VALUE: Visual Analysis for Location and Understanding of Environment

Michele Mazzamuto*, Daniele Di Mauro*, Francesco Ragusa*, Irene D'Ambra*, Andrea Sciuto†, Angelo Aiello†

Flavio Licandro†, Ignazio Infantino‡, Antonino Furnari*, Giovanni Signorello*, Giovanni Maria Farinella*

*University of Catania, †Xenia Progetti s.r.l., ‡ICAR-CNR

***Abstract*—Cultural sites, such as museums and monuments, are popular tourist destinations worldwide, attracting visitors eager to explore the culture, history, and art of specific regions or countries. However, traditional visiting approaches may limit many cultural sites, failing to fully engage visitors. This challenge motivates the development of intelligent systems that integrate computer vision techniques and mixed reality to enhance cultural heritage visits. In this context, we propose VALUE, a mixed reality application designed for cultural environments, offering benefits for both the visitor and the museum manager. Utilizing egocentric image analysis and signals such as gaze, facilitated by the use of smart glasses, VALUE provides visitors with immersive and augmented visiting experiences. It guides them within the museum towards their selected artworks, enabling virtual interaction with observed elements and communication with a virtual guide through voice commands and body movements. For museum managers, VALUE offers analysis of visitor behavior for cultural and commercial proposals, and identification of the most frequented paths. Experiments show that the proposed approaches achieve satisfactory results for cultural heritage attended object detection and visitors localization.**

***Index Terms*—Wearable devices, egocentric vision, cultural heritage, mixed reality**

I. Introduction

Cultural sites attract numerous daily visitors, sparking ongoing interest in developing innovative technologies that support visitors during their experiences. Currently, most cultural sites offer support through conventional means such as information panels, maps, catalogs, and audio guides However, these methods have their limitations, highlighting the necessity for the development of intelligent systems that leverage computer vision and mixed reality to enhance the visitor experience. For example, an intelligent system could localize visitors and suggest a personalized path through the cultural site based on individual preferences. It could also recognize what the visitor is looking at (e.g., statues, artworks, etc.) and provide additional information via holograms. Furthermore, understanding visitors' behavior is crucial for site managers to assess the performance of the site and improve services for visitors. By analyzing patterns in visitor movement and interaction, managers can make data-driven decisions to optimize both the logistic aspects of the site and the overall visitor experience.

Moved by this motivation, we introduce VALUE, an integrated system that includes a variety of wearable devices to aid visitors at cultural sites, along with a backend for analyzing visitor data. In particular, VALUE implements various algorithms to localize visitors within the cultural site, recognize points of interest present in the environment, and detect which point of interest the visitor is looking at by exploiting multimodal signals (i.e., RGB, gaze) captured from the visitors' point of view. Additionally, we developed a virtual guide to enhance the visitor's experience, capable of interacting via voice commands and body gestures. The inferred information is then used to provide different services: 1) Localization, which informs the visitor of their exact location within the environment; 2) Attended Object Detection, which provides additional information about the observed point of interest through augmented reality; 3) Manager Control Panel, which shows managers the collected visitor data, emphasizing the areas visited, the visitors' movement patterns within the site, and the points of interest that garnered the most attention. The proposed system has been tested in Room V of the museum "Galleria Regionale di Palazzo Bellomo" which comprised 7 points of interest. Experiments show that the VALUE system obtains good performance on the tasks of 3DOF localization, Object Detection and Attended Object Detection.

Unlike previous approaches, VALUE unifies wearable technology and back-end analytics in a comprehensive system. This integration provides visitors with real-time, personalized insights into artworks and statues, while site managers receive data-driven information on visitor behavior.

II. Related Work

A. Wearable Assistants in Cultural Sites

Several works explored the use of wearable devices and mixed reality approaches to augment the experience of visitors in cultural sites [3, 22, 25]. Cucchiara and Del Bimbo [5] discuss the use of computer vision and wearable devices for augmented cultural experiences. Fonseca et al. [7] exploited wearable technologies to improve spatial perception and accessibility of the Casa Batlló Museum in Barcelona, enhancing understanding beyond traditional audio-guides. Becattini et al. [2] presented a mobile app that aims to enhance the experience in the museum reducing cognitive load and exploiting gamification. Seidenari et al. [20] introduced a CNN to perform localization and object recognition to develop a context-aware audio guide. Signorello et al. [21] proposed and evaluated a method, useful to explore tourists' flow in an area of interest and extract useful clues that help managers to conduct planning activities. The proposed VALUE system aims to support both visitors and cultural managers providing different services.

