Modeling the Trade-off of Privacy Preservation and Activity Recognition on Low-Resolution Images

Yuntao Wang*
yuntaowang@tsinghua.edu.cn
Key Laboratory of Pervasive Computing, Ministry of Education, Department of Computer Science and Technology, Tsinghua University
Beijing, China

Zirui Cheng*
chengzr19@mails.tsinghua.edu.cn
Department of Computer Science and Technology, Tsinghua University
Beijing, China

Xin Yi†
yixin@tsinghua.edu.cn
Institute for Network Sciences and Cyberspace, Tsinghua University
Beijing, China
Zhongguancun Laboratory
Beijing, China

Yan Kong
ky21@mails.tsinghua.edu.cn
Institute for Network Sciences and Cyberspace, Tsinghua University
Beijing, China

Xueyang Wang
wang-xy22@mails.tsinghua.edu.cn
Institute for Network Sciences and Cyberspace, Tsinghua University
Beijing, China

Xuhai Xu
xuhaixu@cs.washington.edu
Information School, University of Washington
Seattle, WA, USA

Yukang Yan
yanyukanglwy@gmail.com
Department of Computer Science and Technology, Tsinghua University
Beijing, China

Chun Yu
chunyu@mail.tsinghua.edu.cn
Department of Computer Science and Technology, Tsinghua University
Beijing, China

Shwetak Patel
shwetak@cs.washington.edu
Paul G. Allen School of Computer Science and Engineering, University of Washington
Seattle, WA, USA

Yuanchun Shi
shiyc@tsinghua.edu.cn
Department of Computer Science and Technology, Tsinghua University
Beijing, China
Qinghai University
Xining, Qinghai, China

ABSTRACT

A computer vision system using low-resolution image sensors can provide intelligent services (e.g., activity recognition) but preserve unnecessary visual privacy information from the hardware level. However, preserving visual privacy and enabling accurate machine recognition have adversarial needs on image resolution. Modeling the trade-off of privacy preservation and machine recognition performance can guide future privacy-preserving computer vision systems using low-resolution image sensors. In this paper, using the at-home activity of daily livings (ADLs) as the scenario, we first obtained the most important visual privacy features through a user survey. Then we quantified and analyzed the effects of image resolution on human and machine recognition performance in activity recognition and privacy awareness tasks. We also investigated how modern image super-resolution techniques influence these effects. Based on the results, we proposed a method for modeling the trade-off of privacy preservation and activity recognition on low-resolution images.

*The authors contribute equally to this paper.
†denotes as the corresponding author.

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3581425

CCS CONCEPTS

• **Human-centered computing → User studies**; • **Security and privacy → Privacy protections**; • **Computing methodologies → Computer vision**.

KEYWORDS

Privacy, visual privacy, privacy preserving, activities of daily living, ADLs, low-resolution image.

ACM Reference Format:

Yuntao Wang, Zirui Cheng, Xin Yi, Yan Kong, Xueyang Wang, Xuhai Xu, Yukang Yan, Chun Yu, Shwetak Patel, and Yuanchun Shi. 2023. Modeling the Trade-off of Privacy Preservation and Activity Recognition on Low-Resolution Images. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3544548.3581425

1 INTRODUCTION

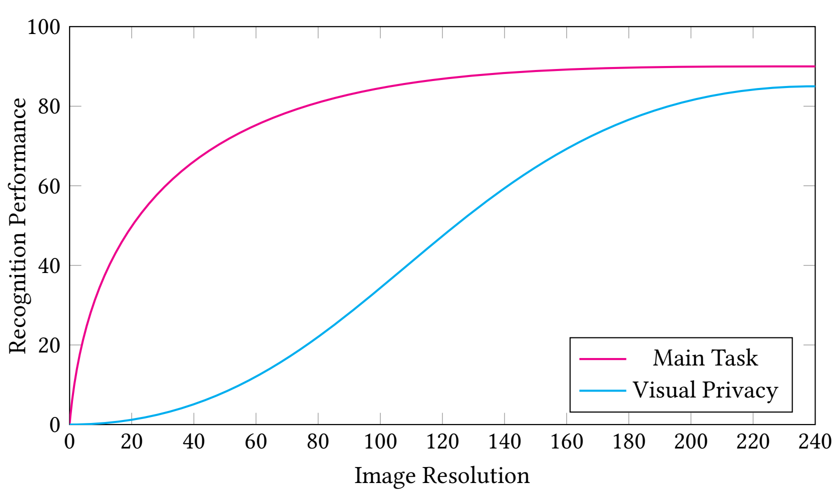
The advances in technological engineering have enabled cameras to be increasingly ubiquitous. Nowadays, many cameras can be manufactured at a low cost in a power-efficient manner, and with small sizes. With the help of artificial intelligence, these cameras are enabled with automatic recognition abilities, providing smart services and applications publicly or privately [6]. However, in realistic scenarios, this brings up a major concern — visual privacy exposure. We expect a vision-based system that can bring intelligent applications while preserving visual privacy.

To achieve this purpose, researchers have explored many post-processing methods, which were often accomplished by decoupling the personally identifiable information (e.g., face) [7, 22, 24, 33, 43, 50]. However, these solutions are not sufficient to process all visual privacy cues [42, 43, 48, 49].

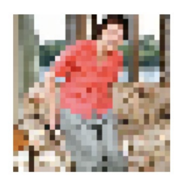
As suggested by related works [10, 47–49, 60], a fundamental solution toward the construction of a privacy-preserving vision-based system is to lower the image sensor's resolution from the hardware level. Thus, machines can achieve applicable performance in the main recognition task (e.g., activity recognition), while preserving visual privacy as much as possible. Related works have proved that a low-resolution image (e.g., 16×12 pixels) possesses sufficient visual features for the main recognition task but not for visual privacy awareness. However, a high-resolution image can provide enough visual features for both of these two tasks. Thus, there is a trade-off regarding the effect of the image resolution on the main recognition task and visual privacy awareness as Figure 1 illustrates. Understanding and modeling such a trade-off will provide guidance for the privacy-preserving vision-based system with low-resolution visual sensors.

In this paper, we focus on a smart home scenario where low-resolution image sensors automatically recognize activities of daily living (ADLs), such as feeding, entertainment, personal hygiene, intimacy, and functional mobility. ADLs recognition system can summarize activities and daily routines on which the ability of a person living independently is assessed; thus is widely used for health monitoring, especially for elderly care [16, 35]. In realistic home environments, the data captured by an image sensor may be single-frame pictures [10, 47] or multi-frame videos [23, 40]. We regard both of them as *images* to model the trade-off between privacy preservation and activity recognition.

We considered such a trade-off as an optimization problem over image resolution. We conducted an online user survey with 115 participants to obtain the most important visual privacy features including nudity, identifiable face, valuable property, and relationship. In this paper, we regarded both the human and the machine as recognizers. Thus, we explored the effect of image resolution on both human and machines' ability in activity recognition and visual privacy awareness on the PA-HMDB51 dataset, which consists of over 500 videos from realistic environments [57]. Specifically, we conducted a user study with 240 participants to investigate the effect of image resolution on human recognition performance. We evaluated the machine's performance on ADLs and visual privacy recognition tasks with cutting-edge machine learning approaches. Finally, we built a modeling method for calculating the trade-off of visual privacy preserving ADLs recognition using low-resolution images. We envision that our method can inspire other vision-based systems that require balancing privacy awareness and machine recognition performance. Overall, the contributions of our paper are two-fold.



(a) 20×20



(b) 30×30

(c) 50×50

(d) 100×100

Figure 1: Demonstration of the effects of image resolution on the performance of the main vision-based recognition task and visual privacy awareness.

1) Using the at-home ADLs recognition as a scenario, we proposed a pipeline to investigate the effect of the image resolution on both human and machine performance on the main activity recognition task and visual privacy awareness.

2) We presented a model for calculating the trade-off of visual privacy preserving activity recognition using low-resolution images. Using the proposed model, we can calculate an optimal resolution range of the image sensor for privacy-preserving activity recognition applications.

2 RELATED WORK

We describe the related work in this section, including visual privacy features and taxonomy, privacy-preserving machine recognition, and balancing the trade-off between privacy preservation and machine recognition.

2.1 Visual Privacy Features and Taxonomy

Privacy is described as "the right to select what personal information about me is known to what people" [21]. Pictures or videos convey a broad spectrum of privacy information, namely visual privacy. While legal and government entities legislated laws and policies on privacy protection [11, 56], their guidance leaves room for intruding visual privacy. Recently, researchers have explored the visual privacy exposure degree, visual privacy taxonomies/features, visual privacy importance, and visual privacy risk assessment using social media image databases [36, 41, 42]. Orekondy et al. summarized 68 kinds of visual privacy features on social media images and then explored the feasibility of evaluating visual privacy exposure degree through machine learning approaches [41, 42]. Li et al. summarized 7 categories, including 22 visual privacy features by crowdsourcing users' descriptions in their photo album [36]. These researches provide fundamental guidelines on taxonomy and the importance of visual privacy, which inspired us to design our user survey to explore the perceived importance of visual privacy in a home environment under varying image resolutions.

2.2 Privacy-Preserving Machine Recognition

A growing number of privacy preserving computation technologies have emerged in recent years, which share the common promise of preserving privacy while also obtaining the benefits of computational analysis [1]. To preserve visual privacy, existing solutions mainly adopted post-processing techniques such as image blurring and encryption techniques for images containing visual privacy information, e.g., human faces [7, 22, 24, 30, 33, 43, 49, 50]. However, these solutions are insufficient to protect all privacy information, including readable addresses, phone numbers, etc. [43, 48, 49].

