NTNU
Kunnskap for en bedre verden

Department of Mechanical and Industrial Engineering

Master's Thesis - robotics & automation

Efficient, accurate, and privacy-preservant object detection in edge devices

Student:
Hallvard Enger Bjørgen

Supervisor at NTNU:
Amund Skavhaug

Trondheim/Esbjerg Spring 2024

Table of Contents

List of Figures i

List of Tables ii

1 Introduction 1

1.1 Background and Motivation . . . 1

1.2 Problem Description . . . 2

1.3 Project Scope . . . 3

1.3.1 Objectives . . . 3

1.3.2 Research Questions . . . 3

1.4 Structure . . . 3

2 Literature Review 5

2.1 Introduction to Object Detection and Visitor Tracking . . . 5

2.2 Visitor Behavior Analysis in Cultural Institutions . . . 5

2.3 Privacy-Preserving Technologies in Surveillance . . . 5

2.4 Case Studies and Applications . . . 5

2.5 Challenges and Future Directions . . . 5

2.5.1 Effectiveness of Object Detection Algorithms . . . 5

2.5.2 Impact of Technology on Visitor Experience . . . 5

2.5.3 Privacy and Ethics . . . 6

2.5.4 Technology Integration . . . 6

3 Conclusion 6

4 Methodology 7

4.1 Labeling . . . 7

4.2 Ethical Considerations . . . 8

4.2.1 Privacy by Design . . . 8

4.2.2 Ethical Use and Data Protection . . . 8

4.2.3 Transparency and Accountability . . . 8

Bibliography 10

List of Figures

List of Tables

Abstract

Placeholders:
What
Why
How
Principal contributions
Principal conclusion

1 Introduction

Citation test: Bjørgen, 2023

1.1 Background and Motivation

The study of visitor engagement in physical spaces, particularly in museums and aquariums, has evolved significantly alongside advancements in technology. Traditionally, understanding how visitors interact with exhibits relied on direct observation, surveys, and manual counting. These methods, while valuable, offered limited insights and were often labor-intensive and prone to biases. The desire to quantitatively measure and analyze visitor behavior has driven the adoption of increasingly sophisticated technologies, from simple passive infrared sensor counters to complex digital surveillance systems.

In the late 20th century, the introduction of video cameras provided a new avenue for tracking visitors, allowing for more detailed observations of visitor behaviour. However, these early systems were primarily used for security purposes and offered limited capabilities for analyzing visitor behavior in depth in a privacy preservant manner. As digital technologies advanced, so did the potential for understanding visitor experiences in more nuanced and meaningful ways.

The turn of the millennium marked a significant shift with the emergence of computer vision and artificial intelligence (AI) technologies. These advancements enabled the development of systems capable of not just recording visitor movements but interpreting them without the need of human interference. Object detection algorithms began to allow for the automatic localization of people across a space, allowing more complex insights to be extracted from the scene.

Today, the field stands on the cusp of a new era, fueled by deep learning and real-time data analysis capabilities. Technologies such as pose estimation, gaze tracking, and behavioral analysis algorithms provide unprecedented insights into how visitors engage with exhibits. These tools can identify not just which exhibits attract the most attention but how visitors physically interact with them, offering a window into the visitor experience that was previously unimaginable.

The current research endeavor, set in an aquarium environment and focusing on the use of advanced machine learning technologies to analyze visitor engagement across different aquariums, is a testament to this technological evolution. By leveraging state-of-the-art object detection to localize the visitors and thus infer their attention, this research aims to provide a detailed understanding of visitor behavior and preferences. This approach not only extends the traditional methodologies of visitor studies but also integrates the latest in technology to offer a holistic view of engagement in cultural spaces.

This historical context underscores the transformative impact of technology on the study of visitor engagement. From the rudimentary tools of the past to the sophisticated systems of today, the journey reflects a broader trend towards increasingly precise, nuanced, and actionable insights into how people interact with their environment. This research represents a continuation of that journey, pushing the boundaries of what can be understood about visitor experiences in cultural institutions.

Visitor attention intelligence in cultural facilities, such as museums or aquariums, offers a wide array of potential applications. By analyzing visitor engagement and behavior, facilities can make informed decisions to enhance the visitor experience. This thesis aims to provide insights into visitor behavior, addressing questions such as:

Questions

1. What are the popularity ratings of various exhibitions within the facility?
2. What are the average and peak visitor counts in front of each exhibition, and how frequently are these peak counts reached?

3. Which exhibition experiences the longest visitor engagement times?
4. To what extent does the physical placement of an exhibition influence visitor engagement and dwell time?
5. Which areas within the facility are most susceptible to crowding or queue formation?
6. Are all parts of the museum being visited?

