Ontology-Based Modeling for Object Segmentation and Eye Gaze Data in VR Art Exhibitions

Delaram Javdani Rikhtehgar[1,*], Batuhan Usta[1] and Shenghui Wang[1]

[1]University of Twente, Enschede, The Netherlands

Abstract

Virtual reality (VR) art exhibitions provide an immersive platform for experiencing art, yet understanding user interaction and perception within these environments remains a challenge. Traditional VR approaches often fail to offer personalized insights into how users engage with specific elements of an artwork. Object detection and segmentation techniques hold potential by enabling the identification and localization of objects within paintings, thereby adding semantic meaning to user eye-gaze data. In this work, we propose an ontology-based modeling framework that links segmented objects in artworks with user gaze data. This framework allows for the integration of numerical data, such as image segments, and measurement data, including user engagement metrics (e.g., eye-gaze patterns), time series (tracking eye movements across exhibits), and qualitative assessments of user experience. By semantically enriching gaze data through associations with specific objects and concepts within the artwork, we are able to generate insights into gaze-based behaviors, such as individual users' visual attention and navigation. These insights enable the exploration of higher-order interpretations, such as visitor behavior and engagement trends, which are critical for improving user experience and optimising exhibition design. Our study demonstrates how this ontology was populated with real-world data from a user study and present further analysis based on recorded numerical data from the virtual environment. Finally, we discuss our findings, limitations, and potential directions for future work.

Keywords

Ontology-based modeling, Eye Tracking, Object Detection, Segmentation, Virtual Reality, Cultural Heritage

1. Introduction

Virtual Reality (VR) is increasingly used in museums thanks to its potential to enhance visitor engagement and educational experiences [1, 2, 3, 4, 5]. These environments often feature virtual agents that guide visitors and provide personalized information about artworks, improving both engagement and attention [6, 7, 8]. VR also addresses geographical limitations by allowing remote access to exhibitions, making art more accessible to a global audience [5]. However, while VR museums can showcase multiple artworks simultaneously, they face significant challenges in efficiently managing and storing large volumes of data related to artworks and visitor interactions.

Knowledge graphs [9] offer a feasible solution for managing this data by storing extensive information about artworks, including artist details, descriptions, and object locations. This structured information is crucial for identifying objects of interest and enhancing user engagement [6, 7, 10]. Despite these advantages, a major challenge remains: determining which specific objects within an artwork capture visitors' attention [11]. Eye-gaze data provides a potential solution, as it reveals where visitors focus their attention, offering insights into their engagement and behavior [12, 13, 14]. Integrating eye-gaze data into knowledge graphs, alongside object location data, allows for precise tracking of user interactions with specific elements within artworks.

However, manual annotation of objects within artworks is time-consuming and inefficient, especially in dynamic virtual environments with frequent content updates. Deep learning (DL) models can automate this process by identifying objects and generating bounding box coordinates [15, 16, 17, 18, 19,

EKAW 2024: EKAW 2024 Workshops, Tutorials, Posters and Demos, 24th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2024), November 26-28, 2024, Amsterdam, The Netherlands.

*Corresponding author.

d.javdanirikhtehgar@utwente.nl (D. Javdani Rikhtehgar); b.usta@student.utwente.nl (B. Usta); shenghui.wang@utwente.nl (S. Wang)

0009-0007-8266-8395 (D. Javdani Rikhtehgar); 0000-0003-0583-6969 (S. Wang)

20]. While bounding boxes provide a general localization of objects, they often lack the precision required to handle overlapping or intricate elements within a painting. Segmentation models, which offer pixel-level accuracy, address these limitations by providing more precise object boundaries [21, 22, 23, 24, 25]. However, segmentation models frequently struggle with providing meaningful class labels for the segmented objects [26]. To address this gap, an auto-annotation system combining both object detection and segmentation is required, delivering both class labels and accurate object boundaries.

In this paper, we propose an ontology-based modeling framework to link segmented objects within paintings to class labels and user gaze data, enriching the semantic context of the visitor's gaze. This approach enables the standardization of key components, such as objects, actions, and user preferences, facilitating interoperability across systems. Additionally, ontology-based models allow for semantic reasoning and automated inference, generating insights into user behavior and optimizing content delivery in personalized experiences [27, 28, 29]. Our work addresses the following research questions:

1. How can segmented objects within artworks and eye-gaze data as a time series be formally represented in a knowledge graph to ensure semantic accuracy and interoperability?
2. What new behavioral insights can be derived when segmented object data and eye-gaze patterns are semantically enriched and integrated within the ontology?
3. How can gaze patterns and object interaction data be efficiently stored, managed, and queried to facilitate the extraction of high-level behavioral trends and visitor engagement metrics?

2. Related Work

Our research integrates multiple key areas, including object detection, segmentation, eye-gaze analysis, and ontology-based systems, to improve user interaction in Virtual Reality (VR) museum environments.

Object Detection and Segmentation Models Object detection and segmentation are key techniques in computer vision, often used together to enhance the accuracy of object localization and boundary detection within images. Faster R-CNN and YOLO are widely adopted object detection models for their efficiency in generating bounding boxes around objects within images [20, 19]. Faster R-CNN offers high accuracy through a two-stage process, while YOLO is faster and better suited for real-time applications, though with slightly reduced accuracy [30, 19].

However, while object detection offers quick localization, it lacks the pixel-level precision needed for complex scenes like artworks. Segmentation models such as Mask Region-Based Convolutional Neural Network (Mask R-CNN) [21], Panoptic Segmentation Model [22], You Only Look Once Version 8 (YOLOv8) Instance Segmentation [23], Fast Segment Anything Model (SAM) [24], and Unified Panoptic Segmentation Network (UPSNet) [25] address this issue by generating pixel-level segmentation masks. While SAM offers significant capabilities with minimal manual annotation, its inability to provide semantic labels limits its use in tasks requiring both localization and classification. To overcome this limitation, we combine the speed of YOLO for fast object detection with the precision of SAM for segmentation, ensuring that user gaze interactions are accurately linked to the objects they observe. Additionally, we use an ontology-based system to assign semantic meaning to the segmented objects, enabling richer and more interactive user experiences.

Eye-Gaze Data Interpretation Eyes can provide more information than just a single focal point. Research shows that eye movement patterns can be further analyzed to reveal the intentions behind tasks in various applications [31, 32]. Eye movement patterns are difficult to extract. Eye-gaze data alone does not provide sufficient information, as the gaze only gives coordinates. To make meaning out of these gazes, semantic context is necessary. Semantic meaning can be provided by utilizing the deep learning models as they provide identification and location to desired sections. Koochaki and Najafizadeh proposed a method to interpret these gazes by incorporating semantic context to make meaningful conclusions [32]. However, few approaches have combined these insights with structured

ontologies. Our work builds on this by using ontologies as a semantic framework to interpret real-time eye-gaze data, enabling dynamic, personalized content adaptation in virtual museums.