Recently, researchers proposed a fundamental solution for a privacy-preserving vision-based system — to lower the image sensor's resolution from the hardware level [10, 40, 47, 49, 60, 61]. Specifically, Miyazaki et al. developed a technology that can accurately detect the flow of people on low-resolution videos in which the faces cannot be distinguished [40]. Dai et al. simulated a privacy-protected smart room prototype and then studied the performance impact of the image resolution from a single pixel to 10×10 pixels [15]. They evaluated that five 10×10 resolution cameras can achieve a fairly high accuracy of 89.6% on recognizing 9 human poses. Ryoo et al. proposed the inverse super-resolution (ISR) method for activity recognition on ultra-low-resolution videos, which also achieved state-of-art recognition accuracy while preserving identifiable personal information [47, 49].

These solutions showed the feasibility of activity recognition on low-resolution images. However, they only assumed an image resolution threshold (e.g., 10×10 [15]) to be able to preserve visual privacy without evidence. Obviously, the lower the resolution is, the better the visual privacy can be preserved. However, a lower resolution will inevitably decrease the amount of information for activity interpretation. It remains unknown how to balance the two adversarial demands on image resolution for recognizing the activity and safeguarding visual privacy. Our work is to answer this question by proposing a mathematical trade-off model and a method to calculate the optimal resolution range.

2.3 Balancing Privacy Preservation and Machine Recognition

Researchers have discussed the trade-off between privacy preservation and activity recognition by quantifying humans' perceptions of privacy features. Some existing works have explored the effect of image resolution on human ability in facial recognition [26, 62]. Harmon and Julesz found that humans are good at facial recognition even when the portrait's resolution is down to 16×16 pixels [26]. Yip and Sinha found that humans can still recognize celebrities' faces on portraits with a resolution of merely 7×10 pixels [62]. Some researchers also explored the impact of blur or pixelize filters at various levels on visual privacy awareness and activity recognition in the context of common workplace activities [8] or crowdsourced behavioral video coding [34]. They concluded the feasibility of achieving activity awareness while preserving visual privacy when tested on human eyes.

Taking human or the machine recognition performances into account, some researchers tried to understand how to balance privacy preservation and recognition performance. Alharbi et al. evaluated the effect of varying degrees of obfuscation on bystander privacy and visual confirmation utility [5]. Hasan et al. studied the relative trade-offs between privacy (revealing and concealing selective attributes of objects) and utility (the visual aesthetics and user satisfaction of the image) of different image transforms [27]. Wu et al. formulated a novel adversarial training framework to learn anonymization transform for input videos such that the trade-off between target utility task performance and the associated privacy budgets is explicitly optimized on the anonymized videos [57].

However, these existing works have three limitations. First, They only tested humans' interpretation ability. An intelligent application relies on the machine for the main recognition task rather than the human. A more comprehensive study is highly demanded to explore how resolution affects both the human and the machine's recognition performance. Second, they mainly regarded the character's face as a privacy feature, which is insufficient to quantify a fine-grained model for privacy-preserving applications. Third, applying post-processing filters to high-resolution images differs from lowering the image sensor's resolution, which can preserve the visual privacy information from the hardware level with fewer on-device computing resources required.

Our work fills the gap mentioned above. We targeted enabling visual privacy-preserving machine recognition applications on low-resolution image sensors. We modeled the effects of image resolution on both the human and machine's ability in activity recognition and visual privacy awareness. Further, we proposed a quantitative survey method to model the importance of comprehensive visual privacy features.

3 PROBLEM DEFINITION AND IMPLEMENTATION PIPELINE

This section offers the mathematical definition of privacy-preserving machine recognition using low-resolution images.

3.1 Problem Definition

Assume $\mathcal{X}$ to be the raw image set in a realistic environment that could be captured by the image sensor, for example, single-frame pictures or multi-frame videos. $f_r(\mathcal{X})$ represents the captured image set from the image sensor at a resolution of $r \times r$. Assume T to be the main recognition task associated with $\mathcal{X}$, in this paper, ADLs recognition. P is the visual privacy awareness task associated with $\mathcal{X}$. There are three main components in our model.

- **Recognition Function**. We define the recognition function of the main recognition task T as $f_T(\cdot)$ and the privacy detection function designed for the privacy feature P as $f_P(\cdot)$. Both $f_T(\cdot)$ and $f_P(\cdot)$ can generate the recognition results given the captured image set $f_r(\mathcal{X})$. To give an example of machine recognition, $f_T(\cdot)$ and $f_P(\cdot)$ can be computer vision models such as artificial neural networks.
- **Evaluation Function**. We define the evaluation function $L_T(\cdot)$ and $L_P(\cdot)$ which take both the outcome of the recognition function $f_T(f_r(\mathcal{X}))$ and $f_P(f_r(\mathcal{X}))$ as input and evaluate the performance of the recognizers according to the ground truth labels $g_T(\mathcal{X})$ and $g_P(\mathcal{X})$.
- **Importance Weights**. Considering the variety of privacy features contained in the captured image set, there may be differences in humans' perceived importance of different privacy features. For a given type of privacy feature P_i in the privacy feature set $\mathcal{P}$, we define a weight coefficient ω_i to denote humans' perceived importance of P_i.

To optimize the trade-off between privacy preservation and activity recognition empirically, many prior works in computer vision have focused on finding suitable measurement metrics and objective functions mathematically [25, 45, 52, 57, 58]. However, most of these aforementioned works ignored humans' mental evaluation and recognition abilities of privacy features, and were thus insufficient. Based on prior works, we regard our research problem as mathematically optimizing the objective function $S(r)$ shown in Equation 1, to reveal the trade-off between privacy preservation and machine recognition with both machine and human factors taken into consideration.

$$S(r) = L_T\left(f_T(f_r(\mathcal{X})), g_T(\mathcal{X})\right) - \lambda \sum_{i=1}^{n} \omega_i L_{P_i}\left(f_{P_i}(f_r(\mathcal{X})), g_{P_i}(\mathcal{X})\right) \tag{1}$$

Here, $\lambda > 0$ is a scaling factor representing the sensitivity ratio of visual privacy preservation over activity recognition performance. The goal of formulating our research problem in the form of Equation 1 is to find out optimal resolution ranges where (1) cameras are limited from obtaining detailed visual information to preserve as much privacy information as possible and (2) cameras are able to capture as much detailed non-private information as possible to improve recognition performance.

3.2 Implementation Pipeline

The implementation pipeline of solving the optimization problem in Equation 1 has been depicted in Figure 2. In this paper, we choose the activities of daily living (ADLs) recognition task in a home environment as our target task T. First of all, we conducted a user study to obtain humans' perceived importance of various privacy features (ω in our formulation), with the main results presented in section 4. To evaluate our model in realistic environments, we utilized the publicly available video dataset PA-HMDB51 which is described in section 5.

The critical step of the whole implementation procedure is to model the recognition abilities of humans and machines under different resolutions. In order to preserve privacy comprehensively, we consider the recognition performance of both the human and the machine for each specific privacy feature P to get an estimation of the evaluation results (L_P in our formulation). The processing logic

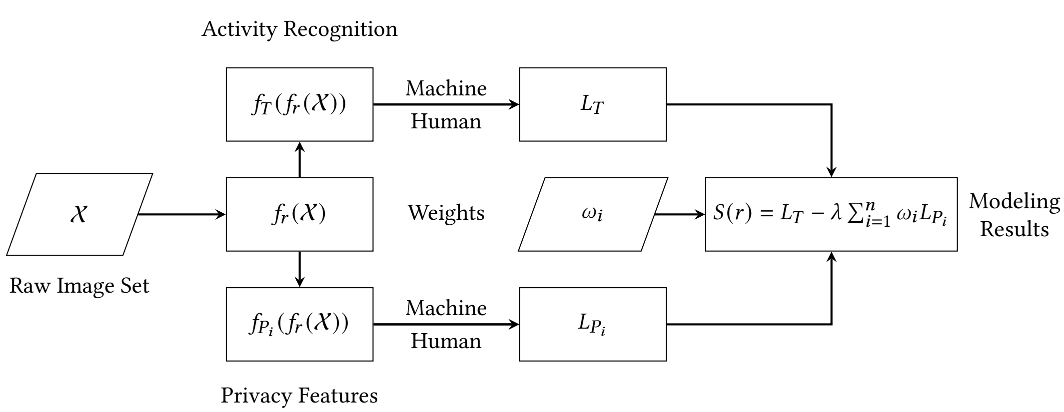


Figure 2: Framework for modeling the trade-off between privacy preservation and activity recognition.

is similar for the activity recognition task to obtain the evaluation results (L_T on our formulation). We conducted a user study to model human recognition performance on those tasks in section 6. Then, we utilized state-of-art computer vision models to finish those recognition tasks under different resolutions in section 7. Also, we provided additional analysis in section 8 to check whether our modeling results are robust against currently state-of-art super-resolution techniques. In the end, we proposed the calculating procedure of our objective function in section 9 to model the trade-off between privacy preservation and activity recognition.

4 QUANTIFYING THE IMPORTANCE OF VISUAL PRIVACY FEATURES

Our first user study aims to understand what visual privacy features users value the most and quantify the importance of those visual privacy features, thus simplifying the to-be-built model (Equation 1). Inspired by related works [13, 36, 42], we obtained 25 visual privacy features that exist in a home environment. Here we divide them into 5 categories as below.

- **Biometric Identification**: identifiable face, gender, skin color, age group, weight group, hair color, eye color, and height group.
- **Personal Marker/Information**: nudity, home address, number/code, medical treatment, physical disability, handwriting, birthday, clothing, and tattoo.
- **Ethnicity**: religion, race, and nationality.
- **Society**: relationship, employment and pet.
- **Safety**: valuable property and living schedule.

4.1 User Survey on Importance of Visual Privacy Features

We recruited 125 participants (66 females, 59 males) from MTurk. They had an average age of 32.7 (s.d. = 14.7). The whole survey lasted around 15 minutes. Each participant who passed the attention check received a 6 USD Amazon gift card.