B. Visitor Localization

Outdoor localization can be effectively handled using Global Positioning System (GPS) devices. However, GPS is unsuitable for indoor environments, prompting the development of various Indoor Positioning Systems (IPS) over the years [6]. These systems rely on devices like active badges [26] and WiFi networks [8], which must be installed as part of the infrastructure. This installation is costly and not always feasible, particularly in the cultural heritage context. Kendall et al.[11] proposed to infer the 6 Degrees of Freedom pose of a camera from egocentric images using a CNN. In [17], the challenge of room-based visitor localization within cultural sites was tackled using egocentric images to provide site managers with behavioral information.

Unlike previous approaches, we aim to achieve punctual localization, positioning the user at a specific 3D coordinate in the environment or at a 2D location on a map. This will enable to offer visitors relevant, context-aware information.

C. Attended Object Detection in Cultural Sites

The attended object detection task involves detecting and recognizing the object observed by people([23, 27]). This task has been addressed from egocentric vision [15, 14] and also by considering gaze [24]. In particular, [24] emphasize the pivotal role of gaze in object recognition and its facilitation of interaction with augmented reality applications

The VALUE system utilizes eye tracking to understand which objects the visitor is looking at, providing additional information about specific objects or details of the observed scene.

III. Experimental Cultural Site and Datasets

VALUE has been tested in Room V of the Galleria Regionale di Palazzo Bellomo museum[1], located in Syracuse. In particular, we acquired and annotated different datasets of egocentric videos useful for designing 3DOF localization, object recognition, and attended object detection algorithms.

Data were collected using various wearable devices (Hololens 2, Vuzix M300 XL, Epson Moverio) and different smartphones (Huawei Mate 20, Motorola G6, Xiaomi 9 Lite, One Plus 8 Pro, Redmi 9).

A. Dataset for Localization

This dataset was acquired using two wearable glasses: Microsoft Hololens 2 and Vuzix Blade. We asked subjects to walk around Room V, covering all possible positions and head rotations at random. The dataset comprises 7 videos. The estimated focal distances are 1024.80 pixels for Hololens 2 and 768 pixels for Vuzix. The videos were recorded at resolutions of 2272×1278 at 30 fps with the Hololens 2 and 1920×1080 at 24 fps with the Vuzix. Frames were used to create a 3D model of Room V using Structure from Motion software (VisualSFM[2]). Frames that were not used by the software in creating the 3D model were discarded from the dataset. The final dataset comprises a total of 88,706 frames, each labeled with: 1) 2D and 3D positions, 2) orientation as Euler angles, quaternions, and rotation angles on the z-axis, 3) estimated focal distance, and 4) camera rotation matrix.

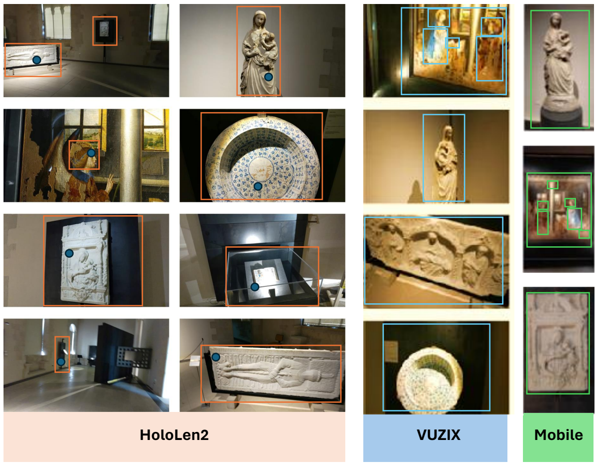


Fig. 1: Examples of labeled images belonging to the Room V captured with different devices. On the left, images acquired with HoloLens 2 labeled also with gaze point (blue). In the middle, images acquired with Vuzix, while on the right, images acquired with smartphone devices.

