AI-Driven UX for Video Object Annotation

by

Md Fazla Rabbi Alam

Supervisor: Duc Tien Dang Nguyen

Thesis

Submitted for the degree of Master in Media and Interaction Design

Department of Information Science and Media Studies

University of Bergen

December 2021

Contents

Acknowledgments

It has been quite a learning experience, a challenge and pleasure working on my thesis. My sincere gratitude to my supervisors Prof. Duc Tien Dang Nguyen for guiding me through this process and providing constructive feedback, new ideas and perspectives. Thanks to my cordial collaborators, Steinar Søreide, CTO and Andreas Teigland Whiteley, Lead developer at Mjoll AS for navigating through the complete design, test and development workflow. Last but not least, I am also grateful to my wife and beloved son for their eternal motivation and support that has given me the strength to continue on and finish my research work.

Abstract

Annotating large sets of image and video data is the elementary task in multimedia information retrieval and computer vision applications. The aim of annotation tools is to relieve the user from the burden of manual annotation as much as possible. In order to achieve this ideal goal, making the annotations workflow as fluid as possible many different functionalities are required . *Motivated by the limitations of existing tools, I have proposed an interactive semi-automated object annotation workflow which is intended to reduce users cognitive load by applying user centered design principles.* This makes the workflow fluid and suitable to be used in different domains. I have incrementally designed, tested and calibrated the workflow based on the user centered design principles. ***A quantitative and qualitative evaluation of the proposed workflow demonstrates that the use of the user centered design principles and semi-automatic modality can potentially reduce human cognitive load by at least one order of magnitude, limiting the user interaction choices and generating visual cues.*** Furthermore, the findings also indicate that user centered design principles help to structure UI components logically and strategically to guide users towards performing desired actions efficiently. However, limiting the interaction choices might have a side effect of lower precision annotation. My contribution to this thesis is to introduce a simple three step hybrid video object annotation workflow to reduce users' cognitive load, adopting user centered design principles following an incremental designing and testing approach.

Chapter 1 Introduction

Video is one of the most prevalent forms of visual media. It is widely used to inform(news), entertain (film, tv series), educate (video lectures) and connect (video conferencing) us, as well as attract our interest via TV commercials and social media posts. Likewise, video is also a crucial modality for AI applications such as self-driving cars, security applications, and patient monitoring in healthcare. One of the fundamental tasks common to those applications is the ability to detect and track objects across the duration of the video. Annotating large sets of image and video data is the elementary task in multimedia information retrieval and computer vision applications. Manual annotation of video object datasets requires an immense amount of human effort due to the dynamic nature of video data. Annotating videos are prohibitively time consuming as labeling only a single object in a single frame can take up to a minute [7,8]. This fact perhaps presents a major roadblock for video content editing and producing high quality interactive video contents. Thus reduction of human annotation costs is an active research topic in multimedia information retrieval and computer vision applications.

In recent years several video annotation tools have been developed with an aim to generate high quality ground truth visual datasets by reducing the human effort and improving the annotations quality. Some of the annotation tools proposed in the literature include computer vision and machine learning techniques that allows users to annotate objects efficiently [33,34,25,26], while others promote the use of crowd-sourcing based platforms to improve the quality of the annotations [15,38,39]. Sorokin and Forsyth [1] has made an influential observation on image labeling through crowdsourcing at a low cost. This approach has revolutionized static data annotation in vision, and since then enabled the affordable labeling for large-scale image data sets [2, 3, 4]. However, a similar approach does not hold true for video despite a corresponding abundance of data. This is due to the dynamic nature of video data which makes frame by frame labeling necessary but in-efficient(cognitive burden and time consuming) for manual labor. Therefore, to reduce the users cognitive load and decision making time, a new breed of solutions are essential to make the annotation workflow as automatic as possible with a smart and adaptive user interface. Thus, focusing on enhancing usability, I am ***designing a proof of concept for an efficient semi-automated video object annotation workflow combining AI and UX, considering the success and limitation of the existing annotation tools***.

User groups

The goal of this research is to improve the usability by simplifying the video object annotation workflow which allows the multiple user groups to build the ground truth dataset economically. The video object annotation user groups range from a novice individual to a group of professionals with the varied usage pattern from the basic usage to the power usage. However, usability remains important for all of them. Thus, it is important that users should feel immersed and in control of the tool which predicts their actions and helps them get things done properly and fast. Keeping that in mind, designing a smooth workflow can guide users and take very little effort to annotate objects and track them throughout the frames.

The primary user groups for the video object annotation tool are professional annotators, video editors and researchers. However with the growing trend of making and sharing video contents in social media platforms, opens up the window for a diverse interested user group for the tool. Taking that into consideration, the broader perspective of this research is to develop a generic adaptive workflow and design a simple learnable annotation interface that will guide the users to intuitively accomplish their tasks and make the system flexible so both novices and experts can choose to do more or less on them.

Research Questions

This project aims to better understand how AI and UX complement each other to improve the usability by simplifying the video object annotation workflow. I have defined a main research question and two sub-questions for this project, which will eventually help to answer the main question. The thesis will refer to the following research questions:

☐ ***RQ1: How state-of-the-art machine learning techniques complement user centered design principles to enhance workflow and usability by reducing users cognitive loads and decision-making time for economically annotating video objects?***

☐ ***RQ2: How does limited choice of interaction reduce the cognitive load and expedite the decision making?***

☐ ***RQ3: Scoping the number of perceived options on screen makes the workflow fluid and the interface more user friendly that allow users to accomplish the task efficiently.***

Purpose of the Research

The fundamental goal is to improve the usability such that the users should feel immersed and in control of the application and they should find it satisfying, if not delightful. In this master thesis, it is researched how AI can complement user-centered design to enhance usability for developing smart video object annotation workflow by reducing decision making time. In particular, recognizing the key areas where design principles can amplify user experience for annotating video objects by minimizing users cognitive load. This is a collaboration project between UIB and Mjoll AS, a Bergen based broadcasting solution provider. In this project Mjoll AS is contributing with relevant domain knowledge while I am harnessing that knowledge to design and test the proof of concept for simplifying video objects annotation workflow.

The objective of the project is to design a simple and intuitive video object annotation user interface applying user centered design principles and exploring various AI techniques to automate the object detection and tracking workflow. Taking the inspiration from "*Hick's law for choice reaction time"*[40] and "*Nielsen Heuristic"*[54], this work aims to leverage state-of-the-art machine learning object detection and tracking techniques, visual perception, visual hierarchy, proximity, contrast and balance to reduce users cognitive loads and decision-making time for economically annotating video objects. Essentially, this workflow will enable users to smartly implant descriptive metadata to the video assets to augment the intelligence of the content.

Short Description of the Prototype

Users react extremely fast, encountering an interface. Their eyes follow predictable reading paths and prefer recognition over recall. Considering these, I have envisioned to scale the most important elements to make the most important information prominent and unmissable for users as they try to achieve goals in their individual contexts by reducing their cognitive load.

To reduce users' cognitive load while annotating objects, I am envisioning a fluid workflow adopting user centered design principles, to be precise "Hick's law for choice reaction time"[40] and "Nielsen Heuristic"[54].

The prototype is designed to be a web application. It is a simple three steps object annotation workflow covering *Nielsen Heuristics[54], visual hierarchy and color and contrast. Key attributes of the app are given below.*

- Web based easy to use video object annotation tool
- Intuitive user interface
- Easily upload files by drag and drop.
- Use object detection models for automatic object detection and tracking
- Single object selector(rectangle)
- Highlight select object
- Simple and optimized interface for video annotation
- Save and export annotated data

Video object annotation prototype V1.0 link :
https://xd.adobe.com/view/45c0fbea-3ba8-42d0-a9c1-b2227eb8c62b-1093/?fullscreen

Research Contributions

Table 1 demonstrates a summary of my key contributions to this research.

Table 1: Key contributions
□ A simple three steps video object annotation workflow.
□ Implemented and tested incremental designing and testing approach.
□ Introducing a hybrid(combining AI and UX) workflow to reduce users' cognitive load while annotating objects.
□ Applied Hick's law for choice reaction time and Nielsen Heuristics to allow users performing video object annotation efficiently.
□ Conducted usability testing to select and structure UI components logically and strategically to guide users towards performing desired actions efficiently
□ Demonstrated user centered design principles and semi-automatic modality reduces human cognitive load by at least one order of magnitude

Thesis Outline

This thesis contains five chapters. Following this introduction, Chapter 2 is a literature review that includes central topics related to image annotation, video object annotation, cognitive load, decision making time based on Hick's Law, Nielsen Heuristics as well as the design methodologies. Lastly, there is a brief review on the other related applications. Chapter 3 discusses the research methods used in this project that includes the research framework, user-centered design, development methodologies and research ethics. Chapter 4, walks through the design, development and evaluation iterations of the application. Chapter 5 is the concluding chapter that summarizes research findings and provides propositions for the future development. Following the main chapters, there is a collection of appendices, which contain supplement documents related to the research work.

Chapter 2 Literature Review

Background Literature Review

With the rising popularity and success of massive data sets in vision, the research community has put considerable effort into designing efficient visual annotation tools. However due to the time-consuming nature for video object detection and tracking, various strategies have emerged to facilitate the annotation task.This chapter briefly reviews related work in designing image and video annotation tools, designing principles, user cognitive loads, Hick's Law and Nielsen Heuristics.

Image Annotation

As Artificial Intelligence (AI) and Machine Learning (ML) are bringing light to progressive technologies, availability of ground truth training dataset is becoming crucial to enhance the performance of the ML algorithm. To enable machines to perceive objects in their natural surroundings, annotated images are required to train the algorithm to learn and predict correctly.

Essentially, image annotation is the most prominent technique used to develop ground truth datasets in computer vision research.

Deng et al. [2] introduced a crowdsourced image annotation pipeline through ImageNet. Torralba et al. [3] presented LabelMe as an open platform for dense polygon labeling on static images. Everingham et al.[5] describe a high quality image collection strategy for the PASCAL VOC challenge. Von Ahn and Dabbish [6] and Von Ahn et al. [9] discovered that games with a purpose could be harnessed to label images. Ramanan et al.[10] shows exploiting temporal dependence in video can automatically generate a static faces data set. Welinder et al. [11] came up with a quality control mechanism for annotation on crowdsourced marketplaces. Vittayakorn and Hays [12] describe quality control measures without collecting more data. Endres et al. [13] study some of the challenges and benefits of building image datasets with humans in the loop. Yet, the similar approaches which assist and motivate users to annotate static images do not apply to dynamic videos, since temporal data is difficult to visualize and edit.

Video Object Annotation

Likewise, significant effort has been made to develop tailored interfaces for video annotation. Yuen et al. [14] proposed LabelMe video, a web-based platform for obtaining high-quality video labels with arbitrary polygonal paths using homography preserving linear interpolation. It can also generate complex event annotations between interacting objects. Mihalcik and Doermann illuminate ViPER[15], a flexible and extensible video annotation system designed for spatial labeling. Huber [16] designed and described a simplified video annotation interface. Ali et al. [17] discussed FlowBoost, a video annotation tool which annotates videos from a sparse set of key frame annotations. Agarwala et al. [20] emphasized on using a tracker as a more reliable, automatic labeling scheme compared to linear interpolation. Buchanan and Fitzgibbon [23] proposes efficient data structures for interactive video tracking. Fisher [28] describes the labeling of human activities in videos. Smeaton et al. [29] discuss TRECVID, a large benchmark video database of annotated television programs. Laptev et al. [31] further testify that using Hollywood movie scripts can automatically annotate video data sets.

While all the literature emphasized on the utility, none of them addressed the usability of the annotation application. All the above literature indicates object detection and tracking are the two most crucial and time consuming tasks for a video object annotation tool. Thus, existing tools are perhaps effective in building large data sets but they are not necessarily user-friendly. In order to scale up to the next generation data sets, a smarter workflow is needed that can annotate high quality, sizable videos without exhausting users. So to optimize the annotation workflow, Amazon Rekognition[50] service can be rendered to automate the object detection and tracking throughout the video and leverage different design principles to reduce users cognitive load and annotation time.

Designing for users

Design has always been around and has evolved with humans for centuries. One of the definitions of design is explained as “an outline, sketch, or plan, as of the form and structure of a work of art, an edifice, or a machine to be executed or constructed.”[18]. Humans have always manipulated the environment around us, shaping it into objects that make sense for us either functionally or aesthetically. Design is everywhere, from the chair we are sitting on, the road we drove to get here, and the coffee machine we use every day. Three main design disciplines; human-computer interaction (HCI), interaction design (IxD), and user experience (UX) design are discussed in this section. These terms came along as computers became part of our professional and private lives. Engineers and researchers have paid significant attention to how computers should be designed for optimal human interaction.

Human-Computer Interaction

The first field to grow out of this research field was human-computer interaction (HCI) [21]. According to Preece et al. “HCI is a multidisciplinary field of study focusing on the design of computer technology and the interactions between the users and computers” [24]. While initially concerned with computers, HCI has since expanded to cover almost all forms of information technology design [25].

Interaction design

Interaction design (IxD) on the other hand, is mainly used today to describe the interaction between humans and products. The Interaction Design Foundation (25) describes IxD as the design of interactive products and services in which a designer’s focus goes beyond the item in development to include the way users will interact with it. Thus, assessing the users’ needs, limitations and contexts, enables designers to customize output to attain specific demands. It is a broader term than HCI, because it does not limit its use to computer interaction. Preece et al. [24] describe the main difference between HCI and IxD to be the scope. IxD casts a wider net, concerning the theory, research and practice of user experience for the manner of products, systems, and technologies. HCI has a narrower focus traditionally, by focusing on design, implementations, and evaluation of interactive computer systems for human use [24]. IxD is concerned with designing any interactive product to support the way the user interacts and communicates and has a pleasant experience. IxD is about creating the user interface (UI). The UI should be designed to create a good dialog between a product and the user, and the connection is the interaction one is designing for. In addition, IxD also focuses on selecting the right elements to include to make the product useful and effective.

User Experience design

In interaction design, the user experience is fundamental. Donald Norman introduced User Experience design (UX) into the research field in the 90' when he was working in Apple [19]. He considered the term interaction design insufficient to explain all the variables regarding what the user perceives. According to Donald Norman [26] "I invented the term because I thought human interface and usability were too narrow. I wanted to cover all aspects of the person's experience with the system including industrial design, graphics, the interface, the physical interaction, and the manual". Designing for how the user will perceive the product or service has come to be referred to as user experience (often abbreviated UX design) and is defined by Norman and Nielsen [19] as follows: "User experience encompasses all aspects of the end-user's interaction with the company, its services, and its product". UX design is about creating and shaping the experience the user receives. It includes all aspects of the experience: physical, sensory, cognitive, emotional, and aesthetic. Preece et al. [24] points out an essential factor in UX design; one cannot design an user experience, one can only design for a user experience. When designing for the experience, it is about putting the user first in every step of development; starting with mapping what they need, what they prefer, how they prefer it, their pain points, making it enjoyable and so on. UX design is all about knowing the user and encompasses all subfields while developing to reach the goal of having a satisfied user.

