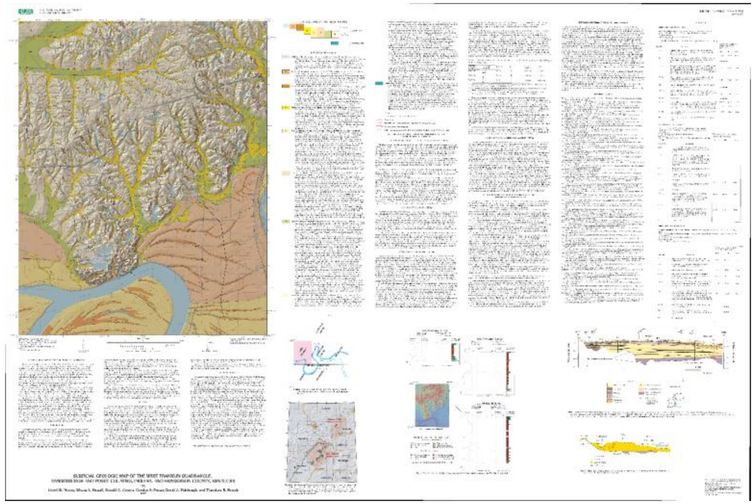
Across all pairs of pixel sets in map content and map keys

Difference between two color spaces

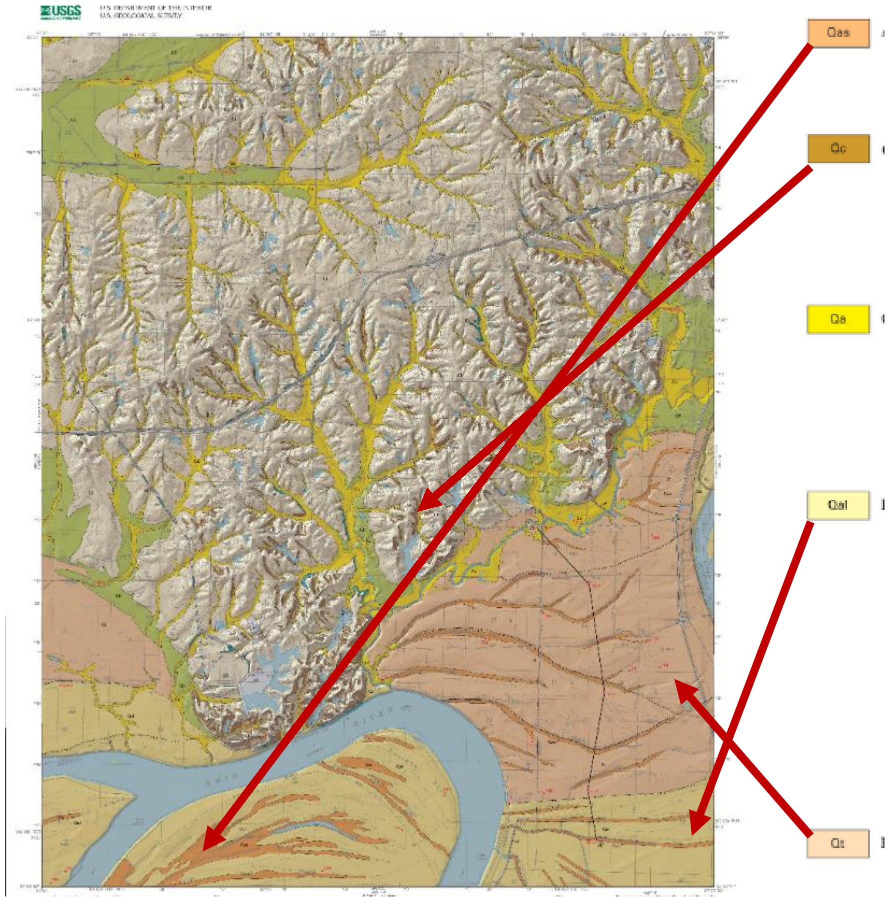
$$Dist_{K,P} = \sum_{k}^{K^s} \sum_{p}^{P^n} \left(\gamma \sqrt{\sum_{c}^{R,G,B} (c_k - c_p)^2} + (1-\gamma) \sqrt{\sum_{c}^{\overline{RG},\overline{GB},\overline{BR}} (c_k - c_p)^2} \right)$$

Methodology: Color Shift due to Shaded Relief

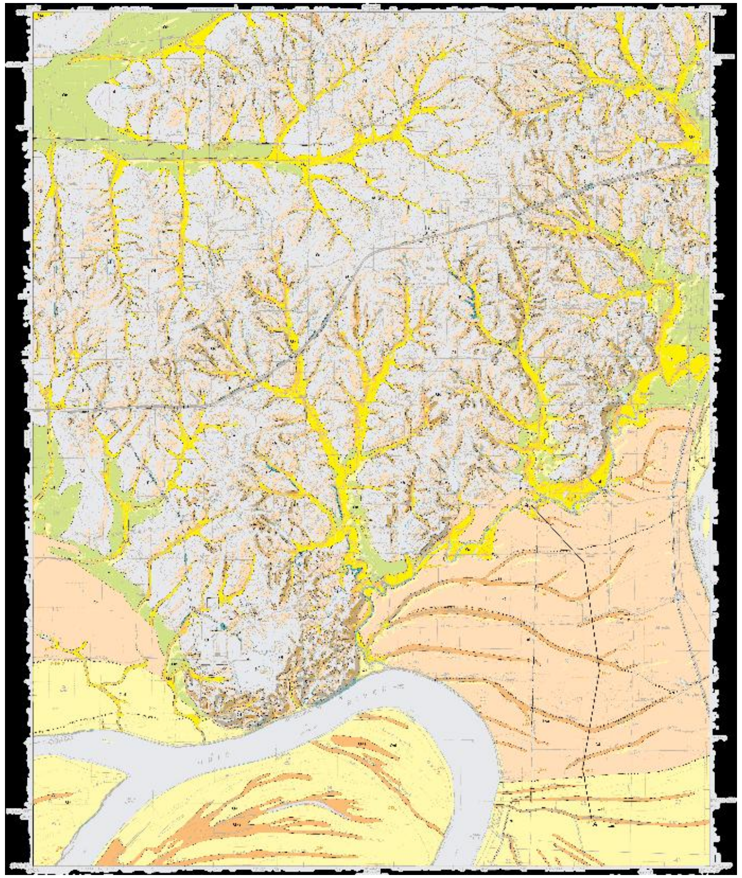
- A severe color-shift example by comparing the image re-colored based on map key



Original map

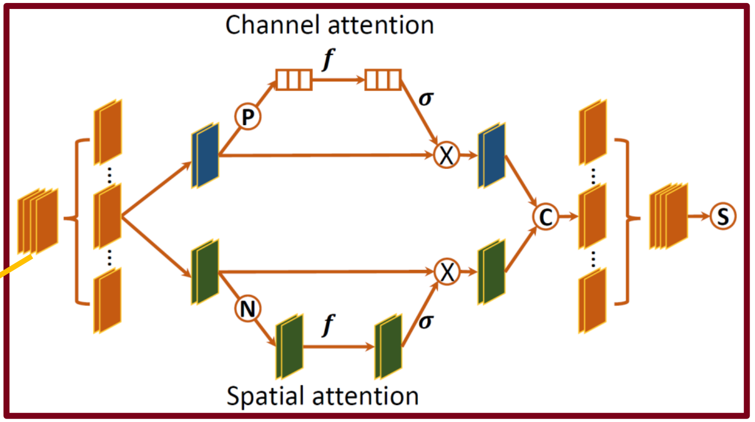


Recolored map

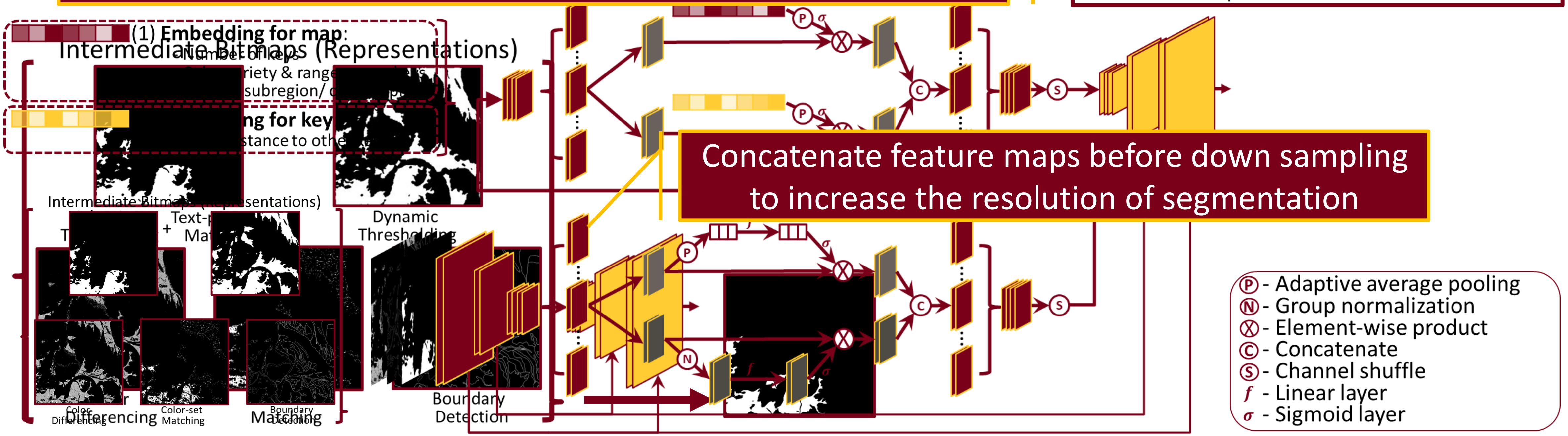


Methodology: Polygon-Recognition Model

- Learn to recognize polygons
  - U-net based convolutional model
  - Two-phase shuffle attention (SA-net)



Split feature maps into two branches to capture channel dependency and pair-wise relationship at the pixel level



Concatenate feature maps before down sampling to increase the resolution of segmentation

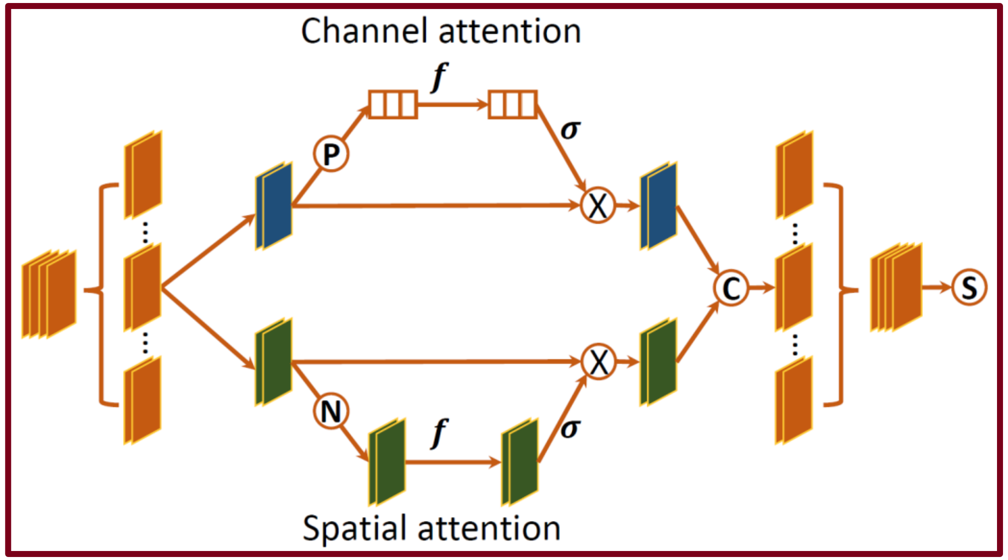
Methodology: Shuffle Attention

- SA-Net splits feature channels into groups and splits each into two branches
  - For channel attention, it employs 2-dimensional adaptive average pooling
  - For spatial attention, it applies group normalization
  - It concatenates the results from the two attentions for each group and shuffles the results from each group back into the feature channels

$$X_1' = \sigma\left(W_1 \cdot F_P(X_1) + b_1\right) \cdot X_1$$

$$F_P(X_1) = \frac{1}{g_h \times g_w} \sum_{i=1}^{g_h} \sum_{j=1}^{g_w} X_1(i, j)$$

$$X_2' = \sigma\left(W_2 \cdot F_N(X_2) + b_2\right) \cdot X_2$$



P - Adaptive average pooling
N - Group normalization
X - Element-wise product
C - Concatenate
S - Channel shuffle
f - Linear layer
σ - Sigmoid layer

Motivating Example: Colors of Map Keys

- The color depends on the geologic time and rock type of the map key

33—SUGGESTED RANGES OF MAP-UNIT COLORS FOR VOLCANIC AND PLUTONIC ROCKS AND FOR STRATIGRAPHIC AGES OF SEDIMENTARY AND METAMORPHIC ROCKS

CMYK* values (K = 0): A = 8%; 1 = 13%; 2 = 20%; 3 = 30%; 4 = 40%; 5 = 50%; 6 = 60%; 7 = 70%; X = 100%

33.1—Suggested range of map-unit colors for volcanic and plutonic rocks*

010	030	050	070	0X0	057	07X	036	047	05X
A60	270	3X0	150	370	5X0	033	055	077	0XX

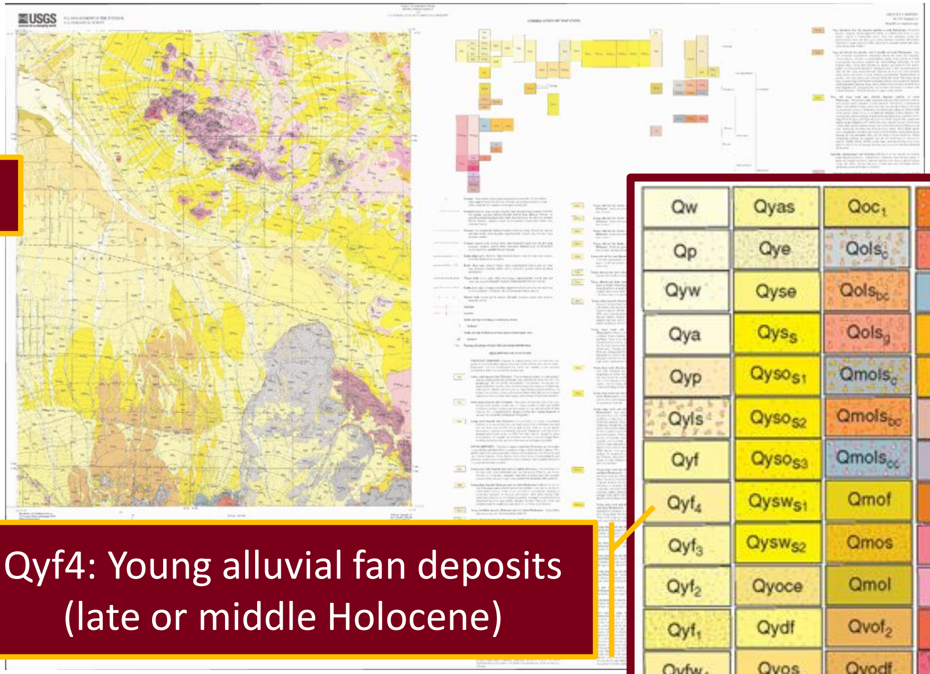
33.2—Suggested range of map-unit colors for stratigraphic ages of sedimentary and metamorphic rocks*

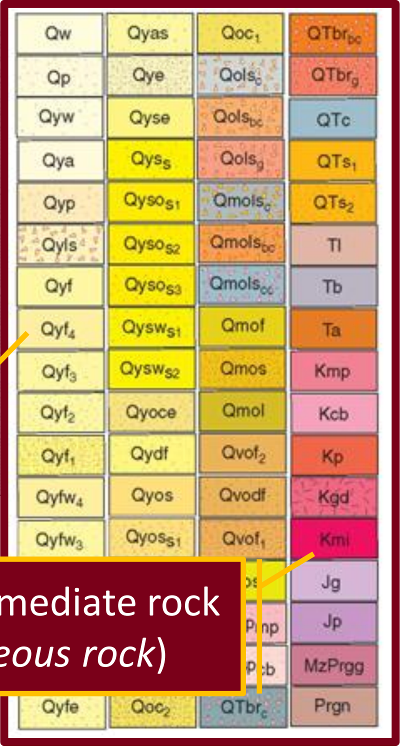
Q 007	001	0A6	005	003			
T 037	0A3	A4X	A37	026	014	A25	024
K 507	104	517	415	406	305		
J 604	202	705	504	303			
Ŧ 602	20A	6A3	402	301			
P 600	300	701	501	40A			
IP 620	4A0	72A	61A	510			
M 431	21A	531	42A	32A			
D 540	220	650	440	330			
S 350	A20	460	34A	230			
O 051	02A	A51	041	031			
€ 054	022	A54	043	A33			
p€ 446	A11	455	344	233	122	121	
	A12	457	346	235	124	A13	
	1A3	537	436	326	324	214	
	1AA	533	433	422	322	211	

Volcanic and plutonic rocks

Sedimentary and metamorphic rocks

STRATIGRAPHIC AGE	SUBDIVISION TYPE	AGE SYMBOL*
Cenozoic	Era	Cz
Quaternary	Period	Q
Tertiary	Period	T
Neogene	Subperiod	N
Paleogene	Subperiod	Pg
Mesozoic	Era	Mz
Cretaceous	Period	K





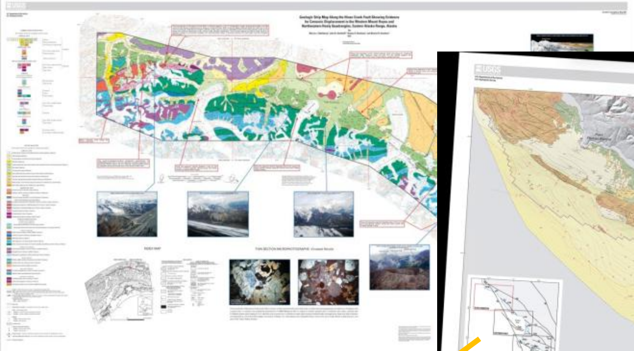
Qyf4: Young alluvial fan deposits (late or middle Holocene)

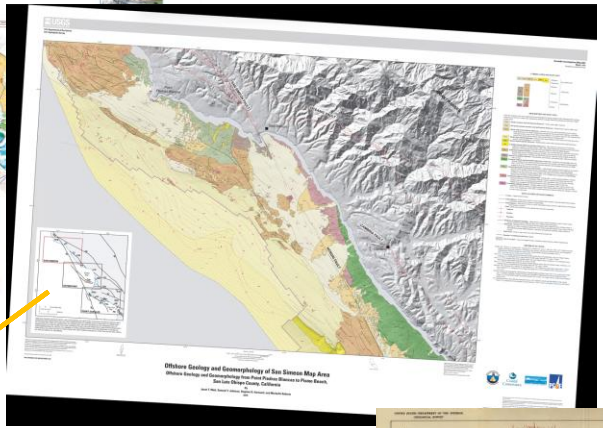
Kmi: Mafic-rich intermediate rock (Cretaceous) (*Igneous rock*)

Reference:
FGDC Digital Cartographic Standard for Geologic Map Symbolization.
Geologic Data Subcommittee, Federal Geographic Data Committee, 2006.

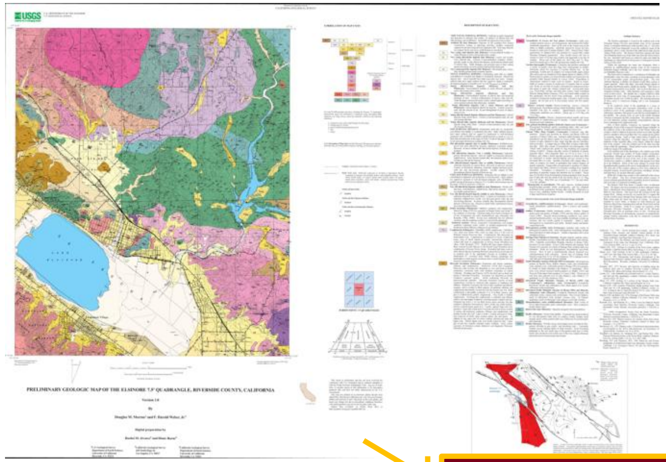
Evaluation: Some Maps in the Dataset

- Maps under various timespans and styles

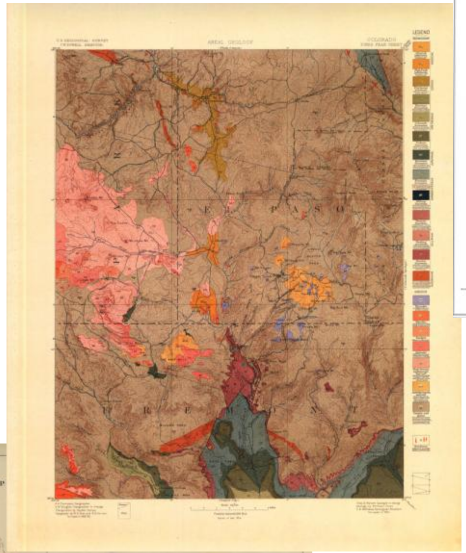




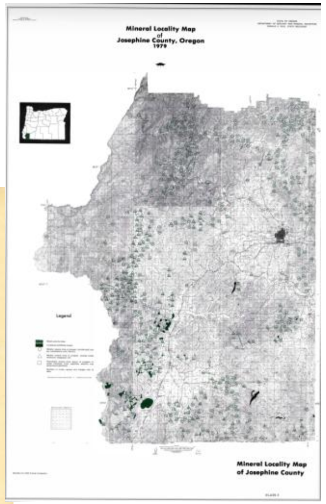
Digital-born geologic maps in 2010s

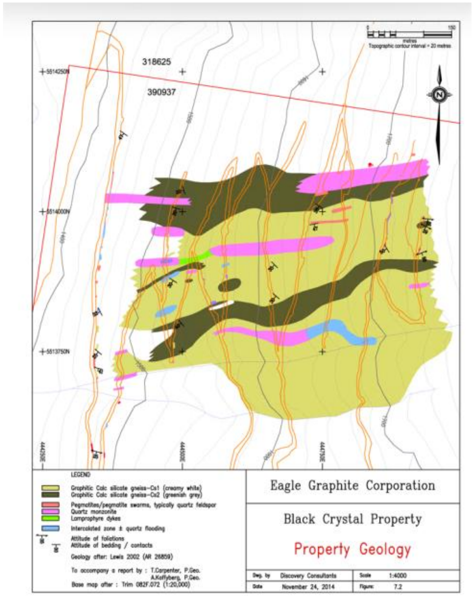


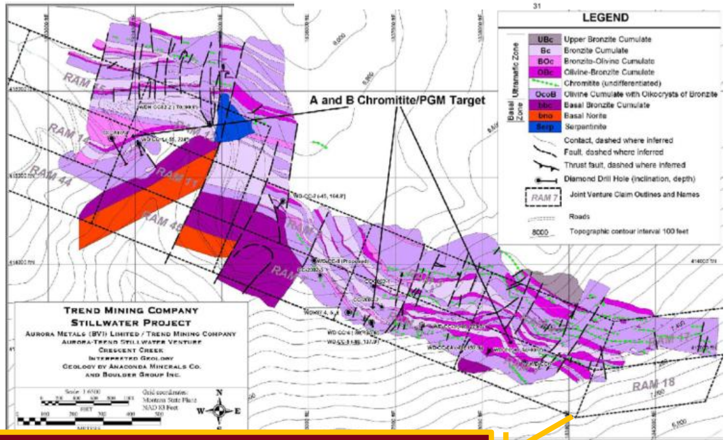
1:24000 geologic maps from state geological survey



Geologic or prospective maps from USGS earlier than 1980s







Geologic or prospective maps from mining company

Evaluation: Performance Breakdown

- Each component contributes to the final performance

Method	Weighted Performance														
Indicator	Precision														
Percentile	10	25	50	75	90	10	2[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	90
LOAM	0.032	0.280	**0.891**	**0.973**	0.986	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	**0.974**
Model Structure															
LOAM -DC/SA	0.035	0.262	0.870	0.970	0.984	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	**0.057**	0.307	0.771	0.938	0.974
LOAM -DC	0.030	0.246	0.838	0.966	0.983	0.409	0.741	0.906	0.966	0.988	0.048	0.304	0.763	0.935	0.972
LOAM -SA	0.031	0.220	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.988	0.051	0.289	0.760	0.938	0.973
Model Input															
LOAM -DT	0.030	0.216	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]73
LOAM -CD	**0.040**	**0.281**	0.825	0.964	0.983	0.241	0.66[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]73
LOAM -CM	0.024	0.156	0.780	0.961	0.984	0.458	0.770	0.921	0.977	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]966
LOAM -BD	0.026	0.185	0.784	0.960	0.987	0.348	0.623	0.839	0.943	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]52
Component															
AT+TM	0.029	0.224	0.740	0.892	0.946	0.337	0.702	0.892	0.982	0.997	0.049	0.272	0.682	0.876	0.940
DT	0.028	0.196	0.678	0.838	0.915	0.458	0.770	**0.930**	0.990	0.998	0.049	0.260	0.672	0.846	0.914
AT	0.024	0.139	0.650	0.868	0.936	0.451	0.760	0.925	**0.991**	**0.998**	0.042	0.210	0.630	0.861	0.928
CM	0.016	0.109	0.451	0.880	**0.989**	0.084	0.286	0.519	0.739	0.861	0.024	0.112	0.377	0.675	0.836

The polygon-recognition model improves precision based upon multiple representations

U-net structure can obtain decent precision

Including color-set matching output is crucial to low-accuracy cases

…n matching and color-set matching tend to have high precision

Color thresholding tends to have high recall

AT: adaptive color thresholding; **TM**: text-pattern matching; **DT**: dynamic color thresholding; **CD**: color differencing; **CM**: color-set matching; **BD**: boundary detection; **DC**: dual-scale information-based channel attention; **SA**: polygon-based spatial and channel attention.

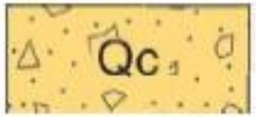
Evaluation: Running Time

- The inference time is around 3 minutes per map key

Stage	Component	Running Time (format: *hh:mm:ss*)			
		Median	Average	Maximum	Overall
Preprocessing of Metadata	Map-Content Detection	0:03:36	0:03:58	0:08:31	1:35:14
	Color Differencing (**CD**)	0:03:16	0:04:53	0:18:22	1:57:14
	Adaptive Color Thresholding (**AT**)	0:01:05	0:01:28	0:04:43	0:35:15
	Conditional Dilation (**AT**)	0:07:41	0:10:26	0:35:41	4:10:25
	Dynamic Color Thresholding (**DT**)	0:01:52	0:02:48	0:10:47	1:07:15
	Text-Pattern Matching (**TM**)	0:07:44	0:30:02	3:25:31	12:00:50
	Boundary Detection (**BD**)	0:00:42	0:00:55	0:03:07	0:21:51
	Color-Set Matching (**CM**)	0:12:23	0:25:08	2:22:09	10:03:19
	Auxiliary-Info Embedding	0:00:18	0:00:20	0:00:43	0:07:52
	Generating Bitmaps for Model	0:01:40	0:02:28	0:08:04	0:59:11
Learning to Recognize Polygons	Model (Training on training dataset)	N.A.			2:53:17
	Model (Inference on testing dataset)	N.A.			11:45:16

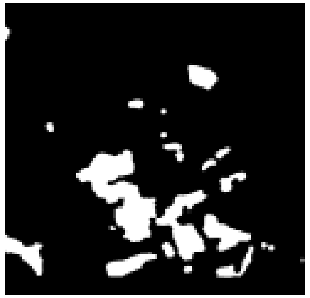
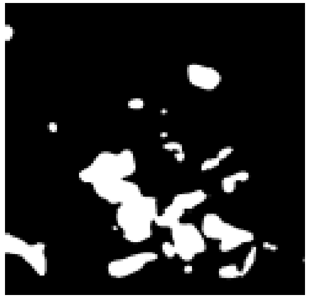
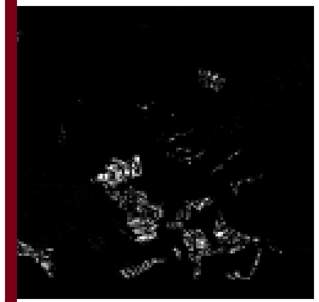
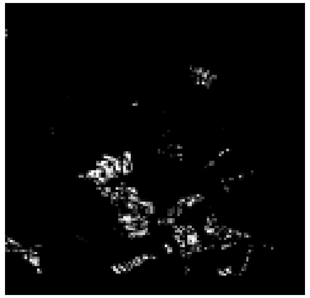
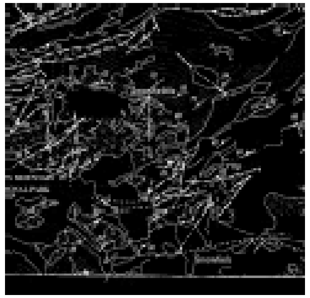
Case Study

- Challenge 1: Various **styles of map keys**

Input	
Ove	Map Key
	
	Qc_poly

Failed to include translucent symbols

Color thresholding can extract rough areas

Output							
Input Channels of LOAM Polygon-Recognizing Model (partial image, overall performance for median precision, recall, and F1 s							Groundtruth
AT+TM	DT	CD	CM	BI			
							
(0.748, 0.844, 0.793)	(0.678, 0.866, 0.760)	N.A.	(0.760, 0.216, 0.336)	N.A.	(0.809, 0.726, 0.765)		

AT: adaptive color thresholding; **TM**: text-pattern matching; **DT**: dynamic color thresholding; **CD**: color differencing; **CM**: color-set matching; **BD**: boundary detection.

Case Study

- Challenge 2: Same color, **different text**

Input		
Raster Image		Map Key
Overview	Partial	
		
		led_poly

Text-pattern matching failed due to cropping

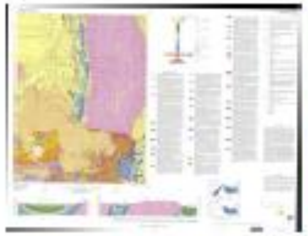
Color-set matching identify distinct usage

Output						
Input Channels of LOAM Polygon-Recognizing Model (partial image, overall performance for median precision, recall, and F1 score)					**LOAM**	**Groundtruth**
AT+TM	**DT**	**CD**	**CM**	**BD**		
(0.382, 1.000, 0.553)	(0.359, 1.000, 0.529)	N.A.	(0.998, 0.630, 0.773)	N.A.	(0.733, 0.995, 0.844)	

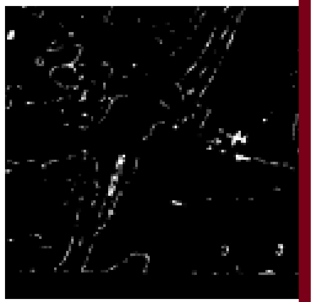
AT: adaptive color thresholding; TM: text-pattern matching; DT: dynamic color thresholding; CD: color differencing; CM: color-set matching; BD: boundary detection.

Case Study

- Challenge 3: **Color shift** between key and content

Input		
Raster Image		Map Key
Overview	Partial	
		
	en	Xq_poly

Dynamic color thresholding compensates false negatives from others

Input Channels of LOAM Polygon-Recognizing Model (partial image, overall performance for median precision, recall, and F1 score)					LOAM	Groundtruth
AT+TM	DT	CD	CM	BD		
						
(0.424, 0.122, 0.189)	(0.573, 0.994, 0.727)	N.A.	(0.862, 0.667, 0.752)	N.A.	(0.894, 0.989, 0.939)	

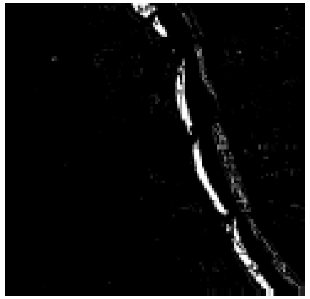
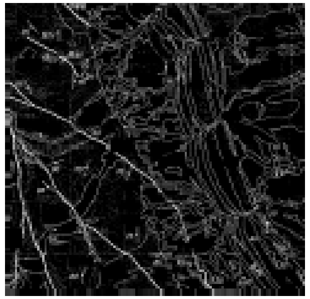
AT: adaptive color thresholding; **TM**: text-pattern matching; **DT**: dynamic color thresholding; **CD**: color differencing; **CM**: color-set matching; **BD**: boundary detection.

Case Study

- Challenge 4: Text **labeled outside** polygon

Input		
Raster Image		Map Key
Overview	Partial	
		
CO_DenverW		TRPI_poly

Text pattern is not used as a direct prompt

Output						
Input Channels of LOAM Polygon-Recognizing Model (partial image, overall performance for median precision, recall, and F1 score)					LOAM	Groundtruth
AT+TM	DT	CD	CM	BD		
						
(0.761, 0.874, 0.813)	(0.633, 0.880, 0.736)	N.A.	(0.289, 0.638, 0.398)	N.A.	(0.937, 0.896, 0.916)	

AT: adaptive color thresholding; **TM**: text-pattern matching; **DT**: dynamic color thresholding; **CD**: color differencing; **CM**: color-set matching; **BD**: boundary detection.

Related Work: U-net vs. fine-tuned SAM

- Given sufficient (> 2% cases) training data
  - U-net can outperform fine-tune SAM with transformer on single-type polygon extraction

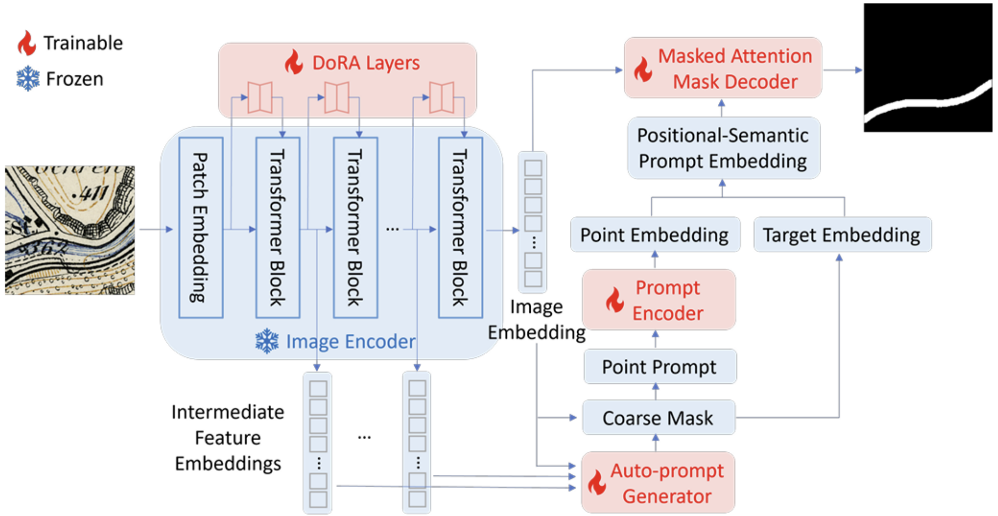
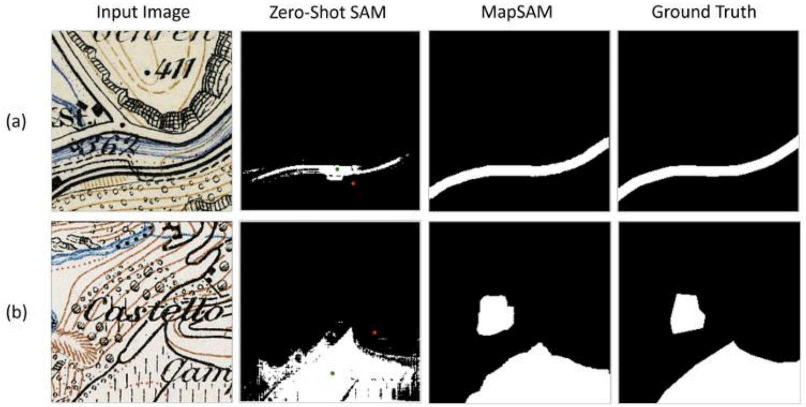


TABLE I
COMPARISON OF THE PROPOSED MAPSAM WITH OTHER BASELINES ON THE RAILWAY AND VINEYARD DATASETS. THE BEST RESULTS OF IOU AND F
ARE IN **BOLD**.

Method	Railway								Vineyard			
	Full (5872)		10% (587)		1% (58)		10-shot		Full (613)		10-shot	
	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU
U-Net	**95.13**	**91.86**	**94.30**	**90.56**	89.68	83.52	74.02	61.43	**84.61**	**77.04**	71.52	60.23
SAMed	91.98	86.31	91.63	85.69	91.75	86.01	84.62	75.44	82.80	74.85	**71.95**	**61.53**
Few-Shot SAM	–	–	–	–	–	–	47.49	35.82	–	–	57.99	46.83
MapSAM	94.06	89.46	93.57	88.71	**92.05**	**86.53**	**87.17**	**78.50**	82.78	74.32	70.52	60.02



Line feature

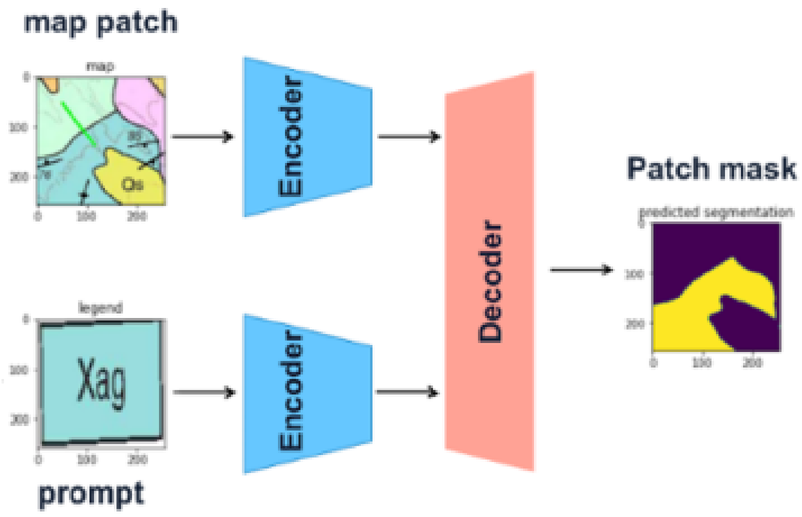
Polygon feature

Related Work: U-net with different patch size

- Enlarging training dataset and decreasing crop size can improve accuracy of U-Net
  – The training dataset here is at least 3 times larger than the one we use

Table 2. Results of experiments with different patch sizes and overlap sizes using the Vanilla U-Net model. All metrics reported here are for "patch-wise" measurement. The parameters with the best performance is in bold.

