Trends and Applications of Computer Vision

Massimiliano Mancini, Giulia Boato,
Nicu Sebe

A.A. 2025/2026

Learning Goals

Hard skills

- Gain knowledge on advanced computer vision topics
- Deepen one topic through a project-based activity
- Learn to do a literature search on a given topic
- Fully understand the principles and limits of the techniques studied
- Get familiar with practicals tools and codes

Soft skills

- Learn to present formal arguments in talk and written form
- Learn to work in team and to communicate with supervisors

Main Areas

Multimedia Forensics (Giulia Boato)	**T1** Multimedia Forensics, deepfake detection, AI-generated media identification
Vision-Language Models (Massimiliano Mancini)	**T2** Vision-Language Models, compositionality, machine unlearning

Team-based project activity

- Work on one subtopic among a list of proposed ones *(autonomously)*
  - work in small groups
  - get deeper knowledge and explore related work
  - set up experiments with pointed tools

- Present results in the form of presentations *(in class)* and written reports
  - the activity is supervised by one lecturer with fixed milestones
  - the activity is monitored together with Prof. **Nicu Sebe**

Timeline

Fix in your calendar the dates of the presentations!

Date	Plan
29/10	Presentation of projects list
31/10	How to do a literature review
05/11	T1 - Multimedia Forensics
07/11	T2 - Vision and Language
12/11	Seminar on T1
14/11	Seminar on T2
19/11	Lab on T1 / in-progress meeting
21/11	Lab on T2 / in-progress meeting
26/11	**Presentations "Related work and project status" I**
28/11	**Presentations "Related work and project status" II**
03/12	Final in progress meeting (Q&A on projects)
17/12	**Final project discussion I**
19/12	**Final project discussion II**

Project milestones - part 1

Presentation on related work and project status *(in class)*:

- present the initial source(s) as well as the most important related work
- state how and why these sources are important for your topic
- expose the tools you have used and initial experiments you made
- max 15 minutes: ~10 related work + 5 project status
- dates: November 26 and November 28

Intermediate meeting (only if needed - to help you, not an evaluation)

- write the lecturer and ask for an appointment in November

Project milestones - part 2

In-progress meetings (*to help you, not an evaluation*):

- present your progresses so far to the supervising lecturer
- explain the difficulties encountered so far, if any
- describe the open issues and discuss how you plan to address them
- date: November 19, November 21, and December 3

Project report write-up *(written report)*:

- summarize and categorize the relevant sources, elaborating on their importance
- summarize the main experiments and results
- be careful when formatting the bibliography
- 5 pages (excluding references)
- deadline December 15 (send the report via email to massimiliano.mancini@unitn.it and giulia.boato@unitn.it)

Project milestones - part 3

Final presentation and demonstration *(in class)*:

- introduce why your topic is relevant
- first, give the theoretical background
- then, give an overview and a small demonstration of your work
- ideally, conclude with an outlook
- max 20 minutes
- dates: December 17 and 19

Elements for grading

- Related work and project status *(presentation in class)* **40 points**
- Project report write-up (written report) **20 points**
- Final project discussion *(presentation in class)* **40 points**

NOTES:

- When preparing the presentations, keep in mind that the audience is composed of lecturers and fellow students. The talks should be structured in such a way that everyone is able to follow.

- **All group members must contribute to the project work and to the presentations.**

Grading elements

Related work and project status *(presentation in class)*

- Was the problem well motivated/introduced? **4 points**
- Was the literature search extensive? **4 points**
- Are the related works well-organized? **4 points**
- Is the choice of related works clear? **4 points**
- Are the main related techniques well explained? **3 points**
- Were the questions answered well? **3 points**

- Was the presentation clear, even for non-experts? **3 points**
- Was the quality of the slides high? **3 points**
- Was there a clear focus for each presented part? **3 points**
- Was there a clear flow/connection among slides? **3 points**
- Is there an equal split of effort when presenting? **3 points**
- Did the presentation respect the given time? **3 points**

Grading elements

Project report write-up (written report)

- Was the problem well motivated/introduced?	**2 points**
- Is the literature search extensive and well-organized?	**2 points**
- Were the performed experiments deep?	**2 points**
- Were the performed analyses well described?	**2 points**
- Are the conclusions sound and well-supported?	**2 points**
- Is the writing correct?	**2 points**
- Is the text clear and easy to follow?	**2 points**
- Is the bibliography formatted correctly?	**2 points**
- Is there a clear flow between sections?	**2 points**
- Are the images/tables well-formatted?	**2 points**

