Guided Visual Analytics for Image Selection in Time and Space

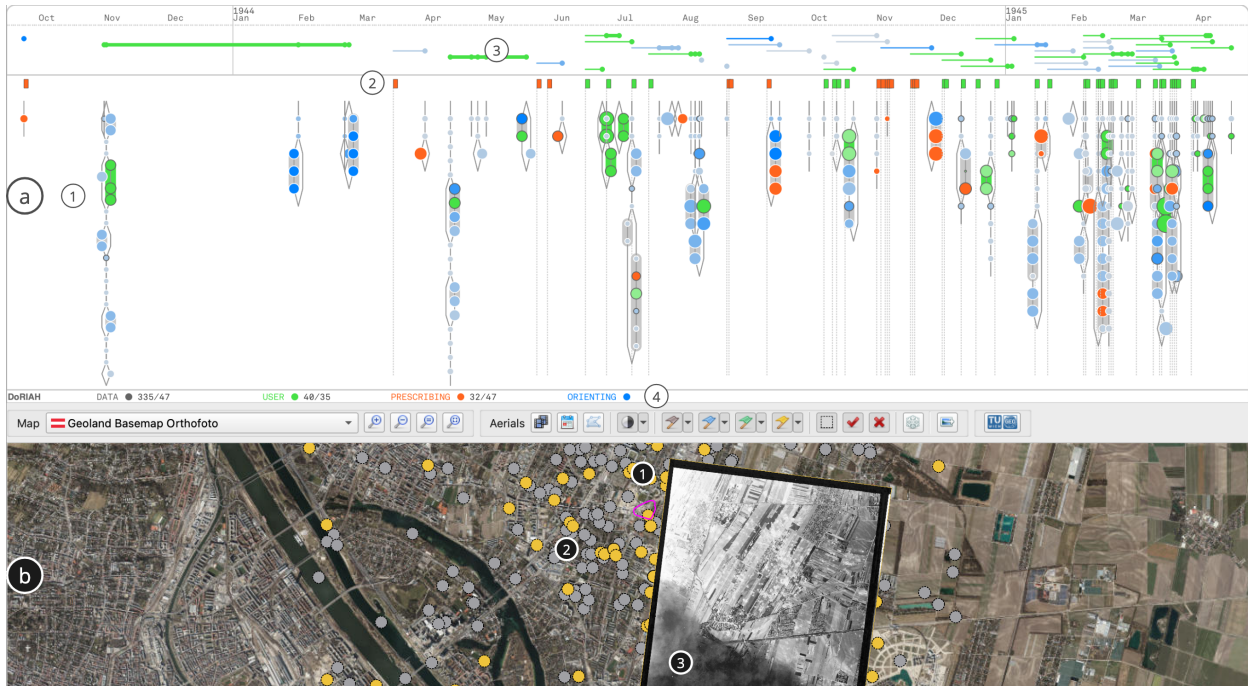Ignacio Pérez-Messina ⓘ, Davide Ceneda ⓘ, and Silvia Miksch ⓘ

Fig. 1: Our system for Image Selection for UXO detection. The timeline visualization (a) shows image and flight metadata (1), attacks and coverage (2), and their aggregated relations (3). The user selection (green), orienting guidance (blue), and prescribing guidance (orange) are represented in different colors and can be switched on/off from the menu (4). In the map view (b), the area of interest in purple (1) and the images (2), are shown. There, photographs can be inspected and manually georeferenced (3).

Abstract—Unexploded Ordnance (UXO) detection, the identification of remnant active bombs buried underground from archival aerial images, implies a complex workflow involving decision-making at each stage. An essential phase in UXO detection is the task of image selection, where a small subset of images must be chosen from archives to reconstruct an area of interest (AOI) and identify craters. The selected image set must comply with good spatial and temporal coverage over the AOI, particularly in the temporal vicinity of recorded aerial attacks, and do so with minimal images for resource optimization. This paper presents a guidance-enhanced visual analytics prototype to select images for UXO detection. In close collaboration with domain experts, our design process involved analyzing user tasks, eliciting expert knowledge, modeling quality metrics, and choosing appropriate guidance. We report on a user study with two real-world scenarios of image selection performed with and without guidance. Our solution was well-received and deemed highly usable. Through the lens of our task-based design and developed quality measures, we observed guidance-driven changes in user behavior and improved quality of analysis results. An expert evaluation of the study allowed us to improve our guidance-enhanced prototype further and discuss new possibilities for user-adaptive guidance.

Index Terms—Application Motivated Visualization, Geospatial Data, Mixed Initiative Human-Machine Analysis, Process/Workflow Design, Task Abstractions & Application Domains, Temporal Data

✦

1 INTRODUCTION

UXO detection is the practice of finding and removing active bombs buried underground from aerial bombings during past armed conflicts, particularly WWII, and ensuring the safety of the area, such as buildings and civilians. It is costly and hazardous but necessary work. Interestingly, aerial photographic reconnaissance flights, which were a crucial

• *Ignacio Pérez-Messina, Davide Ceneda and Silvia Miksch are with TU Wien.*
  E-mail: {ignacio.messina, davide.ceneda. silvia.miksch}@tuwien.ac.at

part of the Allied strategic bombing campaigns, now offer the most valuable source of information for the detection of ordnance, and the analysis of this archival material has become an essential part of UXO detection [29, 34]. UXO detection comprises different stages that may overlap but can be considered independent: image selection, image georeferencing, crater detection, and risk assessment, among others (see Fig. 2). In this paper, we focus on image selection.

In the image selection process, experts must first consider all available imagery in the archives that has some coverage over a specified Area of Interest (AOI) and select the best subset of images that will allow them to create a faithful reconstruction of the AOI, i.e., one that leaves no space for undetected UXOs (false negatives). This process can be costly and time-consuming since even for small AOIs there could be hundreds of relevant images. The resulting image subset must

have good spatial coverage over the AOI but also temporal coverage over known attacks. Thus, image selection is an autonomous and complex task crucial to the success and quality of UXO detection. More in general, image selection belongs to an exciting class of problems involving decision-making that has been barely explored in Visual Analytics (VA) literature, namely, combinatorial optimization (similar to the knapsack problem [24]), where experts must produce a minimal selection while optimizing different criteria (cost and quality of images, spatial and temporal coverage, resilience to poor quality images, etc.) and use their knowledge on site-specific historical events, to assess the actual quality of the photographs selected.

Currently, only general-purpose software such as Excel, Google Earth, and ArcGIS is used for UXO detection, with little automation and no task-specific visualization techniques. Hence, our task was to design a functional tool that experts could effectively integrate into their workflow to improve the quality of their work. For its design, we used guidance as a conceptual framework, as guidance is defined as "the computer-aided process of actively resolving user knowledge gaps" [5] featuring a user-centered approach focused on studying *how* users can be supported and understanding the implications of recommender models on the analysis.

Through a series of semi-structured interviews with domain experts from a partner company, we identified the spectrum of user tasks, derived the knowledge gaps, and defined the necessary guidance to support them. Our guidance-enhanced interface (Fig. 1) is composed of two coordinated views that show the available images in their geographical and temporal dimensions. In the timeline (Fig. 1A) images are organized into a custom glyph representing reconnaissance flights. A guidance model that captures expert criteria *orients* users towards the best candidate images and *prescribes* full image selection sets, using only color-coding as visual variables. The production of a selection by a user follows an iterative *divide and conquer* approach, where the whole time period of a project is partitioned into smaller time frames which can be independently analyzed by temporal zooming.

We evaluated the usability of our interface and the appropriateness of our guidance-enhanced system through a user study with domain experts where they performed image selection over two real-world scenarios, with and without guidance, confirming that our approach helped improve the quality of results produced by experts.

Our contributions are (1) a domain characterization of image selection for UXO detection with a visualization and guidance task analysis (Sec. 3), (2) a guidance-enriched visualization design for time-oriented data (Sec. 4), (3) a qualitative user study of our working prototype (Sec. 5-7), and (4) a model that predicts selection quality in alignment with expert criteria (Sec. 8).