Ontology-Based Systems in Human-Computer Interaction In recent years, ontologies have become vital tools in enhancing human-computer interaction (HCI) by providing structured, semantic representations of knowledge. Chang et al.[33] developed an ontology-based knowledge model for an integrated robot framework, enabling interactive human-robot services. Their model incorporates both general and domain-specific knowledge, structured into five key ontologies: user, robot, perception, environment, and action. This framework was tested in a social robot service, demonstrating its usability and effectiveness. Similarly, Costa et al.[34] introduced HCIO (Human-Computer Interaction Ontology), rooted in the Unified Foundational Ontology (UFO), to address semantic interoperability in HCI. HCIO is composed of three sub-ontologies—Interactive Computer System, User, and Human-Computer Interaction—and serves as a reference model for communication and learning in the HCI domain. Castro et al.[35] extended this work by introducing HCIDO (Human-Computer Interaction Design Ontology), part of the broader HCI-ON (Human-Computer Interaction Ontology Network). HCIDO supports the development of KTID (Knowledge Supporting Tool for HCI Design), facilitating knowledge sharing, especially among novice designers. Freitas et al.[36] explored ontologies to support Adaptive User Interface (AUI) systems, enabling real-time interface adjustments tailored to users with low vision or color blindness. These examples highlight the growing role of ontologies in structuring knowledge for more adaptive and inclusive HCI systems. However, previous work in HCI ontologies does not fully address the needs of dynamic user interaction with visual elements in VR environments.

Multimodal Technologies in Virtual Museums Within the domain of Virtual Heritage, multimodal technologies have proven instrumental in creating immersive and engaging experiences, particularly in virtual museums [37, 38]. These technologies often incorporate conversational agents that act as guides, enhancing information accessibility and user engagement [39, 40, 41]. Head-mounted displays (HMDs) further enrich these experiences by integrating various input sensors, such as those tracking hand movements, speech, and eye and head motions, which collectively enhance the sense of immersion [42]. Among these, eye-tracking technology has emerged as a powerful tool in virtual reality (VR), providing rich insights into user behavior and interaction. Eye-tracking enables the customization of content based on users' gaze patterns, offering a more personalized and adaptive experience [43, 44]. Beyond interaction, eye-tracking also facilitates emotional and learning assessments, adding further dimensions to user experience analysis environments [45, 46].

Although multimodal technologies have significantly enhanced user immersion in virtual museums, there remains a lack of systems that integrate gaze data with semantic representation to provide context-aware, personalized content. Our framework bridges this gap by combining eye-tracking data with an ontology-based system that enriches interactions with semantic meaning.

In summary, while significant progress has been made in object detection, segmentation, and the use of ontologies in HCI, there remains a gap in integrating these techniques to enhance user interactions in VR environments. Our research addresses this gap by combining deep learning for object detection and segmentation with ontology-based semantic representations, all informed by real-time eye-gaze data, to create personalized, adaptive experiences in virtual museum settings.

3. Ontology-based Data Modeling for User Interaction in VR

Motivational context As virtual reality (VR) technology evolves, it creates new opportunities for users to interact with digital environments in personalized and engaging ways. Our VR art exhibition, as shown in Figure 3, featuring 19 paintings, aims to model visitor interactions by integrating pre-processed numerical data from object detection and segmentation with real-time eye-gaze tracking data. The key to achieving this dynamic interaction lies in an ontology-based framework that systematically organizes and models both types of numerical data, allowing for personalized and adaptive user experiences.

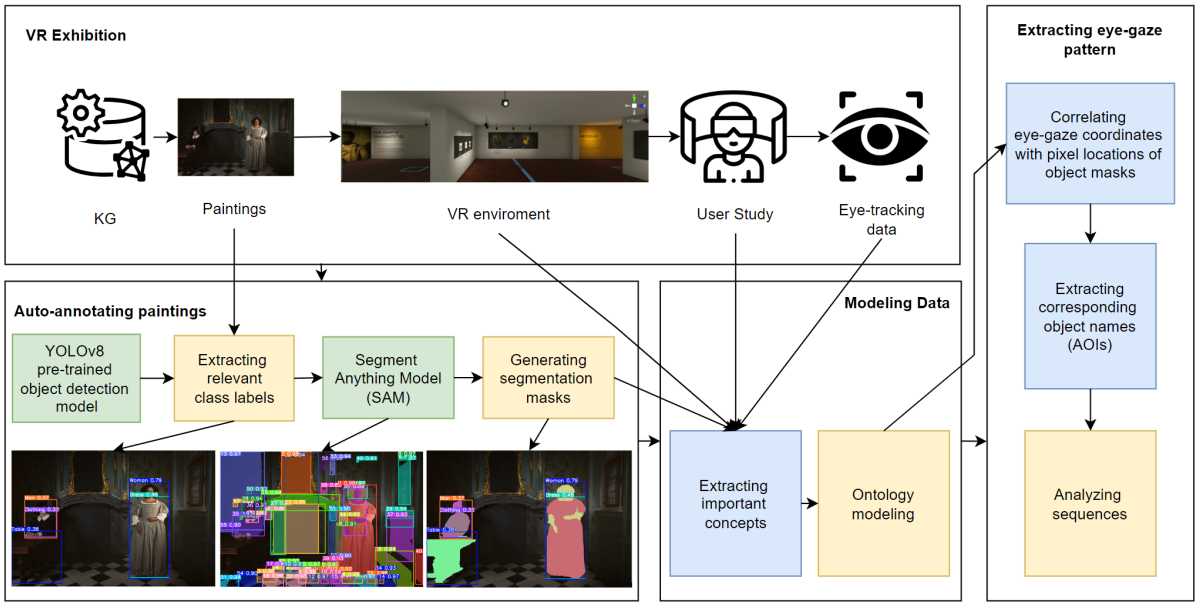


Figure 1: Overview of methodology.

Before a visitor enters the VR environment, each painting undergoes object detection and segmentation. This process identifies key elements, such as figures and objects, and generates numerical data in the form of segmentation masks and spatial coordinates. This pre-processed data is essential for understanding the structure of the artwork and is stored in an ontology to represent the spatial relationships and semantic meaning of each object.

When a visitor enters the VR environment, their eye-gaze data—captured in real time as spatial coordinates—is linked to this pre-processed object data via the ontology. The ontological model plays a critical role in mapping real-time gaze points to specific objects within the paintings by using the pre-stored segmentation data. This allows the system to interpret where the visitor is focusing and deliver tailored information about the object, such as its historical significance or artistic relevance.