Model	Patch Size	Overlap	Best Train F1 Score (%)	Best Validation F1 Score (%)	Difference (%)
Vanilla_Unet	128	3	95.29	80.72	14.57
		5	95.89	80.59	15.30
		10	95.49	81.26	14.23
		15	96.05	82.18	13.87
	256	3	90.58	81.65	8.93
		5	91.24	82.26	8.98
		10	92.18	84.23	7.95
		15	93.54	84.39	9.15
		32	**94.02**	**85.72**	**8.30**
	512	3	78.84	75.43	3.41
		5	70.46	69.86	0.60
		10	74.38	72.21	2.17
		15	78.16	72.76	5.40
	1024	3	23.44	20.02	3.42
		5	16.88	15.04	1.84
		10	24.59	26.76	−2.17
		15	24.66	23.53	1.13



The accuracy reported here is "patch based" instead of per map

Related Work: Feature Extraction from Geologic Maps

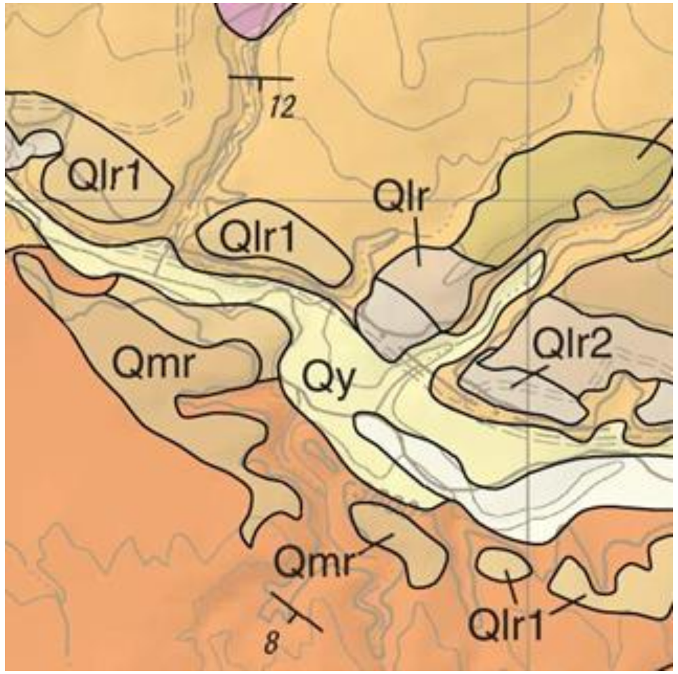
- Related Work
  - Line Extraction (Chiang 2013, Duan 2017, Duan 2020, Xia 2023, Duan 2025)
    - Apply mean shift with Hough transform, convolutional model, or transformer for line detection
    - Require a tailored model for each type of line feature
    - Map keys only demonstrate parts of the line feature
  - Point Extraction (Chiang 2009, Saeedimoghaddam 2020)
    - Apply graphic recognition or convolutional mode to extract road intersections from raster maps
    - Require a tailored model for each type of point feature
  - Polygon Extraction (Luo 2023, Saxton 2024)
    - Apply U-Net with OCR and histogram equalization
    - Do not address the color shift and map styles

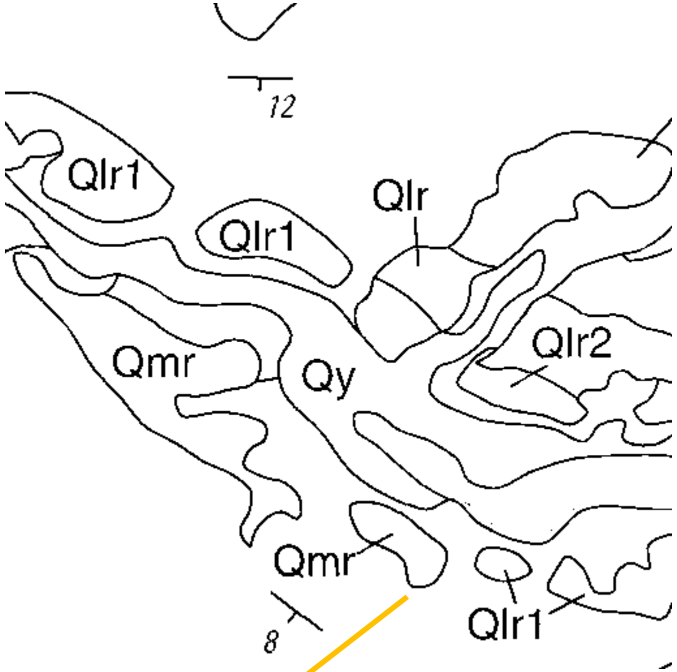
Appendix for Recoloring Historical Maps

Sketch Extraction
Color Style Extraction
Multi-scale Recoloring Model

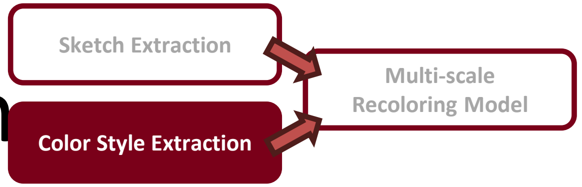
Methodology: Sketch Extraction

- Segmenting polygons with distinct colors
  - Color thresholding
  - Canny-edge detection
  - Color-gradient detection





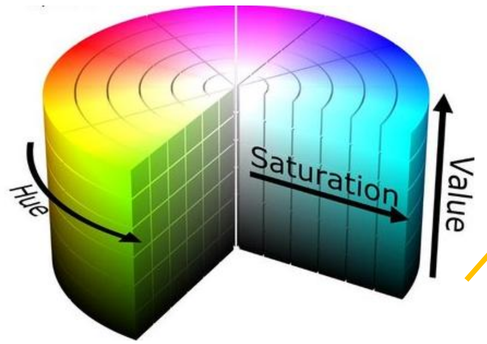
Separate polygons that use statistically significantly different colors



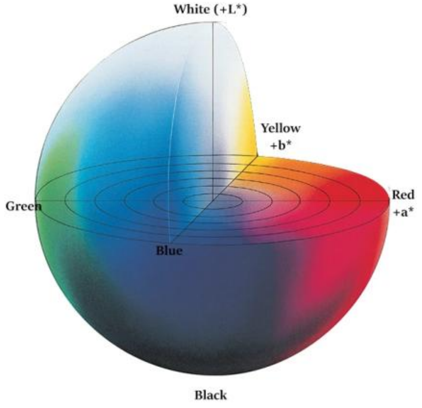
Methodology: Color Style Extraction

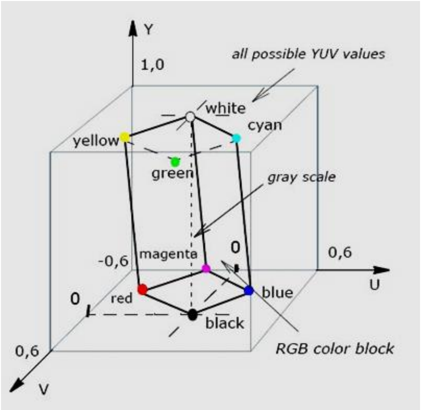
- Extract color spectra from map and keys with explicit palette hint for the image crop
  - For image crop: k-means clustering
  - For map key: color-histogram embedding

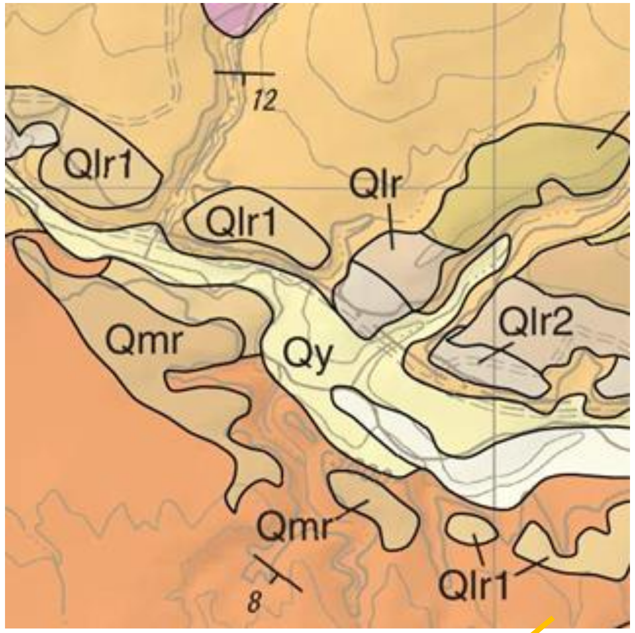
Preserve complex color usage within each map key



Exploit HSV, CIELAB, YUV color spaces to build spectra embeddings







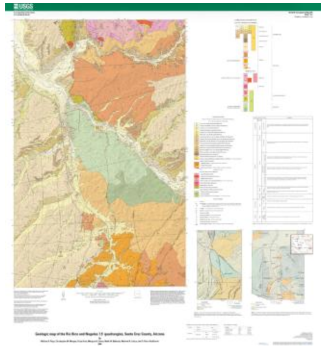
Identify dominant target color distribution from the spectra embedding

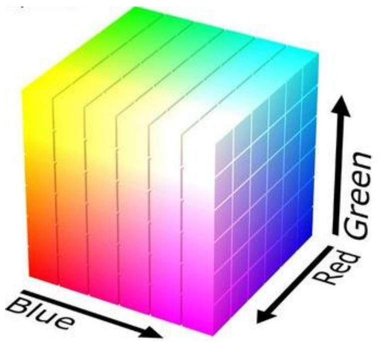
Methodology: Spectra Embedding

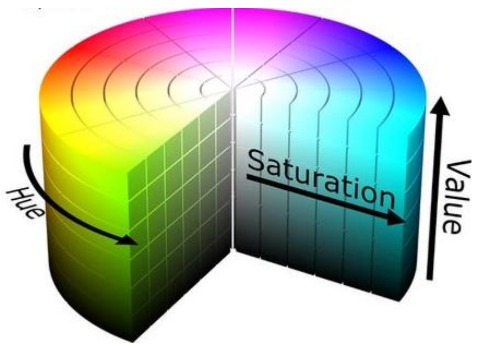
- Perceptual color histograms for the map content and map keys
  - H (HSV): hue
  - L* (CIELAB): lightness
  - U & V (YUV): chrominance

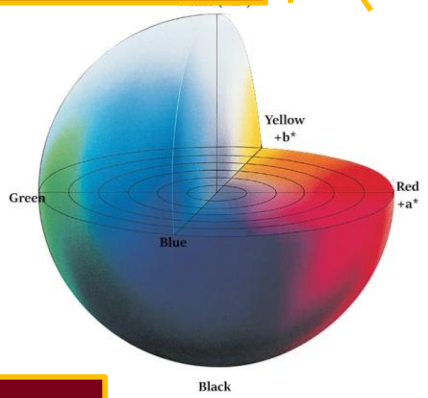
The two perceptual signatures highlight the color-usage difference between map content and keys

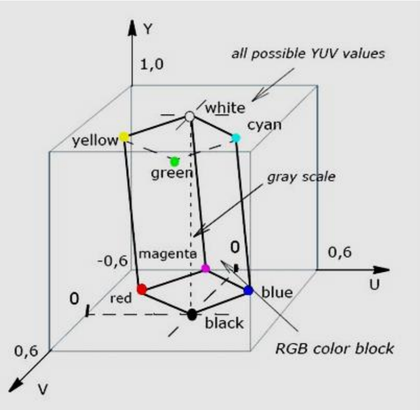
LAB has uniform perception for brightness but chromatic info sensitive to lighting









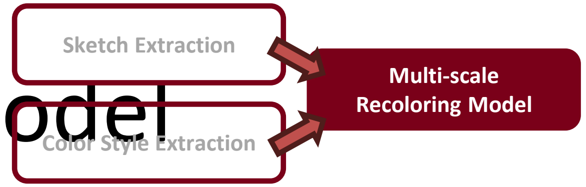


RGB is sensitive to lighting and shadows and highly correlated among channels

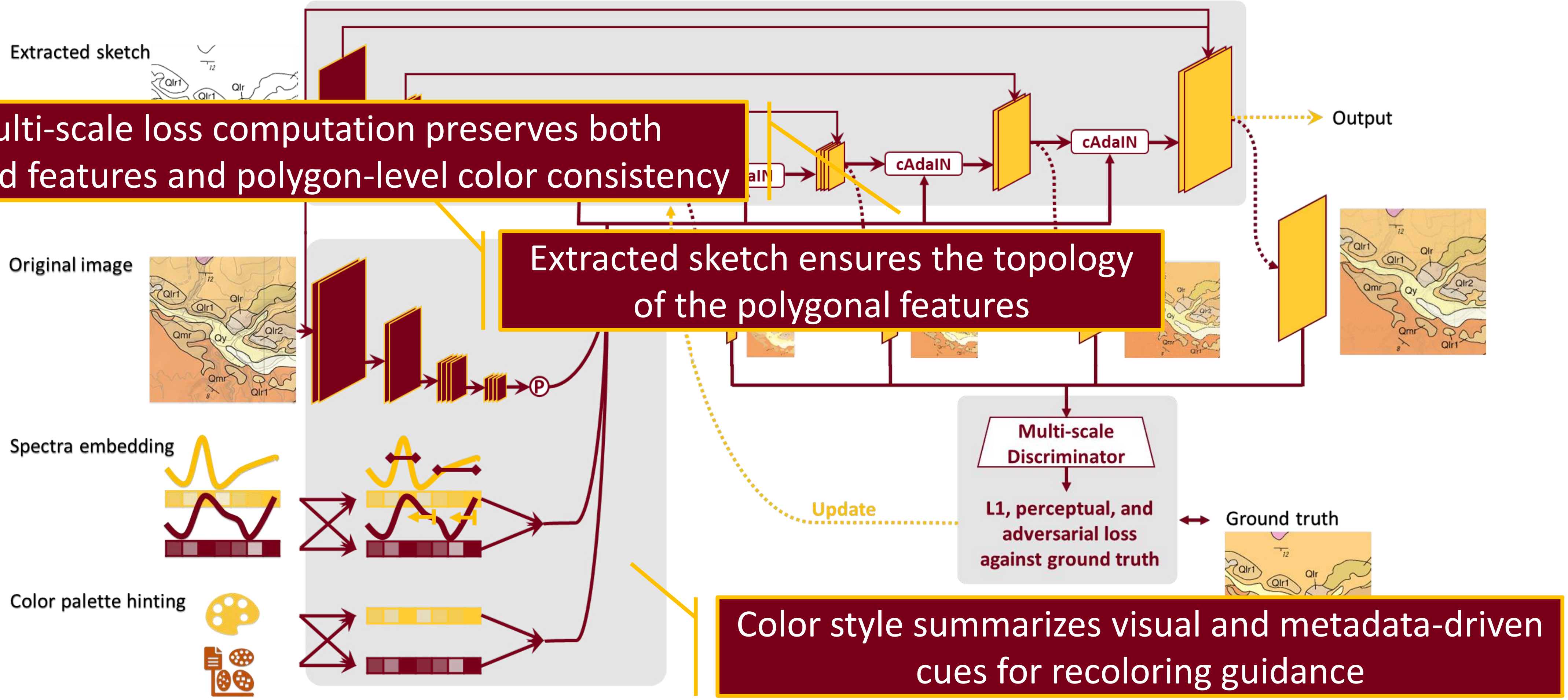
HSV has intuitive color identity but no perceptually uniform chromatic info

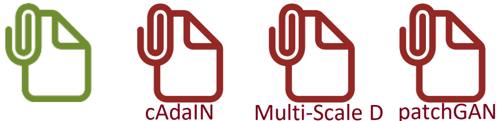
YUV has decorrelated chroma contrast but no perceptually uniform brightness

Methodology: Multi-scale Recoloring Model



- A multi-scale GAN with structural and color style as a guidance for recoloring

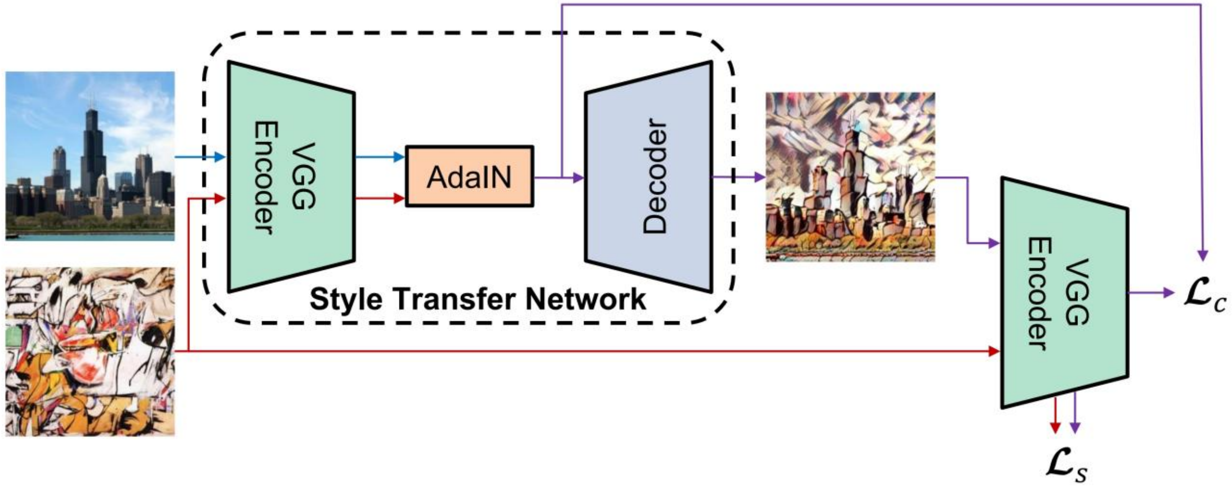


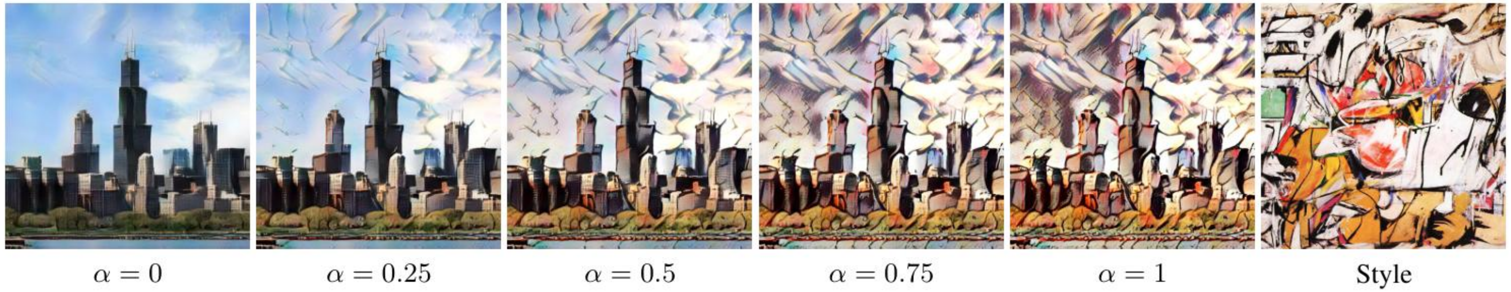


Methodology Ref: Adaptive Instance Normalization

- Adaptive regional style transformation with instance normalization
  - Affine parameters based on the features representations of the image of the targeted style

$$\mathrm{AdaIN}(x, y) = \sigma(y)\left(\frac{x - \mu(x)}{\sigma(x)}\right) + \mu(y)$$

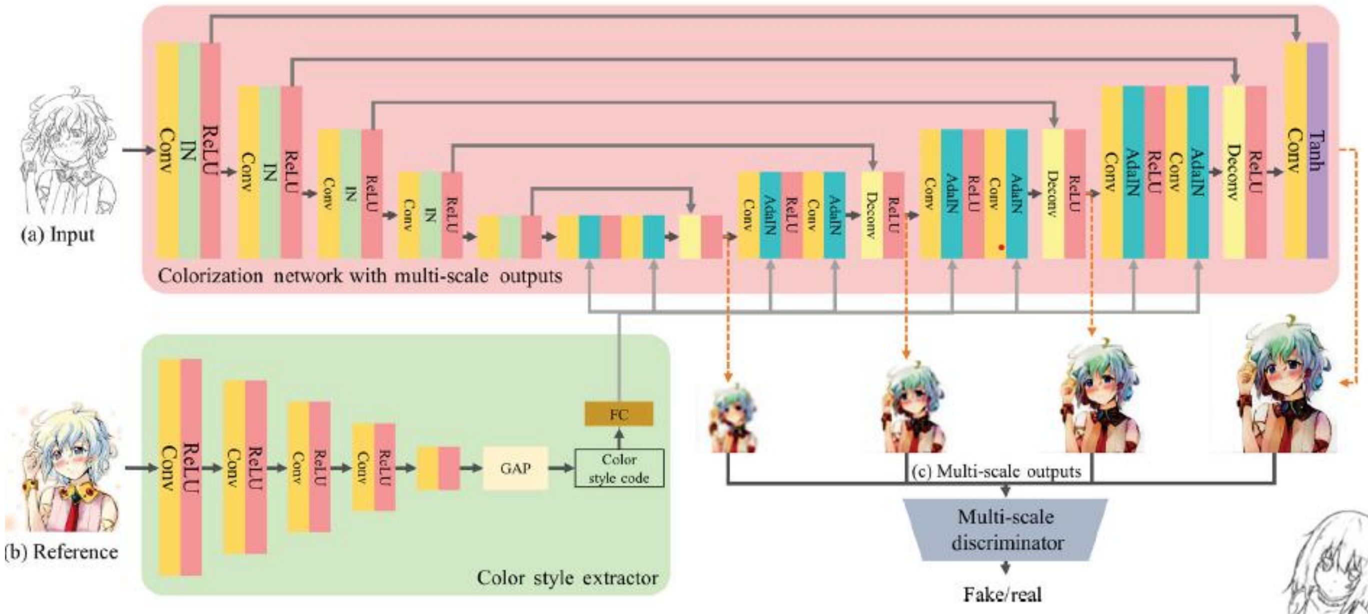




Reference:
Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization. ICCV, 2017.

Methodology Ref: Multi-scale Discriminator

- A GAN using extracted color-style space representation as a reference
  - Multi-scale adversarial loss balances global color composition and local color shading



$$\mathcal{L}_{\text{adv}} = \frac{1}{n}\sum_{i}^{n} \lambda_i E_{y_i}[\min(0, -D(y) - 1)] + \frac{1}{n}\sum_{i}^{n} \lambda_i E_{\hat{y}_i}[\min(0, D(\hat{y}_i) - 1)] + \omega_2 E_y[\|\nabla D(y)\|^2]$$

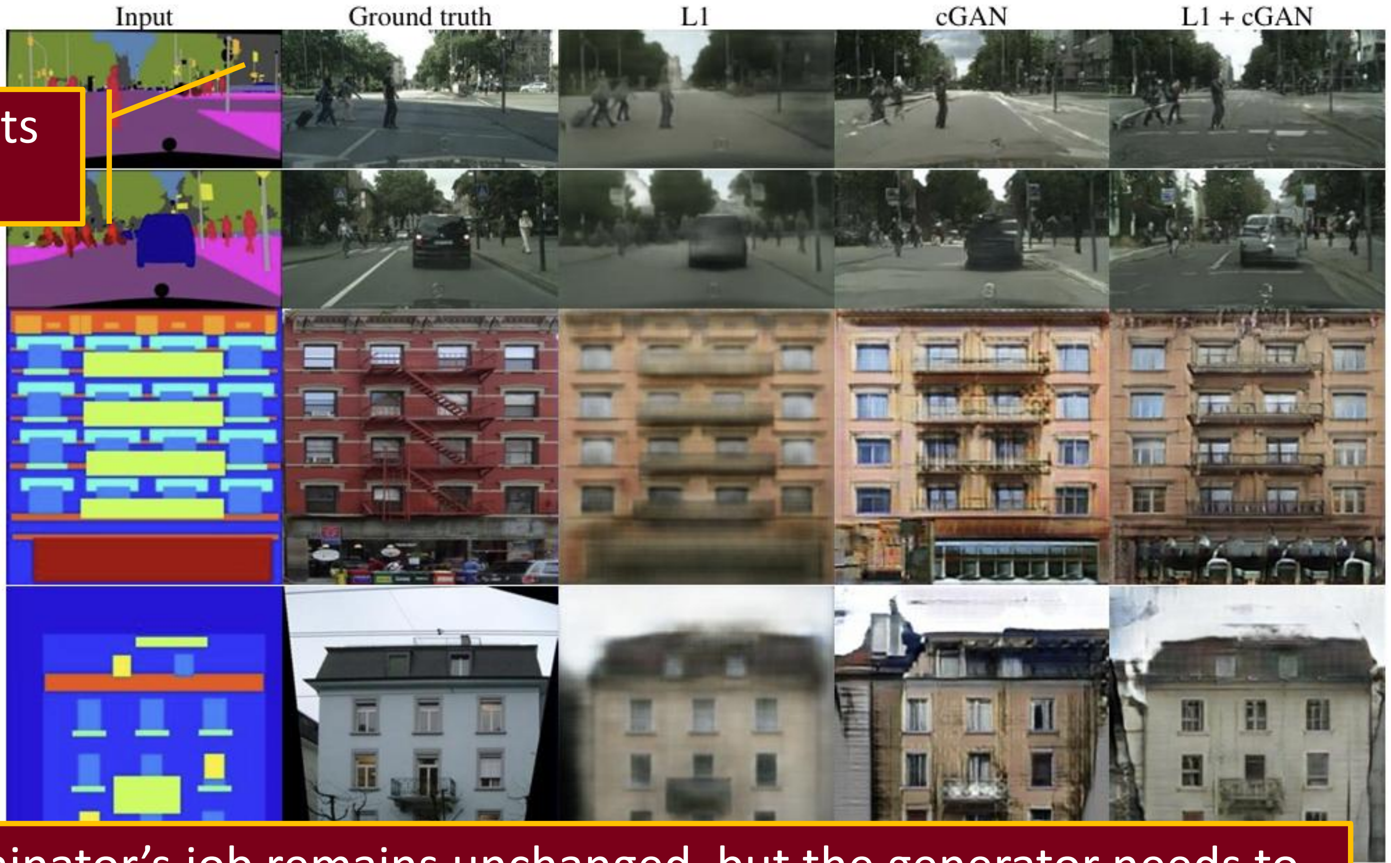
Apply gradient penalty regularization to improve stability

Multiple inputs to the discriminator improve the perceptual fields of the patchGAN

Methodology Ref: PatchGAN

- Combine conditional GAN objective with a traditional L1 loss

Using solely L1 loss tends to blur the results of conditional GAN



$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z))], \quad (1)$$

where G tries to minimize this objective against an adversarial D that tries to maximize it, i.e. $G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G, D)$.

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x, z)\|_1].$$

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G).$$

The discriminator’s job remains unchanged, but the generator needs to fool the discriminator while ensuring L1 loss against the ground truth

Evaluation: PSNR and SSIM Metrics

- PSNR (↑) (peak signal-to-noise ratio)
  - Inversed MSE
  - The larger the better
- SSIM (↑) (structural similarity index measure
  - Combination of luminance, contrast, and structural covariance
  - For balanced objectives and simplicity
    - $\alpha = \beta = \gamma = 1$
    - $C_2 = 2C_3$

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{MAX^2}{\mathrm{MSE}}\right)$$

$$\mathrm{SSIM}(\mathbf{x},\mathbf{y}) = [l(\mathbf{x},\mathbf{y})]^{\alpha} \cdot [c(\mathbf{x},\mathbf{y})]^{\beta} \cdot [s(\mathbf{x},\mathbf{y})]^{\gamma}$$

$$l(\mathbf{x},\mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$$

$$c(\mathbf{x},\mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$

$$s(\mathbf{x},\mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$$

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

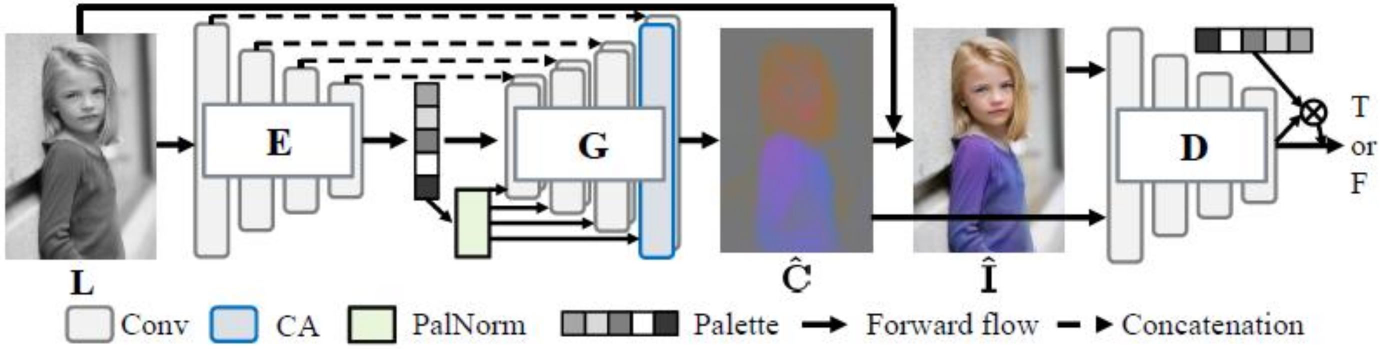
$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\sigma_x = \left(\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2\right)^{1/2}$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)$$

Related Work: Color Palette Hinting for GANs

- Colorize natural grayscale image using the color palette
  - Natural grayscale image itself already have enough statistical cues for recoloring/ colorization
  - Apply color palette from itself or external image with GAN



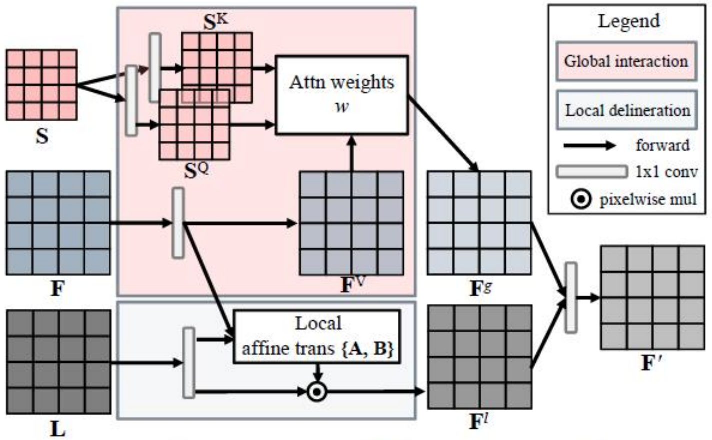


Fig. 4: The illustration of chromatic attention.

(a) (b) (c) (d) (e) (f)

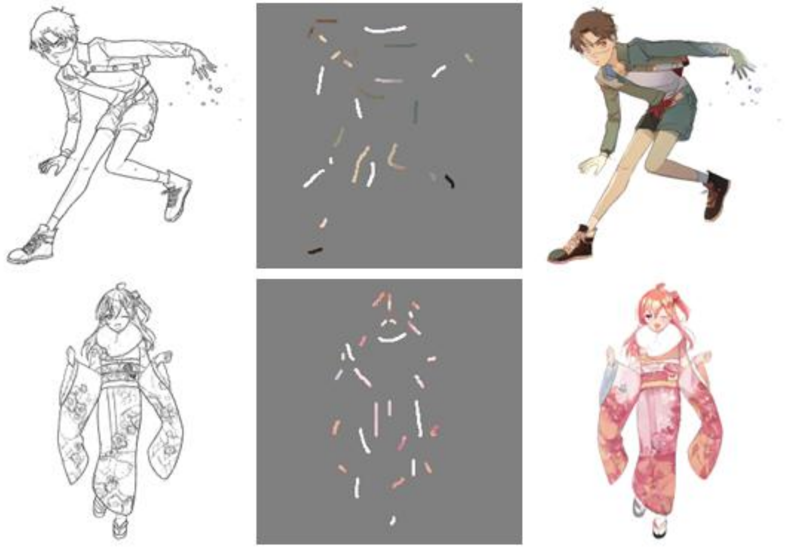
(a) (b) (c) (d) (e)

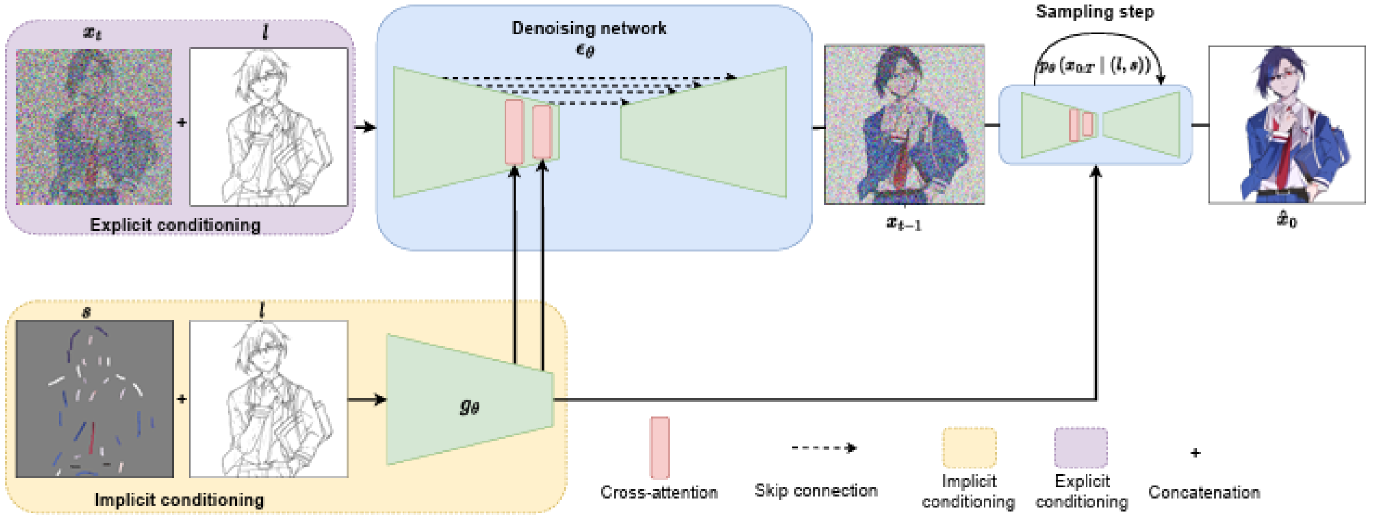
Fig. 5: Ablation studies of chromatic attention (CA). (a) input, (b) wo CA, (c) w Global, (d) w Local, (e) full CA. Please zoom in.

Reference:
PalGAN: Image colorization with palette generative adversarial networks. ECCV, 2022.

Related Work: Diffusion Model for Line Art Recoloring

- Colorizing grayscale line art with user color scribbles
  - A denoising model dedicated for output image quality
  - Require explicit user input

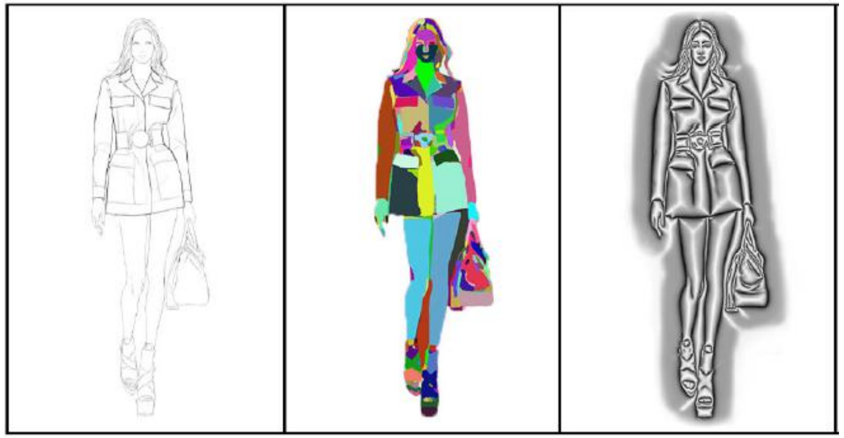




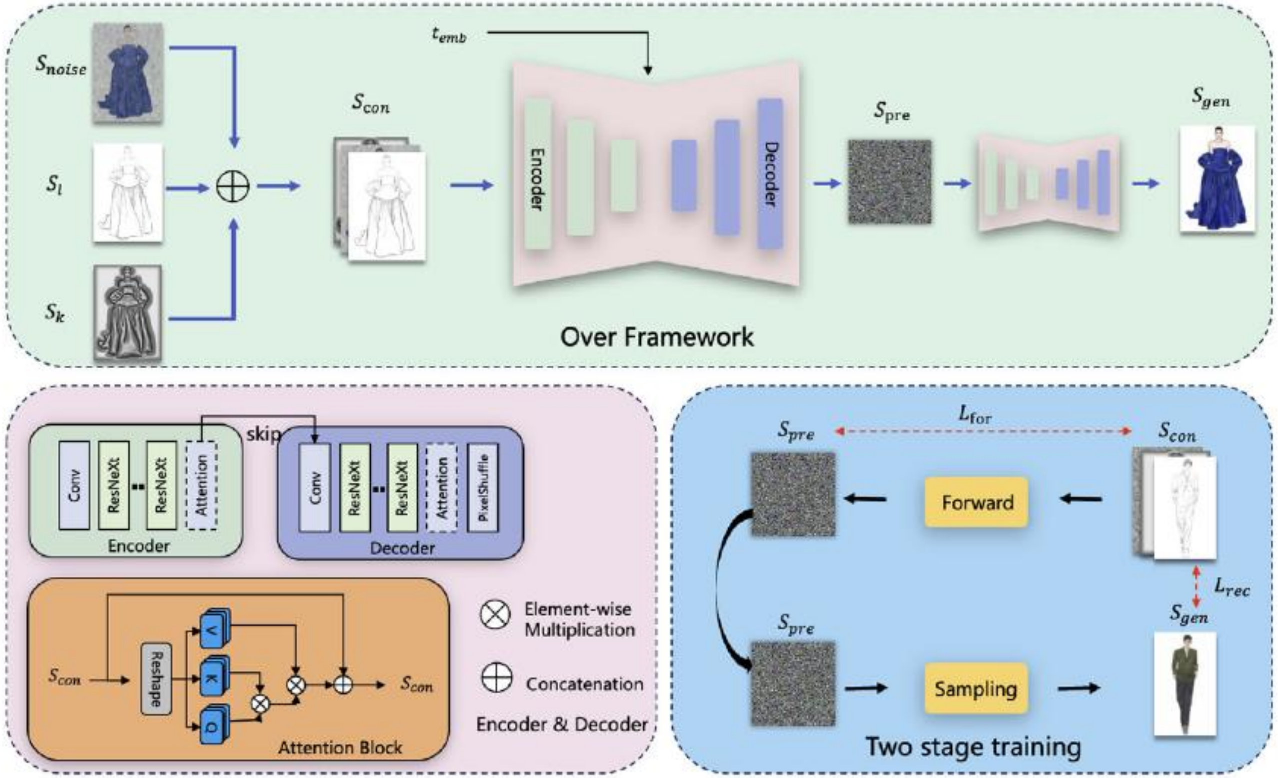
Reference:
Diffusart: Enhancing line art colorization with conditional diffusion models. CVPR, 2023.

