Leveraging Content Sensitiveness and User Trustworthiness to Recommend Fine-Grained Privacy Settings for Social Image Sharing

Jun Yu, *Member, IEEE*, Zhenzhong Kuang, Baopeng Zhang, Wei Zhang, Dan Lin, and Jianping Fan

***Abstract*— To configure successful privacy settings for social image sharing, two issues are inseparable: 1) content sensitiveness of the images being shared; and 2) trustworthiness of the users being granted to see the images. This paper aims to consider these two inseparable issues simultaneously to recommend fine-grained privacy settings for social image sharing. For achieving more compact representation of image content sensitiveness (privacy), two approaches are developed: 1) a deep network is adapted to extract 1024-D discriminative deep features; and 2) a deep multiple instance learning algorithm is adopted to identify 280 privacy-sensitive object classes and events. Second, users on the social network are clustered into a set of representative social groups to generate a discriminative dictionary for user trustworthiness characterization. Finally, both the image content sensitiveness and the user trustworthiness are integrated to train a tree classifier to recommend fine-grained privacy settings for social image sharing. Our experimental studies have demonstrated both the efficiency and the effectiveness of our proposed algorithms.**

***Index Terms*— Privacy setting recommendation, image content sensitiveness, user trustworthiness, deep multiple instance learning, tree classifier, social image sharing.**

I. Introduction

WITH the growing popularity of smart-phones and other mobile devices, high-quality cameras are becoming increasingly ubiquitous and pervasive. As a result, capturing high-quality images has become one part of our daily activities and image sharing has now become very popular on social platforms like Facebook, Myspace, Flickr, and Instagram [1]–[10]. Since social images can intuitively tell when and where a special moment took place, who participated and what were their relationships, the shared images can reveal much of users' personal and social environments and their private lives [1]–[8]. In addition, social network sites may nowadays abuse the technologies of artificial intelligence and facial recognition on automatically tagging objects of interest and human faces [59]–[69]. Thus, privacy protection is a critical issue to be addressed during social image sharing [6].

To ensure privacy, most social sites for image sharing allow users to manually specify coarse-grained privacy settings: whether an image is public, private or visible to their family members or friends. Due to the lack of privacy knowledge, it is not easy for common users to correctly configure privacy settings to achieve their desired levels of privacy protection; also, given the large number of images being shared and the tedious steps needed for privacy settings, some users may not be willing to spend extra time on providing their fine-grained privacy settings for image sharing. To reduce users' additional burdens on configuring the privacy settings manually, it is very attractive to develop new techniques that are able to recommend fine-grained privacy settings for social image sharing.

It is worth noting that the visual properties of the images are the most important resource that can be used to characterize the image content sensitiveness (privacy) [34]–[38]: (a) sharing the images with privacy-sensitive objects (persons and others such as locations) and events may result in unwanted privacy disclosure; (b) visually-similar images often contain similar privacy-sensitive objects and events. Thus performing deep image analysis may play an important role in recommending fine-grained privacy settings for social image sharing and privacy protection [27]–[33].

By assuming that the visual features for image content representation have strong correlations with the image content sensitiveness (privacy), the visual-based approach [27]–[32] has leveraged the hand-crafted visual features (such as SIFT (scale-invariant feature transformation) features, GIST, color histograms) to learn the classifiers to recommend appropriate privacy settings for social image sharing. Because deep learning [39]–[44] has demonstrated its outstanding abilities on extracting high-level features and significantly boosting

Manuscript received June 26, 2017; revised October 20, 2017 and December 19, 2017; accepted December 20, 2017. Date of publication January 8, 2018; date of current version January 30, 2018. This work was supported in part by the National Science Foundation under Grant 1651166-CNS and Grant 1651455-CNS and in part by the NSFC under Grant 61772161. The work of J. Yu was supported in part by the NSFC-61622205 and in part the NSFC-61472110. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Anderson D. R. Rocha. *(Jun Yu and Zhenzhong Kuang contributed equally to this work.) (Corresponding author: Jianping Fan.)*

J. Yu and Z. Kuang are with the School of Computer Science, Hangzhou Dianzi University, Hangzhou 310018, China, and also with UNC-Charlotte, Charlotte, NC 28223 USA (e-mail: yujun@hdu.edu.cn; zzkuang@hdu.edu.cn).

B. Zhang is with the School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China, and also with UNC-Charlotte, Charlotte, NC 28223 USA (e-mail: bpzhang@bjtu.edu.cn).

W. Zhang is with the School of Computer Science, Fudan University, Shanghai 200433, China, and also with UNC-Charlotte, Charlotte, NC 28223 USA (e-mail: weizh@fudan.edu.cn).

D. Lin is with the Department of Computer Science, Missouri University of Science and Technology, Rolla, MO 65409 USA (e-mail: lindan@mst.edu).

J. Fan is with the Department of Computer Science, UNC-Charlotte, Charlotte, NC 28223 USA (e-mail: jfan@uncc.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2017.2787986

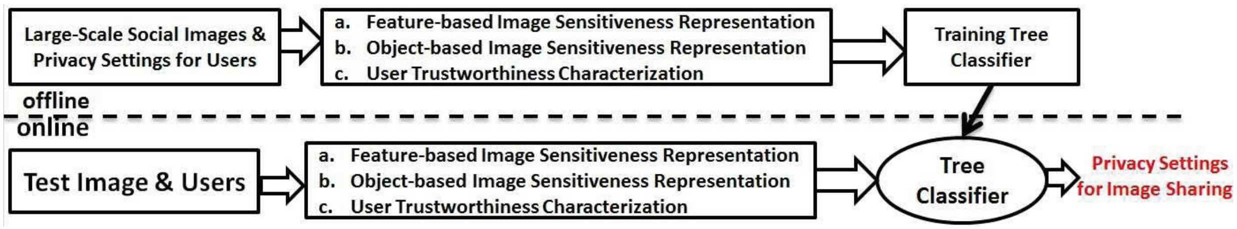

Large-Scale Social Images & Privacy Settings for Users
a. Feature-based Image Sensitiveness Representation
b. Object-based Image Sensitiveness Representation
c. User Trustworthiness Characterization
Training Tree Classifier
offline
online
Test Image & Users
a. Feature-based Image Sensitiveness Representation
b. Object-based Image Sensitiveness Representation
c. User Trustworthiness Characterization
Tree Classifier
Privacy Settings for Image Sharing

Fig. 1. The flowchart for our fine-grained privacy setting recommendation algorithm by considering both image content sensitiveness and user trustworthiness simultaneously.

the accuracy rates for many image understanding tasks, some researchers have leveraged deep learning to train more discriminative classifiers for image privacy prediction [27], [28], [30], [32], and they have found that the deep features can be used to recommend more appropriate privacy settings for social image sharing. One major problem for the visual-based approach is its low interpret-ability [33]: (a) the hand-crafted visual features for image content representation may not have exact correlations with the image content sensitiveness; (b) the deep features may have better interpret-ability at certain levels [54], but they may not be able to exactly represent the appearances of the privacy-sensitive object classes and events in the images. In addition, it is worth noting that the fine-grained privacy settings for social image sharing depend on two inseparable issues simultaneously: (1) sensitiveness of visual content of the images being shared; (2) trustworthiness of the users being granted to see the images. Thus, it is very attractive to develop new algorithms that are able to consider these two inseparable issues simultaneously to recommend fine-grained privacy settings for social image sharing.