By organizing both the pre-processed numerical data (segmentation masks, object coordinates) and the real-time gaze data within the ontology, the system gains the ability to semantically interpret and respond to user behavior. The ontology provides a framework to represent not only static information about the artworks but also dynamic user interactions. This enables the conversational agent to adapt to each visitor's unique engagement patterns, responding to prolonged gazes with more detailed information or suggesting related artworks.

The ontological modeling of these two types of numerical data—pre-processed spatial data from object detection and real-time eye-gaze coordinates—creates a seamless bridge between passive observation and active engagement. The system not only tracks user behavior but also uses it to generate new insights, such as identifying which parts of the exhibition are most engaging or how users interact with different types of artwork. By structuring this information within an ontology, we ensure that the system can manage and query both pre-processed and real-time data efficiently, providing a flexible and scalable solution for modeling user interactions in VR environments.

3.1. Obtaining Segmentation Data (Pre-Processed Data)

To analyze participants' gaze patterns and understand which objects they are focusing on during the VR exhibition, we first needed to obtain detailed numerical segmentation data for the objects within the paintings. This was achieved through a combination of object detection and segmentation models.

We employed YOLOv8,[1] a state-of-the-art object detection model, which can recognize up to 601 class

[1]https://docs.ultralytics.com/models/yolov8/#key-features

labels. YOLOv8 was used to detect prominent objects in the paintings, such as people, buildings, and other significant elements. Each detected object was assigned a class label (e.g., "Person", "Building"), providing semantic meaning to the identified elements within the artwork.

Once the objects were detected, we applied the Segment Anything Model (SAM) from the Ultralytics library[2] to achieve pixel-level segmentation. SAM generated segmentation masks, which define the precise location of each detected object. These segmentation masks were normalized to a coordinate system where all points are represented within the range $[0.00, 0.00]$ (top-left corner) to $[1.00, 1.00]$ (bottom-right corner). This normalization ensures consistency in how objects are represented across different paintings of varying sizes.

Once the objects were detected and segmented, we incorporated the data into our ontology following these specific requirements:

1. Detected Objects as Areas of Interest (AOIs):
    - Each object detected in the painting (e.g., a person or building) must be represented as an Area of Interest (AOI) within the ontology.
    - The AOI is semantically linked to a class label (e.g., "Person", "Building"), which defines the type of object being observed.
2. Segmentation Masks as lists of coordinates:
    - Each AOI must be associated with a segmentation mask that defines the pixel-level location of the object, along with the painting in which the object is located.
    - These segmentation masks are stored in the ontology as a list of normalised coordinates $[[x_1, y_1], [x_2, y_2], \ldots, [x_n, y_n]]$, ensuring the precise location of the AOI can be used to match real-time gaze data during the exhibition.

3.2. Collecting real-time eye gaze data stream

Once the objects in the paintings were represented through normalized segmentation data, the next step involved integrating real-time eye-gaze data captured during the VR exhibition. This process required storing and representing the gaze data in the ontology in a manner that enables dynamic interaction with the previously identified Areas of Interest (AOIs).

In the VR environment, gaze points are initially captured as three-dimensional coordinates $[x_i, y_i, z_i]$, which represent the user's view in the VR space. Each painting is also represented by a 3D array that defines its position within this virtual environment. When the user focuses on a specific painting, the gaze data must be converted from 3D coordinates to 2D coordinates $[x_j, y_j]$, which correspond to the surface of the painting. This transformation allows the gaze data to align with the 2D segmentation masks previously generated for the painting's AOIs. Like the segmentation data, these gaze points are normalized between $[0.00, 0.00]$ (top-left corner) to $[1.00, 1.00]$ (bottom-right corner). This normalization ensures that the gaze data can be directly compared with the normalized segmentation masks of the AOIs, regardless of the painting's original dimensions.

As users explore the virtual environment, their gaze points are continuously recorded. To effectively represent the real-time eye gaze data stream, the following requirements are implemented:

1. Tracking Gaze Interactions:
    - Gaze points are stored in the ontology as part of an Observing action. This action includes the normalized 2D gaze coordinates, the painting being observed, the actor performing the action, and the event within which the action occurs.
    - If a gaze point falls within the coordinates of an Area of Interest (AOI) associated with the painting, the system logs that the user is observing that specific object (e.g., a person or building). This relationship is recorded in the ontology, linking the user's gaze data to the corresponding AOI. This allows for precise tracking of what the user is focusing on at any given moment.

[2] https://docs.ultralytics.com/models/sam/#key-features-of-the-segment-anything-model-sam

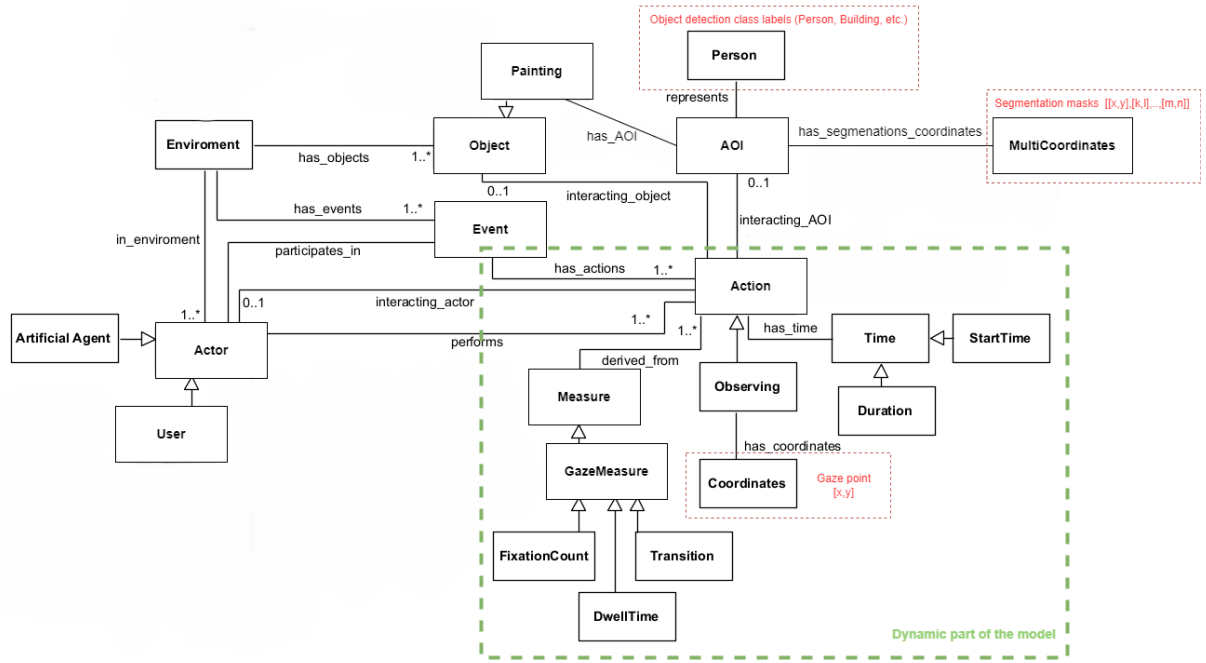


Figure 2: Sub-ontology: Eye-gaze based user behaviour modeling

2. Capturing Temporal Data for Gaze Analysis:
    - The ontology also captures temporal data related to the user's gaze interactions, including the start time of the gaze on each AOI and the duration of the observation.
    - This temporal information is crucial for analyzing gaze patterns, such as the sequence in which objects are viewed or identifying areas where the user's attention is sustained for longer periods. Such insights provide a deeper understanding of user engagement with specific elements in the artwork.

