European Commission

ReInHerit

Redefining the Future of Cultural Heritage, through a disruptive model of sustainability

ReInHerit
Redefining the Future of Cultural Heritage

www.reinherit.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004545

Project

Project Number	101004545
Project Acronym	ReInHerit
Project Title	Redefining the future of cultural heritage, through a disruptive model of sustainability
Starting Date	01/03/2021
Duration in Months	36
Funding Scheme	Coordination and Support Action
Call (part) Identifier	H2020-SC6-TRANSFORMATIONS-2020
Topic	TRANSFORMATIONS-19-2020 *Culture beyond borders – Facilitating innovation and research cooperation between European museums and heritage sites*
Website	www.reinherit.eu

Deliverable

Work Package	WP3 - ReInHerit Toolkit
Task	T3.4
Deliverable	D3.3 - Toolkit Phase I
Dissemination Level	Public
Type of Deliverable	Report
Leader	UNIFI-MICC
Due Date	
Submission Date	
Keywords	Toolkit, Strategy, Digital Tools, Digital Innovation, AI/CV Interactive Tools, Gamification, CH Management

Version History

Version	Date	Author	Notes
V1.0	7/10/2022	Marco Bertini, Alberto Del Bimbo, Paolo Mazzanti, Leonardo Galteri, Filippo Principi, Andrea Ferracani	First version submitted to SC for peer-review
V1.1	20/10/2022	Marco Bertini, Alberto Del Bimbo, Paolo Mazzanti, Leonardo Galteri, Filippo Principi, Andrea Ferracani	Review after comments of peer-reviews
V2.0	09/2023	Marco Bertini, Paolo Mazzanti	Revision - based on the “General Project Review Consolidated Report” No. 2
V2.1	11/2023	Marco Bertini, Paolo Mazzanti, Andrea Oratiou	Revision based on internal review
V.3.0	09/2024	Paolo Mazzanti, Marco Bertini (UNIFI - MICC)	Revision - based on the “General Project Review Consolidated Report” No. 3

Acronyms and abbreviations

European Commission	**EC**
Research Executive Agency	**REA**
Grant Agreement	**GA**
Consortium Agreement	**CA**
Description of Action	**DoA**
Project Coordinator	**PC**
Steering Committee	**SC**
Project Management Team	**PMT**
Work Package	**WP**

Disclaimer

This document reflects only the author's view and the Research Executive Agency is not responsible for any use that may be made of the information it contains.

Contents:

1. Introduction 8
1.1 Scope and tasks 8
1.2 Problem Statement 9
1.3 Objectives 10
2. Toolkit development and toolkit strategy 11
3. Toolkit applications - phase I 13
3.1 Basic technologies 15
3.2 Smart retrieval 16
3.2.1 Unimodal retrieval 17
3.2.2 Multimodal retrieval 20
3.3 Smart video restoration 24
3.4 Strike-a-pose and Face-fit 30
3.4.1 Strike-a-pose 31
3.4.2 Face-fit 36
3.4.3 Co-creation and Ethical use of AI tools 38
3.5 Webinars 45
Appendix 48
References 49

Executive Summary

This deliverable D3.3 “ReInHerit Toolkit Applications Phase (I)” contains a description of the basic technologies used in the development of the toolkit and a detailed description of the first set of digital tools developed in accordance with the results of the D3.1 “National Surveys on current state-of-the-art tool”s and the D3.4 “Consolidated Report on ICT in CH Management”. Above all, D3.3 describes and explains the tools and apps developed in accordance with the D3.2 “ReInHerit toolkit strategy” and based on the key outcomes highlighted and based on primary and secondary research (WP2). These applications, source code, descriptions, associated scientific papers when available, and associated webinars will be made available in the context of the Digital Hub-WP4 Toolkit. The description of the technologies and advanced functionalities of the applications, exploiting in particular AI and CV techniques, will be useful also in identifying and selecting key training topics and curriculum (aimed at professionals) for professional development webinars and co-creative workshops that meet the needs of all stakeholders. Finally, D3.3 provides valuable material to the Consortium to feed and shape the next WPs of ReInHerit project (Digital Hub-WP4, Travelling and Digital Exhibitions WP6 – WP5, and Dissemination, Exploitation and Communication activities-WP7).

1. Introduction

According to the overall goal of WP3 to develop innovative methods and tools for communication and collaboration between museums and cultural heritage sites, and based on the ReinHerit Strategy (D3.2), the goal of the ReInHerit Toolkit is to develop and deliver digital **apps and digital tools** "based on existing, commercially available or open source, core technologies and frameworks in the **fields of AI, IoT, webinars**, and mobile development for CH management" (DoA, PartA, p. 25).

As described in the Toolkit Strategy (D3.2, p. 8), an important and useful trend that has been followed in the development of the Toolkit is **user personalization**, considered as a factor that enables museums and CH sites to move from "talking to the visitor" to "talking with the visitor," turning a monologue into a dialogue, co-creating contents with users stimulating their participation and creativity, creating broader connections with new and younger audiences, solutions that help create collaborative narratives with users enable them to follow the "visitor journey theory," according to which the visitor experience begins before visitors enter the museum and continues after they leave.

In the following subsections of this introduction, the **Scope and Tasks (1.1)** of the ReInHerit Toolkit Application Phase (I) are defined and described. It is highlighted which main requirements of the GA and the Annexes are to be considered for the development of the Toolkit and which tasks derive directly from them.

Subsection 1.2 Problem Statement describes the general objectives to give an idea of the Toolkit development process in relation to the structure of the Digital Hub.

Subsection 1.3 Objectives is concerned with challenges and goals related to development and implementation of the Toolkit and in relation to the past and future WP3 tasks.

1.1 Scope and tasks

The **D3.3 "ReInHerit Toolkit Applications Phase (I)"** contains an initial overview and a description of the digital tools developed in accordance with T3.3 and the Strategy (D3.2) with the aim of publishing and **sharing the digital tools in the context of the Digital Hub (WP4)**. Also related to specific **training webinars** on CH management topics identified in the context of WP2 analysis (i.e technologies, AI/CV, conservation, preservation, etc.) and on what has been developed at this stage

Based on the detailed requirements of T3.4 (DoA, PartA, p. 25), related to D3.3 and to the next deliverables, the development process will result in a series of **mobile apps for the** ReInherit Digital Hub. CH and Museums Professionals will use them to 1) **interact** with the performance **environment** 2) as **intelligent guidance tools t**hat adapt to the actions and interests of a museum or site visitor, understanding both the context of the visit and what the visitor is looking at; 3) as **gamification and learning-based tools** through the use of

techniques such as style transfer and "deep fakes" applied to user-generated content (e.g. turn participants' photos into paintings or paint/sculpt visitors in museum exhibits); 4) as tools for **discovering relationships** and similarities between different objects in collections within the same museum/site and other collections 5) as tools for a participatory **storytelling** experience.

All of these applications will be part of the Toolkit, which will include tools and **open-source codes** useful for their implementation, with associated **skills** and **training** webinars needed for their use and development.

Overall, all the defined **Tasks** and the next **Deliverables** of WP3, related to the Toolkit D3.3 are as follows:

- T3.1 National Surveys on current state-of-the-art tools
- T3.2 Consolidated Report on ICT in CH Management
- **T3.3 ReInHerit Toolkit Strategy**
- **T3.4 Mobile-based Applications**
- T3.5: Story based game development
- T3.6: Training curriculum development

- D3.4 Consolidated Report on ICT Tools in CH Management
- D3.5 Mobile Applications
- D3.6 Story Based Game Report
- D3.7 Demonstrator Mobile Applications
- D3.8 ReInHerit Toolkit Phase (II)
- D3.9 Training Curriculum and Syllabi

1.2 Problem Statement

As already introduced in Strategy D3.3, the overall objective of “Redefining the future of cultural heritage, through a disruptive model of sustainability - ReInHerit” project is to create a model of **sustainable heritage management,** which will foster a digital dynamic European network of heritage stakeholders. This model is based on a digital cultural heritage ecosystem in which all the stakeholders (museums, heritage sites, policy makers, professionals and communities) will be provided with the **tools and resources** to communicate, experiment, innovate and disseminate European cultural heritage.

The ReInHerit Digital Hub was designed as the **interactive and dynamic space** to collect and share **resources** (webinars, tools and documentations) for cultural heritage professionals. The cooperation is taking place within an open-ended design space in the form of an online space that will sustain a digital ecosystem of cultural heritage stakeholders.

The Hub will **include the Toolki**t on the use of digital technologies in museums and cultural heritage sites. It will provide **guidelines**, **prototypes** for developing technology-assisted

immersive performances, digital exhibitions, and educational and smart tourism applications, and **training webinars**.

In accordance with structure of D3.1 and requirements reported in D4.1 and D4.2, the specific section dedicated to the toolkit and related webinars will include the following sections (Figure 1):

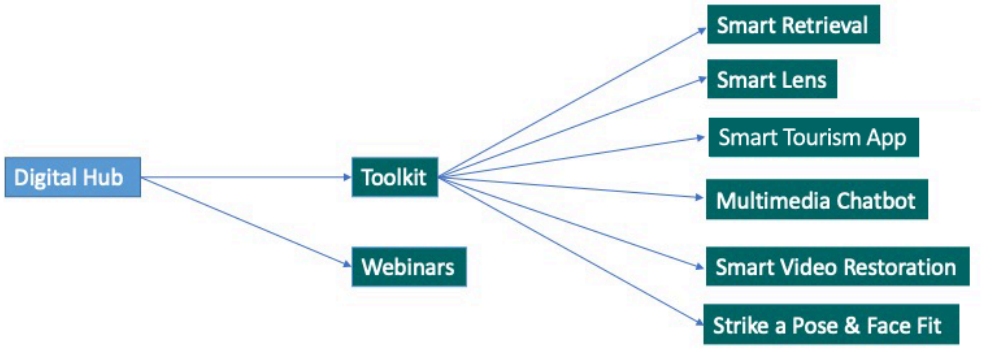


Fig .1: Toolkit tree section summary

The Smart Guide for Tourism app will be the base application for the Smart Tourism goals of the project, addressing in particular the use of digitalisation through computer vision and AI to enhance the tourism experience, through a more accessible and smart guide.

1.3 Objectives

As indicated in the DoA, Part (p. 14) the ReInHerit Toolkit, is a great opportunity for the partners to exhibit in real-time their **collaboration potential,** proving that the suggested model of sustainability is valid. The consortium researched current tools & methods of communication & collaboration between museums and cultural heritage sites, in order to identify a useful gap that the ReInHerit Toolkit can address (T3.1).

Based on the results of the analysis about the available tools available and used today (T3.2), the consortium proceeded to develop a set of innovative tools & practices that will disrupt the current status quo of museums-cultural heritage sites communication and collaboration (T3.4), but prior to this all the consortium members developed a relevant **Toolkit strategy** (T3.3). As a concluding activity, the partners will design **courses and syllabi for the main technologies** (T3.6) and practices identified in T3.1-T3.4, addressing in particular AI, IoT and mobile development. Development of this type of courses is fundamental to transfer technological know-how for capacity building in the cultural heritage sector, and help

reaching project sustainability beyond its conclusion. In T3.4 and T3.5 a multidisciplinary team of experts (archaeologists, museologists, historians etc will ensure its scientific/ historical accuracy and its added value in heritage management), webinars and workshops involving these related multidisciplinary issues will be organized.

The primary objective of this Deliverable is to provide a first detailed description of application development in accordance with the Theoretical Framework and Toolkit Strategy Concept highlighted in D3.2. The following sections will present the development strategy of apps based on AI/CV technologies that are needed to reach the identified targets, motivate them in educational activities and engage them in active participation.

2. Toolkit development and toolkit strategy

In deliverable **D3.2 "Toolkit Strategy"** has been laid out the strategy to be followed in the development of the toolkit, following the insights gained in WP2, through national surveys and focus groups phase (I) and phase (II). The technological state-of-the-art of ICT tools has been presented in deliverable **D3.4 "Consolidated Report on ICT Tools in CH Management",** where relevant open source and commercial solutions related to the tools considered in the questionnaires, as well as examples of applications and installations.

In this section we briefly summarize the **main outcomes** of these other deliverables that are related to the toolkit development presented in this deliverable.

- **Open source development** - national surveys and focus groups (D3.2 "Toolkit Strategy" sections 2 and 3) highlight that the lack of open-source solutions leads to maintenance issues and to the failure of reusing applications by different organizations. All the apps developed in the toolkit, starting from those presented in the next section follow an **open-source approach**; however, availability of source code alone is not sufficient to allow the adoption of such new technologies by **smaller organizations,** so the code of these apps will be complemented by additional documentation and associated **webinars** on the digital hub of ReInHerit, to ease the implementation of these apps also by small organizations with limited resources.

- **AI and Computer Vision (CV) tools** - these advanced technologies are being used in large museums and organizations, creating a technological gap with **smaller organizations** that do not have the **resources** to work with **experts** in these fields. The apps developed in the toolkit exploit AI and CV as founding bases of their use, with the goal of reducing this gap.

- **Interactive tools** - this type of tools is needed to increase the **engagement** of visitors in a **user-centered approach.** Within WP3 an activity is related to **story-based game** development, and in addition to this type of interaction in this deliverable we

present two apps that have the specific goal of increasing the **interaction** of the visitor with elements of the collection of a museum.

- **Mobile-first and web-first apps** - developing apps considering mobile devices as first-class targets allows to follow the **Bring-Your-Own-Device (BYOD)** approach more easily; developing such apps using web-based technologies allows to avoid requesting users to download native apps from some app store, a feature that national surveys have shown to be extremely relevant. These guidelines are followed for **user-oriented applications**, whenever technically possible, in the development of the Toolkit. We have reserved the possibility of dropping the use of web technologies when such a requirement impairs too much the performance of the app; this may happen, in particular, when the frameworks for AI and CV do not provide the full functionality in their web-based version.

- **Digital Transformation & Multidisciplinary approach** - it is important to create a mixed network, providing webinars, hackathons, workshops and trainings. Developing and Piloting the innovative ReInHerit Toolkit (Apps, webinars, training curriculum) as a **mixed innovation process** involving ICT developers and CH professionals to create usable tools, involving visitors in the co-creation and design process. The apps of the toolkit are designed to show possible uses of advanced ICT techniques and form the basis of further development, reducing the need of the users to concentrate technical aspects so to focus more, for example, on **design-thinking** methods, to empathize with the users and understand needs, define the problem, ideate solutions, and test the prototypes.