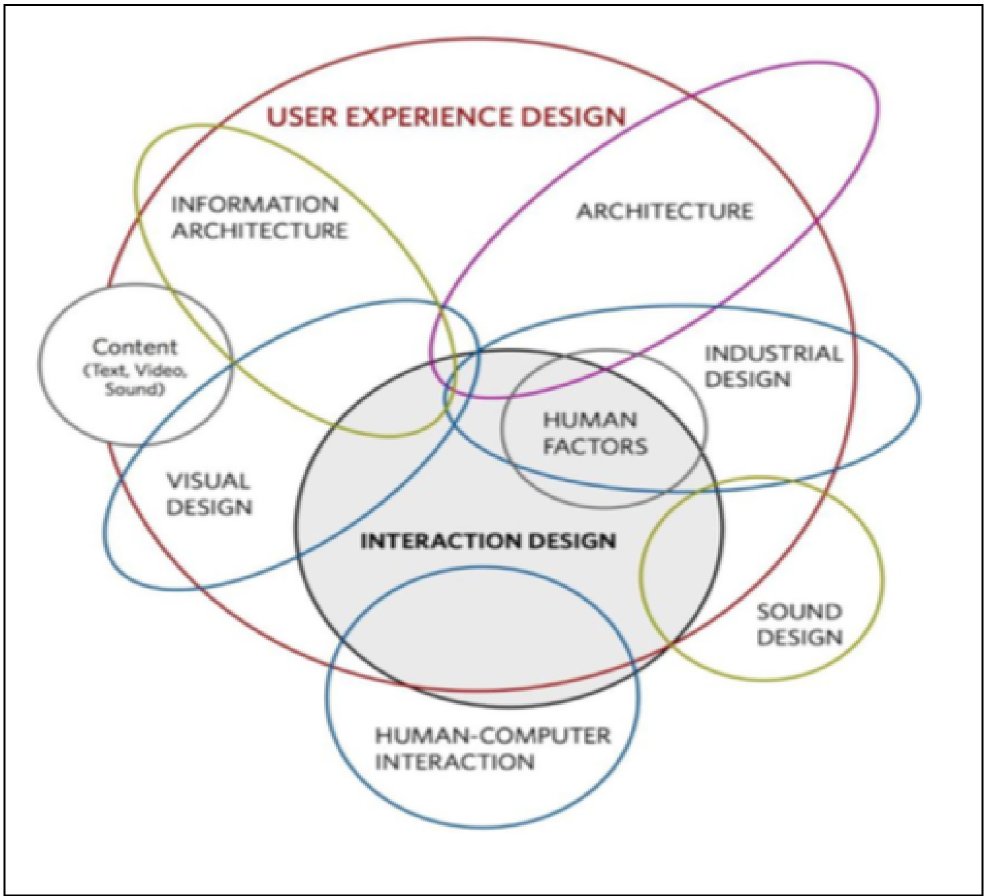


Figure1: The disciplines surrounding interaction design [21]

Which of these fields that are subsets of another is discussed widely, nevertheless there is no global definition of the difference of the terms [25]. I, therefore, decided to use the way Preece et al. [24] differentiate between the terms IxD and HCI, and divide the terms by the amount of subfields underlying them, and put UX design at the top of the hierarchy. Dan Saffer [21] published a diagram of the disciplines in his book “Designing for interaction” that shows the overlapping of the fields (Figure.1). In this model he shows that most of the disciplines fall at least partially under the umbrella of user-experience design, the discipline of looking at all aspects visual design, interaction design, sound design, and so forth of the user’s encounter with a product, and making sure they are in harmony [21].

Cognitive Load

Cognitive psychology is the study of mental processes such as memory, perception or problem solving. Broadly put, cognitive psychology deals with how people think, which is the key element to understanding the user's perception. By paying attention to such mental processes, it is perhaps possible to reduce the amount of mental processing power people need when using a product.

In psychology, cognitive load refers to the mental effort, which is required to learn new information[59]. From the UX design perspective, cognitive load is the mental processing power needed to use a product. The amount of mental processing power or total cognitive load required to use an application, affects users' tasks compilation efforts. If the amount of information that needs to be processed exceeds the user's ability to process it, the overall performance suffers. The cognitive load is too high. So how to deal with this? Users' actual processing power can not be changed. But it is possible to get to know users' limits, and use that to minimise their processing efforts by guiding them throughout the process. Thus I am harnessing Hick's law for choice reaction time[40] and Nielsen Heuristics[54] to design the video object annotation workflow. While Hick's law will narrow down big volumes of information without overloading the user, Nielsen Heuristics will ensure the usability and efficiency.

Hick's Law

Achieving a delightful user experience, first requires to find out the functionalities that will answer user needs; second, to navigate them to the specific functions they need the most. If users struggle with the decision-making process, they may become confused, frustrated, or leave the app. Hick's law for choice reaction time predicts that the time and the effort it takes to make a decision, increases with the number of options.

Hick's Law (or the Hick-Hyman Law) is named after a British and an American psychologist team of William Edmund Hick and Ray Hyman. In 1952, this pair set out to examine the relationship between the number of stimuli present and an individual's reaction time to any given stimulus[40]. Hick's Law describes the positive correlation between time and the offered choices. The time a user takes to make a decision as a result of the possible choices he or she has. Thus increasing the number of choices increases the decision making time logarithmically[40].

Hick’s Law is applicable to any simple decision making that offers multiple options, precisely in a control system environment. In our life when sudden situations arise and alarms are triggered we need to be able to make quick decisions. In such situations we enter the stress zone and get tunnel vision. If we combine that with the other body senses, suddenly it can turn into a critical situation. Thus when response time is critical it is wise to keep the choices to a minimum that speed up the decision making.

Likewise, from the user interface design perspective, in the milliseconds after a person encounters a new app, millions of neurons fire and the brain makes hundreds of subconscious decisions and form aesthetic reactions to the UI within the first 17 to 50 milliseconds after exposure[59]. These impressions might not register in our memory, but they do impact behavior. Thus a cluttered UI can get the user confused to make decisions. In such a situation Hick’s law can be useful to narrow down big volumes of information without overloading the user by presenting specific parts of that process at any one time on the screen.

I am expecting scoping the number of perceived options on screen makes the workflow fluid and the interface user friendly that allow users to accomplish the task efficiently. Thus to design the object annotation workflow I am breaking down choices to small chunks and presenting fewer and clearer options at a time in the user interface. For example, presenting a single object selector and highlighting the selected object can speed up the response times.

Nielsen Heuristic

The word “heuristic” defines a method or process to detect inconsistencies and find solutions for them in a digital product. Detecting early errors is an important step in the UX process as it ensures the usability and efficiency of an application. Heuristic analysis identifies strengths and weaknesses of a workflow and proposes recommendations to ensure the fluidity for structuring a good user experience.

Computer scientist Jakob Nielsen, known as the usability kingpin developed 10 principles for evaluating the usability of user interfaces which is known as Nielsen’s heuristics. These principles define important elements for the user interface composition that should be considered while creating layouts. According to Jakob Nielsen[54] *“A wonderful interface to the wrong features will fail.”* Thus, it has become practical rules for all human-computer interaction and serves as an usability evaluation guideline for professionals.

User interface is one of the most used means of communication between a human being and a machine in the digital world. It generates visual cues to instruct users to perform their actions efficiently. While certain interfaces catch users eye and get their blood pumping others make them confused and increase their cognitive load. Poorly designed user interfaces generate noise in the communication that possibly leads to insecurity and stress for the users. Thus, it is essential that the design in its entirety is considered before, during, and after the development, bringing a simpler and clearer direction for the user to perform tasks.

The more pleasant and fluid the usability (user experience), the greater their efficiency. To achieve this goal for my object annotation interface I have drawn on Nielsen's heuristic evaluation which is presented below (Table: 1) to ensure a user-friendly simplified workflow.

Table 1: Nielsen's heuristics for user interface design

H1	Visibility of system status
H2	Match between system and the real world
H3	User control and freedom
H4	Consistency and standards
H5	Error prevention
H6	Recognition rather than recall
H7	Flexibility and efficiency of use
H8	Aesthetic and minimalist design
H9	Help users recognize, diagnose,and recover from errors
H10	Help and documentation

Similar Applications

Computer vision algorithms require high quality annotated data for a deeper understanding of the actions and interactions of different objects (individuals and groups) in every single video frame. This is beyond just identifying the name and location of the object, as is the case with image annotation. Over the last few years, diverse video annotation tools have been developed with different functionalities to generate ground truth of large scale visual datasets for machine learning models. Most of the tools proposed in the literature include computer vision and machine learning methods that allows users to annotate efficiently while others are promoted to improve the quality of the annotations using crowd-sourcing platforms. However the effective video annotation tool should be user-friendly and able to maximize annotation quality by minimizing human cognitive load.

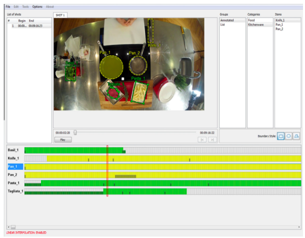
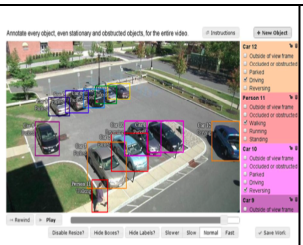
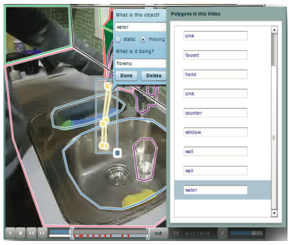
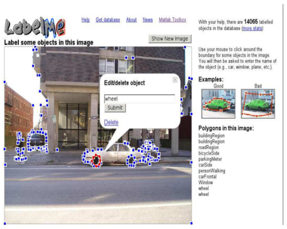
Here I have summarized the key features of Four (4) object annotation tools found in the literature. I have identified some properties that are considered important in an annotation tool. The properties refer to the tool's design, user interactions, supported functionalities and agility .

As it is seen from the following Table: 2 and Figure: 2, each tool possesses a set of important functionalities and properties but lacks others also important for the annotation task. That is why, I am proposing an interactive semi-automated annotation tool that supports easy user interactions during the annotation, and integrates different computer vision modules for object detection and tracking. User centered design principles will make this tool flexible and suitable to be used in different application domains.

Table 2: Key attributes of similar applications

Properties	iVAT[47]	VATIC[15]	LabelME V[14]	LabelME[3]
User Group	Individual	Individual	Individual Group	Individual
Annotation Type	States, Behaviors	States	States, Behaviors	States, Behaviors
Platform	Web based	Web based	Web based	Web based
Boundary Shapes	Ellipse Polygon	Rectangle	Ellipse Polygon	Ellipse Polygon
Interface	Manual, semi-automatic, and automatic annotations via user interaction with various	Optimized for video annotation	Responsive user interface	

	detection algorithms			
Annotation propagation	linear interpolation	linear interpolation	homographs	Not required
Agility	Automated tracking with interpolation for assisting manual annotation	Automatic quality assurance Flexible and suitable to be used in different application domains	Homography-preserving shape interpolation to propagate annotations temporally and with the aid of global motion estimation.	

			
iVAT[47]	VATIC[15]	LabelME V[14]	LabelME[3]
Figure 2: Similar applications interfaces			

Proposed Approach

While all the literature emphasized on the utility, none of them addressed the usability of the annotation workflow and UI. However, several of them [48,51,52] indicate the commonly used detection and annotation strategies are labor incentive and mentally exhausting. Moreover identifying objects and labeling them within a video frame is a time demanding task[53]. To reduce the cognitive load as well as the decision making time, I propose a semi-automatic approach which combines the best of both worlds i.e. the speed of machine learning algorithms and the accuracy of the human eye.

The goal of my work is to design a semi-automatic video object annotation workflow applying user centered design principles, to be precise Hick's law for choice reaction time[40] and Nielsen Heuristic [54]. I am considering splitting the annotation process into two steps: 1) automatically identifying and tracking objects and 2) Labeling the tracked object or new objects. I am introducing automation and human-in-the-loop interaction in both stages, aiming to achieve the highest level of labeling efficiency. Research on automatically identifying and tracking objects is wide. Since automatic tracking and segmentation is not my contribution I am scoping my work into improving usability for smart decision making processes.

I am presuming automatically object detection and tracking will reduce a significant amount of user cognitive load and decision making time for video object annotation. Similarly this process will allow users to act like a curator where they will be allowed either to edit existing labels or manually detect other objects. Thus I am intended to perform user testing by limiting user interactions with the interface which might result in a significant savings of time and effort.

A flexible user interface perhaps allows more powerful annotations, but at the expense of increased annotation effort. Thus to optimize the object detection and tracking process I am considering rendering service from Amazon Rekognition[50]. Amazon Rekognition object detection model is able to identify and track most common objects within a video. Users are allowed to edit or customize those object labels which will eventually feed back to the object detection model and the model will relearn from the correction.

Similarly the user will be able to identify and label new objects manually within the video frames which will be also automatically tracked throughout the video. However in the case of manual object identification and annotation the UI(user interface) will present limited options to the users by aiming to expedite the decision making[40]. For instance, the user can only use a rectangle to select an object instead of having multiple selectors and highlighting the selected object among the clutter. To achieve this goal I am considering Hick’s law for choice reaction time[40] and Nielsen Heuristic[54] to design the user interface. However this approach might have a side effect of lower precision annotation.

Chapter 3 Methodology

In this chapter, discusses the research methods used in this research. In addition to that, I have briefly explained the importance and requirements of the research ethics and the user consent procedures for this thesis.

Methodologies are step-by-step procedures to carry out the research and development activities in different phases of a system development life cycle. A methodology has its own procedures or techniques to support working principles and tools to generate the deliverables [8]. There is a collection of specific techniques and tools for a certain research and development methodology. This chapter discusses the research and development methods used for user studies, requirements collection and system development life cycle. User centered design (UCD) [10] is applied for user studies, requirements collection, evaluation and calibration while Agile development process [1] is used for functional prototype development. Moreover, these methods are fused to enhance the quality of the application and User Experience(UX) [11].

I am developing a web based video object annotation tool combining AI and UX. The core focus of this project is complementing AI with UX to develop smarter workflows by recognizing the key areas where AI can enhance user experience and vice versa and developing a functional prototype. Agile development process and user centered design principles (UCD) are integrated to develop the web based annotation application. To achieve that goal, this work aims to leverage state-of-the-art machine learning techniques, visual perception[12], visual hierarchy[11], proximity[11], contrast and balance[11] to reduce users cognitive loads and decision-making time for economically annotating video objects[13].

At present, the majority of information systems are web-based. Web applications rely on the web as its interaction medium with the end-users to create, exchange, and modify data for transaction requirements[8]. Though web applications live under the umbrella of software systems, they are exclusive regarding user recognition, user environment, communication control, security issues, interface requirements, feedback mechanism, functionality design, and life cycle[8]. As web applications are becoming increasingly important to all aspects of life, how to ensure the success of their research and development is an issue of interest and practical value to practitioners, educators, and researchers [9]. Considering that, I have combined the user centered design principles and Agile development process to conduct the research and development processes and achieve targets related to time, quality and user experience.

Development Methods

The integration of Agile and User-Centered Design(UCD) methods is a fundamental condition to improve the quality of software products and enhance the user experience.

Agile and UCD share the common objective of producing high-quality software although they address it from different perspectives. The intrigue in Agile-UCD has grown over time since the creation of the Agile development process[5]. Literature indicates the collaboration between the Agile development process and the User Experience (UX) can increase the success likelihood of a project by complementing each other [6]. Fundamentally, both approaches are recurring and human centered. User Centered Design (UCD) practices can improve the Agile process, providing structured ways to evaluate end-user requirements [6]. Similarly, the Agile process can improve User Centered Design (UCD) by providing frequent iterations that lead to continuous usability evaluations. The early feedback can be incorporated into the application quickly. The collaboration should include developers, designers, users, product managers, and business analysts. It has been recommended to include a Sprint 0 in the product development life cycle during which the initial user research is performed for UX design [6]. During this initial iteration, user stories should be created.

