DIPA2: An Image Dataset with Cross-cultural Privacy Perception Annotations

ANRAN XU, Interactive Intelligent Systems Lab., The University of Tokyo, Japan
ZHONGYI ZHOU, Interactive Intelligent Systems Lab., The University of Tokyo, Japan
KAKERU MIYAZAKI, Interactive Intelligent Systems Lab., The University of Tokyo, Japan
RYO YOSHIKAWA, Interactive Intelligent Systems Lab., The University of Tokyo, Japan
SIMO HOSIO, Center for Ubiquitous Computing, University of Oulu, Finland
KOJI YATANI, Interactive Intelligent Systems Lab., The University of Tokyo, Japan

The world today is increasingly visual. Many of the most popular online social networking services are largely powered by images, making image privacy protection a critical research topic in the fields of ubiquitous computing, usable security, and human-computer interaction (HCI). One topical issue is understanding privacy-threatening content in images that are shared online. This dataset article introduces DIPA2, an open-sourced image dataset that offers object-level annotations with high-level reasoning properties to show perceptions of privacy among different cultures. DIPA2 provides 5,897 annotations describing perceived privacy risks of 3,347 objects in 1,304 images. The annotations contain the type of the object and four additional privacy metrics: 1) information type indicating what kind of information may leak if the image containing the object is shared, 2) a 7-point Likert item estimating the perceived severity of privacy leakages, and 3) intended recipient scopes when annotators assume they are either image owners or allowing others to repost the image. Our dataset contains unique data from two cultures: We recruited annotators from both Japan and the U.K. to demonstrate the impact of culture on object-level privacy perceptions. In this paper, we first illustrate how we designed and performed the construction of DIPA2, along with data analysis of the collected annotations. Second, we provide two machine-learning baselines to demonstrate how DIPA2 challenges the current image privacy recognition task. DIPA2 facilitates various types of research on image privacy, including machine learning methods inferring privacy threats in complex scenarios, quantitative analysis of cultural influences on privacy preferences, understanding of image sharing behaviors, and promotion of cyber hygiene for general user populations.

CCS Concepts: • **Security and privacy → Privacy protections**; **Usability in security and privacy**.

Additional Key Words and Phrases: Dataset, image privacy, usable security, cross-cultural study

ACM Reference Format:
Anran Xu, Zhongyi Zhou, Kakeru Miyazaki, Ryo Yoshikawa, Simo Hosio, and Koji Yatani. 2023. DIPA2: An Image Dataset with Cross-cultural Privacy Perception Annotations. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 4, Article 192 (December 2023), 30 pages. https://doi.org/10.1145/3631439

Authors' addresses: Anran Xu, Interactive Intelligent Systems Lab., The University of Tokyo, Tokyo, Japan, anran@iis-lab.org; Zhongyi Zhou, Interactive Intelligent Systems Lab., The University of Tokyo, Tokyo, Japan, zhongyi@iis-lab.org; Kakeru Miyazaki, Interactive Intelligent Systems Lab., The University of Tokyo, Tokyo, Japan, kakeru-miyazaki@iis-lab.org; Ryo Yoshikawa, Interactive Intelligent Systems Lab., The University of Tokyo, Tokyo, Japan, ryo@iis-lab.org; Simo Hosio, Center for Ubiquitous Computing, University of Oulu, Oulu, Finland, simo.hosio@oulu.fi; Koji Yatani, Interactive Intelligent Systems Lab., The University of Tokyo, Tokyo, Japan, koji@iis-lab.org.

2474-9567/2023/12-ART192 $15.00
https://doi.org/10.1145/3631439

1 INTRODUCTION

Visual content empowers people's communication and information sharing. This pervasive sharing of images and video online leads to various interpersonal privacy leakages [48]. As a simple everyday example, commoditized digital cameras are a source of privacy conflicts between photographers and bystanders [56]. Researchers in different fields have proposed mitigating similar issues with usable security [33, 49], and different privacy detection methods have been proposed to mitigate e.g. sharing of images that contain passersby who don't want to be recorded [7, 21]. Other typical research areas in this context include identifying content that might be considered privacy-threatening (see e.g. [39, 54]) or studying sharing behaviors in visual media for different user groups including vulnerable populations [4, 8, 24, 56, 68].

While prior ubiquitous computing literature suggests the importance of investigations on privacy protection support, the lack of open resources to facilitate such research remains a substantial burden [23, 54, 63]. Although open image privacy datasets for ubiquitous computing exist [54, 85], they are largely tailored for objective tasks (e.g., detecting if specific privacy-threatening content exists). Other datasets consider people's appearances and behavior in smart home settings (e.g., face, nudity, and relationship) [78]. The dataset establishment was further advanced by using imitated objects to substitute real private objects [63]. While these datasets may benefit building ubiquitous computing applications, the annotations do not include information about human perception, such as how severe people think the threat would be, how strictly people would refrain such content from being shared, and how human factors such as individual characteristics as well as cultural backgrounds influence the perception of privacy. Combining visual annotations with such information, researchers may facilitate image privacy protection in multiple aspects, i.e., providing personalized or culture-specific recommendations on proper risk perception or establishing decision-making management on image sharing based on an individual's social relationships. Thus, an image dataset covering fine-grained object-level annotations on what people considered privacy-threatening, complemented by privacy metrics reasoned by all kinds of people, will benefit researchers in investigating usable security and building ubiquitous and machine learning (ML) applications with considerations on human factors.

This paper presents DIPA2 [1]: an image dataset containing annotations that emphasize user-perceived privacy and security. DIPA2 provides 5,897 annotations on 3,347 different objects in 1,304 images, thus illustrating a broad range of different perceived privacy risks. Each object-level annotation includes the category label of the object (e.g. person, accessory, and clothing) and four additional privacy-related metrics: 1) information type, referring to what type of information the annotated content threatens to reveal; 2) informativeness, referring to how severe the potential privacy risk is; 3) sharing scope as a photo owner, referring to which groups the person would be willing to share the image as its owner; and 4) sharing scope by others, referring to which groups would they allow others to be able to share the image. Further, the dataset combines annotations from people in two countries (Japan and the United Kingdom) to illustrate cultural differences in privacy perception. In this paper, we also discuss the methods of how this was achieved, enabling others to easily extend the dataset to their cultural needs. Each image in DIPA2 was annotated by four annotators, two from Japan and two from the U.K. We also collected demographic information and Big Five personality traits of all the 600 annotators (300 from Japan, 300 from the U.K.) to enable a novel cross-cultural analysis of the dataset. DIPA2 facilitates various types of research in image privacy by offering multidimensional data. The dataset can enable sociologists and quantitative researchers in privacy aspects to drive investigations on how people perceive privacy in images in a quantitative manner as well as ubiquitous computing and user interface researchers to build applications with considerations on image privacy through computer vision (CV) and ML methods. It can also be used to support research on other visual media, such as videos, because of the commonality of visual media.

Together this paper and the dataset itself, DIPA2, offers the following contributions.

[1]The dataset can be downloaded at https://anranxu.github.io/DIPA2_VIS/

- DIPA2 dataset construction including 1,304 images with 5,897 annotations specifically dedicated to user-perceived privacy and security aspects;
- A comprehensive data analysis on DIPA2 including basic statistical analysis and machine-learning baselines to exhibit the potential of the dataset for reasoning privacy threats and anticipating people's sharing intentions;
- A discussion of possible future usage scenarios of DIPA2 for different researchers and practitioners.

This work substantially extends Xu et al.'s work on DIPA [79] in the following three ways: 1) redesigning question descriptions and option choices for privacy metrics after scrutinized justification; 2) approximately doubling the number of annotators than participants recruited in DIPA to cover more persons' opinions in different characteristics and backgrounds; and 3) exhaustive data analysis and comparison with DIPA to help researchers quickly familiarize themselves with the data distribution and explore their studies. DIPA2 presents a versatile dataset for a broad set of researchers and practitioners interested in studying image privacy particularly at the object level with human factors as well as developing applications with image privacy considerations.

2 RELATED WORK

While different fields have different research focuses on image privacy, we present the most closely related work on image content detection and classification, studies on user intent to share images online, and other publicly-available image datasets focusing on privacy.

2.1 Research on Image Privacy

2.1.1 Privacy-threatening Content Detection. With the advance of pattern recognition techniques, researchers attempted to achieve the detection of privacy-threatening content, such as human (faces) [16, 25, 36, 62, 84], computer screens [32], vehicle plates [82], critical documents [73], and personal information [54]. Many of these studies primarily focused on algorithm development, leaving the task of forming rigorous privacy definitions for other research efforts (see Section 2.1.2).

Another research focus lies on protecting the privacy of unintended people captured in photos (i.e., bystanders). In addition, mischaracterization associated with AI fallibility could cause shared ethical concerns between photographers and bystanders [5]. Protection of bystanders' privacy is also an important research topic in ubiquitous computing. Researchers utilized machine-learning approaches to automatically obfuscate the faces of unintended people and then offer tagging suggestions for intended people in photos [58, 77]. Other researchers achieved the protection of bystanders' privacy in ubiquitous videography taken by wearable cameras [7, 21] and on-dash mounted cameras [55]. Their systems integrated activity-oriented recognition to balance the utility of videos and the privacy concerns of bystanders. The detection technologies have been further advanced by Steil et al.'s work, through analyzing deep scene images and eye movement features [67].

While these studies successfully enable us to detect privacy-threatening content in given visual media and user groups, an obstacle exists in the difficulty of conducting follow-up research and comparing existing studies due to the lack of open research resources. We, therefore, argue that a public-available dataset with metrics on image privacy associated with user perception will thus contribute to the development and reproducibility of image privacy detection approaches as well as cross comparisons of proposed methods.

2.1.2 Classification of Privacy-threatening Content. Prior research attempted to systematically summarize categories of privacy-threatening content in images. Yu et al. proposed a recommendation algorithm trained with 268 types of privacy-threatening content on privacy settings during photo sharing [81]. Hoyle et al. conducted an in-depth investigation into privacy norms, considering specific scenarios like household rooms and conditions such as whether privacy-threatening content and people were present [28]. Stangl et al. summarized 21 categories

of image content from people with visual impairments (PVIs), providing a rating score to assess privacy concern levels for each type of content across five different sharing scenarios [66].

Li et al. identified 28 privacy-threatening content categories through analyzing privacy-threatening photos from 116 participants [39]. Their study provided fine-grained analysis from participants, including the likelihood of sharing each privacy-threatening content and the likelihood of sharing images with given recipients (e.g., family, friends, and colleagues). While the study by Li et al. offers a set of privacy-threatening categories with reasonable justification, researchers still encounter difficulties in developing low-level understandings of privacy issues without being able to access actual images. For instance, image privacy concerns may vary depending on scenarios and personal preferences. Deeply understanding user behavior around image privacy thus requires image data that accompany annotations about perceptions of privacy-threatening content from a set of different users. Therefore, we argue a dataset that provides object-level annotations in multiple aspects of privacy will offer researchers resources to explore more complex analyses and inspire more studies in image privacy.

In practice, whether a given content threatens users' privacy highly depends on scenarios and personal preferences. Recent breakthroughs in explainable AI demonstrated promising outcomes in determining whether a combination of visual contents can be considered privacy-threatening from a general perspective [10]. Meanwhile, other studies [75] enabled personalized privacy detection through deep learning through implicit inferences. Different from analyzing whether a category in general would threaten users' privacy, in this work, we aim to understand the privacy issue using the dataset approach. In our dataset, we believe that each image represents a specific scenario which can effectively reduce the ambiguity of the analysis and support the advance of explainable AI. We further provide cultural factors and fine-grained privacy metrics in each annotation to support the explicit understanding of image privacy in human-centered contexts.

2.1.3 Associations Between Image Privacy and Sharing Intentions. Photo sharing is a common activity in online social and communication platforms, but it is also a common cause of privacy leakages [38, 45]. To avoid observations from malicious people, many people share photos with specific groups, such as family, friends, or customized recipients [35]. Researchers have also investigated mechanisms of privacy conflict and proposed solutions for assisting people to better manage photo sharing with other stakeholders (e.g., co-owners [2, 64, 80] and bystanders [42]).

Sharing intentions (i.e., mental models on managing sharing behaviors) depend on objective factors (e.g., content in images [1, 51], demographic [14]) and subjective factors (e.g., personal traits [55]). Prior studies explored users' privacy preferences when uploading photos, and found a large difference in setting private or public tags when the content in images was varied [3]. Different perceptions of privacy protection may cause privacy conflicts when images involve multiple stakeholders, resulting in privacy violations to part of photo owners [68]. People's attitudes toward whether to share an image were influenced by their ownership of the image with different concerns on privacy [8].

Demographic differences are another important factor in photo-sharing behavior. Existing empirical studies have shown that women were more careful about their privacy than men [11, 27, 72]. However, as research also found that women generally shared more images than men [44, 70], this phenomenon, therefore, requires continuous and more holistic explorations. Besides gender, cultural differences can also affect the perception of privacy in images and self-disclosure behavior [6, 17, 41]. As existing studies mainly investigated image privacy in North American regions and culture [37, 39, 76], cross-cultural investigations on privacy preferences can fill unexplored spaces in this field.