In the user survey hosted by Qualtrics [1], we first introduced the smart home scenario where cameras are installed for ADLs recognition. Then we asked the participants to assume that they were living in the demonstrated house/apartment. Then we evaluated the importance of each privacy feature with or without low-resolution to find out what privacy features users value the most and explore the effect of low-resolution on users' perceived importance of privacy features. For the high resolution test, we showed participants five high-resolution (300×300) images. Each image captured one of the five basic daily activities: functional mobility, feeding, intimacy, entertainment, and personal hygiene. We did not control the participants' backgrounds regarding their culture, age, gender, and technical knowledge. In the instruction, we explicitly stated the scenario of visual privacy leakage as their similar pictures were posted on the Internet and thus can be accessed by everyone. For the low-resolution test, we just showed participants the same five images in low-resolution (50×50). Under each resolution, the participant was asked how he/she values the importance of the different visual privacy information listed in the questionnaire. Then, the participants were required to rate the importance of each visual privacy feature using a 100-point slider where 0 stands for not important at all and 100 stands for extremely important. The score of each privacy feature shown on the slider updates along with the participant's choice.

We designed two attention check questions under each condition. Each attention check question requires the participant to slide to a certain score that was generated randomly before each survey. All the questions were provided to the participant in random order.

[1] https://www.qualtrics.com/

4.2 Result

In total, we received 115 valid responses out of 120 total responses, in which respondents successfully completed the survey and passed all attention check questions. We utilized the Wilcoxon signed-rank test ($p < 0.05$) and Friedman test ($p < 0.05$) for statistical analysis since the rating scores are ordinal.

The analysis results are listed in Table 1. We concluded with the following findings.

1. **Lowering the image sensor's resolution can significantly decrease users' concerns about visual privacy**. Table 1 shows the average and the standard deviation of the rating score of each visual privacy feature under two different resolution conditions. On average, people rated visual privacy features with significantly lower importance scores ($Z = -4.02, p < 0.001$) under the low-resolution condition ($avg. = 45.1$) than the high-resolution condition ($avg. = 49.3$).

2. **Identifiable face, nudity, home address, number/code, medical treatment, relationship, employment, valuable property, and living schedule are considered to be more important than other visual privacy features.** Statistic analysis indicates that visual privacy features have significant effects on the human perceived important scores under either the high-resolution condition ($\chi^2(25, N = 115) = 298.5, p < 0.001$) or low-resolution condition ($\chi^2(25, N = 115) = 169.9, p < 0.001$). When we ran the pairwise statistical analysis using Wilcoxon signed-rank test among visual privacy features, we concluded with the following major results. On both high-resolution and low-resolution images, identifiable face, nudity, home address, number/code, medical treatment, relationship, employment, valuable property, and living schedule were considered the most important visual privacy features, since users rated them with significantly higher scores than other features ($p < 0.05$). Among these important privacy features, medical treatment and employment were considered less important ($p < 0.05$).

3. **Nudity, identifiable face, valuable property, and living schedule are the most important privacy features despite the image resolution.** When compared with the high-resolution condition, we observed significantly lower importance scores on features including home address ($p = 0.01$), medical treatment ($p < 0.001$), and relationship ($p < 0.01$) under the low-resolution condition. This finding is reasonable since these privacy features require high-resolution details to interpret. For instance, people were less concerned about the readable texts on low-resolution images. However, nudity, identifiable face, valuable property, and living schedule still lead to the most concerned visual privacy features in the low-resolution setting, with an average score above 57.

Table 1: The statistic of the user rated importance scores of the 25 visual privacy features in 5 categories with and without the low-resolution conditions. $p < 0.05$ indicates significant difference between high and low resolution conditions.

Category	Feature	High Resolution		Low Resolution		Significance
		avg.	std.	avg.	std.	
Biometric Identification	Identifiable Face	60.2	24.3	**57.5**	26.0	$p = 0.13$
	Gender	43.5	29.2	43.4	29.4	$p = 0.81$
	Skin Color	42.0	28.6	43.1	27.3	$p = 0.94$
	Age Group	42.9	25.1	41.2	25.8	$p = 0.35$
	Weight Group	43.9	27.2	40.9	27.2	$p = 0.16$
	Hair Color	36.2	27.4	40.9	28.1	$p = 0.05$
	Eye Color	40.4	28.9	40.3	28.4	$p = 0.90$
	Height Group	37.3	25.8	40.0	27.7	$p = 0.30$
Personal Marker / Information	Nudity	61.6	30.9	**62.9**	29.4	$p = 0.71$
	Home Address	62.8	23.1	55.6	26.1	$p = 0.01$
	Number/code	57.5	25.5	55.6	26.6	$p = 0.79$
	Medical Treatment	60.4	23.2	51.7	25.9	$p < 0.001$
	Physical Disability	52.1	25.1	49.4	26.0	$p = 0.25$
	Hand Writing	52.6	26.4	44.9	27.7	$p < 0.01$
	Birthday	54.2	26.8	44.7	28.5	$p < 0.01$
	Clothing	40.5	27.9	41.5	27.5	$p = 0.94$
	Tattoo	42.2	28.7	39.2	28.6	$p = 0.34$
Ethnicity	Religion	41.8	27.7	44.6	26.6	$p = 0.29$
	Race	40.1	26.5	42.2	27.7	$p = 0.64$
	Nationality	42.1	28.3	41.3	27.5	$p = 0.46$
Society	Relationship	60.3	24.8	**52.9**	25.7	$p < 0.001$
	Employment	58.2	22.8	52.1	25.8	$p = 0.05$
	Pet	37.3	24.4	39.1	27.8	$p = 0.46$
Safety	Valuable Property	64.0	25.0	**59.6**	26.1	$p = 0.34$
	Living Schedule	59.3	24.4	59.1	26.3	$p = 0.10$

Instead of considering all the visual privacy features, we want to explore the most concerned ones that have the highest importance score and are potentially still vulnerable to low-resolution images. Therefore, we chose the most important visual privacy features in each category under the low-resolution condition with a minimum importance score threshold of 50.0. As a result, four visual privacy features including ***nudity***, ***identifiable face***, ***valuable property*** and ***relationship*** were chosen for later user studies and analysis.

5 ADLS DATASET WITH VISUAL PRIVACY FEATURES

This section describes the dataset we used to explore the effect of image resolution on humans' and machines' performance on activity recognition and visual privacy awareness tasks.

5.1 Constructing the ADLs Dataset

In order to evaluate the model in realistic environments, we used the publicly-available PA-HMDB51 dataset for privacy-preserving activity recognition [57]. This dataset consists of about 355 minutes and 51 types of human activity videos collected from realistic environments with various visual privacy features annotated.

In this paper, we mainly focus on activities of daily living (ADLs) in a smart home scenario. Therefore, three of our authors selected the qualified videos from the PA-HMDB51 dataset together with the following requirements. 1) The video represents a home environment. 2) All authors agreed that the main character conducted the same kind of activities. 3) All authors felt comfortable to publish the video online. For instance, due to the internet policy, we only chose men's or kids' topless videos in this study. Then, we divided the human activities in the PA-HMDB51 dataset into five basic kinds of activities of daily living (ADLs) including *functional mobility*, *feeding*, *intimacy*, *entertainment*, and *personal hygiene*. Finally, we obtained 46, 30, 22, 37, and 16 minutes of videos for functional mobility, feeding, intimacy, entertainment, and personal hygiene, respectively.

We randomly split the PA-HMDB51 dataset into a training dataset, a validation dataset, and an evaluation dataset, which accounts for 90%, 5%, and 5%, respectively. Considering the difference of the video duration in the PA-HMDB51 dataset, we divided all the videos into 2-second clips for later training and evaluation without affecting the judgment of the video content. Therefore, there are 226 clips of the videos in the evaluation dataset, with 69, 45, 33, 55, and 24 clips for functional mobility, feeding, intimacy, entertainment, and personal hygiene, respectively.

5.2 Labeling the Privacy Features

Based on the user study results presented in section 4, we annotated each frame and each clip in our dataset with privacy features including *nudity*, *identifiable face*, *valuable property*, and *relationship*. Since privacy features may vary during the video clip, for example, even in the same video clip, the visibility of a person's face may be different, we provided both *frame-level* and *clip-level* labels of for each video in our dataset. First of all, we annotated all of the privacy attributes on each frame of different clips. Then, we annotated each

- **Type of ADLs**: feeding.
- **Nudity**: fully closed.
- **Identifiable face:** yes.
- **Valuable property:** yes.
- **Relationship:** only one person.

- **Type of ADLs**: personal hygiene.
- **Nudity**: semi-naked.
- **Identifiable face:** no.
- **Valuable property:** no.
- **Relationship:** only one person.

- **Type of ADLs**: feeding.
- **Nudity**: fully closed.
- **Identifiable face:** yes.
- **Valuable property:** yes.
- **Relationship:** intimate relationship.

- **Type of ADLs**: entertainment.
- **Nudity**: fully closed.
- **Identifiable face:** yes.
- **Valuable property:** no.
- **Relationship:** only one person.

Figure 3: Examples of the annotated frames in our dataset.

clip according to the frames in the clip for later user studies and machine experiments. The detailed description of both frame-level and clip-level labels are listed below.

