HISTORICAL POSTCARDS RETRIEVAL THROUGH VISION FOUNDATION MODELS

Anis AMRI & Salvatore Tabbone

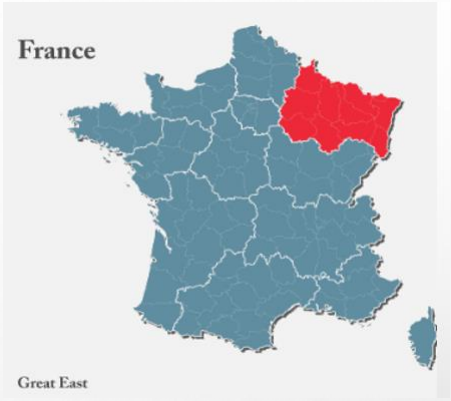
IDMC, University of Lorraine, France

SUMAC '24: Proceedings of the 6th workshop on the analySis, Understanding and proMotion of heritAge Contents.

SUMMARY

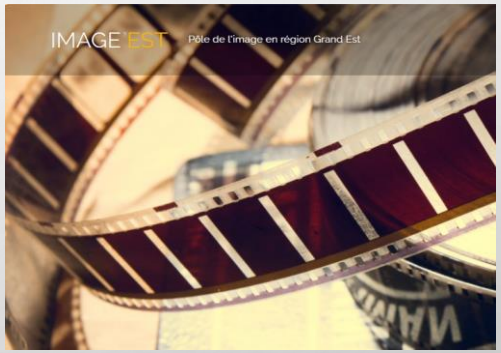
- INTRODUCTION
- METHODOLOGY
- RESULTS

INTRODUCTION



- **Data**: A collection of **historical postcards** from the Grand East of France Region

  From **20th century** belonging to **IMAGE'EST**, an association from **Grand East France**, dedicated to collecting, digitizing, and preserving the audiovisual, cinematographic, and iconographic heritage of the region. This association works to protect and promote local film and photo archives.



- **Objective**: An application that uses computer vision techniques to create a retrieval system for exploring postcard collections.

- **Methodology**:

  Text extraction using **OCR** techniques.

  Use of **Vision Foundation Models** for the **retrieval process**.


Cliché G. Durand, phot.
CHALONS-SUR-MARNE — Ecole d'Arts et Métiers — Promotion 1909-1912


CHAMPAGNE MERCIER, ÉPERNAY. Une partie des chantiers de dégorgement — Part view of the disgorging stands


COURCY. - La Gare


GRAUVES-en-CHAMPAGNE — Centre du Pays


1. - NANCY. - Arc de Triomphe - Rue Héré


LA GRANDE GUERRE. — Le Crime de Reims. — Rue Colbert bombardée et incendiée.
The Crime of Reims. — Colbert street bombarded and burnt down.


Les Vosges.
62. — Lac de Gérardmer. - En plein Lac.


CHAMPAGNE MOËT & CHANDON (Epernay)
MARNE
GILLET
A.LECOUVREUR
COLBERT
REIMS
EPERNAY
CHALONS S-MARNE
STE MENEHOULD
VITRY LE FRANÇOIS
A VISITER
Nos Départements. — 51.
Épernay - Caves Moët et Chandon.

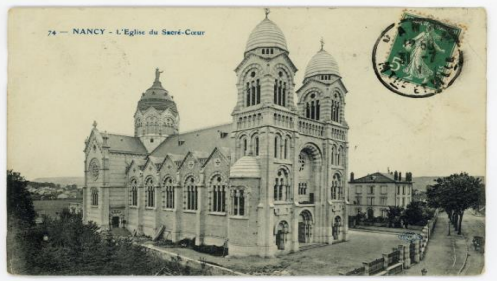

74 — NANCY - L'Eglise du Sacré-Cœur


HERPY (Ardennes)
Souvenir de la Fête des Élections Législatives 1910
A. Wilmet, phot.-édit., Rethel


71. - TROYES. - Pont et Rue de la Cité - Hôtel-Dieu
PHARMACIE
Collection T. G.


BOULT-SUR-SUIPPE. - L'Église du XIIIe Siècle.


MUIZON (Marne). — L'Eglise.