2 BACKGROUND AND RELATED WORK

The expected outcome of image selection is an "optimized subset" of the whole set of candidate images; hence, it involves decision-making. Interestingly, interactive approaches supporting this task appear as an underrepresented topic in the literature, although it is a class of problems that appears in a wide variety of real-world scenarios. Notably, in the review of design papers mentioning decision-making as a goal, no system deals with combinatorial optimization (i.e., optimizing the choice of items in a set) [10]. Moreover, within the domain of UXO detection, the only visualization approach we found is Amor-Amorós et al. [2], which is, nonetheless, limited to image georeferencing.

Guidance in VA   We designed and implemented a guidance-enhanced VA approach to support image selection. Guidance is a mixed-initiative process that refers to approaches (e.g., recommender systems) for enhancing analytical tasks and it describes how assistance can be provided to the user [6]. Guidance approaches can be characterized by the user's knowledge gap it aims to resolve and by the constraints guidance imposes over user action. The latter has been classified into three guidance degrees, from lowest to maximum constraint [5]: *orienting* – usually highlighting interesting elements (e.g., previously recorded patterns in malware detection [35], number of segments as a parameter for the discovery of cyclical patterns [9]);
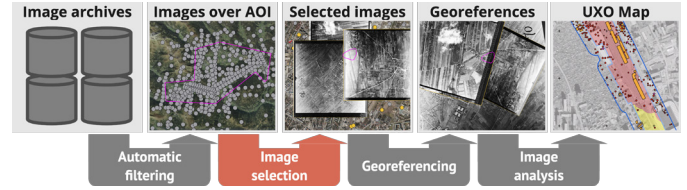

Fig. 2: UXO Detection workflow: Images that have coverage of the AOI are automatically extracted from different archives' catalogs, from which a small set of images is manually selected. Then, images are manually georeferenced and craters are identified to produce the UXO map. In this paper, we focus only on the process of image selection.

directing – providing a ranked set of suggestions, common in applications where different options can be clearly defined (e.g., sampling from a large parameter space, recommending different actions and analysis paths [11]); and *prescribing*, or automated and self-enacted decision-making, more commonly used in analytical systems to provide a starting point or path for analysis (e.g., touring the user through calculated interest points in very large images [15]). These types of guidance were recently abstracted into *system guidance tasks*, coupled to the analysis of user tasks, and further expanded into a typology by Pérez-Messina et al. [23], on which we base our task analysis (Sec. 3). A general framework for providing guidance in practical scenarios is still part of the guidance research agenda. However, there are attempts at implementing such aframework [31].

Guided decision making   Image Selection can be generalized as a broader problem in VA, namely, interactive optimization or Parameter Space Exploration (PSE) [28]. PSE deals with situations where not a single optimal solution but an infinite set of Pareto-optimal options exist, allowing users to interactively search in a wider range of, sometimes generative, solution design options [27]. VA systems supporting this task have already been applied in many domains: architecture and industrial design [1, 3, 33], space lighting design [36], yacht hull design [16], graphical layout design [8, 22], visual design [17], to name some. However, none of these approaches involves discrete combinatorial optimization, and neither explicitly employs guidance. To the latter point, we find exceptions in the work of El-Assady et al. [11] and Sperrle et al. [32], which study the effect of providing different guidance degrees to model building tasks and generalize Speculative Execution as a technique for guided PSE [30].

3 DOMAIN CHARACTERIZATION

The ultimate goal of UXO detection using aerial photographs is minimizing the risks and costs associated with on-site bomb detection and deactivation, for instance, on construction sites where armed confrontations (i.e., during wars) are known to have taken place. The final product of this process is a map showing risk-level zones and the location of identified UXOs. Obtaining this result is costly, either in terms of money and time, and can only be achieved through the meticulous reconstruction in time and space of the AOI by using archival photographic material, historical registries of bombings, and considerable location-specific tacit domain knowledge [13]. We can roughly divide the workflow into three manually-performed analytical stages: Image selection, image georeferencing, and image analysis (crater detection) as shown in Fig. 2. At the start of image selection, a set of all images that overlap with the AOI at some point in time is automatically extracted from international archives. Even for a very small AOI, the number of potentially useful images can be in the order of hundreds. For very small projects (but it varies widely with the distribution of attacks and images in time), a subset of 40-120 images will be selected, covering all known attacks so that a faithful reconstruction may be achieved and reduce the possibility that bombs may have escaped analysis, i.e., maximizing temporal coverage and image quality over the AOI. Experts mostly base their decision on (1) image metadata, (2) image preview (when available) to check if the image effectively covers

the AOI, and its quality (e.g., absence of clouds, damage, etc.), and (3) image temporal position in relation to recorded attacks. Once selected, images are retrieved (purchased from the archives if necessary) and then finely georeferenced (i.e., correctly transformed and placed over the same orthophoto map). Lastly, all images are examined in detail to pinpoint craters, to generate a hazard-level map, with markers indicating possible UXOs.

Problem statement   In this work, we focus on image selection for UXO detection. Abstractly, the problem can be stated as *picking a minimal subset of images of a subject (the AOI) that contains the information of all relevant changes of the subject in a timespan of years (bomb craters and damages) with a focus on particular events (aerial attacks)*. Considering that the value of an image does not only depend on its quality and metadata, but on its relations in time to attacks and to the other selected images, image selection represents a complex task involving unevenly-spaced time-oriented data which can greatly benefit from a VA approach. Our goal is also to provide guidance, as analysts usually face knowledge gaps that could hinder the analysis. Typically, the entire workflow is carried out manually by experts, as there is no commercially available specialized software for this task. Hence, our first aim was to design and implement a guidance-enriched VA tool integrating all the necessary features for the analysis, such as visualizing images and their metadata in time and space and with a clearly defined interactive workflow.

In the following, we present our domain characterization according to the data-users-task design paradigm [21]. We also describe how we designed the guidance workflow by performing guidance task analysis and providing essential definitions.

3.1   Data

A *project* in UXO detection is defined by an AOI (e.g., a construction site): this determines what set of images will be processed for selection and which attack registries (of aerial bombings) are relevant. The primary data source for UXO detection is photographic material (i.e., *aerial images*) captured by reconnaissance flights during WWII, which are stored in the US, UK, and Russian archives. Images correspond to a point in time and an area in space and are identified by the flight that took them, the camera's identification, and a sequence number. The relevant metadata of an image consist of: geographic center position, scale, camera, date (day), precalculated coverage over AOI (not accurate, as images are not properly georeferenced at this point), and owned status (whether they have already been acquired for a previous project). Typically, on reconnaissance flights, there are two types of cameras: *overview* and *detail*, the latter also having chirality (i.e., a pair of left and right cameras). Overview images have a larger scale, meaning they also have less information over a given area than detail images when coverage is fixed. Also, two subsequent detail images with the same chirality (from now on, an *image pair*) can be combined to produce a stereoscopic (3D), higher-resolution image. These properties define a hierarchy of images: Analysts will always prefer image pairs rather than single detail images, and only if detail images are not available will they select overview images. In other words, we can calculate, for each image, its *information value*, which is proportional to the pixels an image/pair contains for a fixed geographical area. These criteria are essential but do not represent hard constraints, as different experts will employ different strategies to deal with specific cases (e.g., some analysts would choose two detailed images from different dates rather than two unpaired images from one flight).

Equivalence classes   Images are related to the AOI spatially and to attacks temporally. If we consider images and attacks as two different types of nodes in a graph, we obtain that the temporal relationships between them form a directed bipartite graph. We say that an attack is *covered* by an image *iff* there is a spatial covering relation between AOI *and* the image date is (for obvious reasons) posterior and "relatively close" in time to when the attack took place. This relative closeness in time, which can be thought of as a *threshold*, depends on the characteristics of the project and can vary from some days to weeks (depending, for instance, on the weather and season). By suggestion of the domain

experts, we considered a 25-day period as our threshold, meaning that an image can cover an attack only if it was taken at most 25 days after the event. By applying this threshold, we obtain the actual coverage relationships between images and attacks. This relation allows us to group images into *equivalence classes*, i.e., two images belong to the same equivalence class if they cover the same attacks. This means that images belonging to the same flight are equivalent, but also that flights can be equivalent. This grouping is useful for visualization design and guidance. However, images that are closer to the attacks (i.e., the older flights of an equivalence class) will be preferred by analysts as their *delay* is smaller.