Motivated by this observation, a new algorithm is developed in this paper by leveraging both the image content sensitiveness and the user trustworthiness to recommend fine-grained privacy settings for social image sharing. First, as shown in Fig. 1, two approaches are developed for image content sensitivity representation: (a) *Feature-based approach*: By adapting the structure of the AlexNet [39]–[41] to our new task for privacy setting recommendation and integrating user-provided images to fine-tune the underlying kernel weights, 1024-D deep features are learned for image content sensitiveness representation. (b) *Object-based approach*: The privacy-sensitive object classes and events are identified automatically and they are used for image content sensitiveness representation. Second, the users on the social network are clustered into a set of representative social groups to generate a discriminative dictionary for user trustworthiness characterization. Finally, both the image content sensitiveness and the user trustworthiness are seamlessly integrated to train a tree classifier to recommend fine-grained privacy settings for social image sharing.

The remaining of the paper is organized as follows. Section 2 reviews the related work briefly; Section 3 introduces our feature-based approach for image content sensitivity representation; Section 4 presents our deep multiple instance learning algorithm for identifying large numbers of privacy-sensitive object classes automatically and using them for image content sensitiveness representation; Section 5 introduces our algorithm for user trustworthiness characterization; Section 6 presents our tree classifier training algorithm for privacy setting recommendation; Section 7 reports the experimental results for algorithm and system evaluation; Section 8 concludes the paper and outlines the future work.

II. Related Work

Many recent works have studied how to leverage machine learning to automate the privacy setting process [11]–[38]. These pioneering researches can be partitioned into three categories: (a) *Tag-based approach*: Social tags are used to learn a classifier for privacy setting recommendation [11]–[18]; (b) *Topic-based approach*: Keywords or topics from users' profiles are used to partition the friend lists into multiple subgroups or circles [19]–[26], and the friends in the same circle are assumed to share similar privacy preferences; (c) *Visual-based approach*: Visual properties of the images (i.e., visual features or object classes) are leveraged to learn a classifier for privacy setting recommendation [27]–[38].

By assuming that the social tags for image semantics interpretation can also be used to characterize the image content sensitiveness effectively, the tag-based approach leverages the social tags for privacy policy inference [11]–[18]. Vyas *et al.* [11] and Squicciarini *et al.* [12], [13] have leveraged the social tags for privacy policy inference and good performances are reported. Klemperer *et al.* [16] studied whether the keywords from social tags can be used to help users create and maintain access-control policies more intuitively. Ravichandran *et al.* [17] studied how to leverage zone tags to predict a user's privacy preferences from the location data (i.e., share the locations or not). Yeung *et al.* [18] have leveraged social tags and linked data for providing access control to online photo albums.

When high-quality social tags are available, such tag-based approach can recommend appropriate privacy settings for social image sharing. Because tagging rich image semantics could be a time-consuming process [7], most images may be associated with low-quality social tags (i.e., noisy tags, missing tags and spam tags). As a result, such tag-based approach may not be able to recommend appropriate privacy settings for social image sharing [8]–[10]. Another shortcoming for the tag-based approach is that it completely ignores the user trustworthiness for privacy setting recommendation, however, the fine-grained privacy settings for social image

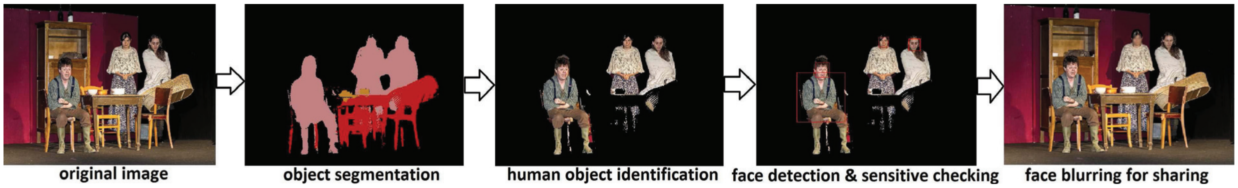

original image
object segmentation
human object identification
face detection & sensitive checking
face blurring for sharing

Fig. 2. The key operations in our iPrivacy system for image privacy protection.

sharing may also change with different users according to their trustworthiness.

To automate the privacy setting process, the topic-based approach uses the keywords or topics from users' profiles for privacy preference prediction [19]–[26]. Fang and LeFevre [21] proposed a privacy wizard to help users grant privileges to their friends. The wizard asks users to first assign privacy labels to the selected friends, and then uses this as the input to construct a classifier to group friends based on their profiles and automatically assign privacy labels to the unlabeled friends. Similarly, Danezis [19] proposed a machine-learning based approach to automatically extract privacy settings from the social context within which the data is produced. Parallel to the work of Danezis [19], Adu-Oppong *et al.* [20] developed privacy settings based on a concept of "social circles" which consist of clusters of friends formed by partitioning users' friend lists. However, such topic-based approach considers only the user trustworthiness but completely ignores the image content sensitiveness, thus it may not be able to recommend appropriate privacy settings for image sharing because the fine-grained privacy preferences may also change according to the image content sensitiveness.

The visual properties of the images are recognized as the most important information source that may significantly affect the privacy settings for image sharing [27]–[38]. Zerr *et al.* [27] and [28] were the first team to leverage the visual features for supporting privacy-aware image classification, where a large number of participants are asked to label images into two categories: private *vs.* public. A classifier is learned from user-labeled images and meta data are also integrated to achieve better performance on privacy-aware image classification. Squicciarini *et al.* [29] have exploited both SIFT features and facial recognition to achieve automatic image privacy prediction. More recently, Tonge and Caragea [30] have first integrated the deep features for image privacy prediction, Spyromitros-Xioufis *et al.* [32] have recently leveraged user-dependent images and privacy settings to support personalized privacy-aware image classification. Both teams have found that the deep features can yield remarkable improvements on the performance as compared with other hand-crafted visual features such as SIFT, GIST and color histograms. One shortcoming of the visual-based approach is that it just considers the image content sensitiveness but completely ignores the user trustworthiness for privacy setting recommendation, however, the fine-grained privacy settings for social image sharing may also change with different users according to their trustworthiness.

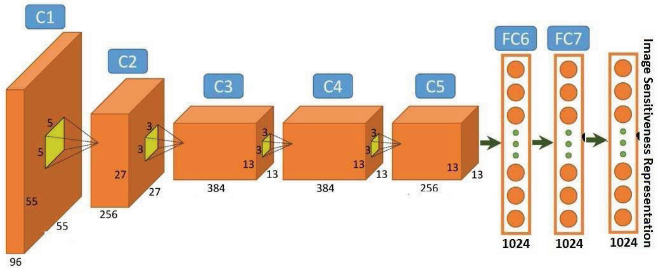


Fig. 3. The flowchart of our feature-based approach to extract 1024-D features for image sensitiveness (privacy) representation.