74 — NANCY - L'Eglise du Sacré-Cœur


La Lorraine Illustrée.
73. — NANCY. - L'Église du Sacré-Cœur et la Cure d'Air Saint-Antoine.


Charot. — L'Eglise.
Mille baisers
Alice Barbier

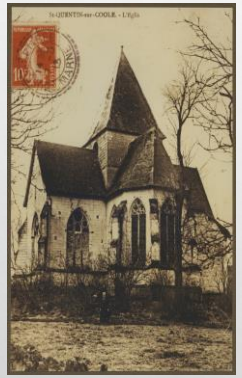

FRESNES — L'Eglise

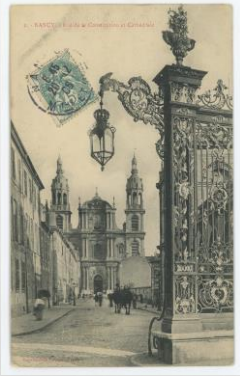

GIREY-SUR-BLAISE — L'Eglise — La Place

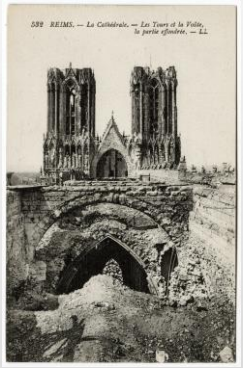

RAISNES-en-SAINTOIS (Vosges). - Rue de l'Eglise


284 Environs de REIMS
GUEUX — L'Eglise et le Lac


LA CHEPPE (Marne). — L'Eglise.


St-QUENTIN-en-COOLE - L'Eglise



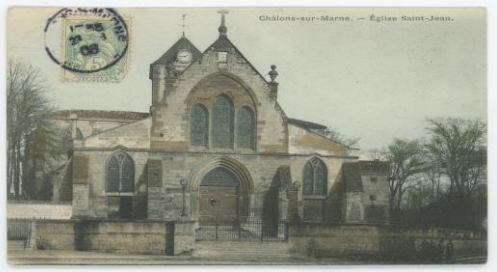

592 REIMS. — La Cathédrale. — Les Tours et la Voûte, la partie effondrée. — LL.


135. - NANCY. — Eglise St-Nicolas


81. — Reims. - Place Royale.


CHALONS-sur-MARNE - La Cathédrale


160. REIMS — La Cathédrale et la Rue Lebergier


Châlons-sur-Marne. — Église Saint-Jean.


26. - NANCY. - La Gare - Eglise Saint-Léon - Faubourg Stanislas


REIMS — Rue Colbert


TROYES — L'Avenue de la Gare et l'Entrée de la Rue Thiers, vue prise du haut de la Nouvelle Gare
S. Bruncair, édit., Troyes


BAROUES
7a. GERARDMER. Le Canotage sur le Lac.
Clichés et Edition A. Largeron. Bijouterie-Souvenirs
POSTES
5c
RÉPUBLIQUE FRANÇAISE

- ✓ Printed text

- ✓ Printed text
- ✓ Postmark
- ✓ Postage stamp

- ✓ Printed text
- ✓ Postmark
- ✓ Postage stamp
- ✓ Handwritten text

METHODOLOGY

Our application uses computer vision techniques to create a search system designed for exploring historical postcard collections:

- Text extraction using **Optical Character Recognition (OCR):**
  Testing and comparing two solutions:
  - **EasyOCR**
  - **Tesseract-OCR**

- Use of **Vision Foundation Models (VFM)** for the retrieval process:
  Testing and comparison of two models:
  - **CLIP** (86M parameters) by Radford et al. (2021)
  - **DINOv2** (86,5M parameters) by Oquab et al. (2023)

Image Retrieval Systems

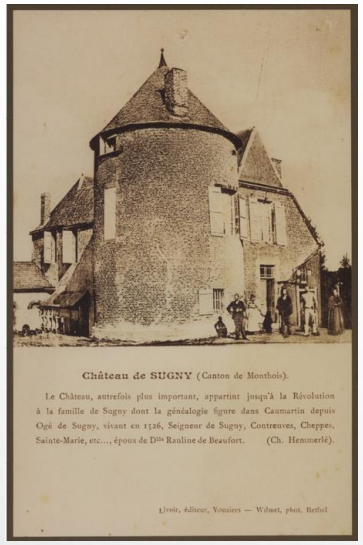
Image retrieval systems have improved a lot, making it easier to find images similar to what users want.