Grading elements

Final project discussion (presentation in class)

- Was the goal of the project well motivated/introduced?	**3 points**
- Are the replication efforts correct (initial experiments)?	**3 points**
- Were the performed analyses deep and critical?	**4 points**
- Did the team include an innovation element beyond replication?	**3 points**
- Are the conclusions sound and well-supported?	**3 points**
- Are limitations/future directions discussed?	**3 points**
- Were the questions answered well?	**3 points**
- Was the presentation clear, even for non-experts?	**3 points**
- Was the quality of the slides high?	**3 points**
- Was there a clear focus for each presented part?	**3 points**
- Was there a clear flow/connection among slides?	**3 points**
- Is there an equal split of effort when presenting?	**3 points**
- Did the presentation respect the given time?	**3 points**

Projects list and description:

10 projects are available, groups of 3 components are expected:

- P1 - Voice Signature: Person-of-Interest Protection for High-Profile Voices
- P2 - A Novel HQ Dataset of fully generated images by LORA Models
- P3 - Improved Fake Video Detection with 3D-ViT
- P4 - Tampering Localization with DINOv2
- P5 - Dual-Domain Analysis of Adversarial Sensitivity for Forgery Localization
- P6 - (Generalized) few-shot learning
- P7 - Test-time adaptation
- P8 - Bias discovery
- P9 - Bias mitigation through compositionality
- P10 - Low-resource vision

Rules

Each group should send us via email:

- the components of the group
(**3 group members are expected**)

- the list of **ALL** projects ordered by preference (*from the most preferred to the least preferred*).

Deadline: October 30st (11.59 am)

Special cases:

If you want to propose your own project, you are **welcome**! But:

1. Send us your (even rough) plan (i.e., title and focus, repo you would start with, planned experiments/analyses) via e-mail **by October 28th (11.59 am)**. In such a way, we can discuss it in class, before going ahead with the assignments.

2. Still share your preferences
(we aim to keep a balanced group assignment across topics)

3. If the project is linked to another course (e.g., Advanced computer vision, Foundation models) share it with us:
   a. The details of the project for the other course
   b. What will be the shared components and the different ones
   c. **Note that we will perform checks: better that the overlap is shared beforehand.**

Feel free to ping us beforehand to refine the proposal!

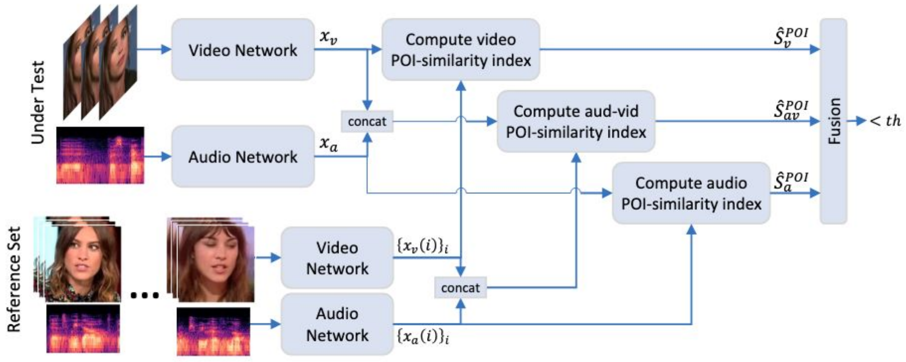
Trends in CV

Projects presentation for T1

Giulia Boato

Project #1 · Voice Signature: Person-of-Interest Protection for High-Profile Voices

- A **specialized voice detector** designed to safeguard the vocal identity of a specific individual, such as a CEO or politician.
- Move beyond the general problem of synthetic vs. real voice detection.
- Build a unique **model**, or "voice signature," for a target person using a dedicated dataset of their speech.
- The model **goal** is to verify whether any given voice message is genuine or an impostor
- **Key challenge:** system's performance with noisy data, typical of real-world voice messages, to ensure its practical effectiveness.



