Content-Based Image Retrieval System for Museum Art

Master's Degree in Computer Vision

C1 Course — Introduction to Human and Computer Vision

Team 1

Team 1

Adrián García Tapia, 1568249

Maiol Sabater Oliveras, 1634564

Júlia Garcia Torné, 1630382

Gerard Vilaplana Pelegrí, 1776578

Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

Global System Pipeline

Goal: Retrieve artworks in a museum database using visual similarity.

Input: Query set (QSD1-W4) and museum database images.

Output: Ranked list of most similar paintings for each query.

Load and preprocess images from the database and query sets → **Apply masks to extract paintings from complex scenes** → **Detect and compute local features** → **Match query descriptors with database descriptors** → **Evaluate performance using MAP@K**

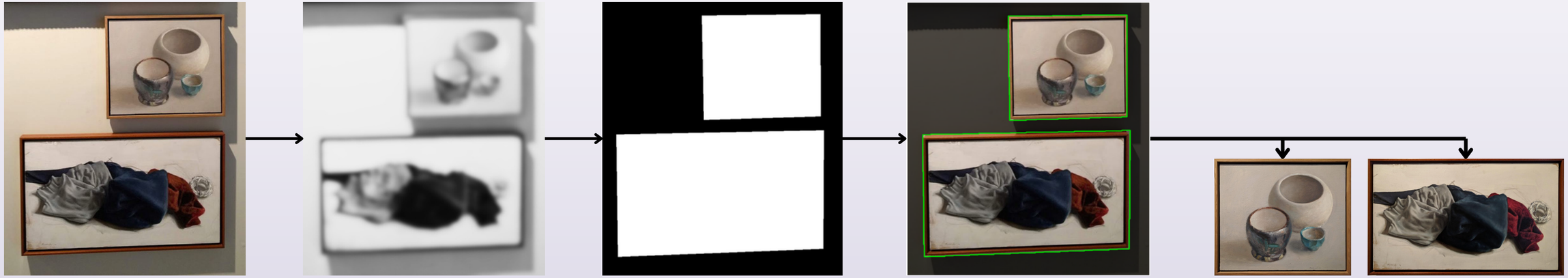
Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

Keypoint Detection and Descriptor Computation

Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

Dataset Loading and Preprocessing

- Images are loaded as **grayscale** and **resized** to a maximum side of **512** pixels.
- **Gaussian blur** applied before resizing to reduce aliasing.
- **Masks** extracted from previous weeks are used to isolate paintings.
- **Connected components** separate multiple paintings within one image.
- Each detected painting is **cropped, masked**, and **stored** individually.



Original Image **Preprocessed Image** **Extracted Mask** **Mask Overlay** **Extracted Individual Pictures**

Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

Mask Extraction

For this step, we used and adapted the **Group 2** code to better fit our specific requirements by allowing the detection of both **horizontal and vertical separations**.

Before Modification

After Modification

Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

Feature Detection and Description

Goal: Extract distinctive visual features from museum and query images.

Techniques tested:

- **SIFT** (Scale-Invariant Feature Transform): detects distinctive keypoints robust to scale, rotation, and illumination changes.
- **ORB** (Oriented FAST and Rotated BRIEF): a fast, rotation-invariant binary descriptor suitable for real-time applications.
- **AKAZE** (Accelerated KAZE): extracts features in a nonlinear scale space, offering a good trade-off between accuracy and efficiency.

Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

Image Matching and Retrieval Pipeline

Brute Force Matcher

- **SIFT:**
  NORM_L2
  NORM_L2SQR
- **ORB:**
  NORM_HAMMING
  NORM_HAMMING2
- **AKAZE:**
  NORM_HAMMING

→

Filter matches using Lowe's Ratio Test

KNN (K=2)

Keep if:
1st match < ratio × 2nd match

→

Verify correctness

Cross-check matches
(DB ↔ Query)
using same Lowe ratio

→

Are there enough verified matches?

Apply a threshold to confirm if image exists in DB.