Prior work has also shown that people's personalities can influence photo-sharing preferences [29, 30]. People who frequently used aggressive and self-deprecating humor were more likely to violate others' privacy by sharing photos [24]. These studies emphasize the importance of introducing human factors when building a public-available dataset to support various human-centered research (e.g., privacy conflicts) in image privacy.

2.2 Public-available Datasets on Image Privacy

Table 1 compares existing datasets on image privacy. Early work provided basic image-level annotations showing whether an image was private or not [65, 83]. Two recent datasets in this field are called VISPR (Visual Privacy Dataset) [54] and PrivacyAlert [85], which provide image-level annotations on what content in their images is privacy-threatening. However, these two datasets do not offer an explicit understanding of why a given image was considered privacy-threatening. Most of the annotations in VISPR and PrivacyAlert were not at the object level, although the researchers of VISPR offered a subdataset of object-level annotations on text-based private information (e.g., identification and receipts) in a later extension [53]. We argue an image privacy dataset should not only provide basic information about what content is considered to threaten privacy, but also illustrate detailed reasoning of why and how the content can potentially influence privacy in multiple dimensions.

A rising trend in the field of image privacy involves the creation of datasets specifically tailored for the assistive technology community. VizWiz-Priv [23] is a large-scale image privacy dataset, centering on private photographs taken by individuals with visual impairments, accompanied by questions for assistance. While VizWiz-Priv removed all privacy-threatening content in their data to avoid privacy violation, BIV-Priv [63] leveraged a prop-replacement method, where actual appearances of private content were substituted with provided props, to ensure the analysis of privacy-threatening data while respecting participants' privacy. In addition, most of the recent datasets [23, 53, 54, 85] employed only a small group of annotators in their building process. While this approach may have contributed to more stable annotations, it may also limit the diversity and richness of perspectives, as privacy concerns can greatly vary based on individual differences [20] and societal factors, such as economic and ethical considerations [50].

Xu et al. [79] recently published DIPA, a dataset tailored toward image privacy that provides object-level annotations on selected images from existing large-scale datasets (OpenImage [34] or LVIS [22]). Each image was assigned one annotator from Japan and one from the U.K. to emphasize cultural influences on privacy concerns. Three extra privacy metrics besides category names are offered in each annotation to illustrate the information type and informativeness level of identified privacy-threatening content, along with annotators' sharing intentions on the given content (Table 2). This work expands DIPA by introducing four privacy metrics to measure more complex privacy issues in images provided by DIPA. We recruited approximately twice as many annotators in the building of DIPA2 to better include individual differences and cultural perceptions in image privacy. Researchers are able to use DIPA2 to more precisely explore privacy issues and corresponding sharing intentions from different representatives.

There exist other types of image privacy dataset construction based on image synthesis [13, 78] and adversarial perturbation [15]. However, they are tailored to defend against malicious attacks from specific machine-based approaches, and they fall outside the scope of this work, constructing a dataset for both CV research and interactive applications.

3 DATASET CONSTRUCTION

3.1 Annotations in DIPA

DIPA [79] includes 5,671 annotations, provided by 177 Japanese annotators and 183 British annotators, for visual contents that are considered privacy-threatening in 1,495 images selected from existing datasets [22, 34]. DIPA provides basic demographic information (age, gender, and nationality) and Big-five personality test results (extraversion, agreeableness, conscientiousness, neuroticism, and openness) about each annotator [59]. For each content identified as privacy-threatening in DIPA, it provides three extra privacy metrics along with a category tag (e.g., person and place identifier). Table 2 provides corresponding question descriptions and options for each privacy metric. The three privacy metrics are composed of "Information type" referring to what kind of information could be inferred from the selected privacy-threatening content; "Informativeness" referring to

Table 1. A comparison of existing image privacy datasets, reproduced from a literature review [40]. Some of the listed datasets also contain images that were considered as not privacy-threatening. Only images that were judged as the equivalent of "private" in these datasets are included for comparison.

Dataset (Year)	Available Annotations	Major Image Content	Number of Private Images	Annotators
PicAlert [83] (2012)	image label (binary)	daily-life images	4,701	81 annotators
YourAlert [65] (2016)	image label (binary)	daily-life images	1,511	27 annotators
VISPR [54] (2017)	image label (category)	daily-life images	22,167	3 annotators
VISPR-extension [53] (2018)	object label information type	text-based private information	8,473	5 annotators
VizWiz-Priv [23] (2019)	object label paired questions in images	first-person viewpoint photos taken by visually impaired people	5,537	1 annotator 2 reviewers
PrivacyAlert [85] (2022)	image label (category)	daily-life images	13,910	8 annotators
BIV-Priv [63] (2023)	image label (category)	first-person viewpoint photos taken by visually impaired people	728 (corresponding with the same amount of videos)	26 participants with visual impairments
DIPA [79] (2023)	object label information type informativeness maximum sharing scope (as a photo owner)	daily-life images	1,495 (annotated twice)	177 annotators from Japan 183 annotators from the U.K.
DIPA2 (**ours**)	object label information type informativeness sharing scope (as a photo owner) sharing scope (by others)	daily-life images	1,304 (annotated four times)	300 annotators from Japan 300 annotators from the U.K.

perceived threatening levels if the content is leaked to malicious people; and "Maximum sharing scope" referring to how broadly the annotators would be willing to share the content if annotators owned it. A comprehensive conclusion of the DIPA data collection process can be found in Appendix A.

3.2 Revised Metrics in DIPA2

While we decided to keep the three privacy metrics used in DIPA (Table 2), we employed revisions and expansions for further data quality and usability improvements. We also added a new metric that describes how broadly annotators would allow others to share a given photo.

- **Information Type.** This metric measures a high-level summary of what type of information the annotated content is considered to threaten. In DIPA, the four choices of "personal identity", "location of shooting", "personal habits", and "social circle" as well as a free-form response were given. However, our pilot study found that "personal identity" and "personal habits" were confusing for some annotators. For example, they were not sure whether "personal identity" referred to any content that could be used to identify a person or only to things like ID cards. The option "personal habits" was complained about that "habit" referred to a long-term tendency that could not be observed by a single image. Annotators in our pilot study also expressed privacy concerns about bystanders appearing in given images. In addition, because privacy-threatening content probably reveals multiple types of information, allowing annotators only to choose one response might generate biases. We thus enabled annotators to choose multiple answers and adjusted the choices to "personal information", "location of shooting", "individual preferences/pastimes", "social circle", "others' private/confidential information", or use a free-form textbox in case none of the choices was appropriate.
- **Informativeness.** This metric estimates the perceived severity of privacy leakages in a 7-Likert scale. In DIPA, the annotation interface asked "How informative do you think about this privacy information for the photo owner?" with a 7-Likert scale mapped into the score ranging from -3 ("extremely uninformative") to 3 ("extremely informative"). Although the authors clarified that "uninformative" referred to a small amount of information related to privacy, it is still possible for annotators to misunderstand that "uninformative"

Table 2. Overview of the captured metrics in DIPA and DIPA2. In DIPA, annotators were forced to choose one answer only to the metrics "Information Type" and "Maximum Sharing Scope". In DIPA2, annotators were allowed to choose multiple answers to "Information Type" and "Sharing Scope" (as a photo owner and as well as by others). In DIPA2, if annotators choose "I won't share it" or "I won't allow them to share it" in the corresponding questions, they were not permitted to choose other options.

Dataset	Metric	Options	Question Description
DIPA [79]	Information Type	personal identity location of shooting personal habits social circle free-form answers	Assuming you want to seek privacy of the photo owner, what kind of information can this content tell?
	Informativeness	7-Likert scale (−3: extremely uninformative | 0: neutral | 3: extremely informative)	How informative do you think about this privacy information for the photo owner?
	Maximum Sharing Scope	I won't share it at all family or friend public broadcast programs free-form answers	Assuming you are the photo owner, to what extent would you share this content at most?
DIPA2 (**ours**)	Information Type	personal information location of shooting individual preferences/pastimes social circle others' private/confidential information free-form answers	Assuming you want to seek the privacy of the photo owner, what kind of information can this content tell (please select all that apply)?
	Informativeness	7-Likert scale (−3: strongly disagree | 0: neutral | 3: strongly agree)	How much do you agree that this content would describe or suggest the people associated with this photo (e.g., the owner of this photo or the person in the photo) in respect of what you chose in the previous question? Higher scores mean the more informative the content is.
	Sharing Scope (as a photo owner)	I won't share it close relationship regular relationship acquaintances public broadcast program free-form answers	Please assume it is a photo related to you, and answer the following questions. Who would you like to share this content to (please select all that apply)?
	Sharing Scope (by others)	I won't allow them to share it close relationship regular relationship acquaintances public broadcast program free-form answers	Please assume it is a photo related to you, and answer the following questions. Would you allow the group you selected above to repost this content (please select all that apply)?

meant there was no privacy-threatening information because no direct illustration was written. We changed the description of the question (Table 2) as well as the choices. The question was rephrased to "How much do you agree that this content would describe or suggest the people associated with this photo?" The new 7-Likert scale maps into the score ranging from -3 ("strongly disagree") to 3 ("strongly agree").

- **Sharing Scope (as a photo owner).** This question asks what recipient groups annotators would be willing to share the annotated content with if they were its owners. DIPA used four categories of "I won't share it at all", "family or friends", "public", and "broadcast programs" in their question and required annotators to choose the maximum extent of sharing. However, we identified two defects in their question designs. People may think of specific recipient groups differently. For instance, while some people are more willing to share photos with their families, others might be reluctant to do so believing families are not close relationships to them [69]. The authors of DIPA also assumed that intended sharing with a broader recipient group (e.g., public) must indicate the willingness of sharing in smaller groups (e.g., family and friends). This is not necessarily true when people hope to keep their secrets in a closer relationship and are willing to share with others [75].

To mitigate people's subjectivity to these two factors, we replaced specific recipient groups with general descriptions of interpersonal relationships. It allows annotators to choose multiple recipient groups from "I won't share it" and "close relationship", "regular relationship", "acquaintances", "public", "broadcast program" and a free-form input. The term "broadcast program" refers to a scenario where a give photo would be featured in a TV program. This represents a sharing scenario that extends beyond being "public", where the images are deliberately disseminated to a wide audience, rather than simply being accessible to the general public. Researchers can further explore the relationship between general recipients and specific recipients by analyzing personal traits and preferences in sharing context.

- **New Metric: Sharing Scope (by others).** Besides collecting sharing intentions as photo owners, other work also measured people's mental models when others want to share their photos [8, 9, 24, 68]. We then asked annotators to respond how broadly they allowed the content to be reposted by their selected recipients in the previous question. In this new question, annotators were required to choose multiple recipient groups from "I won't allow them to share it", "close relationship", "regular relationship", "acquaintances", "public", "broadcast program" and a free-form input.

In addition to these four metrics, we asked annotators to fill the following information to understand their demographics and personality traits at the beginning of annotation tasks:

- **Age**. We required each annotator to provide their specific age to analyze the influence of age on perceived privacy threats. We only allowed people above 18 to participate in our tasks.
- **Gender**. The annotators were asked to tell us their gender for data analysis. We provided three options, including "male", "female", and "not prefer to say".
- **Nationality**. We required annotators to claim their nationality in a free-form textbox.
- **Frequency of sharing photos**. To investigate the influence of sharing habits, we required annotators to tell us their frequency of sharing their own photos online through a question saying "How often do you share pictures taken by you online?" with 5 options ("Never", "Less than once a month", "Once or more per month", "Once or more per week", and "Once or more per day").
- **Big-five personality questionnaire**. We used personality traits as an indicator to distinguish individuals. To quickly identify annotators' personalities, we used a 10-question short version of the Big-five personality questionnaire proposed by Rammstedt et al. [59]. Each of the five personality factors (extraversion, agreeableness, conscientiousness, neuroticism, and openness) was determined by two questions on a scale from 1 to 10.

3.3 Data Collection

We included all 1,495 images provided in DIPA during our data collection process. These images were originally chosen from two public datasets, OpenImages [34] and LVIS [22], along with object names in their original datasets. Given their wide usage in diverse research and application areas, we assumed that releasing unobfuscated versions of these images would not introduce additional privacy breaches even though our focus lies in image privacy. Furthermore, the annotations may provide knowledge on how to discover privacy-threatening images, enabling researchers to take necessary precautions to avoid privacy violations when constructing new large-scale image datasets. Following the same approach from DIPA [79](Appendix A), we recruited crowdworkers aged 18 and above from Japan and the U.K. through two crowdsourcing platforms: CrowdWorks (a crowdsourcing platform based in Japan) [19] and Prolific [57], respectively. We employed an inclusion criterion that the nationality of participants must be either Japanese or British so that our data collection focused on these two different cultural background groups.