The findings from these questions can inform strategic improvements and interventions to enhance the quality of visitation. Potential strategic actions include:

Strategic actions

1. Enhance the lighting around exhibits that gain increased popularity during specific times, such as daylight hours, to maintain or boost visitor interest.
2. Reevaluate the positioning of artworks or exhibitions based on popularity data, ensuring high-value or popular items are placed in strategic locations to maximize visibility and engagement.
3. Implement environmental controls, such as automatic window opening, to regulate airflow and temperature in response to room occupancy exceeding a certain threshold for a designated period.
4. Automate the management of exhibit areas based on real-time visitor presence, such as closing off sections near closing time to streamline the exit process if no visitors are detected.
5. Develop a system integration plan for safety and operational efficiency, enabling immediate response to incidents like falls into water features or ensuring all visitors have vacated the premises before locking down the facility.
6. Utilize visitor count data to identify discrepancies between ticket sales and actual attendance, flagging potential security concerns such as unauthorized entry.

These strategies, informed by data-driven insights into visitor behavior, can significantly enhance operational efficiency, safety, and the overall visitor experience in cultural institutions.

1.2 Problem Description

Recent advancements in object detection technology have significantly enhanced applications across various fields, including security surveillance, retail analytics, and visitor tracking in cultural institutions. Despite these advancements, achieving high-quality and efficient object detection, particularly in the context of visitor localization within museums and aquariums, presents distinct challenges. Key among these is the ability to accurately detect and localize individuals in environments where lighting conditions are variable or suboptimal and where objects or individuals may be partially occluded.

In environments with insufficient lighting, conventional image sensors struggle to capture the detailed visual information required for accurate object detection and segmentation. Although these sensors can detect infrared (IR) light, which extends beyond the visual spectrum visible to the human eye, leveraging this capability introduces its own set of challenges. IR sensors can indeed infer the presence and movement of people within a space, but they fall short when it comes to segmenting individuals from one another or from the background. The difficulty arises from the inherent fuzziness of borders in the infrared spectrum.

In cultural spaces like museums and aquariums, lighting conditions are often designed with the preservation of exhibits in mind rather than optimal visibility for image capture. This can lead to areas within these institutions where low light or shadows impair the effectiveness of standard image sensors. While these sensors are capable of detecting infrared (IR) light, which is invisible to the human eye, relying solely on IR sensors for visitor localization introduces its own set of challenges. Although IR sensors can broadly detect movement and presence, the fuzzy nature of individual outlines in the infrared spectrum complicates the differentiation process of individuals within a group or accurately segmenting a person from the background becomes problematic.

In summary, this project aims to address the multifaceted challenges of implementing an effective and efficient visitor localization system in museums and aquariums. These challenges include adapting to variable lighting conditions, accurately identifying and tracking individuals in crowded or complex scenes, and ensuring the privacy of all visitors. Overcoming these obstacles is crucial for leveraging visitor localization technology to enhance museum management and the overall visitor experience.

1.3 Project Scope

The scope of this project encompasses developing a sophisticated visitor localization system tailored for use in museums and aquariums. This system aims to address the inherent challenges posed by the scene and the need for maintaining visitor privacy.

1.3.1 Objectives

The primary objectives of this project are to:

1. Develop an object detection system that performs reliably under the indoor lighting conditions, characteristic of museum and aquarium environments.
2. Implement privacy-preserving technologies within the localization system to ensure that visitor data is anonymized and secure.
3. Evaluate the system's accuracy and efficiency in real-world settings, identifying areas for improvement and potential scalability.
4. Explore the integration of the localization system with existing museum management systems to enhance the visitor experience through personalized content and navigation aids.

1.3.2 Research Questions

The research will address the following critical questions, which are pivotal to understanding the challenges and opportunities in object detection within cultural institutions:

1. What are the current challenges in object detection technologies, especially in low-light and crowded environments found in museums and aquariums?
2. How can object detection systems be designed to respect and ensure the privacy of visitors while collecting useful localization data?
3. In what ways can the accuracy and efficiency of these systems be improved to provide real-time insights into visitor behavior and exhibit engagement?
4. What potential does such technology hold for transforming museum and aquarium management and the overall visitor experience?

1.4 Structure

The remainder of this document is structured as follows to provide a comprehensive overview of the project and its findings:

1. **Literature Review:** An examination of existing research and technologies in object detection and visitor localization, highlighting gaps this project aims to fill.
2. **Methodology:** Detailed description of the system design, development process, and the methods used for testing and data analysis.
3. **Results:** Presentation of the findings from the implementation of the localization system in a museum or aquarium setting.
4. **Discussion:** Analysis of the results in the context of the research questions, including the implications for museum management and visitor experience.
5. **Conclusion and Future Work:** Summary of the project's contributions to the field and suggestions for future research directions.