How It Works:

They use vector spaces to store image features. When you search for an image, the system extracts features from your query and compares them to its database to find the best matches.

Types of Retrieval:

- **Category-Level Retrieval:** Finds images in the same category as your query.
- **Instance-Level Retrieval:** Searches for specific objects, even if they are shown in different conditions (like time or angle).



An Input Image

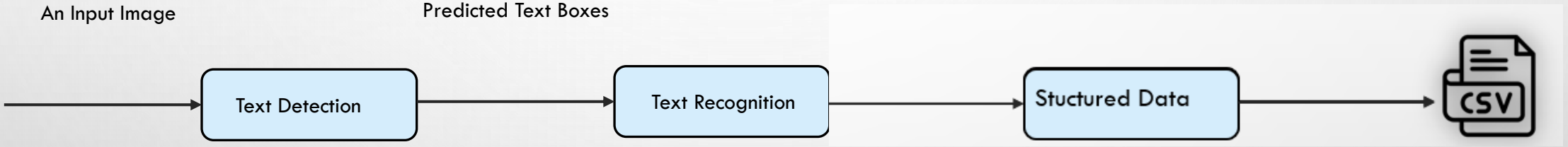
Predicted Text Boxes

Château de SUGNY (Canton de Monthois).

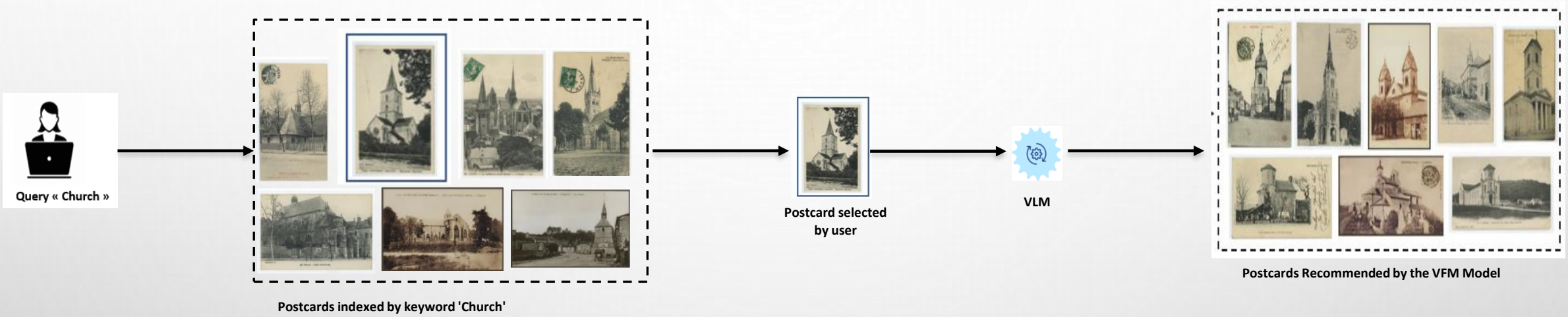
Le Château, autrefois plus important ; appartint jusqu'à la Révolution à la famille de Sugny dont la généalogie figure dans Caumartin depuis Ogé de Sugny vivant en 1526, Seigneur de Sugny, Contreuves, Cheppes, Sainte-Marie, etc.., époux de Dlle Rauline de Beaufort. (Cl. Henmerlé).

Livoir, éditeur, Vouziers Wilmet, Rethel, phot,

Postcard #325
Text Extract: Château de …
Title Card: Château de Sugny
City/village: Sugny (Ardennes)
Keyword: Château



THE PROCESS OF EXTRACTING TEXT FROM HISTORICAL POSTCARDS USING THE OCR SYSTEM.



Example scenario of image searching: When a user asks a question, such as "church", they are presented with a collection of postcards indexed by the keyword "church". After selecting a postcard, the system subsequently recommends a collection based on the VFM.