- **Appropriate use for generative artificial intelligence**
  In accordance with the strategy of the Toolkit (D3.2 section 5) , it is essential to develop generative artificial intelligence (GenAI) tools following a model that takes into account important critical aspects, such as the **scientific accuracy** of chatbot results and the **ethical implications** related to the use of **personal** and **training data.** In this regard, the solutions developed will seek to avoid errors and hallucinations by relying on quality content provided by experts, through the development of appropriate prompts capable of directing responses to validated datasets. sectors, including education. Particular attention will be given to the General Data Protection Regulation (GDPR) and the protection of the user's personal data. (D3.8/D3.4)

- **Considerations on ethical issues**
  The ethical annex in D3.2 contains for each application of the toolkit a series of information regarding training datasets, user data, data security and copyright. In general the applications do not store personal data and have been designed to avoid the leakage of personal information. When acquiring the appearance of the users, e.g. for gamification purposes, the information acquired can not be used to recognize a person, nor it is used to infer any personal information such as the emotional state of the player. The fairness of the models has been evaluated

according to the data provided in the model cards (see Sect. 3.4.3 in this deliverable and D3.2 ethics annex).

3. Toolkit applications - phase I

During the first phase of the development of the ReInHerit toolkit, **4 applications** have been developed, following the list presented in D3.2. These applications are:

- **Smart retrieval** - this is a **web application** that can be used to provide advanced search functions for multimedia archives. It provides **content-based image retrieval** (CBIR) facilities, i.e. search images based on their content. Search can be performed in two ways: **1)** use a **textual description** to search images; **2)** use a **combination of reference** image and associated textual description to search for images. The novelty of this application is due to the implementation of the computer vision component, i.e. the neural network used to associate text describing the desired content of the image and the pixels of the image. Within ReInHerit, a **novel approach** to perform conditioned image retrieval has been developed, i.e. the second type of search described above, that allows to **search an image using an image example** (a visual reference) and an **additional text**, expressed in natural language, that describes a modification w.r.t. the content of the reference image. The first type of image search is implemented using the same type of neural network, used in this case for uni-modal search: text-to-image retrieval (and also image-to-image retrieval). The system can be used also to perform tagging (i.e. image annotation), using a zero-shot learning approach, i.e. an approach where it is not necessary to train explicitly the network to recognize a specific content. This is extremely **beneficial for small museums** that may not have the large resources needed to collect training data required to train explicitly a neural network to recognize a visual concept.
  This application has won the **Best Demo Honorable Mention** award at the Computer Vision and Pattern Recognition (CVPR) 2022 conference, one of the most important conferences for Computer Vision. Three scientific papers related to the development of this app and its scientific aspects, including comparisons with previous state-of-the-art approaches, have been published: [Baldrati2022a], [Baldrati2022b] and [Baldrati2022c].

- **Smart video restoration** - this is a **web-based application** that allows to r**estore archive videos** that have been degraded. Analog videos of historical archives often contain severe visual degradation due to the deterioration of their supports (either magnetic tape or film) that require costly and slow manual interventions to recover the original content. Within ReInHerit we have developed a **novel neural network** that uses a multi-frame approach and is able to deal with degradations such as severe tape mistracking, which results in completely scrambled frames. The network can be used to reduce also other types of tape defects that are less severe than tape

mistracking. It can also be used on films to reduce scratches and traces of mold. The novelty of this application is in the neural network that reduces the artifacts due to such degradation processes.

As part of the ReinHerit Toolkit, this tool employs a neural network that uses a multi-frame approach to restore analog videos from historical archives in the arts and culture field in a timely and cost-effective way.

Development of this application has been carried-on in collaboration with Istituto Luce, one of the most important historical video archives in Europe.
A scientific paper related to this application has been published at ACM Multimedia, the foremost international conference on multimedia [Agnolucci-2022].

- **Strike-a-Pose and Face-fit** - These two applications are designed to employ **gamification and interaction** with an artwork to increase the **engagement** of the visitors of a museum. The apps are based on body pose and face expression recognition, respectively, using computer vision. Gamification is the process of exploiting strategies and game dynamics into use scenarios different from games. It has already been proven able to enhance skills and competences in a variety of domains from industry training to entertainment, and its application to cultural heritage is an opportunity to engage visitors to museums content. The goal is to help museums to move from the traditional "look and do not touch" toward a **"play and interact" approach**. In these apps the user is requested to replicate the pose or the facial expression of a painting, once the challenge is completed he gets information about it and new media generated by the apps (videos and images) that can be shared on social networks to help increase the engagement of the museum. Following the web-first/mobile-first approach the apps can be used on a mobile phone and have been implemented as web apps,but have been developed also to be used as an installation within the museum. The novelty is in the interaction design and in the **sharing of the media and user generated content,** through the functionality of the apps,on the social networks.

  As part of the ReinHerit Toolkit, the applications are designed for the cultural heritage domain and exploit gamification techniques in order to improve fruition and learning of museum artworks. A scientific paper related to this application has been published at **ACM Multimedia**, the foremost international conference on multimedia [Donadio-2022].

  In this section the applications are described, avoiding excessive technical details that are anyway available in the associated scientific publications.

3.1 Basic technologies

The development of all the apps described below has followed a **spiral model** (see Fig. 2), to manage the development **risks**, especially considering the technological risks imposed by the web-first/mobile-first goal, that imposes computational constraints and limits the selection of available frameworks for the most advanced AI and CV applications, w.r.t. desktop/server-based development.

This means that apps are developed in phases adding **new functionalitie**s to the basic functions developed in the previous phase, and considering the technological risks for their implementation.

Regarding the third-party frameworks and libraries used, **open-source projects** have been selected, considering their licenses so that they do not hinder future development of new apps by the end users of the toolkit.

Concurrent versioning system has been used to track development by all the software developers and Github will be used to maintain and distribute the code in the future, collecting all the resources in the Digital Hub of ReInHerit, to avoid duplicating the maintenance of a versioning system and the associated costs.

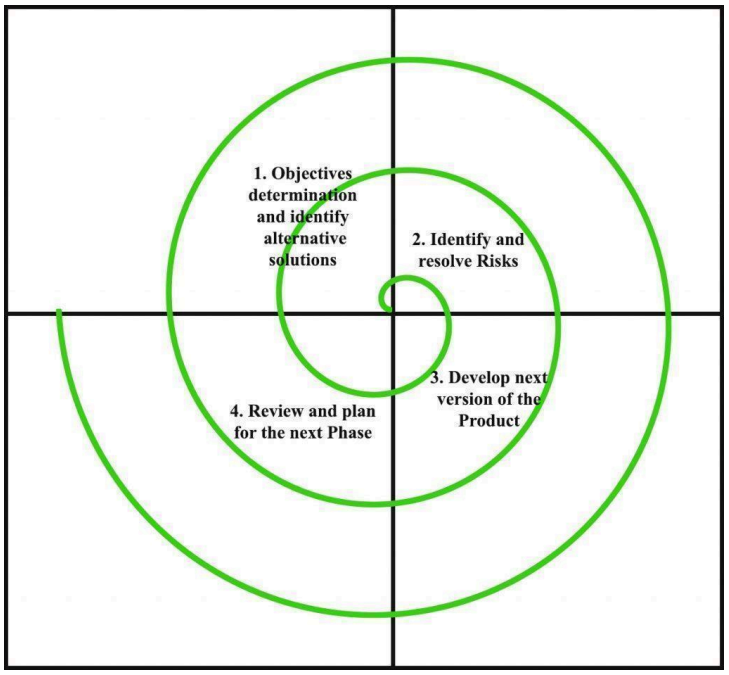


Fig. 2 - spiral model for software development. The Radius of the spiral at any point represents the expenses(cost) of the project so far, and the angular dimension represents the progress made so far in the current phase.

The **main libraries** that have been selected are:

- **PyTorch** to develop AI and CV components that run server-side. This is motivated by the fact that the scientific community is particularly invested in its use and it is typically faster to develop new networks.
- **Tensorflow** to develop AI and CV components that run client-side for web and mobile apps. This is motivated by the fact that this framework provides a TensorflowJS version for Javascript application (thus suitable for web apps) and a TFLite version suitable for mobile apps developed natively, e.g. to implement Android apps.
- **OpenCV** to implement basic computer vision functionalities. This library supports server-side and client-side development, addressing also web-based apps using Javascript bindings and native mobile apps for iOS and Android development.
- **Kivy** is used to implement cross-platform GUI apps, allowing the use of a single codebase to deploy apps on Windows, Linux, macOS, iOS and Android. This allows the implementation of both desktop apps and mobile apps.
- **JQuery** is used as a basic framework for web applications, considering its large popularity.
- **Flask** is used to implement REST calls for the server-side services of web apps.

In general app development is based on **Python** and **Javascript** language, adding **Java** for native **Android** applications.

3.2 Smart retrieval

Image Classification and Content-Based Image Retrieval (CBIR) are fundamental tasks for many domains, and have been thoroughly studied by the multimedia and computer vision communities. In the cultural heritage domain, these tasks allow to **simplify** the **management** of large collections of images, allowing to annotate, search and explore them more easily and with lower costs.

The app is a **server-side application** with a minimal client-side web app that shows how to use it. The server-side app must be integrated by the users.

In recent years the problem of CBIR has been addressed using **convolutional neural networks** (CNNs), that compute visual features representing the visual content of images and videos. Such features are then used to index and search multimedia archives. These networks are typically used in an unimodal fashion, i.e. only one media is used to train and use a network. This may limit the types of application that can be developed and may also reduce the performance of the networks. Several recent works are showing how using multi-modal approaches may improve the performance in several tasks related to visual information.

In [Radford2021] it has been shown that the CLIP neural network[1], a model trained using an image-caption objective alignment on a giant dataset made of 400 million (image, text) pairs,

[1] OpenAI CLIP: https://openai.com/blog/clip/

obtains impressive results on several downstream tasks. The authors pointed out that, using only textual supervision, CLIP model learns to perform a wide set of tasks during pre-training including OCR, geo-localization, action recognition and many others. This task learning can be leveraged via natural language prompting to enable zero-shot transfer to many existing dataset. We have thus selected CLIP as a backbone of our smart retrieval app, to solve two types of CBIR: unimodal search (text-to-image or image-to-image) and multimodal search (text+image to image).

3.2.1 Unimodal retrieval

For the first type of search we exploit the zero-shot capabilities of CLIP in the artworks domain. To evaluate the performance of our app we used the **NoisyArt dataset** [DelChiaro2019] which is originally designed to support research on **webly-supervised recognition of artworks** and **Zero-Shot Learning** (ZSL). Webly-supervised learning is interesting since it allows to greatly reduce annotation costs required to train deep neural networks, thus allowing cultural institutions to train and develop **deep learning methods** while keeping their budgets for the curation of their collections rather than the curation of training datasets. In Zero-Shot Learning approaches visual categories are acquired without any training samples, exploiting the alignment of semantic and visual information learned on some training dataset. ZSL in artwork recognition is a problem of instance recognition, unlike the other common ZSL problems that address class recognition. Zero-shot recognition is particularly appealing for cultural heritage and artwork recognition, although it is an extremely challenging problem, since it can be reasonably expected that museums have a set of curated descriptions paired with artworks in their collections.

To get a better idea of how CLIP behaves in the artworks domain we started with a classification task using a shallow classifier and CLIP as the backbone.

Subsequently, thanks to the descriptions of the artworks in the dataset, we performed experiments in the field of zero-shot classification where CLIP was able to demonstrate its abilities in this task.

To evaluate the performance of our system we performed experiments on the tasks of **artwork-to-artwork** and **description-to-artwork** retrieval obtaining very promising results and superior performance to a ResNet-50 pre-trained on ImageNet, i.e. the method that we developed has a much better performance than that obtainable with neural networks designed for unimodal approaches.

Fig. 3 shows a selection of images from the NoisyArt dataset, that comprises 3.120 different classes, 70.474 images for training and 1.355 images for test. This dataset has been designed to support research on webly-supervised recognition of artworks. It was also designed to support multi-modality learning and zero-shot learning, thanks to its multi-modal nature; making it a perfect fit to evaluate the performance of the proposed approaches used in Smart Retrieval, in that we'd like to have a good zero-shot performance since it allows to

deploy the system without need for further retraining, and allow us to evaluate the CBIR multi-modal capabilities of the ReInHerit toolkit,using both text and images. The dataset is very challenging because of its web-based sources, which result in very varied versions of the same artwork taken from different points of view, different cameras, different quality, etc. as shown in the following figure where several variants of the same artwork are shown. Using this specific dataset allow us to correctly and scientifically evaluate the performance of Smart Retrieval app comparing the proposed network w.r.t. competing state-of-the-art approaches. Another relevant benefit of using this dataset is that webly-supervision is an important feature in the cultural heritage applications, since annotated data that is necessary to train supervised models can be acutely scarce. Thus, the ability to exploit abundantly available imagery from web sources to acquire visual recognition models is a tremendous advantage, even considering the scarce supervision of such data [DelChiaro2019b].