As illustrated in Fig. 2, we have recently developed a deep-learning-based approach called *iPrivacy* (image Privacy) [33] which is capable of recognizing human objects in the images, determining their privacy sensitiveness levels and then blur faces of human subjects who have high levels of privacy concerns. However, *iPrivacy* does not consider the effect of users' social behaviors (i.e., user trustworthiness) on privacy setting recommendation, thus it cannot provide fine-grained access control yet, e.g., an image may be fine to be directly shared with close family members while need to be blurred when showing to the public. In addition, face blurring used in our iPrivacy system may protect image privacy at certain level but it may also raise speculations.

III. Feature-Based Approach for Image Content Sensitiveness Representation

As shown in Fig. 3, a feature-based approach is developed to extract discriminative deep features for image content sensitiveness representation. By assuming that the visual properties of the images have hidden correlations with their visual content sensitiveness, the deep features learned for image content representation are further used to approximate the sensitiveness (privacy) of image content. However, it is not a good idea to directly apply the AlexNet (that are optimized for recognizing 1,000 atomic object classes [39]–[41]) to extract 4096-D deep features for our new task of privacy setting recommendation, and the reason is that the task space for privacy setting recommendation is much smaller than that for large-scale visual recognition. Our feature-based approach can work for two scenarios: (a) recognizing two categories for our binary approach, e.g., assigning the image-user pairs (i.e., the relationships between the images and the users) into one of two categories of fine-grained privacy settings: share *vs.* not-share; (b) recognizing four categories for our multi-category approach, e.g., assigning the image-user pairs into one of four categories of fine-grained privacy settings: {completely-share, not-share, partially-share, share-with-blurring}.

Based on this observation, the outputs for the 2 fully-connected layers (i.e., FC6 and FC7) in our deep network are scaled down to 1,024 rather than 4,096 in the AlexNet [39]–[41], e.g., the number of kernel mappings for the 2 fully-connected layers (i.e., FC6 and FC7) are scaled down into 25% of that for the AlexNet [39]–[41], and Dropout [53] is applied to 2 fully-connected layers with a value of 0.5 to prevent over-fitting. In addition, we use the kernel weights for the AlexNet [39]–[41] to initialize the weights for the mapping kernels on our deep network, so that we can use a small number of user-provided images to fine-tune the node weights and achieve acceptable accuracy rates [51].

Given the user-provided images, the predictions of the categories for their privacy settings are calculated and the errors for these user-provided images are calculated. We formulate the training error rate ξ in the form of softmax regression:

$$\xi(W, x, y) = -\sum_{l=1}^{\tau} \mathbb{I}\{y_j\} log \left\{ \frac{exp(W_l^T x_j + b)}{\sum_{i=1}^{\tau} exp(W_i^T x_j + b)} \right\} \quad (1)$$

where $\tau = 2$ or $\tau = 4$ is the total number of categories being recognized for privacy setting recommendation, $\mathbb{I}\{y_j\}$ is the indicator function such that $\mathbb{I}\{y_j\} = 1$ if $y_j = 1$ (i.e., (x_j, y_j) is the positive training image), otherwise $\mathbb{I}\{y_j\} = 0$. The corresponding gradients for the objective function are calculated as $\frac{\partial \xi(W,x,y)}{\partial W}$, and they are back-propagated [51] through our deep network to fine-tune the kernel weights.

By adapting the structure of the AlexNet [39]–[41] to our new task (i.e., recognizing two categories or four categories for fine-grained privacy setting recommendation) and integrating user-provided social images to fine-tune the node weights, we can learn more representative deep network to extract more discriminative deep features for image content sensitiveness representation. For a given image I, its visual content sensitiveness can be precisely represented as a histogram of 1024-D deep features x_s.

IV. Object-Based Approach for Image Content Sensitiveness Representation

As shown in Fig. 4, an object-based approach is developed to achieve more discriminative representations of image content sensitiveness. Our idea is to learn a deep network to automatically identify large numbers of privacy-sensitive object classes and events for image content sensitiveness representation. Our object-based approach contains the following key steps: (a) A category hierarchy is constructed to organize various types of image privacy concerns and their most relevant privacy-sensitive object classes hierarchically; (b) A deep multiple instance learning algorithm is developed to learn the classifier to detect 268 privacy-sensitive object classes automatically; (c) The CRF (conditional random field [50]) models are further learned for predicting 12 privacy-sensitive image events; (d) All these privacy-sensitive object classes and events are used to generate a 280-D discriminative dictionary for image content sensitiveness (privacy) characterization.

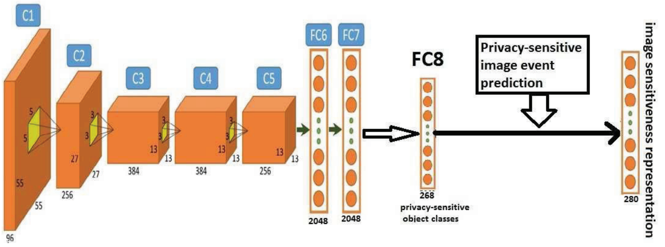

C1
C2
C3
C4
C5
FC6
FC7
FC8
5
5
3
3
3
3
3
3
55
55
96
27
27
256
13
13
384
13
13
384
13
13
256
2048
2048
268
privacy-sensitive object classes
Privacy-sensitive image event prediction
280
image sensitiveness representation

Fig. 4. The flowchart of our object-based approach for image sensitiveness (privacy) representation.

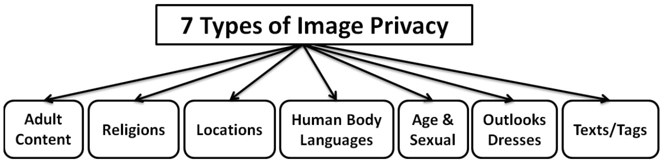


Fig. 5. The category hierarchy for organizing 7 types of image privacy concerns and their most relevant privacy-sensitive object classes.

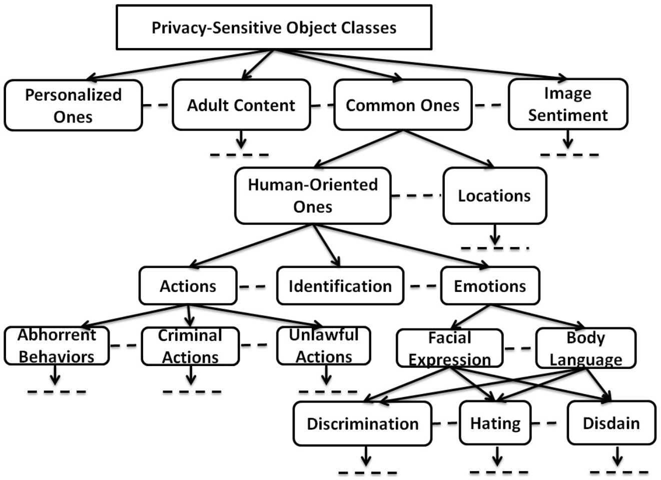


Fig. 6. The category hierarchy for organizing 7 types of image privacy concerns and their most relevant privacy-sensitive object classes.