RESULTS

EVALUATION METRIC

✓ **Character Error Rate (CER):** Calculates the percentage of characters that are incorrectly recognized or missing.

$$\mathbf{CER} = \frac{S+D+I}{S+D+C},$$

Where **S** is the number of substitutions, **D** is the number of deletions, **I** is the number of insertions, **C** is the number of correct characters, and **N** is the total number of characters in the references.

✓ **Jaccard Similarity:**

$$\mathbf{J(A,B)} = \frac{|A \cap B|}{|A \cup B|},$$

Where **A** represents the ground truth text and **B** represents the recognized text.

EVALUATION METRIC

✓ Mean Average Precision (mAP):

$$\text{mAP@} = \frac{1}{I}\sum_{q=1}^{I} AP@k(q)\,,$$

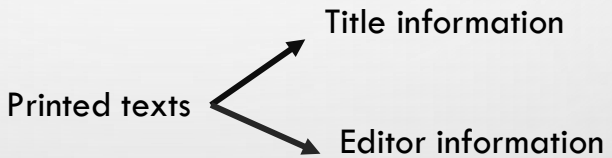
Where **I** is the number of query images in the set and **AP@k** Average precision at rank k for query image **q,** defined as:

$$\text{AP@} = \frac{1}{\min(m_q,k)}\sum_{i=1}^{\min(mq,k)} P_q(i).rel_q(i),$$

Where $\boldsymbol{m_q}$ is the number of ground-truth index images matched with query image q; $\boldsymbol{n_q}$ is the number of predictions made by the retrieval method; $\boldsymbol{P_q}$ is the precision at rank **k** for query image **q** (this evaluates the proportion of retrieved relevant images among the top **k** retrieved images for a given query image); $\boldsymbol{rel_q(i)}$ **is** a relevance indicator function which equals **1**if the result at rank **I** is relevant and equals to **0** otherwise.

Comparison of EasyOCR and Tesseract-OCR

	Jacc. Sim	CER
EasyOCR	0. 63	0.42
Tesseract-OCR	0.29	0.75



Comparison between different approaches and their configurations

	ENCODER	OTHER Configurations	mAP_{PI}@5	mAP_{SC}@10
Clip	VIT-B-32	openai	39.41	52.25
		laion400m_e31	43.65	55.19
		laion400m_e32	43.65	56.23
		laion5b_s13b_b90k	47.43	56.75
	VIT-B-16	openai	48.90	56.51
		laion400m_e31	53.57	55.48
		laion400m_e32	54.04	57.39
		laion2b_s34b_b88k	46.71	57.55
	VIT-L-14	openai	48.46	57.85
		laion400m_e31	49.13	55.25
		laion400m_e32	49.40	56.57
		laion2b_s32b_b82k	47.57	55.82
	VIT-G-14	laion2b_s12b_b42k	52.86	57.59
		laion2b_s34b_b88k	51.48	57.19
	Vit-bigG-14	laion2b_s39b_b160k	54.38	58.75
DINOv2	VIT-S-14		22.48	56.59
	VIT-S-16		20.51	56.59
	VIT-B-14		21.23	56.72
	VIT-B-16		24.97	56.72

Conclusion

- ✓ Historical postcards are a valuable source of information, but traditional research methods do not always take full advantage of their richness. With recent models, we can better extract and analyze visual and textual information. We can also test other VLMs.

- ✓ In our study, we compared two OCR tools for extracting text and found that EasyOCR works better than Tesseract-OCR. We also found that vision models using CLIP are a little better than those using DINOv2.

- ✓ In summary, our application will be beneficial for libraries and museums, making it easier for researchers to access historical data. We will continue to work on optimizing data extraction to improve our results in the future.

Perspectives

In our future work,

- ✓ We plan to use computer vision models to extract a broader and more diverse range of information from postcards.

- ✓ We can also analyze the textual information on postmarks, particularly two key elements: the sending date and the city of origin (as shown in the figure).

- ✓ Another research direction involves applying a Handwritten Text Recognition (HTR) model to extract the addresses written on the back of postcards. This will allow for deeper analysis of the postcards and help establish connections with cultural categories.

- ✓ We also aim to classify the postcards into different categories, taking both visual and textual aspects into account.


THANK
YOU !