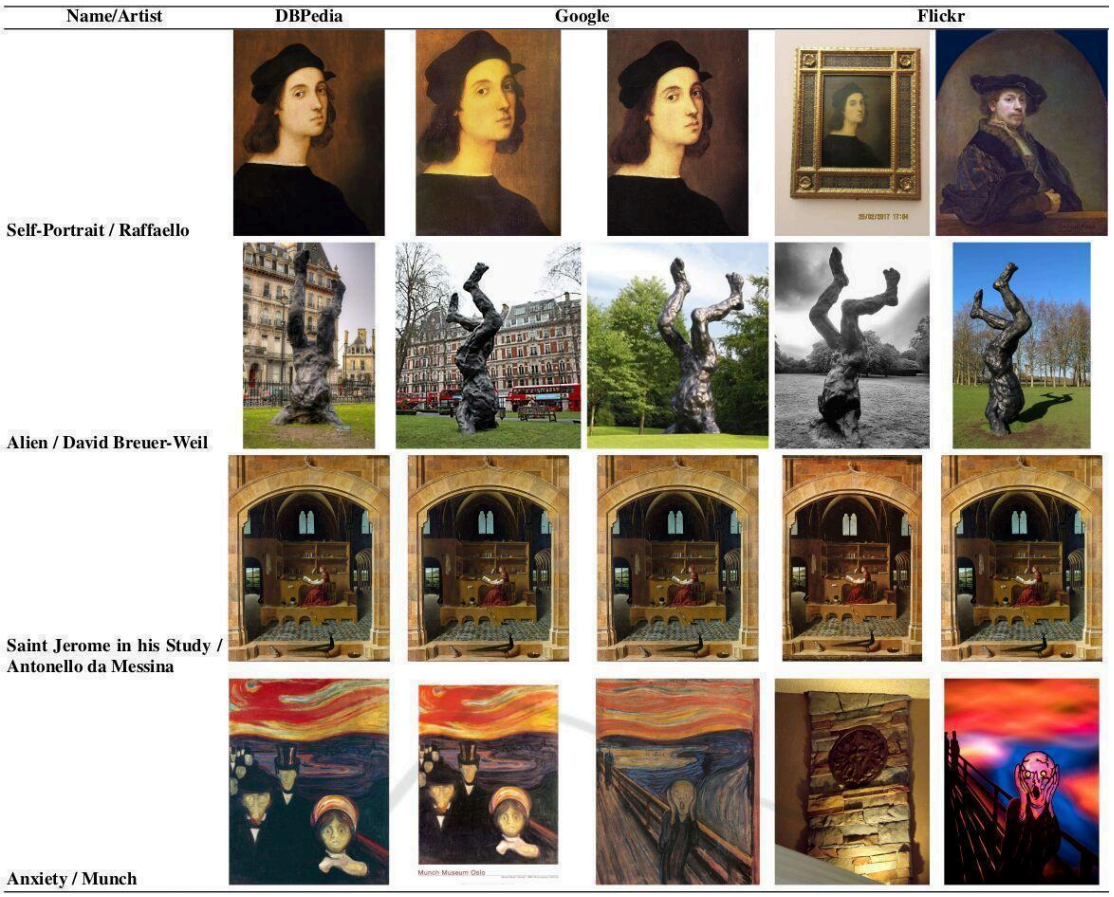


Fig. 3 - NoisyArt dataset - examples of classes and training images. For each artwork/artist pair we show the seed image obtained from DBpedia, the first two Google Image search results, and first two Flickr search.

To understand how CLIP associates text to the parts of the image that refers to an artwork we used a generalization of **gradCAM technique** [Selvaraju2019], that obtains an heat-map visualization showing us the portions of the image that CLIP most closely associated with the description.

Figure 4 shows four examples of such gradCAM visualization. We can see how, using the descriptions in the dataset, CLIP places attention to the portions of the image that are

associated with the artworks shown. This fact made us confident that CLIP would work very well in the domain of artwork.

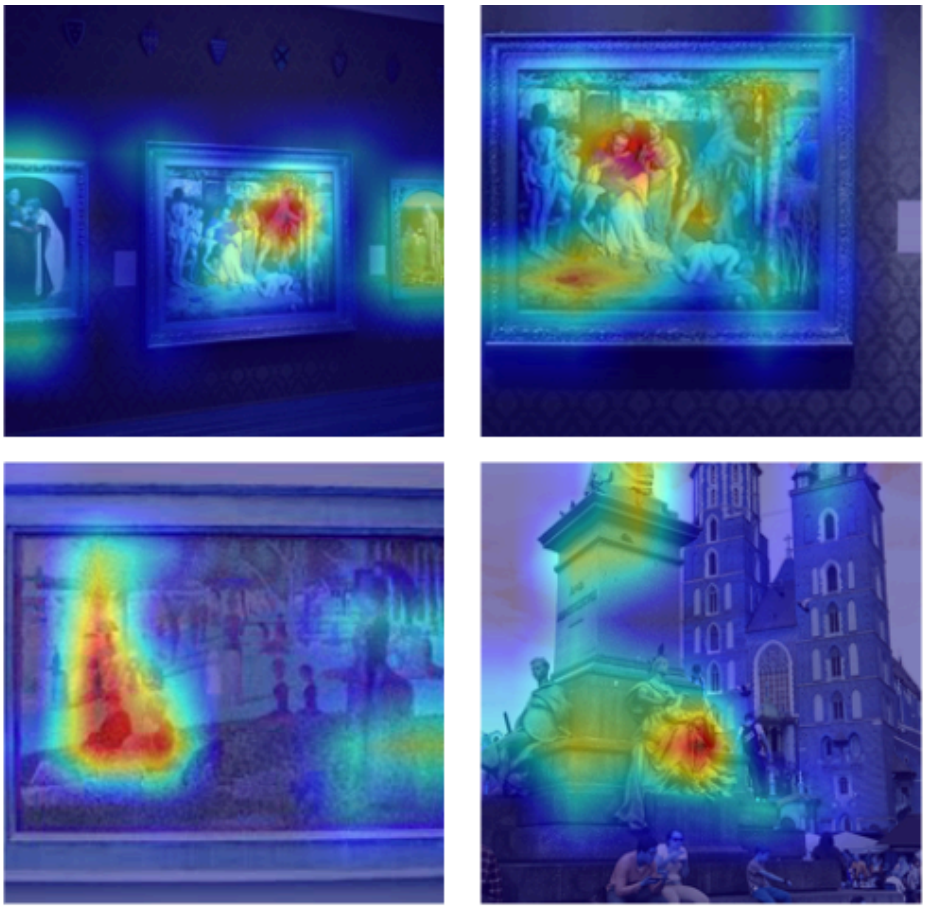


Fig. 4 - Examples of gradCAM visualization on NoisyArt showing which parts of the image received more attention by the CLIP network with respect to the description of the artwork.

To summarize the CBIR performance of the app we report in the following tables the comparison of several variations of use of the CLIP backbone, using text and images as keys for retrieval.

Text-to-image search: this task is akin to **Zero-shot classification**, we use textual queries to retrieve images, therefore we measure the performance of CLIP in terms of recognition accuracy and mean **Average Precision** (mAP). The following table shows the impressive potential of the method when compared to previous state-of-the-art approaches. Best result is highlighted in bold.

Retrieval system	Accuracy	Mean Average Precision
DEVISE RN50	24.79	31.90
EsZSL RN50	25.63	29.89
COS+NLL+ L2 RN50	34.39	45.53

Our (CLIP RN50)	**60.27**	**69.23**

Table 1 - text-to-image performance on NoisyArt dataset of the developed method compared to competing state-of-the-art methods.

Image-to-image search: in this task we compare a baseline using **ResNet** with variations of combinations of features extracted from CLIP. We also try to use CLIP in a zero-shot learning setup. Since the task is purely a retrieval one we report the performance only in terms of mAP.

Again we show that CLIP beats a sensible baseline, but we also show how to greatly improve the basic CLIP performance using the CLIP finetuning component of the developed app, that adapts the CLIP network to datasets in the cultural heritage domain. This component will allow organizations to implement their own CBIR system beating previous state-of-the-art methods.

Retrieval system	Mean Average Precision
RN50 image features	36.32
CLIP image features	46.40
CLIP class (ZSL) + text-to-image	40.54
CLIP class + text-to-image + reranking	47.41
CLIP image feature with our CLIP fine-tuning	**69.60**

Table 2 - image-to-image performance on NoisyArt dataset of the developed method compared to competing baseline and variations of the method.

The details of the system and a thorough comparison of its performance with competing approaches has been detailed in the [Baldrati2022c] paper.

3.2.2 Multimodal retrieval

Considering the impressive results obtained in unimodal search with CLIP, a second type of CBIR has been considered, **combining text and images to express more complex queries**, i.e. allowing users to represent complex visual aspects with image example and then refining their query with high-level expressions using natural language. This type of search is called, in the multimedia and computer vision community, **composed image retrieval**: the unimodal query is extended to an image-language pair. In a small variation, called **conditioned image retrieval,** the additional text may request constraints or add specifications on some attributes of the retrieved results. It must be noted that this type of search is much more

complex than standard CBIR, but is receiving more attention by the scientific community since it allows to extend the effectiveness of CBIR systems by adding some form of user feedback and because it has many possible applications in different domains.

The following figure shows the concept of combined/conditioned image retrieval.

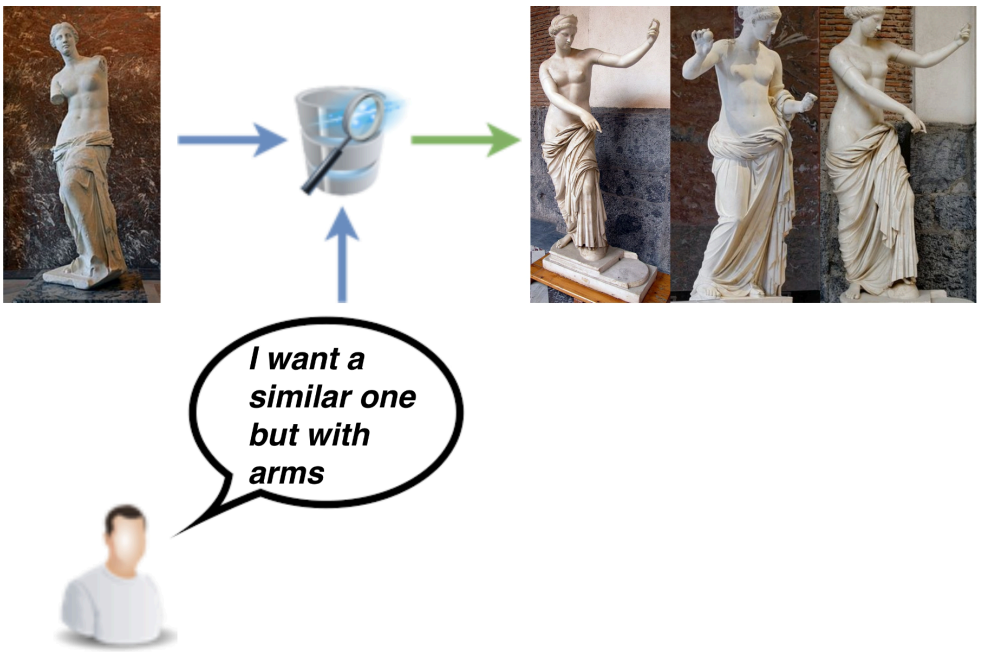


Fig. 5 - example of combined/conditioned retrieval. The text provides a context to the visual query, in this case requesting to change a visual aspect of the reference image.

Since the task is much more complex than the previous one we have developed a novel neural network that combines visual and textual features computed from a CLIP backbone. This network, called **combiner**, learns how to transform the visual features of the reference image using the textual features of the additional query so that they become more similar to the visual features of the objects in the database. In this way, at runtime, there's need to compute only the textual features, an operation that can be performed also on old portable PCs and run the computationally inexpensive combiner network to compute the most similar images. This approach thus allows to easily scale on large datasets, or in the case of small organizations, allow to use the system also with low power servers, or using free/low cost cloud providers. Given the technical complexity of the novel system this section does not include the technical details that are available in [Baldrato2022a] and [Baldrati2022b]. The following figure provides a gist of the architecture of the system; in order to obtain a further performance improvement we have reused the CLIP finetuning component that was developed to implement the unimodal search function presented in the previous subsection.

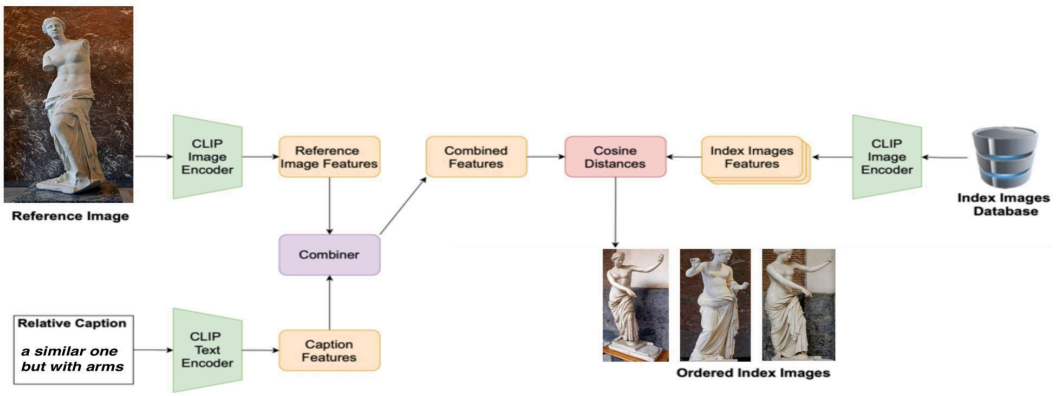


Fig. 6 - system architecture of the application used for conditioned image retrieval. The CLIP image and text encoder have been finetuned with the component developed for unimodal search.

The following figure shows an example of the **web app** used to test the **multimodal retrieval system.** The frontend is developed in Javascript and HTML5, the backend that provides the core of the application is implemented in Python, using PyTorch to implement the CV component and Flask to implement the REST services that let the interaction between the interface and the CV component.

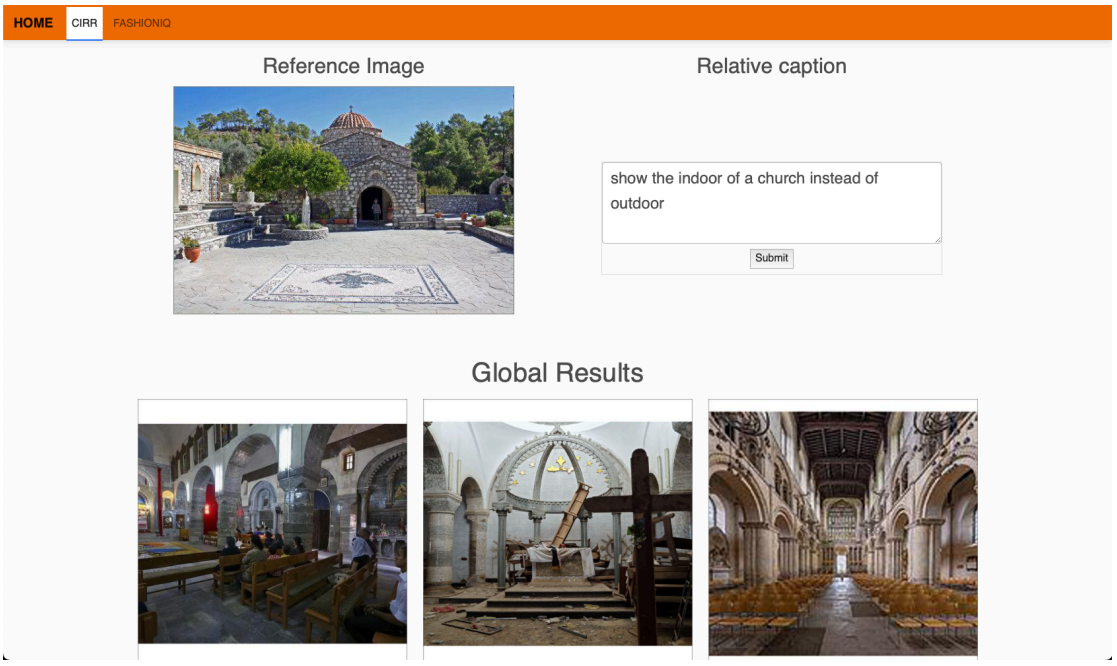


Fig. 7 - example of the web interface of the smart retrieval app for combined CBIR. The figure shows results obtained on the CIRR dataset considering the case of search of images regarding churches.