2 Literature Review

2.1 Introduction to Object Detection and Visitor Tracking

Understanding visitor behavior in cultural institutions such as museums and aquariums is pivotal for enhancing visitor engagement, optimizing exhibit design, and improving overall visitor experience. The study of visitor behavior encompasses a broad range of methods and approaches, from traditional observational techniques to advanced digital tracking technologies.

2.2 Visitor Behavior Analysis in Cultural Institutions

Studies on traditional methods for analyzing visitor behavior (surveys, manual counting, direct observation) and their limitations. Research on the use of digital technologies (video surveillance, mobile tracking) for understanding visitor engagement and flow in museums and aquariums.

2.3 Privacy-Preserving Technologies in Surveillance

Exploration of techniques for ensuring privacy in surveillance, such as real-time image processing (pixelation, blurring), data anonymization, and encryption. Discussion on ethical considerations and legal frameworks governing the use of surveillance technologies in public spaces.

2.4 Case Studies and Applications

Detailed examination of case studies involving the implementation of advanced tracking technologies in museums, aquariums, or similar cultural institutions. Analysis of objectives, methodologies, results, and implications of these studies. Detailed case studies of museums, aquariums, or similar cultural institutions that have implemented advanced tracking technologies. Focus on the objectives, methodologies, results, and implications of these studies. Analysis of how these technologies have impacted visitor experience, exhibit design, and operational decisions. Challenges and Future Directions

2.5 Challenges and Future Directions

Identification of current challenges in the field, such as accuracy in diverse environments, scalability, and balancing privacy with data utility. Speculative insights into future trends in technology development and application in cultural spaces.

2.5.1 Effectiveness of Object Detection Algorithms

Comparison of various algorithms' performance in crowded or complex environments typical of museums and aquariums.

2.5.2 Impact of Technology on Visitor Experience

Studies assessing how tracking technologies affect visitor satisfaction, engagement, and behavior.

2.5.3 Privacy and Ethics

Research addressing the ethical implications of surveillance in public spaces, including visitor perceptions and legal considerations.

2.5.4 Technology Integration

Examples of how cultural institutions have integrated object detection and tracking systems with other technologies.

3 Conclusion

A critical evaluation of existing research, highlighting gaps that the project aims to fill and emphasizing the novelty of the approach, especially the application of privacy-preserving object detection technologies in cultural institutions.

4 Methodology

Two cameras were deployed in a room of acquariums at ”Fiskeri- og Søfartsmuseet” in Esbjerg to take images for building a specialized dataset and to evaluate the effects of developing a highly specialized detector rather than using a general.

Camera setup: Hardware camera can be tweaked by screwing the lens with a mechanical tool to modify it's aperture, which influences it's depth focus. Aperture mechanical setting (camera focus adjustment), depth control. Default not found... images well sharp enough... 50cm to infinity...

Dataset obtainment Dataset was build by capturing images while no visitors were present in the acquarium. Due to the constraint to only operate within opening hours when the facility was open to everyone, a way to cancel image capturing was needed in the case if someone entered the room. One of the goals of the dataset was to have the images taken from the same angle as the device will be used in the future. The device was therefore mounted in the corner of the room, and ssh was used to access the device remotely from a pc in the acquarium. Then, a script was ran to capture an image per second and storing it on the SD card in the device. The choice to store the image locally rather than transmitting it was to not have to worry about data transmission costs and issues. The first iteration of image capture was made with non-optimized camera configurations, turning to the brightness setting of the picamera python module to get the images to a sufficient brightness. Still, some images fell short to the auto focus of the camera focusing on the bright acquariums and thus rendering the rest of the image rather dark. However, these images serve as a way of inspecting the impact of captured image quality on inference performance.

Due to many technical difficulties the first few times images were being captured for the dataset, only the developer and author of this thesis is present in the images[1]. The acquarium could not provide free tickets for additional participants in the experiments, and the cost of bringing friends to enrichen the dataset would then have to be optimized in terms of time spent at the site.

The detector needs to know the ground truth when training and validating on the obtained data. This can be obtained by manually labeling the data. However, a more scientific, robust and scaleable way of labeling is to have the detector do the heavy lifting. Therefore, the dataset was first inferenced by the detector. The data was then manually validated, and finally the images for which the detector did not find any persons was manually labeled. The time to understand the tools and develop this pipeline was similar to what it would take to manually label all the images, but can now be used for future applications also. The approach to label the images is described in section: 4.1.

After the ground truth was identified for all the dataset images, it was then used to evaluate the general-purpose yolov3 and yolov9. Results are discussed in ??. The data was then used to train the detectors. The training process is described in section ??. The trained models were then deployed to the device to evaluate the inference speed and accuracy. The deployment process is described in section ??.

To visualize the improvements and highlight the areas in the image benefitting most from detector improvements, heat maps were generated. The process of generating heat maps is described in section ??.

4.1 Labeling

Ground truth values for the dataset must be obtained before improving the model and evaluating it's effect.