- **Nudity**. The nudity label of each frame included three types that are *naked or semi-naked (topless or bottomless)*, *fully clothed*, and *no person*. A clip is labeled as *naked or semi-naked (topless or bottomless)* if at least one frame of the clip is labeled as *naked or semi-naked (topless or bottomless)*. Otherwise, the clip is labeled as *fully clothed* in a similar way. If every frame is labeled as *no person*, we will finally label the clip as *no person*.
- **Identifiable face**. If more than 70% of a human face is visible, we consider the frame to contain an identifiable face. Therefore, each frame is labeled as *yes*, *no*, and *no person*. A clip with more than one frame labeled as *yes* is labeled as *yes*, otherwise *no*. A clip with every frame labeled as *no person* is then labeled as *no person*.
- **Valuable property**. We only consider safe box, jewelry, watch, ring, and cash as valuable properties. Each frame is labeled as *yes*, *no*, and *no person*. We label clips with at least one frame labeled *yes* as *yes*, otherwise *no*. Clips with no person on any frame are labeled as *no person*.
- **Relationship**. We consider the relationship of all the people presented in the video. There are four types of labels for each frame: *intimate relationship*, *non-intimate relationship*, *only one person*, and *no person*. A video clip is labeled as *intimate relationship* if at least one frame of the clip is labeled as *intimate relationship* and the frames labeled as *intimate relationship* are no less than those labeled as *non-intimate relationship*. Otherwise, a clip is labeled as *non-intimate relationship* in a similar way. A clip with only one person presented is labeled as *only one person* and labeled as *no person* if there is no person existing in the clip.

Examples of the annotated frames in the dataset are demonstrated in Figure 3. Each frame was annotated by at least three of our authors and then cross-checked.

6 EFFECT OF RESOLUTION ON HUMAN’S RECOGNITION PERFORMANCE

After identifying the most important visual privacy features in the first study, we model the effect of image resolution on human performance in recognizing activities of daily living and visual privacy features. We describe the procedure and results in this section.

6.1 User Interface

We developed a web-based user interface as shown in Figure 4. Each problem set in the test for the participants includes one ADLs recognition task and four privacy feature recognition tasks including face, nudity, valuable property, and relationship. The user interface also includes attention-check questions in each test. Responses with incorrect answers to the attention check questions were treated as invalid. A starting page, shown before the testing procedure, introduces the purpose of the user study and requires the participant’s demographic information.

We sampled the image resolutions into seven values including 15×15, 20×20, 30×30, 50×50, 100×100, 160×160 and 240×240. We utilized a randomization strategy on the back-end server so that each participant could view 4 randomly chosen videos, with each video in a random resolution among these seven values. The same video did not appear twice to each participant. In addition, different clips from the same video did not appear to the same participant.

Please watch the video and answer the questions on the right.

1. What is the main character doing?
☐ Functional mobility. ☐ Feeding. ☐ Intimacy.
☐ Entertainment. ☐ Personal hygiene. ☐ No person.
2. Is there any identifiable faces in the video?
☐ Yes. ☐ No.
3. Is there any valuable properties (e.g., safe box, jewelry, watch, ring, cash, necklace)?
☐ Yes. ☐ No.
4. Is the main character fully or semi naked?
☐ Yes. ☐ No.
5. What's the relationship between the characters?
☐ Intimate relationship. ☐ Non-intimate relationship.
☐ Only one person. ☐ No person.
6. Please choose 3.
☐ 1. ☐ 2. ☐ 3. ☐ 4. ☐ 5. ☐ 6.

Figure 4: Example of the web-based user interface. Video clips of different resolutions is displayed on the left side. All tasks are listed on the right side of the web page.

6.2 Participant and Procedure

We recruited 240 participants (105 females, 135 males) with an average age of 22.23 (s.d. = 5.25, ranging from 18 to 30). All participants were required to have healthy eye conditions without any historical disease (e.g., color blindness) and use their laptop or desktop web browser to finish the whole test. The starting page of the web-based user interface introduced the purpose of the study. Participants were required to fill in their demographic information, including gender, age, and historical eye diseases. Following were two practice tests using two 240×240 resolution example videos excluded from the evaluation dataset. Finally, each participant finished the 28 rounds of the test. The user study lasted around 10 minutes. Each participant was offered a 5 USD gift card for compensation.

6.3 Results and Findings

In total, we obtained 6, 720 answer records, with 457 (6.80%) invalid due to the failure of the attention check questions. We utilized One-way ANOVA for the statistic analysis ($p < 0.05$) with independent-samples t-test ($p < 0.05$) as post-hoc analysis. We present our major results and findings below.

Low-resolution images are effective in preserving visual privacy but the effects are highly dependent on privacy features. Figure 5 shows the effect of image resolution on human recognition performance of ADLs, face, valuable property, nudity, and relationship. We observed the significant effect of image resolution on all visual privacy recognition tasks ($p < 0.001$). Further, there is no significant difference between resolutions of 160×160 and 240×240, indicating that resolutions above 160×160 pixels do not further contribute to visual privacy awareness statistically. However, the effect of the image resolution is highly task-dependent. Statistical analysis indicates that the type of privacy features has significant effects on the perception performance ($F_{3,25048} = 427.2$, $p < 0.001$). Specifically, pair-wise comparisons show that human eyes are more sensitive to nudity ($p < 0.001$) when the image resolution is below 50×50 pixels, followed by the relationship task. However, tasks including face identification and valuable property recognition require higher resolution images ($\geq 100 \times 100$ pixels) to achieve higher performance. For example, participants can only identify human faces with an accuracy of 79.2% when the resolution is 100×100 pixels. This is because both face identification and valuable property rely on detailed visual information. Therefore, a low-resolution image sensor can preserve but not fully protect visual privacy from the perspective of a human recognizer.

Lowering the image resolution has a significant negative impact on human recognition performance on ADLs. Results show that there is a statistically significant effect of changed resolution on human ADLs recognition performance ($F_{6,6256} = 278.0$,

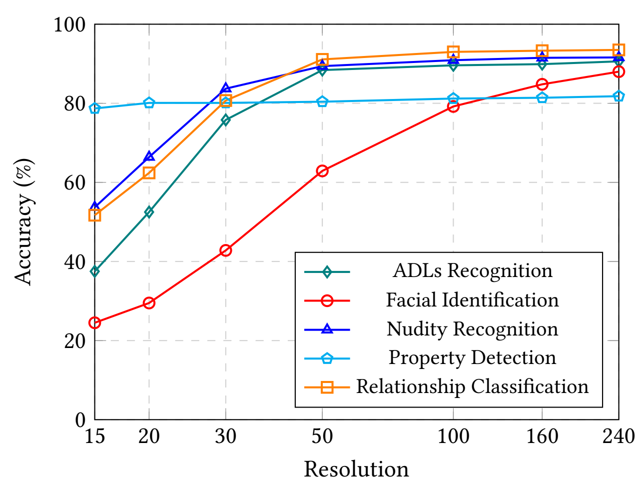


Figure 5: Humans' recognition performance on main activity and privacy feature recognition tasks.

$p < 0.001$). With resolutions lower than 30×30 pixels, human eyes can only recognize the ADLs with an accuracy below 75.8%. When the image resolution increases to 50×50 pixels, participants can recognize the activity with a fair accuracy — 88.4%. However, participants are aware of some privacy features at the resolution of 50×50. For example, they can recognize the relationship and nudity with an accuracy of 91.1% and 89.4%, respectively.

7 EFFECT OF RESOLUTION ON MACHINE'S RECOGNITION PERFORMANCE

In this section, we explore the effect of image resolution on machine's recognition performance of ADLs and visual privacy features. We adopted the open-access cutting-edge deep learning methods as the machine recognizer.

7.1 ADLs Recognition

7.1.1 Training and Evaluation Dataset. We applied data augmentation approaches to the training dataset in section 5, including horizontal flip, and Gaussian Noise, enlarging the dataset by four times. To fairly compare the recognition performance of the machine and the human, we utilized the same evaluation dataset in section 6.

7.1.2 Training and Evaluation Procedure. We utilized both convolutional neural networks and transformer-based models as our ADLs classifiers, including **ResNet50** [28], **Efficient Net** [53] and **Vision Transformer (ViT)** [20]. All the models used here were pretrained with ImageNet dataset [18] that output 1000 probabilistic values. In this experiment, we took every fame of the video clips in our dataset as the model input during our training, validating, and testing procedure. We first scaled the image of low resolution to 512×512 pixels to standardize the input of the model. Then, we fine-tuned the pretrained network using the training dataset with a certain resolution (r) in which the images were all at the resolution of $r \times r$. To transfer the pretrained network model to our application, we added an additional five-node fully connected layer at the end of the network. We used sigmoid as the activation function. Once we finished the training procedure, we evaluated the fine-tuned model using the evaluation dataset under the same image resolution (r). As we have described in section 5, we use the randomly chosen 5% of the total dataset as the validation dataset in our implementation. In order to avoid the over-fitting problem, we used the early stopping method. In other words, we will stop our training procedure when the accuracy on the validation dataset does not rise anymore for 5 successive epochs.

7.1.3 Result. Table 2 shows the effect of image resolution on machines' performance of the ADLs recognition task. Results indicate that **the machine outperforms the human regarding the ADLs recognition task on low-resolution images**. Vision Transformer can maintain an accuracy of 84.4% even when the image resolution is as low as 20×20. However, such a resolution is far from enough for humans to recognize ADLs at an ideal accuracy level. For resolutions above 100×100, both humans and machines can achieve a high accuracy above 90%. Such results show the possibility of constructing a range of image resolutions to preserve visual privacy without bearing great loss in ADLs recognition simultaneously.

7.2 Privacy Features Recognition

7.2.1 Facial Identification. We adopted **InsightFace** [19] for facial identification by testing whether the model can recognize human faces in certain areas of the frames. We used the pretrained **ArcFace** model for facial identification provided by InsightFace. Also, we checked every frame of the video clips in this experiment. The result is shown in Figure 6 as the teal line. Results from the ArcFace model indicate that even the state-of-art models cannot detect any human faces below 50×50 pixels. However, as the resolution increases from 100×100 to 240×240 pixels, machine's facial identification performance significantly increases from 71.0% to 100.0%. Such results imply that identifiable faces can be preserved well against the machine attacker when the image resolution is below 50×50 pixels.