By meeting these ontological requirements, the system ensures that real-time gaze data can be seamlessly integrated with pre-processed segmentation data, enabling dynamic, personalized interactions throughout the VR experience.

3.3. Ontology Components

The ontology models both static elements of the virtual environment (such as paintings and their attributes) and dynamic data related to user interactions (such as gaze patterns and actions). Figure 2 illustrates a sub-ontology designed to represent these relationships. Below, we outline the core elements of the ontology and their roles in eye-gaze-based user behavior modeling.

3.3.1. Static Data Components

The **Environment** segment defines the virtual space, encompassing different VR configurations such as multiple rooms (e.g., three distinct exhibition rooms). It also stores domain context, such as cultural heritage, and high-level contextual information about the exhibition (e.g., an exhibition titled "HERE: Black in Rembrandt's Time"). These environmental attributes are essential for providing context to user interactions and spatial navigation.

The **Object** segment provides detailed information about virtual elements within the environment, primarily focused on the **Paintings** displayed in the exhibition. Each painting is associated with metadata, including the name, description, and narrative related to the artwork. Additionally, specific **Areas of Interest (AOI)** within the paintings are defined, accompanied by sets of coordinates and

relevant semantics. These AOIs can represent particular elements, such as person or buildings. Each AOI is linked to:

- Segmentation masks, which are represented as **MultiCoordinates**, which store the normalized coordinates of the AOI.
- Class labels that provide semantic meaning to the AOIs (e.g., "Person," "Building"). These labels facilitate more meaningful user interaction and content retrieval.
- Cultural knowledge graphs that link certain elements (such as historically significant figures or locations) to broader datasets, enhancing the richness of the interaction.

Efficient Representation of Segmentation Masks In the ontology, segmentation masks are stored as a list of normalized coordinates, which is represented as a single array rather than individual nodes for each coordinate. This compact representation improves:

- Efficiency: Reduces the overhead of managing numerous individual nodes, simplifying data manipulation.
- Query Performance: Makes it easier to perform comparisons and queries involving segmentation masks, especially when matching real-time gaze points with AOIs.
- Semantic Clarity: Storing segmentation masks as arrays aligns intuitively with the way they define spatial areas, making the model easier to understand and navigate.

This design also facilitates easier future adjustments or dynamic modifications without the need to restructure the entire ontology, ensuring the model is flexible and scalable.

3.3.2. Dynamic Data Components

While the **Environment** and **Object** segments are static, representing predefined data about the VR space and paintings, the ontology must also accommodate dynamic data that evolves based on user interactions during the exhibition.

The **Actor** component encapsulates information about users navigating the virtual environment, participating in events, and performing actions. It stores user information, such as their familiarity with the VR environment, knowledge about the exhibition, and demographic details (e.g., age, culture, education, gender, and language). This information is critical for analyzing user engagement and personalizing interactions.

The **Action** concept captures the different actions users perform during the exhibition, including gaze-related **Observing**, or conversation-related actions such as, asking a question, answering a question, commenting, or providing feedback.

To fully analyze gaze patterns and user engagement, the ontology tracks temporal data for each interaction, including **StartTime** and **Duration**. For gaze-related **Observing** action, they capture when the user begins interacting with an AOI and the length of time the user's gaze remains on an AOI, allowing for in-depth analysis of attention patterns, such as prolonged focus on particular areas or sequences of viewed objects.

3.4. Populating and Querying the Ontology for Eye-Gaze Analysis

The following sections illustrate how to populate and query the ontology to connect gaze data with AOIs and extract meaningful insights.

3.4.1. Populating the Ontology with Gaze Data and AOIs

Listing 1 provides a SPARQL query for inserting a new observation (gaze point) into the ontology. The query links the user's gaze point to the corresponding AOI by comparing the normalized gaze coordinates $[X, Y]$ with the MultiCoordinates of the AOI.

Listing 1: Adding a new gaze point (X,Y) and connecting it to the corresponding AOI

```
INSERT {
    # Create a new action instance labeled as 'action_1'
    :action_1 a :Action_Observing ;
        :interacting_object ?object ; # Specify interacting object
        :interacting_aoi ?aoi ; # Specify interacting AOI
        :has_coordinates ?coordXY. # Specify gaze point

    # Connect the actor performing the new action
    ?actor :performs :action_1 .

    # Specify the object as a 'Painting' with AOI ?aoi
    ?object a :Painting ;
            :has_AOI ?aoi .
}
WHERE {
    # Find the AOI related to the coordinates [X,Y]
    ?aoi :has_segmentation_coordinates ?MultiCoordinates .

    BIND((x, y) AS ?coordXY)
    FILTER(contains(?MultiCoordinates, ?coordXY))

    # Find the actor with name 'actor_1'
    ?actor a :Actor ;
           :actor_name "actor_1" .
}
```

3.4.2. Analyzing Gaze Patterns through Temporal Data

Inferences about user behavior, such as decision-making processes and attentional patterns, can be drawn from gaze-based metrics like Dwell Time, Fixation Count, and Transitions between AOIs [47]. Listing 2 shows how to calculate Dwell Time, which represents the sum of gaze fixation durations on a specific AOI.

Listing 2: Calculating Dwell Time

```
SELECT (SUM(?durationValue) AS ?DwellTime)
WHERE {
  ?actor :performs ?action . # An actor performs an action.
  ?action a :Observing .
  ?action :interacting_aoi ?aoi . # The action interacts with an AOI.
  ?aoi a :AOI ; # The AOI is of type AOI and its name is "aoi_1".
       :aoi_name "aoi_1" .
  ?action :has_time ?duration . # The action has a duration.
  ?duration :Duration ; # The duration has a value.
            :value ?durationValue .
}
```

3.4.3. Comparing Actor Behavior Across AOIs

The ontology allows for the comparison of multiple actors' behavior by analyzing how often they interact with different AOIs. Listing 3 presents a SPARQL query that calculates how frequently each actor visits an AOI and compares these observation frequencies across actors.