Related Work: Diffusion Model for Line Art Recoloring

- Exploit diffusion model with colors from the reference image for recoloring
  - Generate region map and skeleton map from the input line art
  - Limited to drawings with extremely similar structure for color reference



(Line Drawing) (Region map) (Skeleton map)

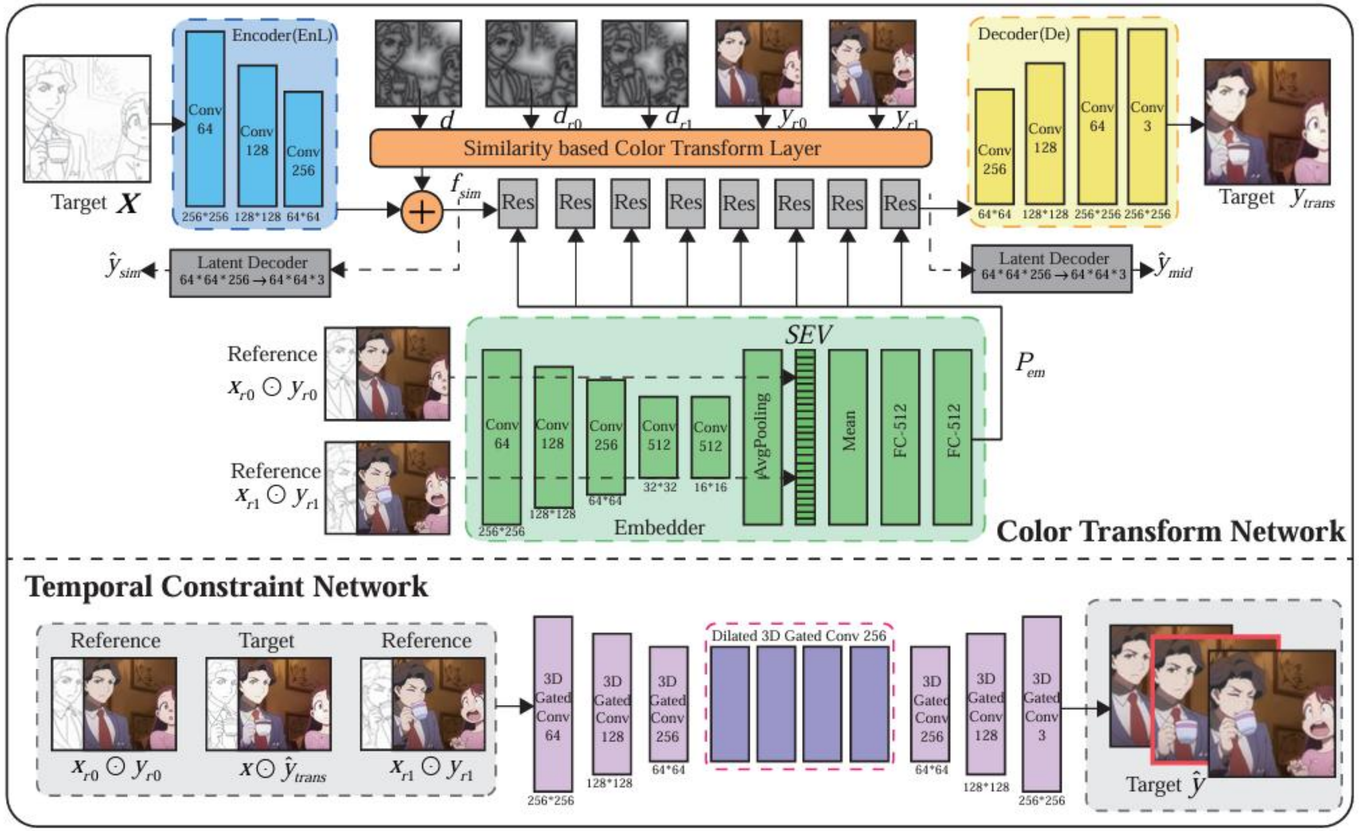


Reference:
Region-assisted Line Drawing Colorization through Diffusion Model. The Visual Computer, 2025.

Related Work: Line Art Video Recoloring

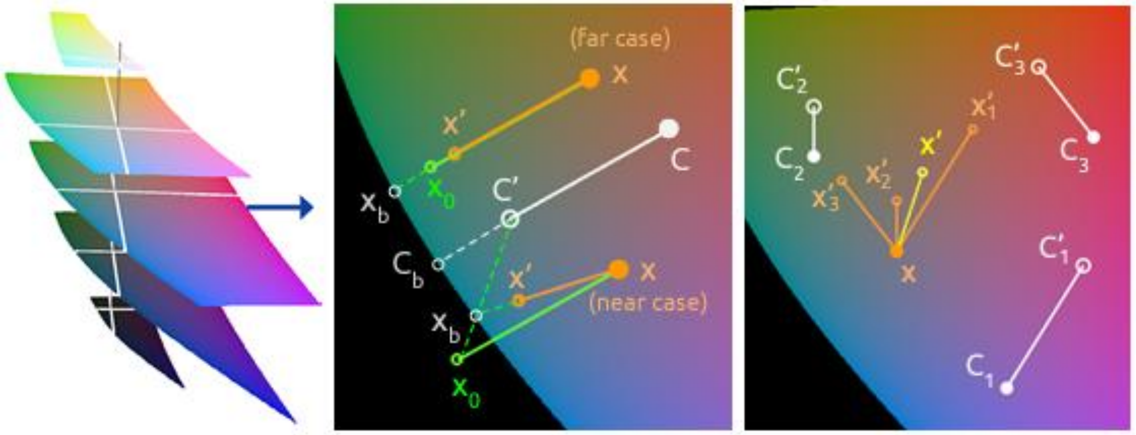
- Use colors from nearby frames as a reference for line art colorization or recoloring
  - Nearby frames tend to have extremely similar structure for color reference

Fig. 1: Comparison of different line art image extraction methods. (a) original color image; (b) Canny [12]; (c) XDoG [13]; (d) Coherent Line Drawing [14]; (e) SketchKeras [15]; (f) distance field map from SketchKeras results.



Related Work: Recoloring Natural Images with Palette

- Apply K-means-based approach to compute palette of the source image
  - Users can change the colors in the palette extracted from a colored natural image
  - Apply monotonic luminance transfer to recolor



(a) Lab slices (b) edit one color (c) RBF blending

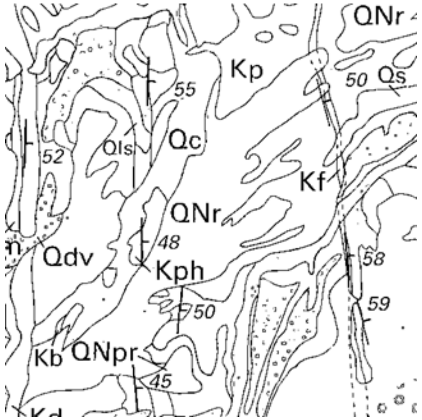
Appendix for Colorizing Draft Maps

Region Map Generation → Conditioned Colorization Model ← Semantic Tag Encoding

Methodology: Region Map Generation

- Hierarchical integration of SLIC superpixel
  - Reconcile into a region map linked to each level of superpixel

Geological units in draft maps exhibit varying degrees of visual and spatial coherence

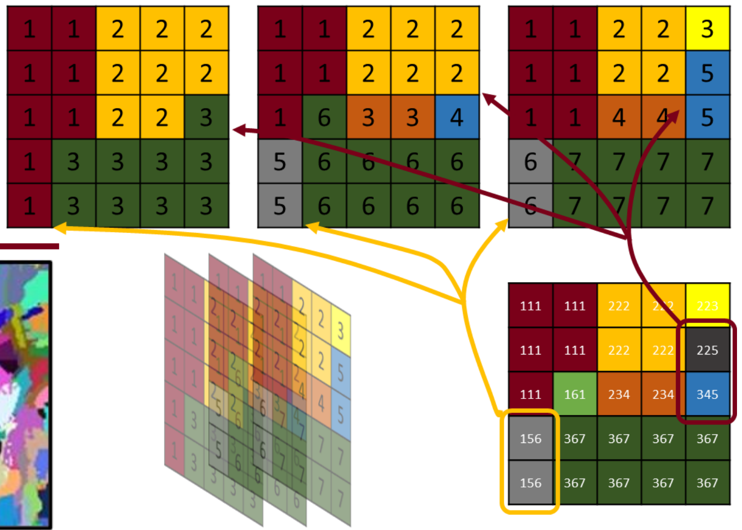


Visual coherence

Higher level: fewer centroids for contrast-oriented separation

Lower level: higher shape compactness for fine-grained boundaries

Spatial coherence

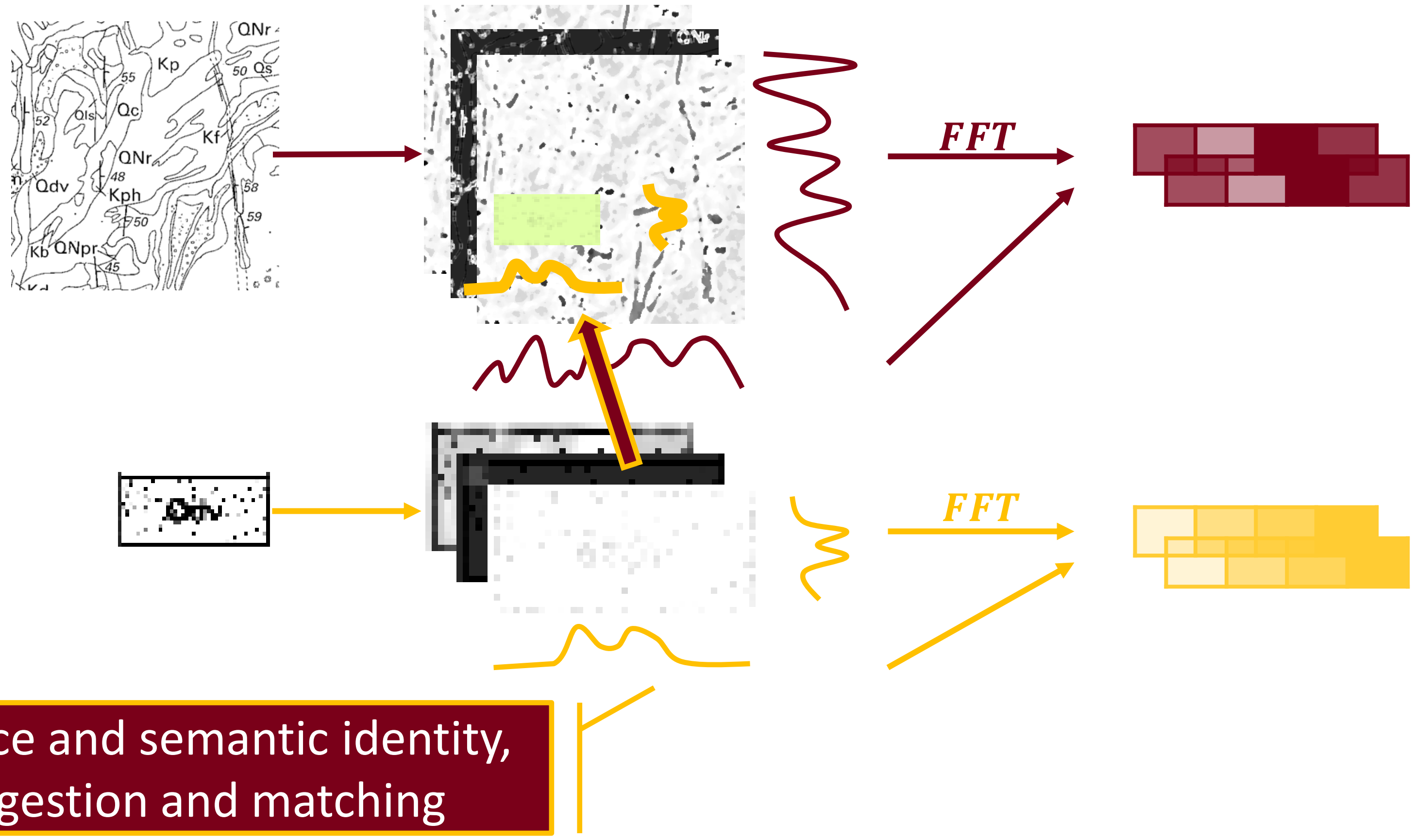


Hierarchical over-segmentation serve as a soft constraint to encourage a flat-color style within each polygon

Methodology: Semantic Tag Encoding

Region Map Generation
Semantic Tag Encoding
Conditioned Colorization Model

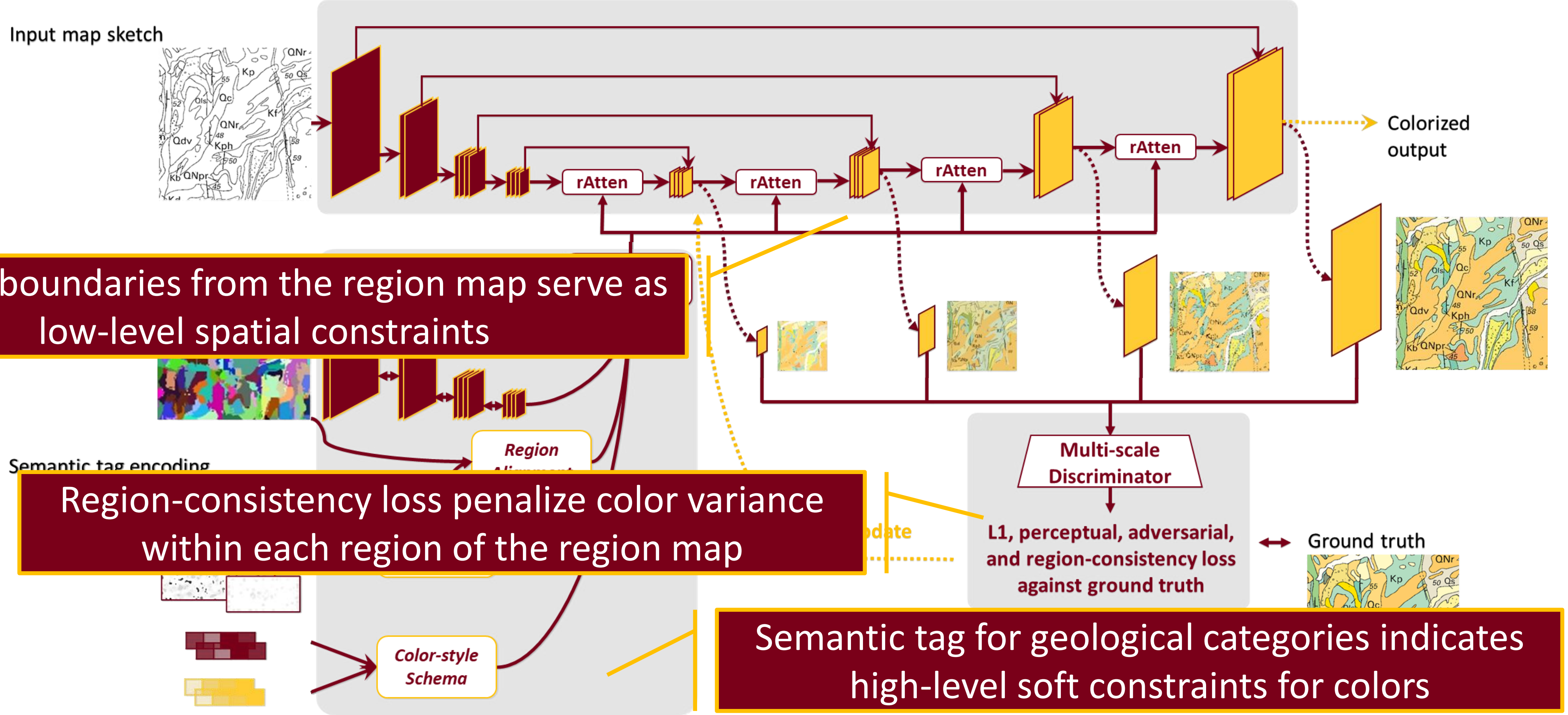
- The ideal color for a map key depends on its meaning following the convention
  - Pattern filtering for the map content and map keys



Embed visual appearance and semantic identity, supporting color suggestion and matching

Methodology: Conditioned Colorization Model

- A multi-scale GAN with semantic-region alignment as a guidance for colorization



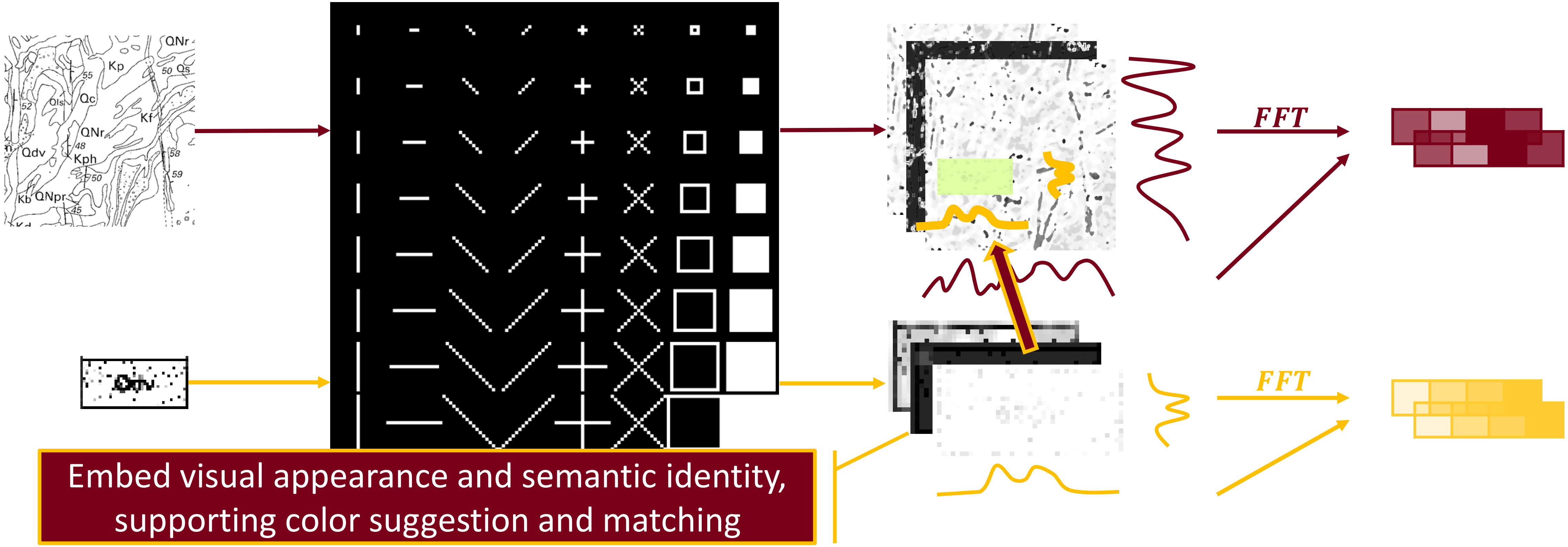
Polygon boundaries from the region map serve as low-level spatial constraints

Region-consistency loss penalize color variance within each region of the region map

Semantic tag for geological categories indicates high-level soft constraints for colors

Methodology: Semantic Tag Encoding

- Pattern filtering for the map content and map keys
  - 64 convolution kernels of various sizes and patterns
  - First 4 magnitudes of Fast Fourier Transform components

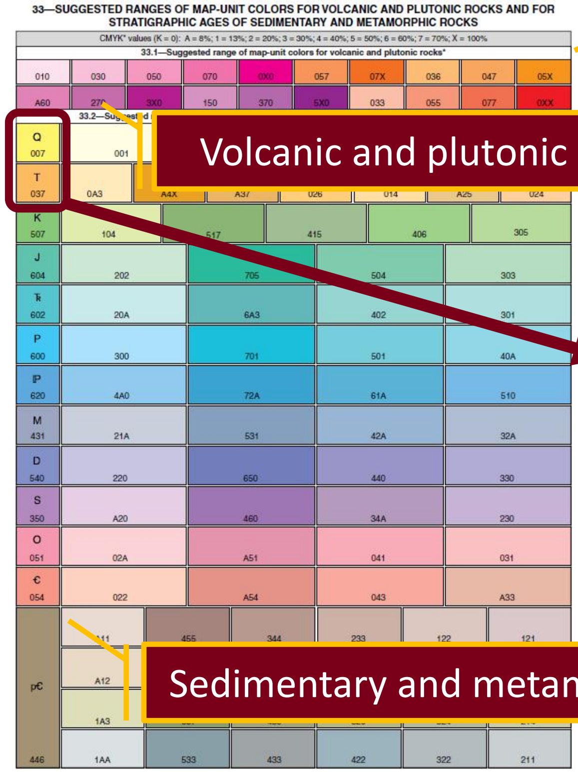


Embed visual appearance and semantic identity, supporting color suggestion and matching

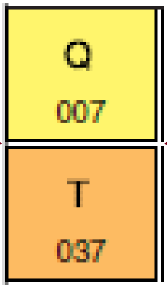
Methodology: Color Schema and Text Convention

- The text for polygonal features in USGS geologic maps has 2 parts
  - Geologic time (geologic age) with some subdivision
  - Rock type (lithology) followed by formation name or relative position

REF NO	STRATIGRAPHIC AGE	SUBDIVISION TYPE	AGE SYMBOL*	KEYBOARD POSITION FOR "FGDCGeoAge" FONT*
32.1	Cenozoic	Era	Cz	{ (left curly bracket = shift–left square bracket)
32.2	Quaternary	Period	Q	No keyboard substitution needed (or, use Helvetica)
32.3	Tertiary	Period	T	No keyboard substitution needed (or, use Helvetica)
32.4	Neogene	Subperiod	N	No keyboard substitution needed (or, use Helvetica)
32.5	Paleogene	Subperiod	Pz	: (colon = shift–semi-colon)
32.6	Mesozoic	Era	Mz	} (right curly bracket = shift–right square bracket)
32.7	Cretaceous	Period	K	No keyboard substitution needed (or, use Helvetica)
32.8	Jurassic	Period	J	No keyboard substitution needed (or, use Helvetica)
32.9	Triassic	Period	Tr	^ (caret = shift-6)
.10	Paleozoic	Era	Pz	| (vertical line = shift-backslash)
.11	Permian	Period	P	No keyboard substitution needed (or, use Helvetica)
.12	Carboniferous	Period	C	No keyboard substitution needed (or, use Helvetica)
.13	Pennsylvanian	Period	ℙ	* (asterisk = shift-8)
32.14	Mississippian	Period	M	No keyboard substitution needed (or, use Helvetica)
32.15	Devonian	Period	D	No keyboard substitution needed (or, use Helvetica)
32.16	Silurian	Period	S	No keyboard substitution needed (or, use Helvetica)
32.17	Ordovician	Period	O	No keyboard substitution needed (or, use Helvetica)
32.18	Cambrian	Period	Є	_ (underscore = shift-hyphen)
32.19	Precambrian	Era	pЄ	= (equal sign)
32.20	Proterozoic	Eon	P	< ("less than" sign = shift-comma)
32.21	Late Proterozoic	Era	Z	No keyboard substitution needed (or, use Helvetica)
32.22	Middle Proterozoic	Era	Y	No keyboard substitution needed (or, use Helvetica)
32.23	Late Middle Proterozoic	Era	Y^3	` (accent grave)
32.24	Middle Middle Proterozoic	Era	Y^2	~ (shift–accent grave)
32.25	Early Middle Proterozoic	Era	Y^1	! (exclamation point = shift-1[one])
32.26	Early Proterozoic	Era	X	No keyboard substitution needed (or, use Helvetica)
32.27	Late Early Proterozoic	Era	X^3	@ ("at" sign = shift-2)
32.28	Middle Early Proterozoic	Era	X^2	# (pound sign = shift-3)
32.29	Early Early Proterozoic	Era	X^1	$ (dollar sign = shift-4)
32.30	Archean	Eon	A	No keyboard substitution needed (or, use Helvetica)
32.31	Late Archean	Era	W	No keyboard substitution needed (or, use Helvetica)
32.32	Middle Archean	Era	V	No keyboard substitution needed (or, use Helvetica)
32.33	Early Archean	Era	U	No keyboard substitution needed (or, use Helvetica)
32.34	pre-Archean	Eon	pA	> ("greater than" sign = shift-period)



The color schema only serves as a soft constraint since there may be overwhelming similar keys

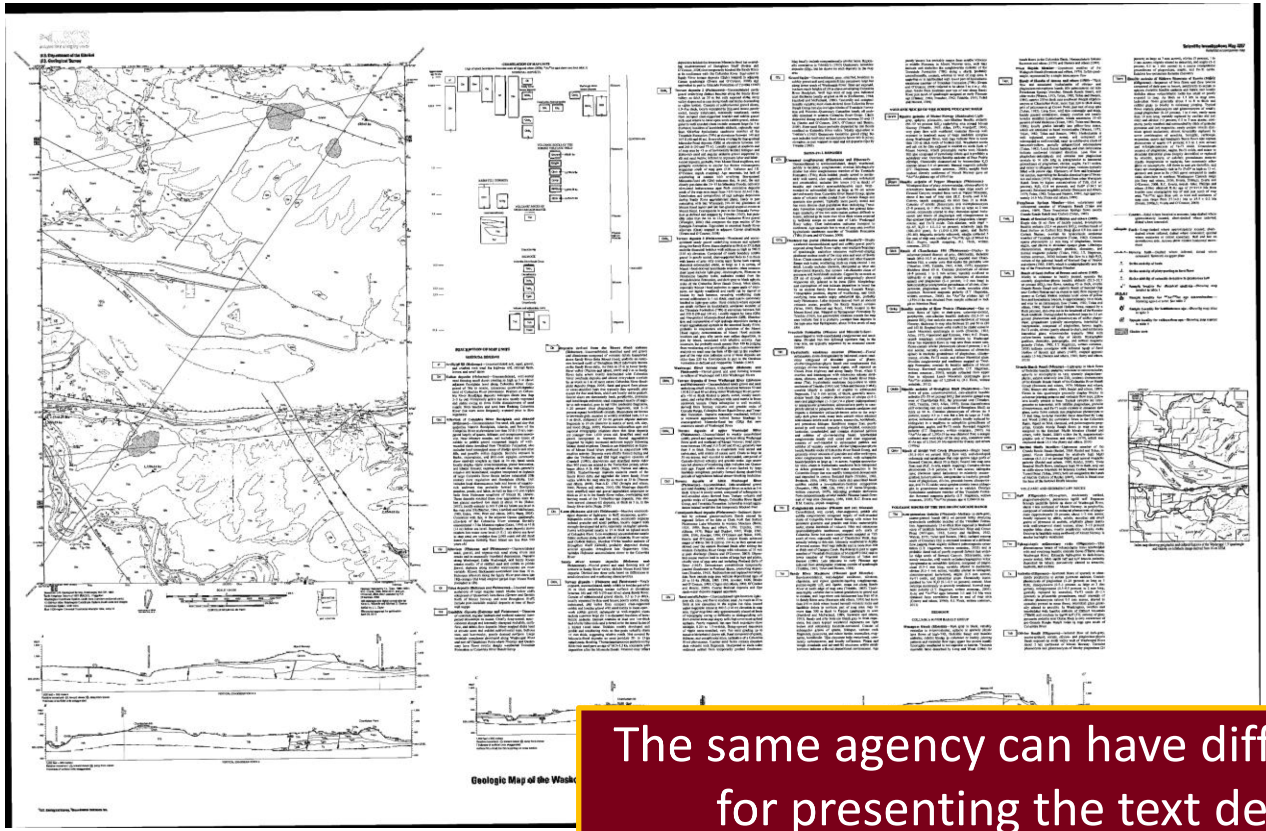


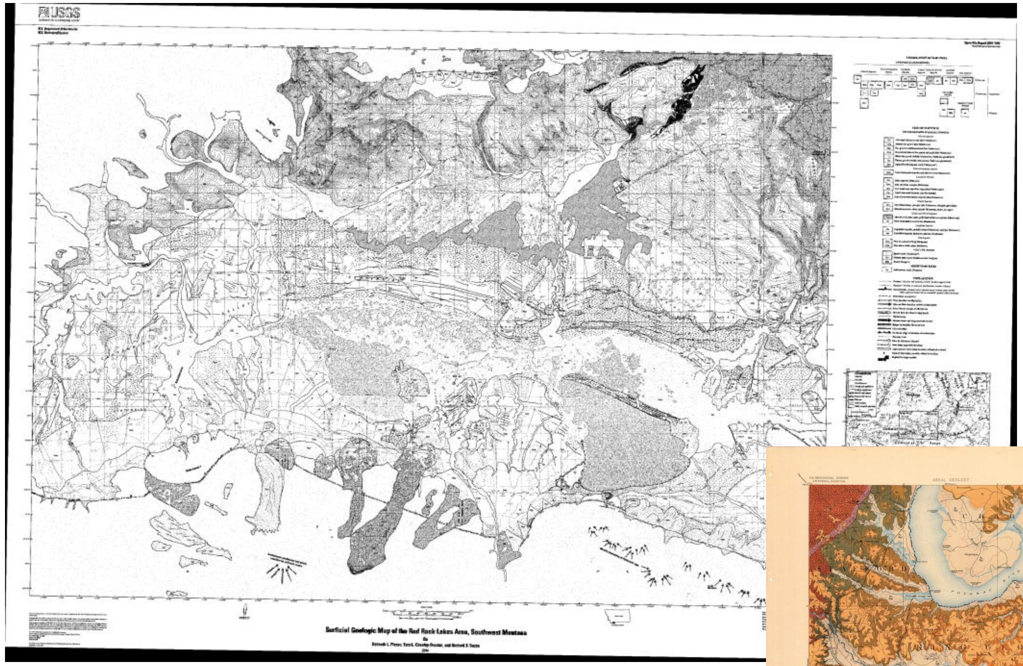
STRATIGRAPHIC AGE	SUBDIVISION TYPE	AGE SYMBOL*
Cenozoic	Era	Cz
Quaternary	Period	Q
Tertiary	Period	T
Neogene	Subperiod	N
Paleogene	Subperiod	Pz
Mesozoic	Era	Mz
Cretaceous	Period	K

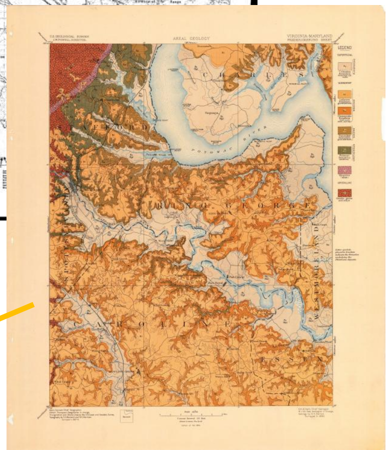
Reference:
FGDC Digital Cartographic Standard for Geologic Map Symbolization.
Geologic Data Subcommittee, Federal Geographic Data Committee, 2006.

Methodology: Why Not Using OCR?

- Information provided in the text description from map legend varies significantly across maps





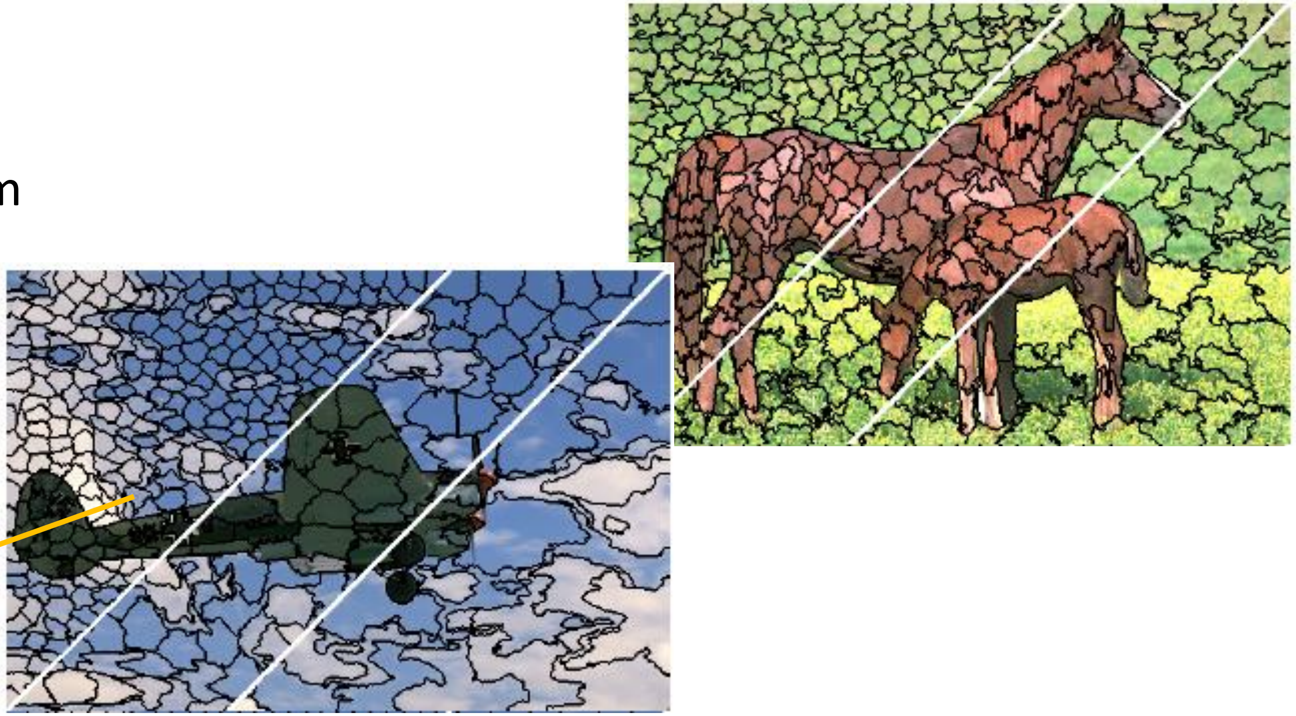


The same agency can have different schema or convention for presenting the text description in map legend

Methodology Ref: Superpixel with fast SLIC

- Simple Linear Iterative Clustering (SLIC)
  - Two parameters for this clustering algorithm
    - Number of centroids (number of polygons)
    - Shape compactness (less color oriented)

The three slices have settings from left to right with a decreased number of centroids



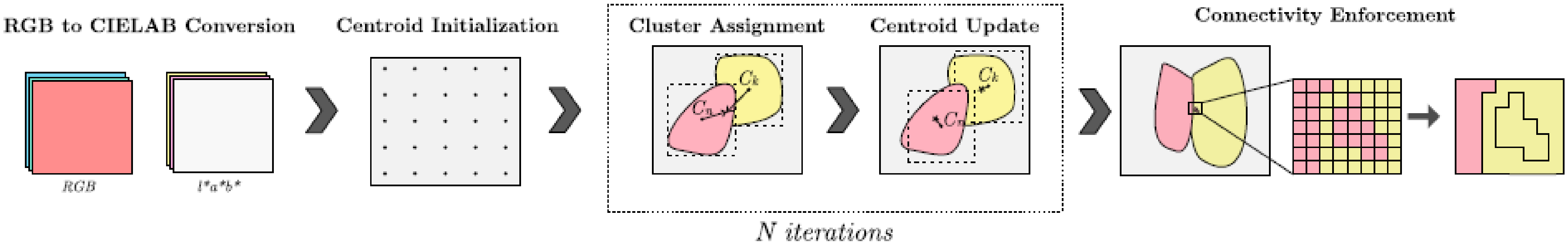
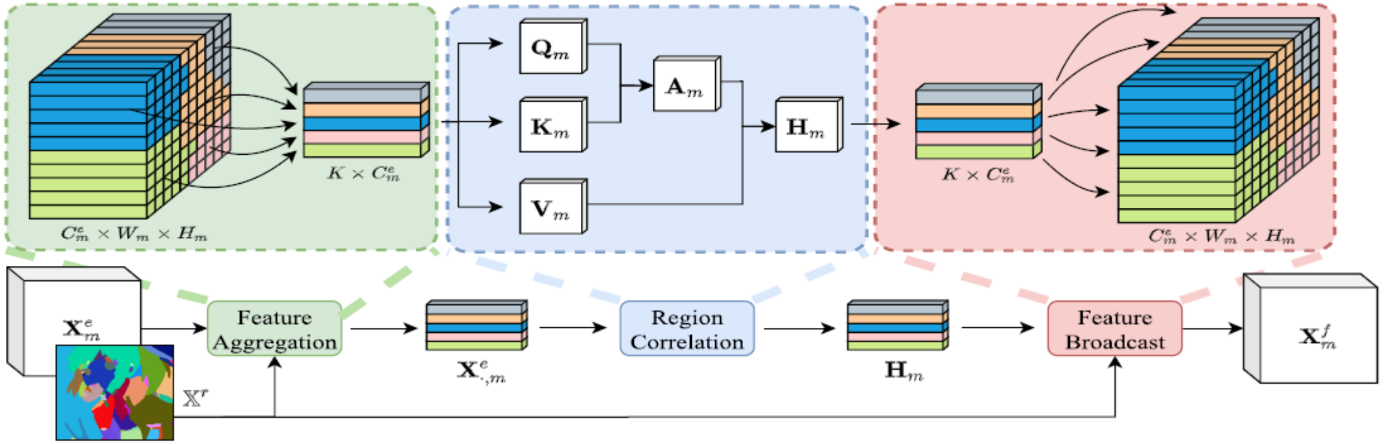
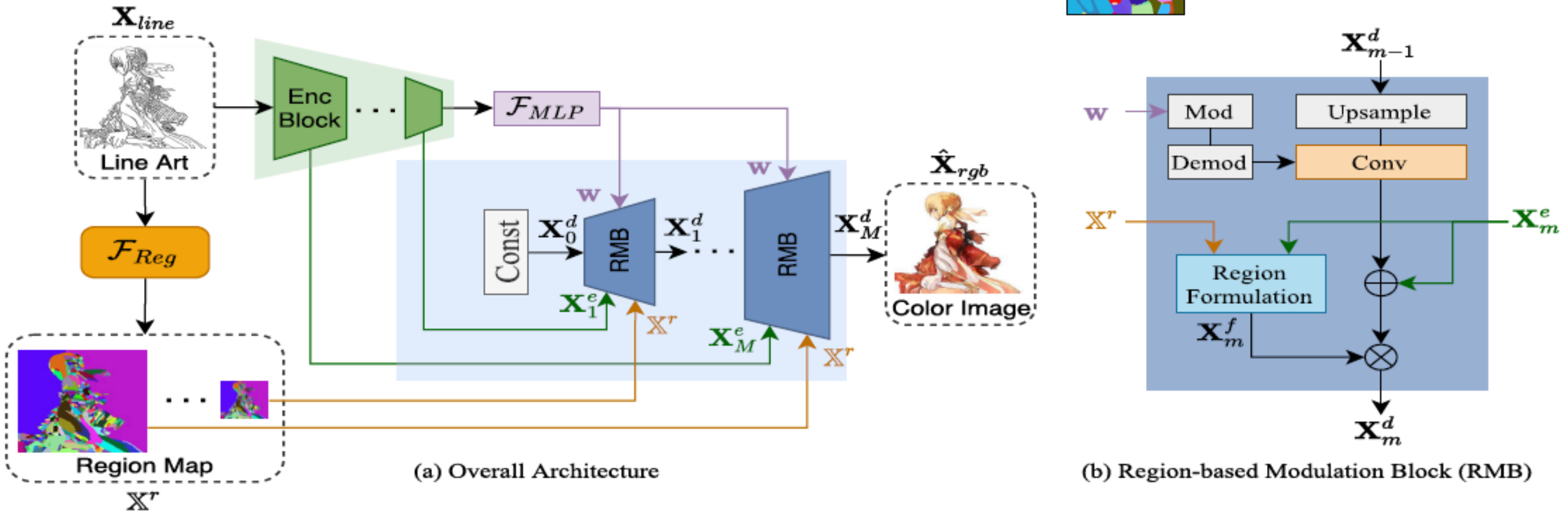