To **evaluate the performance of the system** with respect to competing approaches we have used two standard datasets employed for this task, CIRR, that contains a large variety of images, and FashionIQ that contains images related to fashion. The selection of these

datasets that are commonly used by the scientific community working in this field allow us to evaluate the performance of the system fairly with respect to a large number of competing approaches. The system has the current state-of-the-art performance for the task of conditioned/combined image retrieval.

Method	Recall@K				R_{subset}@K		
	$K = 1$	$K = 5$	$K = 10$	$K = 50$	$K = 1$	$K = 2$	$K = 3$
TIRG† [30]	14.61	48.37	64.08	90.03	22.67	44.97	65.14
TIRG+LastConv† [30]	11.04	35.68	51.27	83.29	23.82	45.65	64.55
MAAF† [45]	10.31	33.03	48.30	80.06	21.05	41.81	61.60
MAAF+BERT† [45]	10.12	33.10	48.01	80.57	22.04	42.41	62.14
MAAF–IT† [45]	9.90	32.86	48.83	80.27	21.17	42.04	60.91
MAAF–RP† [45]	10.22	33.32	48.68	81.84	21.41	42.17	61.60
ARTEMIS [59]	16.96	46.10	61.31	87.73	39.99	62.20	75.67
CIRPLANT † [19]	15.18	43.36	60.48	87.64	33.81	56.99	75.40
CIRPLANT w/OSCAR † [19]	19.55	52.55	68.39	92.38	39.20	63.03	79.49
Proposed approach (CLIP-RN50)	40.91	74.53	84.77	97.35	70.22	87.80	94.46
Proposed approach (CLIP-RN50x4)	**44.82**	**77.04**	**86.65**	**97.90**	**73.16**	**88.84**	**95.59**

Table 3 - performance of retrieval of our novel multimodal system compared to competing state-of-the-art approaches on CIRR dataset. The best result and the second best results are obtained by our system, considering a variation of the CLIP backbone. Retrieval performance is reported in terms of Recall @ K, i.e. we evaluate if the system is able to recall relevant results in the first K positions.

The application has received the **Best Demo Award Honorable Mention at the Computer Vision and Pattern Recognition (CVPR) 2022 conference,** the most prestigious conference on computer vision. The system is currently under further development with the goal of easing the creation of training dataset, and possibly avoiding even the necessity of creating such datasets, to ease the adoption of this type of multimodal search by small organizations that can't afford the creation of training datasets.

Fig. 8 - Best Demo Awards news on Reinherit's social channels (Facebook)

Scientific impact:
Smart retrieval has been demoed live at the IEEE Computer Vision and Pattern Recognition (CVPR) 2022, the foremost scientific conference on computer vision and AI applications, over several days. The system was used by the attendees of the conference.
Smart Retrieval obtained the BestDemo Honorable mention award by the scientific committee of the demo track.
Associated papers describing the apps:

- A. Baldrati, M. Bertini, T. Uricchio, A. Del Bimbo, "Effective conditioned and composed image retrieval combining clip-based features", Proc. of CVPR 2022
  The paper has received 111 citations, as of Aug. 2024.
- A. Baldrati, M. Bertini, T. Uricchio, A. Del Bimbo, "Conditioned and composed image retrieval combining and partially fine-tuning clip-based features", Proc. of CVPR 2022
  The paper has received 68 citations, as of Aug. 2024.
- A. Baldrati, M. Bertini, T. Uricchio, A. Del Bimbo, "Exploiting CLIP-based multi-modal approach for artwork classification and retrieval", Proc. of International Conference Florence Heri-Tech: The Future of Heritage Science and Technologies, 2022
  The paper has received 3 citations, as of Aug. 2024.

3.3 Smart video restoration

Historical videos constitute an important part of the cultural heritage of a society. The availability of this video content often is hampered by numerous artifacts and degradations due to technological limitations and aging of the recording support that limit its distribution and fruition by the general public. Normally the restoration of these videos is conducted frame by frame by experienced archivists with commercial solutions, thus at **great economic and time cost.**

For this reason, some works tried to restore historical video archives more rapidly and without human aid. For example, **DeOldify** [Deoldify] is an **open-source tool** for old films restoration addressing in particular colorization of black and white movies. The scientific community working on CV and multimedia has been active, in very recent years, to develop novel solutions for video restoration, e.g. [Iizuka-2019] relies fully on 3D convolutions and on source-reference attention for frame colorization. [Wan-2022] proposes a recurrent transformer network that localizes defects in an unsupervised manner.

Istituto Luce Cinecittà is an Italian society responsible for the preservation and distribution of the **Archivio Storico Luce**[2]**,** the largest Italian historical video archive dating from throughout the 1900s and comprising a variety of sources, provided us with some analog videos from this archive. These videos present several system intrinsic and aging-related types of degradations typical of analog video tapes. An example of such degradation is shown in the following figure. This poses a severe risk for the long term **maintenance** of

[2] https://www.archivioluce.com

these types of materials and in the future similar archives risk losing tens of years of archive materials, up to the years in which digital video recording became the common storage support.

Fig. 9 - example of frame with degradation due to aging of the support. Standard old film restoration can not solve this type of problem. Video provided by Istituto Luce.

Analyzing the existing approaches for video restoration, we can observe that they focus on standard structured defects such as scratches and cracks, so they are not capable of restoring the particular types of artifacts that analog video restoration requires. Unfortunately, when considering real-world archive videos, there is no clean high-quality version of them to use as ground truth for supervised learning. Consequently we created a synthetic dataset as similar as possible to the real-world videos to train and evaluate our system.

Starting from high-quality videos of the Harmonic dataset [Harmonic-2019] we used **Adobe After Effects** to randomly add several types of degradations, such as:

- Gaussian noise, resembling the tape noise that is typical of analog videos;
- white artifacts simulating tape dropouts;
- cyan, magenta and green horizontal lines resembling chroma fringing;
- horizontal displacements, similar to tape mistracking artifacts; this is the most complex error that can be encountered.

As with the real-world videos, all these artifacts vary over time and occur with different intensity, positions and combinations for each frame. The following figure shows an example of the synthetic dataset created to develop the restoration app. On the left the high quality frame, on the right the version with the artifacts generated using After Effects. The combination of frames is used to train the neural network.

Fig. 10 - example of synthetic training dataset. Left) original high-quality image; Right) degraded version created with Adobe AE. The network is trained to obtain the image on the left from the image on the right, used as input.

For the sake of simplicity the details of the network architecture are not reported in this document, as they are available in the associated scientific publication [Agnolucci-2022]. The novel neural network was designed using Transformer networks and processing a group of consecutive frames so that the restoration of a ruined frame is helped by the contextual data provided by the preceding and following ones (even if they are degraded as well). Given the complexity of the task the computation cannot perform in real-time and must be executed as a batch process.

Then this synthetic dataset was used to train our restoration model and evaluate its comparison with DeOldify, which is currently the only publicly available tool for this type of task. We measured the performance of our method using three standard full-reference visual quality metrics: 1) PSNR; 2) SSIM [Wang-2004]; 3) LPIPS [Zhang-2018]. The quantitative results are reported in the following table. For a fair comparison, we re-trained DeOldify from scratch using our training data. Our model achieves the best performance by a large margin.

Method	PSNR	SSIM	LPIPS
DeOldify	11.56	0.451	0.671
Our method	**34.78**	**0.939**	**0.063**

Table 4 - Comparison of the ReInHerit restoration app w.r.t. DeOldify. Higher values of PSNR and SSIM are better. Lower values of LPIPS are better. Best results are highlighted in bold.

The following figure shows examples of the results obtained by the neural network. On the left of each image is shown the input frame, on the right the desired output and in the middle the actual output of the network. These examples have been obtained from the synthetic dataset since it is the only way to have “gold standard” frames.

Fig. 11 - examples of video frames restoration: left) input: degraded frames; middle) output of the network; right) ideal result, i.e. original high quality frame used to produce the degraded version.

As mentioned above, when applying the system to real world video archives we cannot rely on high-quality reference images, they simply do not exist and the degraded frames are what is available in the archive. The following figures show what happens when the video restoration system is applied in a **real-world usage scenario,** showing an example of materials provided by Istituto Luce.

Fig. 12 - example of results obtained by smart video restoration app on real-world archive video. Video courtesy of Istituto Luce.

The smart video restoration app is composed of two parts: the **back-end**, implemented using PyTorch provides the computer vision functionalities that restore the video using a novel neural network architecture, these CV services are provided to the front-end by a REST API implemented in Flask. The **front-end** is a web app accessible through a browser that allows users to submit degraded videos, start their restoration and then download the restored versions. Screenshots of the web app are presented in the following, showing the whole **process to restore** a video.

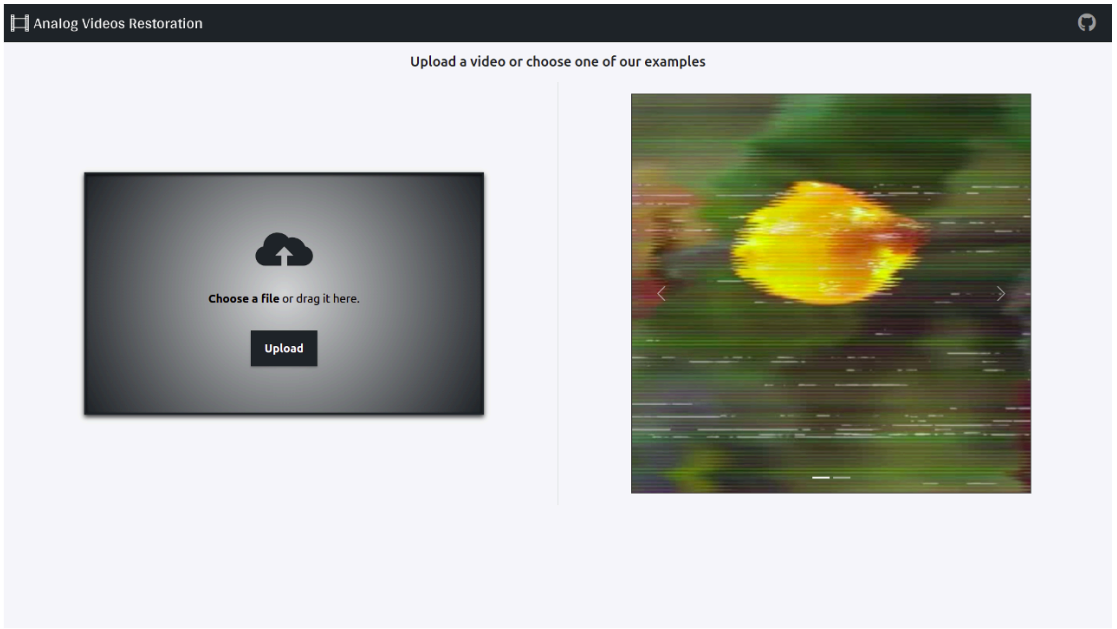


Fig. 13 - The user can upload a file or choose one of the given examples

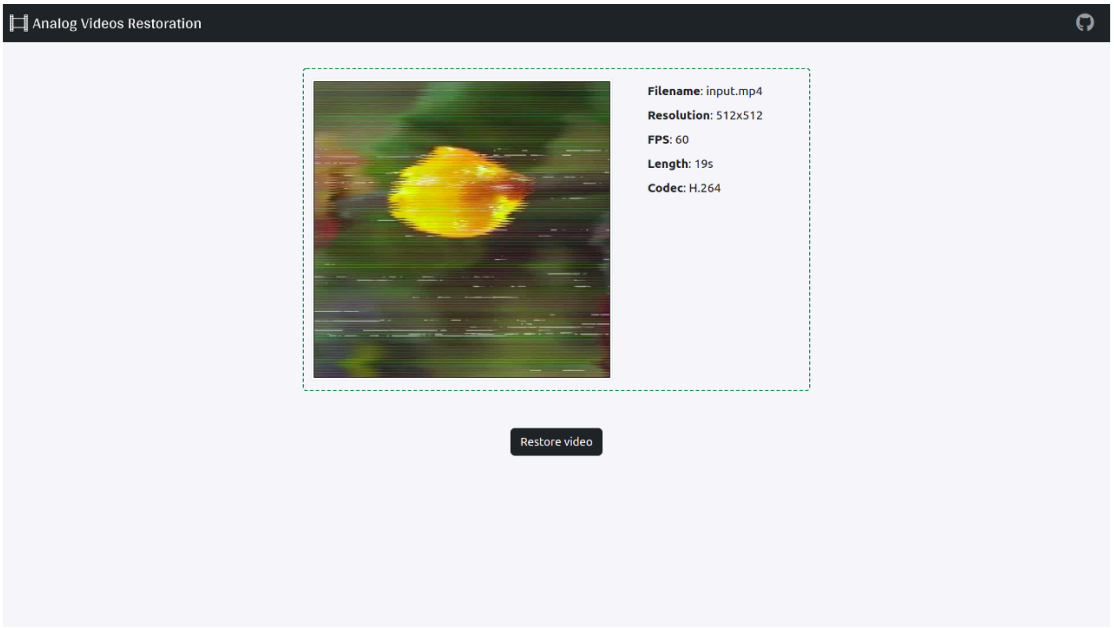


Fig. 14 - An overview of the chosen video provides some information about it, the user can check if the metadata is correct and then can start the restoration process.