A. Category Hierarchy

The critical challenge to be conquered here is to identify various types of image privacy concerns and their most relevant privacy-sensitive object classes, so that we can use them to achieve more effective characterization of image sensitiveness (privacy). To tackle this challenge, as shown in Fig. 5 and Fig. 6, a category hierarchy is constructed to organize various types of image privacy concerns and their most relevant privacy-sensitive object classes hierarchically. In this paper, we focus on 7 types of privacy concerns and their most relevant privacy-sensitive object classes: (a) adult content; (b) locations; (c) religions; (d) age and sexual orientation; (e) human body languages (such as facial expressions); (f) human outlooks and dresses; (h) texts and identifiable personal tags.

We use an example of outlook privacy to identify its most relevant privacy-sensitive object classes. For example,

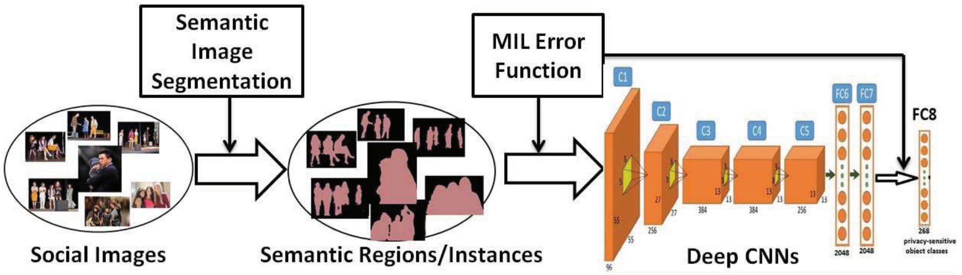


Fig. 7. The flowchart for our deep multiple instance learning algorithm.

a person may wear heavy makeup sometimes or have face acnes during a period. Under different facial conditions, they may have different levels of privacy concerns. In order to capture these outlook privacy concerns, the most relevant privacy-sensitive facial classes include: (1) skin smoothness; (2) skin softness; (3) face shape (roundness); (4) face sizes; (5) face acne; (6) wrinkles; (7) bags under-eyes; (8) heavy makeup; (9) shiny; (10) skin elastic; (11) mustache; (12) facial expressions, etc. In our current implementations, we have identified 268 privacy-sensitive object classes.

Our category hierarchy can allow us to identify the most relevant keywords (text terms) to precisely describe various types of privacy concerns and their most relevant privacy-sensitive object classes, so that we can use these keywords to crawl large-scale training images from multiple social sites. It is worth noting that our category hierarchy contains sufficient numbers of privacy-sensitive object classes to quantify the image content sensitiveness in a fine-grained level, which can allow us to recommend fine-grained privacy settings for social image sharing.

B. Deep Multiple Instance Learning of Object Detectors

Even deep learning has demonstrated its outstanding performances on many image understanding tasks, it requires large-scale manually-labeled training images [39]–[44], but it is a laborious task to label large-scale object regions manually for learning the object detectors. In this paper, a deep multiple instance learning algorithm is developed to directly leverage the coarsely-labeled images (i.e., object labels are coarsely given at the image level rather than at the region level) for learning the object detectors. Our deep multiple instance learning algorithm takes the following steps as illustrated in Fig. 7: (a) Each social image is first segmented into a set of semantic object regions (instances); (b) A noise-or model is used to define the error function for supporting deep multiple instance learning, e.g., learning the deep network and the object detectors jointly in an end-to-end manner.

Deep CNNs have shown their strong ability on supporting pixel-level image classification (i.e., semantic image segmentation by assigning one particular semantic label to every pixel in an image) [45]–[49]. To extract semantic object regions from the images, we have trained a deep network in an end-to-end way to enable pixel-level prediction and classification, and a CRF (conditional random fields) model [50] is further learned to merge the neighboring pixels with the same labels to generate semantic object regions [48]. As shown in Fig. 8, by integrating deep CNNs with CRF models for semantic image segmentation, we can identify semantic object regions precisely.

Multiple instance learning (MIL) [55]–[58] has been used to deal with the issue of coarse labeling by treating each image (which may contain multiple objects) as a bag of instances. In our deep multiple instance learning algorithm, we use the kernel weights from the AlexNet [39]–[41] to initialize the mapping kernels on our deep CNNs, so that we can use a small number of user-provided training images to fine-tune the kernel weights effectively and achieve acceptable accuracy rates. As illustrated in Fig. 7, a special unit is inserted into our deep CNNs to support multiple instance learning and handle the issue of coarse labeling. Our special unit for multiple instance learning contains two components: (a) Noise-or model [55] is used to predict the image labels (bag labels) from the instance labels (region labels); (b) The visually-similar image regions are assumed to share the same object label. Given an image region or object proposal, the prediction of its privacy-sensitive object class is calculated, the error function for our deep multiple instance learning algorithm contains two parts:

$$
\begin{aligned}
J(W) \\
= \sum_{k=1}^{N} \Bigg\{ \frac{\lambda}{R} \sum_{j=1}^{R} \Delta_{bag}(B_j^k, \overline{B_j^k}) + \frac{1}{R^2} \sum_{h=1}^{R} \sum_{l=1}^{R} \delta_{hl} \kappa(x_h^k, x_l^k) \\
\Delta_{instance}(y_h^k, y_l^k) I\{y_l^k\} log \left\{ \frac{exp(W_k^T x_l^k + b)}{\sum_{l=1}^{N} exp(W_l^T x_j^l + b)} \right\} \Bigg\} \quad (2)
\end{aligned}
$$

where $N = 268$ is the total number of privacy-sensitive object classes being recognized, R is the number of training images for each object class, $\kappa(x_h^k, x_l^k)$ is used to characterize the visual similarity between two image regions from the same image (bag), δ_{hl} indicates that we only consider the visual similarity among the image regions from the same image, $\Delta_{bag}(B_j^k, \overline{B_j^k})$ is used to calculate the difference between the predicted bag (image) label $\overline{B_j^k}$ and the given bag label B_j^k for the kth image (bag), and $\Delta_{instance}(y_h^k, y_l^k)$ is used to calculate the difference on their labels between two visually-similar image regions (x_h^k, y_h^k) and (x_l^k, y_l^k) from the same image (bag) B_j^k. We use the noise-or model to predict the label $\overline{B_j^k}$ for a given positive bag (image) B_j^k by cumulating all the predictions for its image regions (instances).

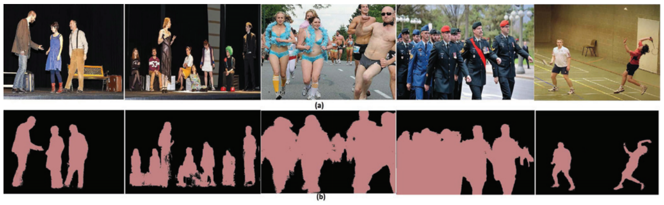


Fig. 8. Our results on semantic image segmentation: (a) original images; and (b) object regions.

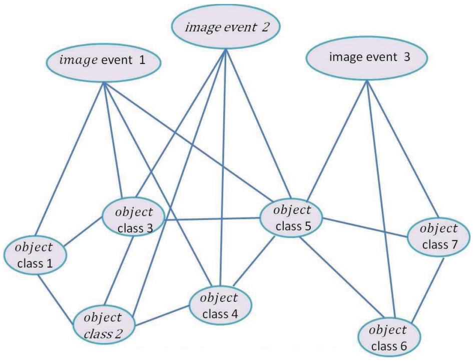


Fig. 9. The two-layer CRF models for image event prediction.

