Mem. S.A.It. Vol. 96, 33

Memorie della

Multiplying Research Outcomes through the Application of AI Tools to Cultural Heritage

A. Guidazzoli[1], D. Sforzini[1], M.C. Liguori[1], R. Pansini[1], S. Caraceni[1], G. Pedrazzi[1], and G. Fatigati[1]

HPC Department - Cineca, Via Magnanelli, 6/3, 40033 Casalecchio di Reno, Bologna, Italy e-mail: a.guidazzoli@cineca.it, e-mail: d.sforzini@cineca.it, e-mail: m.liguori@cineca.it

Received: 21-10-2024; Accepted: 14-01-25

Abstract. This paper explores the application of Artificial Intelligence (AI) tools to cultural heritage research within the framework of the Italian National Recovery and Resilience Plan. Researchers at Cineca investigated open-source AI tools suitable for High Performance Computing (HPC) environments, focusing on object detection, multimodal analysis, Named Entity Recognition (NER), and other areas.

Key words. AI tools, Object Detection, High Performance Computing, Named Entity Recognition, Multimodal analysis

1. Introduction

Artificial intelligence (AI) is being actively experimented within Digital Humanities on diverse cultural heritage resources, including printed documents, manuscripts, and artifacts. Mapping advanced AI technologies in open science is essential for making National Heritage accessible to researchers and the general public.

Within the framework of the National Recovery and Resilience Plan investment M1C3 1.1 “Strategic and digital platforms for cultural heritage”, the Ministry of Culture has involved Cineca[1] in the design and implementation of I.PaC, the Infrastructure and Digital Services for Cultural Heritage. Cineca’s research foresees, among other tasks, surveying technical and scientific literature to identify effective AI applications in digital humanities field. Evaluating the identified tools involves assessing model training accuracy, adaptability to cultural contexts, and usability within High Performance Computing (HPC) environments, since Cineca’s supercomputing infrastructure supports the neural network training. The research focused on mining open-source repositories, such as GitHub and Hugging Face[2], which offer pre-trained models and community support. Continuous technology scouting ensures that tools remain up to date and suitable.

The scouting activities considered the following areas:

[1] CINECA website

[2] Hugging Face website

Method	Tool	Description
Object detection	Detectron2	Based on annotated examples, it allows the identification of the same objects in new images. The limitation is the need for many already annotated examples, which are not easily available in the artistic field. From the SGoaB project, we obtained an annotated dataset for experimentation.
Object detection - Yolo (You Only Look Once)	YOLO	Acronym for You Only Look Once is an alternative to Detectron2 has been the state of the art for Object Detection for years now and enjoys a large community that constantly updates it. Yolo v5 runs in MultiGPU and Multinode mode and it generates a variety of visual and textual report statistics useful for the researcher.
Multi Modal Analysis (Zero Shot Learning)	CLIP	Through CLIP, in the version developed by OpenAI, it is possible, given a vector of discretionary classes, to assign a probability that the class is present in the image. This is a foundational model trained on 400 million images, each associated with its own description. The limitation of the model is that it does not assign a probability to a single detected object but considers the whole image.
Multi Modal Analysis (Zero Shot Object Detection)	OWL-ViT	OWL-ViT is an extension of CLIP that allows for the identification of the bounding box for each detected object and its corresponding class. It overcomes the limitations of previous systems by allowing the detection of the same object multiple times within the image and assigning a probability of belonging to any class that is included in the query vector.
Multi Modal Analysis (Zero Shot Object Detection)	RegionCLIP ViLD	Other methods tested as alternatives to OWL-ViT but present some technical issues.
Manual annotation of images	labelImg	It is a tool for manual annotation that allows creating annotations for an image or editing those created automatically.