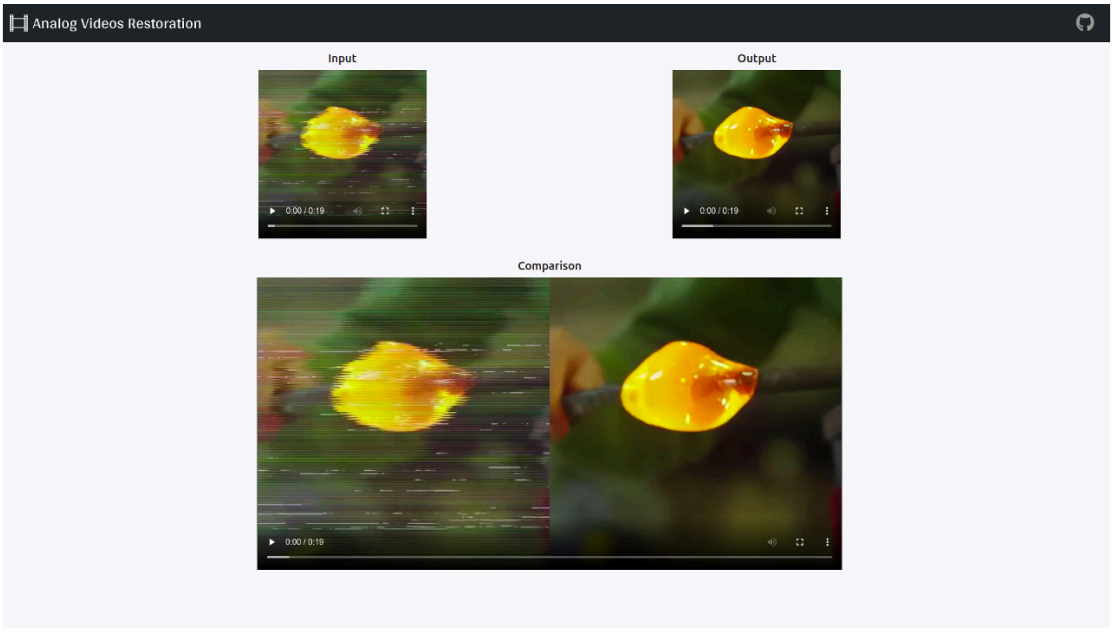


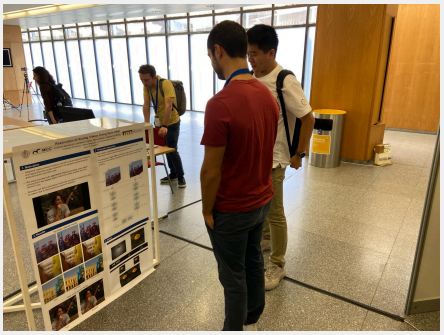
Fig. 15 - When the restoration process ends, the user can see and download the restored video and the comparison with the degraded one.

Since the restoration process is computationally expensive, the CV component must be executed on a server with appropriate GPU. It is also possible to use GPU servers commonly available from cloud providers such as Amazon AWS, Google Cloud, Microsoft Azure.

The video restoration app has been demoed live at ACM Multimedia 2022 conference, the foremost scientific conference on multimedia

Scientific impact:

Smart video restoration has been demoed live at ACM Multimedia 2022 conference, the foremost scientific conference on multimedia.




Associated papers describing the apps:

- L. Agnolucci, L. Galteri, M. Bertini, A. Del Bimbo, "Restoration of analog videos using Swin-UNet", Proc. of ACM Multimedia 2022
  The paper has received 2 citations, as of Aug. 2024.
- L. Galteri, L. Seidenari, P. Bongini, M. Bertini, & A. Del Bimbo, "Lanbique: Language-based blind image quality evaluation", ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 18(2s), 1-19.
  The paper has received 1 citation, as of Aug. 2024.

3.4 Strike-a-pose and Face-fit

Gamification is the process of exploiting strategies and game dynamics into scenarios that are not a game [Robson-2015]. It has already been proved to be useful to enhance skills and competences in a variety of domains such as marketing, industry training and entertainment. Certainly, also cultural heritage can benefit from a gamification approach which represents an opportunity to **engage visitors** to museums contents through the design of more entertaining, social and challenging **digital learning scenarios** [Karahan-2021], [Khan-2020], [Bonacini-2022], [Paliokas-2020], to help museums move from the traditional "look and do not touch" toward a "**play and interact"** approach. In fact, it has been observed that the availability of tools like gamified e-guides to visitors contributes to the sustainability of museums [Bieszk-Stolorz-2021]].

The goal of the two applications is to challenge the user to analyze and **replicate artworks** with their **own body and face**, obtaining 1) **information on the artworks** that are replicated and 2) **personalized artwork** and media representations that can be shared on **social networks**.

3.4.1 Strike-a-pose

Strike-a-pose is a **web application** which performs analysis and evaluation of **human poses** compared to poses present in famous paintings or statues. It is inspired, in spirit, by the

VanGo Yourself [3] (that does not use computer vision) and the ARTLens installation at CMA[4], using computer vision to evaluate how near is the pose of the person to the artwork, and adding a **challenge** that asks users to replicate more than one artwork.

Strike-a-Pose can be made available on the visitors' smartphone, following the **"Bring Your Own Device" (BYOD)** approach that has become more common in museums since the COVID-19 pandemic; the system can be used also in a dedicated environment in a gallery, using a standard PC equipped with a large screen and a camera. The same code base is used for the two setups, easing the maintenance of the application. The goal of having the application executing also on mid-level phones means that there's no need for powerful workstations in case it is used as an installation. The design of the interface adapts to the two different modalities, providing both a vertical interface suitable for a mobile phone and a horizontal one for installation and desktop.

The application exploits a **gamification paradigm** with the **educational** purpose of getting users interested in works of art using fun. Once registered, the user is challenged to reproduce in sequence the poses of some artworks from the museum's collections. The skeleton of both the artwork and the visitor can be displayed on the screen in order to facilitate the user in matching the various points and segments. **Matching the poses** provides the **descriptions of each artwor**k. The poses to be matched are organized in sets of **challenges**, e.g. challenges to replicate poses using the whole body, using only the torso (e.g. to allow also wheelchair users to interact), or any other type of challenge that is considered interesting by the museum curators (e.g. based on thematic collections). Once all the poses have been matched, the application allows the user to **generate a video** that can be saved for any social sharing. The video shows the user matching process and the overall interactive experience lived at the museum. The basic application can be adapted to provide variations of the gamification, e.g. **introducing a competition** between different users. An example of screenshots of the basic app are shown in the following figure.

[3] https://vangoyourself.com

[4] https://medium.com/cma-thinker/strike-a-pose-selfies-and-infinitekusama-at-the-cma-2c3b16671329

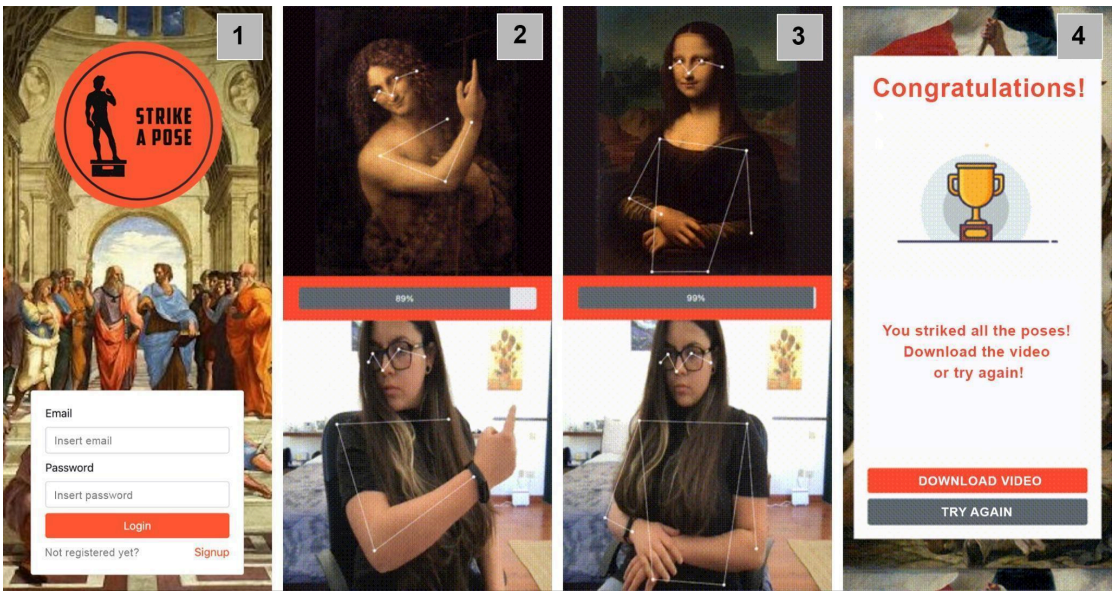


Fig. 16 - Strike a Pose App for smartphone (pilot version). 1) Login. 2-3) The user trying to strike the pose in the painting (playing in "easy" mode, with visible skeletons). 3) Challenge completed: download the video.

The application has been developed in **JavaScript** on the client side and in **Python** on the server side. Pose detection on the human bodies is achieved using **TensorflowJS** detection API exploiting the pose detection model, MoveNet. **MoveNet** is a very fast and accurate model that detects 17 key points of a body. The model is used in the variant "Lightning" intended for latency-critical applications and runs faster than real time (30+ FPS) on most modern desktops, laptops, and phones. The model runs completely client-side in the browser; this allows us to run the whole computer vision task on the device of the user, providing a better user experience thanks to the reduced latency for the pose analysis. Server-side an SQLLite database is used to store artworks' collections, challenges and artworks' metadata and descriptions. Communication between the knowledge-base and the interface is ensured through RESTful APIs developed in Flask. The video is created server side. Details on pose representation and matching are provided in the associated scientific publication [Donadio-2022].

The base interface, implemented in HTML can be adapted by different users, maintaining the computer vision functionalities, so as to allow **customization** by different museums. An example of such customization (pilot version) is shown in the following figures:

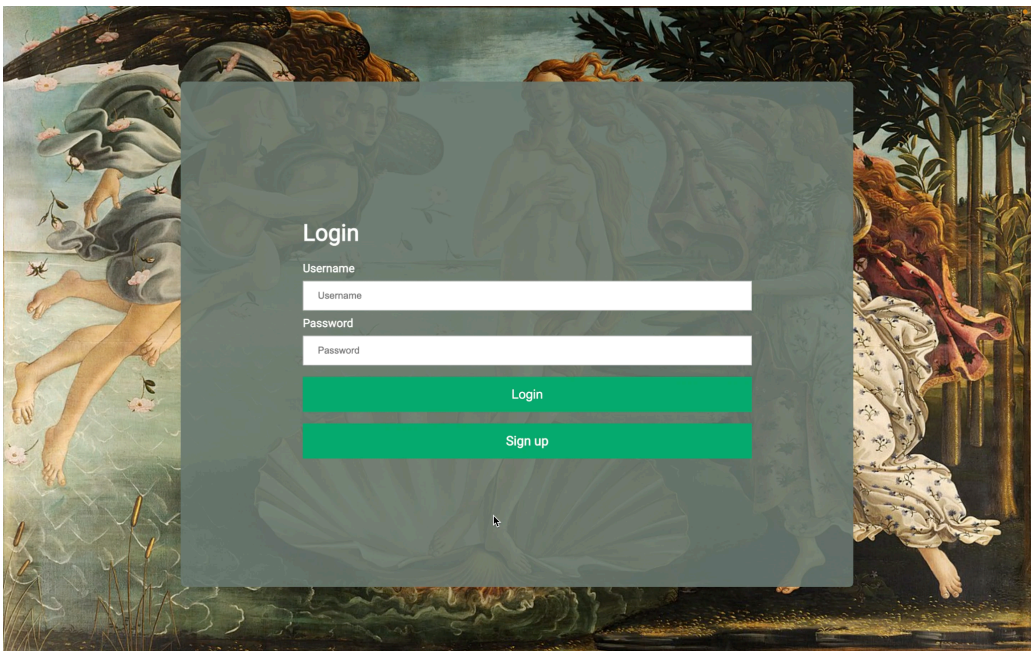


Fig. 17 - customized login screen (pilot version)

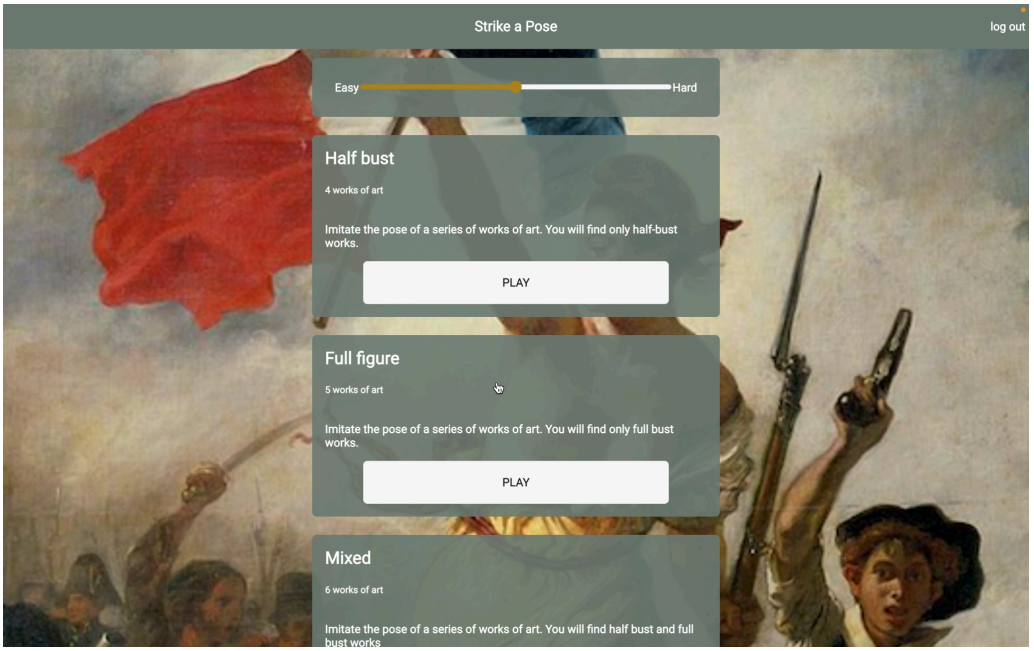


Fig. 18 - customized challenge selection (pilot version)