User Centered Design(UCD) and agile are two major development processes which ensure an application provides good user experience. Multifold benefits of Agile Software Engineering have led to it becoming a mainstream development methodology [1]. However, Agile alone does not necessarily address the usability of the application. Likewise, the need for a good User Experience (UX) has become more evident, and so efforts have been made to integrate usability practices from UX design into Agile. According to Nielsen and Norman, UX is a broad aspect that refers to all interactions that a user makes with a company, its products and services [2]. UX represents a family of user centric development approaches that prioritizes the user needs instead of the system. A common UX development approach is User Centered Design (UCD)[11].

The goal of the User Centered Design (UCD) is to enhance the usability, such that a user finds an interface easy to navigate. The term “usability” also refers to the methods, which can be used to improve the design of an interface [3]. Usability is considered an important factor for any application. A lack of usability increases the users cognitive load and reduces work efficiency[10]. Thus, when users encounter difficulties on a web application, they are presumably responding by abandoning the application [3]. Therefore, many practitioners have been propelled to find the compatible ways to integrate usability practices into applications which are developed through Agile [4].

Agile and User Centered Design (UCD) share some common goals, which can be considered as good starting points for an integration of the two[5]. However due to the time and resource constraints, I am only focusing on User Centered Design (UCD) principles for this thesis.

User-Centered Design

As opposed to working features, the priority of User Centered Design (UCD) and User Experience (UX) is user satisfaction. Significant resources are allocated for extensive user research at the beginning of the project [6]. The entire process is followed by design iterations, consisting of prototyping and evaluation. However, the iterations are longer than a typical Agile sprint.

User Experience (UX) design emphasizes specialized methods of end-user research before the application is developed [7]. Some widely used user research methods for gathering and understanding design requirements are: Focus Groups, Heuristic Evaluations, Comparison Study, User Interview, Observation study. To analyze design requirements, User Experience (UX) makes use of practices such as Personas and Scenarios. Moreover At the end of each cycle, UX designers conduct usability evaluations on the design with end-users. This process generates feedback on usability goals and calibrates accordingly.

For my project I have conducted a user observation study followed by a semi structured interview to understand the existing video object annotation workflow. Based on the study I have developed a Persona to identify the behavioral components of the users. Similarly, I have conducted research on the existing video object annotation tools to identify the strength and limitations of the existing tools.

Integrating UX and Agile

The integration of User-Centered Design and Agile methods is a rudimentary condition to refine the software products quality and enhance the user experience.

Agile and UX share the common goal of producing high-quality software although they approach this goal from different perspectives. The interest in Agile-UX has increased over time since the creation of the Agile development process [5]. Literature indicates the collaboration between the Agile development process and the User Experience (UX) can increase the success likelihood of a project by complementing each other [6].

Fundamentally, both approaches are recurring and human centered. User Centered Design (UCD) practices can improve the Agile process, providing structured ways to evaluate end-user requirements [6]. Similarly, the Agile process can improve User Centered Design (UCD) by providing frequent iterations that lead to continuous usability evaluations. The early feedback can be incorporated into the application quickly. The collaboration should include developers, designers, users, product managers, and business analysts. It has been recommended to include a Sprint 0 in the product development life cycle during which the initial user research is performed for UX design [6]. During this initial iteration, user stories should be created.

Evaluation Methods

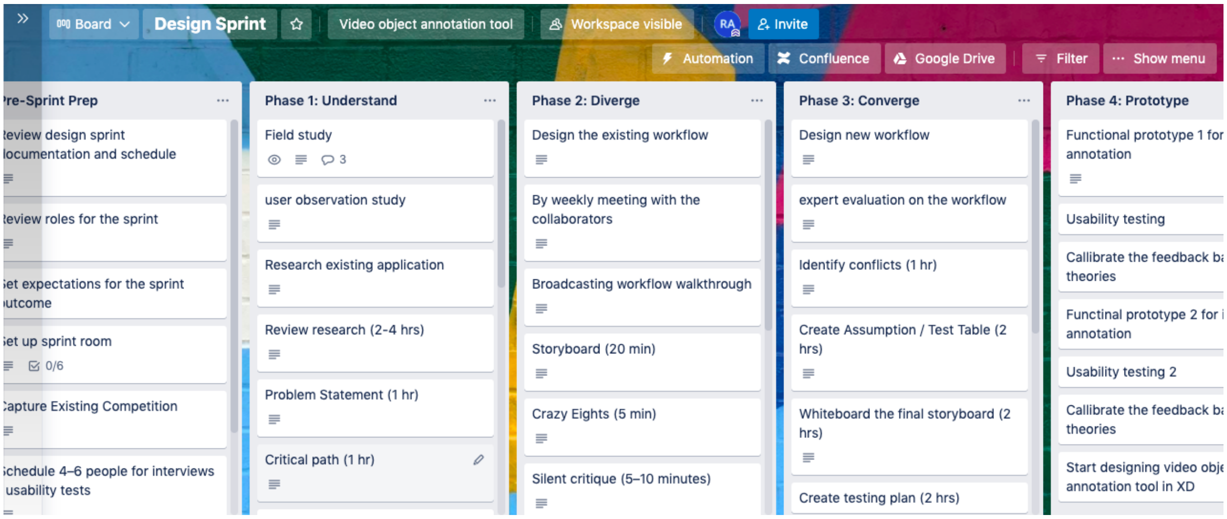
Evaluation and alteration evolve together in a user-centered design process. Thus, each iteration will allow me to take the design towards betterment, involving user feedback. There are diverse evaluation methods available to apprise my design. Among others, usability testing[26] and heuristic evaluation[25] methods will be useful to evaluate the utilities of the proposed solution from diverse perspectives.

An user-centered usability testing[26], will allow the potential users to test and evaluate the designs and functionalities of the prototype. Likewise, a heuristic evaluation[25] will include an expert insight from the alternative viewpoint. Since the potential solutions are very domain specific and hard to anticipate different integration perspectives with the existing workflow, it is worthwhile to evaluate the prototype with domain experts in addition to users. Both users and experts suggestions, criticisms and enhancements will result in revision of the prototype.

Research Methods

My master thesis is a collaboration project with Mjoll AS, a Bergen based broadcasting solution provider. Mjoll obtains comprehensive domain knowledge on the relevant broadcasting technologies, workflow and customer requirements. Moreover, they have a tradition of employing a user-centred approach to development. Thus, Mjoll is contributing with extensive domain and development knowledge while I am harnessing that knowledge to design and test functional prototypes for automating user centered video objects annotation workflow.

One of the senior developers from Mjoll AS is assigned to supervise this project, who holds expatriates equally in user-centred activities and development processes. After having the first meetup with him I have decided to parallelly work on brainstorming, research on existing tools, requirements formulation, prototyping and user evaluation(Figure:4). We have agreed for a by-weekly sprint. Due to the COVID, we could not meet in person instead we chose to use Google Meet for our main communication channel. Similarly, we have opened a discussion channel in Slack(Figure:5). Moreover, I have maintained a by-weekly meeting log in Google doc, where all the discussion and action points are documented(Figure:5). At the same time I have used the Trello project management tool as a Scrum board(Figure: 3) where all the backlogs, assigned tasks, on progress tasks and done tasks are maintained.



're-Sprint Prep
- ꞌeview design sprint ꞌocumentation and schedule
- ꞌeview roles for the sprint
- ꞌet expectations for the sprint ꞌutcome
- ꞌet up sprint room 0/6
- ꞌapture Existing Competition
- ꞌchedule 4–6 people for interviews usability tests

Phase 1: Understand
- Field study 3
- user observation study
- Research existing application
- Review research (2-4 hrs)
- Problem Statement (1 hr)
- Critical path (1 hr)

Phase 2: Diverge
- Design the existing workflow
- By weekly meeting with the collaborators
- Broadcasting workflow walkthrough
- Storyboard (20 min)
- Crazy Eights (5 min)
- Silent critique (5–10 minutes)

Phase 3: Converge
- Design new workflow
- expert evaluation on the workflow
- Identify conflicts (1 hr)
- Create Assumption / Test Table (2 hrs)
- Whiteboard the final storyboard (2 hrs)
- Create testing plan (2 hrs)

Phase 4: Prototype
- Functional prototype 1 for annotation
- Usability testing
- Callibrate the feedback bꞌ theories
- Functional prototype 2 for ꞌ annotation
- Usability testing 2
- Callibrate the feedback bꞌ theories
- Start designing video objꞌ annotation tool in XD

Figure 3: Trello scrum board

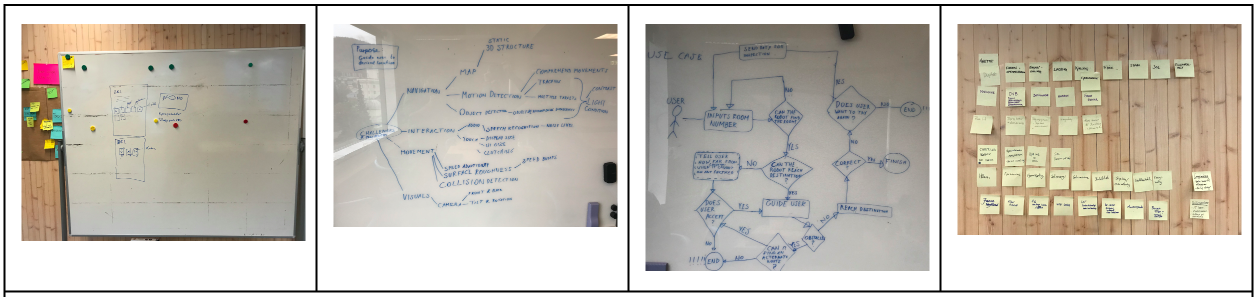
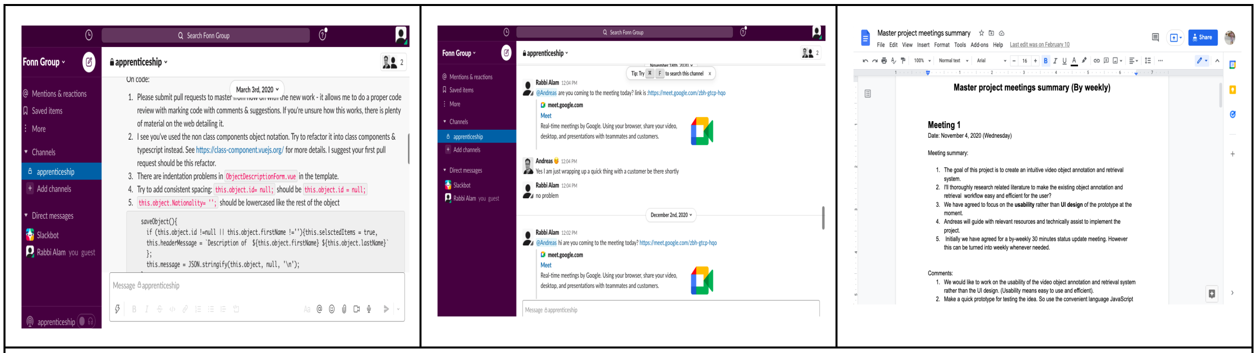


Figure 4: Brainstorming and workflow defining process



On code:

March 3rd, 2020

1. Please submit pull requests to master from now on with the new work - it allows me to do a proper code review with marking code with comments & suggestions. If you're unsure how this works, there is plenty of material on the web detailing it.
2. I see you've used the non class components object notation. Try to refactor it into class components & typescript instead. See https://class-component.vuejs.org/ for more details. I suggest your first pull request should be this refactor.
3. There are indentation problems in `ObjectDescriptionForm.vue` in the template.
4. Try to add consistent spacing: `this.object.id= null;` should be `this.object.id = null;`
5. `this.object.Nationality= '';` should be lowercased like the rest of the object

```
saveObject(){
  if (this.object.id !=null || this.object.firstName !=''){this.selsctedItems = true,
  this.headerMessage = `Description of  ${this.object.firstName} ${this.object.lastName}`
  };
  this.message = JSON.stringify(this.object, null, '\n');
```

November 18th, 2020

Rabbi Alam 12:04 PM
@Andreas are you coming to the meeting today? link is :https://meet.google.com/zbh-gtcp-hqo
meet.google.com
Meet
Real-time meetings by Google. Using your browser, share your video, desktop, and presentations with teammates and customers.

Andreas 12:04 PM
Yes I am just wrapping up a quick thing with a customer be there shortly

Rabbi Alam 12:04 PM
no problem

December 2nd, 2020

Rabbi Alam 12:02 PM
@Andreas hi are you coming to the meeting today? https://meet.google.com/zbh-gtcp-hqo
meet.google.com
Meet
Real-time meetings by Google. Using your browser, share your video, desktop, and presentations with teammates and customers.

Master project meetings summary (By weekly)

Meeting 1
Date: November 4, 2020 (Wednesday)

Meeting summary:

1. The goal of this project is to create an intuitive video object annotation and retrieval system.
2. I'll thoroughly research related literature to make the existing object annotation and retrieval workflow easy and efficient for the user?
3. We have agreed to focus on the **usability** rather than **UI design** of the prototype at the moment.
4. Andreas will guide with relevant resources and technically assist to implement the project.
5. Initially we have agreed for a by-weekly 30 minutes status update meeting. However this can be turned into weekly whenever needed.

Comments:
1. We would like to work on the usability of the video object annotation and retrieval system rather than the UI design. (Usability means easy to use and efficient).
2. Make a quick prototype for testing the idea. So use the convenient language JavaScript

Figure 5: Communicating with the collaborator

The project is divided into two parts(Figure: 6). During the first part, I have conducted user study and researched similar object annotation tools to identify the pros and cons of the existing tools, such as workflow, usability, accessibility, learnability and so on. These findings are then used as a framework for a more in-depth investigation and formulate functional requirements. Then I have developed several low-fi prototypes using pen-papers, storyboard and AdobeXD and evaluated them with heuristic evaluation and user testing. The entire research and development processes are iterative followed by agile and user centered design principles. I envisaged the Listen-Solve Problem-Develop project lifecycle which demonstrates the following steps in Table:3 for my project. Likewise I have used usability testing[26] and heuristic evaluation[25] as evaluation methods.

Table 3: Project lifecycle
□ A separate “up front” period of user research and initial requirements gathering.
□ An iterative prototyping stream where the functional requirements are reviewed in the Sprint planning.
□ Iterative usability testing with constant feedback throughout the development phase.

