



CultMedia: Data Collection and Annotation



Lorenzo Baraldi, Rita Cucchiara
Almagelab Dipartimento di Ingegneria Enzo Ferrari

University of Modena e Reggio Emilia, Italy



UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

softtech-ict
Centro Interdipartimentale di Ricerca
Softtech: ICT per le Imprese

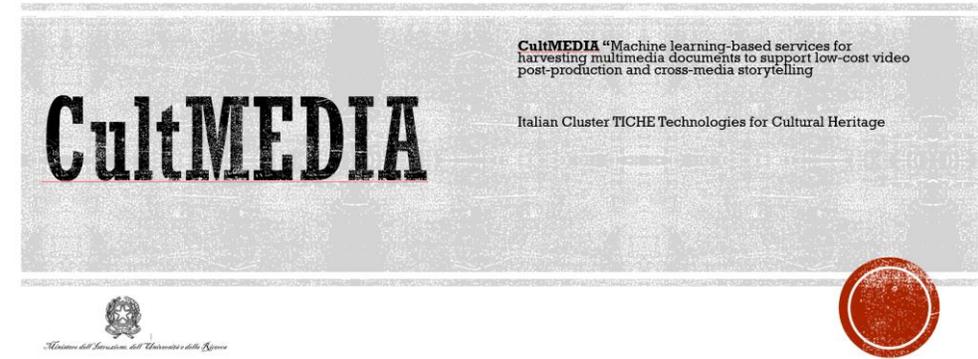
 **Almagest** Lab

The Goal

Creation of a (medium-to-)large-scale dataset oriented to the Cultural Heritage domain and suitable for automatic understanding tasks, such as:

- ✓ **Artwork identification and retrieval**
(I can detect, locate and identify an artwork)
- ✓ **Automatic Artwork description and retrieval with natural language queries**
(I can describe an artwork, and retrieve similar ones from other natural language descriptions)
- ✓ **Detection of attributes and relationships inside the artwork**
(I can identify the people/objects represented in the artwork, and the relationships between them)
- ✓ **Visual grounding of descriptions**
(I can use that knowledge to ground and justify the descriptions)

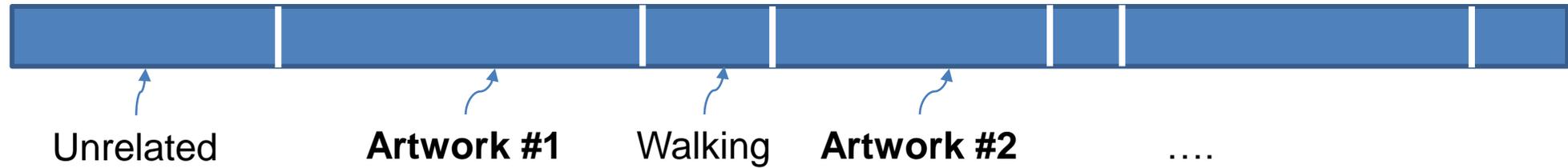
Strong link with the **re-use spirit** of the project.



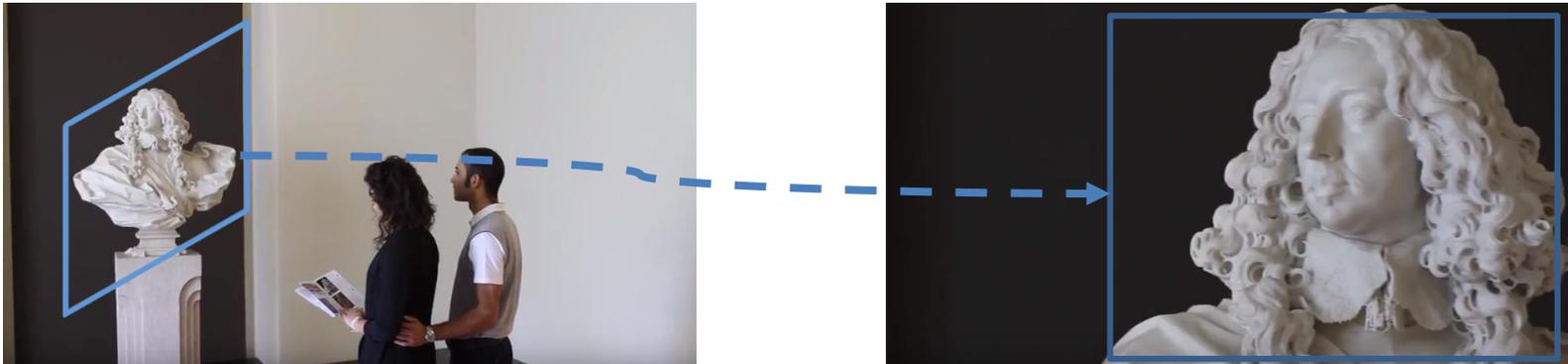
«A man with hat holding a glass of wine»
Soldato con Calice (N. Tournier)

Data annotation (#1)

- ✓ Temporal segmentation of the input video
to isolate the temporal extent of each artwork and unrelated temporal segments



- ✓ Artwork detection
i.e. annotate the bounding box of the artwork, frame by frame (exploiting the semi-automatic annotation given by the optical flow)



Data annotation (#2)

- ✓ Annotation with metadata

i.e. author, name of the artwork, year, style, ...

- ✓ Captions

a. describing the content of the artwork without leveraging any cultural background

b. describing the content and the context of the artwork by leveraging a specific cultural background



«A man with hat holding a glass of wine»

«A caravaggesque painting in which a soldier seems to establish a cultured dialogue with the spectator, descending, into the daily life of an inn, echoes of the classical tradition of the myth of Bacchus »

Data annotation (#3)

✓ Annotation of the details

detection and description of the components of the artwork (objects and people) with actions and attributes



Nerone (person), standing

Agrippina (person), laying

✓ Grounding of captions

i.e. connecting people, objects, attribute and actions in a natural language sentence

“**Nerone** is **standing** in front of a **Agrippina**, who **lays** on a bed.”

Annotation interface

- ✓ Re-use and extension of an existing annotation interface, born for video retrieval and annotation
- ✓ Online demo at: <https://www.neuralstory.it>
- ✓ ...with several additions:
 - object/artwork detection component
 - metadata collection component
 - captioning and grounding component

Thanks to the “Città Educante” project



Schedule

A first round of data collection and annotation will take place from the next February:

- 5th to 9th February
- 12th to 16th February
- 28th May to 1st June

With five to seven annotators in each round.

Each round will be scheduled as follows:

- ✓ 1st day: training on the interface, collection and validation of sample annotations on synthetic data
- ✓ 2nd day: visit to the Estense Gallery (Modena), and collection of the data
 - before and after the
- ✓ 3rd day → 5th day: annotation and cross-validation



Thank you!

lorenzo.baraldi@unimore.it

rita.cucchiara@unimore.it

<http://www.aimagelab.unimore.it>

<http://www.neuralstory.it>



Thanks for the Ongoing collaboration with **Facebook AI Research (FAIR)**



Thanks to **“CULTMEDIA ”** Project Machine Learning-based services for harvesting multimedia documents to support low-cost video post-production and cross-media storytelling , **Italian Cluster TICHE Technologies for Cultural Heritage (MIUR)**.