Image Source for SLIC: FastSLIC: Optimized SLIC Superpixel. preprint, 2020.
Along with the github page/ source code: https://github.com/Algy/fast-slic

Methodology Ref: Region Map Colorization

- Enforcing the single-color assignment for each polygon in the region map
  - Generate region map from line art
  - Treat it as a hard constraint



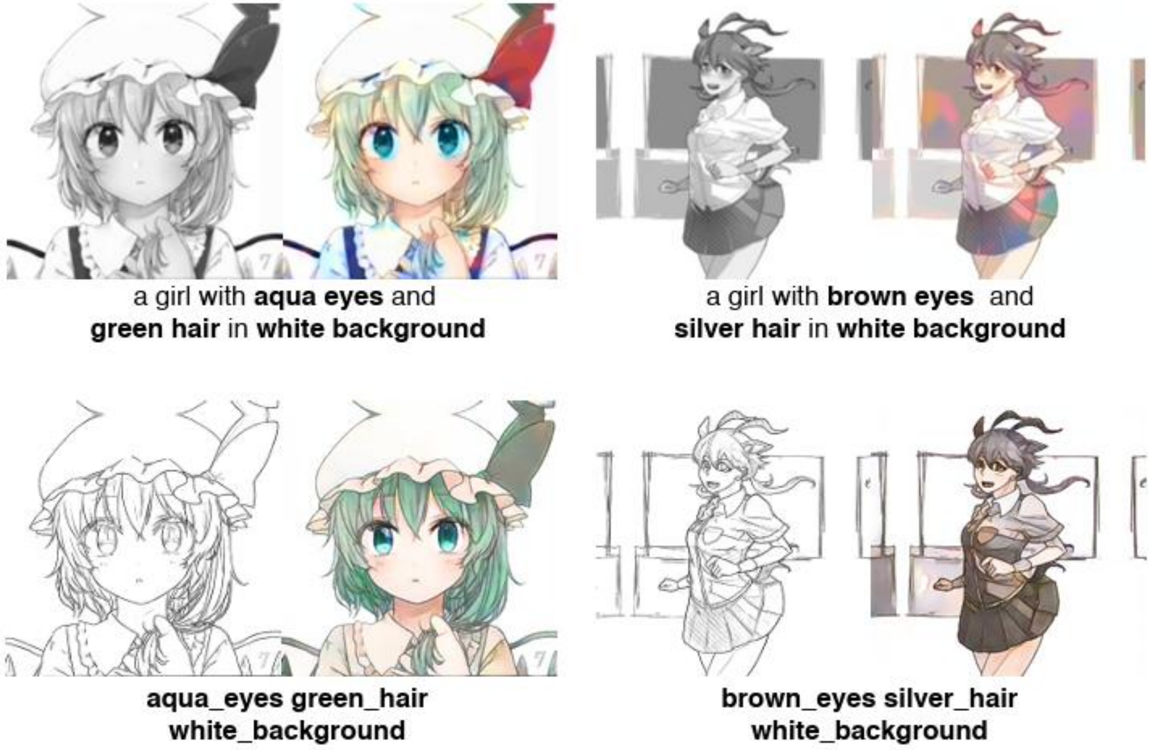


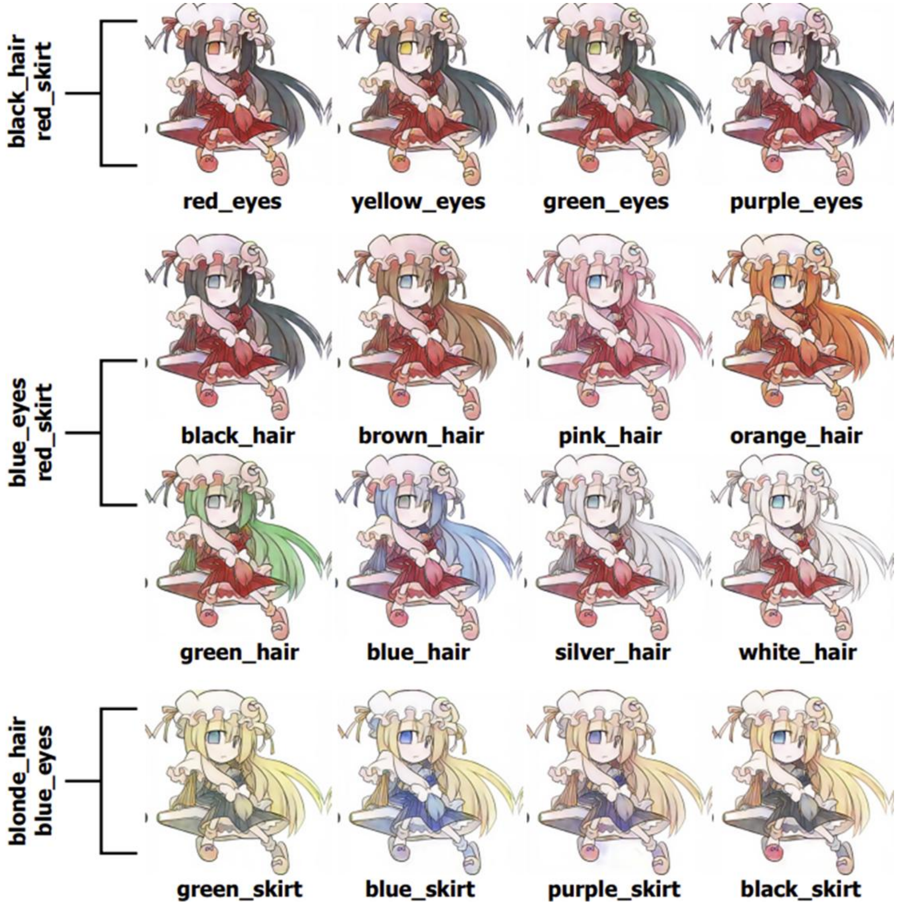
(a) Overall Architecture

(b) Region-based Modulation Block (RMB)

Evaluation: Tag-based Line Art Colorization

- Derive clear color separation among regions from the line art
  - Address limited and precise target components in the tag

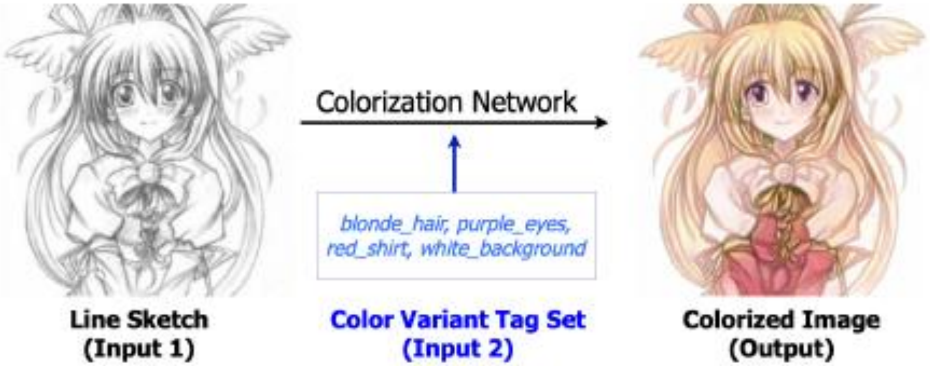


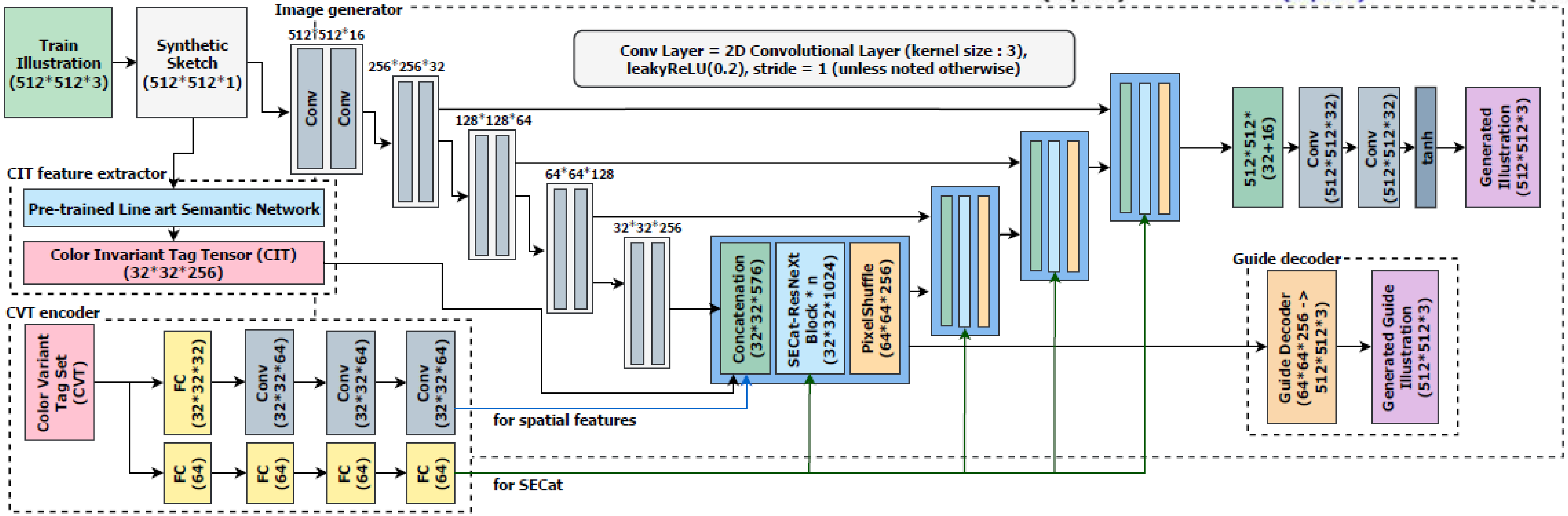


Reference:
Tag2pix: Line art colorization using text tag with secat and changing loss. ICCV, 2019.

Related Work: Tag-based Line Art Colorization

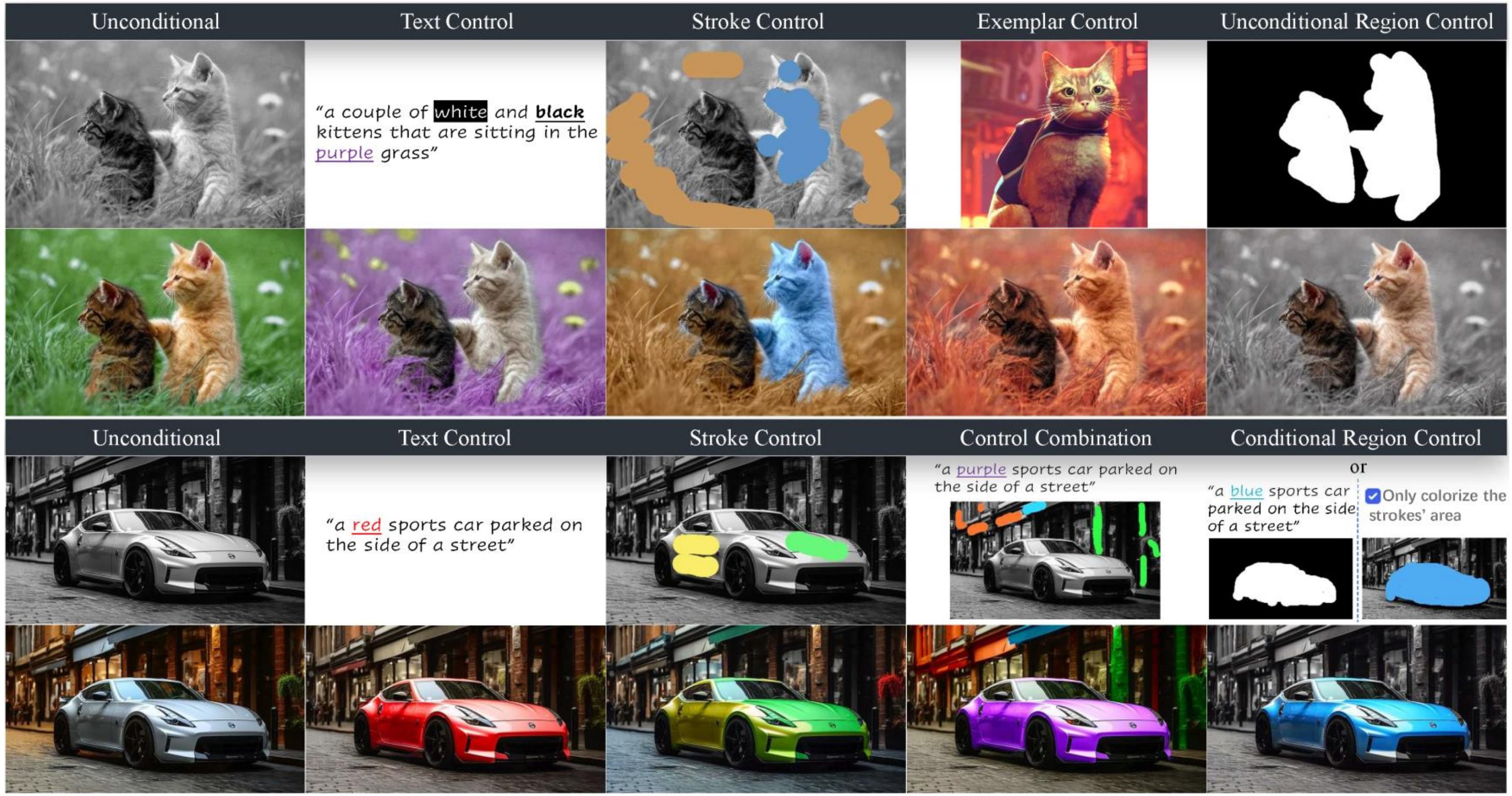
- An encoder-decoder with simple text descriptions
  - Use text tag as a guidance for line-art colorization





Related Work: Interactive Image Colorization

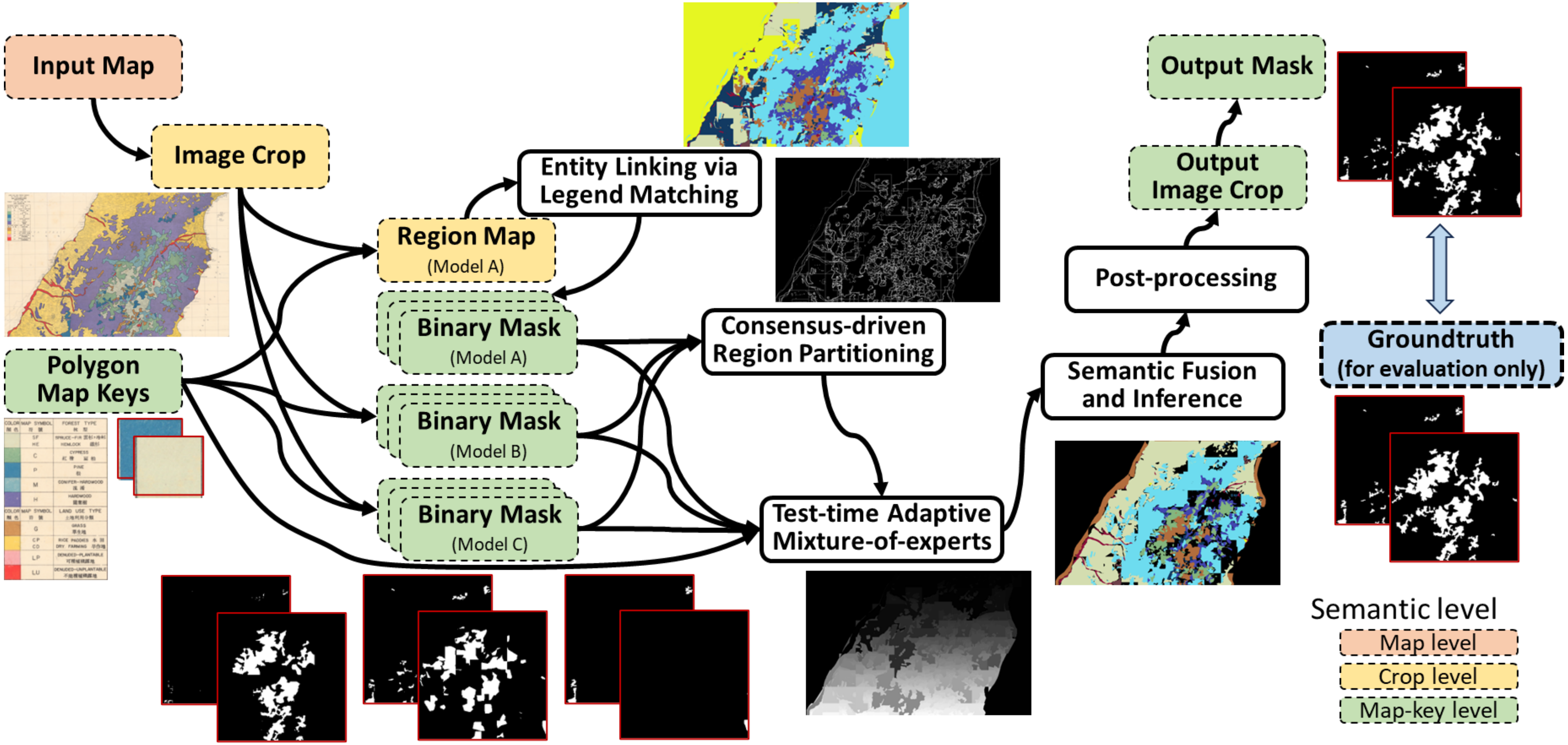
- Apply multimodal diffusion-based framework to take distinct human inputs



Appendix for Generalizing Polygon Digitization

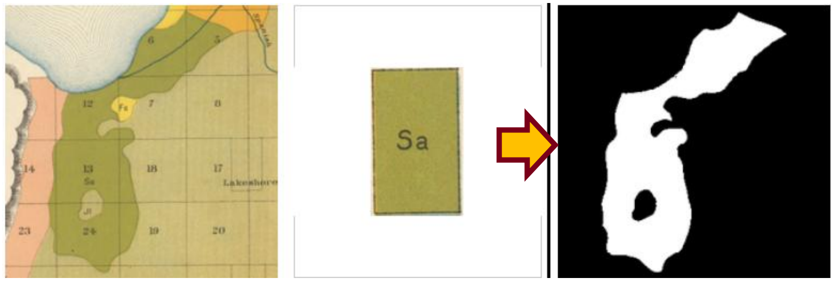
Methodology: Overview and Workflow

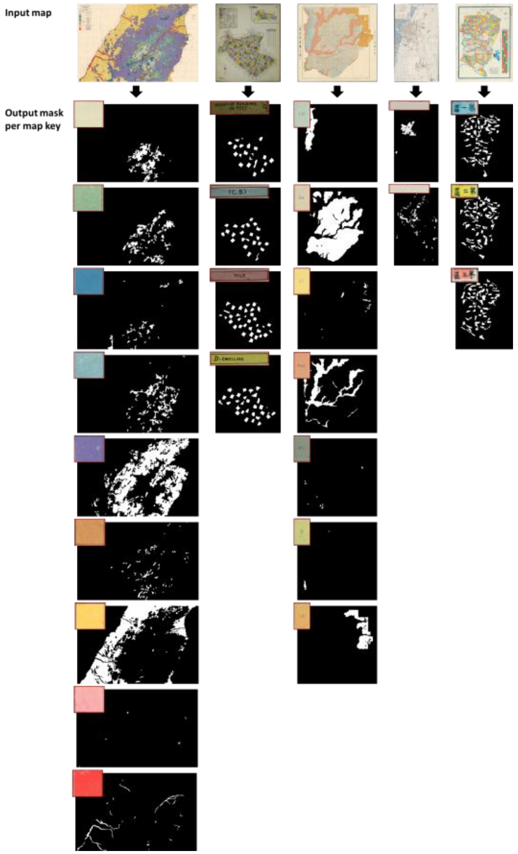
- Reconcile solutions from distinct methods at test time



Methodology: Prompt to Vision-language Models (VLMs)

- Attach the cropped map tile and a key-value set containing the name and image of each polygon map key in the map legend

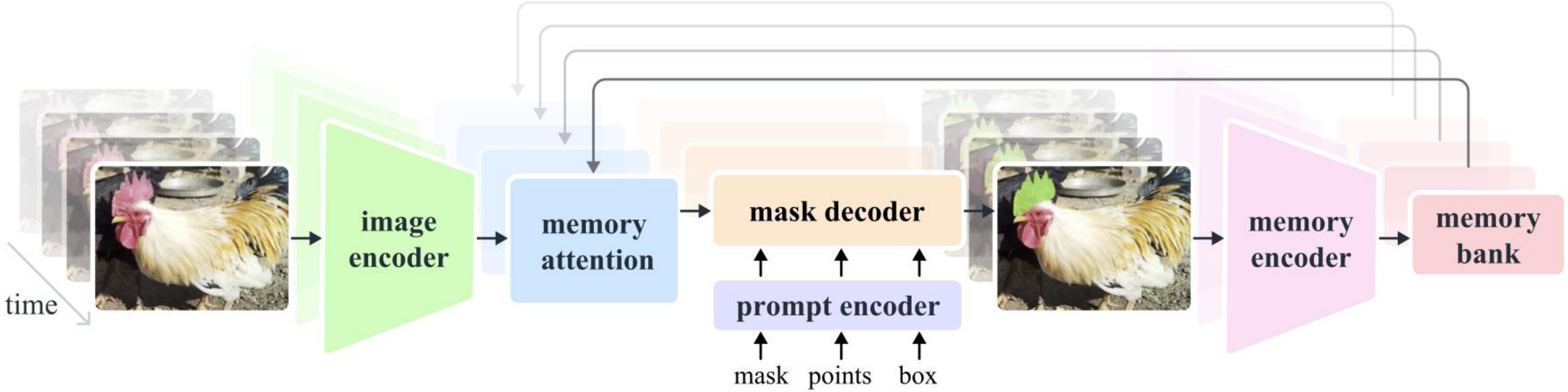





```
**GOAL**
Identify fine-grained polygonal regions in this specific MAP TILE
that match colors and textures for each of the MULTIPLE provided
legend item crops.

**Context**
- This image is a TILE (sub-section) of a larger map.
- Tile Dimensions: {tile_w}px width x {tile_h}px height.
- **Pixel Coordinates**: Return pixel coordinates normalized to a
   {COORD_SCALE}x{COORD_SCALE} scale
relative to the TOP-LEFT of this tile.

**Input Description**
1. Map Tile image. Irrelevant/Out-of-bound areas are black.
2. A list of Legend item crops, each with a unique Name.

**Instructions**
1. For EACH legend item provided:
     a. Understand its major colors, textures, markings, and text
        patterns.
     b. Identify fine-grained regions (pixels) in the map tile that
        match or are similar.
     c. Build fine-grained multi-polygon boundaries for the
        identified regions (can have holes if needed).
2. Organize the results into a JSON object where keys are the
   Legend Names.

**Output JSON Format**
JSON Format: {{
     'LegendName_A': [ [[x1, y1], [x2, y2], ...], ... ],
     'LegendName_B': [ ... ]
}}

**Constraints**
- Do not hallucinate points.
- Ignore masked-out (black) areas.
- If the feature is cut off by the tile edge, trace the edge to
   enclose the multi-polygon boundaries.

**Input Data**
- Map Tile: Provided below.
- Legend Items: Provided below.
```

Methodology: SAM2 with Entity Linking

- We apply SAM2 to **segment anything** from the raster image
  - Conform to its pure design concept and main usage
  - Do not exploit its memory attention for our case
- Then apply polygonal entity linking to polygon map key(s) with a color threshold
  - Viewed as an **advanced color thresholding and matching**

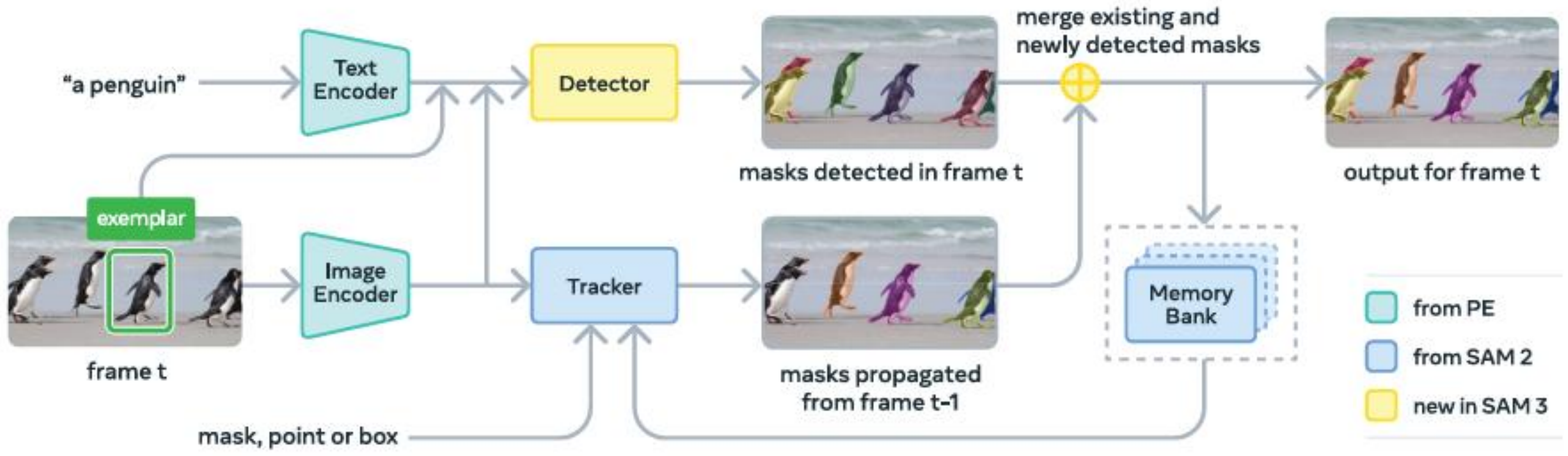


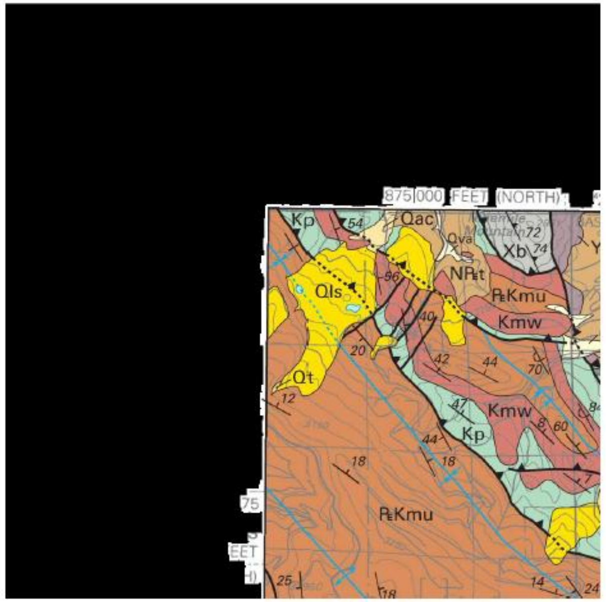
Reference:
Sam 2: Segment anything in images and videos. arXiv, 2024.

Methodology Ref: Prompt to SAM3

- SAM3 can directly take image exemplar as **concept input**
  - Provide ontology and precise information about local instance

- Indicate one polygon map key as a positive image exemplar with the remaining map keys as negative image exemplar

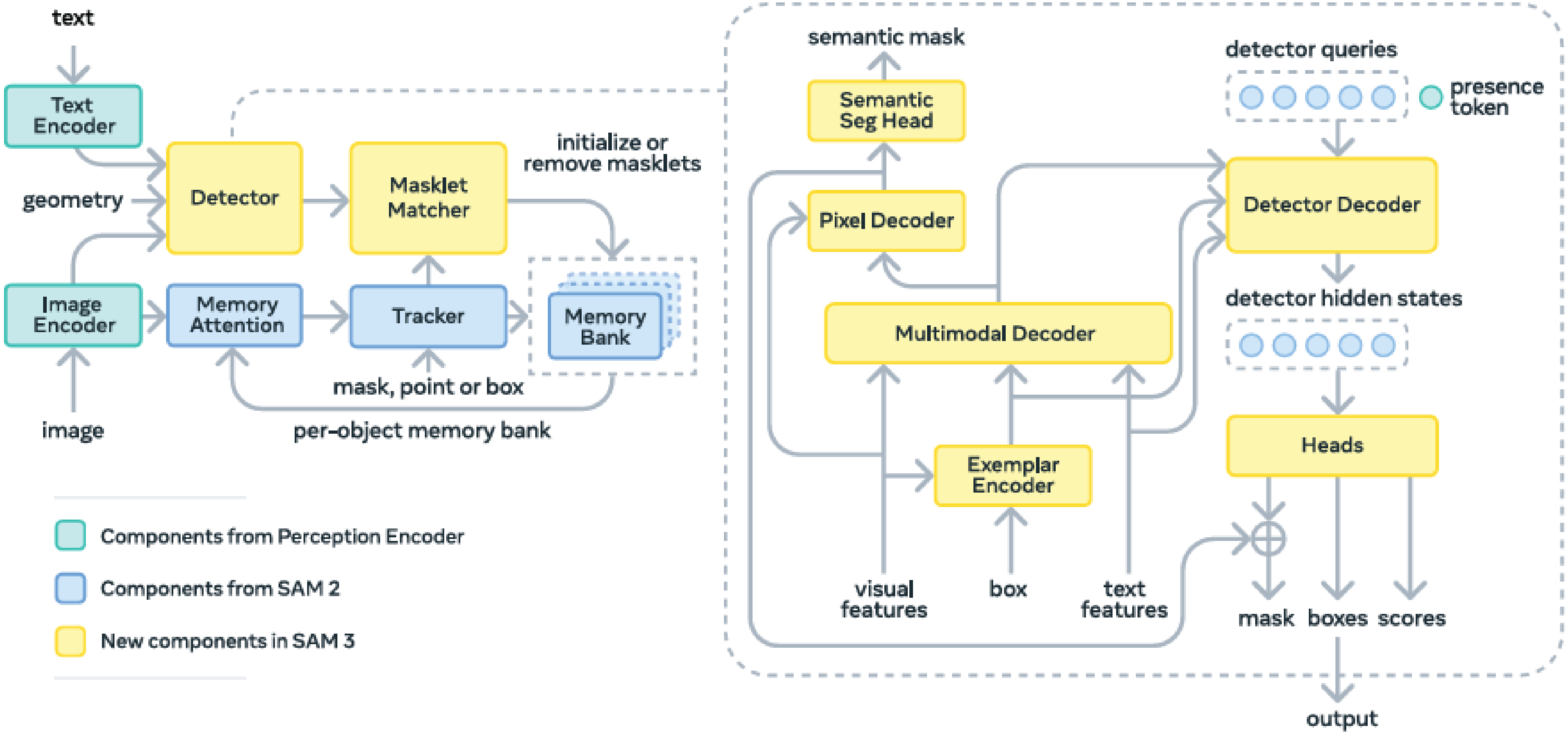




Reference:
Sam 3: Segment anything with concepts. arXiv, 2025.

Methodology Ref: SAM3

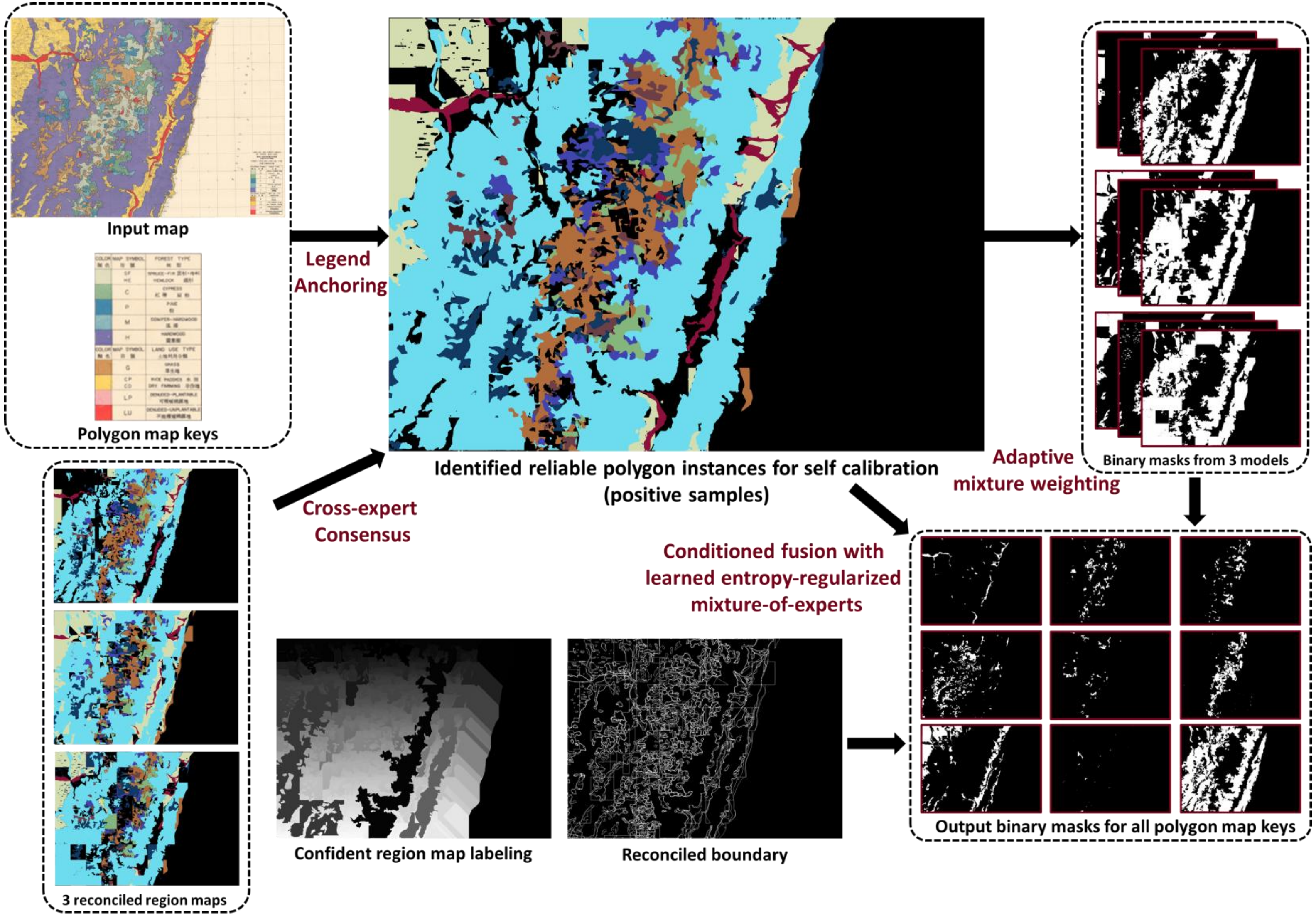
- SAM3 is built on SAM2 and extends to take image exemplar as its input



Reference:
Sam 3: Segment anything with concepts. arXiv, 2025.