3.2   Task analysis

As preparation for the design, we analyzed the user's tasks, which were also used to characterize the necessary guidance to support them. We gathered the necessary information during a structured user workshop, in which we interviewed domain stakeholders about their expectations and requirements. We complemented this information with on-site interviews, with two additional domain experts who helped us to develop our design. The task analysis was then carried out by describing the user tasks and then the necessary guidance tasks were designed using the methodology by Pérez-Messina et al. (see [25] where an extended description of the user and guidance tasks can be found), arriving at the user-guidance task diagram in Fig. 3.

User tasks   The user task analysis displays an iterative structure (Fig. 3): The analysis starts with loading the data (UT0) and ends (the output) with a selected subset of images (UT6). Inside this process, the main user loop (UT1-5) takes place. In each iteration, one or more images are added (or removed from) the selection, which will become the result when the user decides to end the analysis (UT6). The overall analysis strategy resembles a divide-and-conquer approach, as the whole time period under analysis is split in each iteration into more manageable, sometimes causally independent subperiods (divide), in UT2, and then images are selected to cover each of them (conquer). We define the *target* of all user tasks as *a single or group of images* that are visually selected for inspection. Thus, all four search types (*lookup*, *browse*, *locate*, *explore*) are defined in relation images and their location in the representation (i.e., *path*). Next, we list and describe each task from Fig. 3.

UT0   **Produce** the *data* for the project by loading the images, attacks, and area of interest.

UT1   **Explore** the timeline *overview* by getting a sense of the distribution of flights and attacks and decide on a strategy to approach the analysis.

UT2   **Locate** a *timeframe* on the timeline in which to work, and zoom into.

UT3   **Browse** candidate images from *flights* of the focused timeframe.

UT4   **Lookup** the actual attack *coverage* of an image.

UT5   **Lookup** the metadata and geographically positioned preview of an *image* for selection.

UT6   **Produce** the *selection* of an image or a whole subset of images and exporting the result for further analysis.

Guidance tasks   Having analyzed the user tasks, we complemented them with guidance. Each user task hides a potential knowledge gap concerning its target or path of analysis, e.g., the images for selection or the time periods where they are found [25]. We identify two crucial user tasks, which are subject to the target unknown knowledge gap: the explore task (UT1) and the browse task (UT3) [4, 5]. While browsing the user does not know a priori what its target is but only that one may be found in the location where the search is being conducted. According to the guidance task typology [23], browsing tasks can only be directly supported by the *indicate* orienting guidance task, which reduces the target unknown knowledge gap without reducing user freedom.
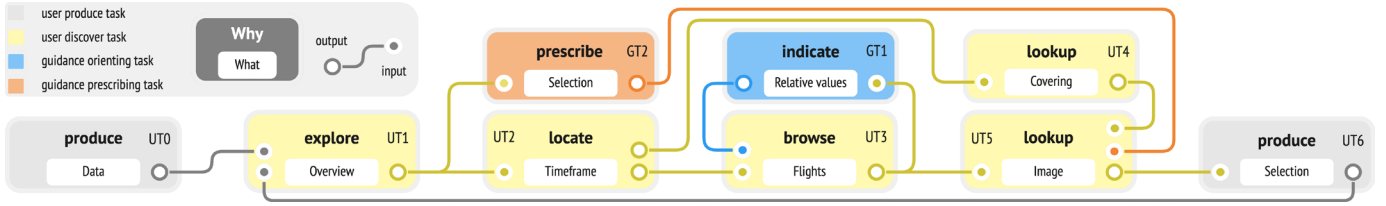
Fig. 3: User and guidance task schema of our guided VA system for Image Selection. Tasks are represented by their *why* and *what* dimensions [4, 23], and their input/output relations. Note that, although the whole process is represented as an iterative loop, the user is free to change tasks at any moment. The guidance prescribe task GT2 changes the user task from UT1 explore to UT5 lookup because of its disruptive nature [23].
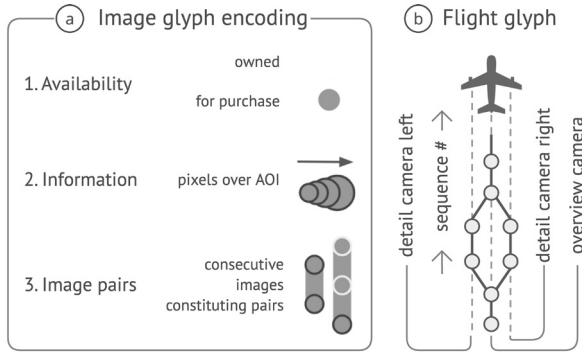

Fig. 4: Encoding of images and metadata. Images are visualized as circular glyphs (a) encoding three aspects: (1) availability status (presence of border), (2) amount of information (area of the circle) and (3) possible pairs (thicker link between the images). Image glyphs of the same flight are embedded into a flight glyph (b), where they are ordered vertically by sequence number and horizontally by camera type, a layout that resembles the reconnaissance flight.
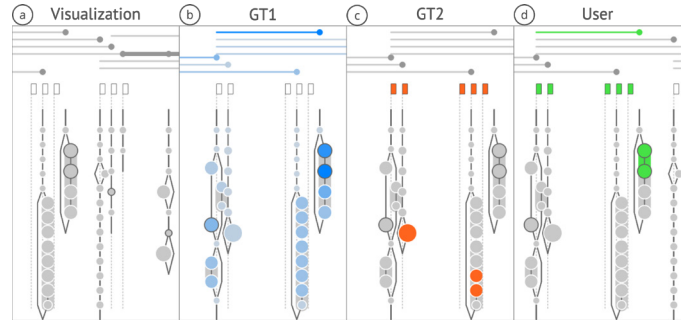

Fig. 5: The three-layer color schema to visually separate data, guidance, and user agency. A first layer in gray-scale contains the images, flight, attacks and equivalence class encodings (a). The color layers are superimposed in order: Orienting guidance in blue-scale shows the model's assessment of every image (b); prescribing guidance marks with orange its image selection and attack coverage (c); and, in the foremost layer, the user selection and attack coverage is set it green (d).

An explore task is subject to both target and path unknown knowledge gaps as it is involved in the generation and weaving of strategies and hypotheses [4, 5]. In interactive optimization problems this relates to the search for local optima from a global perspective. Although orienting guidance can also support this task [23], we follow the design guideline of the methodology for task-driven guidance design [25] and aim for the highest guidance degree (prescribe), which solves these knowledge gaps by providing a full solution to the problem.

Hence, we provide the following two guidance tasks:

GT1 Indicate the *relative value* of every image according to the model (relative to the images in its temporal vicinity, i.e., the locally normalized value). This task does not change the user task but only adds the information from the model's assessment to the visualization.

GT2 Prescribe a full *selection*. This task takes image and attack data and provides a direct answer to both target and path knowledge gap, delivering the user directly to the last search task of the interaction loop (UT5), where the user only needs to verify the suitability of the suggested images.

4 SYSTEM DESIGN

The main focus of the visualization interface is to allow the direct manipulation of the images and ease the selection process with guidance. The interface, shown in Fig. 1, consists of two coordinated views: a visualization of image metadata in time (Fig. 1a), and a geographic map for spatial data and image content previewing (Fig. 1b). The coordination between the two windows has the effect that filters, hover, and selection operations on the timeline visualization are also applied to the images displayed on the map. A video showcasing a usage scenario is available as supplemental material.

Implementation The system was developed as a plugin for QGIS to integrate it into the domain experts' workflow. The timeline visualization runs in JavaScript inside QGIS utilizing Qt for the integration and p5.js as visualization language and turf.js for the calculation of spatial coverage and geometric operations. The map visualization runs in QGIS-native Python.

4.1 Visualization

The main concerns addressed in the visualization design were driven by the gaps in the partner company's workflow: (G1) *having an overview of the image dataset in time* while (G2) *making them easily accessible for inspection*, and (G3) *visualizing the temporal coverage relations between images and attacks*. Hence, visualizing images in time and their relations to attacks (also determined by time) was our guiding principle. We decided to use linear time as the basic structure of the visual space, so images and attacks could enter into chronological relation by their position. However, the timestamp of attacks and images was only of day-granularity, thus needing of a different principle to organize images taken in the same day. The linear sequence imposed by the reconnaissance flights themselves was useful to abstract an ordering for same-day images.