Figure 1 shows the interface we used in the annotation process. We translated all text into Japanese for annotators recruited in CrowdWorks [19]. Participants were first asked to provide their basic demographic

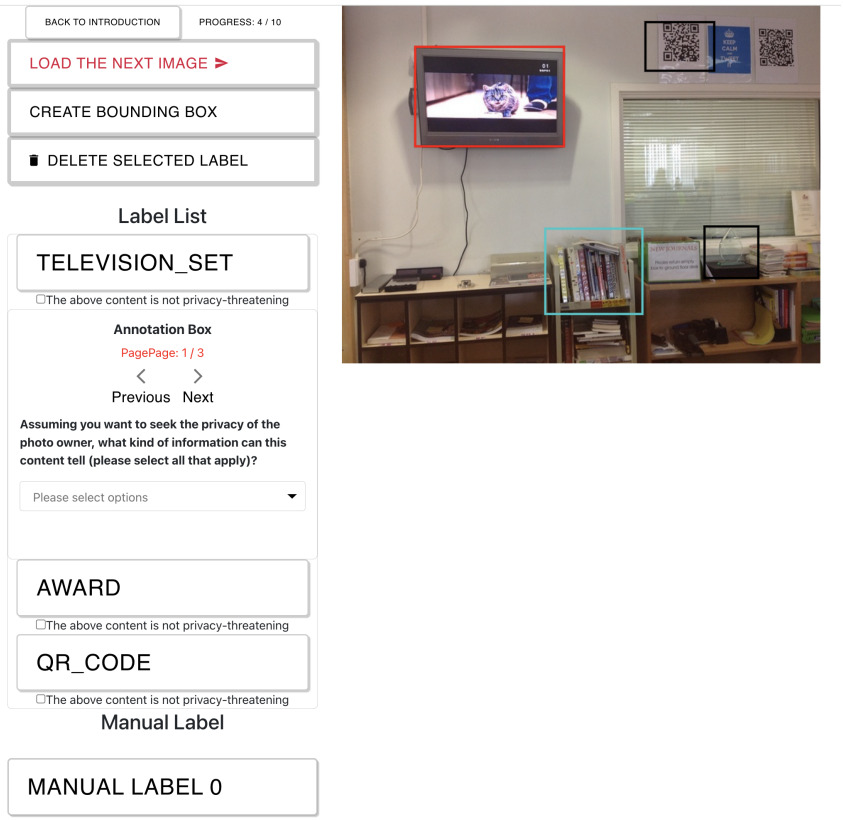

Fig. 1. The annotation interface used in the study. Annotators may select specific labels (e.g., TELEVISION_SET, AWARD, QR_CODE) provided by DIPA under the "Label List" to identify them as "privacy-threatening" and then rate them on each item listed in Table 2 (information type, informativeness, sharing scopes); these corresponding visual contents were by default surrounded by black bounding boxes (see Appendix A for details on how these specific labels and bounding boxes were generated in DIPA). The red bounding box indicates the object currently selected for annotation, corresponding to the chosen label name (here, the "TELEVISION_SET" label is selected, so the television is surrounded by a red bounding box). The light blue bounding box (here, surrounding the books) indicates new potentially privacy-threatening areas, as created by the user in this annotation session. Each manual bounding box creates a corresponding numbered label under "Manual Label" and requires the annotator to input the name of the annotated object. Annotators provide data for each of the items listed in Table 2 (information type, informativeness, sharing scopes). If annotators consider the chosen area risk-free, they may skip to the next area by ticking the checkbox below the corresponding label.

information (age, gender, and nationality), and the frequency of sharing photos, as well as finish the Big-five personality questionnaire [59]. We utilized 4,642 contents, including manual annotations, that were annotated as privacy-threatening by DIPA's annotators in 1,495 images as the default choices for annotators (Appendix A). This would reduce the burden of finding qualified content and examine if the previous annotations would hold even after additional data collection. Annotators were also able to manually highlight content if they found any privacy-threatening area that did not provide a bounding box yet.

Each image would be annotated four times, by two annotators each from Japan and the U.K., resulting in approximately twice the number of annotators as DIPA [79]. Each annotator was assigned 10 images in their task

Table 3. Demographic information of the participants.

	Age Range					All
	18–24	25–34	35–44	45–54	55–	
CrowdWorks (Japan)						
Male	6	37	50	31	18	142
Female	10	52	51	28	10	151
Not prefer to say	1	4	1	1	0	7
All	17	93	102	60	28	300
Prolific (the U.K.)						
Male	32	58	35	27	22	174
Female	11	38	29	23	20	121
Not prefer to say	1	3	0	1	0	5
All	44	99	64	51	42	300

Table 4. Self-claimed frequency of sharing their own photos online.

	Never	Less than once a month	Once or more per month	Once or more per week	Once or more per day
CrowdWorks (Japan)	53	138	65	37	7
Prolific (the U.K.)	58	137	58	38	9
All	111	275	123	75	16

(except one task contained 5 images). We only approved annotators who finished annotations on all assigned images. Because all the images prepared for annotations had previously been identified as privacy-threatening in DIPA, it was highly probable that an annotators would claim at least one image as privacy-threatening out of 10 images in their tasks. Therefore, for quality control, we rejected annotators who did not annotate any privacy-threatening visual content in assigned images. We paid them approximately 3 dollars in their local currency for successful task completion. The data collection procedure above was approved by our institutional review board.

4 DIPA2 DATASET

4.1 Basic Statistics

We followed general practices of previous dataset papers in demonstrating the overall properties of DIPA2 [22, 23, 34, 85]. In the following, we detail the annotators' demographic information [39], general annotation results [23, 85], bounding box distributions [22, 34], and cross-cultural analyses of the four privacy metrics specific to DIPA2.

4.1.1 Demographic Information. To ensure that each image has four annotators, we recruited 345 and 317 participants from CrowdWorks [19] and Prolific [57], respectively. After filtering participants who did not finish their tasks or provide any annotations, we obtained 300 valid annotators from each platform. All valid annotators from CrowdWorks and Prolific claimed that they were above 18 years of age, and their nationalities were Japanese and British, respectively. Table 3 and Table 4 details the demographic information, and estimated frequency of sharing photos of our annotators, respectively.

Fig. 2. The distribution of bounding boxes on privacy-threatening content in DIPA2. Each circle represents a bounding box, positioned relative to the center of the image to which it belongs. The diameter of each circle mirrors the bounding box's size in proportion to the image it is contained within. The coloring scheme reflects the aspect ratio of bounding boxes, which is determined by the width-to-height ratio.

4.1.2 General Annotation Results. We collected 18,950 annotations in total. Among them, 13,053 annotations reveal annotators' perspectives on not treating visual contents as threats to privacy. This led to 5,897 annotations for 3,347 visual contents that were perceived as potential privacy risks in 1,304 images. Specifically, participants from CrowdWorks (Japan) and Prolific (the U.K.) identified 1,159 and 912 images that they claim contain at least one privacy-threatening content, respectively.

Among the 3,347 visual content, 1,507 of them were annotated as privacy-threatening once; 931 of them were annotated twice, 499 of them were annotated three times; 202 of them were believed to threaten privacy by all assigned annotators. The remaining 208 unique contents were identified from 222 manual bounding boxes by filtering them in an overlap threshold of 50%. Seven manually identified visual contents were annotated twice by our annotators.

We categorized all visual contents according to 22 privacy-threatening categories perceived in the image preparation process of DIPA [79] (see Appendix A for details on how these categories were derived). Table 5 details the distribution of all visual content, including those annotated as not privacy-threatening with our results. Our binomial test confirmed that the likelihood of annotating privacy-threatening content by British annotators was significantly lower than those by Japanese annotators (p<.001, 95%CI: [0.36, 0.39]), aligning with findings in previous research on the cultural influences on perceiving privacy risks in image [6, 12, 17, 18, 41, 43, 61, 71, 74, 79]. In 13 out of 22 major privacy-threatening categories and other categories, Japanese annotators tended to annotate more privacy-threatening content in given images. British annotators were only more sensitive when observing a specific category of "table".

4.1.3 Distribution of Bounding Boxes on Privacy-threatening Content. We present the distribution of all bounding boxes annotating privacy-threatening content in DIPA2 for researchers aiming to develop object-level machine-learning models. This distribution is analyzed according to three specific relative parameters, as follows.

Table 5. The distribution of the 22 privacy-threatening content categories in DIPA2 (Appendix A). The statistics present how many times the visual contents belonging to the categories were annotated as "privacy-threatening". For instance, we can observe that Japanese annotators identified 39 visual contents under the category "pet" to be privacy-threatening (32 by one annotator and seven by both annotators). Meanwhile, 34 visual contents of "pet" category were not perceived as privacy-threatening content by any annotator. The column "Both Platforms" represents the distribution of annotations regardless of the platform, thus ranging from 0 times to 4 times. The RP (rate of privacy) represents the percentage of content reconfirmed as privacy-threatening when compared to DIPA. The bold font represents statistically significant results confirmed by our binomial tests ($p < .05$), which means that annotators from Japan and the U.K. made different decisions on deciding a specific category of content as privacy-threatening. Contents that could not be classified into the 22 categories are referred to as "other categories" in the table. Manual annotations are included in "other categories".

Category	**Japanese Annotators**				**British Annotators**				**Both Platforms**					
	0	1X	2X	RP	0	1X	2X	RP	0	1X	2X	3X	4X	RP
person	420	396	405	66%	937	407	102	42%	288	379	300	178	76	76%
place identifier	25	41	17	70%	57	21	5	31%	20	29	27	5	2	76%
identity	0	9	3	100%	5	7	0	58%	0	3	8	1	0	100%
vehicle plate	16	40	43	84%	11	45	43	89%	1	13	31	34	20	99%
food	29	8	1	24%	31	5	2	18%	29	2	5	1	1	24%
printed materials	28	22	10	53%	30	20	10	50%	19	16	12	10	3	68%
screen	49	48	36	63%	46	50	37	65%	24	31	35	29	14	82%
clothing	103	70	29	49%	146	52	4	28%	77	79	31	13	2	62%
scenery	28	26	9	56%	51	10	2	19%	26	22	16	0	1	62%
pet	34	32	7	53%	58	11	4	21%	26	35	8	2	2	64%
book	35	32	7	66%	52	17	5	30%	21	35	16	2	0	72%
photo	7	8	1	56%	11	5	0	31%	6	6	3	1	0	63%
machine	21	10	2	36%	28	4	1	15%	18	10	5	0	0	45%
table	59	12	2	10%	55	17	1	25%	46	22	2	3	0	37%
electronic devices	27	4	2	18%	29	4	0	12%	23	8	2	0	0	30%
cosmetics	14	5	1	30%	18	2	0	10%	12	7	1	0	0	40%
toy	17	10	1	39%	23	4	1	18%	14	11	2	1	0	50%
finger	20	11	0	35%	27	4	0	13%	17	13	1	0	0	45%
cigarettes	21	8	1	30%	19	11	0	37%	13	14	2	1	0	57%
musical instrument	22	17	3	48%	33	8	1	21%	18	18	5	1	0	57%
accessory	41	30	9	49%	53	25	2	34%	26	35	15	4	0	68%
sum of the above	1,050	853	593	58%	1,535	739	222	39%	753	797	535	289	122	70%
other categories	1,094	883	390	54%	1,524	681	164	36%	862	916	404	210	80	65%
sum of the all	2,144	1,736	983	56%	3,059	1,420	386	37%	1,615	1,713	939	499	202	67%

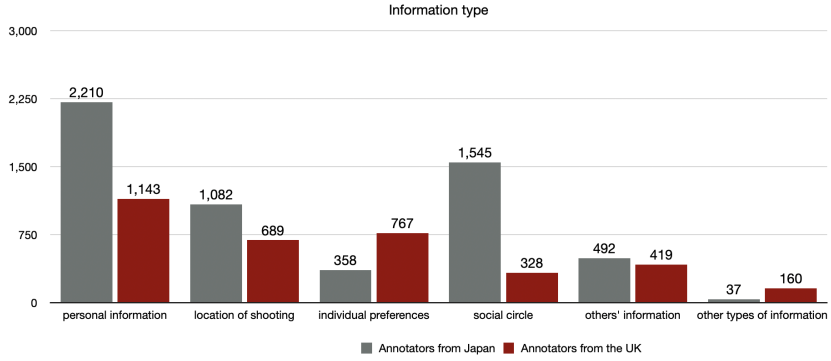

Fig. 3. The distribution of information type of privacy-threatening content in the dataset. On average, 1.55 and 1.60 answers were provided by annotators from CrowdWorks (Japanese) and Prolific (British), respectively.

