

PROJECT TITLE: IDENTIFYING NATURE REPRESENTATION ON SOCIAL MEDIA

EXTENDED ABSTRACT:

🧠🌿 **Think: How much time do you spend in front of a screen vs. in nature?**

Real-world human–nature connections are declining rapidly, with potentially severe consequences for how societies value and protect nature. At the same time, new — and widely overlooked — forms of nature connection are emerging in virtual spaces.

The **BIG-5** project [big-5.eu] investigates how digital platforms shape nature experiences and values, and how digital encounters translate into real-world environmental care and action.

A core challenge for the project is to automate the detection of nature values across large volumes of heterogeneous social media content. This master’s project aims to design a pipeline based on Foundation AI models capable of identifying nature-related content in images efficiently and at scale.

The goals pursues three main goals:

- **Goal 1.** Evaluate state-of-the-art VQA (Visual Question Answering) models to identify which model best answers the question *“Is there any nature content in this image?”* or similar formulations.
- **Goal 2.** Fine-tune the best-performing model so it aligns with the broader and more flexible digital concept of *“nature”* defined by the **BIG-5** project.
- **Goal 3.** Design an optimal processing pipeline, informed by the results of the previous goals, to automatically identify nature-related content across large collections of social media images.

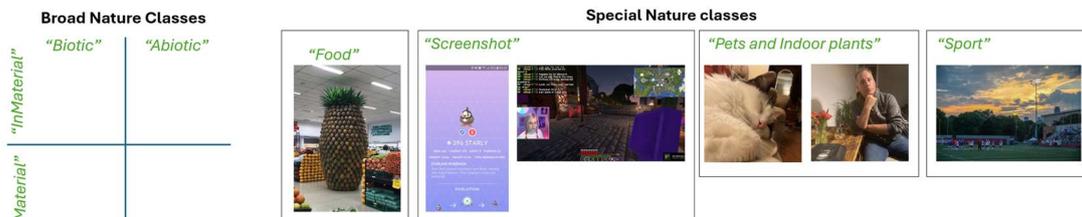
To achieve these goals, we propose the following milestones:

1st Milestone. The student will participate in a one-day workshop at the **BIG-5 Lab** to refine and operationalize the project’s working definition of *“nature”* and to define the full range of nature representations to be classified. This session will help clarify multiple categories of nature-related content, such as,

A Nature value content: *“bird”*
different representations



and different taxonomies:



The outcome of this workshop should be a **draft of a thematic prompt collection** able to cover a wide range of nature-related concepts.

BIG-5 will provide the following data:

- **A hierarchical set of nature-based annotations** for the ImageNet and MSCOCO datasets.
- **An annotated dataset of 8200 images**, with basic annotations such as: (a) Nature: Yes/No; (b) material/immaterial; and (c) biotic/abiotic (2 subject annotations per image).
- **A dataset of 2 million unannotated images** collected from various social media platforms, including X (Twitter), Weibo and Instagram

2nd. Milestone. Design the Experimental Setup to find the best-performing model for detecting nature-related content. The suggested methodology is the Open-ended Benchmarking of Vision-Language Models (VLMs) proposed in [1] adapted to the specific question “*Is there nature content in the image*” as well as the explainability requirements of the project. In this step of the project the nature-related prompt collection must be refined to be properly used in the experiments.

Suggested datasets and models:

Datasets: MSCOCO [2], ImageNet [3], Places [4] and **BIG-5** datasets.

Models: a selection of VLMs included in the original benchmark, plus additional models, with attention to ethical considerations:

- Models fine-tuned on VQA2: BLIP, X2-VLM.
- Multipurpose models evaluated in a zero-shot manner BLIP-2
- Conversational models, such as: InstructBLIP, LLaVA, MISTRAL.

To run the experiments, there may be opportunities to use *MareNostrum Computing Resources*.

3rd Milestone. Identify **the most suitable Foundation Model** for generic nature identification and explanation, while simultaneously defining **a strategy for fine-tuning this model to enhance its performance**. The evaluation will be based on the BIG-5 extended definition of “nature,” which emphasizes adapting the concept of nature to digital contexts.

4th Milestone. Design a complete processing pipeline that specifies all necessary considerations—models, datasets, prompting strategies, and evaluation criteria—to achieve high-performance large-scale nature identification across both vast numbers of images and a wide range of visual variations.

Bibliography

[1] Open-Ended VQA Benchmarking of Vision Language Models by exploiting classification datasets and their semantic hierarchy. Simon Ging, María A. Bravo, and Thomas Brox. In ICLR 2024 (<https://github.com/lmb-freiburg/ovqa>)

[2] Microsoft COCO: Common Objects in Context. Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, C. Lawrence Zitnick. In ECCV 2014. (<https://arxiv.org/abs/1405.0312>)

[3] ImageNet: A Large-Scale Hierarchical Image Database. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, Li Fei-Fei. In CVPR 2009. (https://image-net.org/static_files/papers/imagenet_cvpr09.pdf)

[4] A 10 million Image Database for Scene Recognition. B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba. IEEE Trans. On PAMI, 2017. (<http://places2.csail.mit.edu/>)