The gradients for the objective function are calculated as $\frac{\partial J(W)}{\partial W}$, and they are back-propagated [51] through our deep CNNs to fine-tune the kernel weights. After such deep CNNs is available, it is used to detect 268 privacy-sensitive object classes from the images being shared.

C. Predicting Privacy-Sensitive Image Events

Some object classes may not be sensitive individually, but their co-occurrences in the same image may convey privacy-sensitive image event. Thus it is very attractive to leverage such object co-occurrences to infer the appearances of the most relevant privacy-sensitive image events. First, an object co-occurrence network is constructed and it consists of two key components: (a) object classes; and (b) their inter-class co-occurrences in large-scale social images. Second, as illustrated in Fig. 9, over the object co-occurrence network, a two-layer CRF (conditional random field [50]) model is learned for predicting privacy-sensitive image events. In Fig. 9, the first layer is used to interpret the appearances of the object classes and their co-occurrences (i.e., our object co-occurrence network), the second layer is used to interpret the appearances of the most relevant privacy-sensitive image events, e.g., the co-occurrences of some object classes in the images are sufficient to indicate the appearances of the most relevant privacy-sensitive image events. Thus our two-layer CRF models are used to learn the conditional probabilities over the appearances of the object classes, their co-occurrences and the appearances of the most relevant privacy-sensitive image events.

For a given training image I, our deep multiple instance learning algorithm can effectively extract a set of object classes O from the training image I, and their co-occurrences X can further be identified from our object co-occurrence network, we can estimate the probability $P(y_j|O, X, \Theta_j)$ for the appearance of the most relevant privacy-sensitive image event y_j as:

$$P(y_j|O, X, \Theta_j) = \frac{1}{Z(\Theta)} exp\left(F_j(y_j \mid X, O, \Theta_j)\right) \tag{3}$$

where $F_j(y_j \mid X, O, \Theta_j)$ is the classifier for the jth privacy-sensitive image event y_j given the set of object classes O and their co-occurrences X, $Z(\Theta)$ is the partition function and it is defined as:

$$Z(\Theta) = \sum_{j=1}^{12} exp\left(F_j(y_j \mid X, O, \Theta_j)\right) \tag{4}$$

Such two-layer CRF models are learned from a set of training images and they are then used to predict the presences of the most relevant privacy-sensitive image events in the social images being shared when the appearances of the object classes and their co-occurrences are determined. In our current work, we focus on learning the two-layer CRF models to predict 12 privacy-sensitive image events.

D. Image Content Sensitiveness Representation

The privacy-sensitive object classes and events, which are identified from large-scale social images and frequently occur in the private (not-share) images, are selected to generate a 280-D discriminative dictionary D_I for image content sensitiveness representation. For a given image I, its privacy-sensitive object classes and event are first detected (by our deep multiple instance learning algorithm and our two-layer CRF models). By projecting its privacy-sensitive object classes and event over such 280-D discriminative dictionary, the content sensitiveness (privacy) of the given image I can precisely be represented as a 280-D histogram (i.e., a 280-D bag of privacy-sensitive object classes and events x_s). Like object bank [52] for image content representation, such 280-D bags of privacy-sensitive object classes and events can characterize the image content sensitiveness effectively. For a given image, its 280-D bag of privacy-sensitive object classes and events x_s is very sparse.

V. User Trustworthiness Characterization

The intuition is that how the image owner interacts with others in the social networks show the hints on whom they would share with what types of images. For example, if a user interacts with his/her friend Bob frequently, it is likely that the user would share more images with his/her friend Bob. Yet another example is that if a user usually shares the images of family events only with his/her family members, it is likely the user will do the same for the new images of family events. Therefore, our goal is to characterize these social behaviors and study their correlations with the fine-grained privacy settings for social image sharing.

We have explored multiple factors that describe users' social behaviors, which include: (1) types of relationships (such as friends, family members, colleagues) with the image owner which could help distribute images (with various visual content) to different groups of users; (2) closeness of the relationships with the image owner as the image owner may

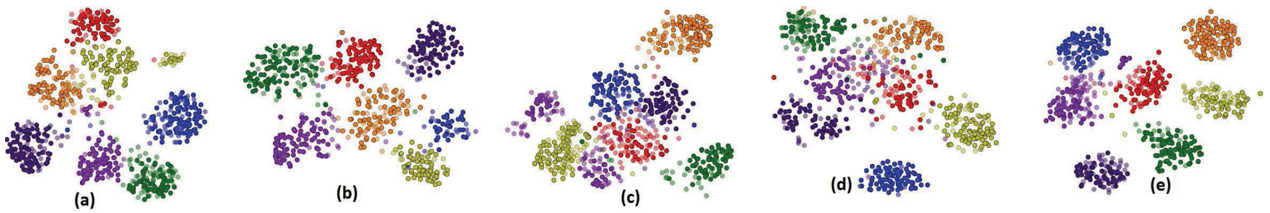


Fig. 10. The clustering results of different users (a)–(e) for multiple smart parts of a social network, where different user groups are represented in different colors.

share sensitive images with people who are very close to him/her; (3) matching scores between the user's interesting topics (from user profiles) and the semantics of the images being shared, e.g., people may have stronger motivations to distribute the image which is very interesting to him/her; (4) interaction intensity between the user and the image owner, which could be another indicator of how likely the image would be shared; (5) user's activity score in social network considering that active users may have a higher chance to distribute the images; (6) stability scores of users' behavior history; (7) reputation scores to assess users' self-representation of honesty and reliability in social network.

In order to learn the effect of these multi-factors on privacy setting configuration, we first analyze individual factor and define the corresponding function to quantify the value of each factor in a fine granularity. Then, we construct a high-dimensional feature vector based on the obtained fine-grained values of all the factors to represent the user's social behaviors, and a joint function is learned to characterize the similarities among the users according to their multi-factors for social behavior characterization. By using multi-factors for user's social behavior representation and treating each user on a social network as one node on a graph, we can use spectral clustering to partition large numbers of users into a set of representative social groups according to their similarities (closeness) on their multi-factors for social behavior characterization. For a given social network with M users, its M users are partitioned into B representative social groups by minimizing inter-group similarity and maximizing intra-group similarity:

$$min \left\{ \Psi(M, B) = \sum_{l=1}^{B} \frac{\sum_{u_i \in G_l} \sum_{u_j \in G^c/G_l} \kappa(u_i, u_j)}{\sum_{u_i \in G_l} \sum_{u_j \in G_l} \kappa(u_i, u_j)} \right\} \quad (5)$$

where $\kappa(u_i, u_j)$ is the kernel function to characterize the similarity (closeness) between two users u_i and u_j on their multi-factors for social behavior characterization, $G^c = \{G_l | l = 1, \cdots, B\}$ is used to represent B groups (clusters) of M users on the given social network, G^c/G_l is used to represent other $B - 1$ groups in G^c except G_l.