→

Retrieval

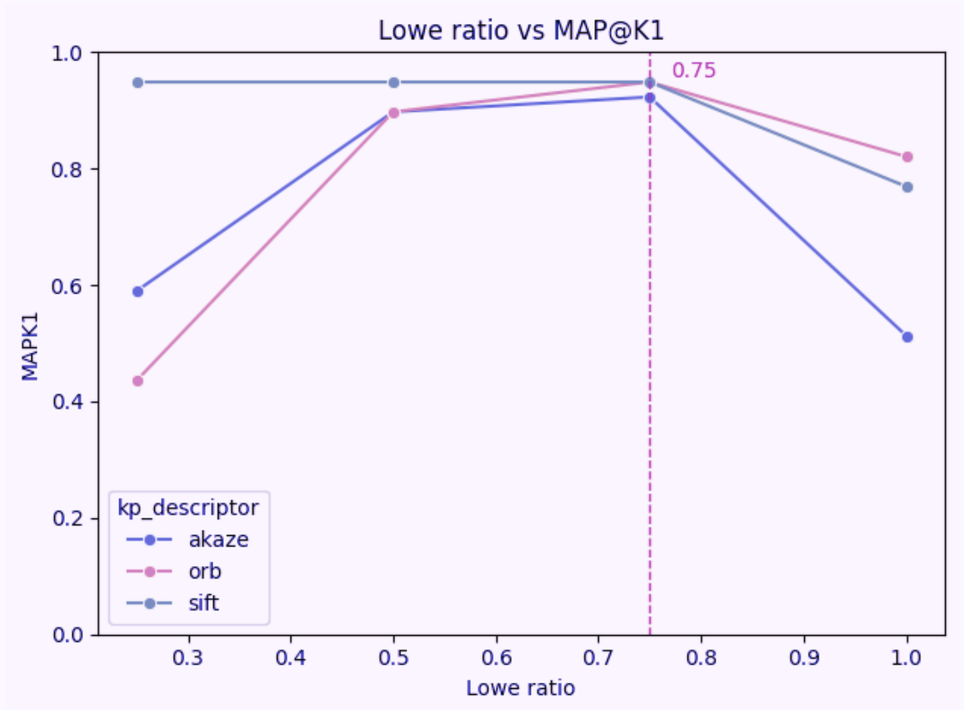
Retrieve the database images sorted by number of good matches.

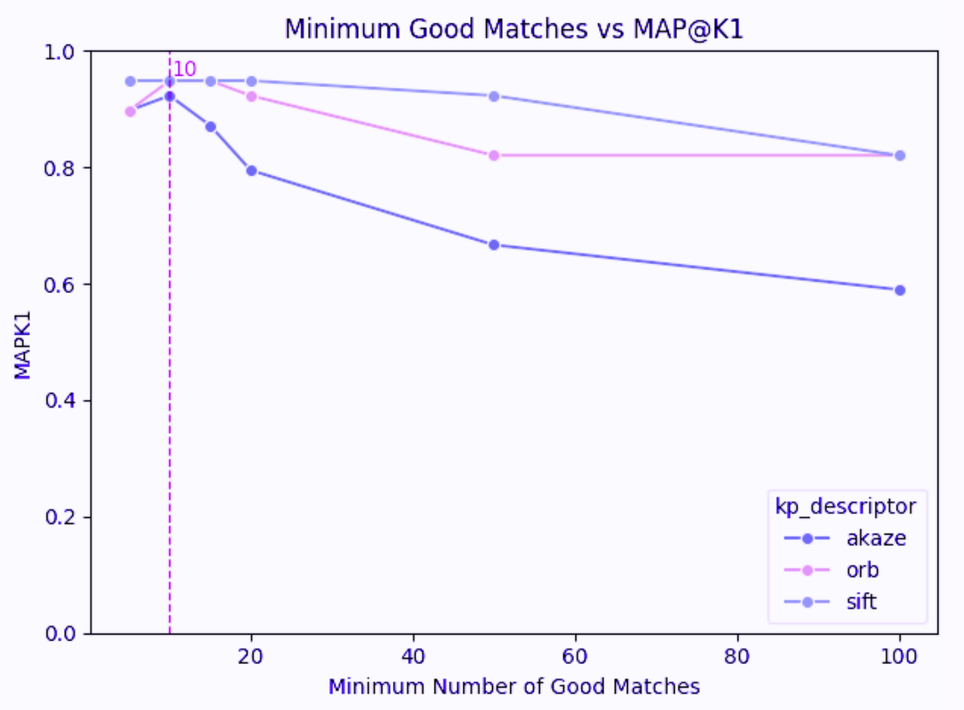
Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