7.2.2 Nudity Recognition. We adopted the pretrained **NudeNet** [2] for binary nudity recognition. This model was trained to detect nude parts of the human body in images. Here we utilized the classifier model to help us make a distinction between safe and unsafe images. We report the result of NudeNet as the orange line in Figure 6. The precision and recall of NudeNet also reveal that it cannot identify any nude parts below the resolution of 30×30 pixels. Under the resolution of 100×100 pixels, NudeNet can recognize frames containing nude parts with an accuracy of 88.0%. Therefore, we conclude that resolutions below 30×30 pixels can effectively preserve the nudity privacy feature.

7.2.3 Property and Object Detection. We adopted **DETR** [9] pretrained on the COCO dataset [3] for property and object detection. Considering the availability of pretrained object detection models, we used the detection performance of DETR on COCO objects as an estimation of machine's recognition performance on valuable properties. We manually annotated the objects which belong to the COCO classes in each frame as ground truth. In our implementation, we first resized videos of different resolutions up to 240×240 pixels. Then we kept bounding boxes with a confidence level above a pre-set threshold (e.g., 0.75) as a result of the model. To evaluate the model performance under different resolutions, we compared the objects detected by the model and the ground truth of each frame one by one to calculate the recognition accuracy. The purple line in Figure 6 shows the recognition accuracy of DETR. Results show that below the resolution of 50×50, DETR fails to detect any object. On images with a resolution of 100×100, DETR can achieve an accuracy of 72.0%. Under the resolution of 160×160, large objects such as the main character can be detected precisely with an overall accuracy of 77.0%. Under the resolution of 240×240, the object detection results are more accurate, and small targets such as bottles and cups can be detected, too. Therefore, DETR can finally achieve an accuracy of 82.0%.

7.2.4 Relationship Classification. We adopted a pretrained cutting-edge social relationship classification model **GRM** [55] on the evaluation dataset with different image resolutions. This model utilized a node message propagation mechanism and a graph attention mechanism to explore the interaction between the person pair of interest and contextual objects. The prerequisite to inferring the

[2]Software DOI: 10.5281/zenodo.3584720
[3]https://cocodataset.org/

Table 2: The ADLs recognition performance of ViT, ResNet50, and EfficientNet, compared with human.

Resolution	Human	Machine		
		ViT	ResNet50	EfficientNet
15×15	37.5%	81.0%	63.9%	52.9%
20×20	52.5%	84.4%	66.3%	63.5%
30×30	75.8%	89.8%	75.1%	68.0%
50×50	88.4%	90.7%	80.5%	74.6%
100×100	89.6%	92.2%	81.5%	75.1%
160×160	89.9%	93.2%	82.0%	80.0%
240×240	90.6%	94.6%	88.8%	83.9%

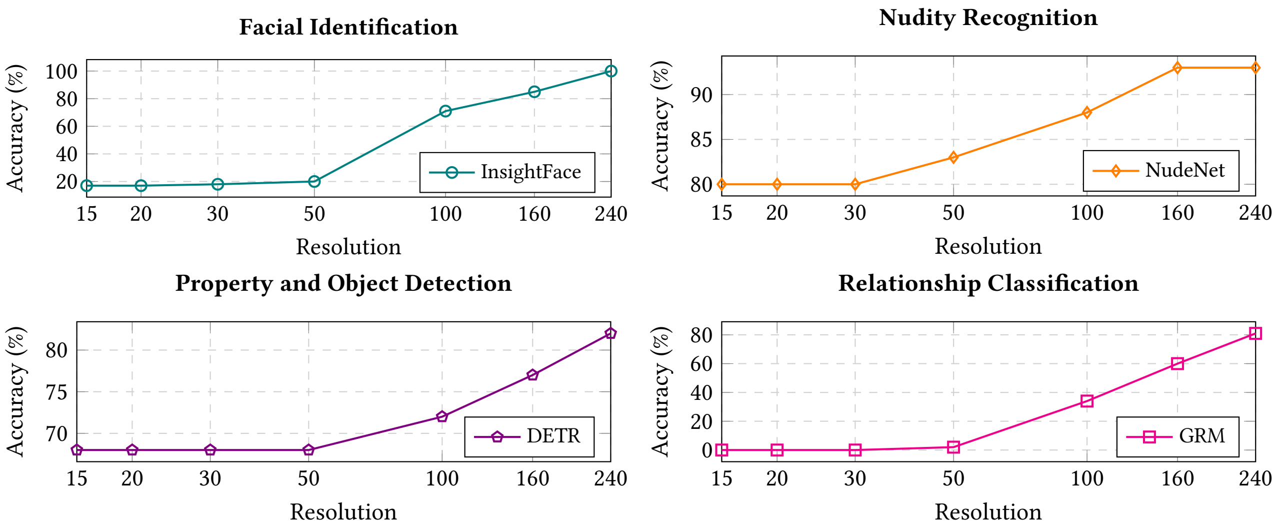


Figure 6: Machines' recognition performance of privacy features.

relationship between people is to obtain the context information using the object detection model. In our implementation, we resized the raw video of different resolutions to 240×240 pixels. Then, we annotated the bounding boxes and classes of different objects using DETR and labeled the bounding boxes of the person pair whose social relationship we wanted to examine. The model took every frame of the raw videos and objects list as input and generated the classification result as output.

The accuracy result we reported as the magenta line in Figure 6 describes the performance of the GRM model on the four-classes social relationship recognition task including *intimate relationship*, *non-intimate relationship*, *no relationship*, and *no person*. As is shown, the GRM model can detect nothing and will classify any input image as the *no person* type under resolutions below 30×30 pixels. Our results here also proved that a low resolution below 30×30 is sufficient to preserve the privacy of social relationships against the cutting-edge machine recognition method. When the resolution is 100×100 pixels, GRM can recognize social relationships in the video with an accuracy of 34.1%. For resolutions of 160×160 pixels and 240×240 pixels, GRM can achieve an accuracy of 60.9% and 80.5%, respectively.

8 JUSTIFY THE INFLUENCE OF IMAGE SUPER-RESOLUTION

Image super-resolution techniques were proposed by researchers to reconstruct a high-resolution image from a low-resolution image [38, 54]. In this section, we justice whether cutting-edge super-resolution techniques influence our results and findings regarding the effects of low resolution on activity recognition and privacy awareness through a user study.

8.1 User Study Procedure and Participant

We adopted one of the cutting-edge image super-resolution methods SwinIR [37] based on Transformer architectures as well as the traditional bicubic method to upscale the videos in our evaluation dataset by four times. Three examples of super-resolution processed videos are shown in Figure 7.

We adopted a similar web-based interface as Figure 4 shows except for changing the attention check question to addition and subtraction test. In this study, we first introduced the purpose and the procedure of our study. Then each participant took 8 trials with each trial having one test on the raw video and one test with videos after super-resolution. In each trial, we first presented each participant with a randomly-chosen raw video in the evaluation dataset and asked them to answer questions of ADLs and privacy features recognition as illustrated in Figure 4. The raw video's resolution was set to a random value among 15×15, 20×20, 30×30, 50×50, 100×100, 160×160 and 240×240. Then, we presented them with the super-resolution videos together with raw videos simultaneously and asked them to answer the same questions. To avoid cross effects between videos under different resolutions, the same raw video did not appear twice to each participant. Further, we also ensured that the participants in this study were different from those who participated in the previous studies.

We recruited 306 participants (123 females, 183 males) with an average age of 21.76 (s.d. = 4.56). The user study lasted around 10 minutes. Each participant was offered a 5 USD gift card for compensation.

Figure 7: Examples of the effect of super resolution on videos of low resolutions including 15×15, 20×20, and 30×30.

8.2 Results and Findings

In total, we received 4,896 test records with 273 (5.57%) of them failed the attention check. Table 3 and Table 4 show the comparison of participants' overall recognition accuracy with or without super-resolution. Results indicate that participants performed better on super-resolution videos than on raw videos. Statistical analysis suggests that when image resolution is below 20×20 pixels, super-resolution techniques can significantly improve human recognition performance on both activity recognition and privacy recognition tasks. But it is worth noting that the improvement in recognition performance brought about by super-resolution technology is still less than that brought about by increasing the resolution itself. Such a finding reveals that super-resolution techniques do not provide enough additional information for humans to enhance their perception ability in both activity recognition and visual privacy awareness tasks.

In terms of the impact of the super-resolution technique on the machine's recognition performance, researchers have proved that super-resolution can slightly facilitate vision-based recognition task such as activity recognition [17, 29], object and text recognition [38, 59]. However, the influence of the super-resolution technique is very limited. The results are still significantly inferior to that with the original high-resolution images [14].

In conclusion, the additional visual information introduced by the image super-resolution technique is insufficient to overcome the effect of resolution on the recognition performance of humans and machines. Therefore, we believe that the effects of image resolution on human (section 6) and the machine's (section 7) ADLs and visual privacy recognition performance are robust against image super-resolution techniques.

Table 3: The statistic of the overall accuracy on main activity recognition with or without super resolution conditions. $p < 0.05$ indicates a significant difference between with or without super resolution conditions.

Resolution	Before		After		Significance
	avg.	std.	avg.	std.	
15×15	0.386	0.487	0.452	0.498	$p < 0.001$
20×20	0.593	0.491	0.706	0.456	$p = 0.002$
30×30	0.803	0.397	0.845	0.362	$p = 0.149$
50×50	0.891	0.310	0.893	0.308	$p = 0.932$
100×100	0.846	0.360	0.898	0.302	$p = 0.046$
160×160	0.899	0.301	0.908	0.289	$p = 0.701$
240×240	0.908	0.289	0.927	0.260	$p = 0.386$

Table 4: The statistic of the overall accuracy on privacy features recognition with or without super resolution conditions. $p < 0.05$ indicates a significant difference between with or without super resolution conditions.