We approached this using simple glyphs for images, which are embedded into a flight superglyph (Fig. 1a1). To show coverage relations between flights and attacks, and the current selection's attack coverage, a *temporal coverage band* (a horizontal strip with a visual encoding for attacks and equivalence classes) was placed in the upper part of the timeline (Fig. 1a2-3).

Another concern was to make the guidance visible and clearly recognizable from the data visualization. We used a layered design pattern with color as the visual dimension to separate the visualization of data from the encoding of guidance and from the state of user analysis. As shown in Fig. 5), the first layer contains all the information coming
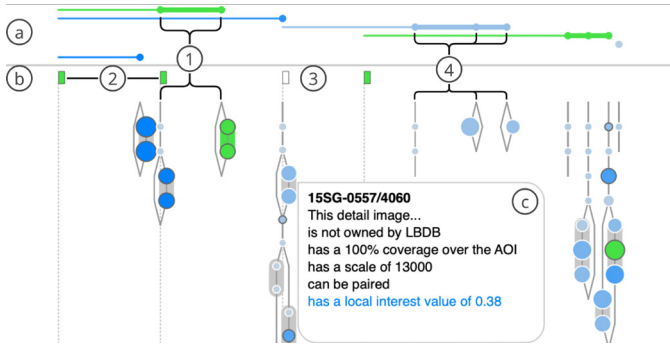
Fig. 6: Temporal coverage band visualizes the flight equivalence classes and their coverage relations to attacks (a), and the attack coverage of the current selection (b). In the example, a pair of images is selected (1) covering two attacks (2), while no images from another equivalence class (4) are selected leaving an attack uncovered (3). The user inspects an image that could cover this attack with the tooltip (c).

from the data and is represented in grayscale (a). Blue, red, and green are used to color the first layer with the guidance assessment of all images (b), guidance prescription (c), and current user selection (d), respectively. Here, hues are used only as categorical variables that differentiate the user agency from the guidance agency, and do not intend to convey any positive/negative valuation.

Next, we describe in detail the timeline layout and the encodings for images, flights, equivalence classes, and attack data.

Timeline   The timeline defines the layout in the horizontal axis for positioning flights and attacks in time (G1). The user can zoom into a time period and filter corresponding flights by swiping over the upper part of the timeline, which corresponds to the temporal coverage band (Fig. 6). In the temporal coverage band flights are linked to the attacks they cover (according to the 25-day rule), acting as a summary of the coverage relations between flights and attacks (G3). These coverage relations are grouped into the equivalence classes defined in Sec. 3.1. For each equivalence class, a line extends from the first attack covered to the last flight that covers it, marking the flights with larger circles (as flights belong only to one equivalence class but an attack can be covered by more than one flight). Attacks have a separate encoding resembling a flag that marks its position in time. The flags's color shows its coverage status (which can also be partial), allowing for quick coverage verification (G3). These encodings aimed to reduce visual clutter and avoid the introduction of new coordinated views, as using a node-link diagram or an adjacency matrix would have supposed, which would have created an unnecessary redundancy of the visual elements.

Images   We used an embedded glyph to represent images within flights (Fig. 4) which effectively represents, along with the position on the timeline and map, all relevant metadata fields described in Sec. 3.1. Each aerial image is represented by a circle encoding its amount of information over the AOI (size of area, Fig. 4a.1) and its owned status (presence of border, Fig. 4a.2). These glyphs are embedded into a (super-)glyph representing the flight that captured the photographs, which arranges the images according to their position within the flight (camera and sequence number, Fig. 4b) and their relation to other images (pairing, Fig. 4a.3). The shape of the flight glyph arises by ordering the images by sequence number on the vertical axis and by the camera (left-overview-right) on the horizontal axis, which is an abstraction that resembles the actual geographical positioning of the image centers along a flight, and it was praised by the domain analysts as a simple and effective representation. Hovering over an image reveals a tooltip with the textual metadata. It was important to use this custom embedded glyph design, as the flight structure conveys relevant information to the analysts and it is the first time it was visually represented, effectively supporting within- and inter-flight image browsing (G2).

Map   In the map (Fig. 1b), images are represented as circles that reveal the aerial photograph (grey/yellow circles represent the un/availability of a preview image) and allow the user to manipulate its position, rotation, and size (which affect its spatial coverage). The purple polygon represents the AOI. The map is necessary for the execution of some tasks, such as image quality assessment and georeferencing, however, we did not consider them in our user task analysis as our focus was on the abstract visualization and guidance parts.

4.2 Guidance

To visually differentiate the types of guidance (orienting and prescribing), we encoded them using different colors (Fig. 5b-c). In the following, we describe how we implemented the guidance tasks described in Sec. 3.

Orienting guidance   To provide orienting guidance for the browsing task (UT3), we need to *indicate* the relative importance of images (GT1). We achieve this by using an interest function (described in Sec. 5) that calculates the interest of the aerials considering the image metadata and their vicinity's precalculated interest. The resulting interest value is encoded using different saturations of blue. The preliminary interest value $preInterest(I) = owned \cdot paired \cdot information$ is calculated for each image by weighting its information value (spatial coverage and level of detail) by an `owned` factor and a `paired` factor. Then, the final interest value is calculated by normalizing each image value to the highest value found in its 25-day temporal neighborhood, as these represent possible alternative candidates.

Prescribing guidance   Prescribing guidance provides a direct solution to the task of finding a sound subset of images (GT2), using a heuristic function. The guidance is shown by adding the encoding of the selected aerials to the color layer of the prescribing guidance (red). To calculate a solution, we build upon the interest function already defined for orienting guidance and add an additional constraint for the equivalence classes. The solver takes the most interesting candidate image/pair from each equivalence class to cover all attacks (as not all equivalence classes are necessary to this end). Also, attacks that are covered by the prescribed images are also visualized in red, to show the effect of the guide's selection. The user can either pick or discard the suggested images or build upon them and improve the solution. Typically, the employed heuristic, which was used for the user study, does not arrive at Pareto-optimal solutions. In Sec. 8 we show how it can be improved using a second model which was developed by analyzing the study's results.

5   EXPLORATORY MODEL FOR SELECTION QUALITY

When designing the guidance-enhanced system, one of our goals was to determine what would constitute with sufficient precision a good solution to the image selection task, or in other words, capture the quality of a set of images selected by the user in order to evaluate user selections and create a better guidance model. To tackle this challenge, we created two models in two subsequent phases: First, we created an initial *exploratory model* to characterize and assess the quality of a set of aerial images. This first model was based on the results of our semi-structured interviews with two domain experts, during which we identified factors that they deemed useful to pursue the image selection task effectively. The model is composed of multiple partial metrics (indexes) that are used to evaluate different aspects of an image set. This first model proved useful to understand what was essential for the expert's criteria. This initial model was also used and evaluated during a first user study. According to the results gathered in this study, we could develop a second, more advanced, and precise, model, described in Sect.8. Interestingly, many metrics we considered for this first model resulted to be of secondary importance in our second model. However, some of the initial metrics which instead considered temporal relations between data cases were the most meaningful for an effective image selection. We describe our initial exploratory model in the following.

Partial indexes   As a first exploratory model, we devised a set of partial metrics to measure the desirable qualities of a selection, shown in Table 1. We call them indexes as all of them are normalized between

Index	Description
Intrinsic indexes	
Detail	Ratio of detail images within the selection (as opposed to overview images)
Information	Average information of images within the selection (normalized)
Pairing	Ratio of images that constitute image pairs (as opposed to single images)
timeCoverage	Ratio of time covered by images, within the whole time-length of the project
Resilience	Average number of images by which an attack is covered (up to a maximum of 5, normalized)
bestShot	Average score for each attack, where the score is the maximum image value found within its covering image set (normalized)
Overall	Average of all the above
Guidance indexes	
prescribedIndex	Ratio of selected images which were also being prescribed
orientingIndex	Average of the interest value per selected image given as orienting guidance.

Table 1: Description of the partial metrics used as an exploratory model to evaluate user study results.