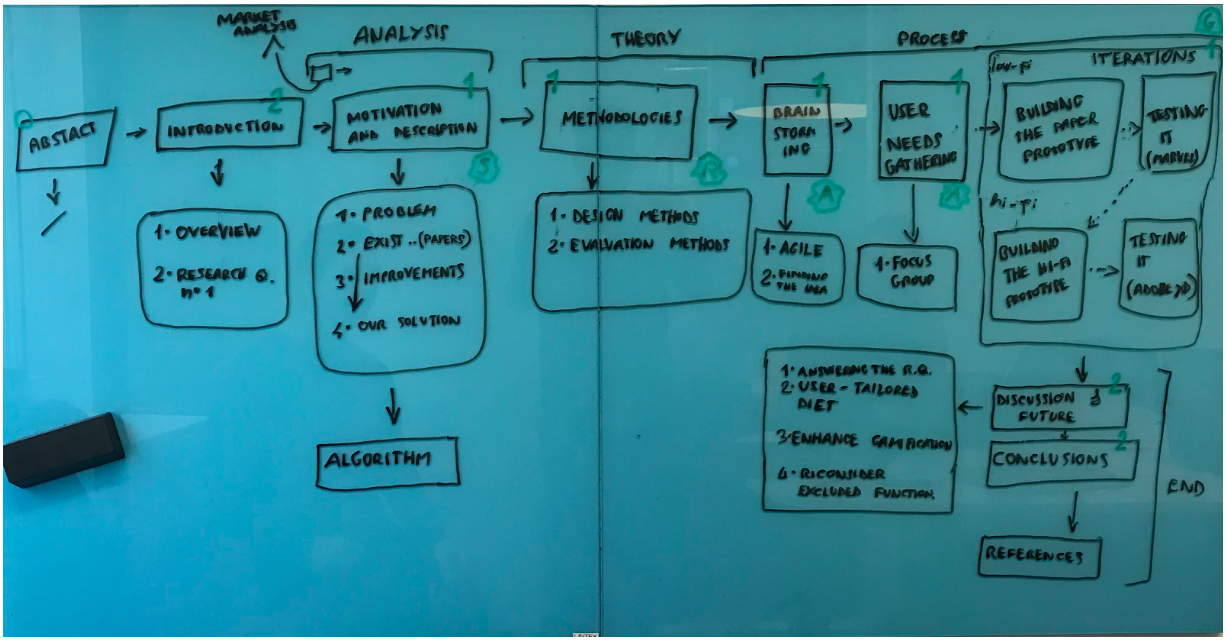


Figure 6: Workflow of the project lifecycle

Research Ethics

It is an important obligation for a researcher to protect their research subjects and the data. Thus, while conducting the research, maintaining good research ethics has been my key priorities. The primary goal of good research ethics is about protecting the subjects and their data. This includes being open about the goal of the research, and what I am trying to accomplish out of it.

Safe Research

My first priority was to apply for approval to the Norwegian Centre for Research Data (NSD) to conduct the research. This is important to ensure the research topic is safe, and the methods are used appropriately. I followed their guidelines on what to include in the consent form and how to collect, store, and plan for data handling (Norwegian Centre for Research Data, 2018). In my application to NSD described how and why to conduct the field study to ensure the General Data Protection Regulation (GDPR). GDPR is a set of rules for the protection of the users' privacy and right to their personal data, that regards everyone who is handling personal data (European Commission, 2019).

Consent

Prior to the user study, interviews, and usability tests, the respondents were given a consent form to read and sign before we proceeded. The consent forms were customized to each research method. It included an explanation of the research project, why the research is conducted, how to use the data, and how to ensure their data's security. Participants had time to read it properly and ask questions before they signed the consent to contribute.

In the observation study, I gave information and received consent orally as I did not collect any personally identifying information. The reason I chose not to ask for signatures on paper as with no personal information saved, signing the consent form is considered an unnecessary complication concerning the recruiting process. However, for the usability testing I have provided them with a written form describing the research and data protection policies and took consent from them before the test began. Users anonymity is preserved and no demographic data is collected.

Chapter 4 Solutions and Evaluations

Designing solutions

The following chapter presents the design iterations for the prototype development and discusses different iterations conducted throughout the research process. The iterations indicate the different phases of the project.

Prototypes

In this research, different prototypes have been developed to present the concept of image and video object annotation tools to minimizing cognitive load and maximize usability by combining AI and user centered design principles.

Prototypes are a pivotal part of the design process and a practice applied in all design disciplines. A prototype is a process to evaluate and validate a concept by putting an early version of the solution in front of real users and collecting feedback as quickly as possible. According to Preece et al. a prototype is an early sample or model of a product created to test a concept or process[24]. The purpose of a prototype is to design a tangible model for the potential solutions and validate the concepts instead of going through the entire development cycle. It is an iterative process that allows the designers to refine the proposed solutions based on real user feedback.

There are several types of prototypes; low-fidelity, mid-fidelity and high-fidelity where the fidelity of the prototype refers to the level of details and functionality built into a prototype[24]. A low-fidelity prototype is considered as a basic representation of a concept which allows validating the concept early in the design process. It generally has limited function, limited interaction, and prototyping efforts. According to Rudd et al. “low-fidelity prototypes are developed to demonstrate concepts, design alternatives, and screen layouts, rather than to model the user interaction with a system” [56]. It is a fast, simple and affordable way of validating a concept. In this project, I started with drawing low-fidelity wireframes on paper to quickly demonstrate the potential functionalities and intended behaviors of the application. A wireframe is a conceptual model of the potential look and workflow of the application [24]. Due to these attributes, low-fidelity prototypes are appropriate for evaluating the concept of the application at a very early stage.

On the other hand, a high-fidelity prototype is a visualization of the concept or product of higher complexity. According to Rudd et al. a high-fidelity prototype is functional and interactive, so it can be user-driven and has a navigational scheme[54]. The high-fidelity prototype is supposed to look and feel like the final product so that it can be used for exploration and testing. Whereas, a mid-fidelity prototype is somewhat in-between the low- and high- fidelity prototype. In my case, it had the digital wireframes but did not include navigation for interactivity.

In this project, I used the digital prototyping tool Adobe Experience Design (Adobe XD) [37] for creating a mid- and high-fidelity prototype. I decided to use Adobe XD because it is free and efficient in making interactive interfaces. Adobe XD is a UX/UI design and collaboration tool, among the few free softwares Adobe System delivers[55].This is a wireframing and prototyping tool to create and test interactive prototypes. In Adobe XD, it is possible to simulate a real web or mobile application by linking different sketches. It provides simple scaling and editing of elements which makes the development faster. Moreover, with a cloud-based system, XD enables quick sharing for collaboration and usability testing[55]. Thus I have chosen Adobe XD for this project to develop different prototypes.

Design Iterations

The video object annotation tool design process is evolved through four iterations (Table 4) from an idea sprint to find a concept, outline the workflow on paper, implement it digitally and eventually make it interactive. The evaluation of each iteration is used to refine the next iteration. In the video object annotation tool design process, a comprehensive user study has been conducted at the first iteration to identify a user need and have generated ideas to meet that need. A low-fidelity prototype is developed to conduct a concept test and validate the idea in the second iteration. Finally, a high-fidelity prototype is constructed and evaluated based on the feedback from the previous iterations.

Table 4: Design iterations

Iteration	Objective
One	Define the concept. Review the existing workflow User observation study

Two	Develop an improved workflow Develop a Low-fidelity prototype Evaluate the concept.
Three	Develop a Mid-fidelity prototype Defining design requirements.
Four	Develop a Hi-fidelity prototype. Evaluate the prototype with usability testing or heuristic evaluation.

Design Iteration one - Defining the concept

In the UIB Masters thesis concept pitching seminar, Mr Steinar Søreide, CTO, Mjoll AS, has stressed the importance of smart techniques for extracting descriptive video metadata in the fastest changing broadcasting workflow. The concept of increasing necessity for rich video metadata in interactive and intelligent video contents has instigated the rolling of this research project. The initial research idea begins with the focus on harvesting descriptive metadata for video assets. However, after reviewing relevant research works and several ideation iterations the initial idea is boiled down to design a video object annotation workflow emphasizing on the usability.

As my contribution to this project is to enhance the usability, an observation study can be an efficient technique for a deep understanding of users' contexts. Usability is a measure of how well a specific user in a specific context can use a product or design to achieve a defined goal effectively, efficiently and satisfactorily. In the first instance, the primary focus is to understand the importance of the descriptive metadata in the broadcasting workflow. Thus, a scenario based digital observation study has been conducted with a professional video editor to perceive the notion of metadata for interactive video content creation. Furthermore, a persona has been created based on the data absorbed from the study.

The primary idea of developing a simplified video object annotation workflow initiated from the perspective of rich video assets for the broadcasters. One of the important tasks of journalism has always been storytelling. In today’s fast-paced, content-everywhere world, broadcasters and journalists need innovative and easy-to-use storytelling media workflow that creates distinctive and cutting-edge video content to increase audience engagement in a wide range of platforms. The broadcasting industry is experiencing comprehensive change due to the shifting audience and consumption patterns fostered by the diffusion of the Internet. Delivering an engaging experience to the viewers, broadcasters must produce high quality video contents, which requires a high level of interactions with the elements embedded within video scenes. Thus, acquiring comprehensive information about the scene is essential for achieving those interactivity. Specifically, if the information about relevant objects present in a scene is known, content creation and distribution can achieve new heights of efficiency. Findings from an empirical study conducted by Kallinikos, J. et. all within the British Broadcasting Corporation (BBC), indicates the usefulness of rich metadata by stressing that, “Descriptive video metadata rises to be an important coordinate medium that provides the cognitive resources for identifying and managing video content within and across the workflow”[58].

However due to the dynamic nature of the video asset it is a complex and time consuming task to annotate video objects and track them throughout the frames. Thus a simplified workflow is essential to reduce the users cognitive load and time to annotate video objects by keeping the users in the center throughout the development process. Thus the entire research process kicks off by investigating some of the existing video object annotation tools and theories behind their workflows. Primary research leads to an observation study followed by a semi-structured interview which allows to develop a user persona and initial requirements for the tool.

Observation Study

To validate the idea of necessity of video object annotation in the broadcasting workflow I have decided to go through an observation study with the professionals which is in fact the first formar step of this research work.

As previously mentioned, this research project was rolled out focusing on finding smarter ways to augment intelligence to the video assets for the broadcasting workflow. Keeping that in mind I started to read through relevant research to understand the ground concept from theoretical perspectives. Research indicates, descriptive video metadata is essentially turning to be the key coordinator for providing the cognitive resources to identify and manage video contents within and across the workflow[58]. However due to the dynamic nature of the video assets, extracting embedded descriptive information from the video and tracking them throughout the frames is a complex and tedious task. Thus a smart workflow is required to ease the process for augmenting descriptive video metadata. To achieve that goal, it is important to have a clear overview on the existing metadata extraction workflow practiced by the broadcasters to find out how well my intended design potentially flows in the context. From that perspective the entire research process was kicked off by conducting a digital observation study on a professional video editor which allowed to develop a user persona and initial usability requirements for the proposed workflow.

My initial plan was to conduct at least three (3) on-site observation studies on professional video editors but due to the pandemic situation, it was difficult to find interested participants and could not manage to conduct the study in a natural setting. Conducting observation study in the natural setting is an important medium of capturing the user context; as such environmental influences on the user behaviour(distractions, cognitive load etc). Likewise, leveraging a deep understanding of users' contexts ensures high usability by guiding users through the easiest and least labor-intensive route. Therefore, I tried several methods(personal contacts, snowball methods, contacting different broadcasters) to recruit participants. But due to the unconventional nature of the user group(professional video editors from the broadcasting industry) and unusual circumstances, I barely managed to conduct one digital observation study.

The digital observation study was conducted for understanding of the user contexts. The entire session was recorded and notes were taken as well. It was a conceptual task based study, where the user was provided with some scenarios and he performed some specific tasks in different contexts(level of difficulties and automations).The entire session was divided into 3 main parts which started with overall workflow centric questions and ended up with some task specific questions. In between 3 scenario based activities were performed to understand the present workflow and user interaction behaviours towards that. Overall performance is evaluated based on the level of difficulties encountered and level of automation used to perform the tasks.

The observation study was quite useful and insightful to understand the usefulness of metadata to augment intelligence to the video assets and the overall context of the user interaction towards finding objects within a video. Overall performance is evaluated based on the level of difficulties encountered and the level of automation used to perform the tasks. However, the user made an action plan beforehand and determined the sequence of the action to perform the tasks of this study. The user explained each step during performing each task.

A summary of scenarios, goal and overall findings are described below:

Scenario 1: The user was asked to take a video clip and a voice audio clip and make a news content using those two media.

Goal: The goal of this task was to observe user behaviours towards video analysis and identify critical activities to perform the task. To be specific, documenting useful data extracting and applying processes and tracking down critical user interactions to complete the task. Moreover, I looked for the level of difficulties encountered and level of automation used to perform the tasks.

Findings: The user randomly selected 1 audio and 1 video file from the archive. At first, he listened to the audio track to check the quality and information of it followed by the video clip. Once he analyzed those clips, he edited with the audio clip first and layedover the vision chopped from the video clip. Once the editing was made, he checked the semantics and quality of the new content several times and calibrated accordingly. Finally, he stored the content by selecting relevant metadata schema and filled up metadata forms. This task demonstrated the basic broadcasting video editing workflow where the entire process was manual and relatively simple. However it was a very skillful task which requires precise human interactions. This basic task indicates, a user friendly interface is the key to achieve the high level of precision and efficiency of the task.

Scenario 2: The user was asked to take some documentary video clips captured in multiple video cameras and create one documentary video and store it with relevant descriptive data. Afterwards make an angle specific frame search from that video. For example: find all the underwater scenes from that video and mark the timeline with colors for that.

Goal: The goal of this task was to observe the overall process of video analysis techniques, editing, metadata cataloging and searching specific components from that video by using those metadata. Level of difficulties encountered and the level of automation used to perform the tasks was observed as well.

Findings: For this task the user randomly picked 3 documentary video clips from the archive, edited and made a single documentary video. Then filled metadata forms and saved the file. However searching specific frames from the clip was not very straightforward. There were no automated frame extraction processes supported by the existing infrastructure. Thus automated speech to text search queries were made with the word “Under water” but the result was not very precise because subtitles were not available for all the frames. This task indicated the lack of automation for extracting information from a video and confirmed the current process heavily relies on manual efforts for finding specific frames. However the overall task was difficult and time consuming.

Scenario 3: The user was asked to take a football match video and find out all the ‘Goals’ scored in that match and marked the timeline with colors for that. Apart from that, find if there were any ‘Penalty’ scored among those goals and mark them as well with separate color.
Goal: The goal of this task was observing the user behaviours while finding specific information from a video. Moreover, I looked for the level of difficulties encountered and level of automation used to perform the tasks.
Findings: For this task the user picked full length football video clips from the archive and made automated speech to text search queries with the word “Goal” which identified all the frames where the word “Goal” was mentioned. However, the challenge of this task was finding the exact frames where the actual “Goal” scored. As the commentators used the word “Goal” several times so the search picked up all the frames where “Goal” was mentioned. Moreover, the user had to inspect the entire video manually to check whether there were any “Penalty” scored in the match. This task was very difficult and time consuming with very little support for automation.

Key Findings

To extract the most important findings and insights from the study, the user 's approach towards performing those tasks are evaluated based on the level of difficulties encountered and level of automation used to perform each task. Important findings are discussed below:

Key findings
● Existing video editing workflow is an havely occupied with manual tasks which increases the task performing difficulties.
● Video editing demands high precision and close attention to detail.
● Data extraction and cataloging is yet a labor intensive and tedious task for video contents.
● User friendly interface and easy navigation is the key to achieve the high precision and efficiency of the task.
● Existing user interface offers many functionalities which is useful for a power user but overwhelming for the general users
● Low learnability for the general users. Learning the system quickly and getting to the point of optimal (plateau) requires significant time and effort for them.
● The overall video editing process is context specific and error prone.