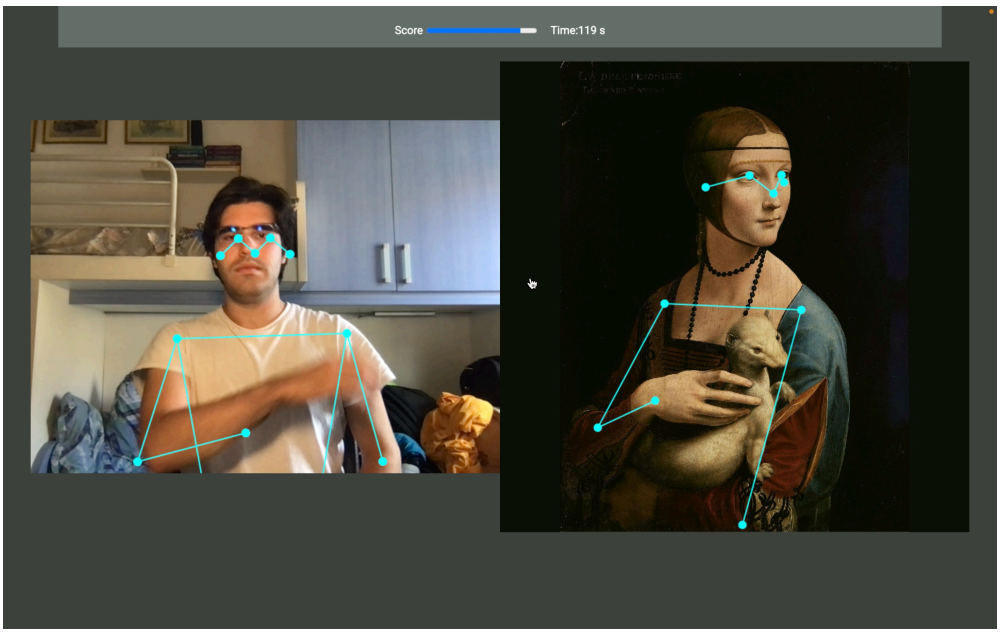


Fig. 19 - horizontal interface for installations - customized template (pilot version)

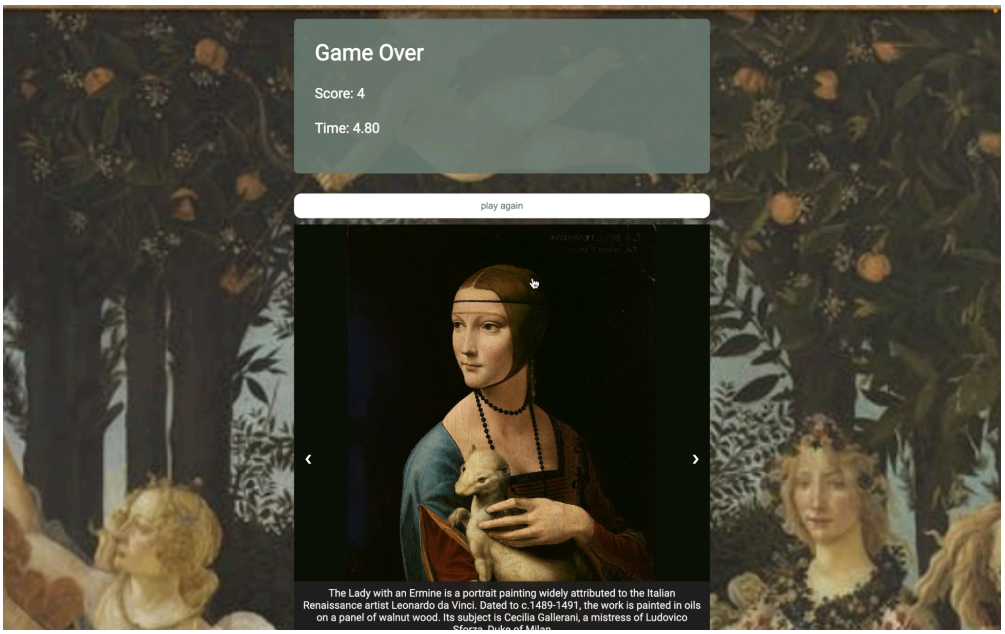


Fig. 20 - customized end-game screen with informations about the artworks
The following figure shows a variation of the basic app to allow competitions between users. (pilot version)

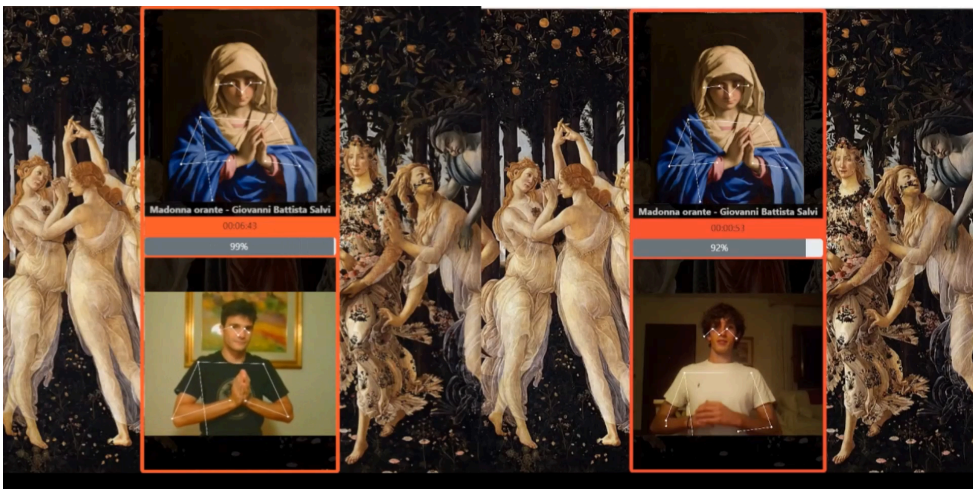


Fig. 21 - Strike-a-pose (pilot version)in a competitive setup. Two users attempt to complete the same challenge in less time.

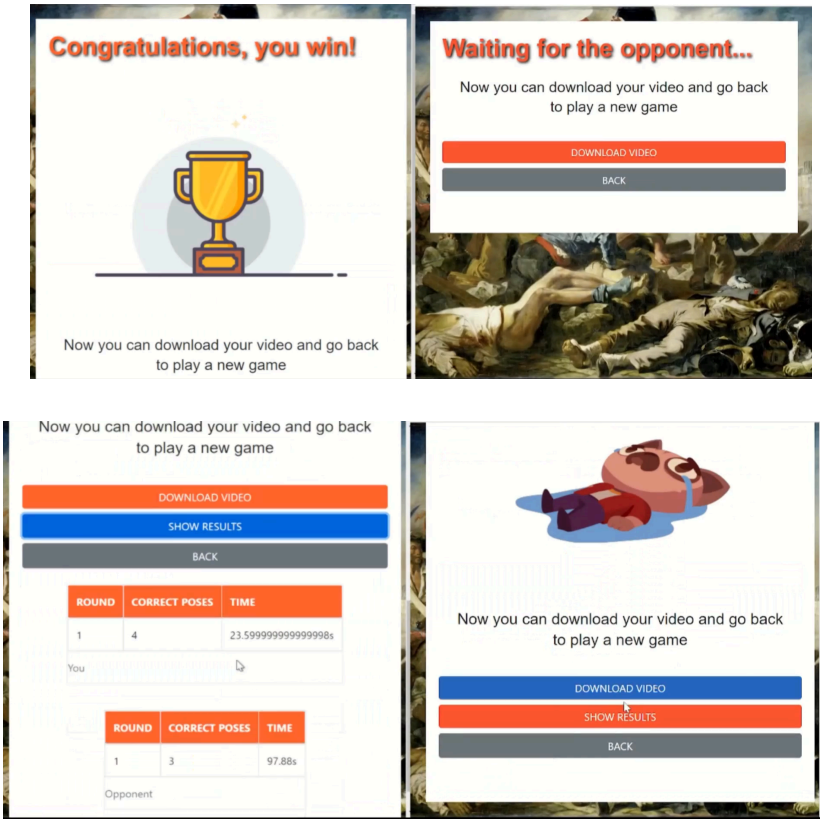


Fig. 22 - Strike-a-pose (pilot version) in a competitive setup: results of the challenge for each user.

3.4.2 Face-fit

Face-fit follows the idea of Strike-a-pose, i.e. asking the user to **replicate an artwork**, but concentrating on **facial expressions** that require a much more refined matching.

Also this application can be used on a **mobile phone** or on a PC, but due to technical limitations of some required computer vision libraries the Javascript version for mobile phones need to relieve some image processing functionality to a server. For this reason the app has been developed using two codebases: a **Javascript** one, with **TensorflowJS**, for mobile phones and a **Python** version using **OpenCV** and **Tensorflow** for the desktop app for museum installations.

The application asks the users to **replicate the pose** of the **head** and the **expression of some portraits** by famous painters and transfer the face of the user on the artworks, generating a new image. The application was designed through a usability study carried out following an iterative design approach with three groups of 5 people [Nielsen-1993]. The user places himself in front of the **smartphone or installation equipped with a camera.** He is presented with a series of portraits' paintings in a vertical carousel. The user can choose the artwork to match. At that point the application presents a ghost image of the user's face that the user must try to super-impose on that of the painting to find a perfect match, see following figure. The **ghost image solution** was the result of our usability study which solved some issues related to how to keep the user at the same time concentrated on the task without losing the fun of the game. At first, in fact, we had provided some visual suggestions to find the right pose but they distracted the user from the painting and therefore from the game.

Fig. 23 - Face-Fit App (pilot version) for museum installation: select an image.

A faster than real-time face mesh prediction network is used to obtain 468 3D points for each face, also when using mobile phones.

The points are used to compute the pose of the whole face. Once the pose is matched, the position of eyes, eyebrows and mouth is matched. When both pose and facial expression

match, the face of the user is substituted to that of the painting and the description of the artwork is provided.

Once the pose is matched the user obtains **information on the artwork** and can **download** the **generated images** for sharing on **social networks**.

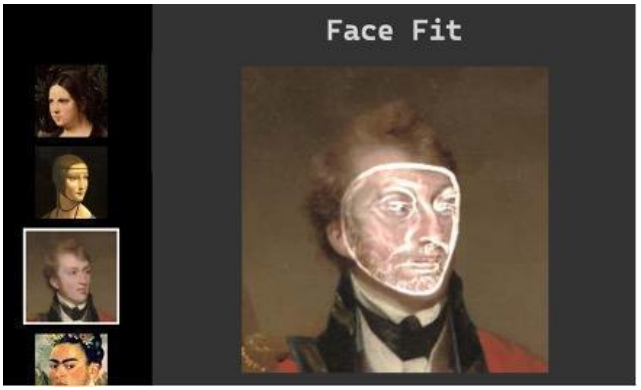


Fig. 24 - user interaction: the ghost image gives feedback to the user to change his pose and expression to better match that of the artwork. (pilot version)

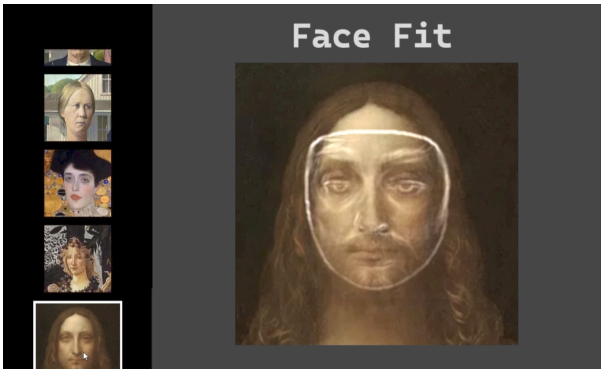


Fig. 25 - Matching the pose and expression with Leonardo's Salvator Mundi and generation of the image merging the user face in the artwork. The image is emailed with info on artwork to the user. (pilot version)

Scientific impact:
Both **Strike-a-pose** and **Face-fit** have been demoed live at ACM Multimedia 2022 conference, the foremost scientific conference on multimedia, obtaining the Best Demo Honourable Mention award for the engaging museum experience they provide; the award was given by the scientific committee of the demo track of the conference. The apps were used by the attendees of the conference over several days of demos.

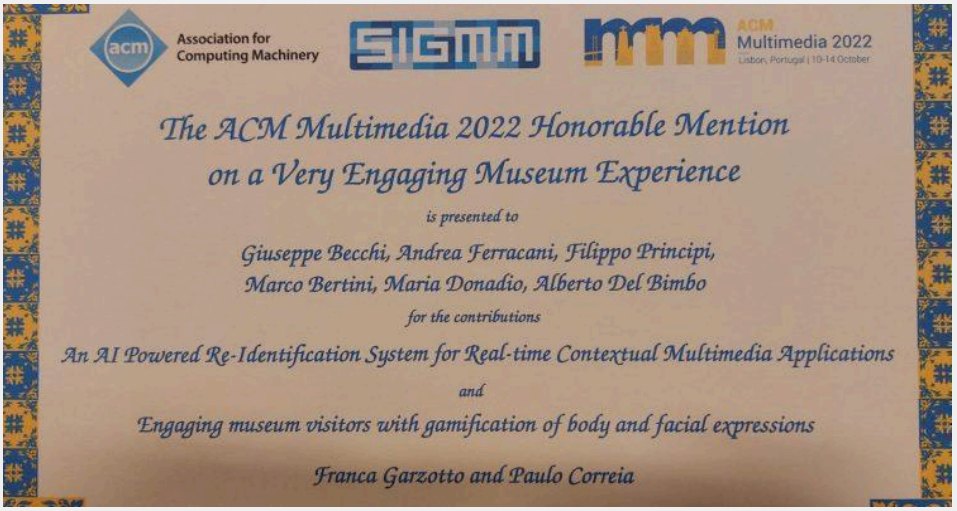


Fig. 26 - ACM Multimedia 2022 Best Demo Honourable Mention award to the paper [Donadio-2022] for the very engaging museum experience provided by the Strike-a-pose and Face-fit apps.
Associated papers describing the apps:

- M.G. Donadio, F. Principi, A. Ferracani, M. Bertini, A. Del Bimbo, “Engaging museum visitors with gamification of body and facial expressions”, Proc. of ACM Multimedia 2022
  The paper has received 4 citations, as of Aug. 2024.