References:

1. Paper1
2. Paper2

Project #2 • A Novel HQ Dataset of fully generated images by LORA Models

- This project investigates the capabilities of SoA detectors to identify images generated by LoRA generators

- **GOAL**: Generate a dataset of Images created by LoRA Models
- **Follow bias-free generative methodologies** during generation
- **Test state-of-the-art detectors** that have never encountered these models during training
- **Report and analyze Pros and Cons** of each Methods, propose improvements.
- **Select a subset of the proposed LORA Models**. Prioritize the most realistic ones.

References:

1. Paper 1
2. Paper 2
3. Paper 3
4. Paper4

Lora Models:

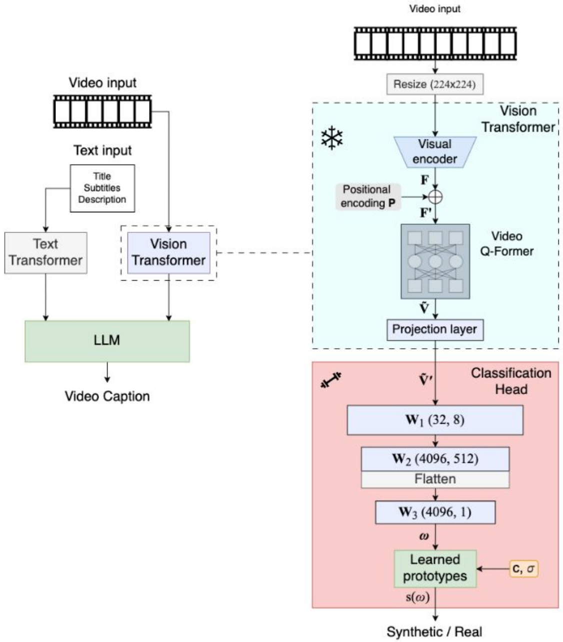
1. Model1
2. Model2
3. Model3
4. Model4
5. Model5
6. Model6
7. Model7
8. Model8
9. Model9
10. Model10
11. Model11
12. Model12
13. Model13

Project #3 · Improved Fake Video Detection with 3D-ViT

- **Evaluate and improve** the generalization capabilities of a **3D-ViT fake video detectors**
- Train on real videos and PyramidFlow
- Improve data augmentation
- **Evaluate on a large set of videos unseen during training** against the detector from the SoA based on DinoV2

References:

1. [Paper1]
2. [Paper2]

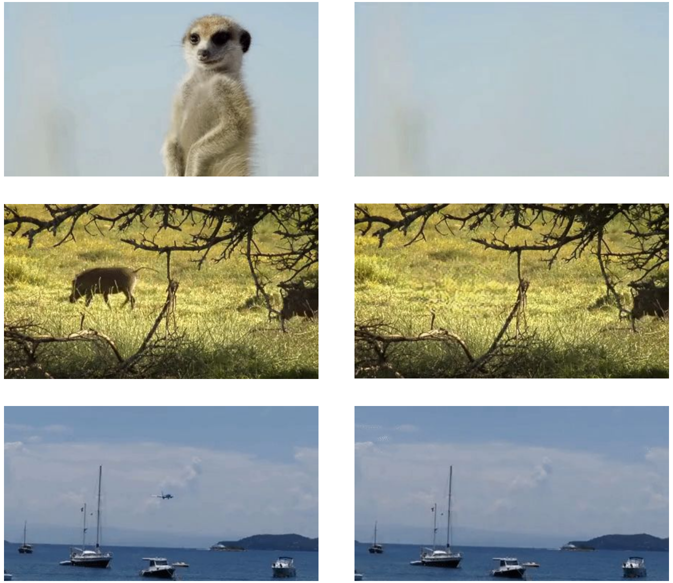


Project #4 · Tampering Localization with DINOv2

- Building on **DINOV2's** proven ability to produce high-quality semantic segmentations without explicit labels, **adapt it to effectively distinguish between authentic and inpainted regions** within video frames.