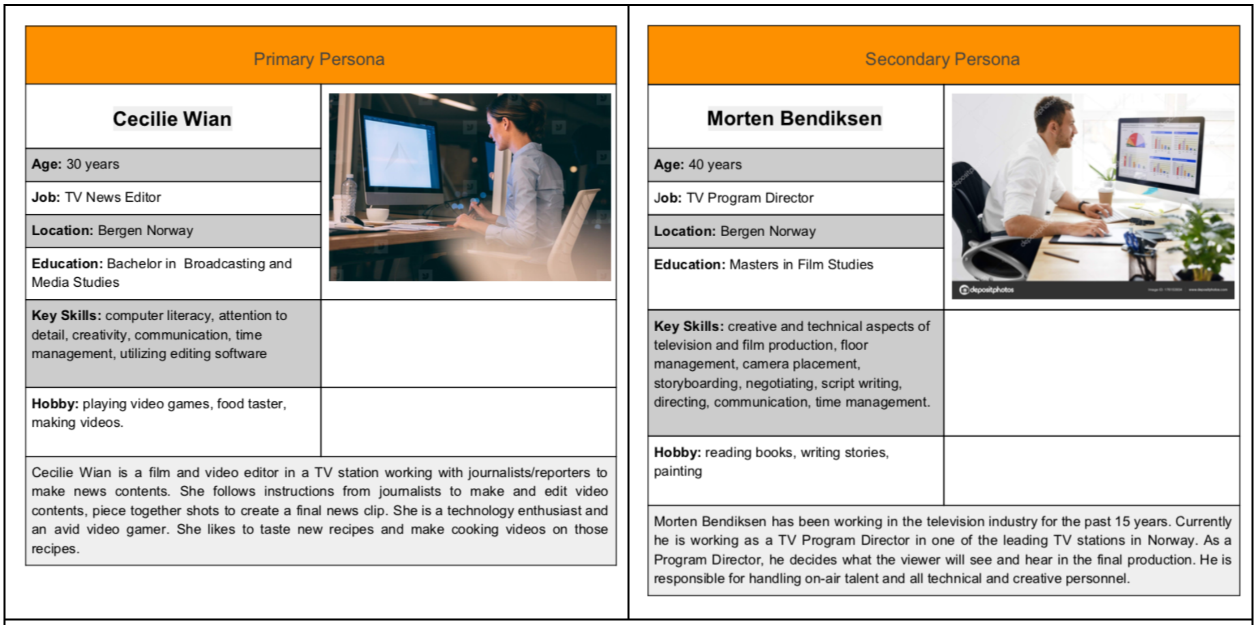
● One of the difficulties or time consuming tasks in the current workflow is to go through the entire video several times to identify the appropriate frames. This is a difficult repetitive task for the editor.
● Users need to fill out descriptive metadata forms manually even though some metadata schemas are used for that. This is a tedious task for the editor.
● Each interaction in the current workflow generates additional metadata which in turn can be useful to simplify the complex tasks. However, the current workflow does not allow any defined process to capture those user interaction data.

Even though I have conducted an observation study focusing on the professional power users, the broader perspective of my research is to develop a generalized adaptive workflow which can cater the needs of a diverse user group. To achieve that objective, the user interface has to be simple and intuitive. The interface should provide clear concise functionalities with proper visual guidance which will enable the user to easily accomplish a task and learn the workflow fast.

Personas

Based on the data collected from the observation study, I have developed a persona that is the next step of this thesis. I chose to create personas because it is a way of making collected survey data into something relatable. It is a way to empathize with potential users through the development phases that can increase usability, utility, and general appeal. Personas are fictional characters, which are created based upon the primary research in order to represent the different user types that might use the service or product in a similar way. Creating personas help to understand the user needs, experiences, behaviours and goals. According to Pruitt et al. a persona is a fictional character with a detailed description that represents a user or customer of a product [57]. Personas are constructed on real data collected from potential users. In this project a primary and a secondary persona are constructed(Figure7) from the primary research on the existing annotation tools and a digital user observation study. A user profile is created to decide who are the primary users for the video object annotation tool from the research and observation study data.

Key characteristics of the defined personas:		
Identity and photo	Status	Goals and tasks
Skillset	Requirements and expectations	Relationships



Primary Persona

Cecilie Wian

Age: 30 years

Job: TV News Editor

Location: Bergen Norway

Education: Bachelor in Broadcasting and Media Studies

Key Skills: computer literacy, attention to detail, creativity, communication, time management, utilizing editing software

Hobby: playing video games, food taster, making videos.

Cecilie Wian is a film and video editor in a TV station working with journalists/reporters to make news contents. She follows instructions from journalists to make and edit video contents, piece together shots to create a final news clip. She is a technology enthusiast and an avid video gamer. She likes to taste new recipes and make cooking videos on those recipes.

Secondary Persona

Morten Bendiksen

Age: 40 years

Job: TV Program Director

Location: Bergen Norway

Education: Masters in Film Studies

Key Skills: creative and technical aspects of television and film production, floor management, camera placement, storyboarding, negotiating, script writing, directing, communication, time management.

Hobby: reading books, writing stories, painting

Morten Bendiksen has been working in the television industry for the past 15 years. Currently he is working as a TV Program Director in one of the leading TV stations in Norway. As a Program Director, he decides what the viewer will see and hear in the final production. He is responsible for handling on-air talent and all technical and creative personnel.

Figure7: Personas

Requirements

After settling on a concept, application requirements are specified. Requirements are made to identify the objectives to include in the development[24]. The goal for collecting requirements is to ensure the usability for better user experience. According to Preece et al. requirements are apparently statements about an intended product or service which states what it should do and how it should perform [24]. The goal of specifying requirements is to establish a sound understanding of the users' needs. To specify requirements is essential to keep track of the goals, and to show what one is working against accomplishing. It helps to remember the direction of the project and make decisions along the way. One of the key focuses of the requirements activity is to make the requirements clear, specific and unambiguous[24].

Software engineering consists of two types of requirements. A functional requirement describes the specifications of the product's functionality, what it should do, while non-functional requirements describe the constraints of the system and its development[24]. The functional requirements are often specifications regarding the scope of work, product, or functions. The non-functional requirements are based on usability, experience, security, and performance. Legal and security requirements and other operational requirements mentioned within non-functional requirements.

The requirements have been adjusted over different iterations. There are three top level activities identified from the observation study which I have initially decided to keep as requirements of this application and break it down into components in different iterations.

- Importing media from the archive
- Mark the identified object in a clutter
- Annotate the object or fill out the metadata form

Design Iteration two - Low-fidelity prototype

From the user study in the first iteration, it is quite clear that storytellers today require modern user interfaces with a high focus on usability and workflow. When users first encounter an interface, they should be able to find their way easy enough to achieve objectives without relying on outside knowledge. A high usability design guides users through the simple and least labor-intensive route. Thus I have decided to focus on developing a new breed of semi-automated video object annotation workflow, harnessing the power of AI and user centered design principles.

In iteration two, the designing and solution process begin. The requirements defined in Design Iteration one, are helpful to start envisioning the application and start drawing simple sketches. The first prototype is a low-fidelity prototype created on paper, which provides the overview of the proof of concept (Figure 9). This iteration is initiated with several brainstorming sessions with an usability expert from Mjoll AS. During these brainstorming sessions we have thoroughly discussed the existing video object annotation workflow and tried to identify the areas of improvements. For these brainstorming sessions I have used pen, paper, sticky notes, white board and marker to draw quick workflows. The findings from my observation study was the departure point of this process followed by relevant discussion with the domain expert from Mjoll AS. Then he quickly walked through some domain concepts and the necessity of the usability in the existing workflow.

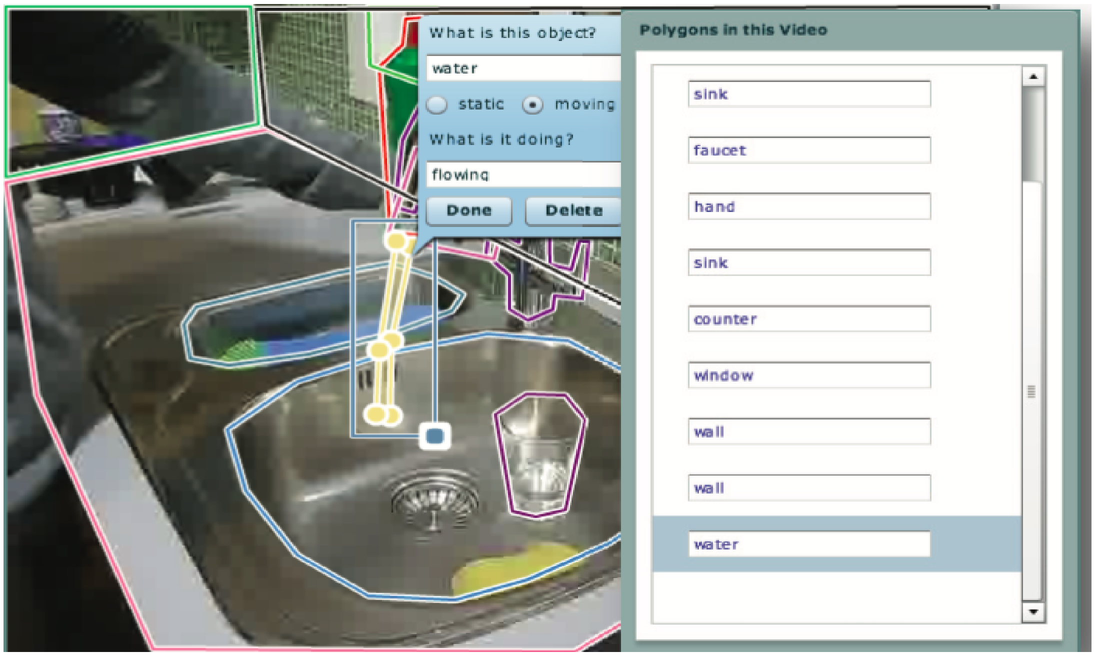
Afterwards I have sketched several workflows for testing the proof of concept. Usability testing and heuristic evaluations are carried out for the concept testing at the end of this iteration.

After several brainstorming sessions and sketching workflows, I have decided to do some research on the existing video object annotation tools and theories behind it. So I went through several of them and finally landed on the relevan four(4) which can contribute to my research. They are

- □ LabelMe image annotation tool [3]
- □ LabelMe video annotation tool [14]
- □ VATIC[15]
- □ iVAT [47]

First Functional Prototype

Scaling video annotation is remarkably laborious than image annotation. So, I extensively read those research papers and use the available annotation tools to understand their functionalities and improvement areas. Afterwards I started developing my proposed object annotation workflow. However due to the nonlinear nature of the video asset it is a complex task to annotate video objects and track them throughout the frames. So, I found it difficult to prototype a video object annotation application right away. I decided to start prototyping an Image annotation as a first step. The image is static so it is easy to label an object and does not frame propagation. So I took inspiration from the research work by Torralba, A et. al, LabelMe[3], an open platform for dense polygon labeling on static images and developed my first interactive prototype of image annotation tool.



What is this object?

water

○ static ● moving

What is it doing?

flowing

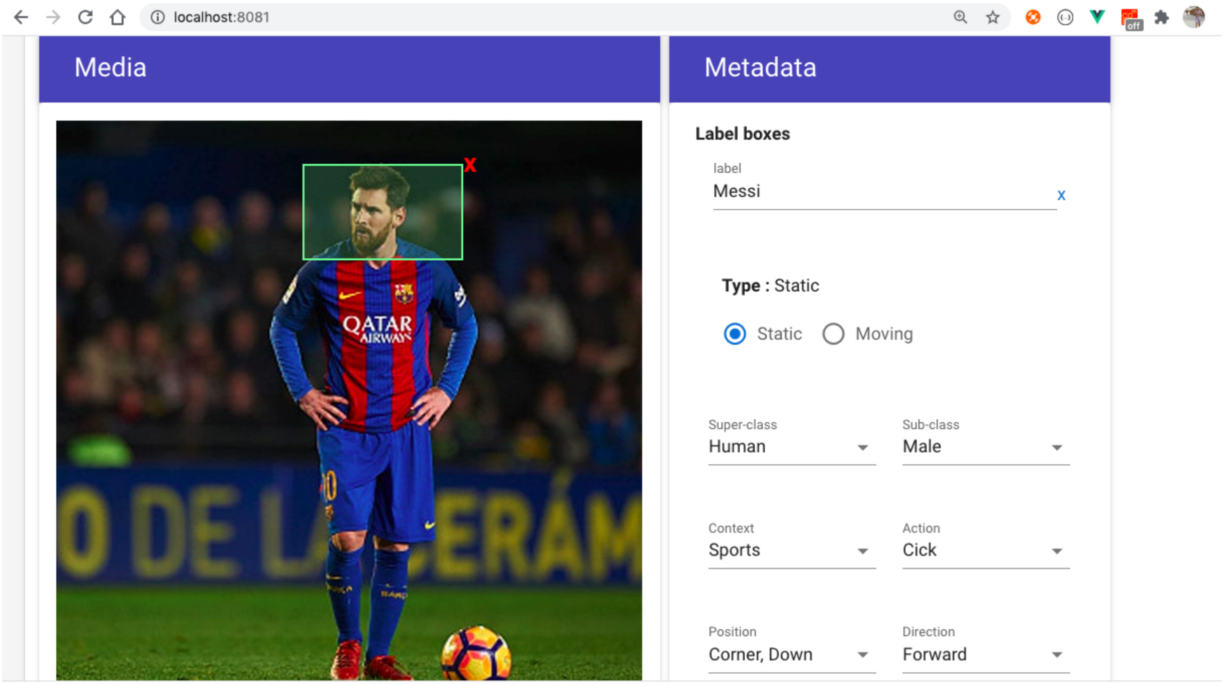
Polygons in this Video

- sink
- faucet
- hand
- sink
- counter
- window
- wall
- wall
- water

Figure 8: LabelMe annotation tool

This interactive prototype is developed by using Vue.js[60],a Progressive JavaScript Framework and Vuetify[61], a complete UI framework built on top of Vue.js. This code can be found in the mentioned github repo [https://github.com/coolrab/real-time-object-recognition.git]

There are two parts of this interface. The media panel contains the image which needs to be annotated and the metadata panel contents the metadata form. After selecting the object the user needs to fill out the fields and save the form. Users can choose from the dropdown for most of the fields of the form. Providing options are considered to reduce users cognitive load and expedite the process.