Listing 3: Comparative analysis of AOI observation frequencies

```
SELECT ?actor ?aoi (COUNT(?action) AS ?observationCount)
WHERE {
  ?actor :performs ?action .
  ?action a :Observing .
  ?action :interacting_aoi ?aoi .
  ?aoi a :AOI .
}
GROUP BY ?actor ?aoi
ORDER BY DESC(?observationCount)
```

3.4.4. Extracting Gaze Sequences

To analyze the sequential patterns in which users observe different AOIs, Listing 4 retrieves gaze observations in chronological order. This query provides information on the start time and duration of each observation, allowing the system to track the order in which AOIs are viewed and analyze users' attention shifts.

Listing 4: Extracting transitional sequences of visiting AOI

```
SELECT ?actor ?aoi ?startTimeValue ?durationValue
WHERE {
  ?actor :performs ?action .
  ?action a :Observing .
  ?action :interacting_aoi ?aoi .
  ?aoi a :AOI .

  # Fetching the Start Time
  ?action :has_time ?startTime .
  ?startTime a :StartTime .
  ?startTime :value ?startTimeValue .

  # Fetching the Duration
  ?action :has_time ?duration .
  ?duration a :Duration .
  ?duration :value ?durationValue .
}
ORDER BY ?startTimeValue
```

4. Case study: Analyzing gaze patterns and user engagement in virtual exhibitions

4.1. VR exhibition setup

In 2020, the Museum Rembrandthuis[3] organized a special exhibition titled "HERE: Black in Rembrandt's Time." To preserve this physical exhibition, we developed a virtual exhibition using Unity, featuring a selection of 19 paintings arranged across three thematic rooms (see Figure 3). Each painting, along with its accompanying textual descriptions, was modeled as a separate virtual object, enabling us to precisely track and analyze visitor interactions.

The development of this VR experience allowed us to address a key challenge in cultural heritage: understanding how visitors engage with individual elements of an exhibition. By utilizing advanced

[3] https://www.rembrandthuis.nl/en/

eye-tracking technology, we were able to monitor and interpret the specific elements that captured users' attention throughout their virtual visit.

Figure 3: VR Exhibition "HERE: Black in Rembrandt's Time" features 19 paintings organized into three different rooms. Each painting and its accompanying text description are modeled as individual virtual objects.

4.2. User study and eye-tracking

To assess user interaction with the virtual artworks, two user studies were conducted with a group of 24 high school students (aged 16–18). Participants were equipped with HTC VIVE Pro Eye headsets, allowing for real-time tracking of their eye movements during the exhibition. The eye-gaze data collected included timestamps, painting identifiers, and x-y gaze coordinates for each painting. This comprehensive dataset formed the foundation for our gaze behavior analysis.

Our study contributes to VR exhibition research by providing an ontology-driven framework to link eye-tracking data to specific objects within the paintings. This allows for a detailed analysis of how users visually explore different areas of the artwork, enhancing our understanding of their attention patterns.

4.3. Analysis of Gaze Patterns and Engagement

Using object detection models like YOLOv8 and segmentation techniques via the Segment Anything Model (SAM), we detected key objects within each painting and linked them to semantic class labels (e.g., "Man," "Dress," "Table"). An example of a segmented painting is depicted in the top left of Figure 4. This image showcases five objects, each uniquely colored for enhanced visualization: a man, a woman, a piece of clothing, a dress, and a table. These objects were stored as Areas of Interest (AOIs) in our ontology, providing a structured framework for integrating gaze data with the content of the paintings. The segmentation results and class annotations were overlaid onto the paintings, allowing us to study user gaze interactions with specific objects.

Table 1 displays the total time each participant spent on each type of objects. We observe diverse behaviors among users: some participants did not look at all the identified objects (e.g., participants P2, P8, and P19). Others spent very little time on the identified objects, like participants P8 and P15, while some, such as P22, spent considerable time on the majority of identified objects. Additionally, some participants exhibited mixed patterns in their interaction with the objects; for instance, participant P17 spent significant time on certain objects and very little time on others.

Table 2 provides information about the number of paintings each participant viewed, as well as the number of paintings in which they spent more time looking at objects rather than the background. The data shows that in the majority of paintings, most participants primarily focused on the background. Percentages of time spent on objects confirm this, with exceptions noted for P2, P14, P17, and P19, who focused more on objects than the background. However, the total and average number of times participants looked at objects exceeded those for the background across all participants. This indicates that while participants spent the majority of their time on the background, they visited the objects more frequently than the background.

Such insights can inform the design of personalized virtual guides or interactive agents that adapt to users' interests, offering tailored recommendations or additional information based on their gaze behavior. For example, if a participant shows prolonged interest in a particular object, the virtual agent could provide additional historical or artistic context, enriching the user's experience. This ability to

Table 1
Total time (ms) spent by participants on each class of AOIs and the background (areas not classified as AOIs are considered background) in the paintings. Bold numbers indicate the classes that received the highest gaze time.

P	Background	Man	Woman	Clothing	Human face	Bust	Building	Picture frame	Fedora	Tree	Person	Dress	Table
1	438.49	28.61	**107.30**	34.64	0.27	4.04	0.58	0.26	1.77	3.25	0	1.21	9.02
2	30.23	0.02	8.41	25.23	0.19	0	**43.12**	0	0	0	0	2.01	9.90
3	522.03	49.55	62.09	99.55	7.26	14.12	1.84	**130.07**	4.00	0	2.36	12.02	0.65
4	1547.11	9.95	**303.76**	281.89	95.63	5.37	9.75	11.14	23.42	0.13	4.17	5.55	2.69
5	675.01	69.11	**219.54**	61.85	79.19	58.23	0.33	8.93	4.85	0	1.80	6.86	14.88
7	339.70	20.82	48.56	**110.68**	2.57	38.24	0.95	34.14	0.14	0	0.13	1.23	0.24
8	132.04	0	**4.19**	2.93	0	0	0.78	0	0	0	0	0	0
9	1976.27	111.63	**997.90**	166.95	80.88	3.29	44.58	9.03	12.67	0.34	0.11	9.77	1.31
10	522.08	24.71	90.34	45.34	2.16	2.26	0.83	0.16	**236.65**	0.18	0.02	4.06	5.69
11	214.08	65.65	**80.76**	7.25	4.33	0.73	0.02	0.40	0	0.07	0	0.09	2.75
12	2450.84	316.09	454.62	**535.32**	5.17	8.15	37.93	51.65	251.38	0.10	1.92	2.81	47.21
13	589.24	57.32	42.75	**129.88**	1.16	6.35	3.43	128.21	46.18	0	0	14.69	0.56
14	127.09	44.79	40.29	26.05	13.31	0.20	4.44	**48.91**	0.42	0	0	22.55	0.79
15	223.81	4.53	11.08	**21.76**	4.74	1.86	1.24	2.44	0.65	0.59	0	2.05	2.61
16	1742.37	7.07	141.35	**316.07**	27.23	0.60	4.34	7.10	0.89	0.35	0.57	6.48	0.20
17	221.54	18.69	122.89	**140.89**	4.70	42.43	1.43	15.25	0.37	0	0	3.56	2.33
18	621.20	17.33	**138.09**	133.58	5.64	4.43	0.84	49.24	0.02	0.07	0.40	4.93	2.03
19	113.30	3.67	5.96	**166.77**	0.23	0	0	0.13	0.36	0	0	0	0
20	162.25	1.08	**47.76**	29.64	27.32	12.05	0.69	0	14.89	0.18	0.24	0.78	0.18
21	1080.97	60.41	**417.57**	63.54	4.76	63.98	26.15	98.03	103.89	0.15	2.12	10.29	0.15
22	13768.23	1009.50	**2881.16**	974.72	881.14	101.15	158.37	768.61	5.57	0.12	24.18	64.64	191.96
23	430.78	7.06	33.08	**52.75**	1.17	0	0.84	0.44	10.15	0.03	0.39	1.27	0.15