B. Dataset for Object Detection

The dataset was captured using different devices. In particular, we used a Microsoft Hololens2, a Vuzix Blade and several smartphones (i.e., Huawei Mate 20, Motorola G6, Xiaomi 9 Lite, One Plus 8 Pro and Redmi 9) to acquire 24 videos in Room V. We labeled this dataset using VGG Image Annotator tool (VIA)[3]. VIA enables the drawing of regions in both regular (e.g., rectangle, circle) and irregular (polygonal) shapes. The taxonomy of the elements to be labeled in Room V includes 7 points of interest (i.e., Statue, Slab, Tomb-Front, Tomb-Back, Plate, Book and Annunciation) and 8 details relating to the painting “L’Annunciazione” (i.e., Angel, Virgin, Body-Angel, Column, Detail window 1, Detail-ruined, Window-right and Lectern).

C. Dataset Acquired for Attended Object Detection

The annotated dataset for object detection acquired with Hololens 2, comprises the gaze signal obtained with the eye tracking. This signal is represented by (x, y) coordinates of the gaze. During the labeling process, these coordinates are classified based on whether the user was looking at an object—i.e., the coordinate falls inside a bounding box—or marked as

[1] https://www2.regione.sicilia.it/beniculturali/palazzobellomo/index.html

[2] http://ccwu.me/vsfm/

[3] https://www.robots.ox.ac.uk/~vgg/software/via/

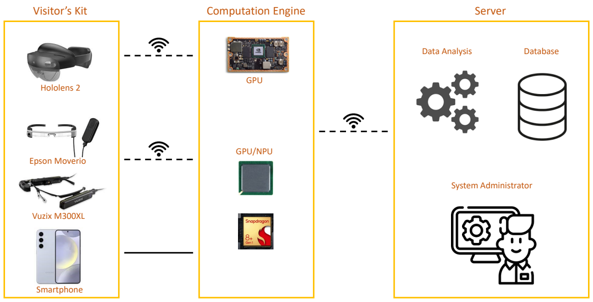


Fig. 2: The VALUE systems is composed of 3 main components: 1) Visitor's kit, 2) Computation Engine and 3) Server.

"Other" if the user's gaze did not fall on any of the objects, i.e., the gaze is outside any object of interest (see Figure 1).

D. Dataset Acquired for Verbal Interaction

In order to train a conversational agent capable to utilize a knowledge base comprising details about objects or their components to address visitor inquiries, we collected a storytelling dataset. The agent should be capable to understand the context of user questions and provide what it determines to be the most accurate response. A dataset composed by a training set and a test set was gathered: the first split contains 106 sentences spoken by a diverse group composed by 4 women and 4 men aged between 21 and 70, along with 5 children aged 8 to 13, and a test set consisting of 35 sentences spoken by 3 adults and 2 children.

IV. Architecture and Services

A. Architecture

Figure 2 illustrates the high level architecture of the proposed VALUE system which is composed of the following components:

- **Visitor's kit:** wearable devices such as Hololens 2, Vuzix M300XL, Epson Moverio, and smartphones are provided to the visitors of the cultural site. These devices serves to capture images and video from the user's perspective as well as to provide digital content through Augmented Reality.
- **Computation engine:** can be internal, utilizing NPUs in smartphones, or external, employing GPUs like Jetson TX2 connected through Wifi network.
- **Server:** for data storage, analysis, configuration, and user profile management. The server is dedicated to managing, processing, and storing data collected during cultural site visits.

B. Services

1) Localization and Attended Object Detection: The system performs 3D localization of visitors by processing the acquired egocentric video with a Triplet algorithm [10] to extract visual features, and then with a K-NN to assign the position value and orientation of the closest training image. The recognition of the points of interest observed by visitors is carried out using different algorithms depending on the used wearable device. Specifically, when using a Microsoft HoloLens 2, object detection is performed using Faster R-CNN [19] (implemented via the Detectron2 framework). Subsequently, we leverage gaze information obtained from eye tracking to filter the detected objects and select the one where the 2D gaze position lies within its bounding box. When using Vuzix, Moverio, or a smartphone, an Android native app performs object detection based on the YOLOv5 model[4]. YOLOv5 is an extension of YOLOv3 [18]. The app selects the observed point of interest by considering the object that is in front of the user.