Label boxes

label: Messi

Type : Static

● Static ○ Moving

Super-class: Human

Sub-class: Male

Context: Sports

Action: Cick

Position: Corner, Down

Direction: Forward

Figure 9: Iteration 1, functional prototype for image annotation tool with a metadata form

Usability Testing 1

Usability Testing-1	
Number of participants	5
Age group	25 - 40
Number of task performed	3

To evaluate the first object annotation prototype, I have conducted a usability testing with 5 users who are recruited through personal connection. According to Nielsen et. al [54], a usability test conducted with about 5-8 representatives is enough to find about 80% of usability issues. Those users range between 25 and 40 years of age. They are asked to perform 3 tasks followed by a few feedback questions at the end. User testing is performed in a quiet room and primary information is given about the application prior to the test. I tracked task completion time and observed the overall approach towards the task. Tasks users were asked to perform are:

Question No	Usability Test Questions
Q1	**Import/Open an image.**
Q2	**Mark an object using polygons[Ellipse, Polygon, Rectangle].**
Q3	**Fill out metadata from and save it.**

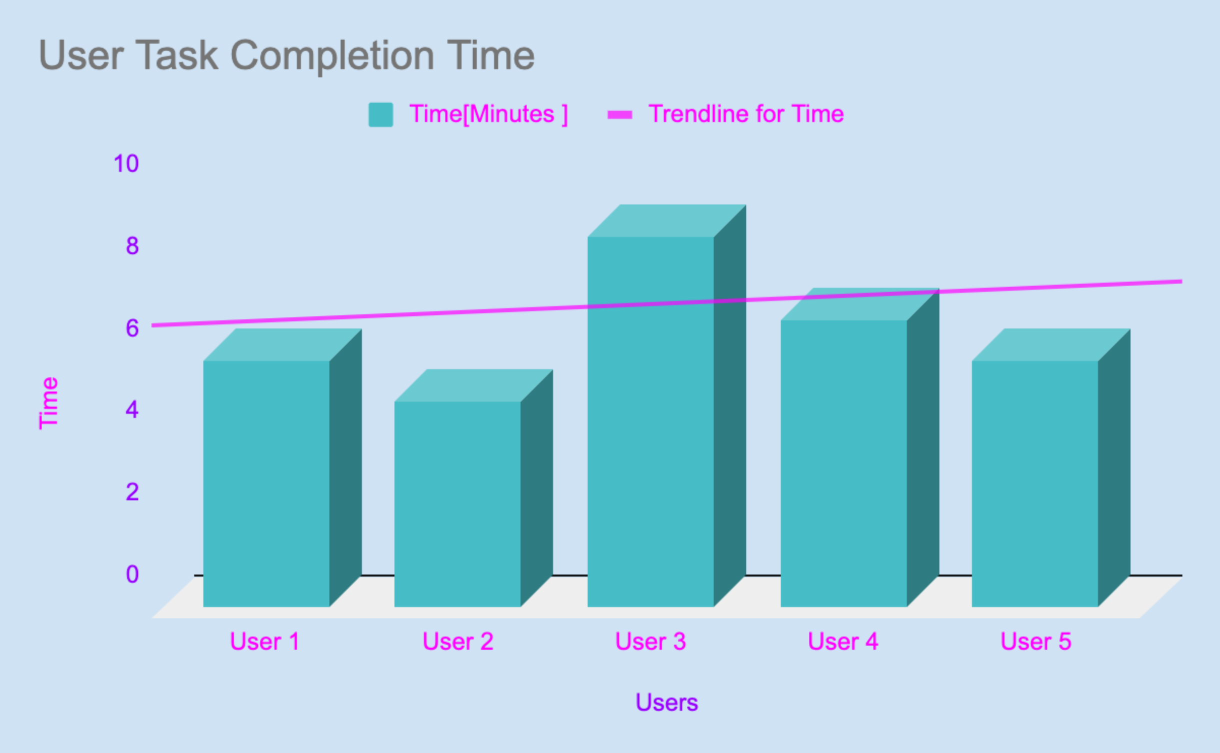


Figure 10: User task completion time for Iteration 1

By analysing task performance time it is observed that on an average it took 6 minutes to perform the tasks. From overall observation and the follow up questions some common patterns have been identified. Likewise users made some recommendations to improve the usability.

User Feedback

Task 1: Import / Open an image.
☐ This task was performed quickly by all the users. However the user had to follow several steps[4 clicks] to perform the task. The steps are File>Open>Select the image>Open.
☐ One user made a recommendation to make a Drag and Drop option to open the file easily.

Task 2: Mark an object using polygons.
This task took a long time to finish because, ☐ Most of the users were confused to choose between rectangle and ellipse as a selector.
☐ Users were confused whether they should select the entire object or part of it. For example, select the face or entire person?

Task 3: Fill out metadata from and save it.
This task seemed the most confusing to the users and took the most time to finish because
☐ Long metadata form to fill out.
☐ Users did not understand the meaning of the input titles(super-class, sub-class, context etc).
☐ Users did not find any information or clue about the input fields.
☐ Selecting from several options took them a long time to decide. For example, selecting the position of an object can take a long time to choose from the options.
☐ Users do not get any cue which data is associated with which object.
☐ Moreover users found it the most tiring and boring task to do.

Findings from the first iteration indicate that the existing object annotation workflow increases the user's cognitive load. I evaluated the first iteration based on the impact of the task and the effort needed to complete them. From the usability testing, it is quite understandable that filling up metadata form is a tedious and time consuming task. Moreover automation services can be harnessed to collect useful data about the objects, such as position of the object.

Second Functional Prototype

Since findings from the first iteration indicate that the existing object annotation workflow increases user's cognitive load, so considering users feedback and several research, I am planning to try a few new steps and remove some existing steps to check if that works better for the user to complete a task.

The primary goal of the second iteration is to fix the workflow based on the user feedback from the first iteration. Secondary goal is to test the object highlighting feature based on Hick's Law.

After several discussions with the expert from Mjoll As on the users feedback of the first image annotation prototype, I decided to follow the simple core 3 steps workflow for the object annotation which are 1. Easy way to upload the media. 2. Draw a box on the identified object. And 3. Tag/annotate the object.

From the conversation with the expert, I have assumed that most of the other functionalities can be automated with the existing technologies and decided to remove the entire object metadata form and connect the annotation input field with the object framing action.

From the literature it is found that highlighting is another way to use Hick's Law. Distractions can often act like having more choices which leads to slow response time. So, reducing distractions is one of the aims in the decision-making context. Thus, apart from the feedback from the first iteration, I have tested highlighting objects in the second iteration. This is a concept to stand out the selected object from the clutter and connect it with the relevant tags to speed up the response times.

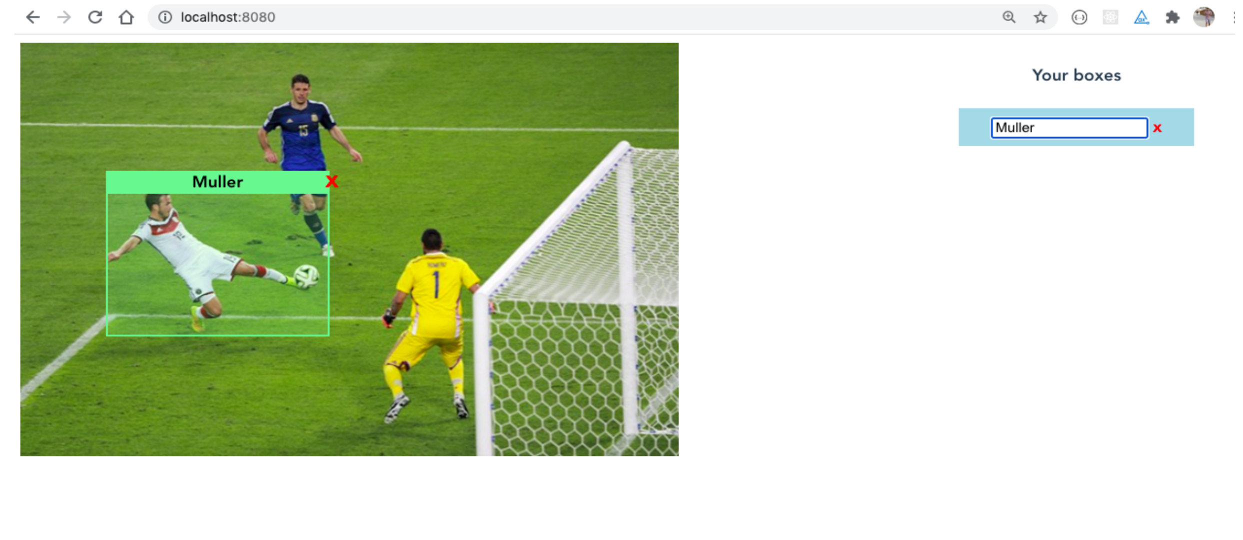


Figure 11: Iteration 2, functional prototype for image annotation tool with a label input field

Iteration 2: image annotation tool with simple 3 steps which are

- □ Import a media
- □ Identify an object
- □ Label the object

Usability Testing 2

Usability Testing-2	
Number of participants	5
Age group	25 - 40
Number of task performed	3

Task 1: Import/Open an image.

Action Taken:
With the new Drag and Drop option this task was performed quickly by all the users. Adopting Nielsen's heuristics Recognition rather than recall I have implemented the drag and drop options which helped the user to recognize rather than remember by reducing the information. This option also complies with the Usability Heuristics Visibility of system status as it provides visual cue during uploading the media.

Outcome:

- □ **Visual cue**
- □ **Improves usability**
- □ **Better user experience**

Task 2: Mark an object using polygons.

Action Taken:
After analyzing the feedback from the first iteration I did some research and adopted the reducing distractions and highlighting concept from Hick's Law. Reducing distractions refers to limiting choosing options. In this case I have kept only one selector and removed all other options. Likewise highlighting concept refers to making a few important options to stand out among the cluttered to speed up the response times. For example, the selected object is directly connected with the input field. So if the input is changed it is visible to the object immediately. This feature also complies with the usability heuristics "Visibility of system status" and "User control and freedom."

Outcome:

- □ **Visual cue.**
- □ **Reduce distraction.**
- □ **Reduce task performance time.**

Task 3: Fill out metadata from and save it.

Action Taken:
Users feedback from the first iteration refers to this task as the most unpleasant and cognitive load task in the entire object annotation workflow. So after carefully doing some research and conducting several brainstorming sessions with the expert from Mjoll AS, I have decided to replace the existing metadata form with a single input field.

Outcome:

- □ **Significant reduction of time.**
- □ **Reduces users cognitive load**
- □ **Easy to understand and perform the task**

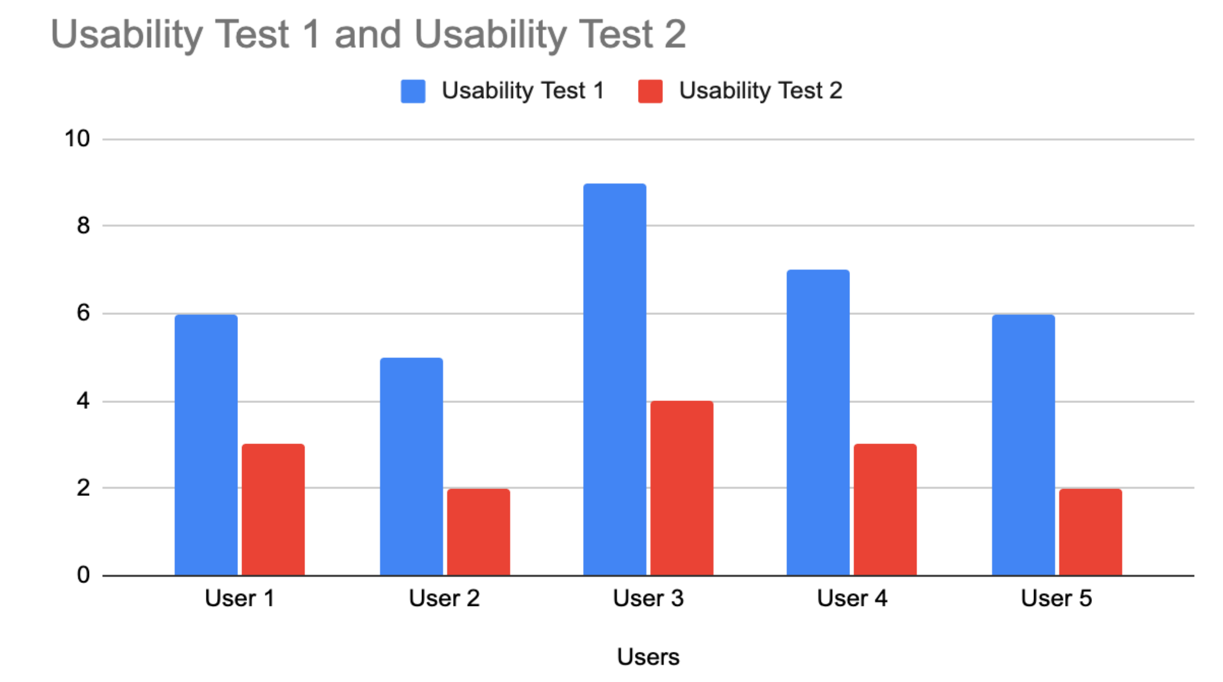


Figure 12: Comparison between users task completion time for Iteration 1 and 2

Figure 12 certainly demonstrates the significant reduction of the task performance time. Considering the users feedback from the first usability testing I have calibrated the workflow based on the user centered design principles, to be precise Hick's law and Nielsen's heuristics that improves the usability and reduces the task performance time by 50%. In the iteration 2 I have simplified the metadata form into one input field. Likewise allowed the user to use only a rectangle to frame the object. At the same time use highlighter to stand out the selected item in a crowd which provides a clear visual cue to the user. By adopting all these features user tests demonstrated a significant hike in the user experience.

Hi-fi Prototype for Video Annotation Tool

Due to the dynamic nature of the video, it seemed ambiguous to develop a functional prototype for the video object annotation tool. Thus I have embraced an incremental designing and testing approach. I developed and tested a functional image annotation tool in the first two iterations and feedback from those iterations is used to design a hi-fi prototype for the video object annotation tool. I have used Adobe XD[37] prototyping tool for designing the workflow. This prototype is evaluated by the usability expert.

While I have started designing the layout of the video object annotation tools, I have focused on the basic design principles based on the Hike's law and Nielsen's heuristics. Design principles are widely applicable laws, guidelines, biases and style considerations that are applied with discretion for a design. It is the fundamental piece of advice to make easy-to-use, pleasurable designs. Design Principles are applied during select, create and organize elements and features in the user interface. To design an effective user interface it is vital to minimize users' cognitive loads and decision-making time. It helps to search out ways to boost usability, influence perception, increase appeal, teach users and make effective design decisions in projects.