Table 2
Number of visited paintings (#p), instances where participants focused more on objects than the background (#f), mean percentage of time spent on objects (%o), and the frequencies of visits to objects (#o) and background (#bg).

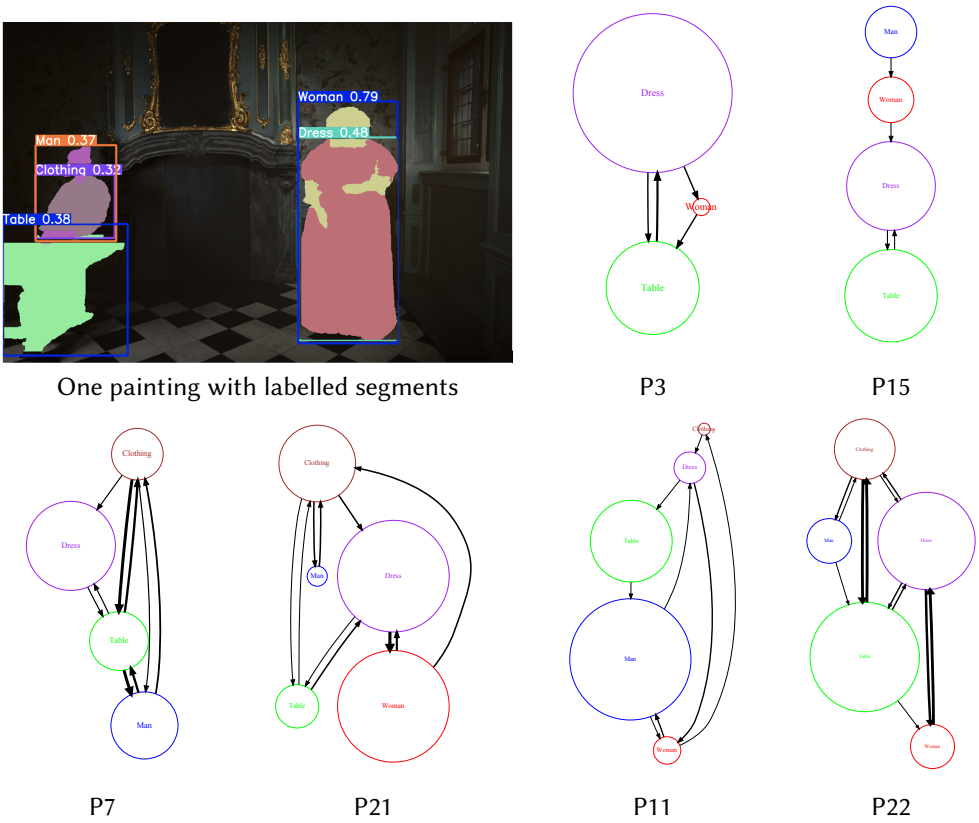
P	#p	#f	%o	#o	#bg		P	#p	#f	%o	#o	#bg
1	17	6	0.3016	267	165		13	18	4	0.4073	282	157
2	9	5	0.6901	160	80		14	18	10	0.6132	462	247
3	19	6	0.4191	416	266		15	17	3	0.1841	151	78
4	20	5	0.3138	694	419		16	20	4	0.2189	1097	603
5	18	8	0.4361	494	269		17	17	9	0.6134	370	233
7	19	6	0.4095	219	131		18	20	6	0.3459	370	217
8	2	0	0.0866	46	34		19	6	2	0.6095	76	26
9	20	6	0.3629	610	410		20	18	4	0.4378	289	143
10	20	4	0.3594	319	186		21	19	7	0.3955	493	274
11	12	5	0.4364	142	87		22	20	6	0.3373	1587	888
12	20	5	0.3888	698	414		23	18	4	0.1667	344	193

dynamically adapt to user behavior represents a significant advancement in virtual cultural heritage experiences.

4.4. Gaze transition sequences

Figure 4 illustrates the diverse gaze transition sequences of participants while viewing Painting D5. Using the SPARQL query provided in Listing 4, we extracted the transitional sequences for each participant who spent time observing this painting. From the sequences, we observed that participant P3 primarily focused on the dress of the lady on the right, briefly shifted attention to her face, then moved to the table on the left. Their gaze alternated between the table and the dress. Similarly, participant P15 scanned briefly from the man and woman to the woman's dress, then back to the table before moving on to other paintings. In contrast, participants P21 and P22 dedicated substantial time exploring each object in the exhibition. P21 focused more on the lady and her dress on the right, whereas P22 showed interest in the man and the table on the left.

These insights offer a deeper understanding of how users' attention moves across different elements



One painting with labelled segments P3 P15

P7 P21 P11 P22

Figure 4: An example painting with detected AOIs (labeled segments) is shown alongside different participants' eye gaze transition sequences. The size of the nodes represents the amount of time a participant looks at the corresponding object. Arrows indicate the direction of the eye gaze transitions from the starting object to the ending object, and the thickness of the edges reflects the frequency of these transitions.

of a painting, providing valuable data for curators and designers to optimize exhibit layouts and content presentation. By linking gaze data with AOIs and analyzing gaze durations, our system is capable of generating high-level behavioral trends, which can be used to enhance user engagement. These findings contribute to the broader goal of creating more interactive and personalized cultural heritage experiences in VR environments.

5. Discussion and conclusion

A critical limitation in this study is the performance of the object detection model, particularly for lower-resolution paintings. Despite predefined class labels, some objects were misclassified as background, skewing the analysis of user gaze patterns. Improving object detection accuracy – by incorporating additional class labels and domain-specific datasets, particularly for classical artworks – will be essential for enhancing the precision of gaze-based analysis in future work.