Grid Search: Finding the Optimal Parameters

Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

Matcher Parameters vs. MAP@K1





Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

Optimal Descriptor Parameters

We performed a **grid search** to explore multiple combinations of descriptor parameters systematically.

Best descriptor setups:

ORB

WTA_K: 3
n_features: 500
scale_factor: 1.2
n_levels: 6
(...)

AKAZE

threshold: 0.0001
descriptor_size: 0
descriptor_channels: 3
(...)

Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

System Evaluation (MAP@K)

Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

Evaluation metric

- MAP@K (Mean Average Precision) measures ranking performance.
- High MAP@K → relevant database images retrieved at top positions.

SETUPS

ORB

WTA_K: 3
n_features: 500
scale_factor: 1.2
n_levels: 6
(...)

AKAZE

threshold: 0.0001
descriptor_size: 0
descriptor_channels: 3
(...)

RESULTS

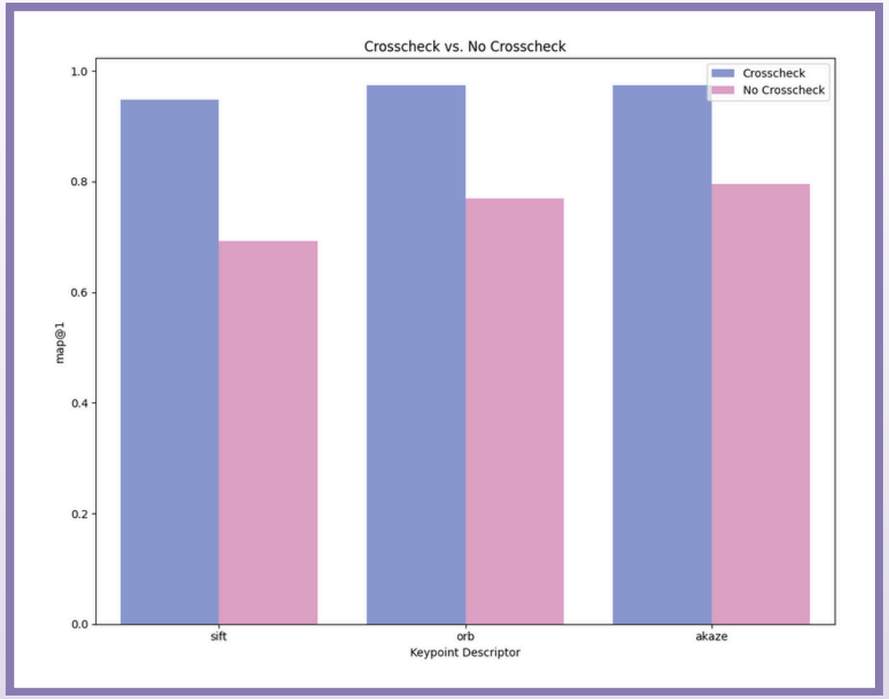
MAP@K1: ≈ 0.974
Correct Predicted Paintings: 38/39

Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

Results and Observations

Key Findings:

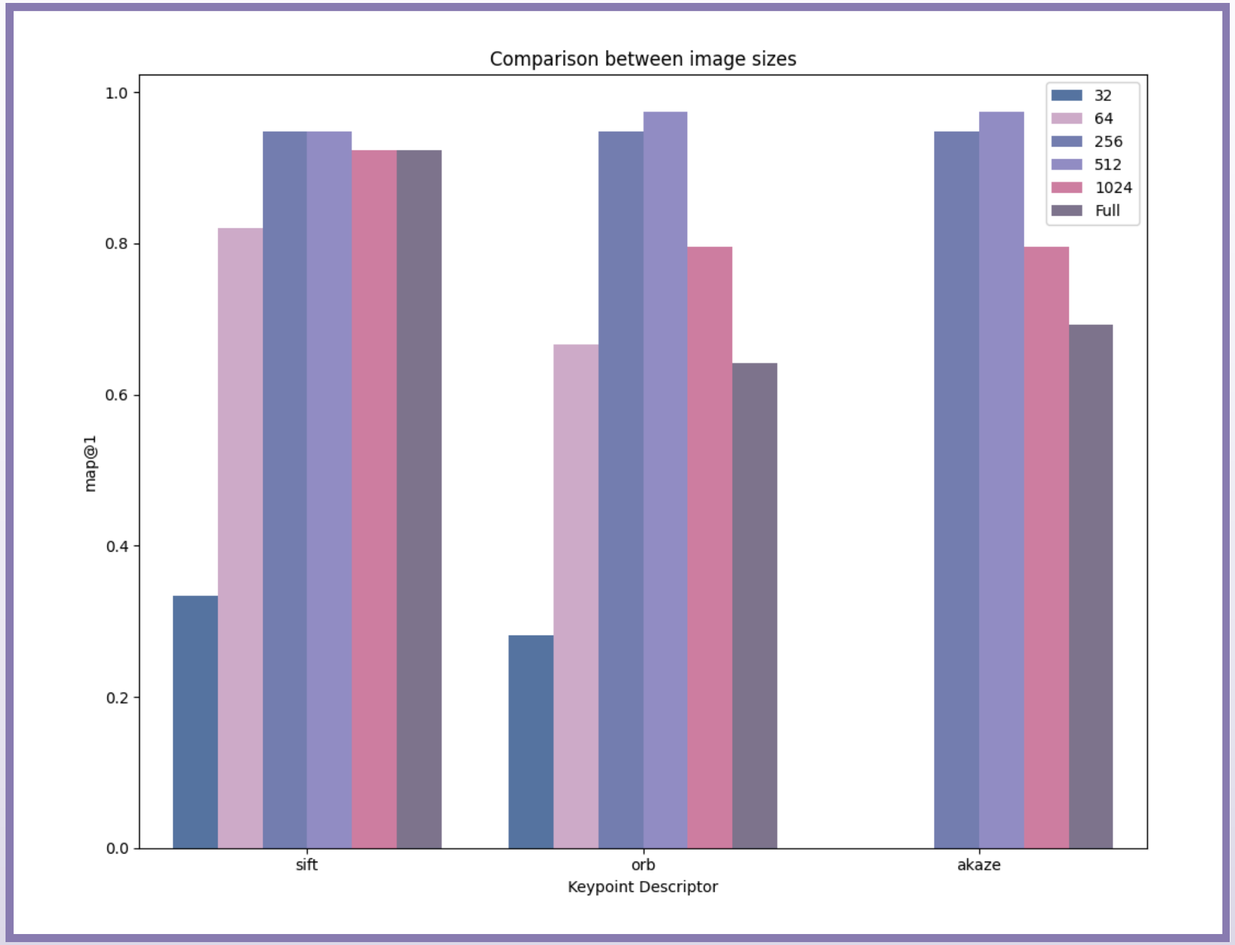
- When it comes to **AKAZE**, lower ***"threshold"*** → more meaningful keypoints and improved retrieval accuracy.
- **Image downscaling + Gaussian blur** → remove noise and minor details, keeping only **essential visual information**.
- By tuning the **Lowe ratio** and **minimum good matches** close to their optimal values, the grid search provides enough variance to achieve **remarkably strong results.**
- As shown in the plot on the right, **Crosscheck** significantly improves retrieval performance.



Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

Results and Observations

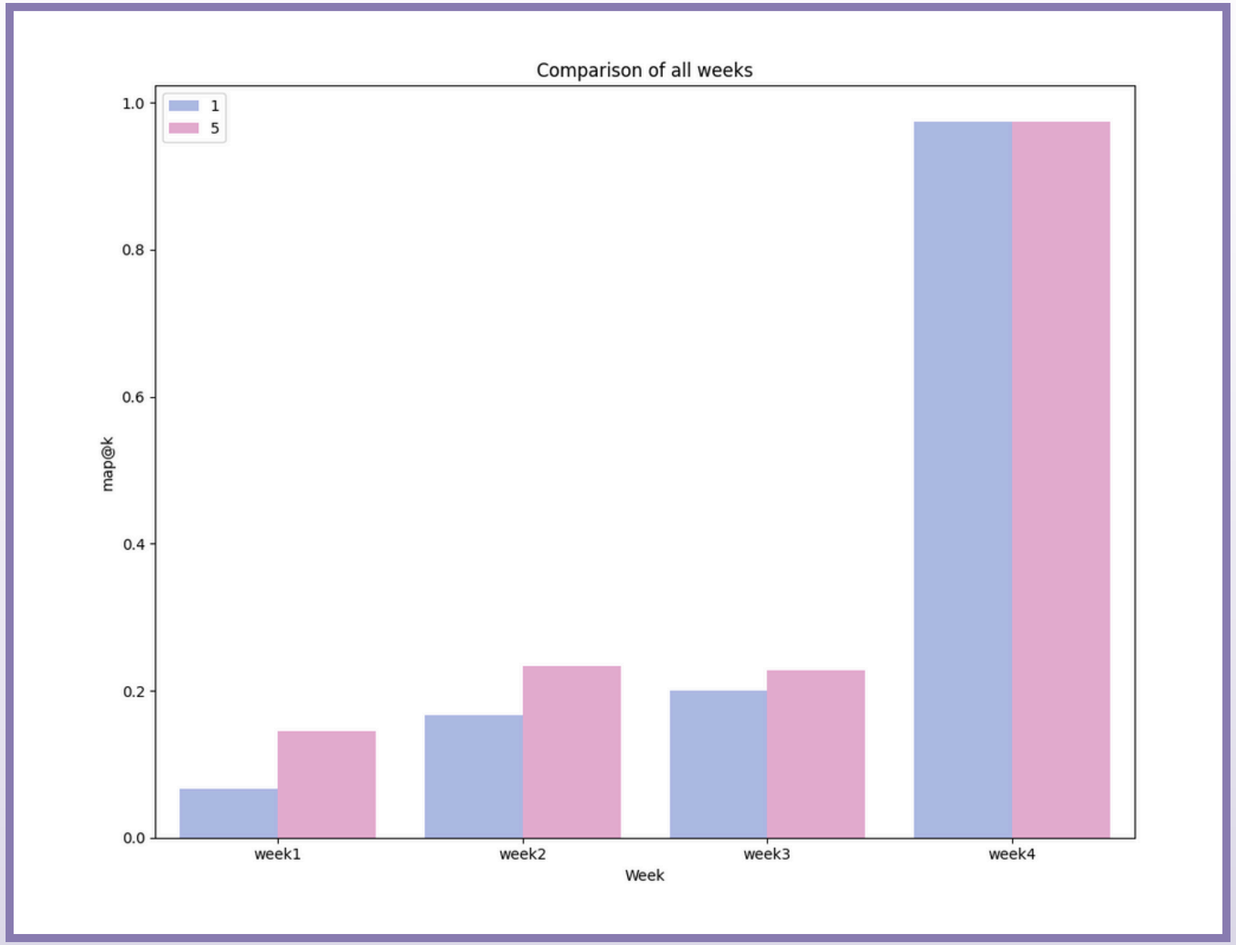
Key Findings:



Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

Results and Observations

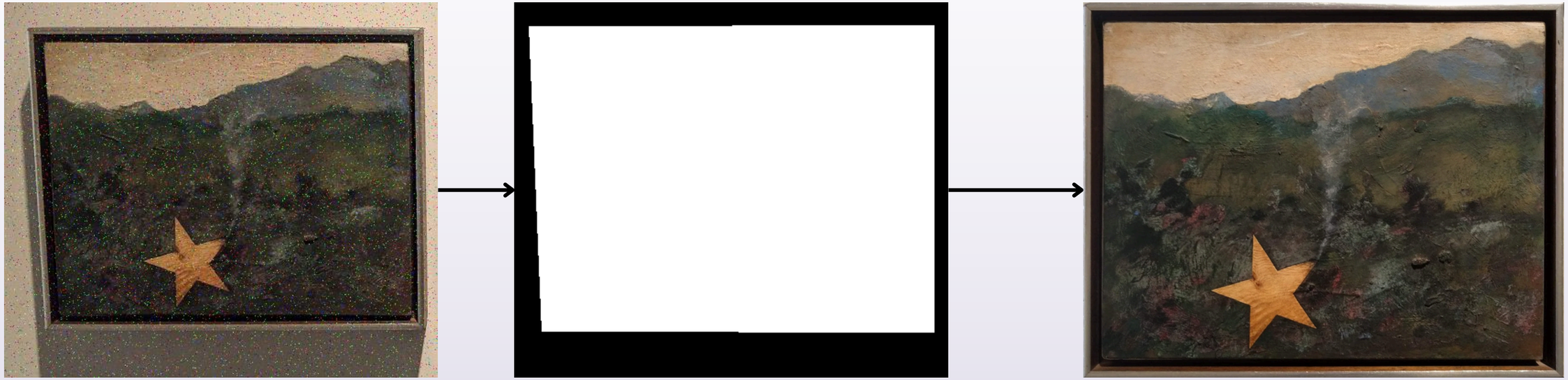
Key Findings:



Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

Results and Observations

Qualitative results

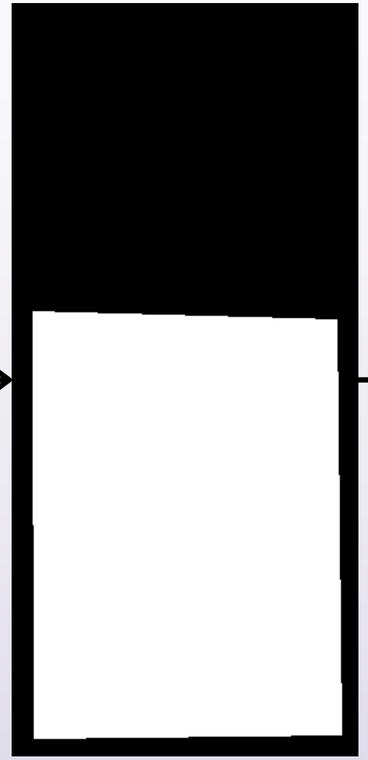


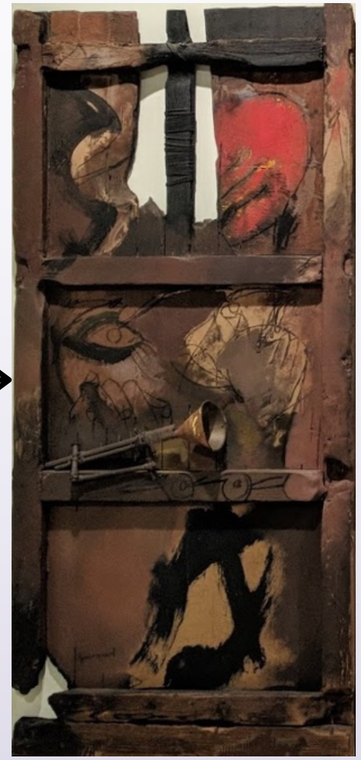
Query Image | Extracted Mask | Retrieved Image

Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

Results and Observations

Qualitative results



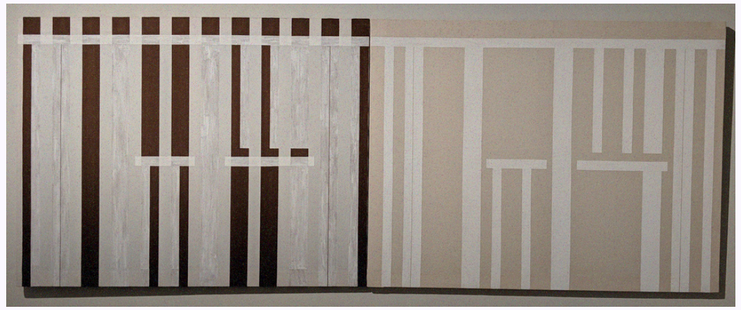


Query Image | **Extracted Mask** | **Retrieved Image**

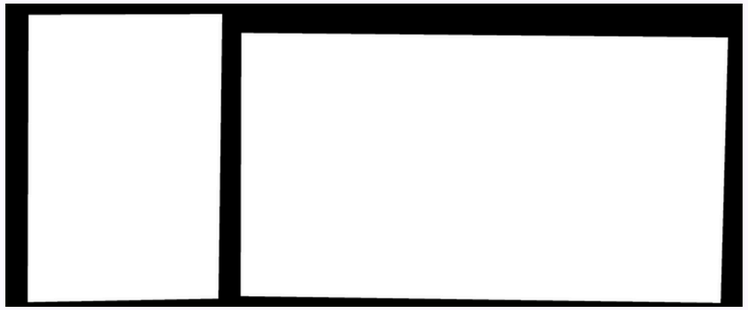
Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

Limitations

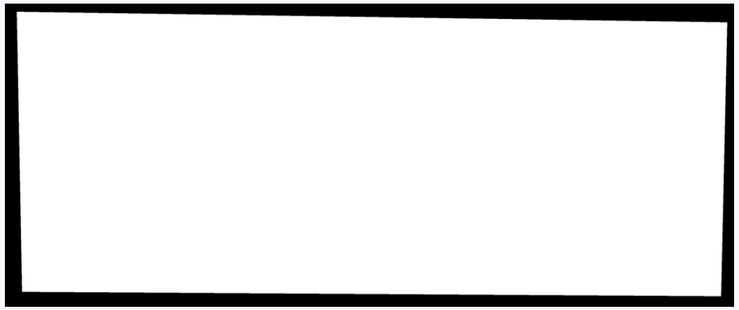
- The method is **mask-dependent** — if the mask fails, the descriptors also fail.
- Incorrect masks lead to inaccurate feature extraction and matching.
- **Future work:** explore deep learning–based segmentation methods to generate more accurate masks and improve overall robustness.



Original Image



Predicted Mask



GT Mask

Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

Limitations

- When the painting surface is flat or lacks texture, feature detectors fail to identify keypoints. As a result, **no reliable descriptors** are generated, and matching becomes impossible. **Blurry or uniform regions** make comparison ineffective.

- **Future work:** investigate feature extraction techniques robust to low-texture or uniform surfaces.

Problematic Samples

Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

Conclusions

- **ORB** and **AKAZE** achieve near-perfect MAP@K1 with optimal settings.
- **SIFT**, while slightly less accurate, remains reliable and stable for retrieval.
- Good **matching conditions** are more important than choosing the right detector.
- The modified mask extraction method **successfully isolates paintings** in most cases.
- The overall retrieval pipeline is **robust and effective**, despite minor limitations.

Group 1:
Gerard Vilaplana Pelegrí,
Adrián García Tapia,
Maiol Sabater Oliveras,
Júlia Garcia Torné

ANY QUESTIONS?