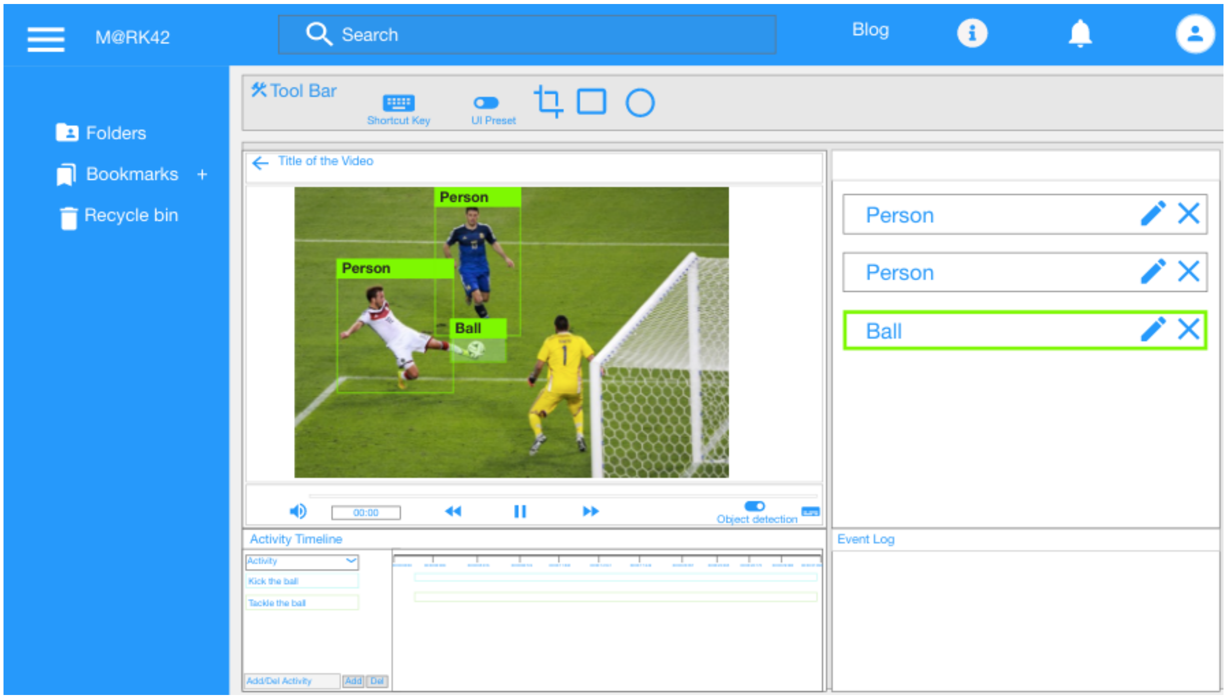
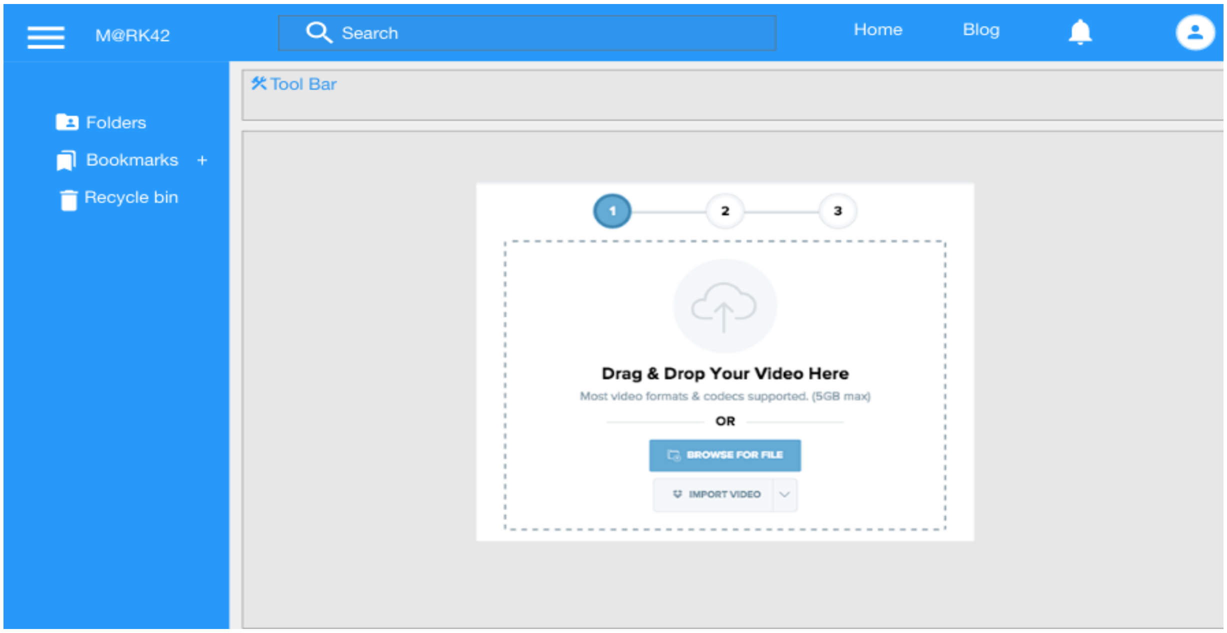


Figure 13: Iteration 3, Hi-fi prototype for video object annotation tool following user centered design principles.

From the user feedback from the last two iterations, I have designed the workflow focusing on the basic 3 steps.

- ☐ Easy way to import a media
- ☐ Single object selector
- ☐ Highlight the selected object

However, to design these 3 step functionalities I have considered following Nielsen's heuristics[54] to structure UI components to minimize users' cognitive load. Table:5 demonstrates a brief overview of the annotation interface.



Drag & Drop Your Video Here

Most video formats & codecs supported. (5GB max)

Figure 14: Drag and drop media

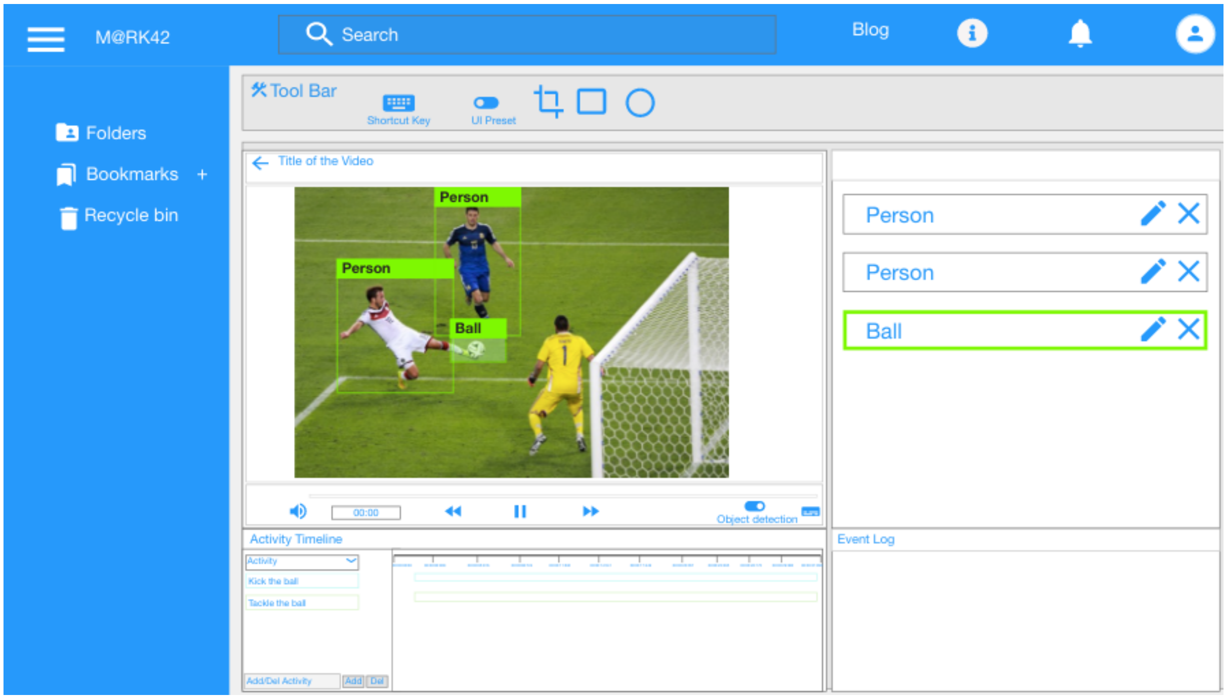


Figure 15: Iteration 3, Hi-fi prototype for video object annotation tool following user centered design principles.

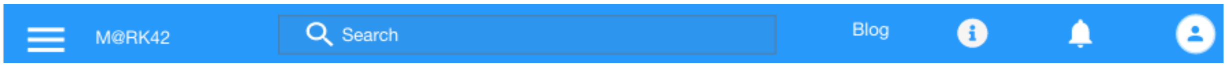


Figure 16: Navigationbar

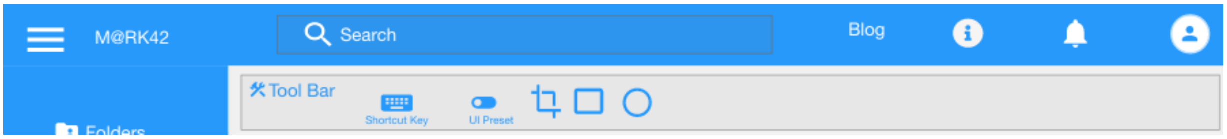


Figure 17: Toolbar

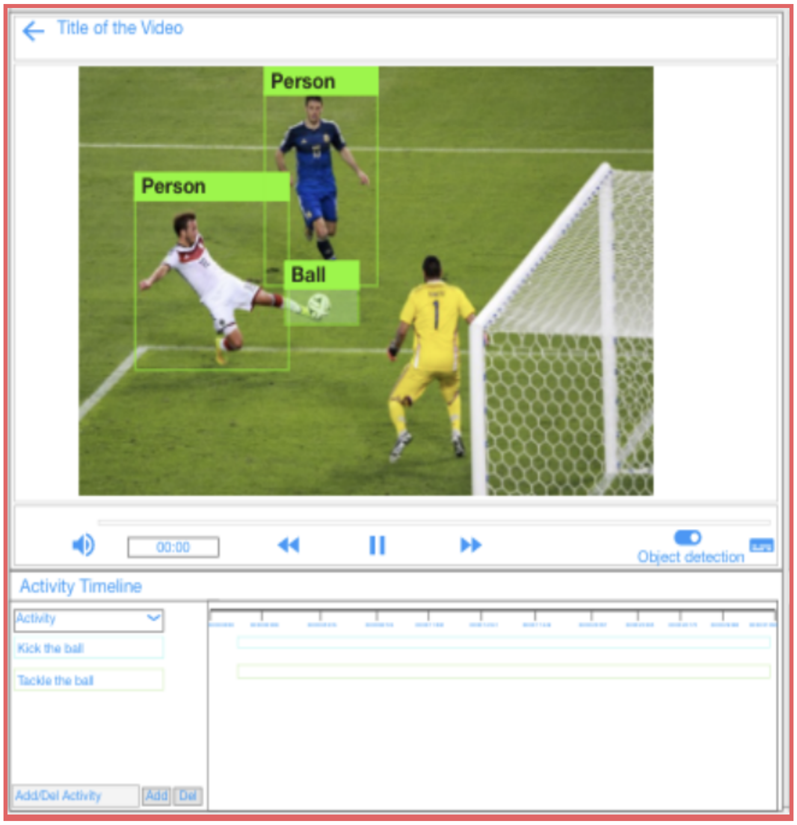


Figure 18: Media panel and activity timeline

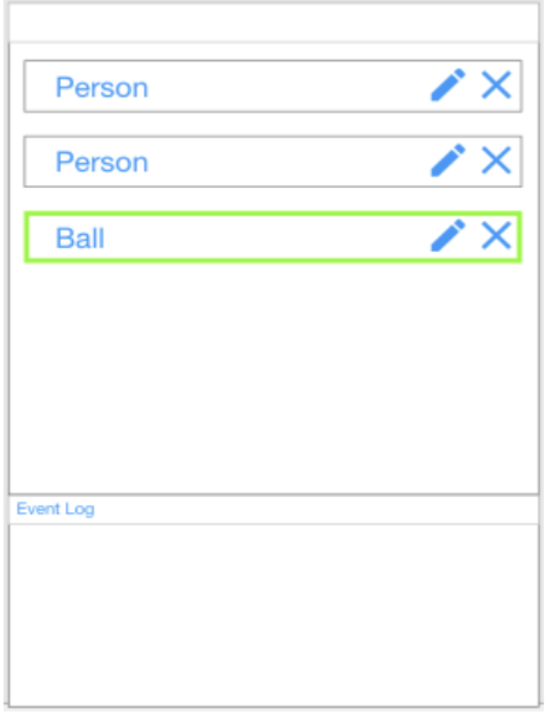


Figure 19: Metadata form and event log

Visual Hierarchy

Visual hierarchy is the principle of arranging elements to point out their order of importance. Designers structure visual characteristics. For example, by looking at the menu icon, users can easily understand information.Users' perceptions can be influenced by laying out elements logically and strategically to guide them towards desired actions. A robust visual hierarchy guides users towards a functionality and gives them clear visual cues.

An effective visual hierarchy helps inform, impress and persuade users to perform tasks efficiently. So, to ensure better usability for an app, it is crucial to structure UI components to minimize users' cognitive load. Hierarchy is a visual design principle which is used to show the importance of different UI elements by manipulating some characteristics. The building blocks of visual hierarchy are size, color, contrast, alignment, repetition, proximity, whitespace and texture style.

Users react extremely fast, encountering an interface. Their eyes follow predictable reading paths and prefer recognition over recall. Considering these, I have envisioned to scale the most important elements to make the most important information prominent and unmissable for users as they try to achieve goals in their individual contexts.

Color Contrast

Color is an integral part of our lives and an important element for visual communication. In an UI design color is one of the core components to grab attention and stimulate interest in ways that would be difficult to create by any other means. Next to application functionalities, the color scheme is factor number two for effective user experience and performance. So choosing the right colors for a design is vital because it can affect user's moods, behavior, and stress levels.

Color theory and the psychological effects color can have on usability and efficiency is a complex topic. But there are certain aspects that can be addressed on a more universal level. Things like the common meanings of the three traditional primary colors are red, green and blue and cultural variations in color.

Using two of three colors is one of the simple design principles that has proven its worth over time. However red and green have been used for so long to signal permit and caution that are ingrained in our psyche correspondingly. Thus blue calms while red puts us in alert mode. So to keep the design simple and neutral I have chosen to use blue as the primary color for the UI with white, green and grey respectively.

Table 5: User interface design following usability heuristics

Nielsen’s heuristics for user interface design		Features included	Checked
H1	Visibility of system status	Added marker and tag to the selected object.	☑
H2	Match between system and the real world	All the relevant terms, concept and icons are used	☑
H3	User control and freedom	Close and edit buttons are added for each input field to fosters a sense of freedom and confidence	☑
H4	Consistency and standards	The app adheres to general standards, so users know what to expect, learnability is increased, and confusion is reduced. Ex. a magnifying-glass icon stands for search.	☑
H5	Error prevention	Added prompt message before delete a tag	☑
H6	Recognition rather than recall	Minimizing the user's cognitive load by making elements, actions, and options visible. Ex. activity log and activity timeline	☑
H7	Flexibility and efficiency of use	Allowed users to tailor frequent actions. Ex. keyboard shortcuts	☑
H8	Aesthetic and minimalist design	Adopted visual hierarchy and color contrast principles to design minimal and simple interface	☑
H9	Help users recognize, diagnose,and recover from errors	Used color contrast to help users recognize, diagnose, and recover from errors	☑
H10	Help and documentation	Added information icon and blog in the navigation bar.	☑

Chapter 5 Conclusion and Outlook

The following chapter summarizes the overall research process and findings followed by answering the research questions. Furthermore recommendations and future scopes are briefly discussed at the end.

Summary of the process

As stated early within the research, the framework for the thesis is to form a decent user experience by improving usability. The primary goal for the application is to scale down user cognitive load and higher cognitive process time by utilizing user-centered design principles. However, the process of making users complete object annotation tasks efficiently is not necessarily straightforward. In my research, I tried to figure out how the combination of AI and design principles can be useful to reduce users' cognitive load while navigating through a complex workflow. Thus, I have incrementally designed and tested a semi-automatic interactive video annotation workflow, motivated by the limitations of existing annotation tools.

The first phase of the user-centered design (UCD) method has helped to specify the context of use. The data from the observation study helped to create a persona and set the bare-bones requirements as starting points for the application. Building the persona became useful to contextualize the user behaviour towards using the tool. However, the personas were set aside during the testing phase. The personas would have included more value if I could have conducted more onsite observation studies. Due to the pandemic, finding appropriate users was very difficult for this project.

The next step in the process specifies the context of potential use cases and the requirements. I wanted to have several expert interviews but ended up only conducting one. The input I received from the industry expert helped me, among other things, to define and focus on the core usability to include. The expert helped me to walk through the process from the brainstorming to evaluate the prototypes. If there were a normal situation and managed to have several professional observation studies in the real environment and test the prototypes with them then it might have given more insight on where to focus more for the process simplification.

In the next step of the UCD process, I began to include the end-users more within the process. I have developed functional prototypes from the concept to a more lifelike application design, which was very useful to refine the new workflow. Additionally, this was useful to work out how the users interacted with the prototype. The tactic of going out doing user testing worked well to seek out their reaction on the new workflow. It is an efficient method to induce quick feedback from users and refine the workflow accordingly.

In the final step of the UCD process, the high-fidelity interactive prototype is developed to conduct a more complementary usability test. Most of the features are taken from the previous prototypes as those were refined from the users feedback. However, due to the time constraints, I have conducted expert heuristic evaluation for this prototype.I received critical feedback from the expert on the prototype which can be useful for the further development of the application. Moreover, it is important to observe if the workflow is easy to navigate and reduces users cognitive load to perform a task.