0 and 1, for simplicity. The first set, the *intrinsic indexes*, measure properties of image metadata or relations between images and attacks. Each of these indexes aims to capture a partial aspect of the experts' criteria for selecting images, and thus none suffices alone as a comprehensive quality measure. Also, although a high value in each of the metrics is desired in a good-quality image set, it is not necessarily desirable for a set to reach the maximum value in all metrics. This is due to economic reasons (each image selected is resource-consuming, in terms of analysis time and money if the image needs to be purchased) and to expert knowledge that goes beyond the available data (e.g., the presence of unregistered bombings). The indexes Detail and Information measure metadata features of the selection; Pairing, the pair relation between selected images; timeCoverage, the distribution of the selection in time; Resilience and bestShot, the relations between selected images and attacks. We also devised the *guidance indexes* prescribedIndex and orientingIndex, to measure the similarity of a user's selection to the images suggested by our guidance model, to further investigate the results of the user study.

6 EVALUATION

To evaluate our VA tool for image selection and assess the effects of guidance, we performed a user study where we asked six domain experts to solve two real-world tasks using our system. After the study, we asked them to answer two Likert scale questionnaires about the system's usability. Finally, we had a semi-structured group discussion to gather more feedback about the system and the guidance. We inspected and analyzed the study results (the images selected by the participants) using the partial indexes described above. In addition, we assessed the quality of the submitted solutions with domain experts, who ranked them by quality and gave us a detailed explanation of their ranking decisions. We structured the study into two sessions, one in which participants received guidance and one without guidance. At the beginning of the study, we introduced the prototype with a 10-minute onboarding presentation. The onboarding was followed by a 5-minute in-platform tutorial, where the participants could learn the essential interaction with the tool (how to filter, browse and select images) and familiarize themselves with the tool. As experts in UXO detection are very rare, and because visualization experts were deemed not suited for evaluating this system for their lack of domain knowledge, we did not aim for a quantitative analysis.

6.1 Task-based User Study

Participants   We recruited six domain experts actively working in the field of UXO detection, not including experts E1-2 which were involved in the system design process. We asked each of them, in a 1-hour long session, to perform image selection on two different real-world datasets and answer a questionnaire after each task. The participants had varying levels of expertise in the specific task of image selection: 1 novice, 1 medium-experienced, 3 experts, and 1 senior expert with extensive knowledge of the Vienna area, where our test scenarios were situated.

Data   We used two real projects to identify possibly active bombs located around the center of Vienna, Austria. Due to time constraints, we reduced the projects to only the last year of imagery available (1945), which accounts for about a third of the whole time period of the projects. We selected these projects for having small AOIs within an urban area, which diminishes the geospatial complexity of images with highly varying coverage ratios but accentuates the complexity in the temporal dimension, with a dense distribution of attacks and flights.

Tasks   The participants were asked to perform the same task, namely, *to select the best-quality minimal subset of aerials that covers all attacks*, in both projects. When solving the first task, no guidance was made available to the participants, while during the second task, the guidance was set visible. Both tasks had a maximum time length of 20 minutes, during which no feedback was given to the participants except for technical questions, matters for which they were also provided with a cheat sheet regarding visualization and guidance encodings and interactions (provided as supplemental material). The participants were not asked to think aloud.

6.2 Questionnaires

After each task, the participants were asked to answer a 7-point Likert scale questionnaire to assess their evaluation of the system (questionnaires and anonymous answers are provided as supplemental material). The first questionnaire was constituted of 5 questions to evaluate the five user search tasks identified during the task analysis (Fig. 3a, UT1-5). Thirteen additional questions were selected from the ICE-T questionnaire [37]. The second questionnaire, which was given after the second task, was constituted of 12 questions to evaluate the effectiveness of the guidance the participants received, 3 questions directly corresponding to our proposed guidance tasks (Fig. 3b, GT1-2) and the same 5 user task-specific questions from the previous questionnaire. The latter questions were repeated to identify possible perceptual effects of guidance on the analysis. Every question was accompanied by a short example to facilitate its understanding. After the task sessions, a focus group session was performed where additional qualitative impressions, comments, and feedback was gathered.

7 RESULTS

7.1 Questionnaires

As shown in Fig. 7, the participants evaluated positively the visualization and the prototype, averaging 6+ points on the 7-point Likert scale, except for the 5.5 scores of the exploration task (UT1). This lower score can be explained by its higher abstraction level, requiring a higher cognitive load that is not completely alleviated by the visualization itself. The lowest scoring dimension in ICE-T was *confidence*, particularly in the "highlighting of unexpected data issues"(however, the design of the user study did not consider testing for such cases).

Guidance was particularly well appreciated for its *visibility* (it can be "easily identified and distinguished from the rest of the visual environment") but not for its *adaptiveness* (as the guidance was not designed to react to user selections). During the semi-structured feedback session, participants generally favored orienting guidance over prescribing

Left (T1 questionnaire):

	User Tasks					Insight										Time			Essence				Confidence	
avg.	5.7	6.2	6	6.5	6.5	5.5	5.3	5.5	6.2	5.8	6.2	5.7	6.2	5.5	6.2	5.5	5	5.7	4.2					
	UT1	UT2	UT3	UT4	UT5	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14					

Right (T2 questionnaire):

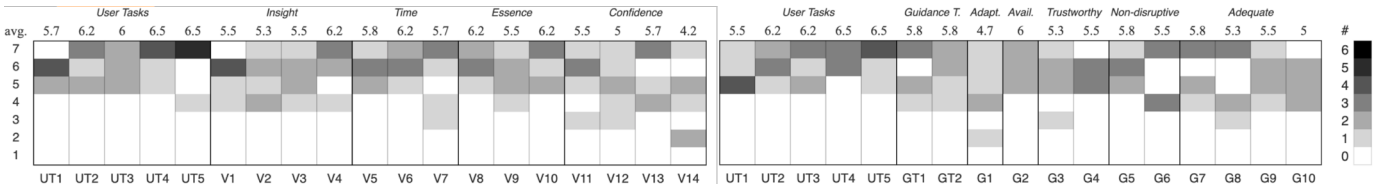|	User Tasks					Guidance T.		Adapt.	Avail.	Trustworthy		Non-disruptive		Adequate		
avg.	5.5	6.2	6.2	6.5	6.5	5.8	5.8	4.7	6	5.3	5.5	5.8	5.5	5.8	5.3	5.5	5
	UT1	UT2	UT3	UT4	UT5	GT1	GT2	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10

Fig. 7: Participant scores for the T1 questionnaire (left, focusing on the visualization) and T2 questionnaire (right, focusing on the guidance) on a 7-point Likert scale. The distribution of scores for each question is represented on a grayscale (rightmost). User Tasks questions (UT1-5) were repeated in the second questionnaire to test for perceptual independency of user and guidance tasks.

guidance, as the former *"gives more freedom"*, makes the best candidates *"easier to see"* and allows to *"pick the first spotlights"*, making the overall process *"faster with guidance."* This result is in line with the results obtained by El-Assady et al. [11], where solutions achieved with prescribing guidance were perceived by users as less valuable, due to the users' lack of agency, even though prescribed results had the highest quality according to the indexes.

The small average variation of task-specific question scores between with and without guidance questionnaires suggests that user and guidance tasks are perceived as different and can be measured independently, thus supporting the assumptions of task-based approaches, particularly for guidance. This is also an important consideration for including guidance in the nested workflow model for VA design and validation [12].

7.2 User study

Selection quality   We calculated for each participant's image set the partial indexes described in Sec. 5 to investigate the effect of guidance on the quality of the solution and measure how much the participants followed the provided guidance (Fig. 8a-g). We observe an increase in quality for most participants when tasks were supported with guidance, which is shown by a general increase in the `overall` index. Particularly, *four participants averaged worst than guidance in T1 (without guidance) and improved to a better-than-guidance average in T2 (with guidance).* Noteworthy cases are P1 (consistently averaging better than guidance in both tasks), P3 (consistently averaging below guidance in both tasks), and P5 (which shows the most significant quality improvement). P5 was an experienced analyst whose solutions were consistently top-ranked by experts (see Sec. 7.3); the improvement, in this case, can be explained with its compliance with the provided orienting guidance in T2 (Fig 8i), as this guidance was designed to optimize the criteria reflected by the partial metrics. By looking at the guidance-related metrics, we can characterize P3 as heavily biased towards distrusting guidance [20]. Also, P3's results in both tasks were qualitatively low according to metrics and experts alike. This indicates low compliance with the given task. Although a small fraction of the P3's selection was also part of the (invisible) prescribing guidance in T1, in T2, P3 selected none of the images suggested by the guidance, not improving its overall quality. The results of P1, the most knowledgeable and experienced analyst, shows that an effective human-computer collaboration may be unfeasible if the guidance model is not as proficient as the user.