Fig. 1: List of tools studied over the years for object detection (Detectron2, YOLO), multimodal analysis (Clip, Owl-Vit, RegionCLIP, Vild) and manual image annotation (LabelImg)

Object Detection, NER (Named Entity Recognition), OCR (Optical Character Recognition), HCR (Handwritten Character Recognition), Speech to Text, Text to Speech, Geocoding, Language Translation and AI Image Enhancement. For each area of analysis, a state of the art was drawn up with the following aims: identifying state-of-the-art

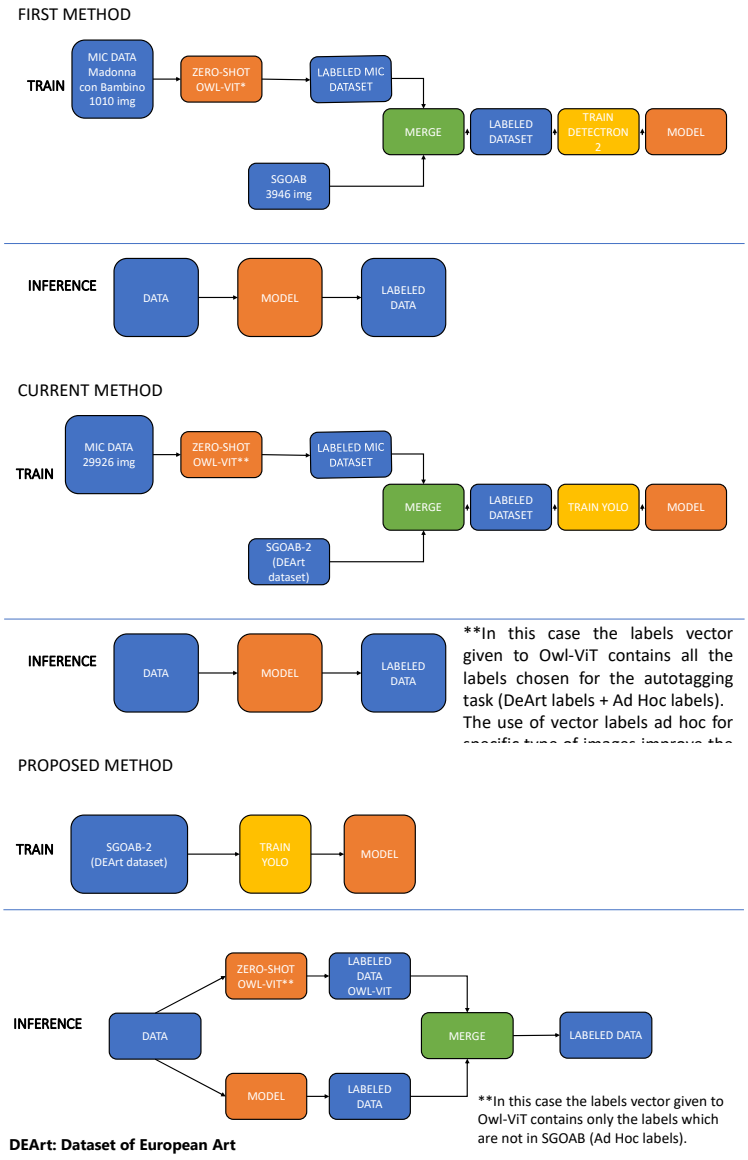

Fig. 2: Workflow

technologies; assessing the immediate operability of the tools identified in the literature; assessing the adaptability of other tools not yet tested in the Digital Humanities field.

In particular, in order to find the most suitable technologies for the different applications, the models were valued considering having a fully pre-trained and ready-to-use model; the possibility of using these models in an HPC environment; the possibility of using GPUs.

In addition, priority was given to finding models that could be used in the Python environmentn rather than exclusively on APIs belonging to the models, so that a more flexible pipeline could be created.

The present paper focuses mainly on object detection, multimodal analysis and NER (Named Entity Recognition). Other resources about the scouting activities are available over other areas of analysis[3].

2. Object Detection and Multimodal Analysis

Object Detection is a computer technology related to computer vision and image processing. Its purpose is to detect instances of semantic objects of a certain class (such as humans, buildings or cars) in digital images and videos. The tool must be able to find one or more objects within the examined image and delineate the rectangle that marks the boundary within which the detected object is located. Multimodal Analysis, besides understanding the content of an image, allows the analysis and correlation of information from multiple input modalities (typically text and images) with artificial intelligence models, such as CLIP and OWL-ViT. The Object Detection and Multimodal Analysis scouting activity for creating new workflows for Cultural Heritage began at Cineca in 2023 using Detectron2, CLIP and OWL-ViT on the SGoaB Datasets.