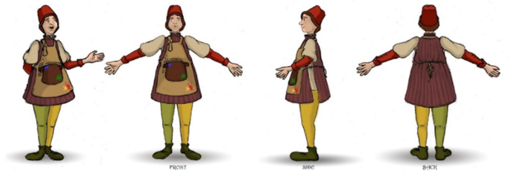


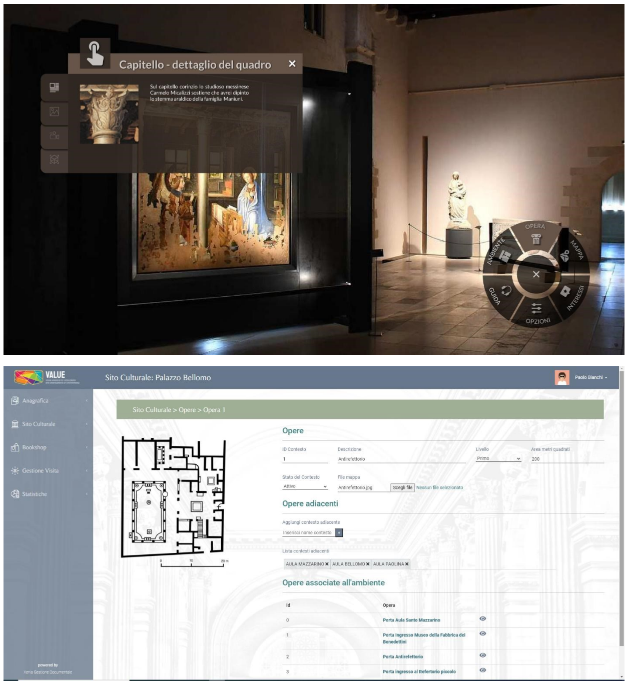
Fig. 3: Sketch of the conversational agent

2) Verbal Interaction: The user can interact with the avatar through verbal interaction, answering information requests on artwork and author, the historical context, and other related details included in system knowledge built by documents provided by the domain experts. Natural language is the most intuitive and efficient way for humans to communicate. Using verbal interaction allows clear and direct communication between the person and the avatar, reducing misunderstandings and enhancing the user experience. Moreover, it could create a sense of familiarity and comfort, which is essential for building rapport and trust between the user and the avatar [16]. Verbal interaction based on natural language understanding [12] requires a robust speech recognition module and a conversational agent capable of managing the evolution of speech between the user and the avatar [1]. The conversational agent (chatbot) considers the context (domain), the dialogue examples given in the learning phase (by rules and stories), and previously given answers by detecting and recognizing the "intent" of the user. The sentence with the highest confidence rate is the chatbot's output. Finally, the output textual answer could generate the audio file for a standard speech synthesizer in the user language. The agent, which appears in the scene as an AR model, was sketched to be alike the painter Antonello da Messina (see Figure 3).

3) Augmented Reality and Visitor Interaction: The system offers visitors an intuitive and seamless augmented reality interface.

- HoloLens 2 has a fully augmented reality interface, enabling users to interact with menus and artistic-cultural elements through gestures, gaze, and voice commands. Visitors can engage in verbal interaction with a conversational agent embodied as the 3D model of the

[4] https://github.com/ultralytics/yolov5



Sul capitello corinzio lo studioso messinese Carmelo Micalizzi sostiene che avrei dipinto lo stemma araldico della famiglia Maniuni.