Despite these challenges, our work contributes to the field of knowledge management for numerical modeling by integrating segmented object data and real-time eye-gaze data into a structured, ontology-based framework. This allows for a comprehensive analysis of user interactions in virtual art exhibitions. By linking eye-gaze data to specific Areas of Interest (AOIs), identified through object segmentation, we provide a semantic layer that enhances the understanding of user behavior. This enriched data supports personalized virtual guides that can adapt in real-time to visitor engagement, offering tailored experiences based on gaze patterns.

Our approach also demonstrates the value of gaze data in informing knowledge management practices, particularly by capturing visitor behavior at a granular level. The identification of different visitor

types based on gaze patterns enables the development of personalized, automated tours, and managing complex datasets to support adaptive, data-driven experiences. Additionally, insights into demographic engagement patterns, such as shorter attention spans in younger visitors, can inform the design of more effective cultural heritage applications.

Managing the large volume of gaze data within the ontology required balancing data granularity with memory efficiency. To address this, we propose selectively storing key metrics such as fixation counts, transitions between AOIs, and time spent on objects in the future. This would ensure that meaningful engagement data is retained while minimizing the system's data footprint, enhancing the scalability of the framework for larger exhibitions and more complex datasets.

In conclusion, our ontology-based framework successfully integrates segmented object data with gaze patterns, offering a novel approach to understanding user behaviors in VR exhibitions. While further improvements in object detection are necessary, our methodology contributes to the broader field of knowledge management by semantically enriching gaze data and enabling adaptive, personalized user experiences. This work demonstrates the potential for integrating complex measurement datasets and behavioral data into knowledge management systems.

References

[1] A. Puig, I. Rodríguez, J. L. Arcos, J. A. Rodríguez-Aguilar, S. Cebrián, A. Bogdanovych, N. Morera, A. Palomo, R. Piqué, Lessons learned from supplementing archaeological museum exhibitions with virtual reality, Virtual Reality 24 (2020) 343–358.

[2] L. Gao, B. Wan, G. Liu, G. Xie, J. Huang, G. Meng, Investigating the effectiveness of virtual reality for culture learning, International Journal of Human–Computer Interaction 37 (2021) 1771–1781.

[3] M. Trunfio, M. D. Lucia, S. Campana, A. Magnelli, Innovating the cultural heritage museum service model through virtual reality and augmented reality: The effects on the overall visitor experience and satisfaction, Journal of Heritage Tourism 17 (2022) 1–19.

[4] S. Machała, N. Chamier-Gliszczyński, T. Królikowski, Application of ar/vr technology in industry 4.0, Procedia Computer Science 207 (2022) 2990–2998. URL: https://doi.org/10.1016/j.procs.2022.09.357.

[5] H. Lee, T. H. Jung, M. Tom Dieck, N. Chung, Experiencing immersive virtual reality in museums, Information & Management 57 (2020) 103229. URL: https://doi.org/10.1016/j.im.2019.103229.

[6] B. Fasel, L. Van Gool, Interactive museum guide: Accurate retrieval of object descriptions, in: Lecture Notes in Computer Science, 2007, pp. 179–191. URL: https://doi.org/10.1007/978-3-540-71545-0_14.

[7] C. Bailey-Ross, S. Gray, J. Ashby, M. Terras, A. Hudson-Smith, C. Warwick, Engaging the museum space: Mobilizing visitor engagement with digital content creation, Digital Scholarship in the Humanities 32 (2016) 689–708. URL: https://doi.org/10.1093/llc/fqw041.

[8] J. B. Schreiber, A. J. Pekarik, N. Hanemann, Z. Doering, A. Lee, Understanding visitor engagement and behaviors, The Journal of Educational Research 106 (2013) 462–468. URL: https://doi.org/10.1080/00220671.2013.833011.

[9] M. Kejriwal, What is a knowledge graph? domain-specific knowledge graph construction, in: Domain-Specific Knowledge Graph Construction, 2019, pp. 1–7. URL: https://doi.org/10.1007/978-3-030-12375-8_1.

[10] P. Weinzaepfel, G. Csurka, Y. Cabon, M. Humenberger, Visual localization by learning objects-of-interest dense match regression, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019. URL: https://doi.org/10.1109/cvpr.2019.00578.

[11] F. Barth, H. Candello, P. Cavalin, C. Pinhanez, Intentions, meanings, and whys, in: Proceedings of the 2nd Conference on Conversational User Interfaces, 2020. URL: https://doi.org/10.1145/3405755.3406128.

[12] P. P. Morantes, S. A. Penarete, G. Arbelaez, M. Camargo, L. Dupont, Understanding museum visitors' experience through an eye-tracking study and a living lab approach, in: 2016 International

Conference on Engineering, Technology and Innovation/IEEE lnternational Technology Management Conference (ICE/ITMC), 2016. URL: https://doi.org/10.1109/ice/itmc39735.2016.9025900.

[13] D. Javdani Rikhtehgar, S. Wang, H. Huitema, J. Alvares, S. Schlobach, C. Rieffe, D. Heylen, Personalizing cultural heritage access in a virtual reality exhibition: A user study on viewing behavior and content preferences, in: Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization, 2023, pp. 379–387.

[14] S. Wang, D. Kulyk, D. J. Rikhtehgar, D. Heylen, C. Rieffe, Correlating eye gaze with object to enrich cultural heritage knowledge graph, in: CEUR workshop proceedings, volume 3632, Rheinisch Westfälische Technische Hochschule, 2023.

[15] S. Zhang, Z. Zhang, L. Sun, W. Qin, One for all: A mutual enhancement method for object detection and semantic segmentation, Applied Sciences 10 (2019) 13. URL: https://doi.org/10.3390/app10010013.

[16] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014. URL: https://doi.org/10.1109/cvpr.2014.81.

[17] R. Mottaghi, X. Chen, X. Liu, N. Cho, S. Lee, S. Fidler, R. Urtasun, A. Yuille, The role of context for object detection and semantic segmentation in the wild, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014. URL: https://doi.org/10.1109/cvpr.2014.119.

[18] J. Dong, Q. Chen, S. Yan, A. Yuille, Towards unified object detection and semantic segmentation, in: Computer Vision – ECCV 2014, 2014, pp. 299–314. URL: https://doi.org/10.1007/978-3-319-10602-1_20.

[19] R. Padilla, S. L. Netto, E. A. B. da Silva, A survey on performance metrics for object-detection algorithms, in: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), 2020, pp. 237–242. URL: https://doi.org/10.1109/IWSSIP48289.2020.9145130.