Answering the Research Questions

A quantitative and qualitative evaluation of the hybrid object annotation workflow indicates that adopting user centered design principles and semi-automatic modality can reduce human cognitive load by at least one order of magnitude, limiting the user interaction choices and generating visual cues.

Table 5 demonstrates the summary of the research findings that allows to draw a conclusion on the research questions. In this research, question 2 and 3 are the subquestion(supliment) to answer the main research question.

Table 6: Summary of the research findings

Research Question	**Results of analysis**	**Checked**
RQ1: How state-of-the-art machine learning techniques complement user centered design principles to enhance workflow and usability by reducing users cognitive loads and decision-making time for economically annotating video objects?	□ Adopting object detection AI services ensures the initial object detection. □ Tracks the objects throughout the frames □ Makes the workflow flexible. □ Reduces users' aggregated work load. □ Improve usability and annotation efficiencencies □ Users perform as a curator instead.	☑
RQ2:	From the usability testing on design iteration 1 and	☑

How does limited choice of interaction reduce the cognitive load and expedite the decision making?	2, it is observed that limiting interaction options(allowing only one object selector) and introducing visual cues(highlighting the selected object) reduces users **task compilation time 50%** by lowering decision making complexities.	
RQ3: Scoping the number of perceived options on screen makes the workflow fluid and the interface more user friendly that allow users to accomplish the task efficiently.	☐ Conducted incremental usability testing to select and structure UI components logically and strategically to guide users towards performing desired actions efficiently. ☐ From the usability testing on design iteration 1 and 2, it is identified reducing the UI components and making annotation forms simple users gain efficiencies significantly.	☑

Conclusion

In this master's thesis, motivated by the limitations of existing annotation tools, I have researched how the combination of AI and user-centered design approach can reduce users' cognitive load and decision making time for video object annotations. The aim of designing a user centered video annotation tool is to relieve the user from the cognitive burden of manual annotation as much as possible. To attain this ideal goal, manifold functionalities are required in order to make the annotations workflow as fluid as possible. The methods used in the research were observation study, expert interview, user testing, and expert evaluations.

The combination of research methods provides useful feedback in order to investigate how the application can contribute to reducing users cognitive load while performing annotation tasks.I have incrementally designed, tested and calibrated the workflow based on the user centered design principles. A quantitative and qualitative evaluation of the proposed workflow demonstrates that the use of the user centered design principles and semi-automatic modality can potentially reduce human cognitive load by at least one order of magnitude, limiting the user interaction choices and generating visual cues. Furthermore, the findings also indicate that user centered design principles help to structure UI components logically and strategically to guide users towards performing desired actions efficiently. However, limiting the interaction choices might have a side effect of lower precision annotation which could be an interesting perspective for the future research work.

Future Work

Regarding the future work for the application, I am assuming the application will be significantly improved by integrating AI based object detection models. The immediate next step for the further development is to refine the design based on the suggestions from the expert evaluation and conduct several usability tests with the real users. Then render services from Amazon Rekognition[50] and make a functional prototype for usability testing with professional users.

Based on the user feedback, the first version or MVP of the semi-automated video object annotation app could be developed and open for the beta testing.

References

[1] Sorokin, A., & Forsyth, D. (2008). Utility data annotation with Ama- zon Mechanical Turk. Urbana, 51, 61, 820.

[2] Deng, J., Dong, W., Socher, R., Li, L., Li, K., & Fei-Fei, L. (2009). Im-ageNet: a large-scale hierarchical image database. In Proc. CVPR (pp. 710–719).

[3] Russell, B., Torralba, A., Murphy, K., & Freeman, W. (2008). La- belMe: a database and web-based tool for image annotation. In- ternational Journal of Computer Vision, 77(1), 157–173.

[4] Kumar, N., Berg, A. C., Belhumeur, P. N., & Nayar, S. K. (2009). Attribute and simile classifiers for face verification. In ICCV.

[5] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisser- man, A. (2010). The pascal visual object classes (voc) challenge. International Journal of Computer Vision, 88(2), 303–338.

[6] Von Ahn, L., & Dabbish, L. (2004). Labeling images with a computer game. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 319–326). New York: ACM Press.

7. L. Castrejon, K. Kundu, R. Urtasun, and S. Fidler. Annotating object instances with a polygon-rnn. In CVPR, 2017.

8. L.C. Chen, S. Fidler, A. Yuille, and R. Urtasun. Beat the mturkers: Automatic image labeling from weak 3d supervision. In CVPR, 2014.

[9] Von Ahn, L., Liu, R., & Blum, M. (2006). Peekaboom: a game for locating objects in images. In Proceedings of the SIGCHI confer- ence on human factors in computing systems (pp. 55–64). New York: ACM Press.

[10] Ramanan, D., Baker, S., & Kakade, S. (2007). Leveraging archival video for building face datasets. In IEEE 11th international con- ference on Computer vision, 2007. ICCV 2007 (pp. 1–8). New York: IEEE Press.

[11]Welinder, P., Branson, S., Belongie, S., & Perona, P. (2010). The multi- dimensional wisdom of crowds. In Neural information processing systems conference (NIPS) (Vol. 6, p. 8).

[12] Vittayakorn, S., & Hays, J. (2011). Quality assessment for crowd- sourced object annotations. In J. Hoey, S. McKenna, & E. Trucco (Eds.), Proceedings of the British machine vision conference (pp. 109–110).

[13] Endres, I., Farhadi, A., Hoiem, D., & Forsyth, D. (2010). The benefits and challenges of collecting richer object annotations. In CVPR workshop on advancing computer vision with humans in the loop. New York: IEEE Press.

[14] Yuen, J., Russell, B., Liu, C., & Torralba, A. (2009). LabelMe video: building a video database with human annotations. In Interna- tional conference of computer vision.

[15] Vondrick, C., Patterson, D. and Ramanan, D., 2013. Efficiently scaling up crowdsourced video annotation. International journal of computer vision, 101(1), pp.184-204.

[16] Huber, D. (2011). Personal communication.

[17] Ali, K., Hasler, D., & Fleuret, F. (2011). Flowboost–appearance learn- ing from sparsely annotated video. In IEEE computer vision and pattern recognition.

[18] Design. (2019). In Dictionary.com Unabridged. Based on the Random House Unabridged Dictionary. Random House, from https://www.dictionary.com/browse/design

[19] Norman, D., & Nielsen, J. (2019). The Definition of User Experience (UX). Retrieved 8 March 2019, from https://www.nngroup.com/articles/definition-user-experience/

[20] Agarwala, A., Hertzmann, A., Salesin, D., & Seitz, S. (2004). Keyframe-based tracking for rotoscoping and animation. ACM Transactions on Graphics, ACM, 23, 584–591.

[21] Saffer, D. (2010). Designing for interaction. creating innovative applications and devices. Retrieved from http://www.gbv.de/dms/ilmenau/toc/602565936.PDF

[23] Buchanan, A., & Fitzgibbon, A. (2006). Interactive feature tracking using kd trees and dynamic programming. In CVPR 06, Citeseer (Vol. 1, pp. 626–633).

[24] Preece, J., Sharp, H., & Rogers, Y. (2015). Interaction design: beyond human-computer interaction (4th ed.). West Sussex.

[25] Interaction Design Foundation. (n.d.). User Experience (UX) design. Retrieved 13 February 2019, from https://www.interaction-design.org/literature/topics/ux-design

[26] Norman, D. A. (1988). The psychology of everyday things. The Psychology of Everyday Things. New York, NY, US: Basic Books.

[28] Fisher, R. (2004). The pets04 surveillance ground-truth data sets. In Proc. 6th IEEE international workshop on performance evalua- tion of tracking and surveillance (pp. 1–5).

[29] Smeaton, A., Over, P., & Kraaij, W. (2006). Evaluation campaigns and trecvid. In Proceedings of the 8th ACM international workshop on multimedia information retrieval (pp. 321–330). New York: ACM Press.

[31] Laptev, I., Marszalek, M., Schmid, C., & Rozenfeld, B. (2008). Learn- ing realistic human actions from movies. In IEEE conference on computer vision and pattern recognition, 2008. CVPR 2008 (pp. 1–8). New York: IEEE Press.

[32] K.Maninis,S.Caelles,Y.Chen,J.Pont-Tuset,L.Leal-Taixe, D. Cremers, and L. Van Gool. Video Object Segmentation Without Temporal Information. IEEE Transactions on Pat-tern Analysis and Machine Intelligence, 2018.

[33] D. Mihalcik, D. Doermann, The design and implementation of viper (2003).

[34] K. Ali, D. Hasler, F. Fleuret, Flowboost: appearance learning from sparsely annotated video, in: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 1433–1440.

[35] J. Yuen, B. Russell, C. Liu, A. Torralba, Labelme video: building a video database with human annotations, in: 2009 IEEE 12th International Conference on Computer Vision, 2009, pp. 1451–1458.

[36] I. Kavasidis, S. Palazzo, R. Di Salvo, D. Giordano, C. Spampinato, A semi-automatic tool for detection and tracking ground truth generation in videos, in: Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications, ACM, 2012, pp. 6:1– 6:5.

[37] Adobe XD https://www.adobe.com/products/xd.html

[38] I. Kavasidis, S. Palazzo, R. Salvo, D. Giordano, C. Spampinato, An innovative web-based collaborative platform for video annotation, Multimedia Tools Appl. (2013) 1–20.

[39] I. Kavasidis, C. Spampinato, D. Giordano, Generation of ground truth for object detection while playing an online game: productive gaming or recreational working?, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2013, pp 694–699.

[40] Proctor, R.W. and Schneider, D.W., 2018. Hick's law for choice reaction time: A review. Quarterly Journal of Experimental Psychology, 71(6), pp.1281-1299.

[41] P. Voigtlaender, M. Krause, A. Osep, J. Luiten, B. B. G. Sekar,A.Geiger,andB.Leibe.MOTS:Multi-ObjectTrack- ing and Segmentation. CoRR, abs/1902.03604, 2019.

[42] P. Voigtlaender and B. Leibe. Online Adaptation of Convo- lutional Neural Networks for Video Object Segmentation. In British Machine Vision Conference 2017, BMVC 2017, Lon- don, UK, September 4-7, 2017, 2017.

[44] B.L. Wang, C.-T. King, and H.-K. Chu. A Semi-Automatic Video Labeling Tool for Autonomous Driving Based on Multi-Object Detector and Tracker. In 2018 Sixth International Symposium on Computing and Networking (CAN- DAR), pages 201–206. IEEE, nov 2018.

[45] S. Wug Oh, J.-Y. Lee, K. Sunkavalli, and S. Joo Kim. Fast Video Object Segmentation by Reference-Guided Mask Propagation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 7376– 7385, 2018.

[47] Bianco, S., Ciocca, G., Napoletano, P. and Schettini, R., 2015. An interactive tool for manual, semi-automatic and automatic video annotation. Computer Vision and Image Understanding, 131, pp.88-99.

[48] Konyushkova, K., Uijlings, J., Lampert, C.H. and Ferrari, V., 2018. Learning intelligent dialogs for bounding box annotation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 9175-9184).

[49] J. Zhao and S.-c. S. Cheung. Human Segmentation by Geo- metrically Fusing Visible-light and Thermal Imageries. Mul- timedia Tools and Applications, 73(1):61–89, nov 2014.

[50]https://aws.amazon.com/rekognition/?blog-cards.sort-by=item.additionalFields.createdDate&blog-cards.sort-order=desc

[51] Papadopoulos, D.P., Uijlings, J.R., Keller, F. and Ferrari, V., 2017. Extreme clicking for efficient object annotation. In Proceedings of the IEEE international conference on computer vision (pp. 4930-4939).

[52] Papadopoulos, D.P., Uijlings, J.R., Keller, F. and Ferrari, V., 2017. Training object class detectors with click supervision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 6374-6383).

[53] Subramanian, A. and Subramanian, A., 2018. One-Click Annotation with Guided Hierarchical Object Detection. arXiv preprint arXiv:1810.00609.

[54] Nielsen, J. and Molich, R., 1990, March. Heuristic evaluation of user interfaces. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 249-256).

[55]

[56] Rudd, J., Stern, K., & Isensee, S. (1996). Low-versus High-fidelity Prototyping debate. Magazine Interactions, 76–85. Retrieved from https://dl.acm.org/citation.cfm?id=223514.

[57] Pruitt, J., & Adlin, T. (2010). The Persona Lifecycle: Keeping People in Mind Throughout Product Design. San Francisco: Elsevier.

[58] Kallinikos, J. and Mariategui, J.C., 2011. Video as digital object: Production and distribution of video content in the internet media ecosystem. The Information Society, 27(5), pp.281-294.

[59] Tuch, A.N., Presslaber, E.E., Stöcklin, M., Opwis, K. and Bargas-Avila, J.A., 2012. The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments. International journal of human-computer studies, 70(11), pp.794-811.

[60] Vue.js [https://vuejs.org/]

[61] Vuetify [https://vuetifyjs.com/]

Appendix

Usability Testing Consent Form

Applying user centered designing principles to improve usability for object annotation workflow.

This user testing is part of Masters-research at University of Bergen(UIB), Department of Media and Interaction design concerning object annotation tool. The purpose of this usability testing is to find out the easiness of task performance, the pattern of user interaction with the interface and overall decision making process while annotating an object. This usability testing is a part of the Master's research about how to simplify an object annotation workflow to reduce users cognitive load and annotate objects efficiently. It is voluntary to participate in the user testing and the data is collected anonymously and is protected in accordance with the guidelines of the Norwegian Centre for Research Data (NSD). Researcher in charge is Md Fazla Rabbi Alam (md.alam@uib.no).

Introduction

There are two parts of this interface. The Media panel contains the image which needs to be annotated and the Metadata panel contents the metadata form. The user can select an image by clicking on the File>Open from the menu. Once the image is open in the Media panel the user can use polygons to select an object. Once the object is selected the user can start filling out the Metadata form and save it at the end.

The use needs to perform three (3) tasks:

Task 1: Open an image from the desktop.

Task 2: Mark an object using polygons.

Task 3: Fill out metadata from and save it.

Heuristic Evaluation Form

Heuristic-based Usability Evaluation of Object Annotation App

- Video Object Annotation App
  - Link: https://xd.adobe.com/view/45c0fbea-3ba8-42d0-a9c1-b2227eb8c62b-1093/?fullscreen
  - Visit all pages
- Fill the heuristics evaluation table
  - Register how many incidents (numbers) for each usability problem and explain them accordingly. For example, If there are 2 minor and 1 Major incidents then put (2) in the minor cell and (1) in the major cell.
  - If available, provide screenshots to explain each usability problem you identified.

Severity Rates

- **Cosmetic usability problem** = need not be fixed unless extra time is available on project
- **Minor usability problem** = fixing this should be given low priority
- **Major usability problem** = important to fix, so should be given high priority
- **Usability catastrophe** = imperative to fix this before product can be released

HEURISTICS EVALUATION TABLE

Nielsen's Heuristics	**Severity Rate**				
	Cosmetic	**Minor**	**Major**	**Catastrophe**	**Explanation**
Visibility of system status		(1)			1.
Match between system and the real world		(1) (2)			
User control and freedom					
Consistency and standards					
Error prevention					

Recognition rather than recall		(1)			1.
Flexibility and efficiency of use	(1)				1.
Aesthetic and minimalist design	(1) (2)				1.
Help users recognize, diagnose, and recover from errors					
Help and documentation					
Other usability problem(s)					