Sito Culturale: Palazzo Bellomo

Sito Culturale > Opere > Opera 1

Opere

ID Contesto: 1
Descrizione: Antirefettorio
Livello: Primo
Area metri quadrati: 200
Stato del Contesto: Attivo
File mappa: Antirefettorio.jpg Scegli file Nessun file selezionato

Opere adiacenti

Aggiungi contesto adiacente: Inserisci nome contesto +
Lista contesti adiacenti: AULA MAZZARINO ✕ AULA BELLOMO ✕ AULA PAOLINA ✕

Opere associate all'ambiente

Id	Opera
0	Porta Aula Santo Mazzarino
1	Porta Ingresso Museo della Fabbrica dei Benedettini
2	Porta Antirefettorio
3	Porta ingresso al Refertorio piccolo

Fig. 4: Top: Augmented Reality GUI sample. Bottom: Web portal for the cultural manager.

sicilian painter Antonello da Messina (Figure 3). This agent provides contextually relevant responses, guides discussions, and offers thematic insights.

- Vuzix M300XL and Epson Moverio have the interface showcased on the device's display, allowing users to view augmented and non-augmented reality elements. Visitors can interact with these elements using dedicated buttons, the integrated touch-pad, or voice commands.
- Mobile Devices offer augmented reality elements overlaid on real images when pointed at museum scenes. Users can interact with the interface or 3D elements by tapping the display or using voice commands.

Throughout the visit, visitors have the option to access text content, view 3D models related to environments or artworks, and navigate the map (see Figure 4-top).

4) Data Visualization: A web portal complements back-end services, offering different functionalities based on user profiles for configuration and data interaction. Users can access the portal through a browser and execute input commands to interact with the system. The profiles considered are: Administrator, Manager, Operator, Data entry, Shop user. Depending on profiles, some operations are permitted in order to add, edit or delete content or to visualize statistics on visiting patterns (see Figure 4-bottom).

V. Experimental Results

We tested our system to assess the performances of localization and attended point of interest recognition services, which are at the core of VALUE system.

A. Localization

We addressed both 3DOF and 6DOF localization task using three different algorithms: Posenet [11], DSAC* [4], and Triplet [9]. To evaluate performance, we computed the average median position error, average rotational axis error in degrees, average quaternion rotational error in angles, and average error for each Euler angle.

Table I shows the results using the dataset acquired with Hololens2. Among the algorithms tested, Triplet exhibited the most effective performances achieving a median angular error of 37 cm and 4.45 degrees. Instead, Posenet excelled in the 6 degrees of freedom localization task. Given its superior performance, we chose Triplet for the final system.

B. Object Detection and Attended Object Detection

Firstly, we considered the standard object detection task. We tested the object detector Faster-RCNN [19] in the considered Room V of the cultural museum. Given that points of interest encompass both individual artworks and specific details within the "Annunciazione" painting, we employed two distinct strategies:

- Standard Object Detector (OD): We used a standard object detector to recognize various points of interest;
- Cascade Object Detector (COD): which consists of two stages: 1) the primary detector identifies different points of interest, while, the secondary detector focuses exclusively on details related to the "Annunciazione" point of interest.

In general, the algorithms take as input the frame captured by the camera and outputs the bounding box and the corresponding class for each object recognized in the scene. Specifically, the returned information is of the type: (x, y, w, h, c), where x, y are the top-left coordinates, w, h represent the width and height of the bounding box around the object, and c represents the class of the recognized object. We evaluated these approaches using standard COCO mean Average Precision (mAP) [13].

Table II shows the obtained results. Detecting a small object contained within a larger one is an extremely challenging task for an object detector (column 3). Notably, the table shows substantial improvements from OD to COD (column 4), indicating that the second layer of the cascade detector, specifically designed for particular of "Annuciazione" (class 7-14), is capable of overcoming the challenges of recognizing smaller objects within larger contexts.

Secondly, we addressed the task of detecting and recognizing the observed point of interest. The algorithm takes as input a frame and the corresponding gaze signal, acquired through the Hololens2 device, and returns the bounding box of the object being looked at, if present. If the user is not looking at any of the considered objects, then the class "Other" will be returned. The model, trained to solve the object detection task, outputs all the bounding boxes related to the objects present in the scene and their respective classes. Once the list of bounding boxes is obtained, it is possible to cross-reference

TABLE I: Results obtained with the 3 proposed algorithms for the task of punctual localization on Hololens data