Resolution	Before		After		Significance
	avg.	std.	avg.	std.	
15×15	0.558	0.497	0.602	0.476	$p < 0.001$
20×20	0.673	0.469	0.736	0.440	$p < 0.001$
30×30	0.793	0.404	0.823	0.381	$p = 0.038$
50×50	0.851	0.356	0.866	0.340	$p = 0.276$
100×100	0.895	0.305	0.906	0.291	$p = 0.359$
160×160	0.905	0.292	0.913	0.280	$p = 0.488$
240×240	0.921	0.268	0.925	0.263	$p = 0.766$

9 MODELING THE TRADE-OFF OF PRIVACY PRESERVATION AND ACTIVITY RECOGNITION

In this paper, our goal is to present a method to model the trade-off between privacy preservation and machine recognition. We have obtained the estimation results of the main components in Equation 1. In this section, we take all these results into consideration and explain how we can utilize them to model the trade-off between privacy preservation and machine recognition. Based on our modeling results, we further present how to apply our model to applications.

9.1 Build the Model Using the Parameters from the Studies

To summarize, we have investigated users' perceived importance of different privacy features under high or low image resolutions in section 4. We chose users' rating of these privacy features under high-resolution image condition as the importance weight ω_i in the model, which was shown in Table 1. Next, we examined both human's and the machine's recognition performance under varying resolutions in order to obtain an approximation of the evaluation function L_T and L_P in our formulation. In realistic environments, intelligent applications may rely on either humans or machines to obtain private information from raw images. Therefore, we take both human and machine recognizers into consideration to preserve privacy features in a comprehensive way. For the main recognition task T, which is activity recognition in our implementation, the Vision Transformer outperforms all other models including humans even on extremely low-resolution videos from the dataset. Therefore, we choose the Vision Transformer as our final recognition function f_T and the evaluation results of the Vision Transformer L_T have been demonstrated in Table 2. For each privacy feature P_i including nudity, identifiable face, valuable property, and relationship, we found that humans are generally more effective recognizers compared with machines, especially on ultra-low-resolution videos from the dataset. Therefore, we consider humans as the final f_{P_i} in our calculation. The evaluation results of each L_{P_i} we are going to use has been depicted in Figure 5.

9.2 Calculating the Objective Function

Based on the results of L_T, L_{P_i}, and ω_i we have discussed above, we can calculate the objective function $S(r)$ in Equation 1 for each resolution we have sampled. Figure 8 illustrates how the values of our objective function $S(r)$ change with resolutions r. The scaling factor λ in our formulation indicates the sensitivity ratio of privacy preservation over activity recognition which can be flexibly adjusted according to the deployment environment or user experience. Here we have only shown the cases for three different lambda values, including 0.75, 1.00, and 1.25.

As is demonstrated in Figure 8, the value of the objective function $S(r)$ shows a trend of first increasing and then decreasing with the increase of resolution r. For the case where lambda is 1.00, the objective function takes its maximum value at a resolution between 20×20 and 30×30, which indicates a proper resolution for balancing privacy preservation and activity recognition. Such an image resolution value can be easily extended to a certain image resolution range where the trade-off result is also acceptable. However, the objective function takes a low value when the resolution is too low (e.g., 15×15) or too high (e.g., 240×240). The reason behind this is also consistent with our expectations. When the image resolution is too low, although the privacy features can be better preserved, the machine's ADLs recognition performance is far from satisfactory. On the contrary, high image resolution may greatly increase the risk of privacy feature leakage except for improving ADLs recognition performance.

Here we also noticed that as the scaling factor λ increases, the maximum point of the objective function is also shifted to the left in Figure 8. Such a finding shows that a lower resolution of the image sensor is required if users are more concerned with privacy preservation compared with activity recognition performance.

9.3 Applying the Model and the Modeling Method to Applications

In this section, we present how to apply our method and model to privacy-preserving machine recognition applications.

9.3.1 Deployment to a Real Scenario Application. When deploying a real scenario application based on our method, one can install an ultra-low-resolution (e.g., 20×20 pixels) image sensor with an edge computer running a machine learning method for ADLs recognition at home. To apply our framework for quantifying the trade-off between privacy preservation and activity recognition, one first needs to determine the sensitivity indicator λ in Equation 1, which is closely related to deployment environment and user experience. In our ADLs recognition example, the bathroom is a more visual privacy-sensitive location than the kitchen. Therefore, we would expect the image sensor in the bathroom having a lower resolution to preserve more visual privacy. Second, with the development of computer vision technologies, the performance of machine recognition on both activities and privacy features will exceed the current results stated in this paper. Future designers just need to fine-tune the results of the evaluation function L_T and L_P by selecting better recognizers f_T and f_P to consider the results of these technological advances. Third, one can leverage activities' probability distribution regarding the different environments in a home environment which may have an effect on the results of the evaluation function L_T and L_P in our formulation. For instance, personal and toilet hygiene is highly possible to happen in the restroom, while feeding is highly possible to occur in the kitchen. Future designers need to modify their training and evaluation data set according to the probability distribution of activities of daily living (ADLs) in different scenarios.

9.3.2 Generalization to Other Applications. For other computer vision based applications in a real scenario, we believe that our pipeline and method can be easily adapted. For instance, using an always-on low-resolution camera on AR glasses for activity recognition, or using a low-resolution smartphone camera for hand gesture recognition, etc. Even though different applications have their own usage scenarios with different visual privacy features, our method's key idea and basic framework can still be used efficiently. Although low-resolution image sensors can preserve visual privacy from the hardware level, deploying the hardware itself costs a large of human labor and money. Instead of purchasing the low-resolution image sensor, we can simply update the firmware to limit the camera's resolution, turning them to low-resolution image sensors. Further, we can attach an additional layer or lens on top of available commercial RGB cameras. For instance, we can add a piece of frosted glass, or a lens built for the Passive Infrared (PIR) motion sensor to the commodity cameras [4]. Most of these camera

[4]https://en.wikipedia.org/wiki/Passive_infrared_sensor

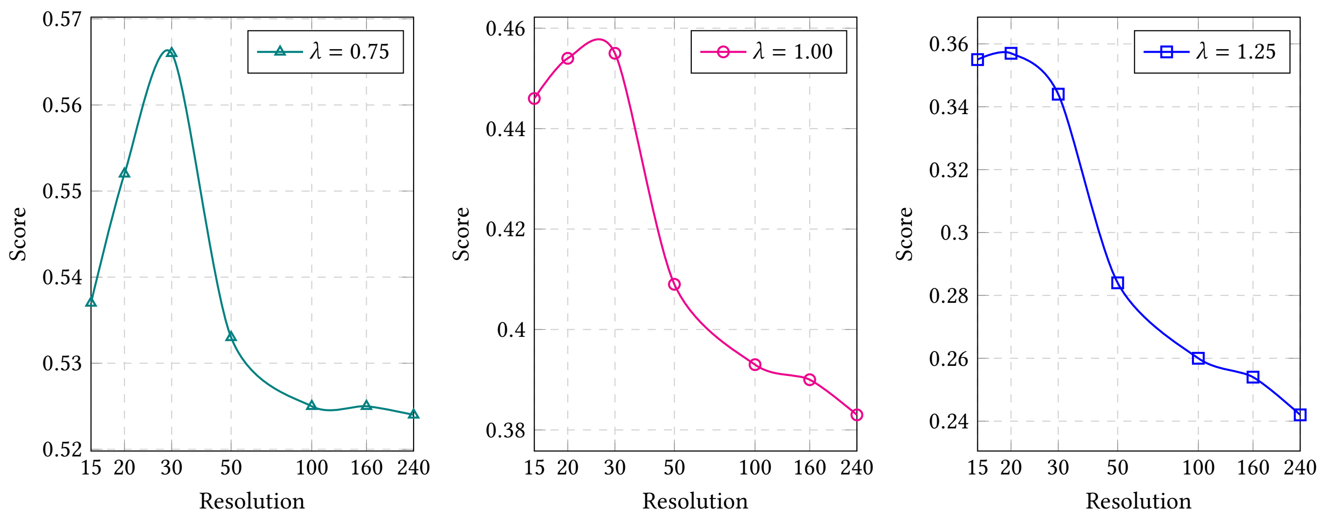


Figure 8: Depicting the objective function based on the results of both humans' and machines' recognition performance.

system parameter selection issues can be discussed and solved in a more generalized way of our methods.

10 LIMITATIONS AND FUTURE WORK

Our work is targeted at modeling the trade-off between visual privacy and the core machine recognition task, e.g., activity recognition in our case. Our purpose is to inspire future work to explore more quantitative methods for privacy-preserving applications. Therefore, future designers can apply or adapt these models according to their applications to preserve users' privacy as much as possible. However, there do exist several limitations of our work and we describe them below.

Utilizing Multimodal Information. We acknowledge that an image sensor deployed at home can only collect images at a fixed position, distance, and field of view after installation. Only with the single modality data captured by images sensors, both machine's and human's recognition performance can be easily affected by the aforementioned factors. We also acknowledge that we didn't take multimodal data, for example, audio data into consideration. Prior works have proved the effectiveness of leveraging multimodal data in activity recognition. With multimodal information, we can alleviate existing algorithms' dependence on images, thus allowing for a lower resolution of image sensors. We expect future research can investigate how the modeling results of the trade-off between privacy preservation and activity recognition can be changed by multimodal information.