Guidance Indexes   The `orientingIndex` and `prescribed Index` metrics (Fig. 8h-i) show how much the users followed the provided guidance in T2, in contrast to how much their solution for T1 followed the guidance calculated interests and suggestions (which was not shown to the participants during T1). We found, in some cases, a negative reaction to prescribing guidance, i.e., the prediction of prescribing guidance for T1 was higher than its effect on T2. This means that, as reported by participants in the group interview, they took the prescribed set as a *"starting point"* for their analysis and tried to improve it. A similar result was found for orienting guidance. This emergent "agonistic behavior" –where the user is put into a state of conflict towards the system suggestions– in guided VA systems was already hypothesized for disruptive degrees of guidance [23]. This leads us to observe that *user-diversion from guidance can also lead*

to improved results, as the guidance solution works only as a starting point for analysis.

Other effects   Although learning effects between tasks cannot be completely ruled out, we assume these did not influence the results of the study since participants were not given any feedback about their solutions and because most participants had long experience with image selection. Dataset-related effects (e.g., differences in the overall valuation of images) were reduced by considering the dataset differences for each partial metric and normalizing the results.

7.3 Expert evaluation

After running the study, we conducted an additional qualitative assessment of the results by showing them to our two domain expert partners E1-2, as they did not take part in the task-based user study. We asked them to rank independently –from what they thought was the best solution to the worst– the image sets submitted by the participants during the study according to their own (tacit) criteria and to comment on them. In addition to the participants' results, we also provided the experts with the solutions suggested by our guidance mechanism (for both projects) and the project *groundtruths* (the actual selection used in these past projects). In the interviews with the experts, and following a similar methodology as the Critical Decision Method for knowledge elicitation [7], we asked them to explain their ranking and their decision criteria clearly. Fig. 9a shows an agreement between experts in what constitutes a good solution, and thus that there are well-defined tacit quality criteria. The top 2 participants (P1 and P5, at the top) were defined unanimously as having provided the best solutions in both tasks. Same for what constitutes an unsuitable solution (0 and 4, at the bottom). According to these rankings, we were able to update the exploratory model we used to evaluate the study results to a non-parametric model that can predict the expert rankings and capture what constitutes a good solution for the domain experts (which we describe in Sec. 5) .

Analyzing the between-task rankings by expert E1 in Fig. 9b (the effect of guidance on user solution quality), we observe there are three position changes in the ranking of participants: P2, P4, and the ground truth itself (7). The quality of the ground truth is independent of our experiment and speaks to the fact that the quality of our partner company's image selection is highly variable and their UXO detection workflow and proofreading protocols can benefit from this work. P2, the participant with the least expertise in the task of image selection, improved *"from a barely usable to a good solution"*, according to E1, suggesting that guidance could resolve knowledge gaps regarding the familiarity with the task (i.e., an experience gap, as P2 had never seen what constituted a possible solution apart from its own until being provided with prescribing guidance in T2). P4 exceeded the time limit during T1 and had to be stopped at a point where the later attacks were not covered, explaining why the selection was deemed unusable (as it would raise ethical concerns to have no selected images after a registered attack). In T2, P4 was able to conclude the task in the allotted 20 minutes and showed a positive reaction to guidance, suggesting that guidance could be used to speed up analysis (although for all other participants, we did not see significant time differences between tasks, with and without guidance).
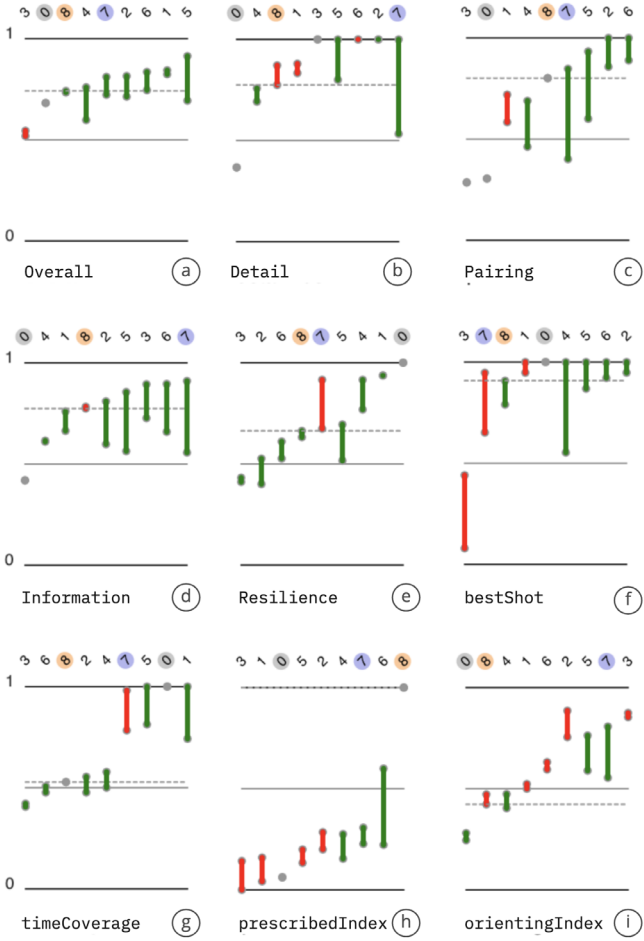
Fig. 8: Difference charts of the partial metrics of participant selections (1-6) for T1-2. The bars show the difference between each participant's performance for a given metric with (T2) and without (T1) guidance. A green bar represents an increase when performing with guidance, while a red bar a decrease. Participants are in each metric sorted by their performance on T2. The ground truth (7, blue), guidance solution (8, orange), and whole dataset (0, gray) are as well included. The dashed line marks the guidance solution score in T2 for reference.

Fig. 9: Comparison of rankings of solution quality produced by experts E1-2 (a), between tasks (b), and between E1 and our quality model M (c). Participant solutions (1-6) and colored by expertise level (darker shade of green is higher expertise). The ground truth (7, blue), guidance solution (8, orange), and whole dataset (0, gray) were as well included. Above each ranking is the number of edge crossings (X#).

8 SELECTION QUALITY MODEL

The expert evaluation provided us with further insight into the study results and a better understanding of the criteria experts used to determine the quality of solutions to the image selection task. Analyzing the collected data along with our partner experts, we could identify that from our partial metrics, timeCoverage and bestShot were the most significant for quality assessment, and building an improved guidance model. We identified six criteria (C1-6) that an image set must comply with, listed in the following in order of importance:

C1 All attacks must be followed at some point in time (not necessarily covered) by an image. Otherwise, the selection is deemed unethical.

C2 All attacks should be covered, when possible, by one (or more) image/s.

C3 Each attack should be covered by, in order of preference, an image pair, a single detail image, or an overview image.

C4 A *wave of attacks* (attacks that are very close in time) should be covered with minimum delay.

C5 Temporal extremes (the period before and after attacks) of the

project, even when there are no registered attacks, should be covered.

C6 If all of the above criteria are matched, no additional images should be selected (i.e., minimal set).