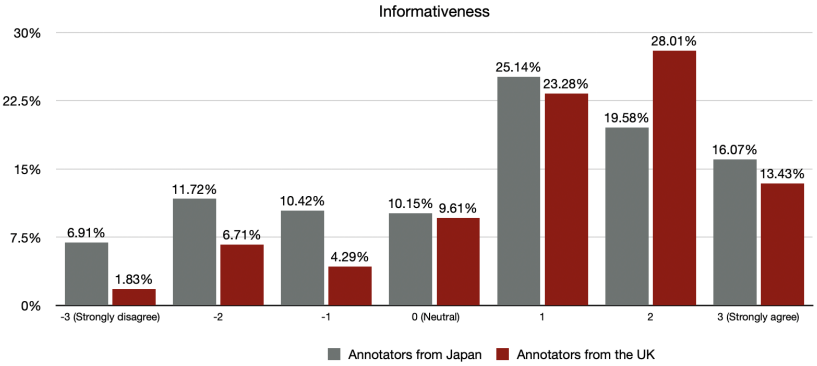

Fig. 4. The distribution of informativeness scores provided by annotators in both crowdsourcing platforms.

Relative Distances to the Center: The distances in this analysis are relative to the length of the diagonal of their images. The median distance for the closest 30% of the points is 0.11. The middle 40% and the farthest 30% of points have median distances of 0.28 and 0.45, respectively.

Relative Size to the Image Size: The sizes of the bounding boxes are expressed as relative values to their image sizes and vary significantly. The median size of the smallest 30% of the boxes is 0.0006. For the middle 40% and the largest 30%, the median sizes are 0.0125 and 0.2522, respectively.

Aspect Ratios: The aspect ratios, denoting the relationship between the width and height of the bounding boxes, are also varied. The median ratio for the smallest 30%, the middle 40%, and the largest 30% of boxes are 0.40, 0.85, and 1.70, respectively.

Figure 2 visualizes the distribution of all bounding boxes within DIPA2, organized according to the aforementioned relative parameter spaces.

4.1.4 Distribution of Privacy Metrics. For the four privacy metrics specified in DIPA2, we present cross-cultural analyses for highlighting the meaning of factoring cultural influences while developing image privacy datasets. Figure 3 shows the distributions of the information types in our annotations. On average, 1.55 and 1.60 responses were provided by CrowdWorks (Japanese) annotators and Prolific (British) annotators in each privacy-threatening content, respectively. While the category of personal information was the most frequent, annotators derived

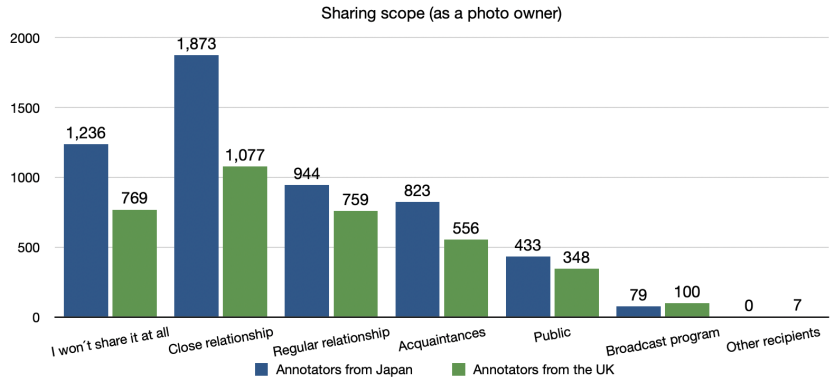

Fig. 5. The distributions of sharing scopes (as a photo owner) provided by the annotators.

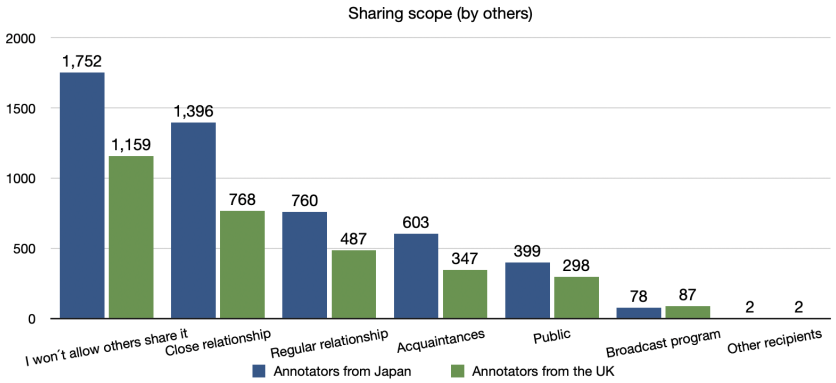

Fig. 6. The distributions of sharing scopes (by others) provided by the annotators.

multiple other types of information in different visual content. Our Chi-square tests for each information type showed significant differences except for "location of shooting" against annotators from two countries (personal information: $\chi^2(1)=33.67$, $p<.001$, location of shooting: $\chi^2(1)=.74$, $p=.389$, individual preferences/pastimes: $\chi^2(1)=474.43$, $p<.001$, social circle: $\chi^2(1)=417.02$, $p<.001$, others' private/confidential information: $\chi^2(1)=21.14$, $p<.001$, others: $\chi^2(1)=189.21$, $p<.001$). Specifically, Japanese annotators annotated more visual contents referring "personal information" and "social circle", while British annotators relatively discovered more contents related to "individual preferences/pastimes" and "others' private/confidential information" in their annotations. These results showed that annotators from Japan focused more on visual content indicating personal information and social circles while annotators from the U.K. paid more attention to those telling individual preferences or bystanders' privacy.

Table 6 summarizes the information types identified across all the object-level annotations. This was a multiple-choice item, i.e. we can observe which information types were chosen in conjunction with other types. Personal information was overwhelmingly more popular than other types, chosen most often together with the location of the shooting (N=670). Future work may deeply investigate interrelationships between different types of information derived from visual content.

Table 6. Co-occurrences of "information type" labels, i.e. how many times was a given type chosen simultaneously with another information type.

	PI	Loc	IP	SC	OPI	Others
Personal information (PI)	3,353	670	540	828	522	35
Location of shooting (Loc)	670	1,771	396	567	273	22
Individual preferences/pastimes (IP)	540	396	1,125	403	267	21
Social circle (SC)	828	567	403	1,873	312	12
Others' private/confidential information (OPI)	522	273	267	312	911	14
Others	35	22	21	12	14	197

Table 7. Co-occurrences of "sharing scope (as a photo owner)" labels. Each cell shows the number of co-occurrences of the labels corresponding to each row and column. (IWS: I won't share it, CR: close relationship, RR: regular relationship, AC: acquaintances, PU: public, BP: broadcast program, Others: other recipients). Annotators were allowed to choose multiple answers except for "I won't share it".

	IWS	CR	RR	AC	PU	BP	Others
I won't share it (IWS)	2,005	0	0	0	0	0	0
Close relationship (CR)	0	2,950	1,367	1,097	424	130	2
Regular relationship (RR)	0	1,367	1,703	1,026	424	131	2
Acquaintances (AC)	0	1,097	1,026	1,379	426	132	2
Public (PU)	0	424	424	426	781	147	0
Broadcast program (BP)	0	130	131	132	147	179	0
Others	0	2	2	2	0	0	7

We next examined the "informativeness" responses (Figure 4). The means of rating results by CrowdWorks (Japan) annotators and Prolific (British) annotators were .58 and 1.11, respectively. We ran a Mann-Whitney U test to evaluate the difference in the responses by Japanese annotators and annotators from the U.K and found a significant effect of nationality ($U = 342.7$, $Z = -10.25$, $p < .001$, $r = .13$). While annotators from Japan marked more privacy-threatening content, British annotators tended to perceive higher threats in contents they selected, which may indicate that they were likely to ignore mild privacy threats.

Figure 5 displays the distribution of sharing intentions when annotators assume that they were photo owners. Most of the annotators were willing to share the given contents with people they were familiar with. Figure 6 summarizes the expected ranges where our annotators were willing to allow their recipients to repost the given contents. Compared with sharing by themselves, our participants were prone to disallow others to repost their images or only permit their recipients to share in a small scope. In both sharing scenarios, the option "broadcast program", indicating an exceptionally broad sharing scope, was infrequently selected.

We also analyze the co-occurrence of sharing scopes (i.e., how many times each of the scopes was chosen simultaneously with other sharing scopes in the multi-option question item). Tables 7 and 8 present the co-occurrence of sharing scopes, as either a photo owner or when shared by others. This allows us to observe, for instance, how the sharing scope of "broadcast program" was a relatively unpopular choice (in Table 7) in conjunction with any other choices. We can observe the overall popularity of all the scopes on the diagonal line,

Table 8. Co-occurrences of "sharing scope (by others)" labels. Each cell shows the number of co-occurrences of the labels corresponding to each row and column. (IWA: I won't allow them to share it, CR: close relationship, RR: regular relationship, AC: acquaintances, PU: public, BP: broadcast program, Others: other recipients). Annotators were allowed to choose multiple answers except for "I won't allow them to share it".

	IWA	CR	RR	AC	PU	BP	Others
I won't allow them to share it (IWA)	2,911	0	0	0	0	0	0
Close relationship (CR)	0	2,164	955	735	366	116	2
Regular relationship (RR)	0	955	1,247	697	377	113	2
Acquaintances (AC)	0	735	697	950	376	115	2
Public (PU)	0	366	377	376	697	131	2
Broadcast program (BP)	0	116	113	115	131	165	2
Others	0	2	2	2	2	2	4

Table 9. Results from the logistic regression model. "age", "frequency of sharing photos" and Big-five personality test results (extraversion, agreeableness, conscientiousness, neuroticism, and openness) are calculated as continuous variables. We mapped answers to "frequency of sharing photos" from 0 to 4 (0: "Never", 1: "Less than once a month", 2: "Once or more per month", 3: "Once or more per week", and 4: "Once or more per day") and use original value of the other two continuous variables ("age" and Big-five personality test results). "gender" and "nationality" were set as categorical variables whose baseline values are "Female" and "the U.K.", respectively. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.

	Estimated coefficients	**Odds Ratios**	**95%CI for coefficients**
age	-0.009^{***}	0.991	$[-0.011, -0.007]$
gender (male)	-0.184^{***}	0.832	$[-0.237, -0.132]$
gender (not prefer to say)	-0.601^{***}	0.548	$[-0.792, -.0410]$
nationality (Japanese)	0.733^{***}	2.081	$[0.678, 0.788]$
frequency of sharing photos	-0.074^{***}	0.928	$[-0.101, -0.048]$
Big-five (extraversion)	-0.019^{*}	0.981	$[-0.034, -0.005]$
Big-five (agreeableness)	0.012	1.012	$[-0.004, 0.029]$
Big-five (conscientiousness)	0.003	1.003	$[-0.012, 0.019]$
Big-five (neuroticism)	0.060^{***}	1.062	$[0.045, 0.074]$
Big-five (openness)	0.059^{***}	1.061	$[0.045, 0.073]$
Akaike Information Criterion (AIC)		32653	

from the upper left table corner to the lower right corner: While 2,005 annotations indicated an image as not suitable to be shared with anyone, 2,950 annotations indicated it could be shared with those in close relationship. The scopes descend in popularity as a function of the openness of the sharing scope, rather intuitively, all the way until the scope "broadcast program" that was selected in just 179 annotations.

Table 10. Distribution of annotations in both our study (DIPA2) and the previous dataset (DIPA [79]. The ranges in the columns represent the number of annotations (e.g., the 6–10 range includes annotators who provided between 6 and 10 annotations).

	Number of Annotators						**All**
	0–5	6–10	10-15	15-20	20–30	30–	
DIPA [79]	122	80	36	30	47	45	360
DIPA2 (**ours**)	218	151	87	65	61	18	600

4.2 Regression Analysis

We conducted statistical analysis to demonstrate example quantitative examinations enabled by DIPA2. We first ran a logistic regression model to predict if visual content can be regarded as privacy-threatening (1 = privacy-threatening, 0 = not privacy-threatening) with the influence of demographic information and personality trait. This analysis illustrates how researchers can utilize DIPA2 to explore the impact of human factors on privacy preferences, like related research introduced in Section 2.1.3.

We used 5,897 annotations of privacy-threatening content and 13,053 annotations of not privacy-threatening content in the regression process to build this model. Our independent variables include basic information about annotators (age, gender, nationality, frequency of sharing photos, and Big-five personality test results). We treated "age", "frequency of sharing photos", and each Big-five factor as continuous variables and regarded "gender" (Male, Female, and others) and "nationality" (Japanese and British) as categorical variables. We weighted the data according to the proportion of "privacy-threatening" content and "not privacy-threatening" content to avoid biases toward the majority class.

Table 9 shows the coefficients, odds ratio, and 95% confidence interval (95%CI) of each independent variable resulting from the logistic regression model. We observe that all independent variables except "agreeableness" and "conscientiousness" were statistically significant predictors. "Age" is a negative predictor in the logistic regression model, which indicates that older people tended to find less privacy-threatening content in our studies. When annotators were male or unwilling to expose their gender, they were inclined to mark down less privacy-threatening content. The difference in nationality created the most significant effect on the odds of annotating content as privacy-threatening. When annotators were from Japan, it increases the odds by approximately 108%, which aligns with our results in the binomial tests. Their sharing habits also had a significantly negative effect on the degree of noticing privacy. Our annotators tended to discover less privacy-threatening content if they shared their photos online more frequently.