Privacy Preserving Methods. In this work, we only use pixelization filters as the privacy-preserving method for the main task. The advantages of using low-resolution images have already been discussed in prior works. Nevertheless, we have to admit that researchers have shown that low resolution alone does not provide enough privacy guarantees. McPherson et al. found that obfuscated images contain enough information correlated with the obfuscated content to enable accurate reconstruction of the latter [39]. Although we have compared the privacy recognition performance of state-of-art machine learning algorithms on low-resolution images, we believe that our evaluation results on low-resolution images leave much room for discussion. We expect future research can explore the effect of more privacy-preserving methods on the trade-off between privacy preservation and activity recognition.

User Survey on Importance of Visual Privacy Features. We acknowledge that our user study in section 4 aims to assess users' perceived importance of visual privacy features. We didn't limit participants' culture, age, gender, or technical backgrounds. However, there are many other factors that may affect participants' perception of privacy. For example, researchers have found that users on Amazon Mechanical Turk, where our participants were from, tend to be more privacy conscious [31, 46], thus are not representative of the general population all over the world. It is also undeniable that the perception of privacy varies substantially across cultures, societies, times, and locations [2–4, 12, 32, 44, 51]. Therefore, our estimation of the perceived importance (ω in our formulation) of privacy features obtained through our user studies is possibly not applicable to populations in different cultural contexts across the world. However, the framework proposed in this paper is meant to inspire future researchers to consider humans' assessments of the importance of different visual privacy features. We expect that there will be more independent works to explore the influence of other factors on humans' perception of privacy.

11 CONCLUSION

Using the at-home activity of daily livings (ADLs) as the scenario, this paper models the trade-off of visual privacy preservation and activity recognition over image resolution. To achieve this purpose, we first conducted a user survey to obtain the most important visual privacy features, including nudity, identifiable face, valuable property, and relationship. Then, using the PA-HMDB51 dataset, which contains videos from realistic environments, we quantified the effect of image resolution on the human's performance on ADLs recognition and visual privacy awareness tasks through a user study. We further modeled the impact of image resolution on the machine's ability to recognize ADLs and visual privacy features using cutting-edge machine learning methods. Finally, we proposed a method with adjustable parameters to model the trade-off of privacy-preserving ADLs recognition using low-resolution images. Using this method, we can calculate an optimal range of image resolution for visual privacy preserving ADLs recognition. We envision that our method can inspire other vision-based systems that require balancing privacy awareness and machine recognition performance.

ACKNOWLEDGMENTS

This work is supported by the Natural Science Foundation of China (NSFC) under Grant No. 62002198 and No. 62132010, Tsinghua University Initiative Scientific Research Program, Beijing Key Lab of Networked Multimedia, Institute for Artificial Intelligence, Tsinghua University, and Beijing National Research Center for Information Science and Technology (BNRist).

REFERENCES

[1] Nitin Agrawal, Reuben Binns, Max Van Kleek, Kim Laine, and Nigel Shadbolt. 2021. Exploring Design and Governance Challenges in the Development of Privacy-Preserving Computation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 68, 13 pages. https://doi.org/10.1145/3411764.3445677

[2] Syed Ishtiaque Ahmed, Md. Romael Haque, Jay Chen, and Nicola Dell. 2017. Digital Privacy Challenges with Shared Mobile Phone Use in Bangladesh. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 17 (dec 2017), 20 pages. https://doi.org/10.1145/3134652

[3] Syed Ishtiaque Ahmed, Md. Romael Haque, Shion Guha, Md. Rashidujjaman Rifat, and Nicola Dell. 2017. Privacy, Security, and Surveillance in the Global South: A Study of Biometric Mobile SIM Registration in Bangladesh. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 906–918. https://doi.org/10.1145/3025453.3025961

[4] Wael S Albayaydh and Ivan Flechais. 2022. Exploring Bystanders' Privacy Concerns with Smart Homes in Jordan. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 446, 24 pages. https://doi.org/10.1145/3491102.3502097

[5] 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. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 72 (sep 2019), 29 pages. https://doi.org/10.1145/3351230

[6] S. Applin. 2017. Amazon's Echo Look: Harnessing the Power of Machine Learning or Subtle Exploitation of Human Vulnerability? *IEEE Consumer Electronics Magazine* 6, 4 (Oct 2017), 125–127. https://doi.org/10.1109/MCE.2017.2714273

[7] Terrance E. Boult. 2005. PICO: Privacy through Invertible Cryptographic Obscuration. *Computer Vision for Interactive and Intelligent Environment (CVIIE'05)* (2005), 27–38.

[8] Michael Boyle, Christopher Edwards, and Saul Greenberg. 2000. The Effects of Filtered Video on Awareness and Privacy. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work* (Philadelphia, Pennsylvania, USA) *(CSCW '00)*. ACM, New York, NY, USA, 1–10. https://doi.org/10.1145/358916.358935

[9] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. 2020. End-to-End Object Detection with Transformers. arXiv:arXiv:2005.12872

[10] Jiawei Chen, Jonathan Wu, Kristi Richter, Janusz Konrad, and Prakash Ishwar. 2016. Estimating head pose orientation using extremely low resolution images. In *2016 IEEE Southwest Symposium on Image Analysis and Interpretation, SSIAI 2016, Santa Fe, NM, USA, March 6-8, 2016*. IEEE Computer Society, 65–68. https://doi.org/10.1109/SSIAI.2016.7459176

[11] U.S. Congress. 1974. The Privacy Act of 1974. Public Low, 88.

[12] Andy Crabtree, Peter Tolmie, and Will Knight. 2017. Repacking 'Privacy' for a Networked World. *Comput. Supported Coop. Work* 26, 4–6 (dec 2017), 453–488. https://doi.org/10.1007/s10606-017-9276-y

[13] R. Jason Cronk. 2016. What Is Haptics? https://enterprivacy.com/2017/03/01/categories-of-personal-information/.

[14] Dengxin Dai, Yujian Wang, Yuhua Chen, and Luc Van Gool. 2015. Is Image Super-resolution Helpful for Other Vision Tasks? arXiv:arXiv:1509.07009

[15] Ji Dai, Jonathan Wu, Behrouz Saghafi, Janusz Konrad, and Prakash Ishwar. 2015. Towards privacy-preserving activity recognition using extremely low temporal and spatial resolution cameras. In *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2015, Boston, MA, USA, June 7-12, 2015*. IEEE Computer Society, 68–76. https://doi.org/10.1109/CVPRW.2015.7301356

[16] C. Debes, A. Merentitis, S. Sukhanov, M. Niessen, N. Frangiadakis, and A. Bauer. 2016. Monitoring Activities of Daily Living in Smart Homes: Understanding human behavior. *IEEE Signal Processing Magazine* 33, 2 (March 2016), 81–94. https://doi.org/10.1109/MSP.2015.2503881

[17] Ugur Demir, Yogesh S. Rawat, and Mubarak Shah. 2020. TinyVIRAT: Low-resolution Video Action Recognition. In *25th International Conference on Pattern Recognition, ICPR 2020, Virtual Event / Milan, Italy, January 10-15, 2021*. IEEE, 7387–7394. https://doi.org/10.1109/ICPR48806.2021.9412541

[18] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA*. IEEE Computer Society, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

[19] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. 2019. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*. Computer Vision Foundation / IEEE, 4690–4699. https://doi.org/10.1109/CVPR.2019.00482

[20] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net. https://openreview.net/forum?id=YicbFdNTTy

[21] Alan F. Westin. 1968. Privacy and Freedom. *Michigan Law Review* 66 (03 1968). https://doi.org/10.2307/1287193

[22] Andrea Frome, German Cheung, Ahmad Abdulkader, Marco Zennaro, Bo Wu, Alessandro Bissacco, Hartwig Adam, Hartmut Neven, and Luc Vincent. 2009. Large-scale privacy protection in Google Street View. In *IEEE 12th International Conference on Computer Vision, ICCV 2009, Kyoto, Japan, September 27 - October 4, 2009*. IEEE Computer Society, 2373–2380. https://doi.org/10.1109/ICCV.2009.5459413

[23] Ziqi Gao, Jianguo Chen, Junliang Xing, Shwetak Patel, Yuanchun Shi, Xin Liu, and Yuntao Wang. 2022. MMTSA: Multimodal Temporal Segment Attention Network for Efficient Human Activity Recognition. https://doi.org/10.48550/ARXIV.2210.09222

[24] Ralph Gross, Latanya Sweeney, Jeffrey F. Cohn, Fernando De la Torre, and Simon Baker. 2009. Face De-identification. In *Protecting Privacy in Video Surveillance*, Andrew W. Senior (Ed.). Springer, 129–146. https://doi.org/10.1007/978-1-84882-301-3_8

[25] Jihun Hamm. 2017. Minimax Filter: Learning to Preserve Privacy from Inference Attacks. *J. Mach. Learn. Res.* 18 (2017), 129:1–129:31. http://jmlr.org/papers/v18/16-501.html

[26] L D Harmon and B. Julesz. 1973. Masking in visual recognition: effects of two-dimensional filtered noise. *Science* 180, 4091 (1973), 1194–1197. https://doi.org/10.1126/science.180.4091.1194

[27] Rakibul Hasan, Eman Hassan, Yifang Li, Kelly Caine, David J. Crandall, Roberto Hoyle, and Apu Kapadia. 2018. Viewer Experience of Obscuring Scene Elements in Photos to Enhance Privacy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173621

[28] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep Residual Learning for Image Recognition. arXiv:1512.03385 [cs.CV]

[29] Mingzheng Hou, Song Liu, Jiliu Zhou, Yi Zhang, and Ziliang Feng. 2021. Extreme Low-Resolution Activity Recognition Using a Super-Resolution-Oriented Generative Adversarial Network. *Micromachines* 12, 6 (2021), 670.