3.4.3 Co-creation and Ethical use of AI tools

Concerning the web-apps Strike-Pose, Face-Fit, ethical issues related to the use of personal data for interaction with users were defined and specified. It must be highlighted that personal data is never stored, shared with any party nor used to train the systems. In particular, facial/body images are not stored to respect privacy, and an appropriate checkbox was included in the apps' privacy policy.

Since these applications should work with users that may have very diverse physical attributes, the neural networks used have been selected so to have a fair performance w.r.t. geographical origin, age, gender and skin color. The machine learning and the computer vision communities have started to deal with the need to improve the transparency of the models used by adopting so-called “Model cards”. Model cards are a form of documentation that provides a standardised way of presenting information about Machine Learning models. They were first introduced by Google in 2018, and have since become increasingly popular across the industry because they provide a concise, holistic picture of a Machine Learning model. They include the following information:

- Model description
- Intended use

- Features used in the model
- Metrics
- Data used for training & evaluation
- Limitations
- Ethical considerations

Considering the neural network used in Strike-a-pose, the model predicts 17 human keypoints of the full body even when they are occluded. The developers of the network have performed a fairness evaluation, analyzing the model performance under different person attributes and categories:

- Gender: Male/Female
- Age: Young/Middle-age/Old
- Skin tone: Medium/Darker/Lighter

evaluating the precision of the detection of the points. The authors of the model have concluded that the model performs fairly (< 5% performance differences between categories)[5].

Considering the neural network used in Face-fit, the goal is the detection of human facial surface geometry from monocular video. As noted in the model card, the predicted face geometry does not provide facial recognition or identification and does not store any unique face representation.The developers of the network have tested the fairness of the system considering gender, skin tones and geographical origins of the users. As reported in the model card[6], the mean absolute error of the predicted mash shows that the model performs fairly across different genders, skin tones and geographical regions.

The two applications do not store any image of the user, and the generated media are sent by email or download, and then immediately deleted. The applications, if executed on the device of the user, request access to the camera, an access that is needed to implement the gamified experience. The privacy policy linked in the app reminds the user that no data is stored and that the purpose of accessing the camera is to implement the gamification aspect of the application.

Further guidance will also be included to inform museums about the use of copyright-free museum images and validated data to be provided and included as additional information for users. [7]

Strike-a-Pose and Face-Fit web apps have been tested and studied during an Hackathon organized in the context of the AI&XR Summer School held in Matera in July 2023 (Fig. 27) A multidisciplinary group of young Ph.D. students worked on the theme "Gamification and Playful Approach," tested the Strike-a-Pose and Face-Fit apps, using open-source codes from Digital Hub.

[5] MoveNet.SinglePose Model card; available at https://storage.googleapis.com/movenet/MoveNet.SinglePose%20Model%20Card.pdf
[6] MediaPipe Face Mash Model card; available at https://drive.google.com/file/d/1QvwWNfFoweGVjsXF3DXzcrCnz-mx-Lha/preview
[7] Please refer to D3.2 - section 6 Ethics Annex

Fig. 27 - AI&XR Summer School held in Matera

During Summer School MICC - University of Florence presented a practical lecture on '*Innovative and sustainable approaches for user engagement and digital interaction with cultural heritage'*. (Fig. 28)

Fig. 28 - Practical lecture, AI&XR Summer School Matera

In order to adopt a sustainable and user-centered approach, the aim was to share the relevant results of the playful approach and user engagement studies conducted within the ReInHerit H2020 research project. Cutting-edge applications based on artificial intelligence developed by the MICC were presented to inspire the students' project proposals. Participants were invited to use, test, and explore the Toolkit in a collaborative and interdisciplinary approach, linking technological and cultural sectors.

During the week, international speakers and famous experts debated and engaged with international students and researchers with different educational content and skills. Young PhD students worked in an interdisciplinary way on the ReInherit web apps Strike-a-Pose / Face-Fit using open-source codes shared by the Digital Hub. This co-creation process added

new technological developments for apps and user interaction scenarios, improving engagement, inclusivity and new design features[8]. (Fig. 29)

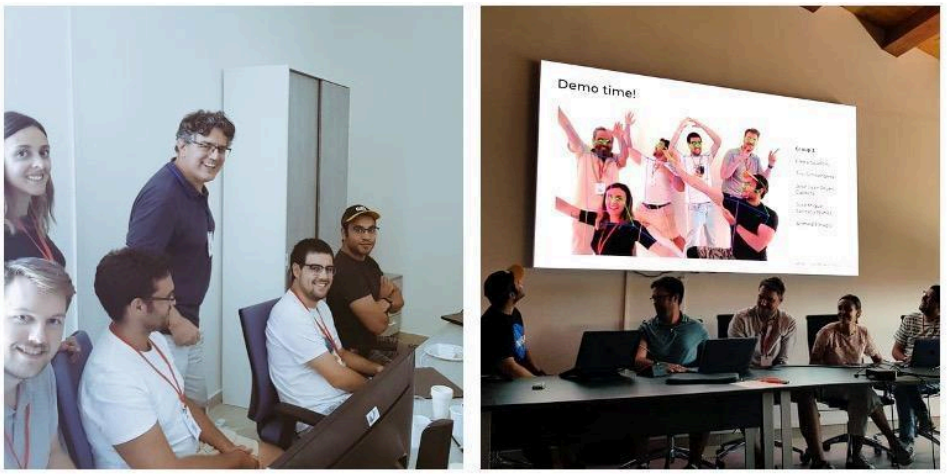


Fig. 29 - Hackathon working Group, AI&XR Summer School Matera

Strike-a-Pose 2.0 [9] proposal aims to increase gamification by guiding the users in interacting with museum artifacts, recreating the characters' pose of famous paintings.(Fig. 30)

Fig. 30 - Strike A Pose v.2.0 - Hackathon Proposal

Starting from the existing application, the team focused on providing more instructions to the user in interacting with the application, improving the codes and design. After downloading the app and going to the museum, the user is invited to find a painting and to select the same painting on the application. (Fig. 31)

[8] ReInHerit Digital Hub https://reinherit-hub.eu/summerschool/
[9] https://reinherit-hub.eu/summerschool/6205f8e2-60aa-46d2-bca3-bc46c9283029

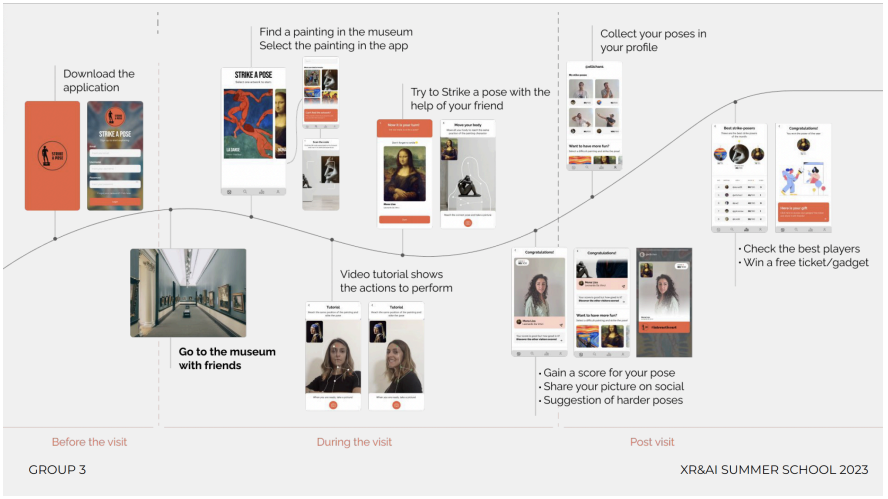
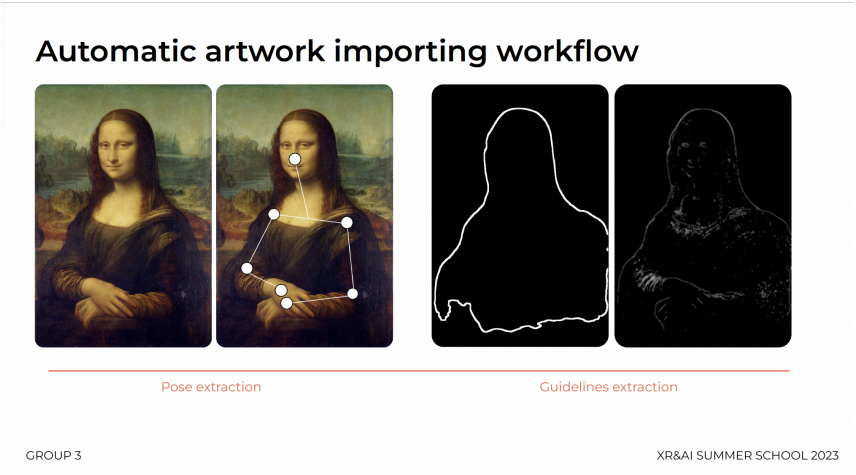


Fig. 31 - Strike A Pose v.2.0 Hackathon Proposal - Description

Then, a brief tutorial shows the user the action to perform in order to recreate the pose of the selected painting before starting the real interaction. During the interaction, some feedback have been designed to guide the user in recreating the character's position:

- an outline shows the silhouette of the character
- a skeleton of the character shows the key points to match
- Some arrows starting from the user body indicate the correct direction to move forward and change their color (from red to green) in relation to the target position. (Fig. 32)



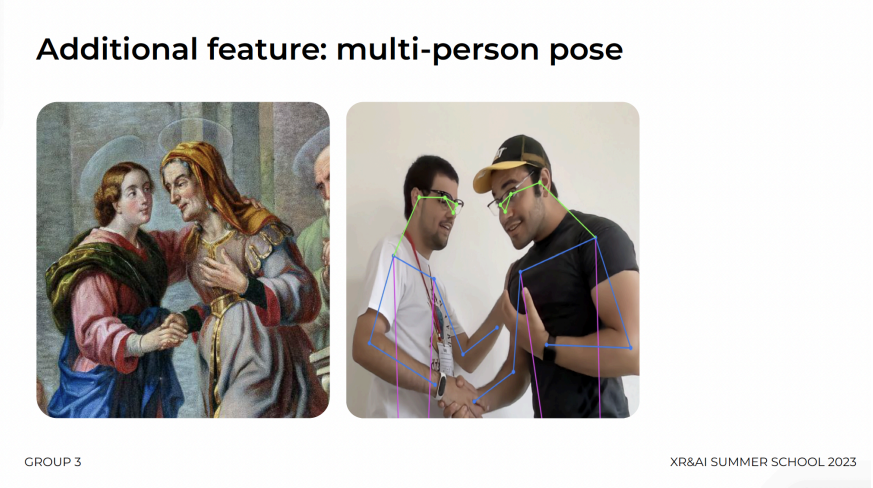


Fig. 32 - Strike A Pose v.2.0 Hackathon Proposal - Additional Features

After recreating the pose in the most accurate way, the user can take a picture of their position, which can be shared on social media with a dedicated hashtag. In addition, the recreated position gains a score, considering the assurance of the reproduction, in comparison to the real artwork. Based on the obtained scores, the application shows a leaderboard of all the visitors by showing the most accurate pose of the month: the best stike-posers! The best strike-poser of the month receives a gift in the form of a museum souvenir or a free ticket to visit the museum again or to give it to a friend. The application also offers the possibility to collect the user's personal poses in a personal profile and to select always more challenging poses, by increasing the game difficulty. The project has the purpose ofIncreasing entertainment, promoting edutainment and stimulating young people to visit museums.

During the hackathon, professionals and experts were invited to chat via the Digital Hub Forum to support students as they work, answering questions and sharing ideas and suggestions in a co-creative and interdisciplinary approach. The Strike-a-Pose 2.0 project proposal was awarded as the best and evaluated according to criteria of coherence, originality and innovation, technological quality and communication skills. (Fig. 33)

Youtube short video: https://youtu.be/IWGtKHc6Pqc?si=L3a3SyN-ntXlGEMe

Fig. 33 - Strike A Pose v.2.0 - Award

These ReInHerit apps testing activities follow the approach already tried out during the Research Fair organised in May 2023 at the Arcada University of Applied Sciences, one of the ReInHerit partners. A station was set up to test the apps and researchers, lecturers and guests were invited to try them out. The aim of this session was to test the functionality of the demo version of the developed apps, in particular their playful approach and the level of user involvement in their use. (Fig. 34)

Fig. 34 Twisting and testing ReInHerit apps at Arcada University of Applied Sciences

This event was also an opportunity to learn from interaction with users how to improve our tools according to their needs and suggestions. User involvement is crucial to mediate meaningful dialogues and pave the way for innovation.

3.5 Webinars

Availability of **source code** and **app documentation** is fundamental for the **maintenance** of the apps developed in the toolkit, since they address the information needs of the developers. In addition, to ease the evolution and adaptation of apps and components to create new applications developed by the users of the Digital Hub, a series of **technical webinars** addressing AI, CV and modern ICT tools in general, will be prepared and made accessible from the Digital Hub. The goal of these webinars is to provide a **higher-level information** regarding advanced ICT **technologies**, that complements the **low-level information** on the specific **implementations** that is available in the code itself and its documentation.