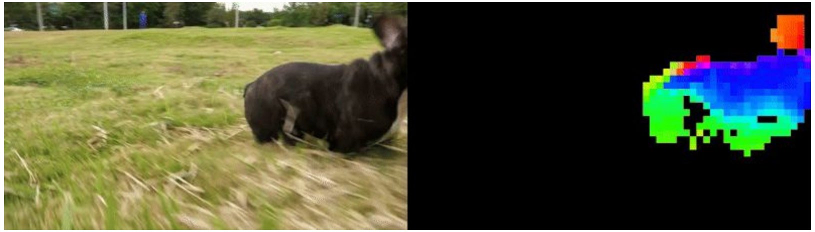
- Work on an available dataset of inpainted videos

- **Evaluate DINO's capability to detected inpainted objects and implement additional improvements.**



References:

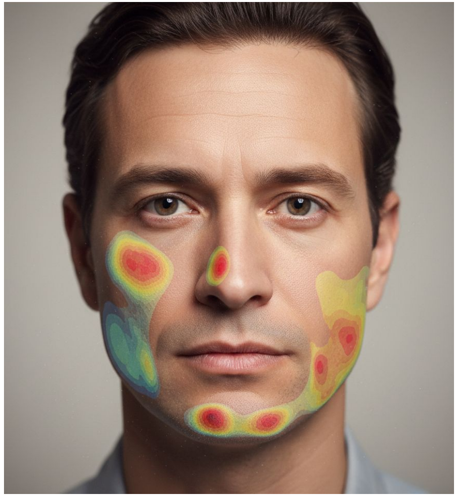
1. Paper1
2. Paper2



Project #5 · Dual-Domain Analysis of Adversarial Sensitivity for Forgery Localization

This project proposes a novel method for forgery localization by correlating it with model-specific adversarial vulnerabilities. Building upon De Rosa et al. (2024), which revealed unique spatial and frequency-based sensitivities in modern vision models, we will systematically generate adversarial perturbations (FGSM/PGD) to probe model weaknesses. The primary output will be **Adversarial Vulnerability Maps**, which quantify the model's output instability on a per-region basis. Our analysis will proceed along two parallel tracks:

1. **Spatial Correlation:** We will assess the spatial overlap between high-vulnerability zones and ground-truth manipulation masks.
2. **Frequency Correlation:** Using Fourier analysis, we will investigate if the spectral profile of a manipulated region is linked to the model's sensitivity in the frequency domain.
3. The goal is to determine if vulnerability itself can serve as a reliable signal for localizing tampered content.



References:

1. Paper1
2. Paper2

Trends in CV
Projects presentation for T2

Massimiliano Mancini

Project #6 · (Generalized) few-shot learning

Motivation: How do we adapt CLIP with few images? Does our choices impact the final results?

1. Report on the main techniques proposed in the literature to perform few shot learning for vision-language models, i.e., learning new concepts from few examples, focusing on parameter efficient fine-tuning. Useful pointers:
   - https://openaccess.thecvf.com/content/CVPR2025/papers/Farina_Rethinking_Few-Shot_Adaptation_of_Vision-Language_Models_in_Two_Stages_CVPR_2025_paper.pd
   - https://openaccess.thecvf.com/content/CVPR2024W/PV/papers/Zanella_Low-Rank_Few-Shot_Adaptation_of_Vision-Language_Models_CVPRW_2024_paper.pdf

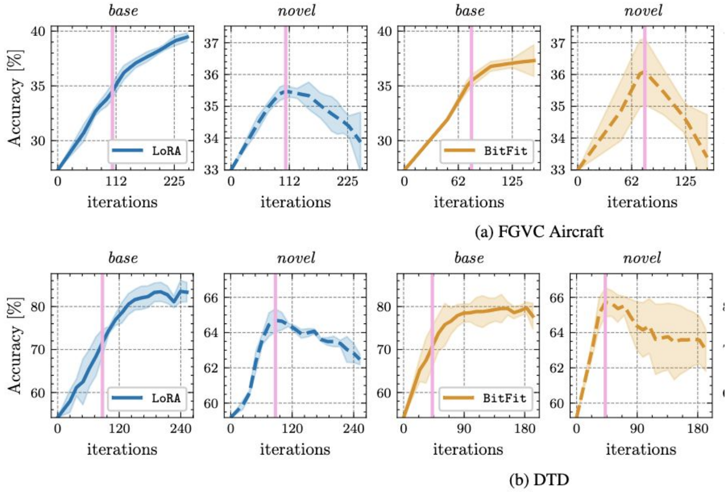
2. Get familiar with FSL for VLMs, using the above repos and techniques.

3. Benchmark them using standard datasets
   a. What are the failure cases?

4. How do its performance change w.r.t.:
   b. Hyperparameters?
   c. The distance of seen/unseen classes?
   d. Do different PEFT strategy show different behaviours?