[20] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, M. Pietikäinen, Deep learning for generic object detection: A survey, International Journal of Computer Vision 128 (2020) 261–318. URL: https://doi.org/10.1007/s11263-019-01247-4.

[21] K. He, G. Gkioxari, P. Dollar, R. Girshick, Mask r-cnn, in: International Conference on Computer Vision (ICCV), 2017, pp. 2961–2969. URL: https://openaccess.thecvf.com/content_iccv_2017/html/He_Mask_R-CNN_ICCV_2017_paper.html.

[22] A. Kirillov, K. He, R. Girshick, C. Rother, P. Dollar, Panoptic segmentation, in: Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 9404–9413. URL: https://openaccess.thecvf.com/content_CVPR_2019/html/Kirillov_Panoptic_Segmentation_CVPR_2019_paper.html.

[23] X. Yue, K. Qi, X. Na, Y. Zhang, Y. Liu, C. Liu, Improved yolov8-seg network for instance segmentation of healthy and diseased tomato plants in the growth stage, Agriculture 13 (2023). URL: https://doi.org/10.3390/agriculture13081643.

[24] L. P. Osco, Q. Wu, E. L. de Lemos, W. N. Gonçalves, A. P. Ramos, J. Li, J. Marcato, The segment anything model (sam) for remote sensing applications: From zero to one shot, International Journal of Applied Earth Observation and Geoinformation 124 (2023) 103540. URL: https://doi.org/10.1016/j.jag.2023.103540.

[25] Y. Xiong, R. Liao, H. Zhao, R. Hu, M. Bai, E. Yumer, R. Urtasun, Upsnet: A unified panoptic segmentation network, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019. URL: https://doi.org/10.1109/cvpr.2019.00902.

[26] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W. Lo, P. Dollar, R. Girshick, Segment anything, in: International Conference on Computer Vision (ICCV), 2023, pp. 4015–4026. URL: https://openaccess.thecvf.com/content/ICCV2023/html/Kirillov_Segment_Anything_ICCV_2023_paper.html.

[27] D. Javdani Rikhtehgar, I. Tiddi, S. Wang, S. Schlobach, D. Heylen, Assessing the hi-ness of virtual heritage applications with knowledge engineering, in: HHAI 2024: Hybrid Human AI Systems for the Social Good, IOS Press, 2024, pp. 173–187.

[28] Y. Zhang, J. Zhu, Q. Zhu, Y. Xie, W. Li, L. Fu, J. Zhang, J. Tan, The construction of personalized virtual landslide disaster environments based on knowledge graphs and deep neural networks,

International Journal of Digital Earth 13 (2020) 1637–1655.
[29] X. Chen, S. Jia, Y. Xiang, A review: Knowledge reasoning over knowledge graph, Expert systems with applications 141 (2020) 112948.
[30] T. Diwan, G. Anirudh, J. v. Tembhurne, Object detection using yolo: challenges, architectural successors, datasets and applications, Multimedia Tools and Applications 82 (2023) 9243–9275. URL: https://doi.org/10.1007/s11042-022-13644-y.
[31] S. Castagnos, P. Pu, Consumer decision patterns through eye gaze analysis, in: Proceedings of the 2010 workshop on Eye gaze in intelligent human machine interaction, 2010, pp. 26–33. doi:10.1145/2002333.2002346.
[32] F. Koochaki, L. Najafizadeh, Predicting intention through eye gaze patterns, in: 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS), 2018, pp. 1–4. doi:10.1109/biocas.2018.8584665.
[33] D. S. Chang, G. H. Cho, Y. S. Choi, Ontology-based knowledge model for human-robot interactive services, in: Proceedings of the 35th Annual ACM Symposium on Applied Computing, 2020, pp. 2029–2038.
[34] S. D. Costa, M. P. Barcellos, R. de Almeida Falbo, T. Conte, K. M. de Oliveira, A core ontology on the human–computer interaction phenomenon, Data & Knowledge Engineering 138 (2022) 101977.
[35] M. Castro, M. Barcellos, An ontology to support knowledge management solutions for human-computer interaction design, in: Proceedings of the XXI Brazilian Symposium on Software Quality, 2022, pp. 1–10.
[36] A. A. de Freitas, M. B. Scalser, S. D. Costa, M. P. Barcellos, Towards an ontology-based approach to develop software systems with adaptive user interface, in: Proceedings of the 21st Brazilian Symposium on Human Factors in Computing Systems, 2022, pp. 1–7.
[37] E. Champion, Virtual heritage: a guide, Ubiquity Press, 2021.
[38] F. Liarokapis, P. Petridis, D. Andrews, S. de Freitas, Multimodal serious games technologies for cultural heritage, Mixed reality and gamification for cultural heritage (2017) 371–392.
[39] D. Liu, Knowledge Graph Driven Conversational Virtual Museum Guide, Master's thesis, University of Twente, 2021.
[40] M. Duguleană, V.-A. Briciu, I.-A. Duduman, O. M. Machidon, A virtual assistant for natural interactions in museums, Sustainability 12 (2020) 6958.
[41] B. De Carolis, N. Macchiarulo, C. Valenziano, Marta: A virtual guide for the national archaeological museum of taranto, in: Proceedings of the 2022 AVI-CH Workshop on Advanced Visual Interfaces for Cultural Heritage. CEUR-WS. org, 2022.
[42] P. Dondi, M. Porta, Gaze-based human–computer interaction for museums and exhibitions: Technologies, applications and future perspectives, Electronics 12 (2023) 3064.
[43] M. Mokatren, T. Kuflik, I. Shimshoni, Exploring the potential of a mobile eye tracker as an intuitive indoor pointing device: A case study in cultural heritage, Future generation computer systems 81 (2018) 528–541.
[44] T. Yi, M. Chang, S. Hong, J.-H. Lee, Use of eye-tracking in artworks to understand information needs of visitors, International Journal of Human–Computer Interaction 37 (2021) 220–233.
[45] M. Pelowski, H. Leder, V. Mitschke, E. Specker, G. Gerger, P. P. Tinio, E. Vaporova, T. Bieg, A. Husslein-Arco, Capturing aesthetic experiences with installation art: An empirical assessment of emotion, evaluations, and mobile eye tracking in olafur eliasson's "baroque, baroque!", Frontiers in Psychology 9 (2018) 1255.
[46] M. Rainoldi, B. Neuhofer, M. Jooss, Mobile eyetracking of museum learning experiences, in: Information and Communication Technologies in Tourism 2018: Proceedings of the International Conference in Jönköping, Sweden, January 24-26, 2018, Springer, 2018, pp. 473–485.
[47] R.-M. Rahal, S. Fiedler, Understanding cognitive and affective mechanisms in social psychology through eye-tracking, Journal of Experimental Social Psychology 85 (2019) 103842.