We also observe that three Big-five factors were significantly related to the annotation of privacy-threatening content. Annotators who were more extroverted showed a proclivity of annotating less privacy-threatening content in our studies. On the contrary, higher scores in "neuroticism" and "openness" resulted in more sensitive observations of privacy-threatening content. These findings offer additional proof that emphasizes the importance of collecting individual differences in image privacy datasets. We expect data scientists to further investigate the in-depth association between these factors with object-level privacy annotations to advance the quantitative research of image privacy such as explainable privacy assistant [10].

4.3 Comparison with DIPA

Most of the images in DIPA [79] could be re-identified as containing privacy-threatening content with a wider range of annotators. Given 1,495 images in DIPA, 1,304 of them were re-identified as privacy-threatening in

our studies. We verified that 72% of the content provided by DIPA could be regarded as privacy-threatening. Specifically, DIPA2 identified 3,347 unique privacy-threatening content while 4,642 privacy-threatening content with 5,671 annotations were identified in DIPA by 360 annotators. Meanwhile, 2,838 out of 3,347 visual content overlapped with annotations in DIPA. The binomial test to compare the probability of annotations by our participants for privacy-threatening contents against those by DIPA showed a significant difference (p<.001, 95%CI: [0.66, 0.69]). This reveals the importance of collecting annotations from multiple annotators on the same images when establishing an image privacy dataset.

On average, each annotator in our study provided 9.82 annotations on content with potential privacy issues, while annotators in DIPA [79] contributed around 15.75 annotations each. The details regarding the number of annotations provided by individual annotators in both DIPA2 and DIPA are listed in Table 10. We noticed a decrease in the proportion of annotators contributing a large number of annotations (e.g., more than 30), leading to a reduction in the overall average number of annotations.

We observed alignments and differences in annotators' responses to privacy metrics. Comparing answers in "information type" with results in DIPA [79], we found that annotators expressed more concerns about privacy related to "social circle" and "other's information", which indicated that allowing multiple choices enables more comprehensive observations. The means of "informativeness" in our studies and DIPA were .77 (SD = 1.73) and .84 (SD = 1.39), respectively. The Mann-Whitney U test on "informativeness" did not show a significant result ($U = 1.67 * 10^8$, $Z = -2.11$, $p = .41$, $r < .001$). As we considerably changed the metrics collecting intended sharing scopes to given visual content, there is no direct comparison we can conduct with DIPA.

5 AUTOMATIC IMAGE PRIVACY ASSESSMENT BENCHMARKING

In the previous section, we report various quantitative properties of DIPA2 and presented an example analysis based on the dataset. In this section, we describe further explorations enabled by DIPA2. We prototype two deep-learning models that predict 1) specific privacy-threatening content, and 2) the classes and scores of four privacy metrics (information type, informativeness, sharing scope (as a photo owner), and sharing scope (by others)) for each privacy-threatening content, in images of DIPA2, respectively.

5.1 Privacy-threatening Content Prediction

We performed classification experiments on privacy-threatening content within DIPA2 by fine-tuning the ResNet-50 model [26], as done for previous image privacy datasets [23, 54]. Moreover, we pretrained this model using the VISPR [54] and VizWiz-Priv [23] datasets to investigate if these existing image privacy datasets could enhance the identification of privacy-threatening content in the DIPA2 dataset.

5.1.1 Dataset. We distributed DIPA2 into a random 65-10-25 split, resulting in 3,833, 590, and 1,474 images allocated to the training, validation, and test sets, respectively. For pretraining with the VISPR and VizWiz-Priv datasets, we leveraged their predefined training sets. To unify the evaluation process, we carefully mapped the categories of both VISPR and VizWiz-Priv datasets to those within DIPA2 (See Appendix B for details).

5.1.2 Methods. The VISPR [54] dataset only provides image-level labels. Also, neither of the other two datasets offers any potential input data beyond RGB images. Therefore, we limited our model implementation and performance evaluation to image-level predictions on DIPA2 under various pretraining scenarios. Our input consisted of only RGB images, and the output was characterized as a 23-dimensional vector. This vector represented the presence of the 22 privacy-threatening content categories identified in DIPA2, along with an additional "Others" category.

For comparative analysis, we devised seven distinct models. Three models were pretrained on VISPR, VizWiz-Priv, and a combination of both datasets, respectively. Subsequent to the pretraining phase, these models were

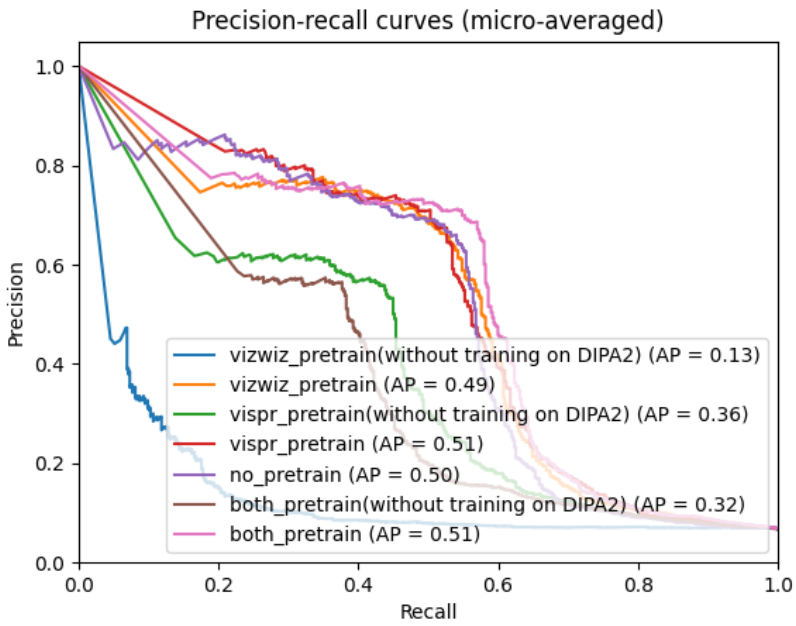

Fig. 7. Precision-recall curves and average precision (AP) scores for detecting privacy-threatening content on DIPA2 across various pretraining scenarios.

then further trained using DIPA2. An additional three models were also pretrained following the same procedure, but they were not subsequently trained with DIPA2. This approach allowed us to evaluate the innate performance of each pretrained model and demonstrate the value of DIPA2 as a novel resource in this field.

Finally, the last model acted as a baseline, being trained exclusively on the DIPA2 without any pretraining.

5.1.3 Results. In accordance with the benchmark established in the VizWiz-Priv [23] dataset, we reported micro-averaged precision-recall curves and average precision (AP) for each variation of the ResNet-50 model, accounting for different pretraining strategies (Figure 7). We finetuned all models for 100 epochs to reach their convergences. The proportion of positive samples in the entire dataset was approximately 0.196, 0.204, and 0.197 for the training, validation, and test sets, respectively (e.g., if an image contains content from two privacy-threatening categories, it will be recorded as having two positive samples, and the remaining 21 categories will each be counted as a negative sample).

Our baseline model, which was trained exclusively on DIPA2 without pretraining, achieved an AP score of 0.50. With the introduction of pretraining, the ResNet-50 model achieved similar AP scores for all versions subsequently trained with DIPA2. These results confirm the challenging nature of predicting privacy-threatening content solely based on RGB image input and common models used for the prediction of image privacy within DIPA2. Moreover, models pretrained but not further trained with DIPA2 exhibited much lower performance. This suggests that DIPA2 successfully enhances the variety and depth of privacy representations, as the presence of a specific content category doesn't necessarily equate to a positive annotation but rather depends on various factors such as cultural backgrounds and individual personality traits. Additionally, the model pretrained on VizWiz-Priv [23] achieved an AP score (0.13) even lower than the proportion of positive samples in the test set

Table 11. Results of applying Resnet-50 to our dataset (validation set). Only annotations referring to privacy-threatening content were selected.

Metric	Accuracy	Precision	Recall	Mean Difference
information type	0.75	0.62	0.59	−−
informativeness	−−	−−	−−	1.24
sharing scope (as a photo owner)	0.72	0.61	0.55	−−
sharing scope (by others)	0.75	0.61	0.56	−−

(0.197), indicating that the characteristics and existences of privacy-threatening content vary distinctly between images taken by normal individuals and PVIs. This experiment emphasizes the distinct aspects of DIPA2 and its capacity to facilitate the study of image privacy recognition in ubiquitous scenarios.

5.2 Privacy Metrics Prediction

In this benchmarking experiment, we also leverage Resnet-50 architecture [26] as in Section 5.1 to predict the four privacy metrics that exist in DIPA2 (information type, informativeness, sharing scope (as a photo owner), and sharing scope (by others)) by giving the information of content and annotators' personality traits. This experiment envisions a scenario of anticipating people's perceptions of privacy protection and sharing intentions after a detector, similar to those discussed in Section 5.1, recognizes potential threats in images.

5.2.1 Dataset. Aligning with the division in Section 5.1, we randomly divided 5,897 annotations that recorded privacy-threatening content in DIPA2 into a 65-10-25 split, resulting in 3,833, 590, and 1,474 images in the training, validation, and test sets.

5.2.2 Methods. For each annotation on privacy-threatening content, we fed the model 1) RGB image (resolution: 224 pixels × 224 pixels), 2) the category name with bounding boxes to locate the visual content, and 3) the annotator's information (age, gender, nationality, frequency of sharing photos, and Big-five personality test results). Therefore, the input is composed of 13 channels in total (3 channels for RGB image, 1 channel for the category name, and 9 channels for the annotator's information). The output was a vector that recorded predictions of four privacy metrics (information type, informativeness, sharing scope (as a photo owner), and sharing scope (by others)).

Figure 8 demonstrates the architecture of our modified model. We made minor modifications to the ResNet-50 model [26], tailoring it to accommodate the specific structure of our input and output data. We converted each channel of the input into a 2-D matrix (resolution: 224 pixels × 224 pixels). Except for the RGB image, each 2-D matrix copied the value of the corresponding variable to the position surrounded by given bounding boxes and set the rest of the matrix values to be 0. We then modified Resnet-50 to input the concatenated matrix with 13 available channels (shape: 224 pixels × 224 pixels × 13 channels) at the first layers. At the end of Restnet-50, we added two fully connected layers to obtain the output that matched the number of total responses of the four privacy metrics.

5.2.3 Results. We finetuned the modified Resnet-50 model on our data for 100 epochs to reach convergence. Table 11 summarizes the accuracy, precision, and recall of our model on each metric except "informativeness". We counted the accuracy, precision, and recall by the weighted average method in multi-label prediction. For the metric "informativeness", we calculated the mean difference between the predicted output and ground-truth results according to the 7-Likert scale (Table 2).

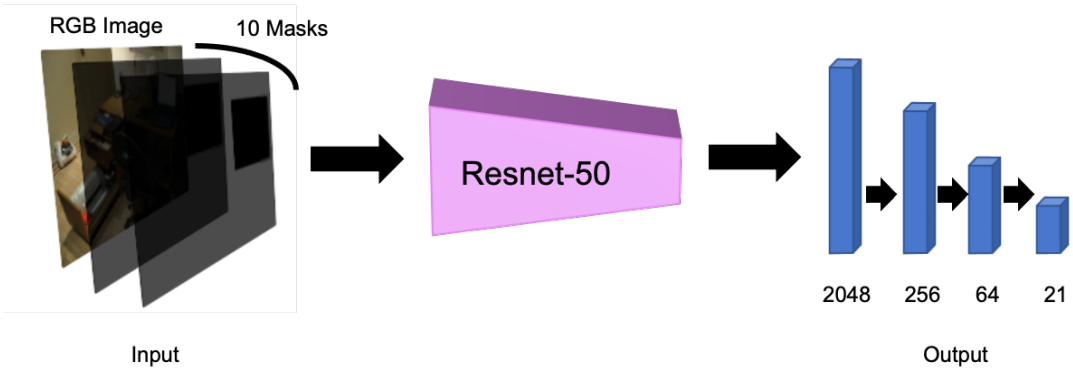

Fig. 8. Simplified network architecture of our baseline model. The input contains 13 channels. Each channel except the RGB data copied the value of the corresponding variable to areas surrounded by given bounding boxes (darker places in the black masks), resulting in 10 masks along with the RGB data. The output of Resnet-50 is a 2048-D vector and would be downsampled by two fully connected layers. The final output is a 21-D vector containing the prediction of each option in the four privacy metrics (index reference: 0-5 -> "information type", 6 -> "informativeness", 7-13 -> "sharing scope (as a photo owner)", 14-20 -> "sharing scope (by others)").