5. Check how performance change w.r.t. the classes considered in the output space
   e. Is there any bias in the predictions?

6. Mitigate the above issues
   f. Can you apply existing approaches for generalized FSL to mitigate the above issues?



Farina, Matteo, et al., CVPR 2025.

Project #7 · Test-time adaptation

Motivation: Test-time adaptation is a topic of growing interest. Does the effectiveness of such techniques rely on crucial assumptions?

1. Report on the main techniques proposed in the literature to perform test-time adaptation (i.e., adapting as test data become available) for vision-language models. Useful pointers:
   - https://proceedings.neurips.cc/paper_files/paper/2022/file/5bf2b802e24106064dc547ae9283bb0c-Paper-Conference.pdf
   - https://github.com/FarinaMatteo/zero

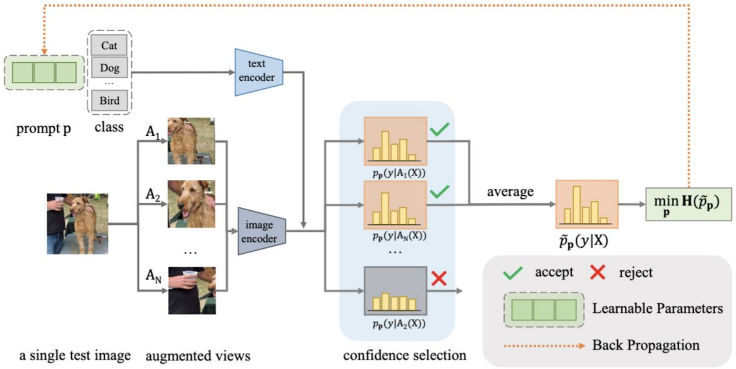
2. Get familiar with TTA for VLMs, using the above repos and techniques.

3. Benchmark them using standard datasets
   a. What are the failure cases?

4. How do its performance change w.r.t.:
   b. Hyperparameters?
   c. The tuning strategy: do different techniques show different behaviours?
   d. The way the network is updated (i.e., tuned for each sample independently or reset after each step)

5. Check how performance change w.r.t. the way the network is updated
   e. Is tuning independently for each sample better than continuous adaptation?
   f. In the latter case: does the data order and quantity matter?

6. Mitigate the above issues
   g. Can you apply existing approaches for robust test-time adaptation?



Shu, Manli, et al., NeurIPS 2022.

Project #8 · Bias discovery

Motivation: Multimodal models inherit the bias of their training set. At which scale can we discover them?

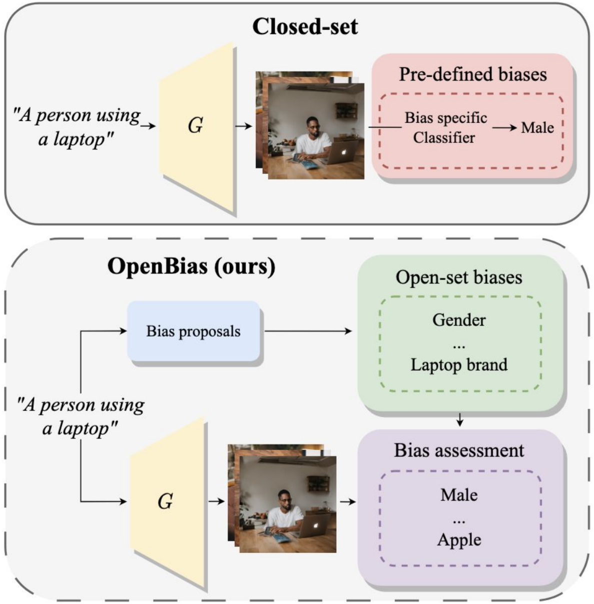
1. Report on the main techniques proposed in the literature to discover biases in generative vision language models. Useful pointers:
   - https://github.com/Picsart-AI-Research/OpenBias
   - https://openaccess.thecvf.com/content/CVPR2024/papers/Kim_Discovering_and_Mitigating_Visual_Biases_through_Keyword_Explanation_CVPR_2024_paper.pdf

2. Get familiar with bias discovery for generative models, using the above repos and references.

3. Run the code to produce some of the reported results.
   a. Are there failure cases?

4. How do the discover biased change w.r.t.:
   b. Number of generated images
   c. LLM or multimodal LLM chosen
   d. Type of input captions
   e. Keyword extractor
   f. Generative model

5. Proof-of-concept:
   g. Can you apply the same techniques to uncover bias in MLLMs?
   h. Do biases on the MLLM impact the results of the previous pipeline?