Methodology: Test-time Adaptive Self Calibration

- Legend anchor & Pseudo label for gating



Methodology: Detailed Workflow

- Legend-based contrastive learning
  1. Feature extraction from map keys and minimal polygon instances
     - Derive region embeddings $\boldsymbol{z_i}$ for each region $\boldsymbol{r_i}$ and legend embeddings $\boldsymbol{z_k^{(L)}}$ (anchors) for each key $\boldsymbol{l_k}$
       – Statistics, color histogram, and color gradient in CIELAB, RGB, and HSV color spaces
  2. Processing expert predictions with vote aggregation
     - Let $v_i^{(A)}, v_i^{(B)}, v_i^{(C)} \in \mathcal{R}^{K+1}$ denote the region-level vote counts from the three experts
       – The index $k \in \{0,1,\ldots,K\}$ corresponds to the class (K map keys) indices, we add one for background class
       – We normalize to obtain the vote count $P_i^{(\cdot)}$ and derive the majority label $\boldsymbol{y_i^{(\cdot)}}$ for for each region $\boldsymbol{r_i}$
         » $P_i^{(A)} = \frac{v_i^{(A)}}{\sum_k v_{ik}^{(A)}},\ P_i^{(B)} = \frac{v_i^{(B)}}{\sum_k v_{ik}^{(B)}},\ P_i^{(C)} = \frac{v_i^{(C)}}{\sum_k v_{ik}^{(C)}}$ and $y_i^{(A)},\ y_i^{(B)},\ y_i^{(C)} \in \{0,1,\ldots,K\}$
  3. Pseudo-label generation from expert agreement
     - We use inter-expert agreement to obtain reliable self-supervision signals at test time
       – A region is considered reliable if at least two experts agree to be assigned a pseudo label $\boldsymbol{\tilde{y}_i}$
         » $y_i^{(A)} = y_i^{(B)}\ or\ y_i^{(B)} = y_i^{(C)}\ or\ y_i^{(C)} = y_i^{(A)}$
       – Regions without sufficient agreement are excluded from pseudo-label supervision

Methodology: Detailed Workflow

- Legend-based contrastive learning
  4. Representation learning via contrastive loss
     - We perform lightweight representation learning to align region features with legend anchors
       – We have contrastive loss
         » $\mathcal{L}_{contrast} = -log \frac{\sum_{j \in P(i)} \exp(sim(z_i, z_j)/T)}{\sum_{j \neq i} \exp(sim(z_i, z_j)/T)}$
         » $P(i)$ is the set of samples (**map-key pixels** or **pseudo-labeled regions**) sharing the same polygon map key as sample i
         » $sim(\cdot,\cdot)$ is cosine similarity
         » T is a temperature parameter that controls the concentration of the contrastive distribution; it is set to 0.2
     - We use two types of samples
       – Pixels extracted from polygon map keys with known labels (**map-key pixels** for $\mathcal{L}_{legend}$)
       – Polygonal instances with pseudo labels obtained from expert agreement (**pseudo-labeled regions** for $\mathcal{L}_{region}$)
         » $\mathcal{L}_{legend} = \frac{1}{|S_{legend}|} \sum_{i \in S_{legend}} \mathcal{L}_{contrast}(i)$ and $\mathcal{L}_{region} = \frac{1}{|S_{region}|} \sum_{i \in S_{region}} \mathcal{L}_{contrast}(i)$
       – We use the combined loss for the overall **representation learning objective**
         » $\mathcal{L}_{train} = \mathcal{L}_{legend} + \lambda_{region} \mathcal{L}_{region}$
         » λ_{region} is a parameter to control the weight of the two losses; it is set to 0.5

Methodology: Detailed Workflow

- Legend-based contrastive learning
  5. Adaptive gating with regularization
     - We employ a gating module that produces weights for different experts per image
       – We have the input feature vector
         » $x = [x_{style},\ x_{agreement}]$
         » x_{style} consists of **global image statistics**, including the mean and standard deviation in various color spaces, etc.
         » $x_{agreement}$ consists of **expert confidence and consistency**, including the mean and median of voting across experts.
       – With the goal for output of
         » $w = (w_A, w_B, w_C), \quad \sum_e w_e = 1 \quad \text{for} \quad e \in \{A, B, C\}$ that index the expert solutions
     - We introduce an entropy-based regularization term to control the sharpness of expert weighting
       – We have the **gating-network learning objective**
         » $\mathcal{L}_{gate} = \lambda_{gate} |H(w) - \tau_g|$
         » $H(w) = -\sum_e w_e \log w_e$
         » $\tau_g = \tau_{base} + \tau_{scale} \cdot \frac{1}{|\mathcal{R}|} \sum_i \frac{1}{3} (1 \left[y_i^{(A)} \neq y_i^{(B)} \right] + 1 \left[y_i^{(B)} \neq y_i^{(C)} \right] + 1[y_i^{(C)} \neq y_i^{(A)}])$
         » The latter part is disagreeing rate
         » τ_{base} is base entropy and set to 0.4, τ_{scale} is expert entropy and set to 0.8, λ_{gate} is gating parameter and set to 0.1

Methodology: Detailed Workflow

- Legend-based contrastive learning

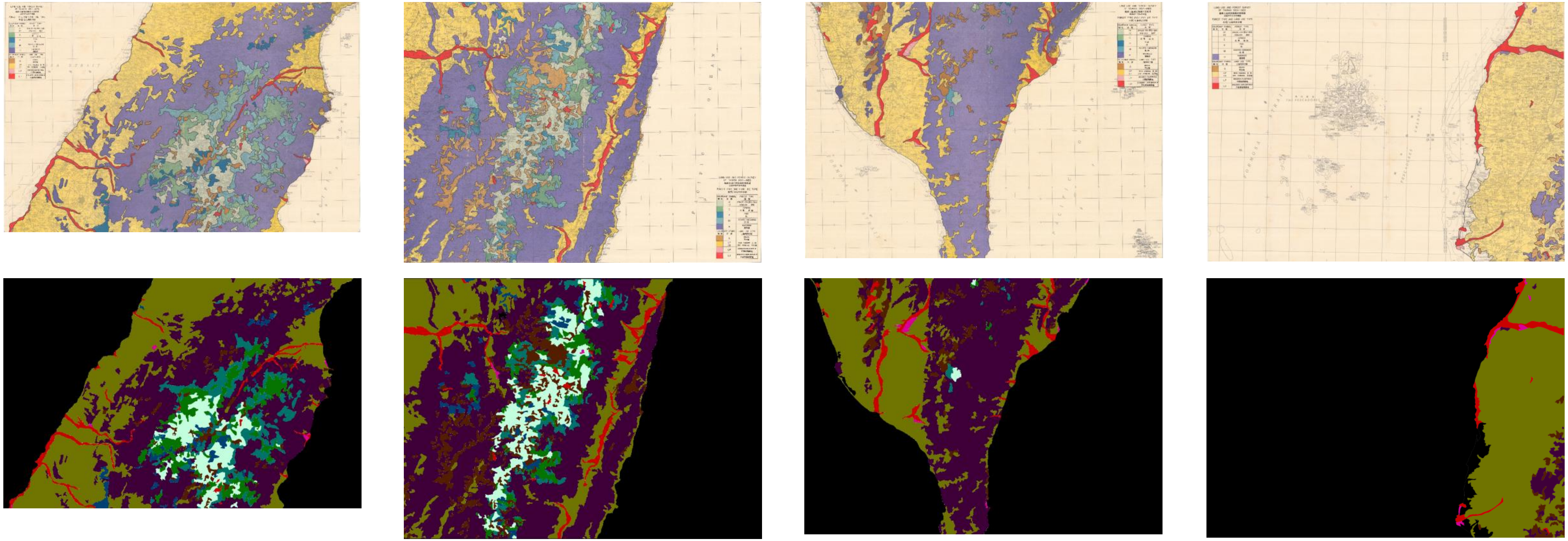
6. The “learning” part
   - We have two learning targets
     – **Representation learning objective**
       » For contrastive legend anchor to update the embeddings $\boldsymbol{z_i}$ and $\boldsymbol{z_k^{(L)}}$
       » $\mathcal{L}_{train} = \mathcal{L}_{legend} + \lambda_{region}\mathcal{L}_{region}$
     – **Gating-network learning objective**
       » For gating the calibration signal to the fused prediction $w = (w_A, w_B, w_C)$
       » $\mathcal{L}_{gate} = \lambda_{gate}|H(w) - \tau_g|$
   - We use a lightweight shallow neural network with a single hidden layer and a softmax
     – This gating module is jointly optimized with the representation learning component during test-time adaptation
       » We train to minimize $\mathcal{L}_{train} + \mathcal{L}_{gate}$ with the gradients applied to their respective parameters
       » It can adaptively adjust expert contributions based on both global style variations and the reliability of expert predictions

Methodology: Detailed Workflow (cont.)

- Legend-based contrastive learning
  7. Adaptive fusion with similarity-based refinement
     - Two levels: **legend-statistic-driven similarity anchor**, and **expert-and-representation-driven fusion**
       – We compute similarity between region embeddings and legend anchor
         » $S_{ik} = sim(\boldsymbol{z_i}, \boldsymbol{z_k^{(L)}})$
       – We derive the fused prediction for each region (minimal polygon instance)
         » $p_i = w_a p_i^{(A)} + w_b p_i^{(B)} + w_c p_i^{(C)}$
         » $p_i \in \mathcal{R}^{K+1}$ is the fused class score vector, and p_{ik} is its k-th entry (the k-th class) corresponding to $\boldsymbol{l_k}$
       – We then obtain foreground score for region $\boldsymbol{r_i}$ to key $\boldsymbol{l_k}$
         » $score_{ik} = p_{ik} + \lambda_{sim} S_{ik}$
         » λ_{sim} is a parameter and set to 1.0 for balanced scoring
     - We rely on the score to determine background assignment
       – We compute background score simply as $p_{i0} + \lambda_{bg}$, with λ_{bg} being background buffer and set to 0.25
       – A region is assigned to the background class if
         » $p_{i0} + \lambda_{bg} > \max(score_{ik})$ for all k>0, or $\max(score_{ik}) < \tau_{bg}$ and $p_{i0} > (1 - \tau_{bg})$
         » τ_{bg} is rejection threshold and set to 0.8

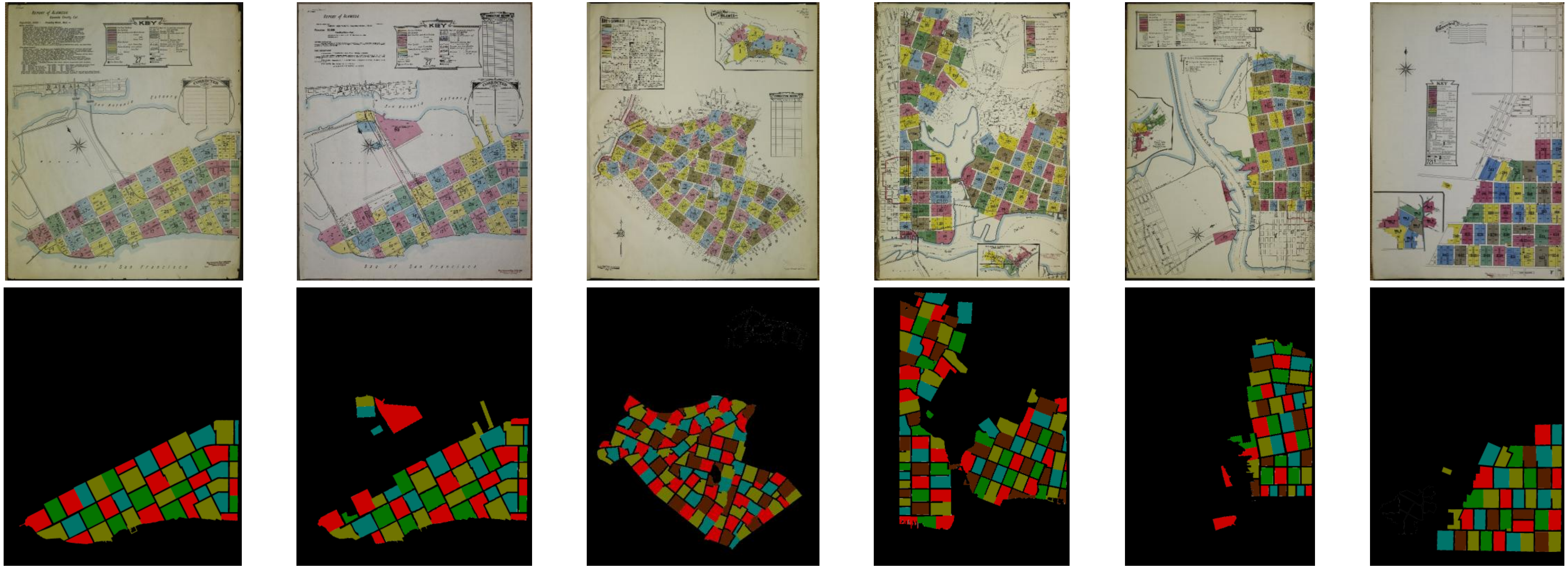
Dataset: Some Cases with its Annotation

- Forest Type (FT) 1954 – 1955
  - Thick polygon boundaries to indicate gradual changes and separation from mountainous areas
  - Support long-term environmental monitoring and landscape change research



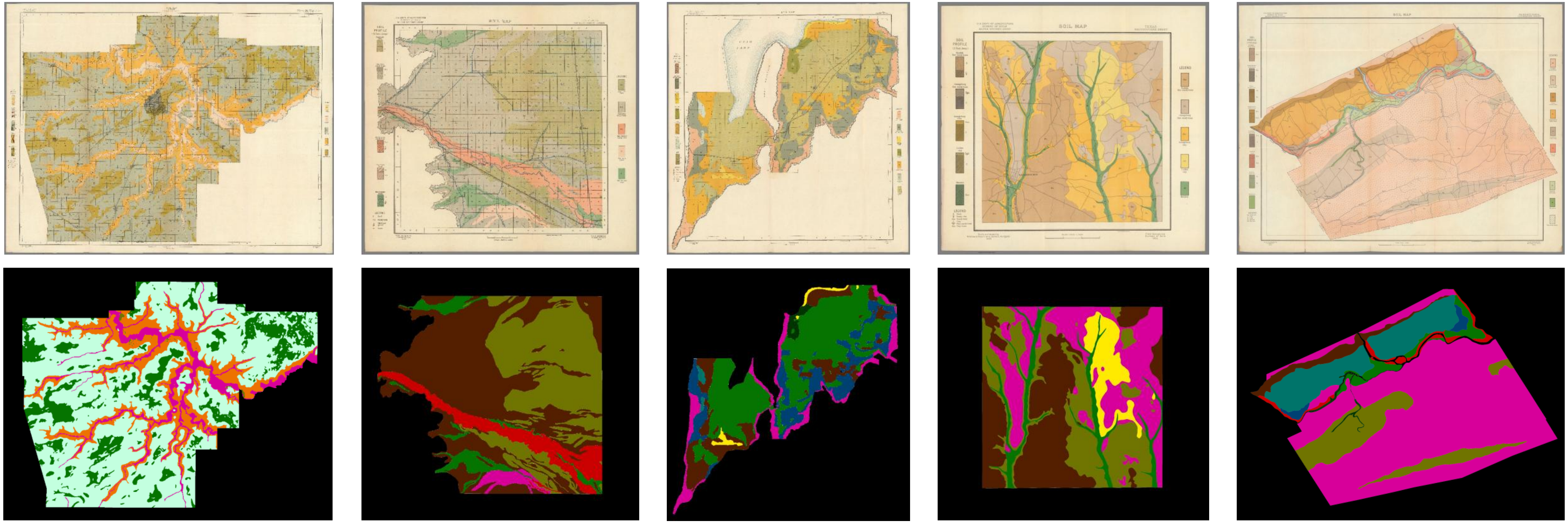
Dataset: Some Cases with its Annotation (cont.)

- Sandborn (SA) 1920 – 1960
  - Information about properties, with overflows to roads, color mismatch, and dense overlap text
  - Support studies or assessment on historical settlement and architecture



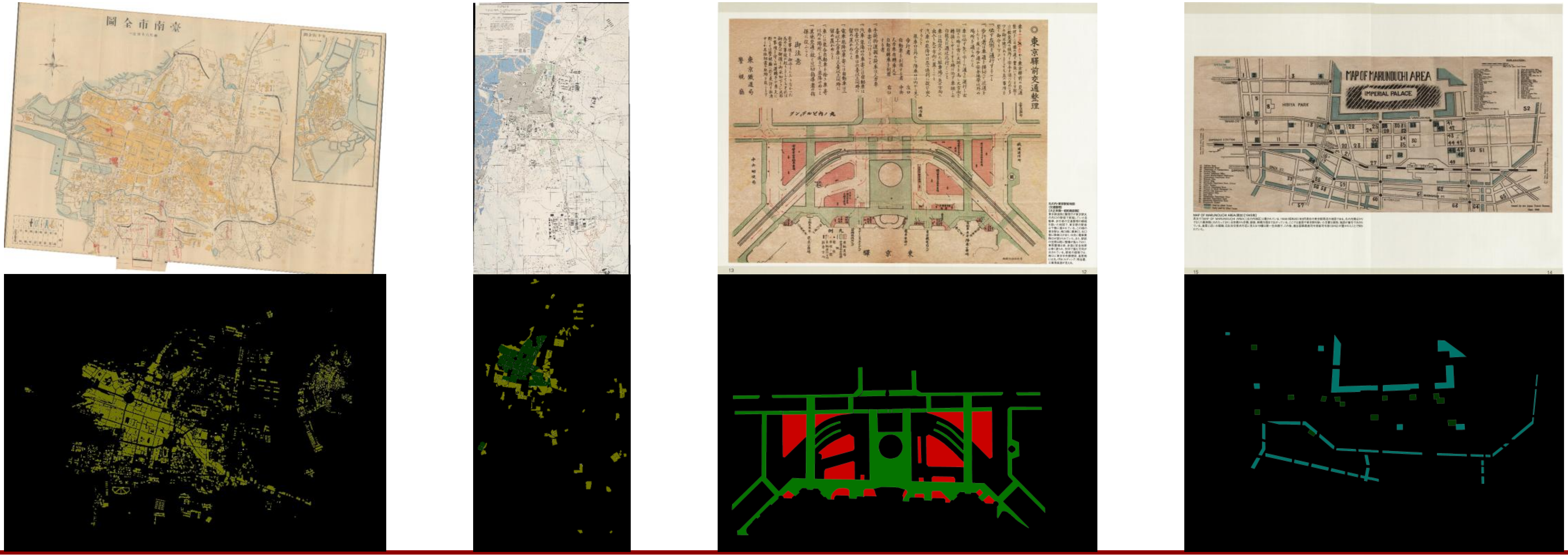
Dataset: Some Cases with its Annotation (cont.)

- Soil (SO) 1903 – 1908
  - Rely on text and patterns for adjacent classes, with ink overflows across dashed boundaries
  - Support large-scale agriculture modeling, infrastructure siting, or hydrologic modeling



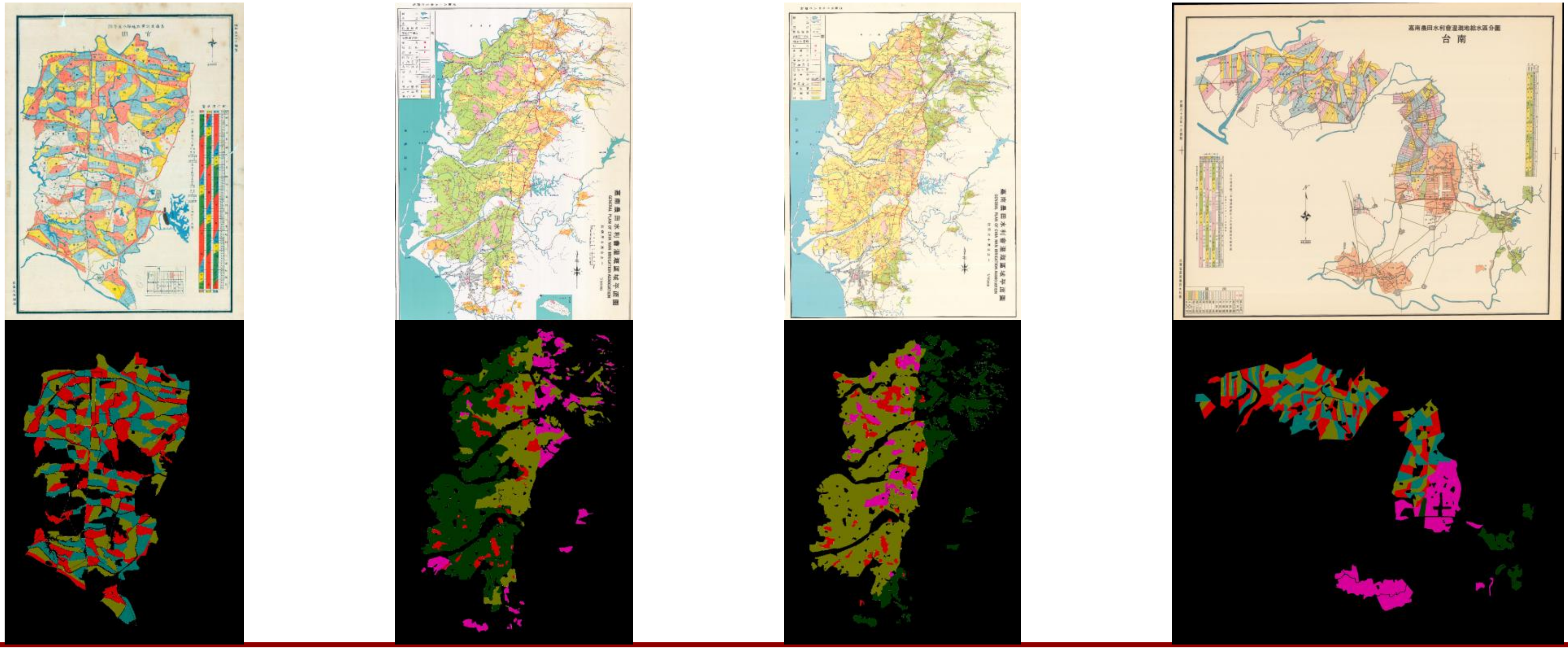
Dataset: Some Cases with its Annotation (cont.)

- Street Plan (SP) 1911 – 1950
  - Record purpose of properties with few map keys, with noticeable CMYK halftone dot patterns under high-resolution scanning and significant ink overflows and severe fading
  - Support studies on historical settlement and urban planning across time and locations

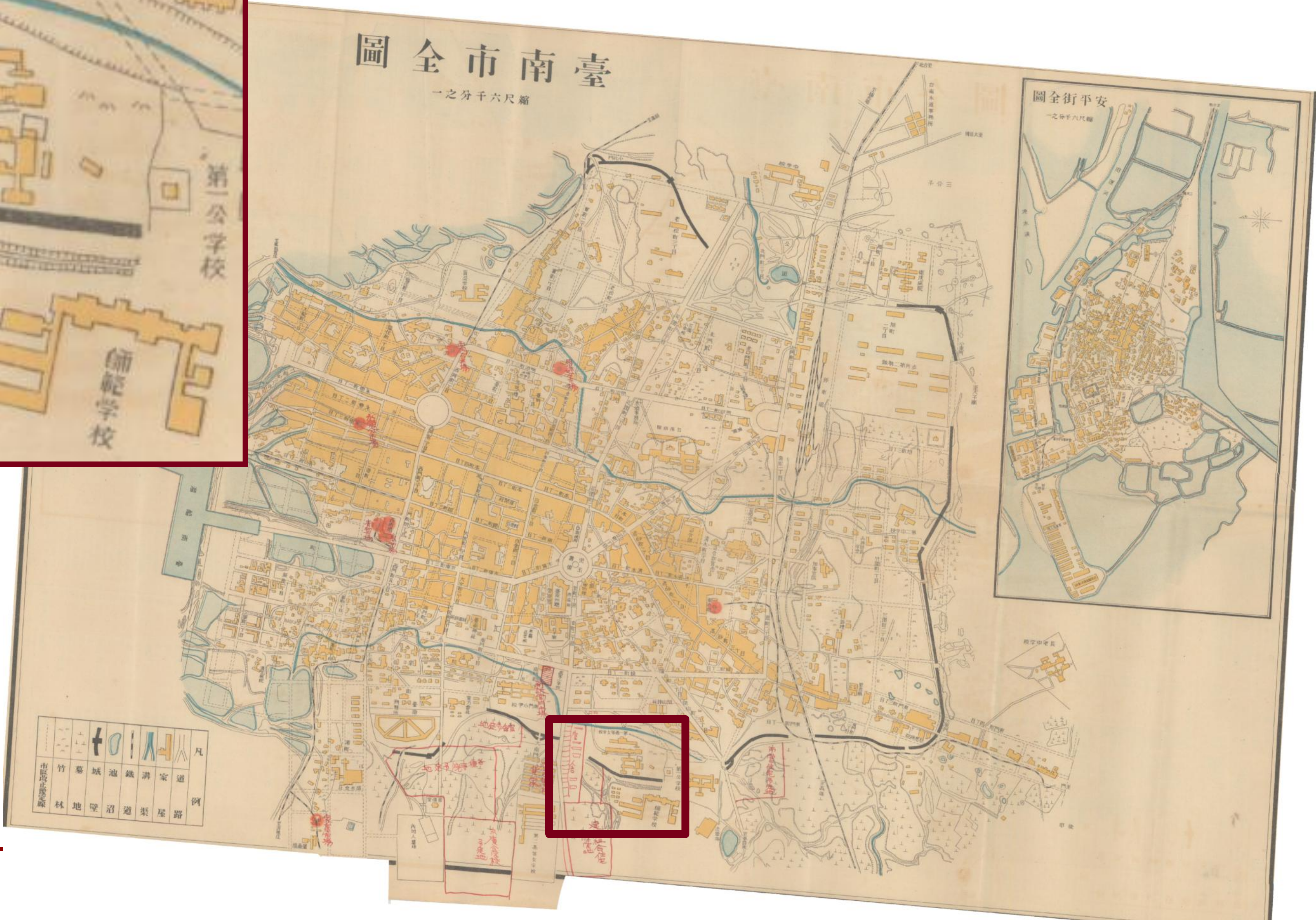


Dataset: Some Cases with its Annotation (cont.)

- Water Resource (WR) 1921 – 1995
  - Irrigation plans for the farmland of regions at different time, with uneven color and ink overflows, and overlaid features (e.g., water channels, and text) across non-explicit polygon boundaries
  - Support studies on historical agricultural planning and settlement across time and locations

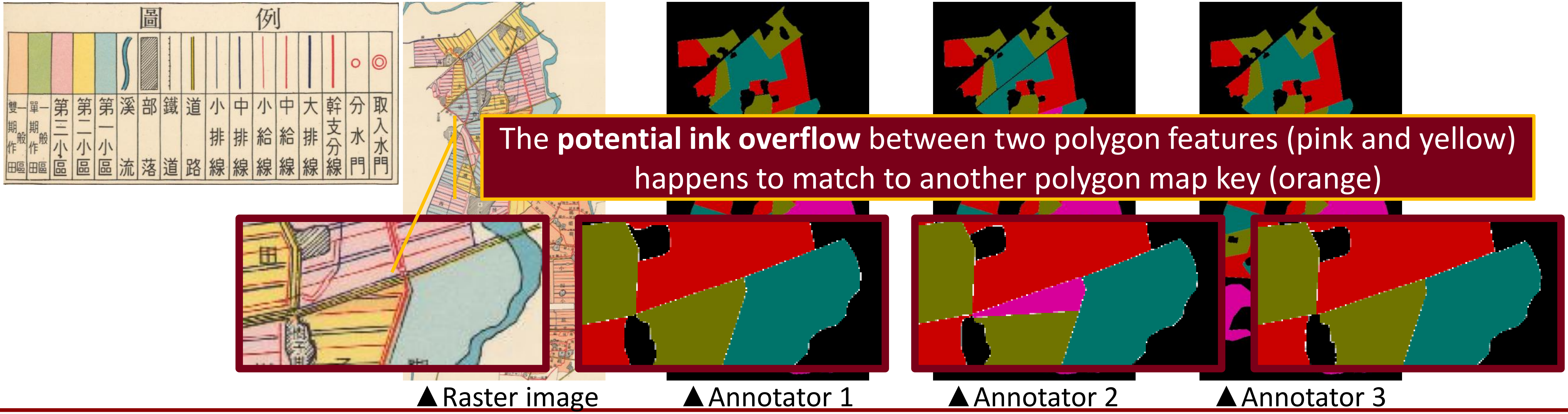



第一高等女学校
第一公学校
師範学校


臺南市全圖
縮尺六千分之一
安平街全圖
凡例
道路
家屋
溝渠
鐵道
池沼
城壁
墓地
竹林
市區改正豫定線

Dataset: Annotation

- A unified guideline is agreed upon prior to annotation
  - covering polygon inclusion criteria, boundary interpretation under degradation, nested or overlapping regions, and ambiguous areas
  - Annotations are conducted independently following these shared instructions
- At least one map from each dataset is annotated by all three annotators.



Evaluation: MMPQ, F1@8, and NBDR Metrics

- MMPQ (↑): Instance-based accuracy
  - Many-to-many panoptic quality
    - Construct a bipartite graph between ground truth and extraction polygon instances
    - Compute Intersection-over-union (IoU) for the linked polygon instances
    - Follow panoptic quality (PQ) to define area-weighted average IoU and component-level F-score

Instance-based accuracy
that combines IoU and F1 score

$$\mathrm{IoU}_c = \frac{|G_c \cap P_c|}{|G_c \cup P_c|}.$$

$$\mathrm{SQ}_w = \frac{\sum_{c \in TP} w_c \cdot \mathrm{IoU}_c}{\sum_{c \in TP} w_c}, \quad w_c = |G_c \cup P_c|;$$

$$\mathrm{RQ}_w = \frac{\sum_{c \in TP} w_c}{\sum_{c \in TP} w_c + \frac{1}{2} \sum_{c \in FP} w_c + \frac{1}{2} \sum_{c \in FN} w_c}.$$

$$\mathrm{MMPQ} = \mathrm{SQ}_w \cdot \mathrm{RQ}_w.$$

Evaluation: MMPQ, F1@8, and NBDR Metrics (cont.)

- F1@8 (↑): pixel-based accuracy
  - Accuracy on polygon geometric alignment with tolerance to small boundary shifts
    - A predicted pixel is considered correct if its Euclidean distance to the nearest ground truth pixel is within a tolerance radius (e.g., 8)
    - The remaining computation follows the general definition of F1 score

Pixel-based accuracy with general F1 score

$$\text{Precision}_r = \frac{\text{TP}_r}{\text{TP}_r + \text{FP}_r},$$

$$\text{Recall}_r = \frac{\text{TP}_r}{\text{TP}_r + \text{FN}_r},$$

$$\text{F1}_r = \frac{2 \cdot \text{Precision}_r \cdot \text{Recall}_r}{\text{Precision}_r + \text{Recall}_r}.$$

Evaluation: MMPQ, F1@8, and NBDR Metrics (cont.)

- NBDR (↓): geometry-based accuracy
  - Estimated post-correction time based on polygon boundary geometry
    - Compute the Average Symmetric Surface Distance (ASSD) between ground truth and extracted polygon boundaries
    - Then normalize it by the size of the polygon boundary geometry

$$\mathrm{ASSD}_i = \frac{1}{|\partial G_i| + |\partial P_i|}\left(\sum_{x\in\partial G_i} \min_{y\in\partial P_i} \|x-y\|_2 + \sum_{y\in\partial P_i} \min_{x\in\partial G_i} \|y-x\|_2\right).$$

Estimated manual correction time based on boundary geometry

$$\mathrm{NBDR}_i = \frac{\mathrm{ASSD}_i}{\sqrt{|G_i|}},$$

Evaluation: Efficiency and Cost

- For API-based methods, their runtime is dominated by external service latency or limitations and may not be algorithmically meaningful

Five out-of-domain datasets:

Method - Tile Size	Avg. Time / Map (min.)	Avg. API Cost / Map (USD)
GLYPH-1024 (Ours)	10.68	0.38
LOAM-1024	**10.12**	N.A.
SAM2-4096	7.08	N.A.
SAM2-2048	**5.37**	N.A.
SAM2-1024	7.55	N.A.
SAM2-0512	28.78	N.A.
SAM2-0256	101.23	N.A.
SAM3-4096	**0.52**	N.A.
SAM3-2048	0.75	N.A.
SAM3-1024	1.68	N.A.
SAM3-0512	5.55	N.A.
SAM3-0256	27.37	N.A.
Gemini-3-flash-4096	**1.93**	**0.04**
Gemini-3-flash-2048	2.95	0.12
Gemini-3-flash-1024	3.58	0.38
Gemini-3-flash-0512	4.90	0.77
Gemini-3-flash-0256	28.04	6.92
Gemini-3.1-pro-4096	**4.34**	**0.34**
Gemini-3.1-pro-2048	14.77	1.13
Gemini-3.1-pro-1024	8.50	6.98
Gemini-3.1-pro-0512	13.02	17.96
Gemini-3.1-pro-0256	28.12	65.76
Gemini-2.5-pro-4096	17.82	**0.39**
Gemini-2.5-pro-2048	9.37	0.86
Gemini-2.5-pro-1024	**5.37**	3.14
Gemini-2.5-pro-0512	11.52	6.36
Gemini-2.5-pro-0256	36.66	57.18
GPT-4o-4096	5.42	0.31
GPT-4o-2048	2.48	**0.19**
GPT-4o-1024	2.82	0.58
GPT-4o-0512	**1.20**	1.54
GPT-4o-0256	2.17	10.38
GPT-4o-0128	15.52	58.46

USGS in-domain dataset:

Method - Tile Size	Avg. Time / Map (min.)	Avg. API Cost / Map (USD)
GLYPH-1024	67.36	1.67
LOAM-1024	**38.66**	N.A.
SAM2-4096	6.90	N.A.
SAM2-2048	**5.60**	N.A.
SAM2-1024	8.69	N.A.
SAM2-0512	24.61	N.A.
SAM2-0256	106.98	N.A.
SAM3-4096	**2.65**	N.A.
SAM3-2048	4.69	N.A.
SAM3-1024	11.55	N.A.
SAM3-0512	37.38	N.A.
SAM3-0256*	79.55	N.A.
Gemini-3-flash-4096	**2.88**	**0.15**
Gemini-3-flash-2048	4.50	0.68
Gemini-3-flash-1024	9.19	1.67
Gemini-3-flash-0512	12.51	6.67
Gemini-3-flash-0256	25.11	32.77
Gemini-3.1-pro-4096	**0.56**	**0.74**
Gemini-3.1-pro-2048	0.92	3.07
Gemini-3.1-pro-1024	22.85	12.48
Gemini-3.1-pro-0512	15.85	33.75
Gemini-3.1-pro-0256	44.51	139.95
GPT-4o-4096	**1.36**	**0.16**
GPT-4o-2048	2.44	0.59
GPT-4o-1024	2.74	2.18
GPT-4o-0512	5.04	8.39
GPT-4o-0256	15.62	41.53
GPT-4o-0128*	291.06	544.79

Processing maps in the USGS dataset is more time consuming and has higher API cost due to its larger amount of map keys per map

Evaluation: In-domain USGS Geologic Map Dataset

- GLYPH has better instance-based accuracy compared to LOAM for in-domain data
  - Worse pixel-based accuracy, but still outperforming other comparative methods

Maintain good accuracy for this in-domain dataset, while generalize to other out-of-domain datasets

Dataset / Metric	GE		
Method - Tile Size	MMPQ ↑	F1@8 ↑	NBDR ↓
GLYPH-1024 (Ours)	0.67	0.67±0.32	15.28±115.11
LOAM-1024	0.61	0.71±0.37	**9.65±53.65**
SAM2-4096	0.24	0.31±0.35	38.50±134.32
SAM2-2048	0.32	0.38±0.36	27.48±124.20
SAM2-1024	0.36	0.39±0.36	24.82±115.70
SAM2-0512	**0.39**	0.42±0.36	20.81±108.15
SAM2-0256	0.38	**0.44±0.36**	**18.31±103.33**
SAM3-4096	0.00	0.00±0.01	56.44±88.46
SAM3-2048	0.00	0.01±0.05	24.04±44.60
SAM3-1024	**0.01**	0.01±0.03	19.47±41.89
SAM3-0512	0.01	**0.01±0.03**	**17.76±41.53**
Gemini-3-flash-4096	0.19	0.14±0.20	14.43±44.06
Gemini-3-flash-2048	0.27	0.26±0.23	13.39±66.10
Gemini-3-flash-1024	0.33	0.41±0.23	**5.02±29.93**
Gemini-3-flash-0512	0.32	0.40±0.23	5.52±21.30
Gemini-3-flash-0256	**0.35**	**0.43±0.26**	6.20±18.58
Gemini-3.1-pro-4096	0.03	0.03±0.10	46.23±141.72
Gemini-3.1-pro-2048	0.12	0.11±0.16	15.40±62.39
Gemini-3.1-pro-1024	0.25	0.36±0.23	4.60±36.94
Gemini-3.1-pro-0512	0.26	0.38±0.22	9.95±96.01
Gemini-3.1-pro-0256	**0.31**	**0.42±0.24**	6.58±33.54
GPT-4o-4096	0.00	0.00±0.02	83.40±159.62
GPT-4o-2048	0.00	0.01±0.04	47.39±131.50
GPT-4o-1024	0.03	0.04±0.08	21.07±71.43
GPT-4o-0512	0.05	0.07±0.11	18.28±68.52
GPT-4o-0256	**0.15**	**0.18±0.18**	**7.25±30.84**
GPT-5.2-pro	0.00	0.00±0.00	N.A.
Claude-sonnet-4.5	0.00	0.00±0.00	N.A.
Claude-opus-4.6	0.00	0.00±0.00	N.A.

Dataset / Metric	GE				
Method - Tile Size	MMPQ	P@8	R@8	F1@8	NBDR
GLYPH-1024 (Ours)	**B**	**B**	**B**	**B**	**B**
LOAM-1024	**A**	**A**	**A**	**A**	**A**
SAM2-4096	E	E	I	G	I
SAM2-2048	D	D	G	E	G
SAM2-1024	D	D	G	D	G
SAM2-0512	**C**	**C**	F	**C**	**F**
SAM2-0256	**C**	**C**	**E**	**C**	**F**
SAM3-4096	**K**	**M**	P	**N**	M
SAM3-2048	**K**	**M**	P	**N**	L
SAM3-1024	**K**	**M**	**O**	**N**	**K**
SAM3-0512	**K**	**M**	**O**	**N**	**K**
Gemini-3-flash-4096	I	I	L	J	I
Gemini-3-flash-2048	H	G	J	H	E
Gemini-3-flash-1024	F	**D**	E	D	**C**
Gemini-3-flash-0512	F	E	D	D	E
Gemini-3-flash-0256	**E**	E	**C**	**C**	F
Gemini-3.1-pro-4096	K	K	O	M	K
Gemini-3.1-pro-2048	J	I	M	K	H
Gemini-3.1-pro-1024	G	F	H	F	**B**
Gemini-3.1-pro-0512	G	E	G	E	D
Gemini-3.1-pro-0256	**E**	**D**	**D**	**C**	E
GPT-4o-4096	K	M	P	N	N
GPT-4o-2048	K	L	P	N	L
GPT-4o-1024	K	J	O	M	J
GPT-4o-0512	K	H	N	L	I
GPT-4o-0256	**I**	**G**	**K**	**I**	**F**
GPT-5.2-pro	K	N	P	N	N.A.
Claude-sonnet-4.5	K	N	P	N	N.A.
Claude-opus-4.6	K	N	P	N	N.A.
ANOVA p-value	$\to 0$	$\to 0$	$\to 0$	$\to 0$	$\to 0$
LSD (α=0.05)	0.02	0.03	0.02	0.02	0.20

Evaluation: Ablation Study

GLYPH: The standard GLYPH pipeline

⇒ **RGB space**: Use RGB color space instead of CIELAB color space, while keeping the remaining components the same

⇒ **Pixel majority**: Apply only pixel-wise majority voting among three expert models

⇒ **Region majority**: Apply only region-wise majority voting among the expert models

- **Psuedo**: The training signal of contrastive learning for color embedding does not consider cross-expert consensus pseudo-labels; it relies only on input map keys

- **Anchor**: The training signal of contrastive learning for color embedding does not consider input map keys; it relies only on cross-expert consensus pseudo-labels.

- **Gating**: The gating in test-time adaptation is not optimized from global style or reliability features; instead, it directly applies uniform weighting for semantic fusion.