The overall results show that it is plausible to predict privacy metrics in DIPA2 by simply inputting all variables into the first layer of the model. The proposed model reached approximately 70% in accuracy and 60% in precision and recall on our privacy metrics (except "informativeness"). Our model was also able to predict similar severity of privacy threats with perceived risks by our annotators.

We note that further improvements in these tasks are beyond the scope of this work. However, the results above are an encouraging outcome for researchers to examine how future machine learning approaches could lead to more successful performance. For instance, our second benchmark study (Section 5.2) suggests that enhanced performance in the future could potentially allow ubiquitous applications to infer users' intended sharing scopes by analyzing their profiles. This capability could greatly benefit various ubiquitous computing contexts, such as wearable camera photography [7], where users often lack the time to manually adjust sharing settings. As the major contribution of this paper is to present our dataset, we, therefore, leave these explorations in future work.

6 DISCUSSION

6.1 Cross-cultural Influences on Image Privacy

Cultural influences are pivotal topics in ubiquitous computing and sensing technologies [31, 60]. Our study also highlights that cultural background is important in assessing which specific objects are considered privacy-threatening in various contexts of online image sharing. Thus, collecting annotations with an emphasis on cultural backgrounds in image privacy datasets contributes to the advance of image privacy protection and implementations of ubiquitous applications in specific contexts. For instance, consider a simple application that aims to warn its users about sharing images when they contain potentially privacy-sensitive content. Such a tool will be exactly as useful as the data it has been trained with: The warnings should be different across different cultures.

Image datasets with privacy annotations from multiple cultures also present an interesting opportunity. If we train a privacy detector with data collected from people in different cultures, the tool could be used to make its users discover new types of privacy-threatening objects in the images that otherwise would be difficult to identify with their own cultural background. In other words, people could learn about what is considered sensitive in

other cultures. For instance, as shown in Figure 3, annotators from Japan relatively paid less attention to exposing their individual preferences. Through a model understanding how people in the U.K. remind themselves of the existence of related visual content, Japanese people might be able to enhance privacy protection for this type of data. Obviously, further challenges remain here as the intended sharing scope also plays a role. Future tools should also be aware of the audiences that are consuming the images. Such audience data is readily available in most social media applications, and combining such information could further improve image privacy detection.

6.2 Comparison between DIPA and DIPA2

The results of our data collection confirmed overall results in DIPA [79]. In other words, most of the images identified as privacy-threatening in DIPA were also perceived as threatening in DIPA2. Further, in DIPA2, we collected richer observations with a multiple-choice option on what information can be derived from different categories of privacy-threatening content compared with the single-choice design in DIPA. Although our dataset maintained a similar distribution as DIPA in the metric of "informativeness", 10% annotations were chosen to the default answer "neutral" in DIPA2 compared with 17% annotations to choose "neutral" in DIPA, which suggested that annotators acquired a better understanding of our new illustration and corresponding choices. We also extend DIPA by collecting image-sharing intentions with multiple-choice items. To achieve this, the annotators provided an average of 1.53 choices per annotation when assuming they were photo owners, and 1.38 choices when allowing others to repost the content. Although annotators normally included smaller ranges (e.g., close relationships) of recipients when they were willing to broadly share the given contents (e.g., public), the dataset enables various other types of analyses as well. For instance, we found 1,074 and 897 annotations indicating annotators would like to block closer relationships and share with broad ranges of people in "sharing scope (as a photo owner)" and "sharing scope (by others)", respectively. This supports researchers in establishing fine-grained analyses of people's sharing preferences by combining visual information and estimated sharing intentions (i.e. "information type" and "informativeness"). For example, mobile participatory sensing systems for ubiquitous computing applications that utilize photos shared by users may integrate models trained by DIPA2 to quickly manage proper sharing scopes to mitigate potential privacy violations.

One major difference of DIPA2 with other image privacy datasets [23, 53, 54, 63, 65, 83, 85] is its emphasis on different opinions on annotating the same image. While presenting more annotations per image advances the understanding of privacy in the realistic world, it results in instabilities when examining existing annotations in DIPA. We observed an approximately 40% decrease in perceived privacy-threatening objects per annotator compared with results in DIPA though we assigned each annotator the same number of images (10 images) in our task as in [79]. The reason for the decrease, we believe, is that annotators were shown more bounding boxes surrounding visual contents that were not privacy-threatening in the study by Xu et al (Appendix A). In comparison to other content that did not indicate any privacy information, annotators in DIPA tended to choose content that is relatively privacy-threatening due to the demand characteristics of finishing their tasks [47]. Especially, in 13 of the 22 privacy-threatening categories summarized by Xu et al., we found that the odds of being identified as privacy-threatening were lower than the odds of finding such content in other categories that were not specified. Our annotators believe more than half of the visual content related to "table", "food", "machine", "electronic device", "fingers", "toy", and "cosmetics" were not privacy-threatening, though these visual contents were identified in DIPA (Table 5). This indicates that perceptions of privacy threats in these categories largely depend on specific scenarios and individual concerns. Researchers may study how to use personalized recommendation models to realize precise recognition of privacy in these categories of visual content. Researchers may also leverage large-scale models [52] to understand complicated mental models behind visual information.

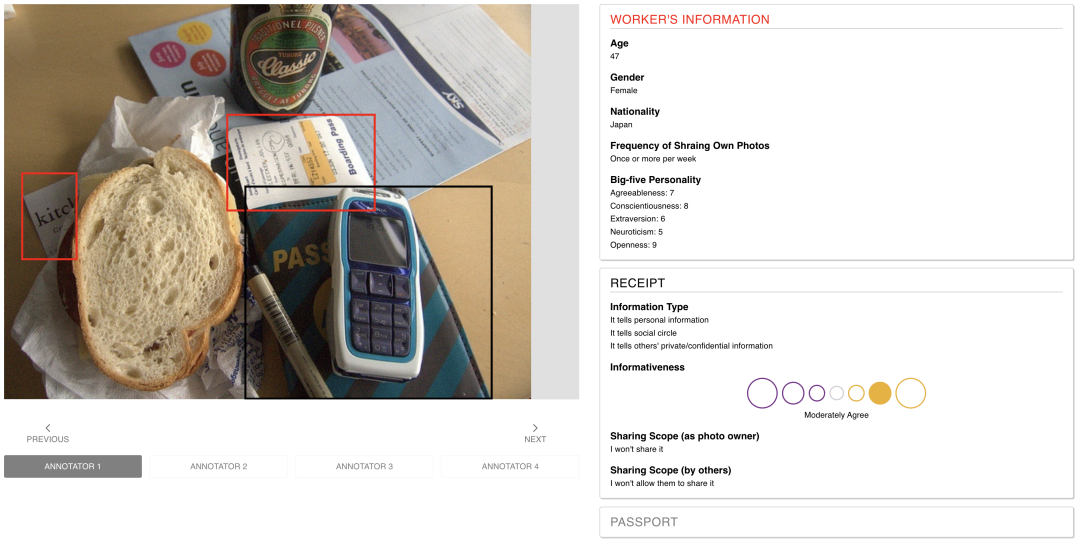

Fig. 9. The online interface for DIPA2. Users can view the demographic information and detailed annotations of the privacy-threatening contents.

6.3 Potential Research and Application Directions in DIPA2

We argue that DIPA2 provides various research directions for machine learning scientists, ubiquitous computing researchers, HCI researchers, and social scientists. Machine learning scientists may develop models to anticipate potential privacy threats in images like using other image privacy datasets [23, 53, 54, 65, 83]. DIPA2 may further improve the explainability of developed models, enabling models to understand what types of privacy-threatening content are contained in specific visual contents and which groups of people will be more sensitive to them. Moreover, privacy detection models can be enhanced to understand how users would like to manage the behaviors of sharing their images by learning the data of intended sharing scopes in DIPA2. Current ubiquitous technologies will be able to offer a privacy-conscious environment by integrating such models to guide users to understand privacy risks and proper sharing behaviors in various scenarios like IoT-based photography.

HCI researchers may further investigate the relationship between privacy and sharing intentions by introducing more assumption scenarios of sharing and conducting comparative experiments based on existing data in DIPA2. They may also extend DIPA2 to other visual media types, such as videography to cover more needs in ubiquitous computing. Researchers who have social science backgrounds may augment the data with more culture-specific properties, such as questions deriving why people regard visual content as privacy-threatening according to their life experiences or social backgrounds.

We also provide an online interface to quickly browse images and all properties in DIPA2 (Figure 9)[2]. The interface presents basic information about each annotator (age, gender, nationality, frequency of sharing photos, and Big-five personality test results). For each privacy-threatening content, it provides detailed metrics measured by each annotator (information type, informativeness, sharing scope (as a photo owner), and sharing scope (by others)). This interface will help both technical users and non-technical users to quickly familiarize DIPA2 and explore the usage of our dataset. In addition, the interface can be easily modified as a tool for educating knowledge of privacy for general user populations. People can suffice their knowledge and reflect on their privacy

[2]The visualization interface can be accessed at https://anranxu.github.io/DIPA2_VIS/visualization

protection practices by reviewing others' perspectives on privacy protection. We hope that our dataset can foster various research and applications around image privacy.

7 LIMITATIONS

There are several limitations in the current DIPA2 to be clarified. Researchers in DIPA [79] identified 22 categories of privacy-threatening content as the baseline to decide if an image should be included in the annotation process. Our study found that more than half of these categories were less likely to contain privacy threats than other categories not listed explicitly. Adjustments of the standard of qualified images to be annotated may facilitate us to derive more annotations and knowledge of privacy.

None of the authors are native English speakers, which might have influenced the quality of annotations due to improper English usage in our question design. While we emphasized cultural differences when building DIPA2, we only recruited annotators from Japan and the U.K. to represent typical Eastern and Western cultures, respectively. It is insufficient to assert that cultural differences will significantly influence the choices when comparing results from annotators in other countries. Future work should expand the dataset with more annotations from different countries to increase the diversity of DIPA2. This work has limitations in utilizing the physical locations to be the cultural self-identification of participants, which is not always true considering the enhanced mobility of modern citizens. As a preliminary work that discovers the cultural consideration in photographic privacy, we believe the results in the paper can provide early-stage insights for the community. In the future, we encourage researchers to address this limitation by conducting a worldwide study and presenting more inclusive and complete results.

Our annotators' demographic information was not distributed based on census data of Japan or the U.K. Although Prolific [57] supported recruiting British participants according to census data, we did not choose to use this function for aligning our recruitment in CrowdWorks [19], where such a function does not exist. As a result, we recruited more female annotators in Japan, and there were more males in our recruitment in the U.K. Some populations may not readily participate in our studies, leading to biases in our data. For instance, senior users may be less likely to join because they are less likely to use crowdsourcing services. Future data collection in our dataset can emphasize the privacy concerns of specific groups to bridge this gap.

Although the four privacy metrics in our dataset supply extra information when inspecting the corresponding visual content, we did not record the mental models of annotators during the data collection process. Future work may directly record the reasoning processes of annotators, and finetune advanced natural language models for reasoning privacy-threatening content and sharing intentions.

The images in DIPA2 were derived from DIPA [79], whose images were filtered from existing datasets to prevent privacy violations. However, since these images are not personal photographs of individuals, the context in which they were taken might be ambiguous to annotators. For instance, a picture depicting a room scenario might be interpreted as either taken in the home of the photo's owner or during a visit to someone else's home. Such ambiguity could influence the judgment of whether a visual content is considered privacy-threatening, without any clear record of how annotators understand it.

An ideal data collection process should be conducting field studies with real photos taken by participants. The advance of generative models [46] may enable annotators to upload their own photos by substituting real content for realistic fake objects. Future studies may investigate this data collection approach to maintain both truthfulness and privacy respect in image privacy datasets.

8 CONCLUSION

We present DIPA2 – a public-available image privacy dataset providing high-level reasoning of privacy threats and corresponding sharing intentions by annotators from different cultural backgrounds. DIPA2 augmented a existing dataset, DIPA [79], by upgrading following measurements on annotated privacy-threatening content.

The dataset contained 1,304 images to be reconfirmed as privacy-threatening by 600 annotators hired from Japan and the U.K. Our annotators provided 5,897 annotations in 1,304 images, each of which is composed of four kinds of information that can be exploited in various research on image privacy: 1) information telling what types of privacy can be indicated in the annotated content, 2) informativeness measuring how severe if privacy leakages happen, 3) willingness of sharing when annotators assuming they were photo owners, and 4) how annotators would allow their recipients to repost the content.

Our data collection aligned with most of the results in the previous dataset while pruning many annotations that could not obtain the confirmation of our annotators. Through data analysis, we verified that demographic information, sharing habits, cultural backgrounds, and personality traits can influence the odds of determining if visual content is privacy-threatening to different degrees. Moreover, we introduced two baseline experiments aimed at predicting privacy-threatening content in images and four specific privacy metrics tailored to DIPA2. We expect DIPA2 would assist researchers and all stakeholders with image privacy to explore research possibilities and understand image privacy better.