[1]Initially, an attempt was made to pass MQTT messages as a way to initialize image capture so multiple cameras could be deployed in several locations, thus speeding up and simplifying the image capturing process. This was discarded due to technical difficulties related to efficiently stopping the image capturing. For a different solution, a recommendation would be to implement a way of communicating to multiple devices however, so one could obtain all images from deployment locations in one shooting. For this single-deployment angle and area project, however, the approach with ssh-ing into the device worked fine.

"Label Studio" was used to label the images. First, the dataset was predicted with a yolo v9 model trained on the COCO dataset. Predictions were mostly decent, but some needed small tweakings and in some cases the persons were not discovered. The detector had close to zero hallucinations due to a sufficiently high confidence rate of 0.5 (: for), but in some cases the fish were identified as human.

4.2 Ethical Considerations

In the deployment of advanced machine learning technologies for visitor localization and engagement analysis, this research proactively addresses privacy concerns through the implementation of image obscuration techniques. These methods ensure that no personally identifiable information is captured or communicated, thus significantly reducing privacy risks associated with visitor tracking in cultural spaces such as museums and aquariums.

4.2.1 Privacy by Design

At the forefront of our ethical approach is the principle of "privacy by design." This concept involves integrating privacy into the development and operation of our tracking technologies from the outset, rather than as an afterthought. By employing image obscuration techniques, such as real-time pixelation or silhouette generation, we ensure that the visual data processed by our system remains anonymous. This method effectively eliminates the possibility of identifying individual visitors from the captured data, thereby safeguarding their privacy.

The application of these privacy-preserving techniques negates the need for explicit consent from visitors for two primary reasons. First, the anonymization process occurs instantaneously as the data is captured, meaning no identifiable information is ever stored or analyzed. Second, the focus of the research is on aggregate behavior patterns rather than individual actions, further distancing the study from privacy concerns.

4.2.2 Ethical Use and Data Protection

Ensuring the ethical use of technology extends beyond privacy considerations to include the responsible handling and protection of any data generated by the system. Although the data is anonymized, we are committed to maintaining high standards of data protection. This includes secure data storage, limiting access to authorized personnel, and employing robust data management policies that comply with relevant data protection laws and guidelines.

The utilization of anonymization techniques also reflects our commitment to minimizing any potential impact on visitor behavior and the overall museum or aquarium experience. By ensuring that the tracking system is unobtrusive and does not compromise privacy, we aim to maintain the integrity of the visitor experience, allowing individuals to engage with exhibits without concern for their privacy.

4.2.3 Transparency and Accountability

While the technical approach effectively addresses privacy concerns, maintaining transparency about the use and purpose of tracking technologies is still essential. Information about the tracking system and its privacy-preserving nature will be made available to visitors, ensuring they are informed about how data is used to enhance the visitor experience.

Furthermore, the project will adhere to an ongoing ethical review process, ensuring that all aspects of the research remain aligned with ethical best practices and respond to evolving technological and societal standards.

In summary, by prioritizing privacy through the use of image obscuration techniques and adopting a comprehensive ethical framework, this research aims to advance the understanding of visitor engagement in a manner that is both innovative and respectful of individual privacy rights. This approach sets a precedent for the ethical application of machine learning technologies in cultural institutions, balancing the benefits of visitor behavior analysis with the imperative of protecting privacy.

Notes

Tried to download/use model from Roboflow, but either image has to be sent to an API which would not retain privacy, or the device has to host an API itself to run the inference... Seems unlikely to be the most preferable solution, as the device would have to set up the service and run it locally. Possibly an interesting solution would be to do this with multiple devices. This supports the master-slave pattern of having multiple weaker computers and have them send to the stronger unit. Setting up private TCP connection between the weaker units and the strong unit and have the images sent to the stronger, so it can detect on them and send information etc... How many weak units do we need in order to make it profitable to have a strong GPU unit to do the processing? This whole systems sounds to be complicating processes, not making the product modular and easy-to-use. Includes a lot of connection/networking to make the weaker units find and connect to strong, physically close device. This task would mean setting up a strong device to host a network to which the weak units might connect to, and send images to. The issue is whenever images are sent, a lot of transmission is used... But the model takes image input size of 416x416. Would it be similar to just downscale the image before sending, or would this give the model less detail to work with?

Will now run several models on datasets from the web, i.e. the CrowdHuman dataset to see their accuracies. Will then deploy the models to device in acquarium to see if the best-performing model is an option in terms of size and inference speed. If it is preferable, I will attempt to increase it's accuracy by accumulating and annotating a specialized dataset for that setting, and training the final layers on the data. Can this be done with a

Bibliography

Bjørgen, H. E. (2023). *Welcome to the era of privacy preserving edge computing devices*. https://youtu.be/gR_ZGTDwFTY (cit. on p. 1).