- **Evidence**: The semantic fusion does not consider expert evidence; it relies only on similarity between region embedding and legend prototypes

- **Similarity**: The semantic fusion does not consider similarity between region embedding and legend prototypes; it relies only on expert evidence

- **Post**: The structural and geometric post-processing is skipped

Dataset / Metric	FT			SA			SO			SP			WR		
Ablation Setting	MMPQ ↑	F1@8 ↑	NBDR ↓	MMPQ ↑	F1@8 ↑	NBDR ↓	MMPQ ↑	F1@8 ↑	NBDR ↓	MMPQ ↑	F1@8 ↑	NBDR ↓	MMPQ ↑	F1@8 ↑	NBDR ↓
GLYPH	0.90	**0.90±0.16**	0.48±1.27	0.69	0.87±0.16	**0.12±0.16**	0.81	0.91±0.18	**0.36±1.12**	**0.30**	**0.59±0.21**	0.79±0.82	**0.65**	**0.82±0.13**	0.15±0.16
⇒RGB space	**0.84**	0.86±0.15	0.53±1.32	0.63	0.81±0.14	0.14±0.09	0.76	0.88±0.25	0.36±1.08	0.13	0.36±0.23	2.62±2.98	0.41	0.73±0.24	0.32±0.30
⇒Pixel majority	0.76	0.82±0.21	0.35±0.92	0.58	0.76±0.22	0.17±0.15	0.72	0.87±0.24	0.37±1.02	0.18	0.49±0.22	0.82±0.85	0.61	0.82±0.18	0.10±0.18
⇒Region majority	0.72	0.77±0.21	0.46±1.01	0.63	0.78±0.20	0.16±0.16	0.75	0.84±0.21	0.36±0.87	0.18	0.52±0.23	0.72±0.84	0.51	0.75±0.17	0.15±0.20
-Psuedo	0.83	0.93±0.12	0.30±0.79	0.64	0.85±0.18	0.14±0.17	**0.81**	0.88±0.22	0.39±1.19	0.15	0.53±0.29	0.89±0.98	0.52	0.73±0.20	0.21±0.22
-Anchor	0.81	0.88±0.18	0.47±1.06	0.66	0.82±0.20	0.16±0.19	0.79	0.89±0.22	0.35±0.96	0.36	0.65±0.23	0.65±0.99	0.49	0.71±0.24	0.21±0.23
-Gating	0.81	0.89±0.17	0.33±0.82	0.78	0.88±0.17	0.11±0.17	0.77	0.88±0.24	0.38±1.10	0.16	0.53±0.24	1.21±1.04	0.66	0.81±0.16	**0.14±0.17**
-Evidence	0.82	0.87±0.22	**0.32±0.85**	0.71	0.79±0.33	0.29±0.54	0.63	0.81±0.27	0.42±1.26	0.15	0.40±0.34	4.29±6.53	0.60	0.74±0.25	0.16±0.20
-Similarity	0.60	0.68±0.25	1.98±9.61	0.46	0.65±0.24	0.18±0.13	0.74	0.79±0.20	0.48±1.01	0.25	0.55±0.26	**0.68±1.05**	0.21	0.47±0.22	0.29±0.25
-Post	0.82	0.89±0.17	0.43±0.96	**0.73**	**0.87±0.15**	0.14±0.17	0.82	**0.89±0.22**	0.39±1.15	0.26	0.56±0.26	0.95±1.10	0.54	0.75±0.21	0.19±0.22

Combining all components (GLYPH) is the only ablation setup to achieve the best or second-best accuracy for most metrics

Evaluation: Ablation Study (cont.)

GLYPH: The standard GLYPH pipeline

⇒ **RGB space**: Use RGB color space instead of CIELAB color space, while keeping the remaining components the same

⇒ **Pixel majority**: Apply only pixel-wise majority voting among three expert models

⇒ **Region majority**: Apply only region-wise majority voting among the expert models

- **Psuedo**: The training signal of contrastive learning for color embedding does not consider cross-expert consensus pseudo-labels; it relies only on input map keys

- **Anchor**: The training signal of contrastive learning for color embedding does not consider input map keys; it relies only on cross-expert consensus pseudo-labels.

- **Gating**: The gating in test-time adaptation is not optimized from global style or reliability features; instead, it directly applies uniform weighting for semantic fusion.

- **Evidence**: The semantic fusio … and legend prototypes

- **Similarity**: The semantic fusic … only on expert evidence

- **Post**: The structural and geon …

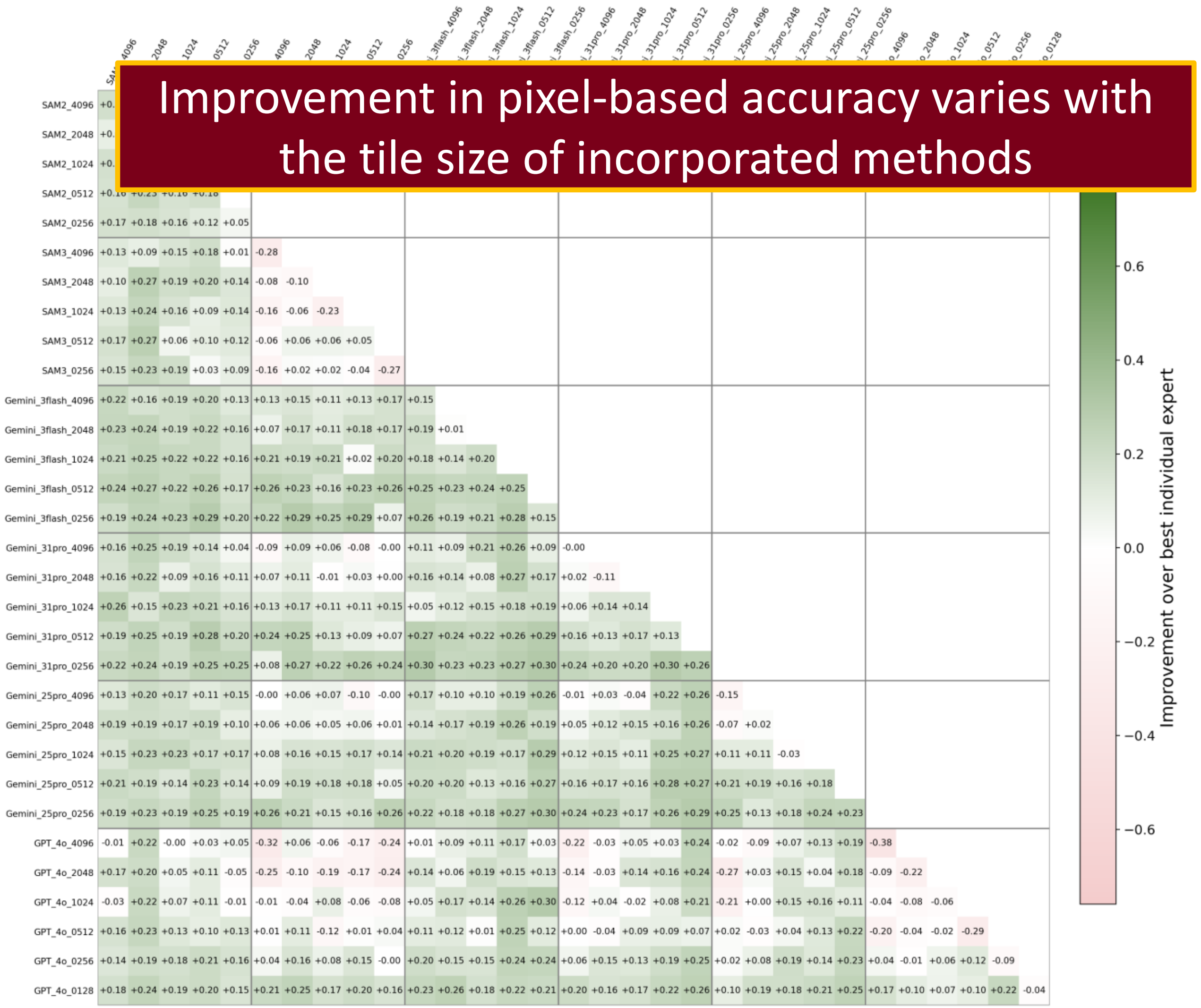
Combining all components (GLYPH) is the only ablation setup to achieve the best tier ("A") across datasets and metrics

Dataset / Metric	FT					SA					SO					SP					WR				
Ablation Setting	MMPQ	P@8	R@8	F1@8	NBDR	MMPQ	P@8	R@8	F1@8	NBDR	MMPQ	P@8	R@8	F1@8	NBDR	MMPQ	P@8	R@8	F1@8	NBDR	MMPQ	P@8	R@8	F1@8	NBDR
GLYPH	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
⇒RGB space	A	B	A	A	A	A	A	A	A	B	A	A	A	A	A	A	B	A	B	A	A	B	A	A	B
⇒Pixel majority	B	A	B	B	A	B	A	B	B	B	A	A	A	A	A	A	A	A	A	A	A	A	B	A	A
⇒Region majority	B	A	B	B	B	B	A	B	A	B	A	A	B	A	A	A	A	A	A	A	A	A	B	A	A
-Psuedo	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	A	B
-Anchor	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	A	A
-Gating	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	A	A
-Evidence	A	A	A	A	A	A	A	A	A	A	A	A	C	A	A	A	A	B	A	A	A	A	B	A	A
-Similarity	C	B	C	C	B	C	B	C	B	B	B	A	C	B	B	A	A	A	A	A	B	A	C	B	B
-Post	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	A	A
ANOVA p-value	e_2^{-9}	e_9^{-3}	e_4^{-12}	e_1^{-7}	e_3^{-2}	e_2^{-4}	e_4^{-1}	e_7^{-5}	e_4^{-3}	e_1^{-2}	e_2^{-1}	e_9^{-1}	e_5^{-9}	e_5^{-1}	e_2^{-1}	e_8^{-1}	e_5^{-1}	e_6^{-1}	e_7^{-1}	e_8^{-1}	e_3^{-4}	e_2^{-1}	e_3^{-11}	e_4^{-4}	e_9^{-2}
LSD (α=0.05)	0.10	0.07	0.09	0.09	0.95	0.13	0.09	0.13	0.11	0.63	0.13	0.11	0.06	0.10	1.00	0.27	0.32	0.31	0.29	1.87	0.15	0.16	0.13	0.15	0.86

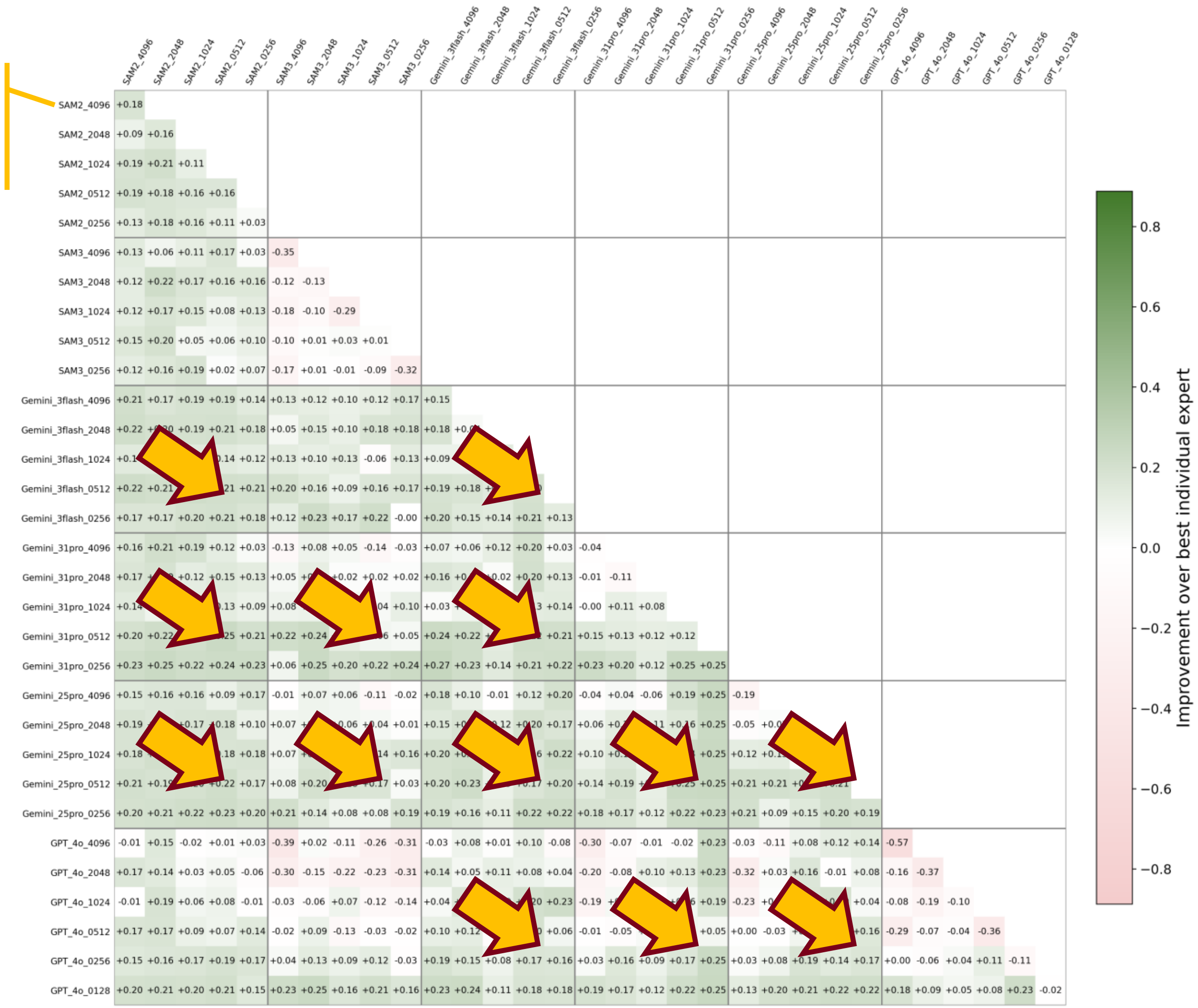
There are statistically significant difference among solutions in MMPQ, R@8, and F1@8 for most datasets

Pairwise Fusion Improvement

Improvement in pixel-based accuracy varies with the tile size of incorporated methods



▲Pairwise fusion improvement over best method (SA, MMPQ)



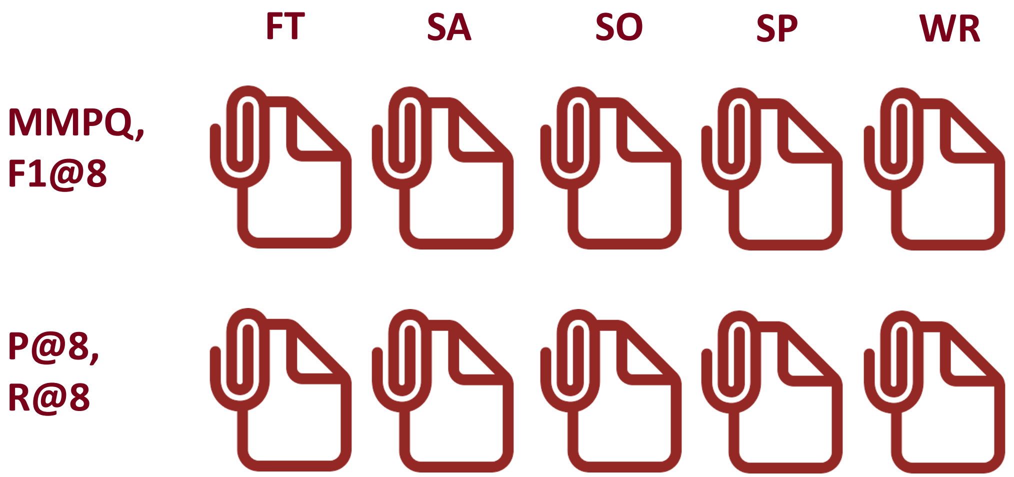
▲Pairwise fusion improvement over best method (SA, F1@8)

Improvement in instance-based accuracy is relatively unanimous **regardless of the adopted methods**

Details

Evaluation: Pairwise Fusion Improvement

- We test GLYPH with model pairs to assess its efficacy to improve from bad solutions
  - Since GLYPH takes 3 inputs
    - We always include LOAM
    - Test all pairs of comparative methods
  - Evaluate the improvement over "the best of the three inputs" for each cell
    - Each block represent a pair of method under different tile sizes



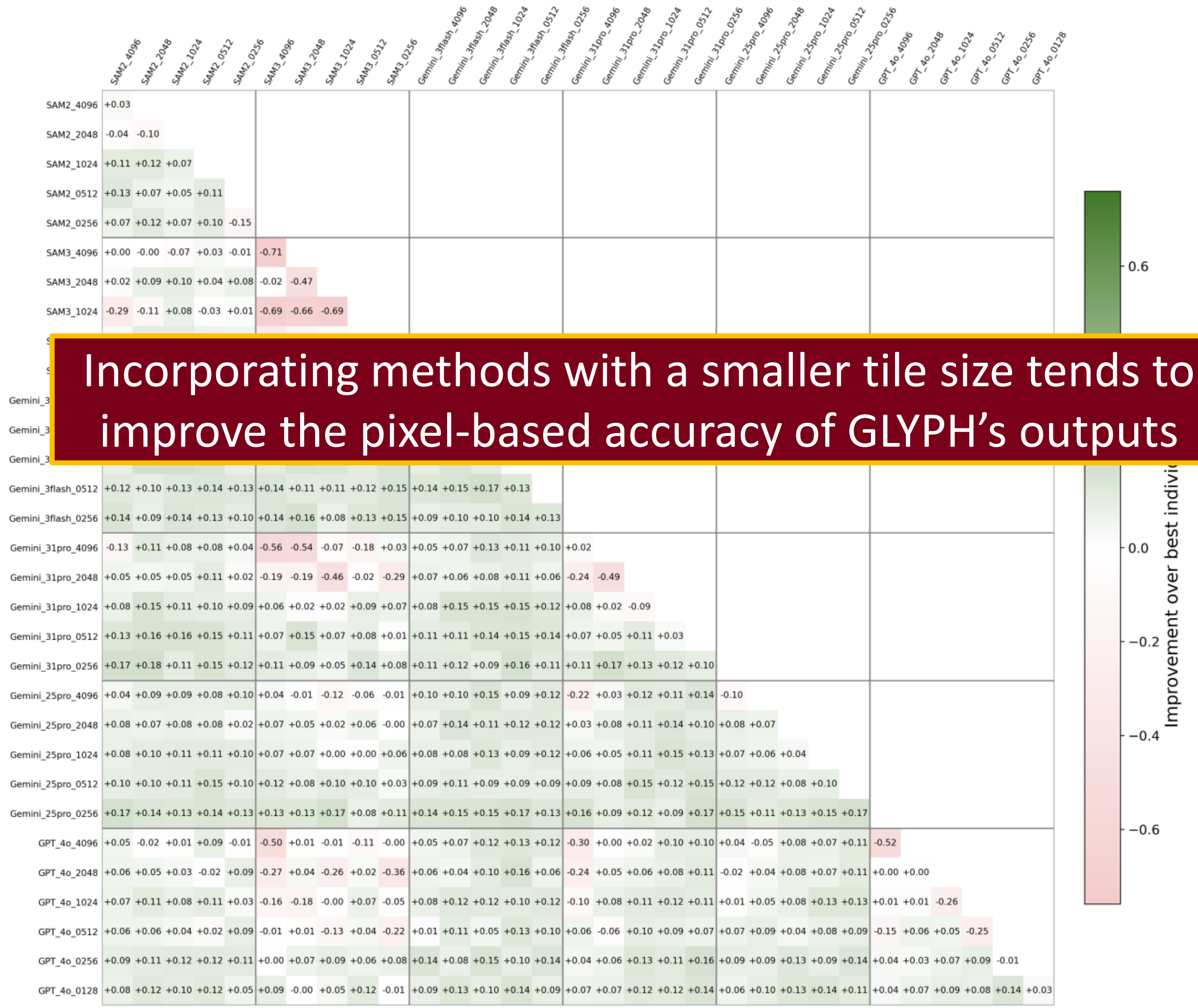
Smaller tile size

	SAM2_4096	SAM2_2048	SAM2_1024	SAM2_0512	SAM2_0256
Gemini_3flash_4096	+0.01	+0.03	+0.01	+0.03	-0.01
Gemini_3flash_2048	+0.03	-0.00	+0.04	+0.04	+0.03
Gemini_3flash_1024	+0.06	+0.04	+0.07	+0.06	+0.02
Gemini_3flash_0512	+0.02	+0.04	+0.06	+0.06	+0.02
Gemini_3flash_0256	+0.11	+0.06	+0.07	+0.07	+0.05

Diagonal cells are the pair of same method and tile size

	SAM2_4096	SAM2_2048	SAM2_1024	SAM2_0512	SAM2_0256
SAM2_4096	+0.18				
SAM2_2048	+0.09	+0.19			
SAM2_1024	+0.17	+0.22	+0.11		
SAM2_0512	+0.16	+0.23	+0.16	+0.18	
SAM2_0256	+0.17	+0.18	+0.16	+0.12	+0.05

Evaluation: Pairwise Fusion Improvement – FT



Incorporating methods with a smaller tile size tends to improve the pixel-based accuracy of GLYPH's outputs

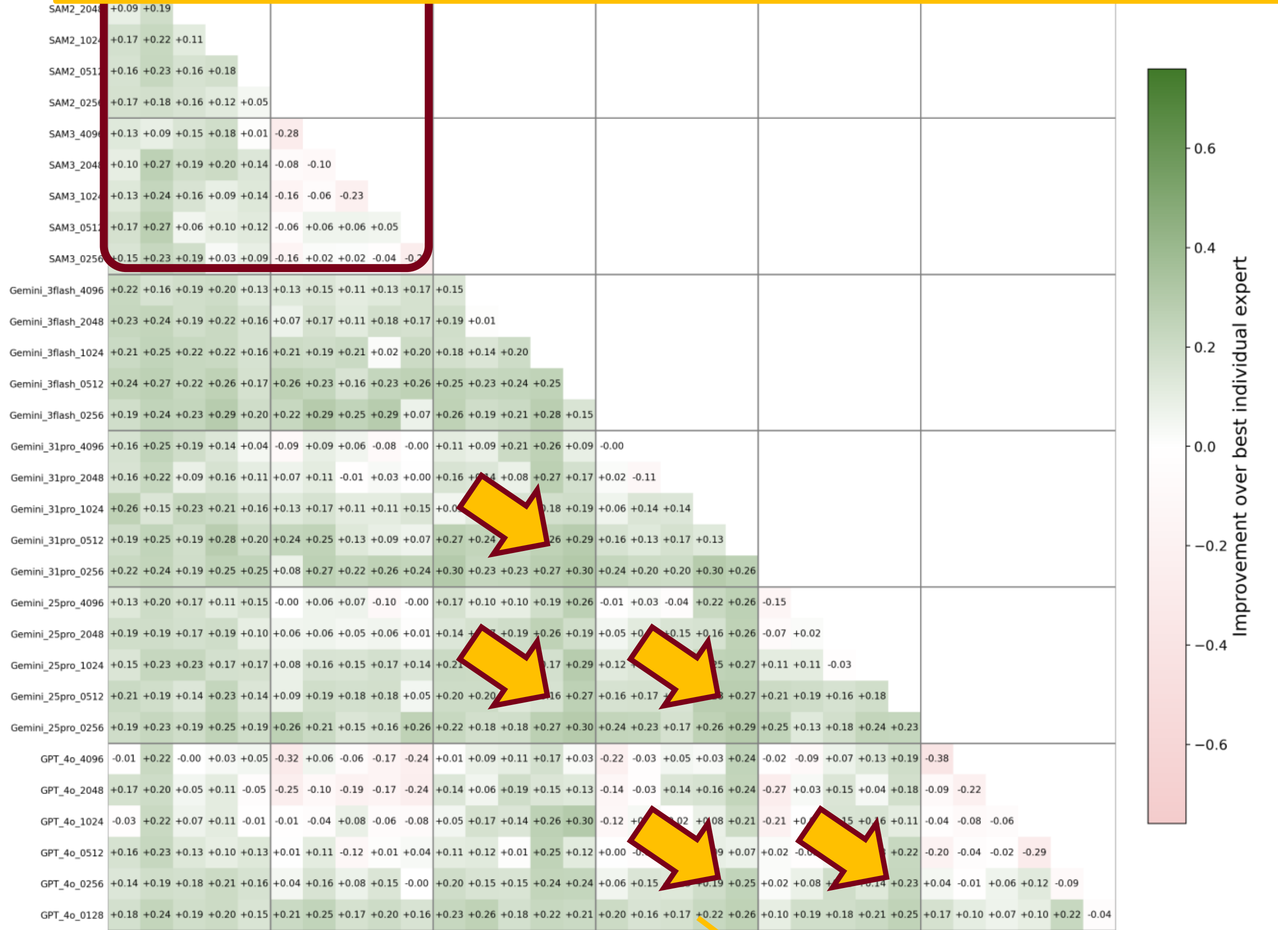
▲Pairwise fusion improvement over best method (FT, MMPQ)

▲Pairwise fusion improvement over best method (FT, F1@8)

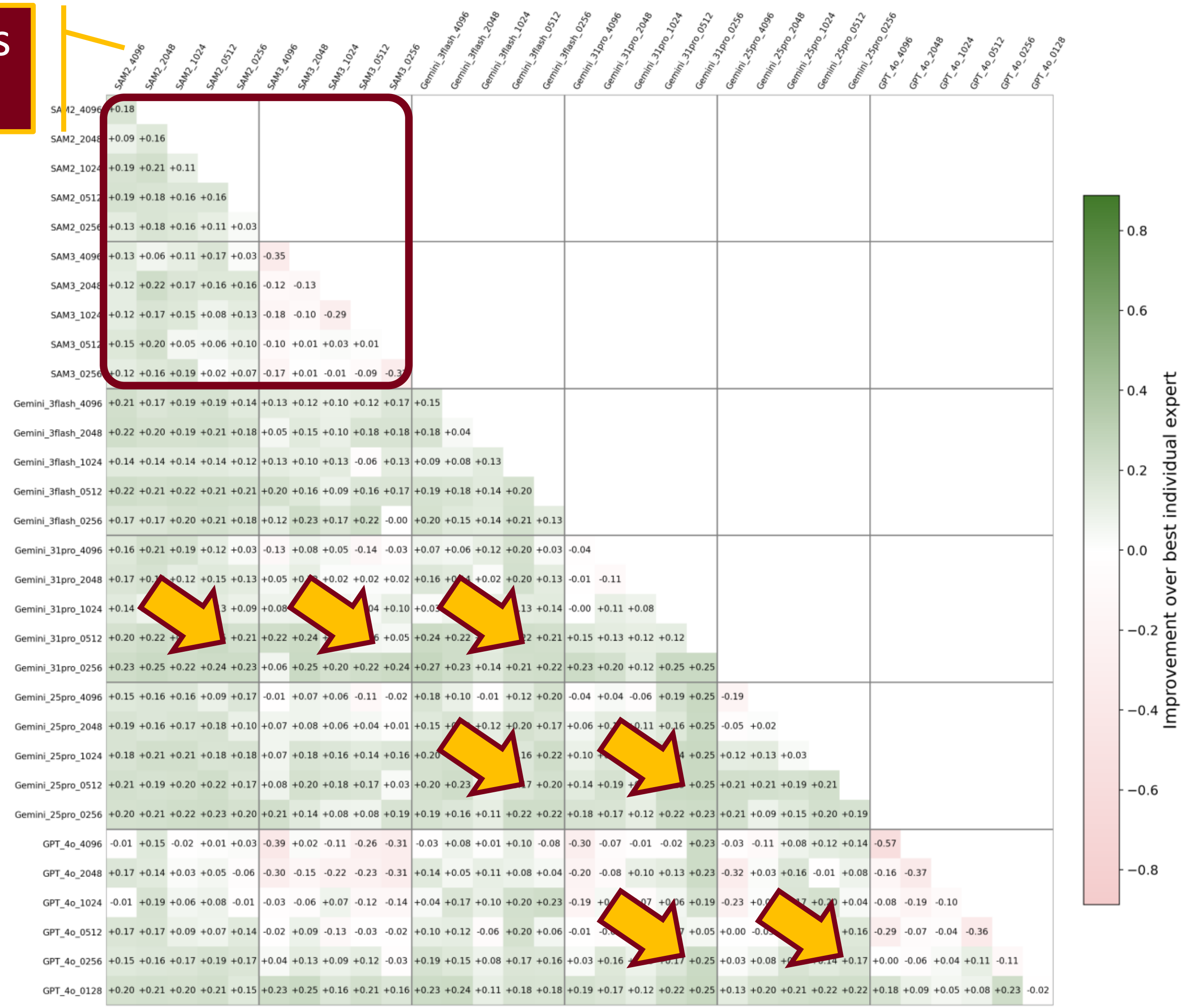
Bring improvement in instance-based accuracy in most cases, especially for methods with tile size smaller than 2048

Evaluation: Pairwise Fusion Improvement – SA

Show GLYPH's ability to improve from multiple solutions **with mediocre standalone performance**



▲Pairwise fusion improvement over best method (SA, MMPQ)

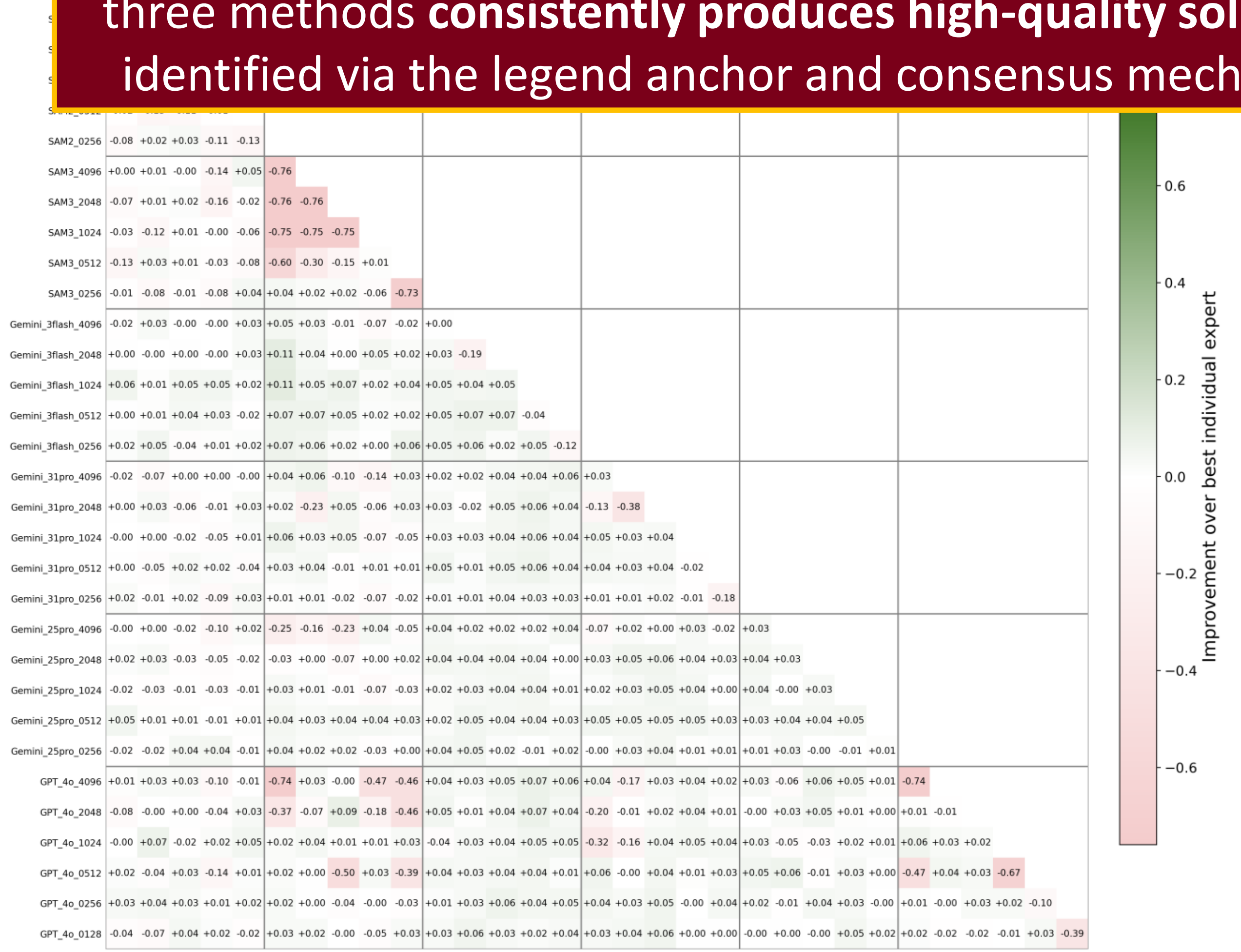


▲Pairwise fusion improvement over best method (SA, F1@8)

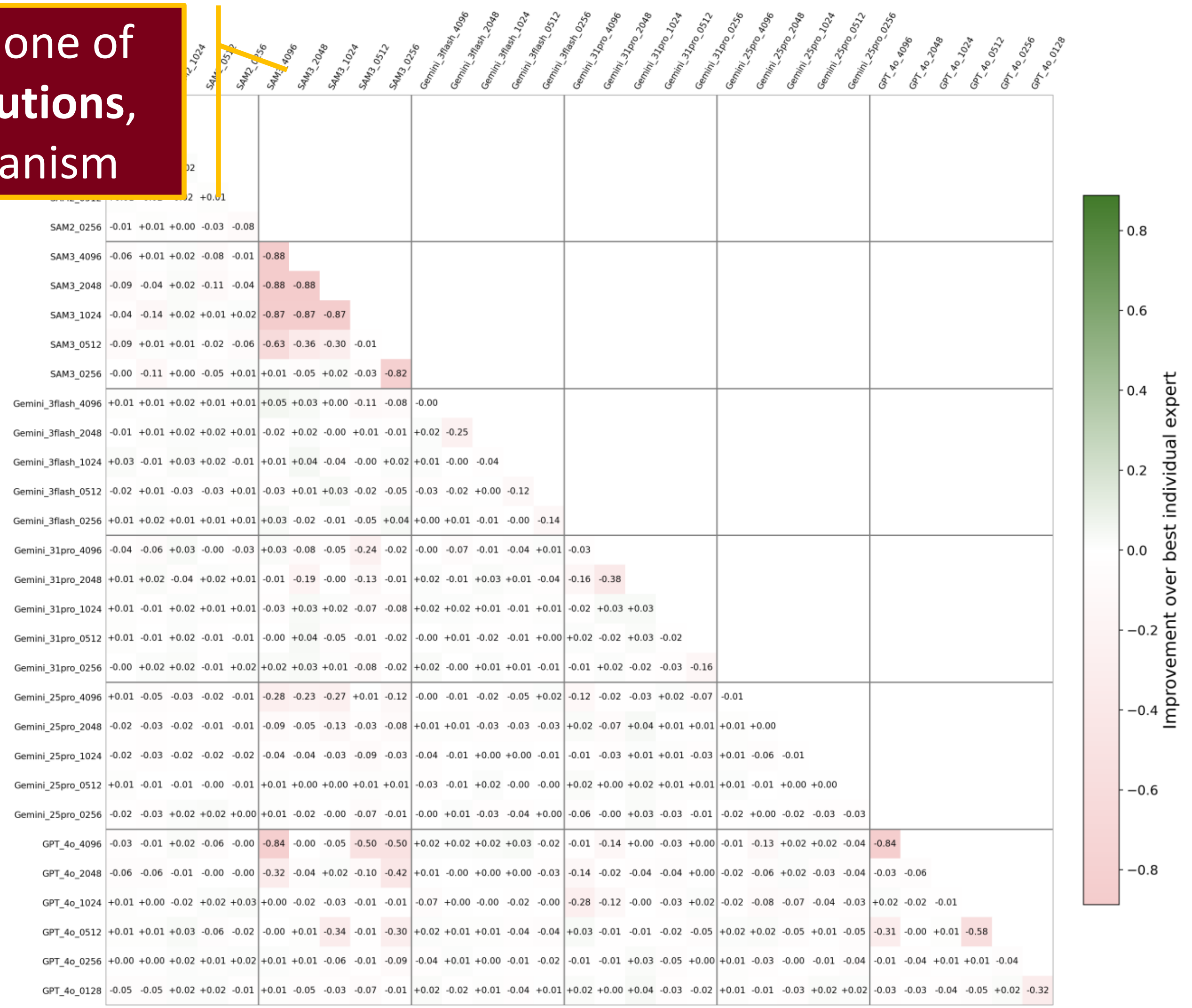
Tile size affects more on VLMs;
SAM series shows less clear pattern

Evaluation: Pairwise Fusion Improvement – SO

Show GLYPH's ability to maintain good accuracy when one of three methods **consistently produces high-quality solutions**, identified via the legend anchor and consensus mechanism



▲Pairwise fusion improvement over best method (SO, MMPQ)



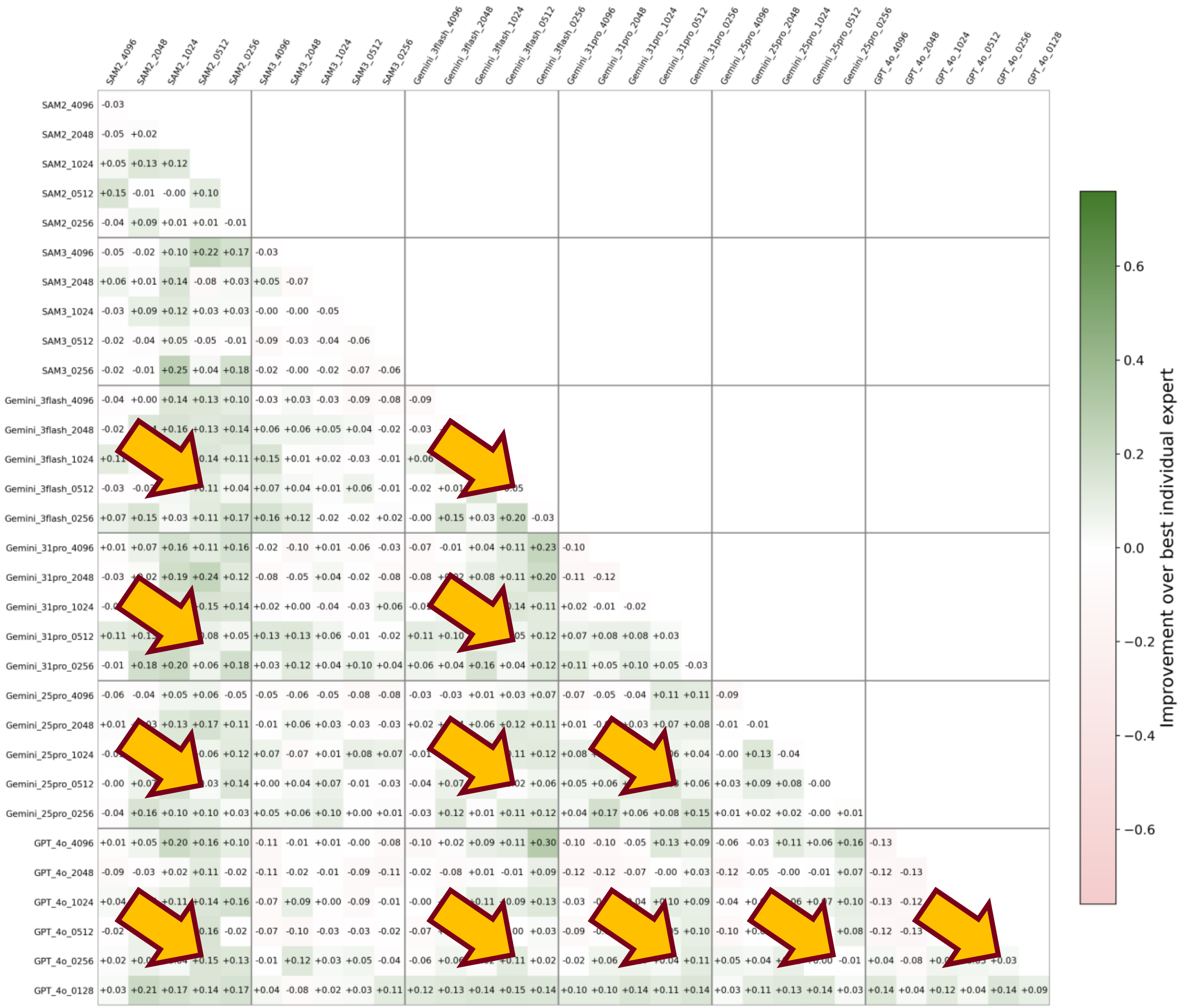
▲Pairwise fusion improvement over best method (SO, F1@8)