Measurement (Error)	Posenet (2017)		DSAC*		Triplet + 1NN	
	Average	Median	Average	Median	Average	Median
6 Degrees of Freedom						
Position (meters)	0.86	0.52	7.75	6.28	1.24	0.73
Quaternion (degrees)	62.17	52.1	131.3	143.2	60.26	55.47
Angle X (degrees)	64.26	14.37	108.62	87.14	59.66	7.76
Angle Y (degrees)	43.09	35.7	39.98	40.16	43.95	40.96
Angle Z (degrees)	63.39	9.4	82.68	77.3	58.88	4.45
3 Degrees of Freedom						
Position (meters)	0.89	0.54	6.61	5.28	0.70	0.37
Angle (degrees)	63.39	9.4	82.68	77.3	58.88	4.45

TABLE II: Results obtained for the task of object detection (OD and COD) and Attended object detection (ATT).

ID	Object/Detail	OD	COD	ATT
0	Madonna del Cardillo	60.66	60.66	58.20
1	Tabernacolo con la Madonna col Bambino	67.71	67.71	66.57
2	Lastra tombale di Giovanni Cabastida r	68.02	68.02	67.95
3	Lastra tombale di Giovanni Cabastida f	68.57	68.57	67.64
4	Piatto fondo	72.25	72.25	72.48
5	Annunciazione	75.57	75.57	78.52
6	Libro d'Ore miniato	78.76	74.34	75.75
7	Dettaglio Arcangelo Gabriele superiore	8.96	91.24	92.78
8	Dettaglio Vergine	5.98	92.13	96.03
9	Dettaglio Arcangelo Gabriele parte inferiore	25.93	94.63	97.93
10	Dettaglio Capitello	2.60	83.64	86.34
11	Dettaglio letto	8.77	89.45	90.75
12	Devoto in basso a dx	13.92	87.43	90.94
13	Dettaglio finestra centrale	6.09	93.32	97.67
14	Dettaglio Sacre Scritture	5.94	69.89	70.39

this data with the gaze data. It is possible to select the object being looked at by the user by selecting the box within which the gaze coordinate falls. Table II - column 5 shows mAP for all the point of interest considered. We also reported qualitative results of our algorithms in Figure 5.

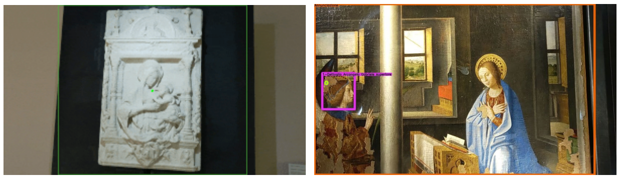


Fig. 5: Attended object detection qualitative results. In green the visitor gaze.

VI. Conclusion

We have presented VALUE, an integrated system to enhance the visitor experience in cultural sites. By exploiting computer vision algorithms and mixed reality, VALUE aims to create a more interactive and informative experience. Simultaneously, it assists site managers by automatically inferring behavioral insights related to the cultural sites. Our experiments, conducted in a real cultural site, underscore the system's robust performance in fundamental tasks such as localizing visitors and recognizing observed points of interest.

Acknowledgment

This research has been supported by the project VALUE (N. 08CT6209090207 - CUP G69J18001060007) - PO FESR 2014/2020 - Azione 1.1.5.

References

[1] Merav Allouch, Amos Azaria, and Rina Azoulay. "Conversational agents: Goals, technologies, vision and challenges". In: *Sensors* 21.24 (2021), p. 8448.

[2] Federico Becattini et al. "Imaging novecento. a mobile app for automatic recognition of artworks and transfer of artistic styles". In: *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection: 6th International Conference, EuroMed 2016, Nicosia, Cyprus, October 31–November 5, 2016, Proceedings, Part I 6*. Springer. 2016, pp. 781–791.

[3] Silvia Blanco-Pons et al. "Design and implementation of an augmented reality application for rock art visualization in Cova dels Cavalls (Spain)". In: *Journal of Cultural Heritage* 39 (2019), pp. 177–185.

[4] Eric Brachmann and Carsten Rother. "Visual Camera Re-Localization from RGB and RGB-D Images Using DSAC". In: *TPAMI* (2021).

[5] Rita Cucchiara and Alberto Del Bimbo. "Visions for augmented cultural heritage experience". In: *IEEE MultiMedia* 21.1 (2014), pp. 74–82.