D'Incà Moreno, et al., CVPR 2024.

Project #9 · Bias mitigation through compositionality

Motivation: We know that vision-language models show biases. Can we address them using compositional properties of the embedding space?

1. Report on the main techniques proposed in the literature to mitigate biases in vision-language models. Useful pointers:
   - https://github.com/chingyaoc/debias_vl

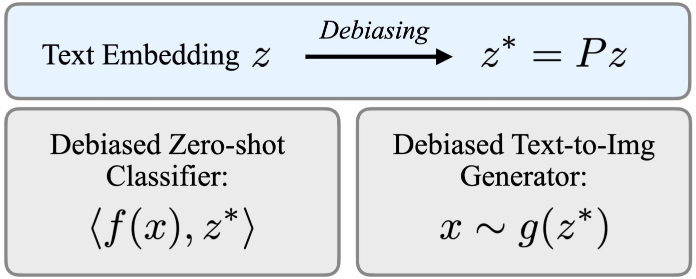
2. Get familiar with debiasing for VLMs, using the above repos and techniques.

3. Benchmark them using standard datasets
   a. What are the failure cases?

4. How do its performance change w.r.t.:
   b. Hyperparameters?
   c. The training loss
   d. The initial robustness of the model it is applied to?
   e. The type of biases

5. Check how performance change w.r.t. the number of biases we want to mitigate
   f. Is one projection enough?
   g. How do projections generalize across biases?
   h. Do projections generalize across application?

6. Mitigate the above issues
   i. Try ways of merging different projections for multi-concept debiasing.



Chuang, Ching-Yao, et al. arxiv 2023.

Project #10 · Low-resource vision

Motivation: Multimodal vision-language models work well on plenty of benchmark. However, specialized domains may need specialized solutions. In the following, we explore the question: can we perform open-vocabulary detection for X-ray images?

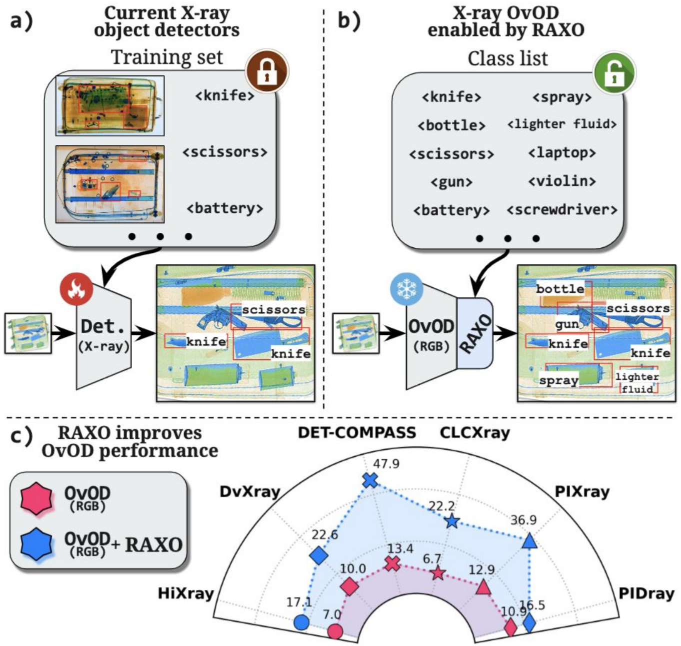
1. Report on the main techniques proposed for object detection in low-resource scenarios and X-ray imaging.
A potential starting point is:
    a. https://pagf188.github.io/RAXO/

2. Get familiar with debiasing for VLMs, using the above repo and referenced techniques there in.

3. Benchmark them using available datasets
    b. What are the failure cases?

4. How do its performance change w.r.t.:
    c. Hyperparameters?
    d. Backbones
    e. Retrieval set
    f. Clustering strategy

5. Investigate the model's assumptions
    g. Is there any constraint on how retrieved data should match target ones (e.g., semantically, style-wise…)
    h. Does the distance between training and test classes impact the performance of the model?
    i. Can you change the source of data to improve the results?



Garcia-Fernandez, Pablo, et al., ICCV 2025.