[30] Panagiotis Ilia, Iasonas Polakis, Elias Athanasopoulos, Federico Maggi, and Sotiris Ioannidis. 2015. Face/Off: Preventing Privacy Leakage From Photos in Social Networks. In *Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security* (Denver, Colorado, USA) *(CCS '15)*. ACM, New York, NY, USA, 781–792. https://doi.org/10.1145/2810103.2813603

[31] Ruogu Kang, Stephanie Brown, Laura Dabbish, and Sara Kiesler. 2014. Privacy Attitudes of Mechanical Turk Workers and the U.S. Public. In *Proceedings of the Tenth USENIX Conference on Usable Privacy and Security* (Menlo Park, CA)

(SOUPS '14). USENIX Association, USA, 37–49.

[32] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara B. Kiesler. 2015. "My Data Just Goes Everywhere: " User Mental Models of the Internet and Implications for Privacy and Security. In *Eleventh Symposium On Usable Privacy and Security, SOUPS 2015, Ottawa, Canada, July 22-24, 2015*, Lorrie Faith Cranor, Robert Biddle, and Sunny Consolvo (Eds.). USENIX Association, 39–52. https://www.usenix.org/conference/soups2015/proceedings/presentation/kang

[33] Marc Langheinrich. 2001. Privacy by Design — Principles of Privacy-Aware Ubiquitous Systems. In *Ubicomp 2001: Ubiquitous Computing*, Gregory D. Abowd, Barry Brumitt, and Steven Shafer (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 273–291.

[34] Walter S. Lasecki, Mitchell Gordon, Winnie Leung, Ellen Lim, Jeffrey P. Bigham, and Steven P. Dow. 2015. Exploring Privacy and Accuracy Trade-Offs in Crowdsourced Behavioral Video Coding. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 1945–1954. https://doi.org/10.1145/2702123.2702605

[35] M. Powell Lawton and Elaine M. Brody. 1969. Assessment of Older People: Self-Maintaining and Instrumental Activities of Daily Living. *The Gerontologist* 9, 3_Part_1 (10 1969), 179–186. https://doi.org/10.1093/geront/9.3_Part_1.179 arXiv:http://oup.prod.sis.lan/gerontologist/article-pdf/9/3_Part_1/179/1466322/9-3_Part_1-179.pdf

[36] Yifang Li, Wyatt Troutman, Bart P. Knijnenburg, and Kelly Caine. 2018. Human Perceptions of Sensitive Content in Photos. In *2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2018, Salt Lake City, UT, USA, June 18-22, 2018*. Computer Vision Foundation / IEEE Computer Society, 1590–1596. https://doi.org/10.1109/CVPRW.2018.00209

[37] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. 2021. SwinIR: Image Restoration Using Swin Transformer. arXiv:arXiv:2108.10257

[38] Xin Liu, Yuang Li, Josh Fromm, Yuntao Wang, Ziheng Jiang, Alex Mariakakis, and Shwetak Patel. 2021. SplitSR: An End-to-End Approach to Super-Resolution on Mobile Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 1, Article 25 (mar 2021), 20 pages. https://doi.org/10.1145/3448104

[39] Richard McPherson, Reza Shokri, and Vitaly Shmatikov. 2016. Defeating Image Obfuscation with Deep Learning. *CoRR* abs/1609.00408 (2016). arXiv:1609.00408 http://arxiv.org/abs/1609.00408

[40] Nobuhiro Miyazaki, Kentaro Tsuji, Mingxie Zheng, Moyuri Nakashima, Yuji Matsuda, and Eigo Segawa. 2015. Privacy-conscious human detection using low-resolution video. In *3rd IAPR Asian Conference on Pattern Recognition, ACPR 2015, Kuala Lumpur, Malaysia, November 3-6, 2015*. IEEE, 326–330. https://doi.org/10.1109/ACPR.2015.7486519

[41] Tribhuvanesh Orekondy, Mario Fritz, and Bernt Schiele. 2018. Connecting Pixels to Privacy and Utility: Automatic Redaction of Private Information in Images. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*. Computer Vision Foundation / IEEE Computer Society, 8466–8475. https://doi.org/10.1109/CVPR.2018.00883

[42] Tribhuvanesh Orekondy, Bernt Schiele, and Mario Fritz. 2017. Towards a Visual Privacy Advisor: Understanding and Predicting Privacy Risks in Images. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*. IEEE Computer Society, 3706–3715. https://doi.org/10.1109/ICCV.2017.398

[43] José Ramón Padilla-López, Alexandros Andre Chaaraoui, and Francisco Flórez-Revuelta. 2015. Visual privacy protection methods: A survey. *Expert Systems with Applications* 42, 9 (2015), 4177 – 4195. https://doi.org/10.1016/j.eswa.2015.01.041

[44] Leysia Palen and Paul Dourish. 2003. Unpacking "Privacy" for a Networked World. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 129–136. https://doi.org/10.1145/642611.642635

[45] Nisarg Raval, Ashwin Machanavajjhala, and Landon P. Cox. 2017. Protecting Visual Secrets Using Adversarial Nets. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2017, Honolulu, HI, USA, July 21-26, 2017*. IEEE Computer Society, 1329–1332. https://doi.org/10.1109/CVPRW.2017.174

[46] Joel Ross, Lilly Irani, M. Six Silberman, Andrew Zaldivar, and Bill Tomlinson. 2010. Who Are the Crowdworkers? Shifting Demographics in Mechanical Turk. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI EA '10)*. Association for Computing Machinery, New York, NY, USA, 2863–2872. https://doi.org/10.1145/1753846.1753873

[47] Michael S. Ryoo, Kiyoon Kim, and Hyun Jong Yang. 2018. Extreme Low Resolution Activity Recognition With Multi-Siamese Embedding Learning. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, Sheila A. McIlraith and Kilian Q. Weinberger (Eds.). AAAI Press, 7315–7322. https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16790

[48] Michael S. Ryoo, Brandon Rothrock, and Charles Fleming. 2016. Privacy-Preserving Egocentric Activity Recognition from Extreme Low Resolution. *CoRR* abs/1604.03196 (2016). arXiv:1604.03196 http://arxiv.org/abs/1604.03196

[49] Michael S. Ryoo, Brandon Rothrock, Charles Fleming, and Hyun Jong Yang. 2017. Privacy-Preserving Human Activity Recognition from Extreme Low Resolution. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, Satinder Singh and Shaul Markovitch (Eds.). AAAI Press, 4255–4262. http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14847

[50] 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. https://doi.org/10.1007/s11042-012-1207-9

[51] Nithya Sambasivan, Garen Checkley, Amna Batool, Nova Ahmed, David Nemer, Laura Sanely Gaytán-Lugo, Tara Matthews, Sunny Consolvo, and Elizabeth Churchil. 2018. "Privacy is Not for Me, It's for Those Rich Women": Performative Privacy Practices on Mobile Phones by Women in South Asia. In *Proceedings of the Fourteenth USENIX Conference on Usable Privacy and Security* (Baltimore, MD, USA) *(SOUPS '18)*. USENIX Association, USA, 127–142.

[52] Jure Sokolic, Qiang Qiu, Miguel R. D. Rodrigues, and Guillermo Sapiro. 2017. Learning to Succeed while Teaching to Fail: Privacy in Closed Machine Learning Systems. arXiv:arXiv:1705.08197

[53] Mingxing Tan and Quoc V. Le. 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA (Proceedings of Machine Learning Research, Vol. 97)*, Kamalika Chaudhuri and Ruslan Salakhutdinov (Eds.). PMLR, 6105–6114. http://proceedings.mlr.press/v97/tan19a.html

[54] Zhihao Wang, Jian Chen, and Steven C. H. Hoi. 2021. Deep Learning for Image Super-Resolution: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 10 (2021), 3365–3387.

[55] Zhouxia Wang, Tianshui Chen, Jimmy S. J. Ren, Weihao Yu, Hui Cheng, and Liang Lin. 2018. Deep Reasoning with Knowledge Graph for Social Relationship Understanding. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, Jérôme Lang (Ed.). ijcai.org, 1021–1028. https://doi.org/10.24963/ijcai.2018/142

[56] WikiPedia. 2019. Privacy Law. https://en.wikipedia.org/wiki/Privacy_law

[57] Zhenyu Wu, Haotao Wang, Zhaowen Wang, Hailin Jin, and Zhangyang Wang. 2019. Privacy-Preserving Deep Action Recognition: An Adversarial Learning Framework and A New Dataset. arXiv:arXiv:1906.05675

[58] Zhenyu Wu, Zhangyang Wang, Zhaowen Wang, and Hailin Jin. 2018. Towards Privacy-Preserving Visual Recognition via Adversarial Training: A Pilot Study. In *Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XVI* (Munich, Germany). Springer-Verlag, Berlin, Heidelberg, 627–645. https://doi.org/10.1007/978-3-030-01270-0_37

[59] Yue Xi, Jiangbin Zheng, Wenjing Jia, Xiangjian He, Hanhui Li, Zhuqiang Ren, and Kin-Man Lam. 2020. See clearly in the distance: Representation learning GAN for low resolution object recognition. *IEEE Access* 8 (2020), 53203–53214.

[60] Mingze Xu, Aidean Sharghi, Xin Chen, and David J. Crandall. 2018. Fully-Coupled Two-Stream Spatiotemporal Networks for Extremely Low Resolution Action Recognition. *CoRR* abs/1801.03983 (2018), 1607–1615. arXiv:1801.03983 http://arxiv.org/abs/1801.03983

[61] Xiangyu Xu, Hao Chen, Francesc Moreno-Noguer, László A. Jeni, and Fernando De la Torre. 2020. 3D Human Shape and Pose from a Single Low-Resolution Image with Self-Supervised Learning. In *Computer Vision – ECCV 2020*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer International Publishing, Cham, 284–300.

[62] Andrew Yip and Pawan Sinha. 2010. Role of color in face recognition. *Journal of Vision* 2 (11 2010). https://doi.org/10.1167/2.7.596