After gathering these criteria, we transformed them into mathematical functions derived from our first exploratory model and integrated them into a second non-parametric model. The new model comprises two metrics, representing the quality of a selection. The first, which captures how effectively (also from an economic perspective) each attack is covered by a set of images S, can be calculated as follows:

$$economyIndex(S) = \sum_{a=1}^{attacks} \frac{bestShot_a \cdot (1 - delay/cvgThreshold)}{selectedCoveringImages_a} \quad (1)$$

where $bestShot$ is the maximum information value found within the selected covering images of an attack, $cvgThreshold$ the threshold defined to the maximum time extent that is admitted for a coverage relation to be established, measures the average achieved quality of coverage for each attack, responding to C2-4 and C6. An exception is added: when there are no selected images after the last attack, this function returns 0 to account for C1. The second metric, which measures how well-distributed in time the selection and how much of the timespan of the project covers (independently of attacks), can be represented as follows:

$$timeCvgIndex(S) = \sum_{i=1}^{selected} saw_{cvgThreshold}(time_i - time_{i-1}), \quad (2)$$

where $saw_{cvgThreshold}$ is a linearly decreasing function with a period of $cvgThreshold$, and $time_i - time_{i-1}$ the temporal distance between image i and the previous image measures how even and expansive is the distribution of the selection in time, responding to C5. These two indexes, which measure competing criteria, are then united in the Quality Index

$$qualityIndex(S) = \frac{timeCvgIndex(S) \cdot economyIndex(S)}{attacks \cdot projectTimespan}, \quad (3)$$

where the optimization problem is represented as the maximization of the area defined by the economy and time coverage of the selection.

8.1 Quality model validation

We revalidated the model with the experts by showing them its output rankings. As shown in Fig. 9c, this improved selection quality model predicts the expert rankings and captures what constitutes a good solution for the domain experts with an error rate not much greater than

the between-expert variance (Fig. 9a). The experts were also able to explain the criteria behind the model as if it were their own rating and agreed with the model's ranking, just noting that the model had a stronger preference for economic solutions rather than more resilient ones.

9 DISCUSSION

Our guidance-enriched system was successfully deployed and evaluated in a user study with real-world scenarios, showing the potential of guidance to improve quality in fine-grained solutions. We used the results of our user study to improve our initial model to better align it with expert criteria. With only a 15-minute introduction, even the least experienced expert was able to produce a relatively good image selection, a result that experts deemed unlikely without our guidance-enhanced prototype. That speaks to the fact that a great part of the knowledge needed to perform the task is already condensed in the design of the visualization and guidance. In contrast, participants valued visualization confidence and guidance trustworthiness the least, unveiling a need for more explainability. This should be, however, handled with care, as some forms of explainability have been shown to build excessive trust in poorly performing systems [20].

We also observed that, as theorized in previous work on guidance, prescribing guidance raises agonistic behavior in users, i.e., some reacted to the guidance suggestion as something to overcome [23]. As the guidance model did not provide Pareto-optimal solutions during the user study, convergence towards prescribing guidance would only help participants who scored less `overall` than the guidance in T1.

Lessons learned　　The guidance design presented here was an instantiation of the methodology for task-driven guidance design by Pérez-Messina et al. [25], which also features more extensive task descriptions for this domain problem as a case study. The positive results of the evaluation support the effectiveness of this methodology, and its modular task nature makes it applicable to other problems of interactive optimization (e.g., domains that can be modeled by the problem-solving loop [18]). Scheduling problems, for instance, fall into a similar category of combinatorial optimization problems, where decisions have to be made about the temporal arrangement of elements to optimize multiple objectives.

Decision-making has been critically considered to be left out of the task literature [10]. In this work, we have tackled a problem involving explicit decision-making through a task-driven approach, implying that decision-making processes can be represented by a combination of *search* and *produce* tasks, where new data (i.e., the output of decisions) is created as a result of solving the search tasks and inputting these results back into the system. Furthermore, in our task analysis (Fig. 3), we considered an iterative task *loop* as its overall structure. To the best of our knowledge, the idea of task loops has not been explored outside interaction loops in games [14], although being fundamental in high-level VA models [5, 19, 26].

Echoing similar results of user studies in guidance-enhanced VA approaches involving decision-making (e.g., [11]), participants favored lower guidance degrees (i.e., users preferred orienting rather than prescribing guidance). This could indicate that users prefer to retain their agency and do not trust fully automated results, even when deemed optimal according to calculated metrics. Hence, prescribing guidance should not be used alone but complemented with the lower guidance degrees. A theoretical schema for such a mixed-initiative scenario has been proposed in the typology of guidance tasks used for our design [23]. More empirical studies to validate these claims are still an open area for research.

Furthermore, participants in our study could quickly get a grasp of the concepts of orienting and prescribing guidance, and differentiate them from the novel encodings of the visual environment, validating guidance degrees as an intuitive and effective framework to communicate design patterns and automated suggestions.

From our experience in this design study, we can confirm the importance and added value of mixed-initiative solutions. We highlight that the combined effect of guidance and direct manipulation afforded by the system was highly valued in the post-study discussion. A fully-automated or purely prescriptive solution would not offer the flexibility and customization that Pareto-front optimization scenarios pose.

Limitations　　In this work, we focused on UXO detection projects with small AOIs, more common within urban areas. Image selection tasks can also be performed on large projects with long AOIs spanning dozens of kilometers (e.g., for highways, power transmission lines, etc.) where selections comprise hundreds of images and even so spatial and temporal coverage might be sparse. This is a scalability issue that calls for a reevaluation of our visualization and guidance model design. Even for our user study, the guidance algorithm used to calculate image selections for prescriptive guidance was on the lower half of the expert rankings, but we were only able to build an improved model that predicted better the expert criteria through our retrospective analysis and expert evaluation of the user study results.

In general, our current system only addresses a few of the recommendations for highly-interactive optimization systems [19], which we believe stand for valuable guidelines. We regret that, due to the participants' time constraints, we could not test orienting and prescribing guidance in separate tasks. Also, a design and implementation of directing guidance could highly benefit our system and is also left for future work. Although the interaction of having to preview images before selecting them is admittedly cumbersome (also pointed out by the study participants), it was requested as a design "feature" by our partner experts, as they did not want that the step of verifying image quality could be neglected.

Future work　　Although our design and evaluation process allowed us to reach a model that satisfies expert criteria, this does not mean the task of image selection can be fully automated, but on the contrary, it means that better guidance can be provided and so improve human-computer collaboration and analysis quality even further. Our aim in the future is to go beyond the guidance described in this study towards *progressive* guidance. Progressive guidance can be defined as adapting the degree of guidance (orienting, directing, and prescribing) to better suit the users' needs. For example, users could tell the guidance the points in time when they consider that an image/pair should be searched for. This could be done prior to analysis (before the direct manipulation phase), explicitly during analysis or even captured as *implicit user feedforward* [5] from the temporal filter task (UT2). With this information, orienting guidance could be turned into directing guidance (by ranking images according to their calculated interest) or prescribing (by selecting the highest ranking image/pair, in case the interest difference with other candidates is larger than a certain threshold). In other words, by collecting more information about the analysis *path* (or progress), a guidance task can be played up or down to reduce the frictions that arise with higher (disruptive) guidance [23] and deliver a smoother human-computer collaboration.

10 CONCLUSION

We have presented a guidance-enhanced VA system for image selection within the UXO detection workflow. We have modeled this task as a multi-objective optimization task, where users can directly change the solutions prescribed by the guidance model. By testing our prototype in controlled setups with domain experts, using guided and unguided versions of the system, we found that guidance had a positive effect by estimating different quality metrics, even when user reaction to guidance was non-uniform. By closely analyzing the user study results with our partner experts, we were able to build an improved model that better aligns with their criteria, which can be used to provide more accurate guidance and improve UXO detection tasks. Domain experts validated the guidance and the VA prototype, and additional challenges for progress-adaptive guidance are discussed for future work.