9 DATASET AND CODE AVAILABILITY

All parts of the implementation of this work are publicly available at https://anranxu.github.io/DIPA2_VIS/, including the dataset, as well as corresponding code for data analysis (Section 4) and machine learning baselines (Section 5).

REFERENCES

[1] Mark S Ackerman, Lorrie Faith Cranor, and Joseph Reagle. 1999. Privacy in e-commerce: examining user scenarios and privacy preferences. In *Proceedings of the 1st ACM Conference on Electronic Commerce*. 1–8.

[2] Paarijaat Aditya, Rijurekha Sen, Peter Druschel, Seong Joon Oh, Rodrigo Benenson, Mario Fritz, Bernt Schiele, Bobby Bhattacharjee, and Tong Tong Wu. 2016. I-pic: A platform for privacy-compliant image capture. In *Proceedings of the 14th annual international conference on mobile systems, applications, and services (MobiSys '16)*. 235–248.

[3] Shane Ahern, Dean Eckles, Nathaniel S Good, Simon King, Mor Naaman, and Rahul Nair. 2007. Over-exposed? Privacy patterns and considerations in online and mobile photo sharing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 357–366.

[4] Tousif Ahmed, Apu Kapadia, Venkatesh Potluri, and Manohar Swaminathan. 2018. Up to a limit? privacy concerns of bystanders and their willingness to share additional information with visually impaired users of assistive technologies. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–27.

[5] Taslima Akter, Tousif Ahmed, Apu Kapadia, and Manohar Swaminathan. 2022. Shared privacy concerns of the visually impaired and sighted bystanders with camera-based assistive technologies. *ACM Transactions on Accessible Computing (TACCESS)* 15, 2 (2022), 1–33.

[6] Mahdi Nasrullah Al-Ameen, Tanjina Tamanna, Swapnil Nandy, MA Manazir Ahsan, Priyank Chandra, and Syed Ishtiaque Ahmed. 2020. We don't give a second thought before providing our information: understanding users' perceptions of information collection by apps in Urban Bangladesh. In *Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies*. 32–43.

[7] Rawan Alharbi, Mariam Tolba, Lucia C Petito, Josiah Hester, and Nabil Alshurafa. 2019. To mask or not to mask? balancing privacy with visual confirmation utility in activity-oriented wearable cameras. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies (IMWUT '19)* 3, 3 (2019), 1–29.

[8] Mary Jean Amon, Rakibul Hasan, Kurt Hugenberg, Bennett I Bertenthal, and Apu Kapadia. 2020. Influencing photo sharing decisions on social media: A case of paradoxical findings. In *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, 1350–1366.

[9] Mary Jean Amon, Aaron Necaise, Nika Kartvelishvili, Aneka Williams, Yan Solihin, and Apu Kapadia. 2023. Modeling User Characteristics Associated with Interdependent Privacy Perceptions on Social Media. *ACM Transactions on Computer-Human Interaction* (2023).

[10] Gonul Ayci, Pınar Yolum, Arzucan Özgür, and Murat Şensoy. 2023. Explain to Me: Towards Understanding Privacy Decisions. *arXiv preprint arXiv:2301.02079* (2023).

[11] Kim Bartel Sheehan. 1999. An investigation of gender differences in on-line privacy concerns and resultant behaviors. *Journal of interactive marketing* 13, 4 (1999), 24–38.

[12] Steven Bellman, Eric J Johnson, Stephen J Kobrin, and Gerald L Lohse. 2004. International differences in information privacy concerns: A global survey of consumers. *The Information Society* 20, 5 (2004), 313–324.

[13] Dmitri Bitouk, Neeraj Kumar, Samreen Dhillon, Peter Belhumeur, and Shree K Nayar. 2008. Face swapping: automatically replacing faces in photographs. In *ACM SIGGRAPH 2008 papers*. 1–8.

[14] Laura Brandimarte, Alessandro Acquisti, and George Loewenstein. 2013. Misplaced confidences: Privacy and the control paradox. *Social psychological and personality science* 4, 3 (2013), 340–347.
[15] Kieran Browne, Ben Swift, and Terhi Nurmikko-Fuller. 2020. Camera adversaria. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. 1–9.
[16] Sen-ching Cheung, M. Vijay Venkatesh, Jithendra K Paruchuri, Jian Zhao, and Thinh Nguyen. 2009. Protecting and managing privacy information in video surveillance systems. In *Protecting Privacy in Video Surveillance*. Springer, 11–33.
[17] Hichang Cho, Milagros Rivera-Sánchez, and Sun Sun Lim. 2009. A multinational study on online privacy: global concerns and local responses. *New media & society* 11, 3 (2009), 395–416.
[18] Kovila PL Coopamootoo. 2020. Usage patterns of privacy-enhancing technologies. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*. 1371–1390.
[19] CrowdWorks. 2022. https://crowdworks.jp. Accessed: 2023-5-15.
[20] Ali Dehghantanha and Katrin Franke. 2014. Privacy-respecting digital investigation. In *2014 Twelfth Annual International Conference on Privacy, Security and Trust*. IEEE, 129–138.
[21] Mariella Dimiccoli, Juan Marín, and Edison Thomaz. 2018. Mitigating bystander privacy concerns in egocentric activity recognition with deep learning and intentional image degradation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT '18)* 1, 4 (2018), 1–18.
[22] Agrim Gupta, Piotr Dollar, and Ross Girshick. 2019. Lvis: A dataset for large vocabulary instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5356–5364.
[23] Danna Gurari, Qing Li, Chi Lin, Yinan Zhao, Anhong Guo, Abigale Stangl, and Jeffrey P Bigham. 2019. Vizwiz-priv: A dataset for recognizing the presence and purpose of private visual information in images taken by blind people. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 939–948.
[24] Rakibul Hasan, Bennett I Bertenthal, Kurt Hugenberg, and Apu Kapadia. 2021. Your photo is so funny that i don't mind violating your privacy by sharing it: effects of individual humor styles on online photo-sharing behaviors. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.
[25] Jianping He, Bin Liu, Deguang Kong, Xuan Bao, Na Wang, Hongxia Jin, and George Kesidis. 2016. Puppies: Transformation-supported personalized privacy preserving partial image sharing. In *46th annual IEEE/IFIP international conference on dependable systems and networks (DSN '16)*. IEEE, 359–370.
[26] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
[27] Mariea Grubbs Hoy and George Milne. 2010. Gender differences in privacy-related measures for young adult Facebook users. *Journal of interactive advertising* 10, 2 (2010), 28–45.
[28] Roberto Hoyle, Luke Stark, Qatrunnada Ismail, David Crandall, Apu Kapadia, and Denise Anthony. 2020. Privacy norms and preferences for photos posted online. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 4 (2020), 1–27.
[29] Daniel Scot Hunt and Eric Langstedt. 2014. The influence of personality factors and motives on photographic communication. *The Journal of Social Media in Society* 3, 2 (2014).
[30] Daniel S Hunt, Carolyn A Lin, and David J Atkin. 2014. Communicating social relationships via the use of photo-messaging. *Journal of Broadcasting & Electronic Media* 58, 2 (2014), 234–252.
[31] Mohammed Khwaja, Sumer S Vaid, Sara Zannone, Gabriella M Harari, A Aldo Faisal, and Aleksandar Matic. 2019. Modeling personality vs. modeling personalidad: In-the-wild mobile data analysis in five countries suggests cultural impact on personality models. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–24.
[32] Mohammed Korayem, Robert Templeman, Dennis Chen, David Crandall, and Apu Kapadia. 2016. Enhancing lifelogging privacy by detecting screens. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. 4309–4314.
[33] Abhishek Kumar, Tristan Braud, Young D Kwon, and Pan Hui. 2020. Aquilis: Using contextual integrity for privacy protection on mobile devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–28.
[34] Alina Kuznetsova, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Malloci, Alexander Kolesnikov, et al. 2020. The open images dataset v4. *International Journal of Computer Vision* 128, 7 (2020), 1956–1981.
[35] Caroline Lang and Hannah Barton. 2015. Just untag it: Exploring the management of undesirable Facebook photos. *Computers in Human Behavior* 43 (2015), 147–155.
[36] Fenghua Li, Zhe Sun, Ang Li, Ben Niu, Hui Li, and Guohong Cao. 2019. Hideme: Privacy-preserving photo sharing on social networks. In *IEEE Conference on Computer Communications (IEEE INFOCOM '19)*. IEEE, 154–162.
[37] Yifang Li and Kelly Caine. 2022. Obfuscation Remedies Harms Arising from Content Flagging of Photos. In *CHI Conference on Human Factors in Computing Systems*. 1–25.
[38] Yao Li and Xinning Gui. 2022. Examining co-owners' privacy consideration in collaborative photo sharing. *Computer Supported Cooperative Work (CSCW)* 31, 1 (2022), 79–109.

[39] Yifang Li, Nishant Vishwamitra, Hongxin Hu, and Kelly Caine. 2020. Towards a taxonomy of content sensitivity and sharing preferences for photos. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. 1–14.

[40] Chi Liu, Tianqing Zhu, Jun Zhang, and Wanlei Zhou. 2022. Privacy intelligence: A survey on image privacy in online social networks. *Comput. Surveys* 55, 8 (2022), 1–35.

[41] Paul Benjamin Lowry, Jinwei Cao, and Andrea Everard. 2011. Privacy concerns versus desire for interpersonal awareness in driving the use of self-disclosure technologies: The case of instant messaging in two cultures. *Journal of Management Information Systems* 27, 4 (2011), 163–200.

[42] Karola Marky, Alexandra Voit, Alina Stöver, Kai Kunze, Svenja Schröder, and Max Mühlhäuser. 2020. " I don't know how to protect myself": Understanding Privacy Perceptions Resulting from the Presence of Bystanders in Smart Environments. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. 1–11.

[43] Bryan A Marshall, Peter W Cardon, Daniel T Norris, Natalya Goreva, and Ryan D'Souza. 2008. Social networking websites in India and the United States: A cross-national comparison of online privacy and communication. *Issues in Information Systems* 9, 2 (2008), 87–94.

[44] Kelly Moore and James C McElroy. 2012. The influence of personality on Facebook usage, wall postings, and regret. *Computers in human behavior* 28, 1 (2012), 267–274.

[45] John A Naslund, Ameya Bondre, John Torous, and Kelly A Aschbrenner. 2020. Social media and mental health: benefits, risks, and opportunities for research and practice. *Journal of technology in behavioral science* 5 (2020), 245–257.

[46] Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. 2021. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741* (2021).

[47] Austin Lee Nichols and Jon K Maner. 2008. The good-subject effect: Investigating participant demand characteristics. *The Journal of general psychology* 135, 2 (2008), 151–166.

[48] Peter Nielsen. 2019. Japanese pop singer was nearly killed because of her selfies. https://twiftnews.com/lifestyle/japanese-pop-singer-was-nearly-killed-because-of-her-selfies/. Accessed: 2023-5-15.

[49] Joseph O'Hagan, Pejman Saeghe, Jan Gugenheimer, Daniel Medeiros, Karola Marky, Mohamed Khamis, and Mark McGill. 2023. Privacy-Enhancing Technology and Everyday Augmented Reality: Understanding Bystanders' Varying Needs for Awareness and Consent. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–35.

[50] Kieron O'Hara. 2016. The seven veils of privacy. *IEEE Internet Computing* 20, 2 (2016), 86–91.

[51] Judith S Olson, Jonathan Grudin, and Eric Horvitz. 2005. A study of preferences for sharing and privacy. In *CHI'05 extended abstracts on Human factors in computing systems*. 1985–1988.

[52] OpenAI. 2023. GPT-4 Technical Report. *ArXiv* abs/2303.08774 (2023).

[53] Tribhuvanesh Orekondy, Mario Fritz, and Bernt Schiele. 2018. Connecting pixels to privacy and utility: Automatic redaction of private information in images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 8466–8475.

[54] Tribhuvanesh Orekondy, Bernt Schiele, and Mario Fritz. 2017. Towards a visual privacy advisor: Understanding and predicting privacy risks in images. In *Proceedings of the IEEE international conference on computer vision*. 3686–3695.

[55] Sangkeun Park, Joohyun Kim, Rabeb Mizouni, and Uichin Lee. 2016. Motives and concerns of dashcam video sharing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 4758–4769.

[56] Blaine A Price, Avelie Stuart, Gul Calikli, Ciaran Mccormick, Vikram Mehta, Luke Hutton, Arosha K Bandara, Mark Levine, and Bashar Nuseibeh. 2017. Logging you, logging me: A replicable study of privacy and sharing behaviour in groups of visual lifeloggers. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (2017), 1–18.

[57] Prolific. 2022. https://www.prolific.co/. Accessed: 2023-5-15.

[58] Moo-Ryong Ra, Seungjoon Lee, Emiliano Miluzzo, and Eric Zavesky. 2017. Do not capture: Automated obscurity for pervasive imaging. *IEEE Internet Computing* 21, 3 (2017), 82–87.