Fig. 1 lists the tools investigated for these tasks. Fig. 2, instead, presents different workflows developed over the years. The training process of AI algorithms needs datasets with millions of annotated images. The lack of such datasets for Cultural Heritage was compensated by adding to the SGoaB project dataset (manually annotated) a "domain" dataset on which we performed automatic annotations through zero-shot tools. The dataset resulting from the union of the previous two was used to train the model. The current workflow and the first workflow are conceptually similar, but they differ in the size of the datasets used (the current ones are larger), the vector of terms used for zero-shot type auto-tagging and the model chosen for the training phase. We are already considering updating the current pipeline. We plan to use only the DEArt dataset for model train and do inference on all MIC data by compensating for the lack of *ad hoc* labels with zero-shot models directly during the inference step.

3. NER

In the field of Cultural Heritage, Named Entity Recognition (NER) is crucial for digitalization and cataloguing. NER, a field of natural language processing (NLP), identifies and classifies named entities in text into predefined categories like names, organizations, and locations. The task is challenging due to the variability and ambiguity of natural language, with entities appearing in different forms.

We implemented and evaluated three different pipelines for the NER task based on different tools.

The first one uses WikiNEuRal[4] and SpaCy(It_core_news_lg)[5] for entity extraction (People, Places, and Organizations) and the "Soggettario di Firenze"[6] Directory for unrecognized entity classification (MISC category). WikiNEuRal is a Bert-based multilingual model, while It_core_news_lg is a model

[3] Video recording of the third Workshop AI, Cultural Heritage and Art - Between Research and Creativity

[4] WikiNEuRal github

[5] Spacy model

[6] "Nuovo soggettario", edited by the National Central Library of Florence (Biblioteca nazionale centrale di Firenze), is a subject indexing tool for various types of information resources. [...] This tool has been created for general and specialized Italian libraries, especially those participating in the National Library Service (SBN), and for museums, multimedia libraries, archives and documentation centres". Nuovo soggettario (sbn.it)

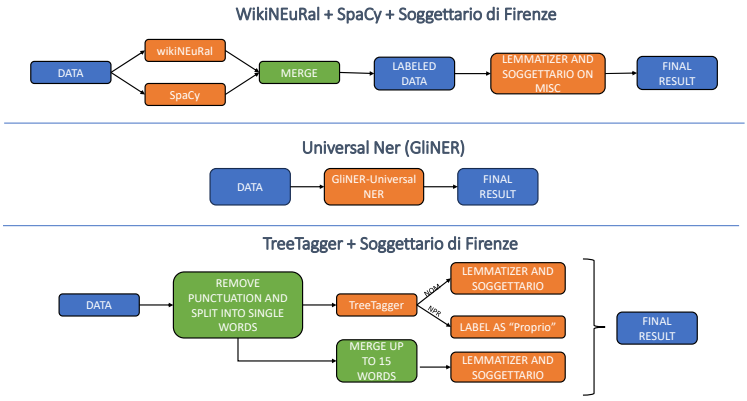

Fig. 3: NER Workflow

of the NLP framework SpaCy, which can perform various tasks, including NER, with the same classes as WikiNEuRal. The second is based on Universal NER, a zero-shot tool that allows finding entities within a text associated with labels passed to the model via a vector of terms which, in this case, would be composed of the categories of the "Soggettario di Firenze" Directory. The strength of this approach is the simplicity of the pipeline, but the results were not encouraging.

The last approach is to find all words, or sets of words, within the text that have a match in the "Soggettario di Firenze" Directory (with their respective category).

From the experiments we conducted, the WikiNEuRal + SpaCy + Soggettario di Firenze method appears to be the best to date. However, we do not exclude implementing an additional pipeline to perform tagging on two different levels. A first level of tagging will be performed with the WikiNEuRal + SpaCy + Soggettario di Firenze pipeline and assigned a "weight," i.e., a high degree of reliability. The second level of tagging (lower weight) will use TreeTagger +Soggettario di Firenze for the purpose of tagging all the remaining terms in the Soggettario di Firenze even if they are not actually entities.

4. Conclusions

The scouting process helped clarify our understanding of the importance of establishing work pipelines that integrate AI tools with control and validation mechanisms in a way that aligns with the expertise of domain specialists. Another important aspect is the integration of various tools, drawing on research from scholars in different disciplines to improve the effectiveness and accuracy of the AI tools' deployment. Additionally, we observed that the complexity and potential of these tools foster better collaboration and synergy between research institutes.

IPaC Infrastruttura e servizi digitali per il Patrimonio Culturale
MiC MINISTERO DELLA CULTURA