Some improvement in MMPQ with no clear pattern

Bounded by high pixel-based accuracy of LOAM

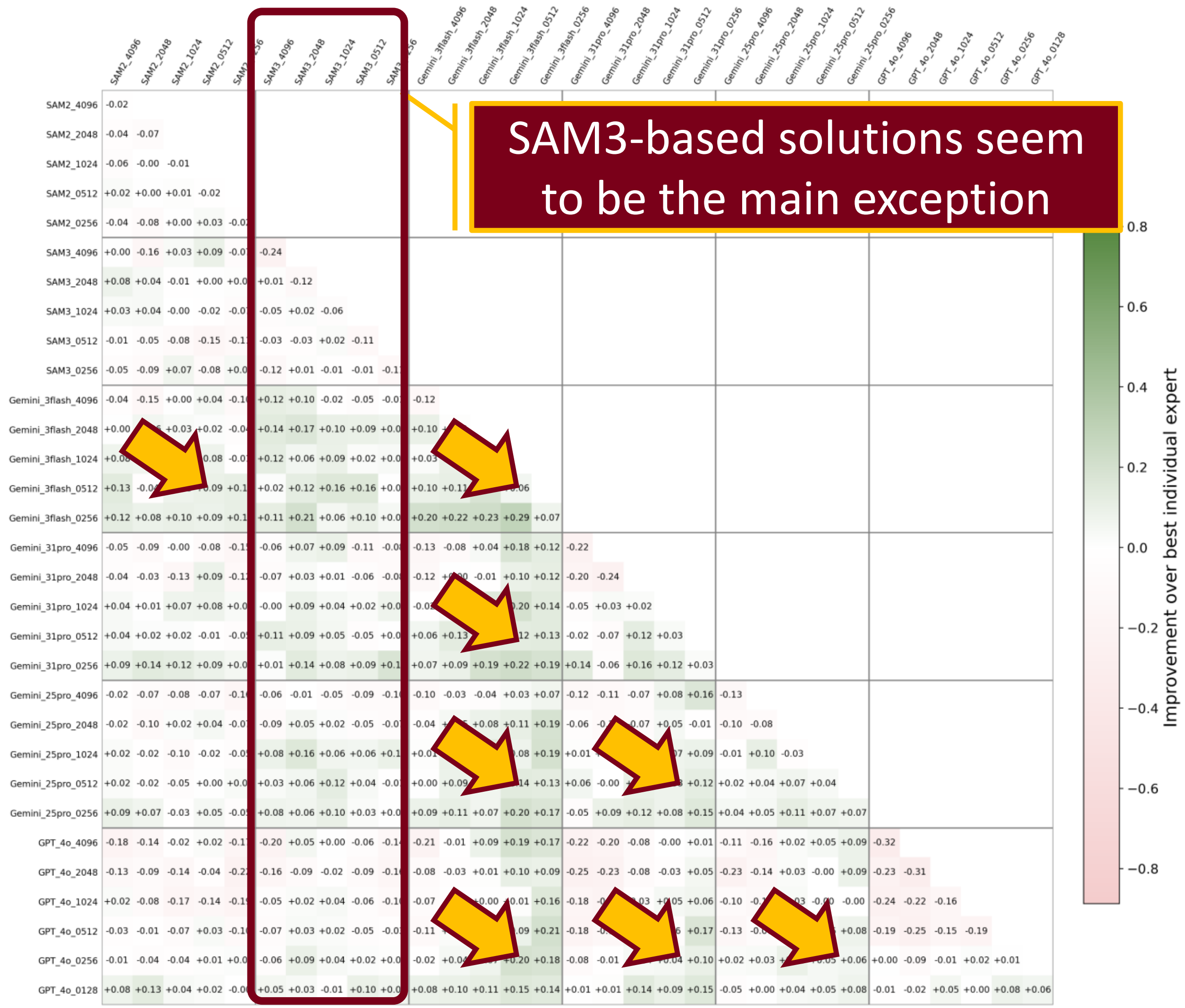
USC Viterbi School of Engineering

Evaluation: Pairwise Fusion Improvement – SP



▲Pairwise fusion improvement over best method (SP, MMPQ)

More improvement in MMPQ with fine-grained solutions

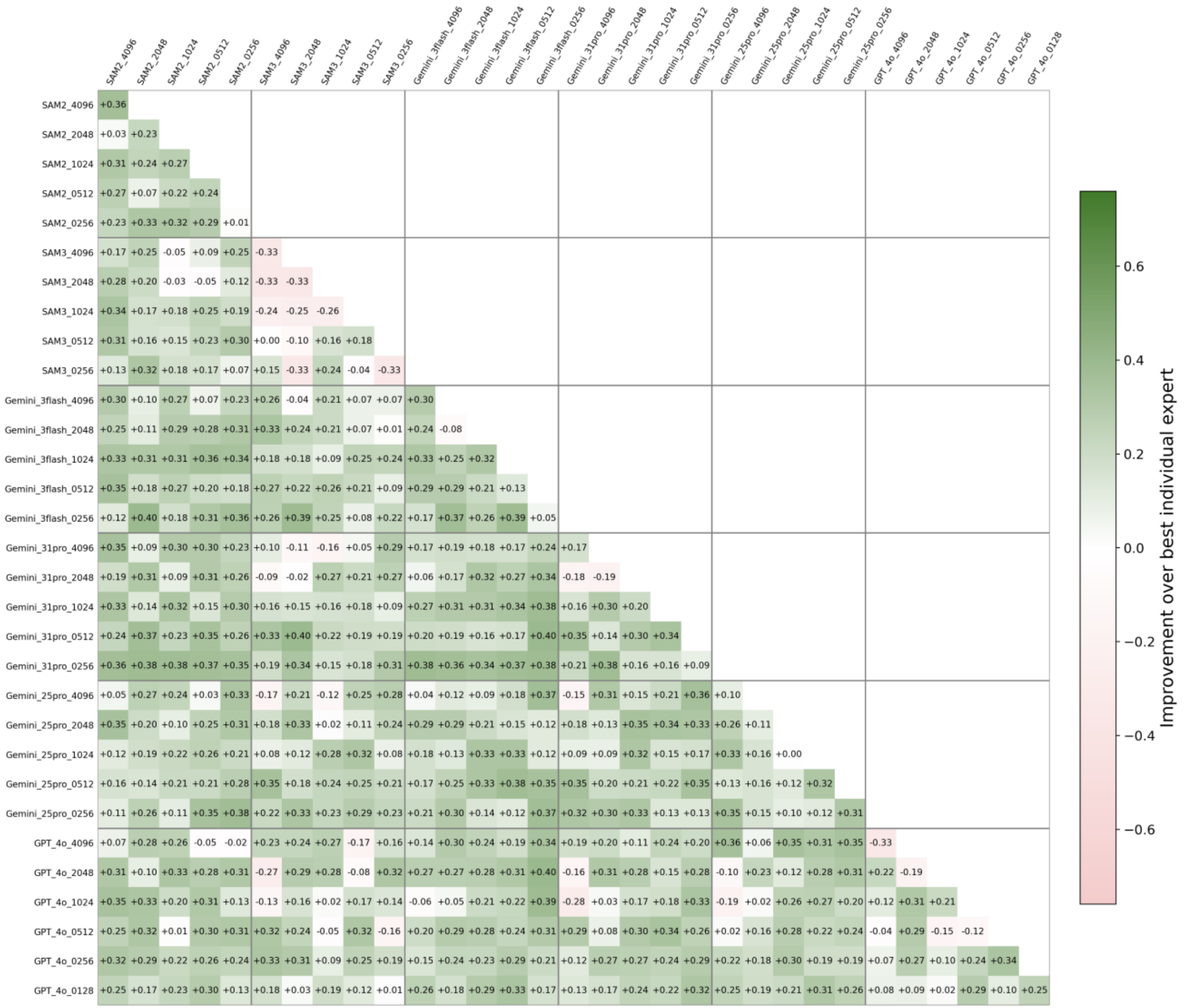


SAM3-based solutions seem to be the main exception

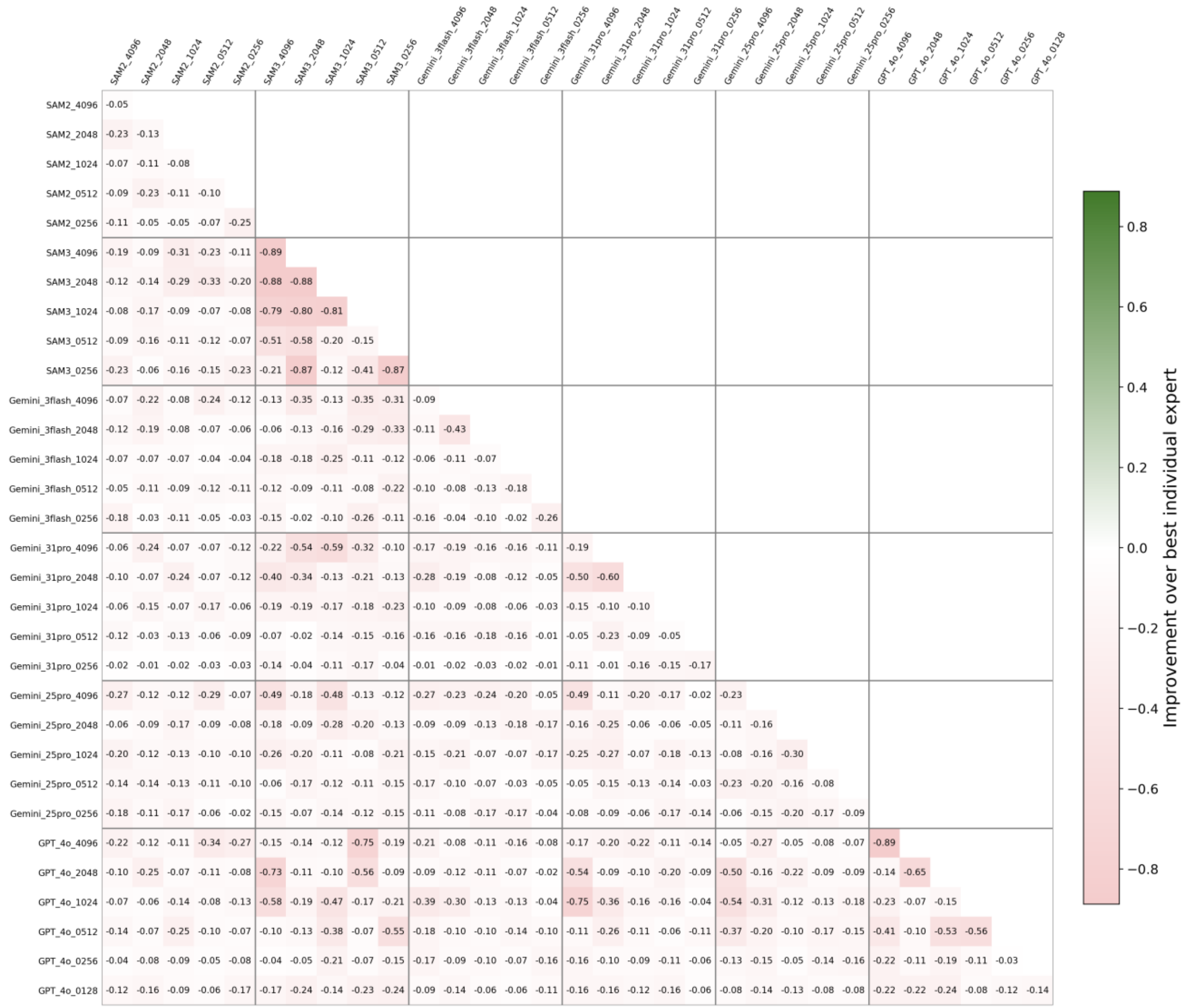
▲Pairwise fusion improvement over best method (SP, F1@8)

Similar trends for F1@8 regarding tile size

Evaluation: Pairwise Fusion Improvement – WR



▲Pairwise fusion improvement over best method (WR, MMPQ)

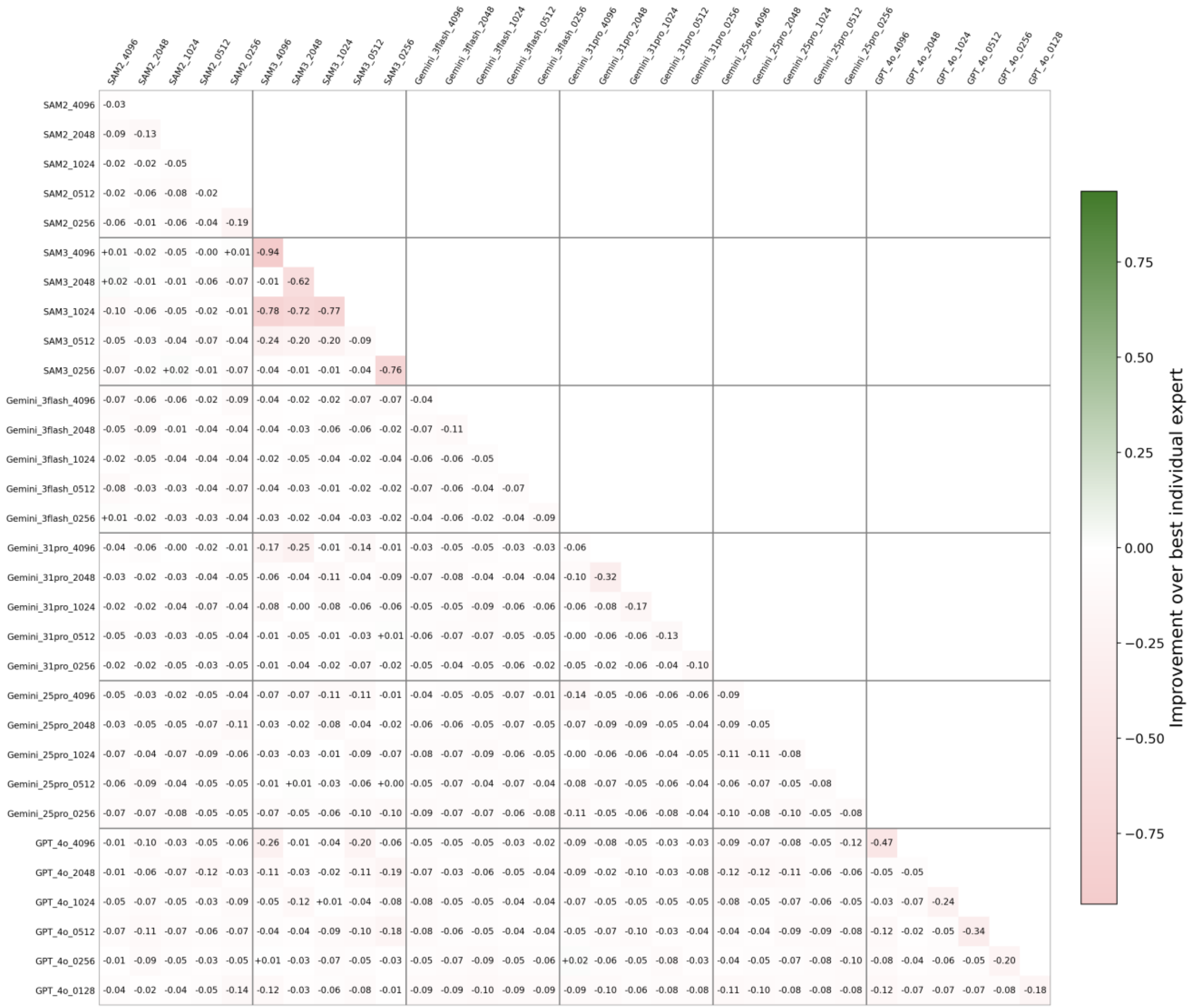


▲Pairwise fusion improvement over best method (WR, F1@8)

Significant improvement in MMPQ with no clear pattern

Bounded by high pixel-based accuracy of LOAM

Evaluation: Pairwise Fusion Improvement – FT (cont.)



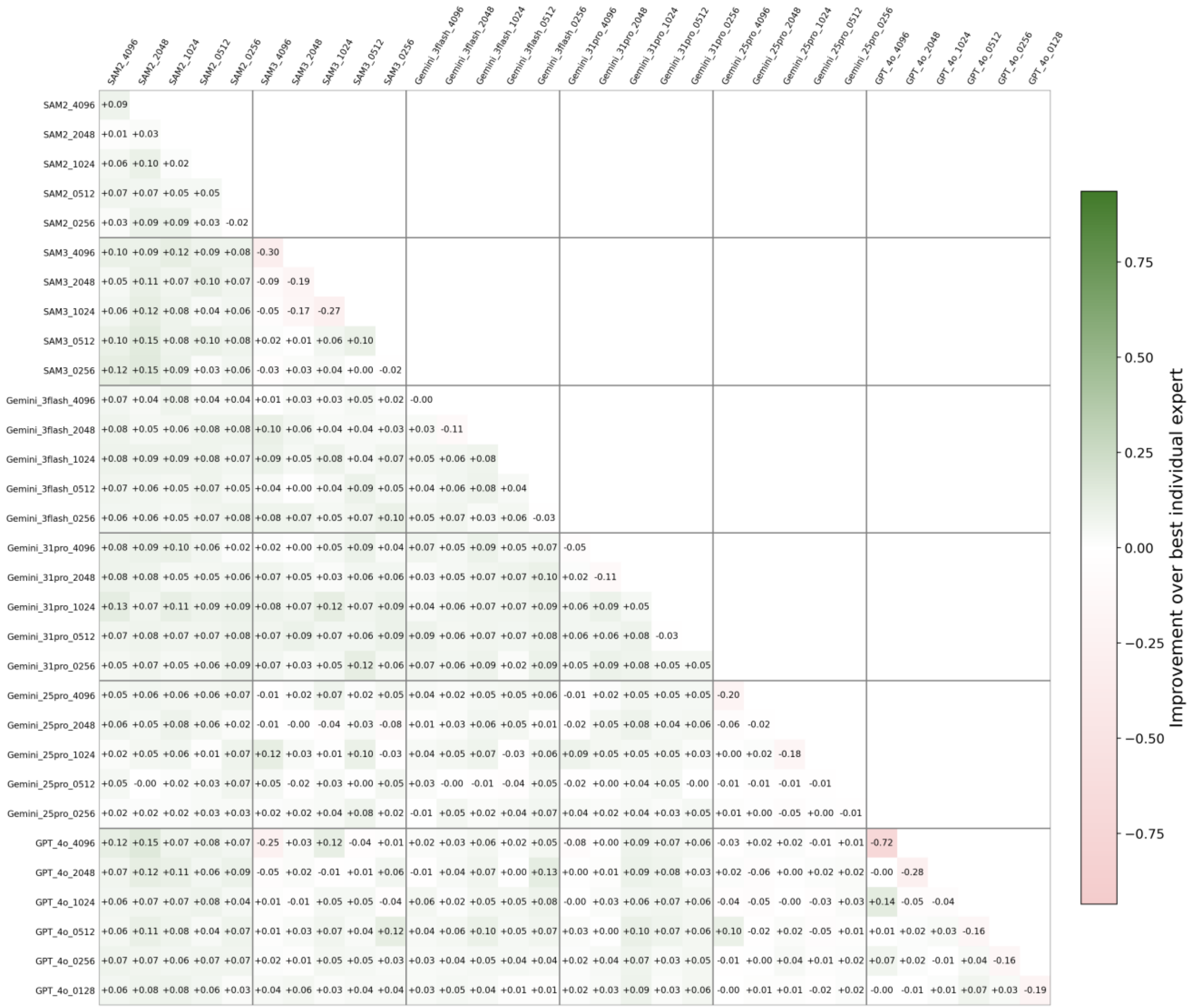
▲Pairwise fusion improvement over best method (FT, P@8)

▲Pairwise fusion improvement over best method (FT, R@8)

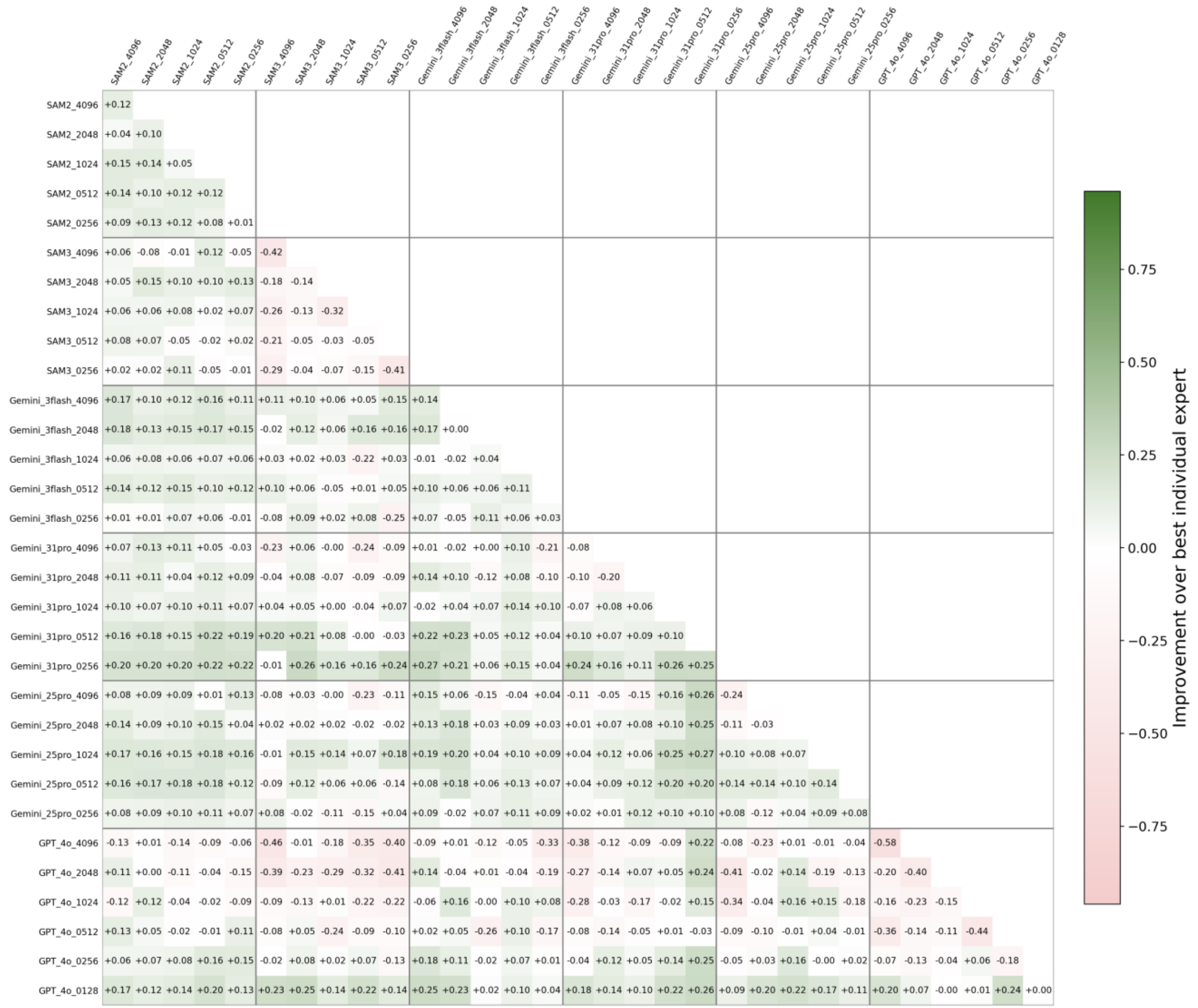
Bounded by high precision (P@8) of LOAM

Improved recall (R@8) with a smaller tile size

Evaluation: Pairwise Fusion Improvement – SA (cont.)



▲Pairwise fusion improvement over best method (SA, P@8)

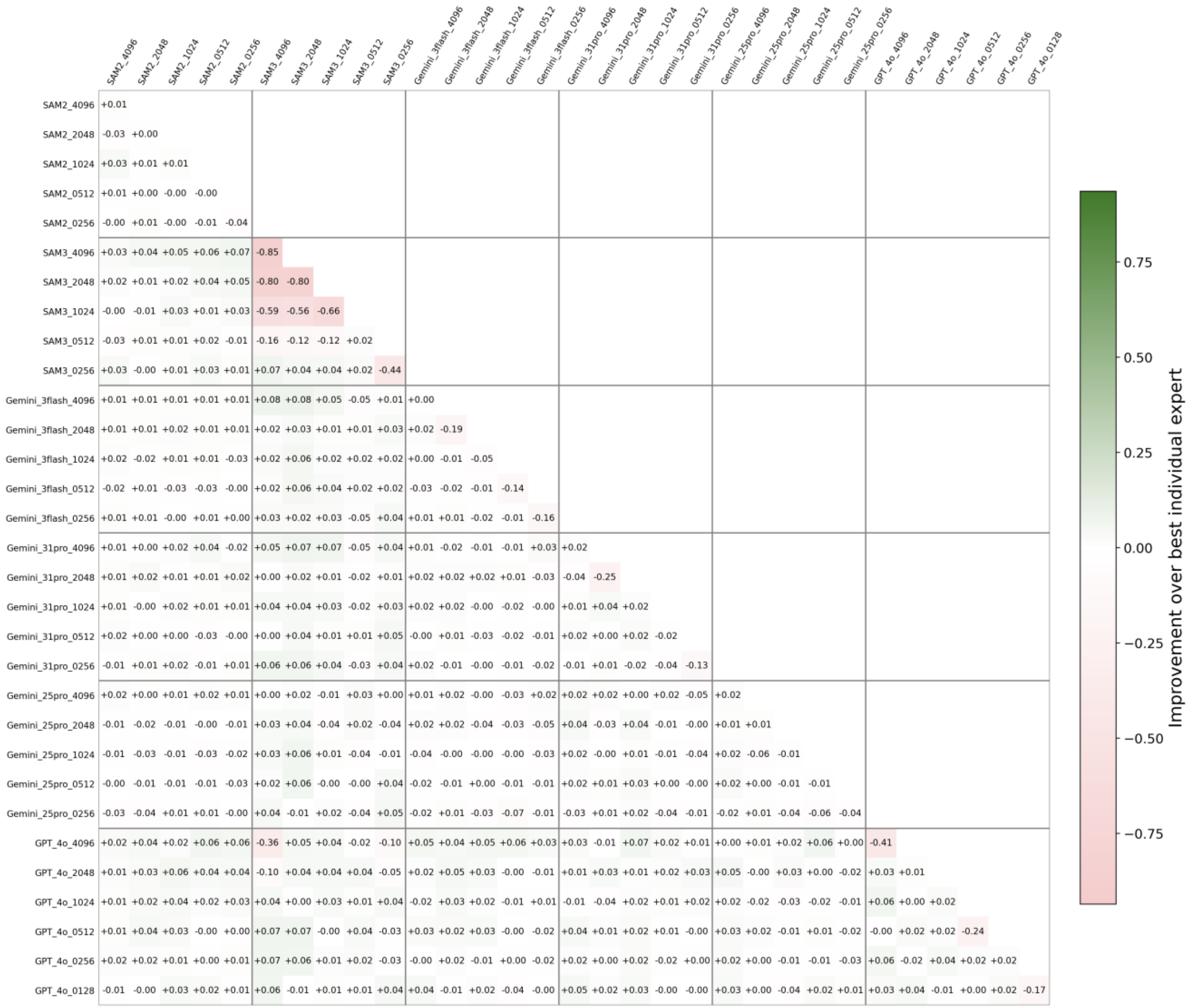


▲Pairwise fusion improvement over best method (SA, R@8)

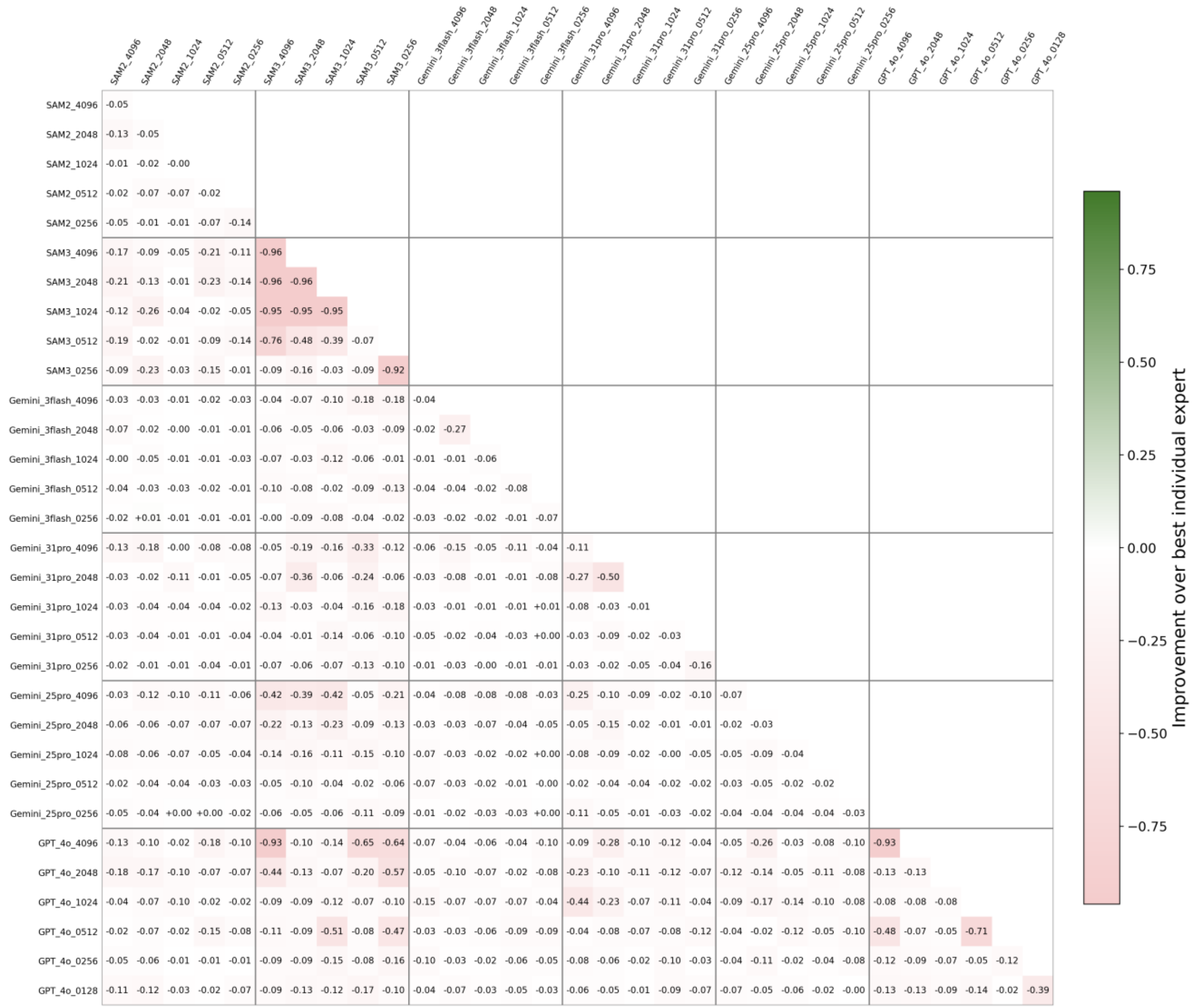
Obtain general improvement in P@8 with no clear pattern

The improvement in R@8 is general but no clear pattern

Evaluation: Pairwise Fusion Improvement – SO (cont.)



▲Pairwise fusion improvement over best method (SO, P@8)

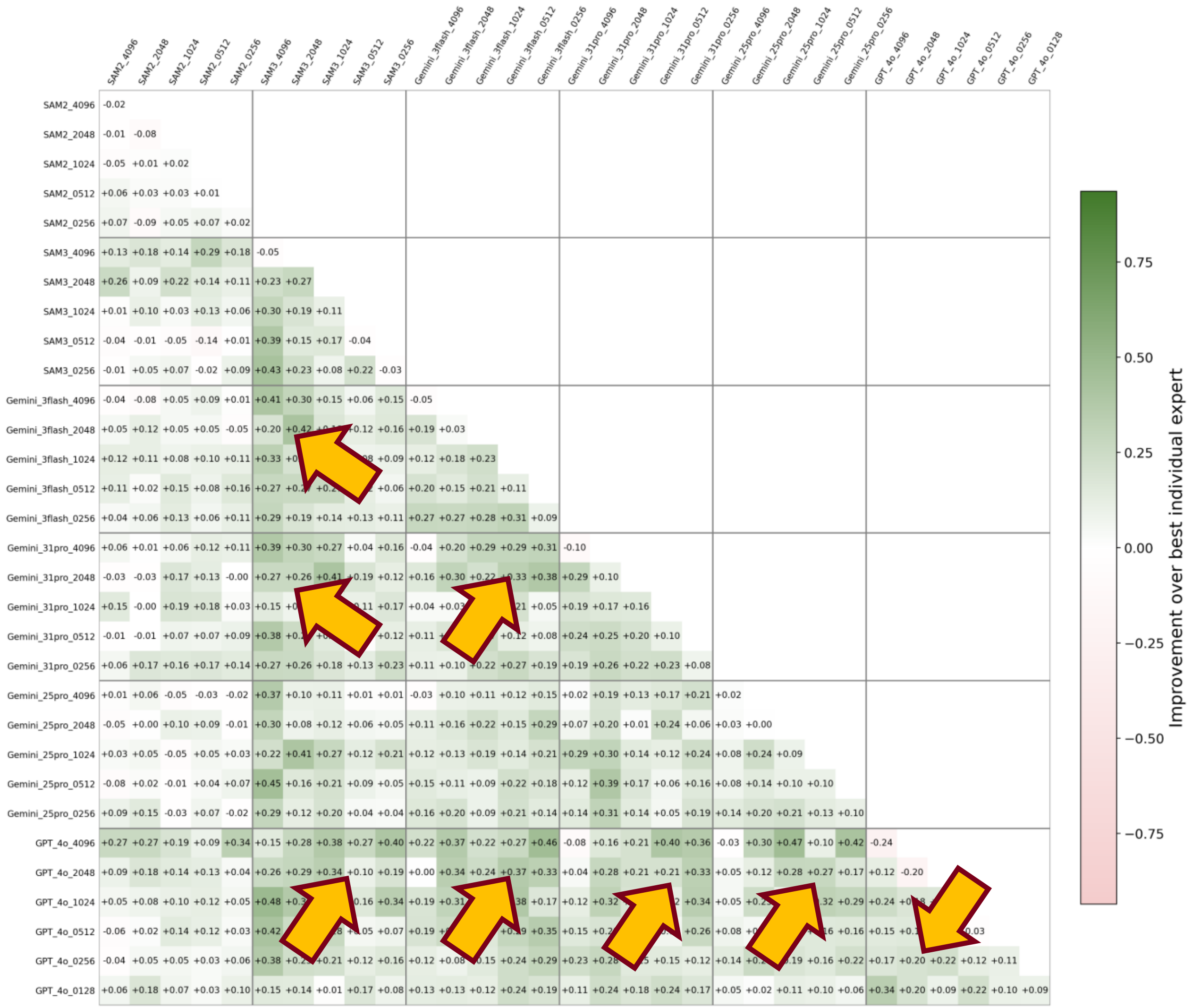


▲Pairwise fusion improvement over best method (SO, R@8)

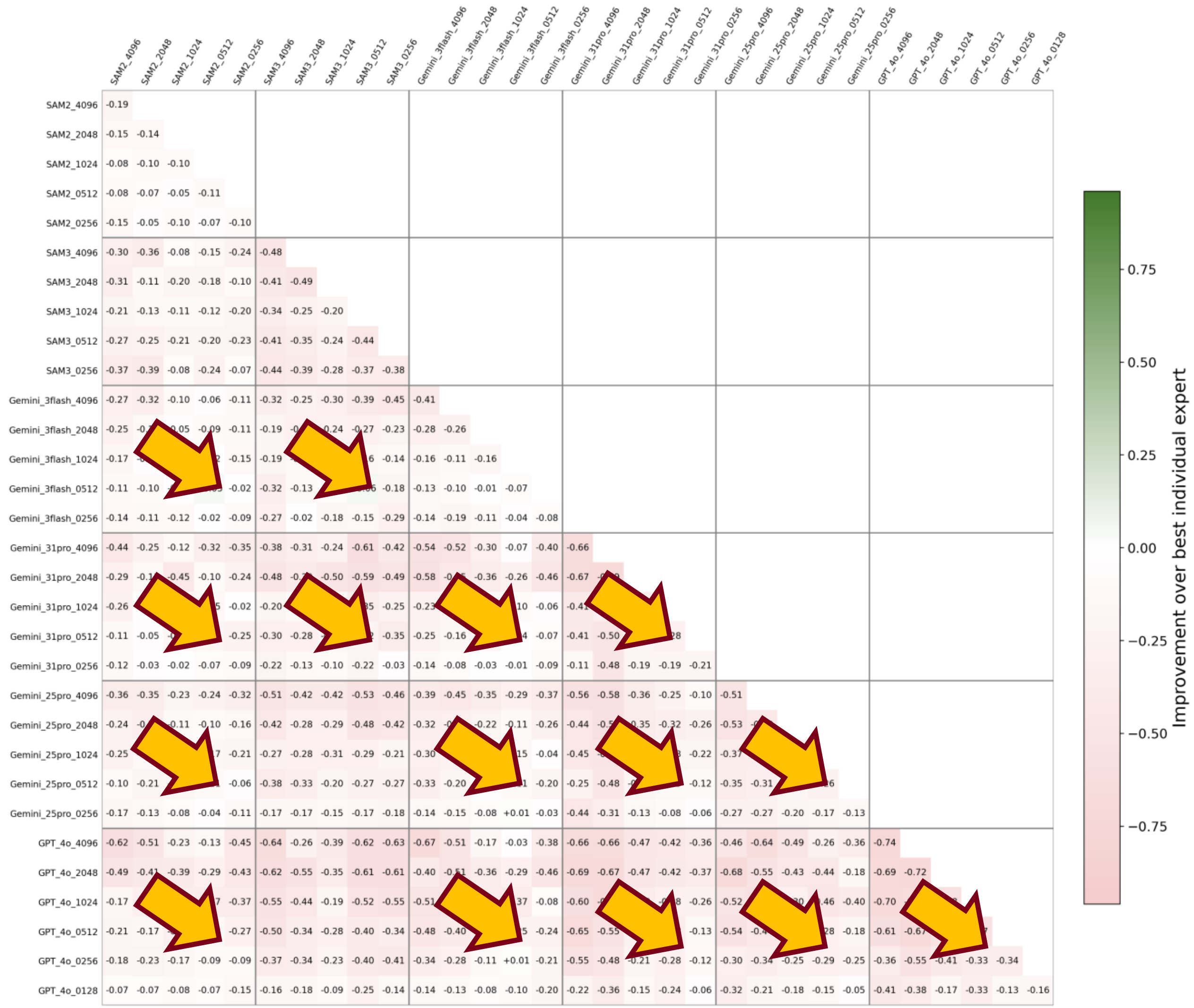
Bounded by high pixel-based accuracy of LOAM

Bounded by high pixel-based accuracy of LOAM

Evaluation: Pairwise Fusion Improvement – SP (cont.)