As shown in Fig. 10, the users on the social network are clustered into B representative social groups (such as altruistic users, cynical users, forgiving users, distrusting users, et al.) according to the similarities (closeness) on their multi-factors for social behavior characterization [1], [16]–[25]. Such representative social groups can effectively characterize the relationships and trustworthiness among the users, thus they can be used to generate a B-D discriminative dictionary D_u for user trustworthiness characterization.

By assigning each user u (characterized by their multi-factors for social behavior characterization) onto one or multiple of B representative social groups in the B-D discriminative dictionary D_u, each user and his/her trustworthiness can be represented as a B-D histogram of representative social groups (i.e., a B-D bag of representative social groups x_u). Such B-D bags of representative social groups can characterize the user trustworthiness effectively (e.g., their closeness or similarities on their multi-factors for social behavior characterization can be used to characterize their trustworthiness in certain accuracy). For each user, such B-D bag of representative social groups x_u is very sparse.

VI. Tree Classifier for Fine-Grained Privacy Setting Recommendation

Without loss of generality, we consider the privacy policies that contain the following components: (a) *Subject S*: a set of users who are socially connected to the image owner u and are granted to access the shared images Φ; (b) *Images* Φ: a set of images shared from u to S; (c) *Action A*: a set of actions granted by u to S on Φ.

The key issue for automating the privacy setting process is to train a classifier for assigning the relationships between the images (represented by their visual content sensitiveness x_s) and the users (characterized by their trustworthiness x_u) into a set of pre-defined categories for fine-grained privacy settings, e.g., assigning the image-user pairs (or the relationships between the images and the users) into a set of pre-defined categories for fine-grained privacy settings.

In this work, two approaches are developed to simultaneously consider both the image content sensitiveness x_s and the user trustworthiness x_u in different modalities for fine-grained privacy setting recommendation: (a) **binary approach:** as shown in Fig. 11, the image-user pairs ($x_s \oplus x_u$) are assigned into one of two categories of fine-grained privacy settings, e.g., *share* or *not-share*; (b) **multi-category approach:** as illustrated in Fig. 12, the image-user pairs ($x_s \oplus x_u$) are assigned into one of four categories of fine-grained privacy settings, e.g., {completely-share, not-share, partially-share, share-with-blurring}.

It is worth noting that: (a) both the feature-based approach and the object-based approach for image content sensitiveness representation can be used to support our binary approach for fine-grained privacy setting recommendation; (b) only the object-based approach for image content sensitiveness representation can be used to support our multi-category approach for fine-grained privacy setting recommendation, e.g., supporting more options (more categories) for fine-grained privacy

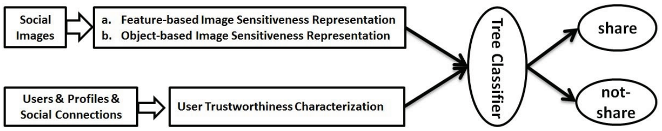

Social Images
a. Feature-based Image Sensitiveness Representation
b. Object-based Image Sensitiveness Representation
Users & Profiles & Social Connections
User Trustworthiness Characterization
Tree Classifier
share
not-share

Fig. 11. The flowchart of our binary approach for privacy setting recommendation.

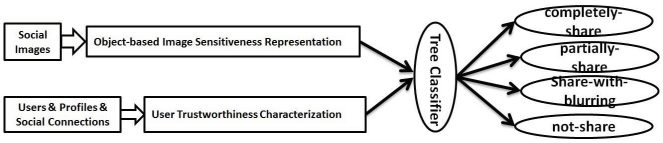

Social Images
Object-based Image Sensitiveness Representation
Users & Profiles & Social Connections
User Trustworthiness Characterization
Tree Classifier
completely-share
partially-share
Share-with-blurring
not-share

Fig. 12. The flowchart of our multi-category approach for privacy setting recommendation.

settings requires deeper analysis and semantic understanding of images.

A. Binary Approach

For a given training image I, its visual content sensitiveness is represented as a 1024-D deep feature x_s or a 280-D bag of privacy-sensitive object classes and events x_s, the privacy setting for one particular user u (who is granted to access the given image I and his/her trustworthiness is represented as a B-D bag of representative social groups x_u) is defined as a binary canonical policy: (a) ***share***: this user u with the trustworthiness representation x_u is granted to see the given image I with the visual content sensitiveness representation x_s, e.g., the image-user pair (the relationship between the image x_s and the user x_u: $x_s \oplus x_u$) is assigned into the category "share"; and (b) ***not-share***: this user u with the trustworthiness representation x_u is not granted to see the given image I with the visual content sensitiveness representation x_s, e.g., the image-user pair (the relationship between the image x_s and the user x_u: $x_s \oplus x_u$) is assigned into the category "not-share".

The goal of our binary approach for fine-grained privacy setting recommendation is to learn a classifier $f(c \mid x_s, x_u, \Theta)$, $c \in$ {share, not-share}, to achieve precise assignment between the image x_s and the user x_u (i.e., image-user pair $x_s \oplus x_u$), e.g., assigning the relationship between the image x_s and the user x_u into one of two categories for fine-grained privacy settings (i.e., share & not-share). Because the features for image content sensitiveness representation and user trustworthiness characterization are in different modalities and they are not comparable directly, it is not a good idea to simply concatenate x_s for image content sensitiveness representation with x_u for user trustworthiness characterization as an unified feature (x_s,x_u). As illustrated in Fig. 13, a tree classifier $f(c \mid x_s, x_u, \Theta)$ is trained to leverage both the image content sensitiveness (x_s) and the user trustworthiness (x_u) in different modalities for fine-grained privacy setting recommendation, where different nodes on the tree classifier can select different features (x_u or x_s) for node classifier training.

The tree classifier $f(c \mid x_s, x_u, \Theta)$ is learned from large-scale training images and their privacy settings that are assigned for a large number of users: (a) large-scale social images and their image content sensitiveness representations x_s; (b) large numbers of users on the social networks and their trustworthiness representations x_u; (c) two categories for fine-grained privacy settings (i.e., the pairwise relationships between the users and the images). Each labeled training sample is defined as: (image-user pair, privacy setting) = (feature pair $x_s \oplus x_u$ for image content sensitiveness representation and user trustworthiness characterization, label c) = ($x_s \oplus x_u$, c), $c \in$ {share, not-share}.

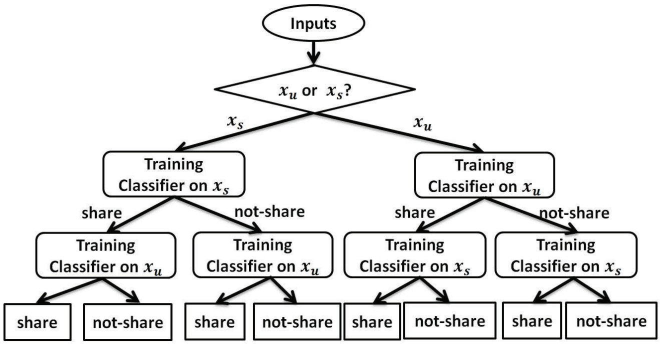

Inputs
x_u or x_s?
x_s
x_u
Training Classifier on x_s
Training Classifier on x_u
share
not-share
share
not-share
Training Classifier on x_u
Training Classifier on x_u
Training Classifier on x_s
Training Classifier on x_s
share
not-share
share
not-share
share
not-share
share
not-share

Fig. 13. The flowchart of our binary approach for tree classifier training.