[59] Beatrice Rammstedt and Oliver P John. 2007. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of research in Personality* 41, 1 (2007), 203–212.

[60] Champika Ranasinghe, Jakub Krukar, and Christian Kray. 2018. Visualizing location uncertainty on mobile devices: cross-cultural differences in perceptions and preferences. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–22.

[61] Philip J Reed, Emma S Spiro, and Carter T Butts. 2016. Thumbs up for privacy?: Differences in online self-disclosure behavior across national cultures. *Social science research* 59 (2016), 155–170.

[62] Mukesh Saini, Pradeep K Atrey, Sharad Mehrotra, and Mohan Kankanhalli. 2014. W3-privacy: understanding what, when, and where inference channels in multi-camera surveillance video. *Multimedia Tools and Applications* 68, 1 (2014), 135–158.

[63] Tanusree Sharma, Abigale Stangl, Lotus Zhang, Yu-Yun Tseng, Inan Xu, Leah Findlater, Danna Gurari, and Yang Wang. 2023. Disability-First Design and Creation of A Dataset Showing Private Visual Information Collected With People Who Are Blind. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[64] Jiayu Shu, Rui Zheng, and Pan Hui. 2018. Cardea: Context-aware visual privacy protection for photo taking and sharing. In *Proceedings of the 9th ACM Multimedia Systems Conference*. 304–315.

[65] Eleftherios Spyromitros-Xioufis, Symeon Papadopoulos, Adrian Popescu, and Yiannis Kompatsiaris. 2016. Personalized privacy-aware image classification. In *Proceedings of the 2016 ACM on international conference on multimedia retrieval*. 71–78.
[66] Abigale Stangl, Kristina Shiroma, Nathan Davis, Bo Xie, Kenneth R Fleischmann, Leah Findlater, and Danna Gurari. 2022. Privacy concerns for visual assistance technologies. *ACM Transactions on Accessible Computing (TACCESS)* 15, 2 (2022), 1–43.
[67] Julian Steil, Marion Koelle, Wilko Heuten, Susanne Boll, and Andreas Bulling. 2019. Privaceye: privacy-preserving head-mounted eye tracking using egocentric scene image and eye movement features. In *Proceedings of the 11th ACM symposium on eye tracking research & applications*. 1–10.
[68] Jose M Such, Joel Porter, Sören Preibusch, and Adam Joinson. 2017. Photo privacy conflicts in social media: A large-scale empirical study. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 3821–3832.
[69] Kimberly Tee, AJ Bernheim Brush, and Kori M Inkpen. 2009. Exploring communication and sharing between extended families. *International Journal of Human-Computer Studies* 67, 2 (2009), 128–138.
[70] Mike Thelwall and Farida Vis. 2017. Gender and image sharing on Facebook, Twitter, Instagram, Snapchat and WhatsApp in the UK: Hobbying alone or filtering for friends? *Aslib Journal of Information Management* 69, 6 (2017), 702–720.
[71] Robert Thomson, Masaki Yuki, and Naoya Ito. 2015. A socio-ecological approach to national differences in online privacy concern: The role of relational mobility and trust. *Computers in Human Behavior* 51 (2015), 285–292.
[72] Sigal Tifferet. 2019. Gender differences in privacy tendencies on social network sites: A meta-analysis. *Computers in Human Behavior* 93 (2019), 1–12.
[73] Lam Tran, Deguang Kong, Hongxia Jin, and Ji Liu. 2016. Privacy-cnh: A framework to detect photo privacy with convolutional neural network using hierarchical features. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 30.
[74] Ho Keung Tsoi and Li Chen. 2011. From privacy concern to uses of social network sites: A cultural comparison via user survey. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*. IEEE, 457–464.
[75] Robyn Vanherle, Hanneke Hendriks, and Kathleen Beullens. 2023. Only for friends, definitely not for parents: Adolescents' sharing of alcohol references on social media features. *Mass Communication and Society* 26, 1 (2023), 47–73.
[76] Nishant Vishwamitra, Yifang Li, Hongxin Hu, Kelly Caine, Long Cheng, Ziming Zhao, and Gail-Joon Ahn. 2022. Towards Automated Content-based Photo Privacy Control in User-Centered Social Networks. In *Proceedings of the Twelveth ACM Conference on Data and Application Security and Privacy*. 65–76.
[77] Zhuo Wei, Yongdong Wu, Yanjiang Yang, Zheng Yan, Qingqi Pei, Yajuan Xie, and Jian Weng. 2018. AutoPrivacy: Automatic privacy protection and tagging suggestion for mobile social photo. *Computers & Security* 76 (2018), 341–353.
[78] Zhenyu Wu, Haotao Wang, Zhaowen Wang, Hailin Jin, and Zhangyang Wang. 2020. Privacy-preserving deep action recognition: An adversarial learning framework and a new dataset. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 4 (2020), 2126–2139.
[79] Anran Xu, Zhongyi Zhou, Kakeru Miyazaki, Ryo Yoshikawa, Simo Hosio, and Koji Yatani. 2023. DIPA: An Image Dataset with Cross-cultural Privacy Concern Annotations. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces*. 259–266.
[80] Kaihe Xu, Yuanxiong Guo, Linke Guo, Yuguang Fang, and Xiaolin Li. 2015. My privacy my decision: Control of photo sharing on online social networks. *IEEE Transactions on Dependable and Secure Computing* 14, 2 (2015), 199–210.
[81] Jun Yu, Zhenzhong Kuang, Baopeng Zhang, Wei Zhang, Dan Lin, and Jianping Fan. 2018. Leveraging content sensitiveness and user trustworthiness to recommend fine-grained privacy settings for social image sharing. *IEEE transactions on information forensics and security (IEEE TIFS '18)* 13, 5 (2018), 1317–1332.
[82] Jinao Yu, Hanyu Xue, Bo Liu, Yu Wang, Shibing Zhu, and Ming Ding. 2020. Gan-based differential private image privacy protection framework for the internet of multimedia things. *Sensors* 21, 1 (2020), 58.
[83] Sergej Zerr, Stefan Siersdorfer, Jonathon Hare, and Elena Demidova. 2012. Privacy-aware image classification and search. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. 35–44.
[84] Lan Zhang, Kebin Liu, Xiang-Yang Li, Cihang Liu, Xuan Ding, and Yunhao Liu. 2016. Privacy-friendly photo capturing and sharing system. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. 524–534.
[85] Chenye Zhao, Jasmine Mangat, Sujay Koujalgi, Anna Squicciarini, and Cornelia Caragea. 2022. PrivacyAlert: a dataset for image privacy prediction. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 1352–1361.

A GENERATION OF PRIVACY-THREATENING CONTENT CATEGORIES AND BOUNDING BOXES

This section cites and outlines the data collection methodology carried out by Xu et al. [79], providing a clear illustration of the source of the privacy-threatening content categories, source images, and corresponding bounding boxes adopted in the annotation process of our research.

Table 12. Twenty-five categories of privacy-threatening content in the first-stage study of DIPA [79], derived from the collection of 1,632 photos with 1,894 annotations from 171 participants from CrowdWoks [19].

Category [% (count)]	**Description**
Person (except bystanders) [43% (809)]	People who are intended to be photographed
Place Identifier [17% (328)]	Road signs, building signs, maps, and other environmental hints that indicate locations
Identity [8% (155)]	Identity information on tickets, passports, nameplates, etc
Home Interior [5% (91)]	Furniture and home decoration that may imply people's habits and locations
Vehicle Plate [4% (84)]	Identifiable information for cars but also for their owners
Bystander [4% (72)]	People who are photographed without permission.
Food [4% (68)]	Close-up views of food or party scenes.
Printed Materials [3% (62)]	Various printed materials containing private information.
Screen [3% (57)]	Computer monitors, smartphone screens, electronic information boards, etc
Clothing [2% (46)]	Clothing that may imply personal identity, habits, and occupations.
Scenery [2% (42)]	Backgrounds which may imply locations or personal information
Pet [1% (23)]	Pet ownership can be private information to some people.
Book [1% (12)]	Books that may imply personal information and preferences
Photo [1% (11)]	Other photos in the captured image
Machine [0% (9)]	Machines used in a workplace or specific areas
Table [0% (5)]	Tables with many personal items
Electronic Devices [0% (5)]	Electronic devices that photo owners' regard as private
Cosmetics [0% (4)]	Personal care products that may reveal owners' habits or locations
Toy [0% (4)]	Toys for children or photo owners
Finger [0% (2)]	Finger close-up that can be used to infer a person's fingerprint
Cigarettes [0% (1)]	Cigarettes or smoking scenes
Accident [0% (1)]	Accident scenes
Musical Instrument [0% (1)]	Instruments or playing scenes.
Nudity [0% (1)]	Naked upper body
Accessory [0% (1)]	Accessories worn by people photographed
Total [100% (1894)]	

Xu et al. employed a two-stage data collection approach in the formation of DIPA. In the first stage, they collected 1,632 private photos featuring 1,894 detailed annotations of privacy-threatening content from 171 Japanese participants. Following this, they derived 25 categories of common privacy-threatening content through the analysis of the collected images and annotations (Table 12).

Excluding three categories ("Nudity," "Accident," and "Bystander") due to their absence in public image datasets, the researchers extracted 2,090 images from two public datasets, OpenImages [34] and LVIS [22], in the second

stage. These images included at least one of the remaining 22 categories, substituting the private images from the first stage. During the annotation phase, each image retained all bounding boxes that were originally provided by the source datasets (i.e., OpenImages or LVIS), regardless of whether they were associated with the 22 predefined categories. Additionally, all visual contents encompassed within bounding boxes were labeled with specific object names, as specified by the original datasets (e.g., “television” as a specific name that belongs to the "screen" category within the 22 categories), assisting the annotators in a clearer understanding of the enclosed objects. Annotators were required to use an interface similar to the one depicted in Figure 1, where they were also able to manually add bounding boxes as needed. 177 Japanese annotators and 183 British annotators, recruited from CrowdWorks [19] and Prolific [57], respectively, identified 1,495 images with 5,671 annotations on 4,642 privacy-threatening content within the selected images, thus forming the DIPA.

In this paper, we incorporated the pre-established 22 categories of privacy-threatening content and all bounding boxes, including those originally sourced from OpenImages and LVIS and those manually added by annotators, from DIPA in our creation of DIPA2.

B CATEGORY MAPPING FROM VISPR AND VIZWIZ-PRIV TO DIPA2

Table 13 provides a comprehensive mapping from the categories defined in VISPR [54] and VizWiz-Priv [23] to those in DIPA2. While VISPR and VizWiz-Priv define 67 and 23 categories respectively, each map to a different subset of 9 categories within the total 23 categories (22 identified privacy-threatening categories and “Others” category) defined in DIPA2. This highlights that the category definitions in DIPA2 encompass a broader scope of privacy-threatening content in realistic scenarios.

Table 13. Mapping between DIPA2 categories and categories from VISPR and VizWiz-Priv datasets.

Categories in DIPA2	Categories in VISPR [54]	Categories in VizWiz-Priv [23]
Person	a1_age_approx, a10_face_partial, a12_semi_nudity, a13_full_nudity, a16_race, a2_weight_approx, a3_height_approx, a4_gender, a39_disability_physical, a41_injury, a5_eye_color, a57_culture, a58_hobbies, a59_sports, a61_opinion_general, a62_opinion_political, a64_rel_personal, a65_rel_social, a66_rel_professional, a67_rel_competitors, a68_rel_spectators, a69_rel_views, a9_face_complete	Object:Face_Reflection, Object:Face, Object:Tattoo
Printed Material	a26_handwriting, a35_mail, a37_receipt, a38_ticket, a82_date_time	Text:Business_Card, Text:Miscellaneous_Papers, Text:Pill_Bottle/Box, Text:Menu, Text:Letter, Text:Newspaper, Text:Poster, Text:Receipt
Book		Text:Book
Screen	a97_online_conversation	Text:Computer_Screen
Vehicle Plate	a103_license_plate_complete, a104_license_plate_partial, a102_vehicle_ownership	Text:License_Plate
Identity	a19_name_full, a20_name_first, a21_name_last, a24_birth_date, a25_nationality, a27_marital_status, a29_ausweis, a30_credit_card, a31_passport, a32_drivers_license, a33_student_id, a46_occupation, a55_religion, a56_sexual_orientation, a70_education_history, a7_fingerprint, a8_signature, a85_username, a90_email, a99_legal_involvement	Text:Credit_Card
Clothing	a18_ethnic_clothing	Text:Clothing
Place Identifier	a23_birth_city, a48_occassion_work, a60_occassion_personal, a73_landmark, a74_address_current_complete, a75_address_current_partial, a78_address_home_complete, a79_address_home_partial	Text:Street_Sign
Electronic Devices	a49_phone	
Others	a17_color, a43_medicine	Object:Suspicious, Text:Suspicious, Object:Pregnancy_Test_Result, Text:Other, Object:Other