11 ACKNOWLEDGMENTS

REFERENCES

[1] A. M. Abuzuraiq and H. Erhan. The many faces of similarity: A visual analytics approach for design space simplification. *RE: Anthropocene, Design in the Age of Humans - Proceedings of the 25th CAADRIA Conference*, 1:485–494, 2020. doi: 10.52842/conf.caadria.2020.1.485 2

[2] A. Amor-Amorós, P. Federico, S. Miksch, S. Zambanini, S. Brenner, and R. Sablatnig. Visual analytics for multitemporal aerial image georeferencing. In *EuroVA@ EuroVis*, pp. 55–59, 2017. doi: 10.2312/eurova. 20171120 2

[3] F. Bao, D.-M. Yan, N. J. Mitra, and P. Wonka. Generating and exploring good building layouts. *ACM Trans. Graph.*, 32(4), 2013. doi: 10.1145/ 2461912.2461977 2

[4] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2376–2385, 2013. doi: 10.1109/TVCG.2013.124 3, 4

[5] D. Ceneda, T. Gschwandtner, T. May, S. Miksch, H.-J. Schulz, M. Streit, and C. Tominski. Characterizing guidance in visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):111–120, 2017. doi: 10.1109/TVCG.2016.2598468 2, 3, 4, 9

[6] D. Ceneda, T. Gschwandtner, and S. Miksch. A review of guidance approaches in visual data analysis: A multifocal perspective. In *Computer Graphics Forum*, vol. 38, pp. 861–879. Wiley Online Library, 2019. doi: 10.1111/cgf.13730 2

[7] L. Cibulski, H. Dimara, S. Hermawati, and J. Kohlhammer. Supporting domain characterization in visualization design studies with the critical decision method. In *IEEE VisGuides: Visualization Guidelines in Research, Design, and Education*, 2022. doi: 10.1109/VisGuides57787.2022.00007 7

[8] N. R. Dayama, K. Todi, T. Saarelainen, and A. Oulasvirta. Grids: Interactive layout design with integer programming. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–13. ACM, New York, 2020. doi: 10.1145/3313831.3376553 2

[9] F. de Luz Palomino Valdivia, H. A. H. Baca, I. S. Solis, M. A. Cruz, and A. M. C. Valdivia. Guided interactive visualization for detecting cyclical patterns in time series. In *Advances in Information and Communication: Proceedings of the 2023 Future of Information and Communication Conference (FICC)*, vol. 1, pp. 391–402. Springer, 2023. doi: 10.1007/ 978-3-031-28076-4_29 2

[10] E. Dimara and J. Stasko. A critical reflection on visualization research: Where do decision making tasks hide? *IEEE Transactions on Visualization and Computer Graphics*, 28(1):1128–1138, 2022. doi: 10.1109/TVCG. 2021.3114813 2, 9

[11] M. El-Assady, F. Sperrle, O. Deussen, D. Keim, and C. Collins. Visual analytics for topic model optimization based on user-steerable speculative execution. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):374–384, 2018. doi: 10.1109/TVCG.2018.2864769 2, 7, 9

[12] P. Federico, A. Amor-Amorós, and S. Miksch. A nested workflow model for visual analytics design and validation. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, pp. 104–111. New York, 2016. doi: 10.1145/2993901. 2993915 7

[13] P. Federico, M. Wagner, A. Rind, A. Amor-Amorós, S. Miksch, and W. Aigner. The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 92–103. IEEE, 2017. doi: 10.1109/VAST.2017.8585498 2

[14] B. Horn, S. Cooper, and S. Deterding. Adapting cognitive task analysis to elicit the skill chain of a game. CHI PLAY '17, p. 277–289. ACM, New York, 2017. doi: 10.1145/3116595.3116640 9

[15] C. Y. Ip and A. Varshney. Saliency-assisted navigation of very large landscape images. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):1737–1746, 2011. doi: 10.1109/TVCG.2011.231 2

[16] S. Khan, E. Gunpinar, and B. Sener. Genyacht: An interactive generative design system for computer-aided yacht hull design. *Ocean Engineering*, 191:106462, 2019. doi: 10.1016/j.oceaneng.2019.106462 2

[17] Y. Koyama, I. Sato, and M. Goto. Sequential gallery for interactive visual design optimization. *ACM Transactions on Graphics (TOG)*, 39(4):88–1, 2020. doi: 10.1145/3386569.3392444 2

[18] J. Liu, T. Dwyer, K. Marriott, J. Millar, and A. Haworth. Understanding the relationship between interactive optimisation and visual analytics in the context of prostate brachytherapy. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):319–329, 2018. doi: 10.1109/TVCG.2017. 2744418 9

[19] J. Liu, T. Dwyer, G. Tack, S. Gratzl, and K. Marriott. Supporting the problem-solving loop: Designing highly interactive optimisation systems. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1764–1774, 2020. doi: 10.1109/TVCG.2020.3030364 9

[20] J. Liu, K. Marriott, T. Dwyer, and G. Tack. Increasing user trust in optimisation through feedback and interaction. *ACM Transactions on Computer-Human Interaction*, 2022. doi: 10.1145/3503461 7, 9

[21] S. Miksch and W. Aigner. A matter of time: Applying a data–users–tasks design triangle to visual analytics of time-oriented data. *Computers & Graphics*, 38:286–290, 2014. doi: 10.1016/j.cag.2013.11.002 3

[22] P. O'Donovan, A. Agarwala, and A. Hertzmann. Designscape: Design with interactive layout suggestions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 1221–1224, 2015. doi: 10.1145/2702123.2702149 2

[23] I. Pérez-Messina, D. Ceneda, M. El-Assady, S. Miksch, and F. Sperrle. A typology of guidance tasks in mixed-initiative visual analytics environments. In *Computer Graphics Forum*, vol. 41, pp. 465–476. Wiley Online Library, 2022. doi: 10.1111/cgf.14555 2, 3, 4, 7, 9

[24] J. Puchinger, G. R. Raidl, and U. Pferschy. The multidimensional knapsack problem: Structure and algorithms. *INFORMS Journal on Computing*, 22(2):250–265, 2010. doi: 10.1287/ijoc.1090.0344 2

[25] I. Pérez-Messina, D. Ceneda, and S. Miksch. A Methodology for Task-Driven Guidance Design. 2023. doi: 10.2312/eurova.20231094 3, 4, 9

[26] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim. Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1604–1613, 2014. doi: 10. 1109/TVCG.2014.2346481 9

[27] A. Schulz, H. Wang, E. Grinspun, J. Solomon, and W. Matusik. Interactive exploration of design trade-offs. *ACM Trans. Graph.*, 37(4), 2018. doi: 10. 1145/3197517.3201385 2

[28] M. Sedlmair, C. Heinzl, S. Bruckner, H. Piringer, and T. Möller. Visual parameter space analysis: A conceptual framework. 20(12):2161–2170. doi: 10.1109/TVCG.2014.2346321 2

[29] E. J. Shepherd. Mapping unexploded ordnance in italy: The role of world war II aerial photographs. *Conflict Landscapes and Archaeology from Above*, pp. 205–218, 2016. 1

[30] F. Sperrle, J. Bernard, M. Sedlmair, D. Keim, and M. El-Assady. Speculative execution for guided visual analytics. In *2018 IEEE Workshop on Machine Learning from User Interaction for Visualization and Analytics (MLUI)*. IEEE Computer Society, Los Alamitos, 2018. doi: 10. 1109/MLUI52768.2018.10075559 2

[31] F. Sperrle, D. Ceneda, and M. El-Assady. Lotse: A practical framework for guidance in visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 2022. doi: 10.48550/arXiv.2208.04434 2

[32] F. Sperrle, R. Sevastjanova, R. Kehlbeck, and M. El-Assady. Viana: Visual interactive annotation of argumentation. In *2019 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 11–22, 2019. doi: 10 .1109/VAST47406.2019.8986917 2

[33] N. Umetani, T. Igarashi, and N. J. Mitra. Guided exploration of physically valid shapes for furniture design. *Communications of the ACM*, 58(9):116–124, 2015. doi: 10.1145/2185520.2185582 2

[34] S. Van der Baan. Keynote-geophysical practice of uxo detection. In *1st Conference on Geophysics for Infrastructure Planning Monitoring and BIM*, vol. 2019, pp. 1–5. European Association of Geoscientists & Engineers, 2019. 1

[35] M. Wagner, A. Rind, N. Thür, and W. Aigner. A knowledge-assisted visual malware analysis system: Design, validation, and reflection of kamas. *Computers & Security*, 67:1–15, 2017. doi: 10.1016/j.cose.2017. 02.003 2

[36] A. Walch, M. Schwärzler, C. Luksch, E. Eisemann, and T. Gschwandtner. Lightguider: Guiding interactive lighting design using suggestions, provenance, and quality visualization. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):569–578, 2020. doi: 10.1109/TVCG.2019. 2934658 2

[37] E. Wall, M. Agnihotri, L. Matzen, K. Divis, M. Haass, A. Endert, and J. Stasko. A heuristic approach to value-driven evaluation of visualizations. *IEEE transactions on visualization and computer graphics*, 25(1):491–500, 2019. doi: 10.1109/TVCG.2018.2865146 6