Given R training images and their visual content sensitiveness representations x_s, their similarities are then calculated and represented as a $R \times R$ similarity matrix $\mathbf{S}$ and its component S_{ij} is used to characterize the similarity between the ith training image and the jth one according to their content sensitiveness representations x_s^i and x_s^j:

$$S_{ij} = \exp\left(-\frac{d(x_s^i, x_s^j)}{\sigma_s}\right) \tag{6}$$

where $d(\cdot, \cdot)$ is the Euclidean distance between two images on their visual content sensitiveness representations x_s^i and x_s^j.

Given T users and their trustworthiness representations x_u, their similarities are then calculated and represented as a $T \times T$ similarity matrix $\mathbf{U}$ and its component U_{kl} is used to characterize the similarity between the kth user and the lth one according to their trustworthiness representations x_u^k and x_u^l:

$$U_{kl} = \exp\left(-\frac{d(x_u^k, x_u^l)}{\sigma_u}\right) \tag{7}$$

where $d(\cdot, \cdot)$ is the Euclidean distance between two users on their trustworthiness representations x_u^k and x_u^l.

Given two categories for fine-grained privacy settings (i.e., share & not-share), the similarity matrix $\mathbf{S}$ and $\mathbf{U}$ can be used as a proxy to determine the separability of training samples (images and users) on two feature subsets in different modalities: (a) image content sensitiveness representations x_s; and (b) user trustworthiness characterizations x_u. As illustrated in Fig. 13, to make the first decision for tree classifier training, the most discriminative feature subset x_{best} is first selected and the associated samples (either R training images or T users) are then partitioned into two categories for fine-grained privacy settings (i.e., share *vs.* not-share). The most discriminative feature subset x_{best} is determined automatically by maximizing the separability:

$$x_{best} = max \left\{ \frac{1}{R^2} \sum_{j=1}^{R} \sum_{i=1}^{R} S_{ij}, \quad \frac{1}{T^2} \sum_{k=1}^{T} \sum_{l=1}^{T} U_{kl}, \right\} \tag{8}$$

If the most discriminative feature subset x_{best} is determined as the image content sensitiveness representations x_s, a binary SVM classifier $f_s(c \mid x_s, \theta)$ is first trained over R training images to obtain the optimal recognition of two categories (i.e., share & not-share) for fine-grained privacy setting recommendation and a binary SVM classifier $f_u(c \mid x_u, \vartheta)$ is then trained over the associated set of users by using the feature subset x_u. If the most discriminative feature subset x_{best} is determined as the user trustworthiness characterizations x_u, a binary SVM classifier $f_u(c \mid x_u, \vartheta)$ is first trained over T users to obtain the optimal recognition of two categories (i.e., share & not-share) for fine-grained privacy setting recommendation and a binary SVM classifier $f_s(c \mid x_s, \theta)$ is then trained over the relevant training images by using the feature subset x_s. As illustrated in Fig. 13, different paths on the tree classifier $f(c \mid x_s, x_u, \Theta)$ have different combinations of two binary SVM classifiers $f_s(c \mid x_s, \theta)$ and $f_u(c \mid x_u, \vartheta)$.

After the tree classifier $f(c \mid x_s, x_u, \Theta)$ is learned, it is further used to automatically configure an appropriate privacy setting between a given image I and one particular user u. For a given image I being shared and one particular user u on the social network of the image owner, as illustrated in Fig. 13, our fine-grained privacy setting recommendation algorithm takes the following key steps to make the decision: (a) our deep network is first used to extract 1024-D deep features or our deep multiple instance learning algorithm is used to detect the privacy-sensitive object classes and event, and the visual content sensitiveness of the given image I is precisely represented by using the 1024-D deep features x_s or the 280-D bag of privacy-sensitive object classes and events x_s; (b) the trustworthiness of the user u is characterized as a B-D bag of representative social groups x_u, e.g., the user's multi-factors for social behavior characterization are used to obtain his/her representative social groups and then project onto the B-D dictionary D_u to obtain a B-D bag of representative social groups x_u for user trustworthiness characterization; (c) our tree classifier $f(c \mid x_s, x_u, \Theta)$ is then used to make the decision {share, not-share} hierarchically according to both the image content sensitiveness (represented by x_s) and the use trustworthiness (characterized by x_u), e.g., project the image-user pair $(x_s \oplus x_u)$ onto the appropriate category for fine-grained privacy settings (i.e., share or not-share).

B. Multi-Category Approach

In our multi-category approach for privacy setting recommendation, for a given training image I, its visual content sensitiveness is represented as a 280-D bag of privacy-sensitive object classes and events x_s, and the fine-grained privacy setting for one particular user u (who is granted to access the given image I and his/her trustworthiness is represented as a B-D bag of representative social groups x_u) is defined as a multi-category policy: (a) ***completely-share***: this user u with the trustworthiness representation x_u is granted to completely see the given image I with the visual content sensitiveness representation x_s, e.g., the image-user pair $(x_s \oplus x_u)$ is assigned into the category "completely-share"; (b) ***not-share/private***: this user u with the trustworthiness representation x_u is not granted to see the given image I with the visual content sensitiveness representation x_s, e.g., the image-user pair $(x_s \oplus x_u)$ is assigned into the category "not-share/private"; (c) ***partially-share***: this user u with the trustworthiness representation x_u is granted to partially see the given image I with the visual content sensitiveness representation x_s, e.g., the image-user pair $(x_s \oplus x_u)$ is assigned into the category "partially-share"; (d) ***share-with-blurring***: this user u with the trustworthiness representation x_u is granted to see the blurring image (for example, the privacy-sensitive object classes are blurred), e.g., the image-user pair $(x_s \oplus x_u)$ is assigned into the category "share-with-blurring".

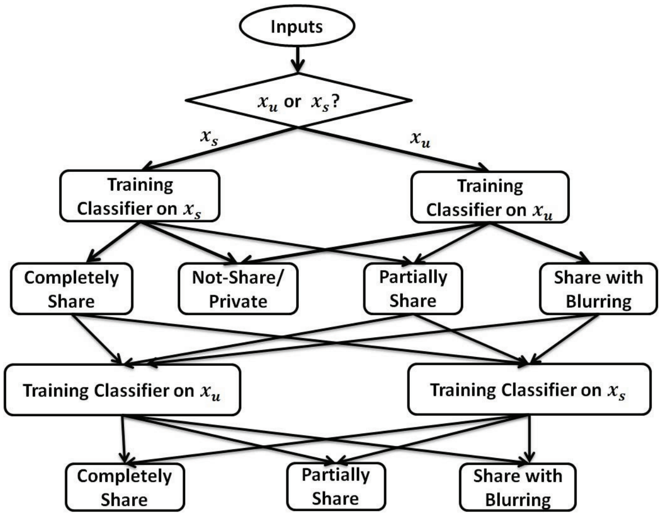


Fig. 14. Our multi-class tree classifier training algorithm for supporting fine-grained privacy setting recommendation.