From the perspective of a **multidisciplinary and collaborative approach**, the primary research conducted in WP2 highlighted that the training needs of professionals are not only exclusively technological. Regarding the Digital Hub webinar topics, the following themes were indicated in D7.5 "Dissemination and Exploitation Plan 2nd Version" to be selected and included in the training curriculum and syllabi, addressing topics of interest for the CH domain in general including also cultural tourism:

- Digital and emerging technology skills
- IPR Intellectual Property Rights
- Traveling Exhibition planning
- Digital Exhibition planning
- Immersive Performance
- Gamification
- Cultural Tourism
- Digital Hub
- Conservation and preservation
- Storytelling skills
- Fundraising skills
- Negotiation and Listening
- Marketing and Communication
- Management
- Leadership
- Soft skills
- Inclusive Museum
- Sustainability and Heritage

From the **technical point of view** webinars will be produced using a specialized service provider that provides some form of visual customization of the events to make them visually part of the ReInHerit website look-and-feel, and will be recorded. Videos and materials will then be stored on the ReInHerit Digital Hub. A technical analysis of webinar service providers has resulted in the selection of **Clickmeeting**[10], because it allows a certain level of personalization (e.g. advanced appearance, customization of the layout of the waiting room and the login page, with logo, colors, background images etc.). Webinars will

[10] https://clickmeeting.com/

be provided on Clickmeeting Platform, creating a waiting room page (Fig. 35) with a ReInherit logo and customized background. Registration form, announcement and recorded videos after the webinar will be posted on the Digital Hub, while the live webinar will be hosted on ClickMeeting during its duration. In addition, the system provides the video recording file, participants' statistics report and a good level of interactive functionalities in the call room (video/audio, chat, uploading and opening files, screen for presentation, switching between files i.e. videos, audios, images, presentation PDF, opening a youtube video links, using white board, using desktop share , using surveys, Q&A, call-to-action, highlighting and denoting with cursor and chat). Clickmeeting is also a webinar platform known by cultural professionals and used by many cultural organizations, e.g. NEMO - *Network of European Museum Organisations* - has been using this platform for years for its webinars[11]. From 2017 to 2019, NEMO managed webinars in collaboration with UniFi-MICC and NEMECH - *New Media for Cultural Heritage*.[12] In 2017-2018 MICC organized the webinars courses of the Erasmus+ Project UMETECH - *University and Media Technology for Cultural Heritage*, using Clickmeeting platform.[13]

An **internal protocol** to develop webinars has been defined, to ease the production of materials and shared among the ReInHerit partners that will produce the webinars.

The webinars will be **delivered from November 2023 to October 2023** on a bi-monthly basis. Two webinars per month were considered sufficient to organise and manage the work, publish news on the digital hub and promote registrations through ReinHeirit project channels. The organisation and distribution of the webinar news was managed by MICC/Unifi, in particular through the NEMECH Regional Competence Centre contact mailing list. This mailing list consists of a huge network of museum professionals who received updates about the webinars. In total, more than approximately 200 experts were involved and participated in various online webinar sessions.

The **organisation** of each webinar includes:

1. Email contact with the project partner to select title, topics, date, and expert speakers;
2. Request for webinar news materials for registration on the Digital Hub (at least 2 weeks before each webinar);
3. Sharing of the news on the ReInherit social media channels and through the contacts and mailing lists of the partners and experts involved;
4. Each webinar is preceded by a test trial session with speakers held by the ReInherit host (MICC) to introduce the platform and organize the timing and structure (slide and video sharing, Q&A via chat, creation of interactive questionnaires for participants)
5. During the official webinar session the host MICC always follows the session for technical support and to give information via chat, start the video recording and

[11] https://www.ne-mo.org/training/nemo-webinars.html
[12] http://nemech.unifi.it/webinar-nemo-nemech-micc/
[13] http://morpheus.micc.unifi.it/umetech/index.php/category/courses/

create screenshots of the webinar, to be used for subsequent publication and promotion activities;

6. After each webinar, useful materials (slides and video recordings) are collected together with statistic reports on attendees generated by the platform and shared with the speakers, partners and coordinators of the ReinHerit Project;
7. After each webinar MICC creates and edits the resource section of each webinar, including images and all training materials.[14]

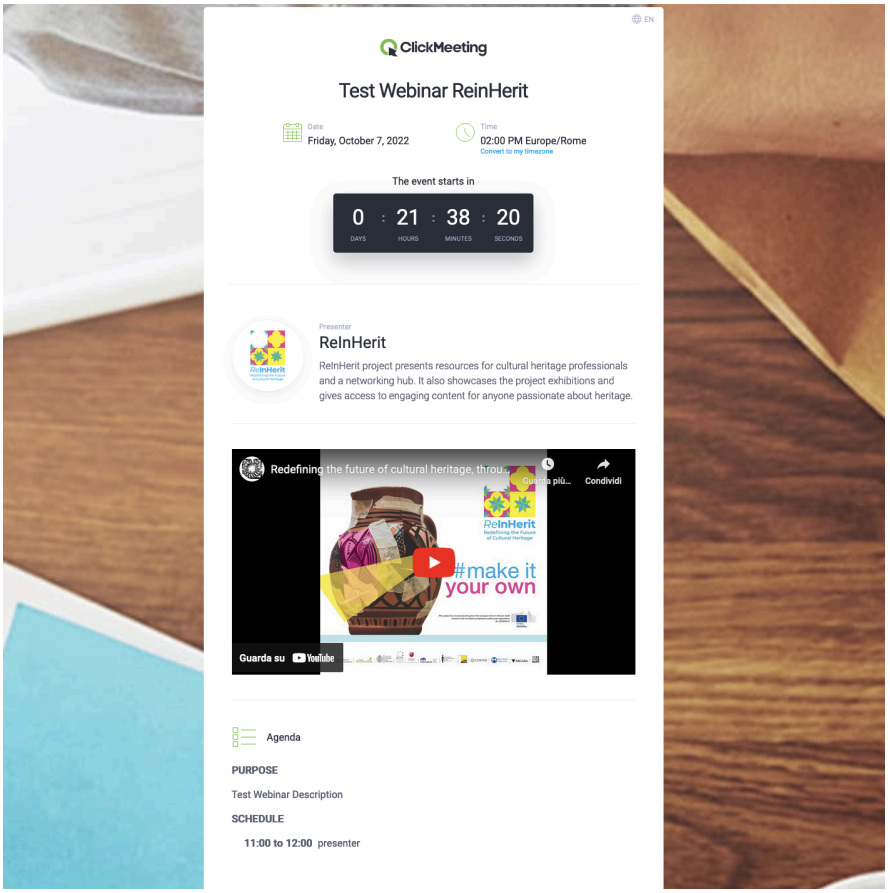


Fig. 35 Screenshot of a Clickmeeting customized "Waiting-Room" page for a ReinHerit webinar test

[14] More details on webinar topics and material are included in - D3.9 - "Training Curriculum and Syllabi"

Appendix

List of **meetings and workshops** associated with the production of this deliverable.

1. Workshop AI+CV for Cultural Heritage (21st February 2022) MICC-UNIFI_ WP3_T3.3
2. Consortium Meeting - 2nd year (28th March 2022)
3. Workshop “Artificial Intelligence and Computer Vision for Cultural Heritage” Graz - AU, in collaboration with MICC-Unifi (WP3), UniGraz (WP4) and GrazMuseum (WP6) - (23-24 May 2022)
4. “IPR training meeting” internal meeting for all Consortium partners (6 July 2022)
5. Meeting on Developing Webinars - in collaboration with BoCCF and UNIFI (11 July 2022)
6. Smart tourism app meeting (WP3/WP5) - In collaboration with MICC-Unifi, Museum of Cycladic Art, ECTN; BoCCF (20 July 2022)
7. WP3 toolkit meeting (WP3/WP6) - in collaboration with Media Integration and Communication Center UNIFI, GrazMuseum (22 July 2022)
8. Digital Toolkit | Workshop, Smart Lens app meeting (WP3/WP6) - - In collaboration with MICC-Unifi, GrazMuseum, Museum of Cycladic Art, BoCCF (23 September 2022)
9. Regular meetings of the tech committee are taking place every month. In collaboration with MICC-Unifi, GrazMuseum, Bank of Cyprus Cultural Foundation, CYENS.
10. ReInHerit Study Visit and Consortium Meeting in Brussels, 8-12 May 2023 Brussels
11. Testing ReInHerit apps | Research Fair at the Arcada University of Applied Sciences, May 2023 - Finland
12. ReInHerit mobility activities, visit and meeting with BoCCF in Florence at MICC Media Integration and Communication Center of the University of Florence, 13 July 2023
13. ReInHerit Hackathon | XR&AI Summer School 2023 Matera IT 17-22 July 2023.

References

- DoA, Part A / B
- D2.1 “Focus Group Report”
- D2.2 “State of the Art Report - Literature Review”
- D2.3 Questionnaires Report
- D2.4 Focus Groups Report Phase II
- D2.5 “CH Management Guidelines”
- D2.6 “A sustainable model of CH management state of the art report”
- D3.1 “National Surveys Report”
- D3.2 “Toolkit Strategy”
- D3.4 “Consolidated Report on ICT in CH Management”
- D4.1 “Requirements Analysis Report”
- D4.2 “Digital hub”
- D7.5 “Dissemination and Exploitation Plan 2nd Version”
- [Baldrati-2022a] A. Baldrati, M. Bertini, T. Uricchio, and A. Del Bimbo. Conditioned and composed image retrieval combining and partially fine-tuning clip-based features. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pages 4959–4968, June 2022.
- [Baldrati-2022b] A. Baldrati, M. Bertini, T. Uricchio, and A. Del Bimbo. “Effective conditioned and composed image retrieval combining clip-based features.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 21466–21474, June 2022.
- [Baldrati-2022c] A. Baldrati, M. Bertini, T. Uricchio and A. Del Bimbo. “Exploiting CLIP-based Multi-modal Approach for Artwork Classification and Retrieval”. In Proceedings of Heri-Tech - The Future of Heritage and Science and Technologies Conference, Springer, 2022
- [Donadio-2022] M.G. Donadio, F. Principi, A. Ferracani, M. Bertin i and A. Del Bimbo. “Engaging Museum Visitors with Gamification of Body and Facial Expressions” In Proceedings of the 30th ACM International Conference on Multimedia (MM '22), October 10--14, 2022, Lisboa, Portugal
- [Agnolucci-2022] L. Agnolucci, L. Galteri, M. Bertini and A. Del Bimbo. “Restoration of Analog Videos Using Swin-UNet” In Proceedings of the 30th ACM International Conference on Multimedia (MM '22), October 10--14, 2022, Lisboa, Portugal
- [Radford2021] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sas- try, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever.” Learning transferable visual models from natural language supervision,” 2021
- [DelChiaro2019] R. Del Chiaro, A. D. Bagdanov, andA. Del Bimbo. Noisyart: Adataset for webly-supervised artwork recognition. In VISIGRAPP (4: VISAPP), pages 467–475, 2019.A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sas- try, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever. Learning transferable visual models fromnatural language supervision, 2021
- [DelChiaro2019b] R. Del Chiaro, A. D. Bagdanov, A. Del Bimbo, Webly-supervised zero-shot learning for artwork instance recognition, Pattern Recognition Letters, Volume 128, 2019
- [NEMO 2023] Digital Learning and Education in Museums - Innovative Approaches and Insights”" Report 2023, NEMO - The Network of European Organizations
- [Selvaraju2019].R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, andD. Ba- tra. Grad-cam: Visual explanations from deep networks via gradient- based localization. International Journal of Computer Vision, 128(2): 336359, Oct 2019
- [Deoldify] 2018. DeOldify. https://github.com/jantic/DeOldify
- [Iizuka-2019] Satoshi Iizuka and Edgar Simo-Serra. 2019. DeepRemaster: Temporal Source-Reference AttentionNetworks for Comprehensive Video Enhancement. ACM Transactions on

Graphics (Proc. of SIGGRAPHAsia 2019) 38, 6, Article 176 (2019)
- [Wan-2022] ZiyuWan, Bo Zhang, Dongdong Chen, and Jing Liao. 2022. Bringing Old Films Back to Life. CVPR (2022).
- [Wang-2004] ZhouWang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. IEEETransactions on Image Processing (TIP) 13, 4 (2004), 600612.
- [Zhang-2018] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and OliverWang. 2018. The unreasonable e>ectiveness of deep features as a perceptual metric. In Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- [Robson-2015] Karen Robson, Kirk Plangger, Jan H Kietzmann, Ian McCarthy, and Leyland Pitt. 2015." Is it all a game? Understanding the principles of gamification." Business horizons 58, 4 (2015), 411–420.
- [Karahan-2021]Sevde Karahan and Leman Figen Gül. 2021. "Mapping Current Trends on Gamification of Cultural Heritage." In Game + Design Education, Özge Cordan, Demet Ar- slan Dinçay, Çağıl Yurdakul Toker, Elif Belkıs Öksüz, and Sena Semizoğlu (Eds.). Springer International Publishing, Cham, 281–293.
- [Khan-2020] Imran Khan, Ana Melro, Ana Carla Amaro, and Lídia Oliveira. 2020. "Systematic review on gamification and cultural heritage dissemination." Journal of Digital Media & Interaction 3, 8 (2020), 19–41.
- [Bonacini-2022] Elisa Bonacini and Sonia Caterina Giaccone. 2022. "Gamification and cultural institutions in cultural heritage promotion: A successful example from Italy." Cultural trends 31, 1 (2022), 3–22.
- [Paliokas-2020] Ioannis Paliokas, Athanasios T Patenidis, Eirini E Mitsopoulou, Christina Tsita, George Pehlivanides, Elli Karyati, Spyros Tsafaras, Evangelos A Stathopoulos, Alexandros Kokkalas, Sotiris Diplaris, et al. 2020. "A gamified augmented reality application for digital heritage and tourism." Applied Sciences 10, 21 (2020), 7868.
- [Bieszk-Stolorz-2021] Beata Bieszk-Stolorz, Krzysztof Dmytrów, Jurgita Eglinskiene, Susanne Marx, Agnieszka Miluniec, Karolina Muszyńska, Grażyna Niedoszytko, Weronika Podle- sińska, Attila v. Rostoványi, Jakub Swacha, René Larsen Vilsholm, and Senija Vurzer. 2021. "Impact of the availability of gamified e-guides on museum visit intention." Procedia Computer Science 192 (2021), 4358–4366. https://doi.org/10. 1016/j.procs.2021.09.212 Proc. of International Conference Knowledge-Based and Intelligent Information & Engineering Systems.
- [Nielsen-1993] J. Nielsen. 1993. Iterative user-interface design. Computer 26, 11 (1993), 32–41.
- [UNESCO 2023] UNESCO 2023 *"Guidance for generative AI in education and research"* https://unesdoc.unesco.org/ark:/48223/pf0000386693
- [Villaespesa Murphy 2020] Elena Villaespesa, Oonagh Murphy, "AI: A Museum Planning Toolkit" - The Museums + AI Network - Goldsmiths, University of London New Cross London SE14 6NW https://themuseumsai.network/toolkit/