[6] Kevin Curran et al. "An evaluation of indoor location determination technologies". In: *Journal of Location Based Services* 5.2 (2011), pp. 61–78.

[7] David Fonseca et al. "Assessment of wearable virtual reality technology for visiting World Heritage buildings: an educational approach". In: *Journal of educational computing research* 56.6 (2018), pp. 940–973.

[8] Yanying Gu, Anthony Lo, and Ignas Niemegeers. "A survey of indoor positioning systems for wireless personal networks". In: *IEEE Communications surveys & tutorials* 11.1 (2009), pp. 13–32.

[9] Elad Hoffer and Nir Ailon. *Deep metric learning using Triplet network*. 2018. arXiv: 1412.6622 [cs.LG].

[10] Elad Hoffer and Nir Ailon. “Deep metric learning using triplet network”. In: *Similarity-Based Pattern Recognition: Third International Workshop, SIMBAD 2015, Copenhagen, Denmark, October 12-14, 2015. Proceedings 3*. Springer. 2015, pp. 84–92.

[11] Alex Kendall, Matthew Grimes, and Roberto Cipolla. “Posenet: A convolutional network for real-time 6-dof camera relocalization”. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 2938–2946.

[12] Diksha Khurana et al. “Natural language processing: State of the art, current trends and challenges”. In: *Multimedia tools and applications* 82.3 (2023), pp. 3713–3744.

[13] Tsung-Yi Lin et al. “Microsoft COCO: Common Objects in Context”. In: *Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V*. Vol. 8693. Lecture Notes in Computer Science. Springer, 2014, pp. 740–755.

[14] Michele Mazzamuto et al. “Weakly supervised attended object detection using gaze data as annotations”. In: *International Conference on Image Analysis and Processing*. Springer. 2022, pp. 263–274.

[15] Michele Mazzamuto* et al. “Learning to Detect Attended Objects in Cultural Sites with Gaze Signals and Weak Object Supervision”. In: *J. Comput. Cult. Herit.* 17.3 (Apr. 2024). ISSN: 1556-4673. DOI: 10.1145/3647999. URL: https://doi.org/10.1145/3647999.

[16] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. “Computers are social actors”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1994, pp. 72–78.

[17] Francesco Ragusa et al. “Egocentric Point of Interest Recognition in Cultural Sites.” In: *VISIGRAPP (5: VISAPP)*. 2019, pp. 381–392.

[18] Joseph Redmon and Ali Farhadi. “Yolov3: An incremental improvement”. In: *arXiv preprint arXiv:1804.02767* (2018).

[19] Shaoqing Ren et al. *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*. 2016. arXiv: 1506.01497 [cs.CV].

[20] Lorenzo Seidenari et al. “Deep artwork detection and retrieval for automatic context-aware audio guides”. In: *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 13.3s (2017), pp. 1–21.

[21] Giovanni Signorello et al. “Exploring geo-tagged photos to assess spatial patterns of visitors in protected areas: the case of park of Etna (Italy)”. In: (2018).

[22] Manuel Silva and Luis Teixeira. “Developing an extended reality platform for immersive and interactive experiences for cultural heritage: Serralves museum and coa archeologic park”. In: *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE. 2020, pp. 300–302.

[23] Francesco Tonini et al. “Object-aware Gaze Target Detection”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 21860–21869.

[24] Takumi Toyama et al. “Gaze guided object recognition using a head-mounted eye tracker”. In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. 2012, pp. 91–98.

[25] Mariapina Trunfio, Salvatore Campana, and Adele Magnelli. “Measuring the impact of functional and experiential mixed reality elements on a museum visit”. In: *Current Issues in Tourism* 23.16 (2020), pp. 1990–2008.

[26] Roy Want and Andy Hopper. “Active badges and personal interactive computing objects”. In: *IEEE Transactions on Consumer Electronics* 38.1 (1992), pp. 10–20.

[27] Daniel Weber et al. “Gaze-based object detection in the wild”. In: *2022 Sixth IEEE International Conference on Robotic Computing (IRC)*. IEEE. 2022, pp. 62–66.