▲Pairwise fusion improvement over best method (SP, P@8)

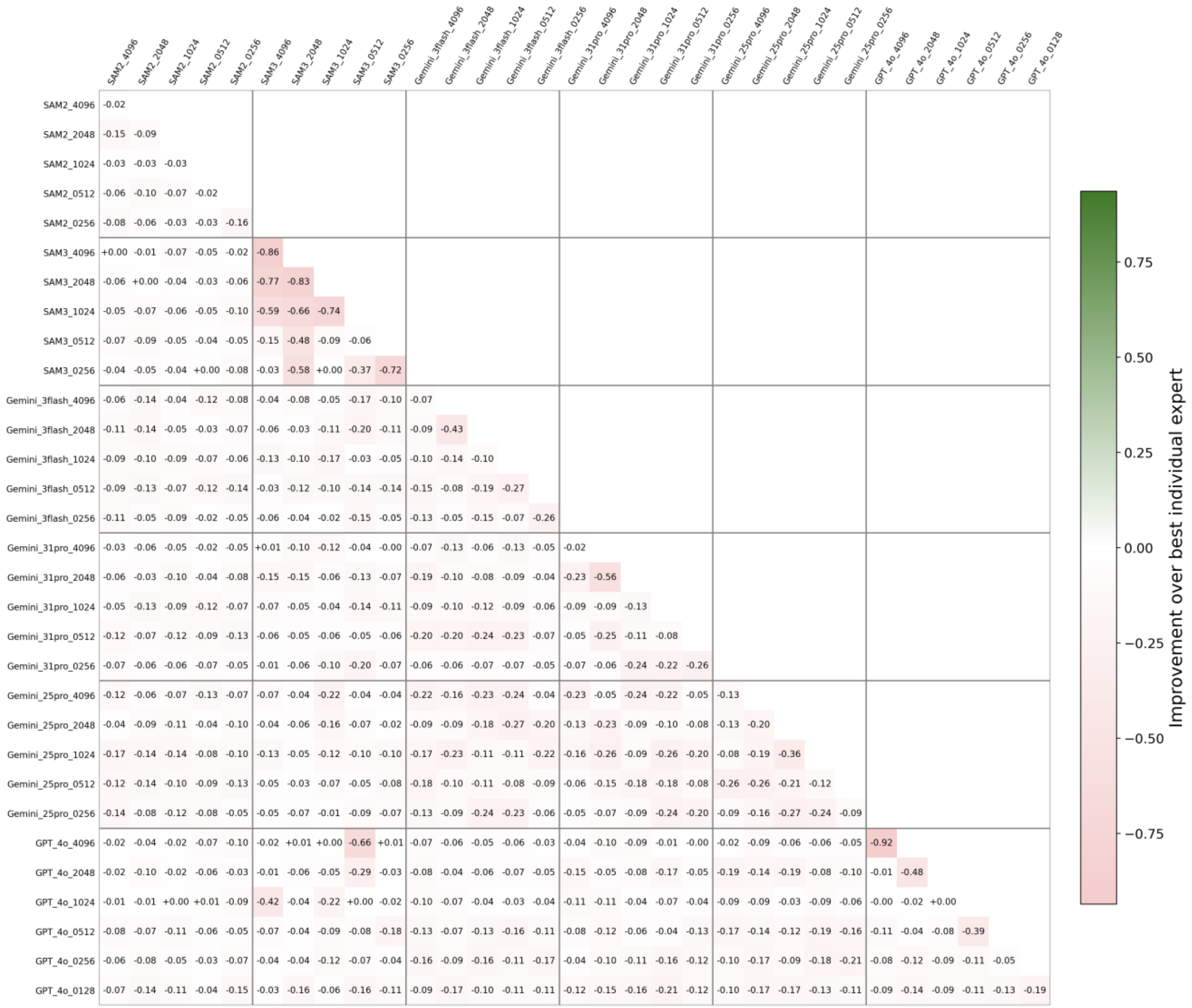


▲Pairwise fusion improvement over best method (SP, R@8)

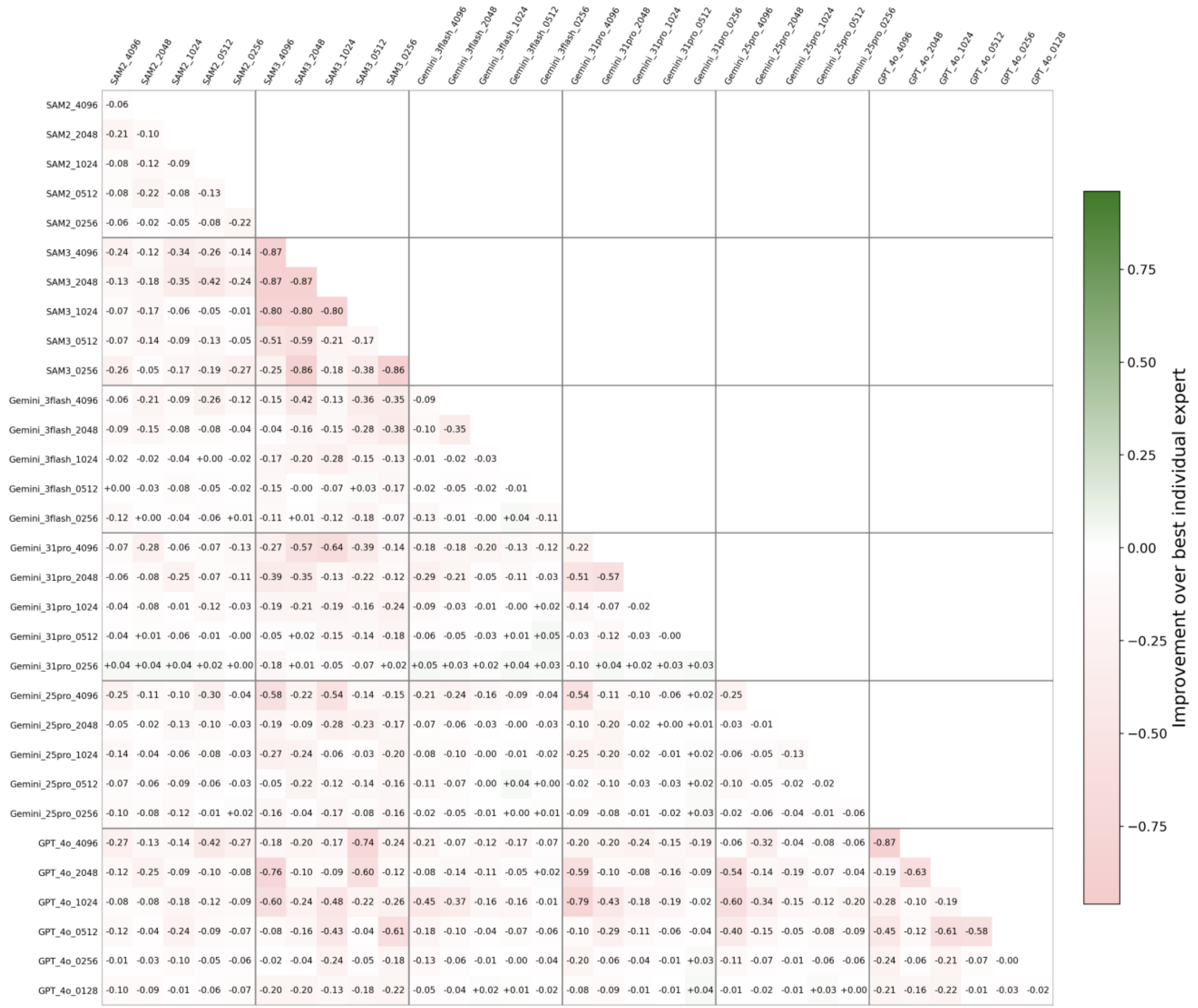
Obtain improvement with a slightly shifted pattern

Bounded by extremely high R@8 (and low P@8) of LOAM

Evaluation: Pairwise Fusion Improvement – WR (cont.)



▲Pairwise fusion improvement over best method (WR, P@8)



▲Pairwise fusion improvement over best method (WR, R@8)

Slight decrease in P@8 compared to LOAM

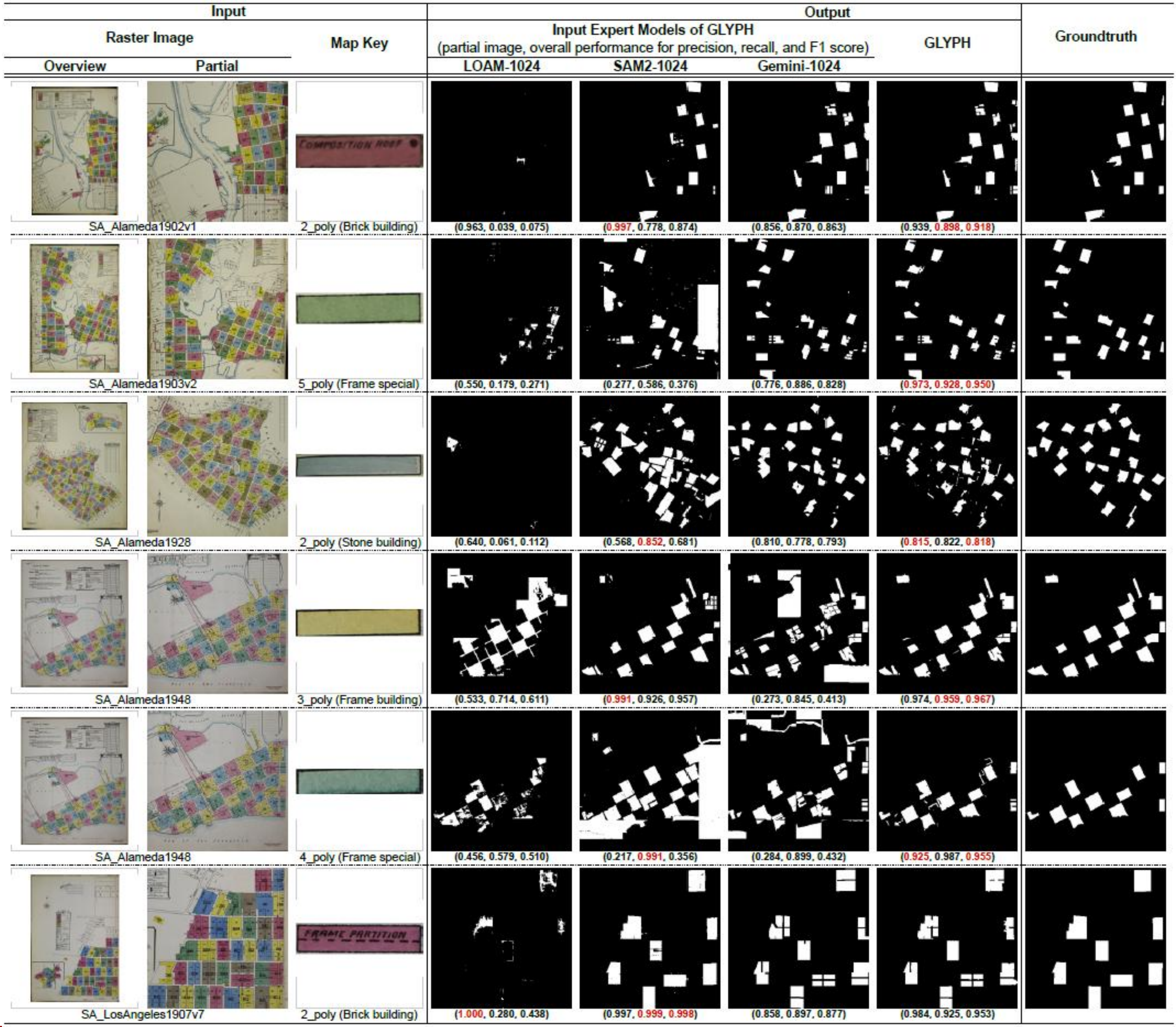
Bounded by high pixel-based accuracy of LOAM

USC INFORMATION SCIENCES INSTITUTE

USC Viterbi School of Engineering

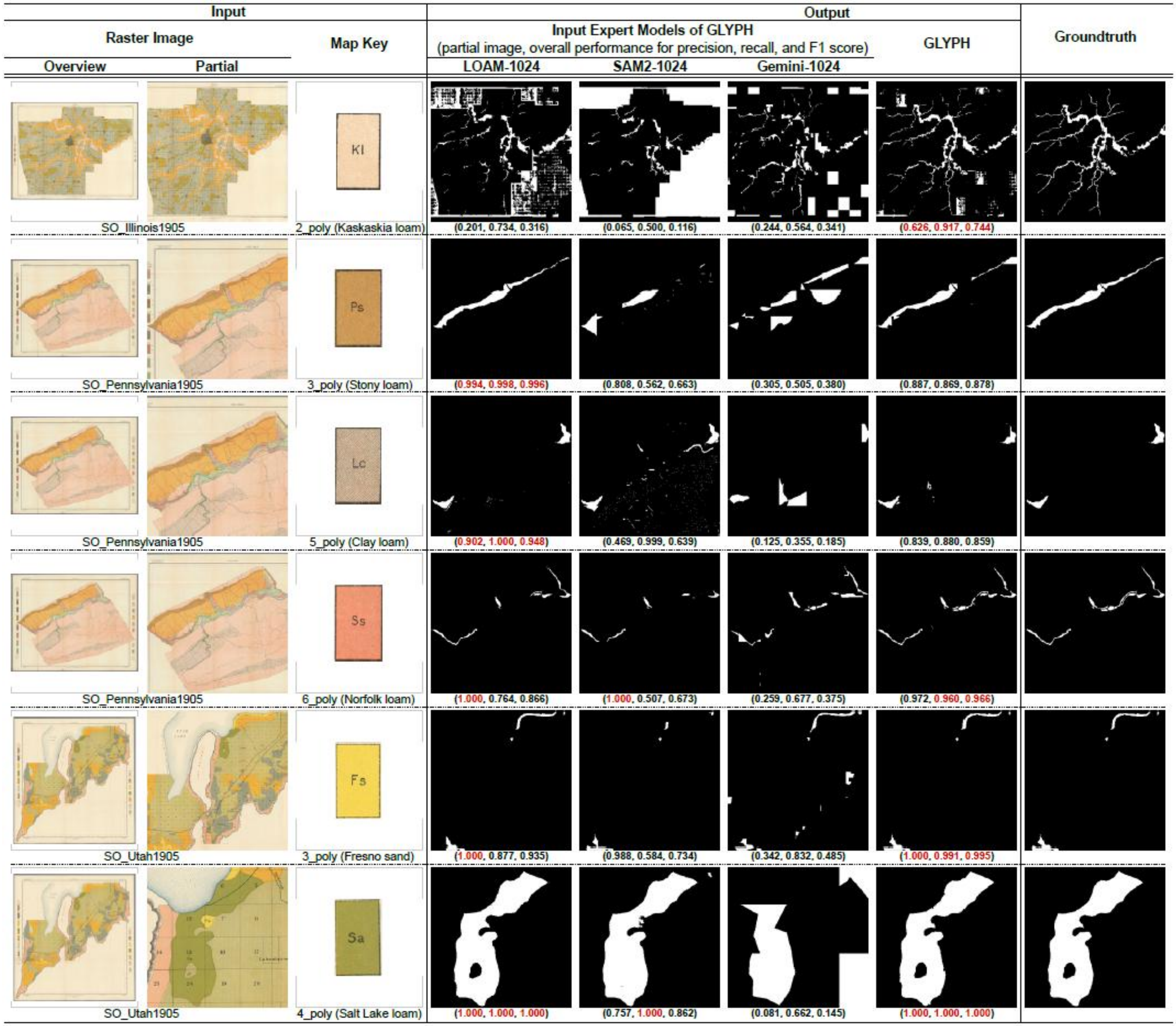
Evaluation: Case Study

- Input
  - LOAM often fails in recall due to the strong text interference
  - SAM2 performs well in geometric delineation but exhibits unstable assignment across categories
  - Gemini 3 Flash captures semantic intent but produces polygon boundary irregularities and occasional merging
- GLYPH
  - combines two or more expert models with low accuracy into decent outputs



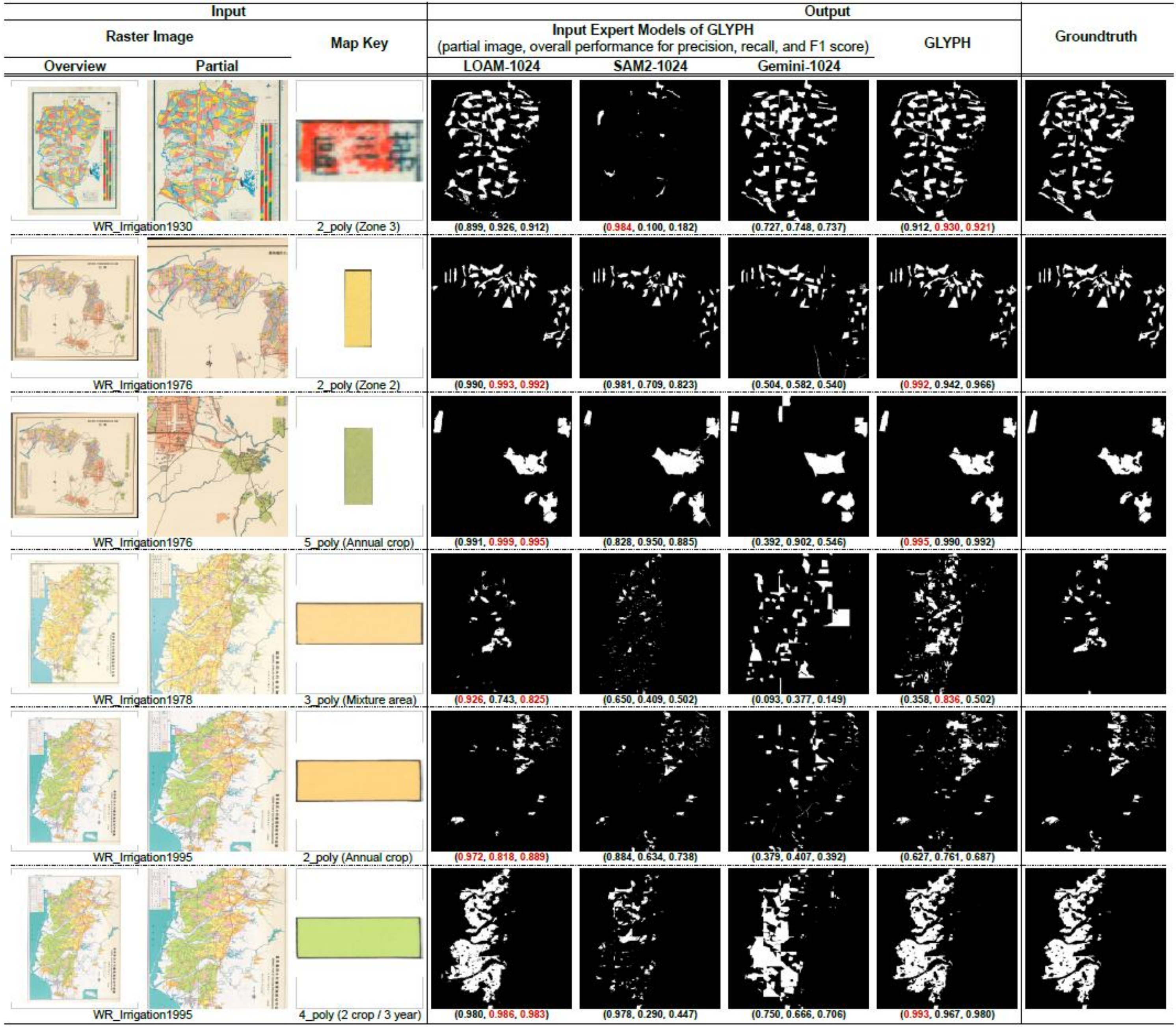
Evaluation: Case Study

- Input
  - LOAM frequently achieves near-perfect precision in high-contrast map keys but over-extension
  - SAM2 captures polygon boundary continuity well has several small segments
  - Gemini 3 Flash has fragmented masks with moderate recall but poor geometric coherence
- GLYPH
  - Achieve qualitative improvement by recovering boundary continuity while suppressing interior noise



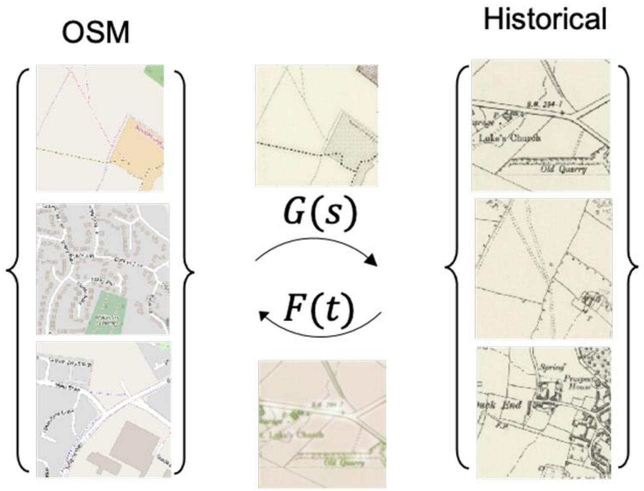
Evaluation: Case Study

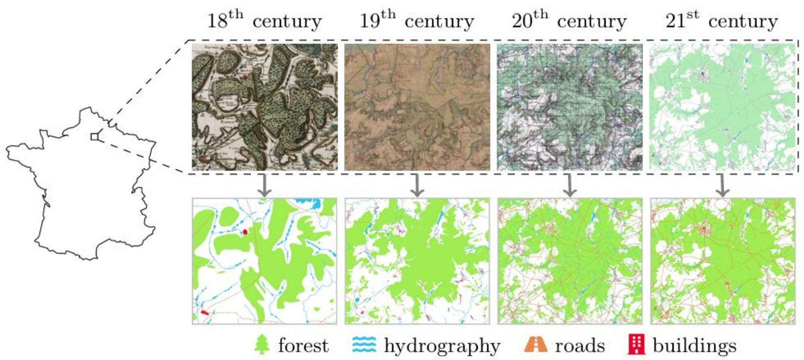
- Input
  - LOAM maintains structural coherence and achieves the most balanced results
  - SAM2 tends to have limited recall with cluttered polygon features
  - Gemini 3 Flash produces mediocre results with poor geometries
- GLYPH
  - Correct under-segmentation while removing spurious fragments
  - Emphasis on segmentation leads to an increase in false positives for ambiguous map keys or polygons accidently linked due to overlaps

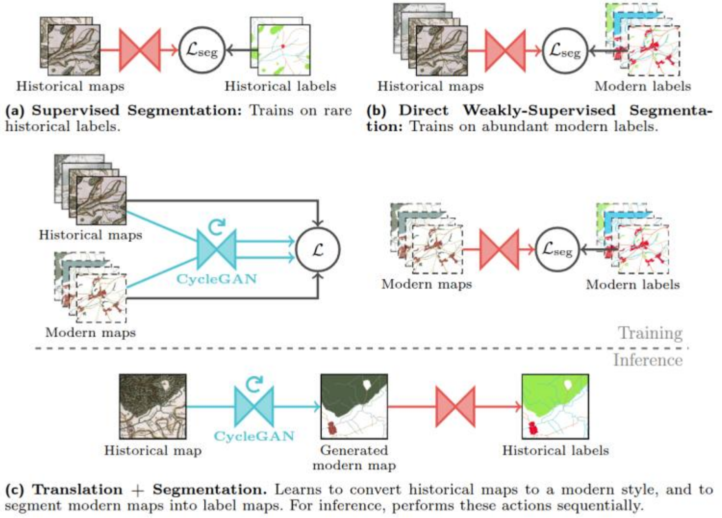


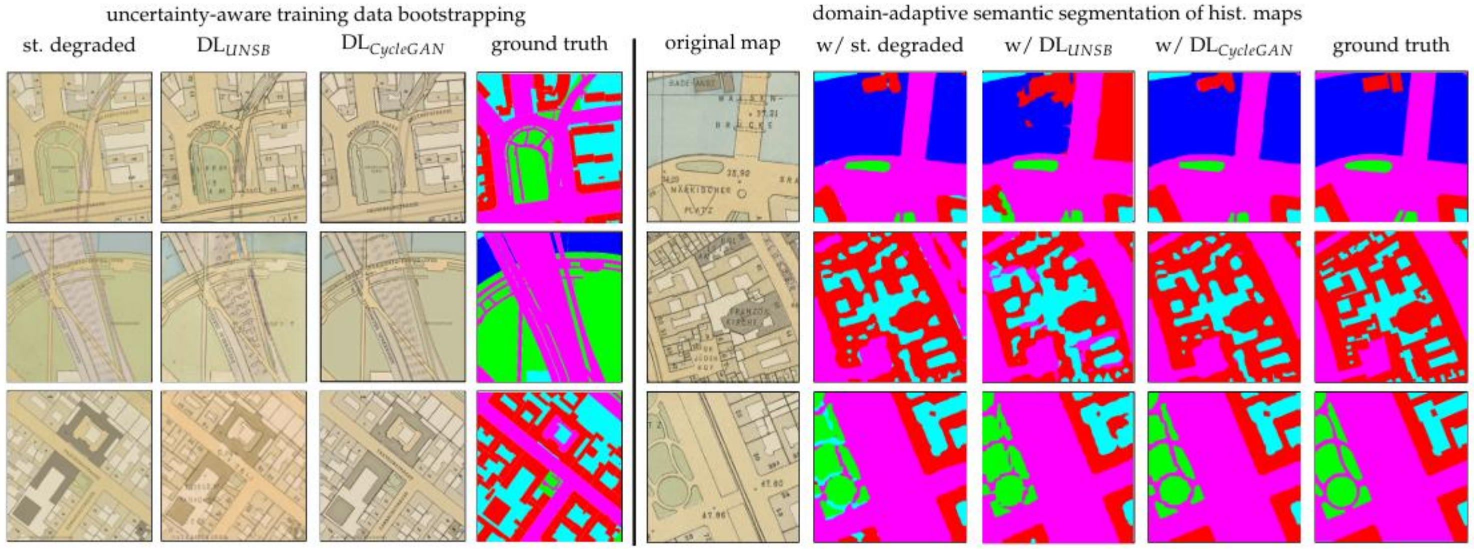
Related Work: Synthetic Data Generation

- Generate synthetic maps following certain cartographic style









Ref: Synthetic Map Generation to Provide Unlimited Training Data for Historical Map Text Detection. ACM SIGSPATIAL, 2021.
Ref: Segmenting France Across Four Centuries. arXiv, 2025.
Ref: Automatic Uncertainty-Aware Synthetic Data Bootstrapping for Historical Map Segmentation. arXiv, 2025.

Related Work: MapSAM2

- Turn image tiles into a time series
  - Belonging to the same map series
  - Augmentation with artifacts
- Leverage the memory attention in SAM2
  - To process image tiles with seen map styles
- Require training data for YOLO detection
  - Improvement when training data is extremely limited

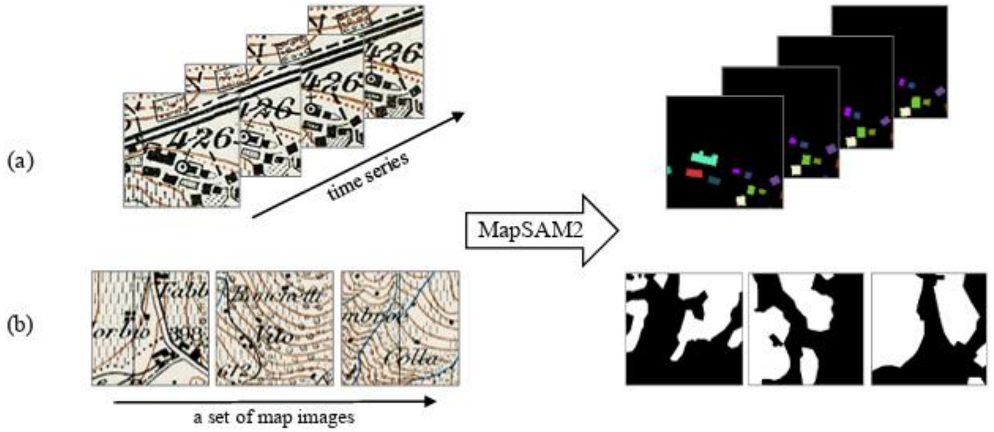
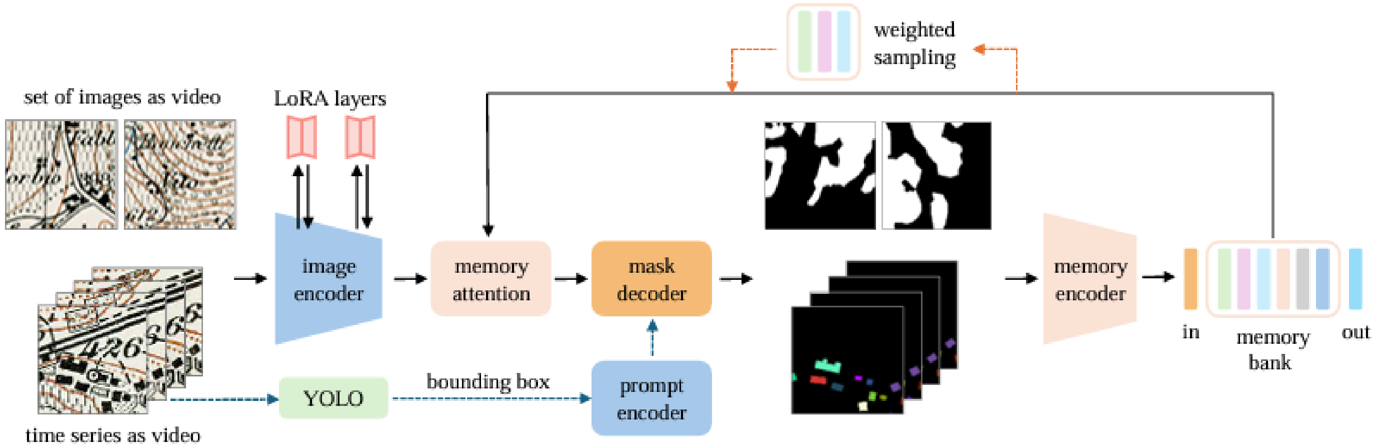


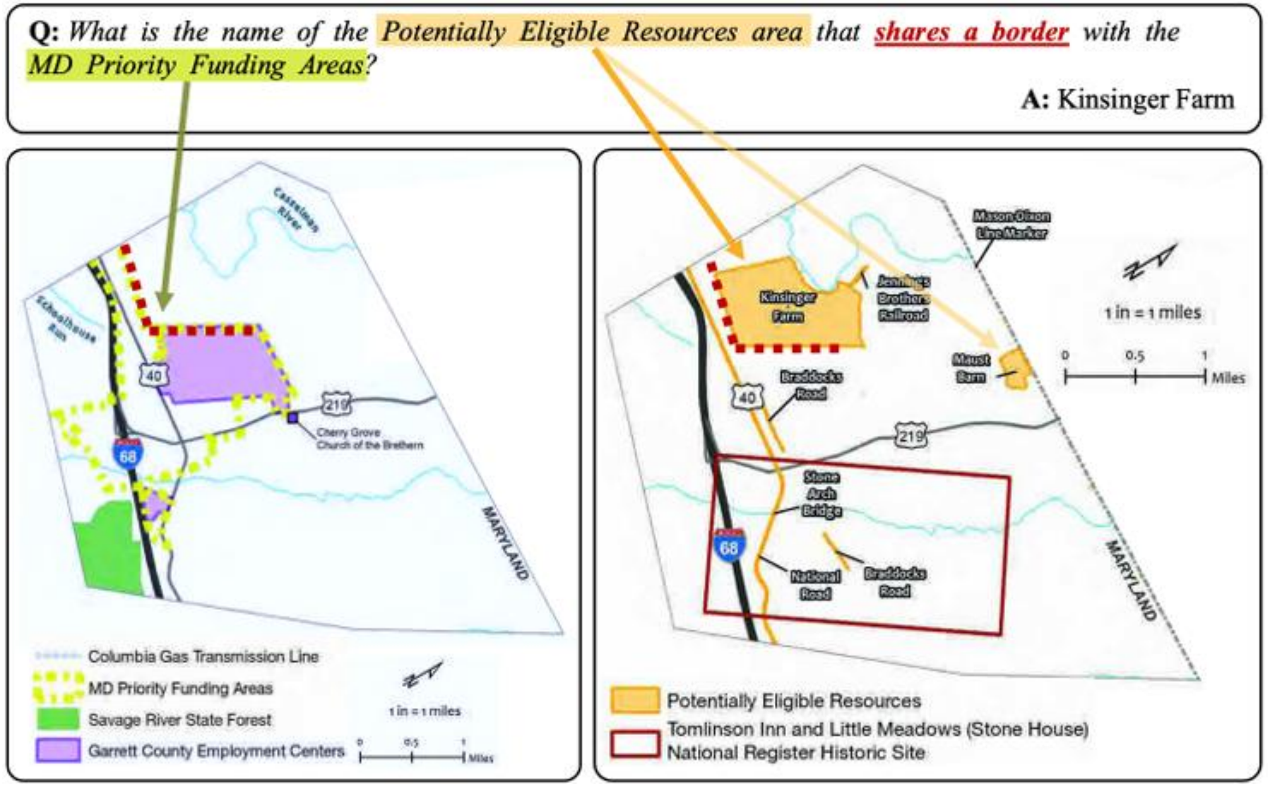
Table 1: Image segmentation accuracy (IoU) on the Siegfried Railway, Vineyard, and ICDAR 2021 Building Block datasets under full and few-shot settings.

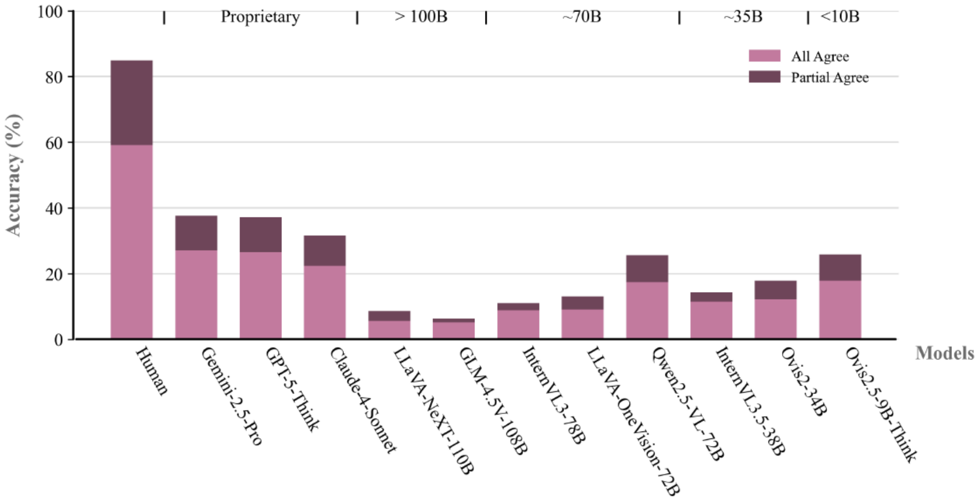
Method	Railway (5872)				Vineyard (613)		Building Block
	Full	10%	1%	10-shot	Full	10-shot	10-shot
U-Net (Ronneberger et al., 2015)	**91.9**	**90.6**	83.5	61.4	77.0	60.2	60.0
PerSAM (Zhang et al., 2023)	–	–	–	5.9	–	22.7	16.0
Few-Shot SAM (Wu et al., 2023)	–	–	–	35.8	–	46.8	15.5
SAMed (Zhang & Liu, 2023)	86.3	85.7	86.0	75.4	74.9	61.5	70.3
MapSAM (Xia et al., 2025)	89.5	88.7	**86.5**	**78.5**	74.3	60.0	71.1
MapSAM2 (Ours)	90.9	89.8	84.7	73.0	**77.3**	**67.6**	**75.8**



Related Work: VLM Reasoning on Maps

- VLMs have shown promising results in map VQA
  - We use their adopted VLMs for comparative methods





	Overall (500)	Border (71)	Distance (91)	Equal (54)	Intersect (80)	Orientation (89)	Within (115)
Human Average	**84.87**	**89.00**	**78.28**	**89.10**	**85.53**	**91.80**	**88.08**
Proprietary LVLMs							
Gemini-2.5-Pro	**38.20**	32.39	25.27	33.33	28.75	**71.59**	**35.34**
GPT-5-Think	37.20	25.35	**27.47**	**44.44**	**31.25**	69.32	28.45
Claude-Sonnet-4	31.60	**33.80**	23.08	37.04	22.50	56.82	21.55
Open Source LVLMs							
LLaVA-NeXT-110B	8.60	4.23	10.99	11.11	16.25	0.00	9.48
GLM-4.5V-108B	6.40	5.41	2.15	21.57	6.17	1.16	7.83
InternVL3-78B	11.00	1.41	4.40	12.96	5.00	34.09	7.76
LLaVA-OneVision-72B	13.00	9.86	10.99	5.56	8.75	29.55	10.34
Qwen2.5-VL-72B	25.60	11.27	14.29	25.93	17.50	55.68	25.86
InternVL3.5-38B	14.20	11.27	8.79	14.81	2.50	36.36	11.21
Ovis2-34B	17.80	25.35	13.19	25.93	26.25	2.27	18.97
Ovis2.5-9B-Think	25.80	12.68	20.88	24.07	22.50	51.14	21.55

Ref: FRIEDA: Benchmarking Multi-Step Cartographic Reasoning in Vision-Language Models. ICLR, 2026.

End of Appendix