As illustrated in Fig. 14, the goal of our fine-grained privacy setting recommendation algorithm is to learn a multi-class tree classifier $f(c \mid x_s, x_u, \Theta)$, $c \in$ {completely-share, not-share, partially-share, share-with-blurring}, to achieve precise assignments between the image-user pairs $(x_s \oplus x_u)$ and multiple categories (i.e., completely-share, not-share, partially-share, share-with-blurring) for fine-grained privacy settings. We use similar techniques as introduced above to train our multi-class tree classifier: (a) The most discriminative feature subset x_{best} is first selected automatically; (b) The multi-class SVM classifier is then trained to achieve optimal partitioning of training samples (images or users) by using the most discriminative feature subset x_{best} (x_s or x_u); (c) Different paths on the multi-class tree classifier $f(c \mid x_s, x_u, \Theta)$ have different combinations of the multi-class SVM classifiers $f_s(c \mid x_s, \theta)$ and $f_u(c \mid x_u, \vartheta)$. Some experimental results for supporting multi-category fine-grained privacy setting recommendation are illustrated in Fig. 15, where the special category for "not-share/private" is not demonstrated because no further operations are required.

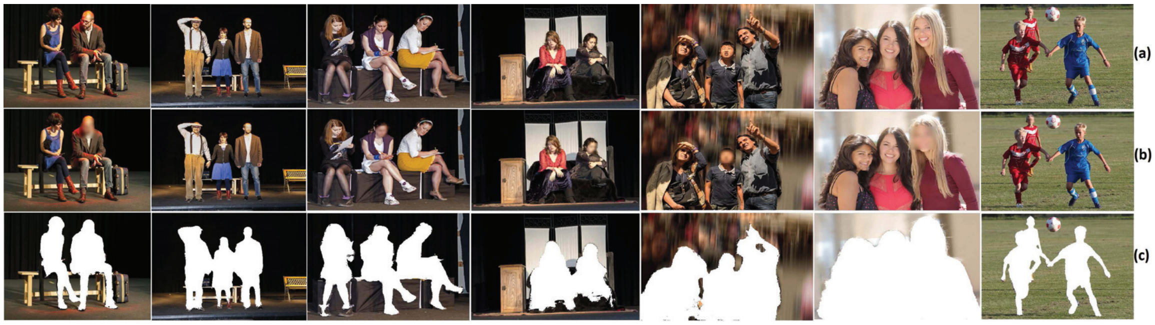


Fig. 15. Examples to illustrate our multi-category approach for fine-grained privacy setting recommendation: (a) completely-share; (b) sharing with blurring (where the sensitive objects are blurred); (c) partially-share by replacing the objects with white silhouettes.

VII. Algorithm and System Evaluation

We have evaluated our proposed algorithms over: (a) our large-scale social images; and (b) two public image sets, PicAlert[1] and Mirflickr.[2]

We have collected 800, 000 social images and their privacy settings are labeled, where 90, 000 images are treated as the test images and others are used as the training images. In the following experiments, we focus on evaluating the effectiveness of our fine-grained privacy setting recommendation algorithm when different types of features are used for image content sensitiveness representation. Specifically, we compare the predicted privacy settings generated by our proposed algorithms with the original privacy settings provided by humans. For 90, 000 test images, we partition them into 900 subsets according to their scores on the correspondences between the image content sensitiveness (privacy) and the visual features for image content representation. We tested our privacy setting recommendation algorithm by using three approaches for image content sensitiveness representation, e.g., low-level visual features, high-level deep features and privacy-sensitive object classes and events.

For a given image set with R test images which are shared with T users under different privacy settings f, its privacy disclosure is defined as:

$$\mathbb{S}=\frac{1}{R\times T}\sum_{j=1}^{T}\sum_{l=1}^{R}\delta(c,\hat{c})\|f(c\mid x_s^l,x_u^j,\Theta)-\overline{f(x_s^l,x_u^j)}\| \quad (9)$$

where $f(c \mid x_s^l, x_u^j, \Theta)$ is the predicted privacy setting for the lth given image with the visual content sensitiveness representation x_s^l to the jth user u with the trustworthiness x_u^j, $\overline{f(x_s^l, x_u^j)}$ is the human-defined privacy setting assigned between the jth user u with the trustworthiness x_u^j and the lth given image with the visual content sensitiveness representation x_s^l, $\delta(c,\hat{c})$ is used to emphasize that the privacy disclosure is counted differently for various situations.

For our binary approach: (a) when both the predicted privacy setting $\hat{c}$ for the given image I and its human-defined one c are *not-share*, $\delta(c,\hat{c})=0$; (b) when both the predicted privacy setting $\hat{c}$ for the given image I and its human-defined one c are *share*, $\delta(c,\hat{c})=0$; (c) when the predicted privacy setting $\hat{c}$ for the given image I is *not-share*, but its human-defined one c is *share*, $\delta(c,\hat{c})=0.5$, such privacy disclosure is the punishment to avoid cheating from the system, e.g., without this penalty, the system may easily achieve low privacy disclosure by recommending the privacy setting for each image as "not-share"; (d) when the predicted privacy setting $\hat{c}$ for the given image I is *share*, but its human-defined one c is *not-share*, $\delta(c,\hat{c})=1$. Thus $\delta(c,\hat{c})$ is defined as:

$$\delta(c,\hat{c})=\begin{cases}0, & \text{c} = \hat{c} = \text{not-share}\\ 0, & \text{c} = \hat{c} = \text{share}\\ 0.5, & \text{c = share}, \ \hat{c} = \text{not-share}\\ 1.0, & \text{c = not-share}, \ \hat{c} = \text{share}\end{cases} \quad (10)$$

For our multi-category approach, $\delta(c,\hat{c})$ is defined as:

$$\delta(c,\hat{c})=\begin{cases}0, & \text{c} = \hat{c} = \text{not-share}\\ 0, & \text{c} = \hat{c} = \text{completely-share}\\ 0, & \text{c} = \hat{c} = \text{partially-share}\\ 0, & \text{c} = \hat{c} = \text{share-with-blurring}\\ 0.5, & \text{c = completely-share},\\ & \hat{c} = \text{partially-share}\\ 0.5, & \text{c = completely-share},\\ & \hat{c} = \text{share-with-blurring}\\ 0.5, & \text{c = completely-share},\\ & \hat{c} = \text{not-share}\\ 1.0, & \text{c = not-share}, \ \hat{c} = \text{partially-share}\\ 1.0, & \text{c = not-share}, \ \hat{c} = \text{share-with-blurring}\\ 1.0, & \text{c = not-share}, \ \hat{c} = \text{completey-share}\end{cases} \quad (11)$$

A. Comparison on Deep Features from Different Networks

By assuming that the low-level visual features extracted for image content representation can also be used to characterize the image content sensitiveness effectively, the low-level visual features are directly used to learn the classifier for privacy setting recommendation, and we have compared two approaches: (1) our feature-based approach: the structure (FC6 and FC7) of the AlexNet [39]–[41] is scaled down and adapted to our new task (i.e., assigning the image-user pairs into two categories of fine-grained privacy settings: share *vs.* not-share or four categories of fine-grained privacy settings: {completely-share, not-share, partially-share, share-with-blurring}) and only 1024-D deep features are extracted

[1] http://l3s.de/picalert/ustudydata
[2] http://s16a.org/mirflickr

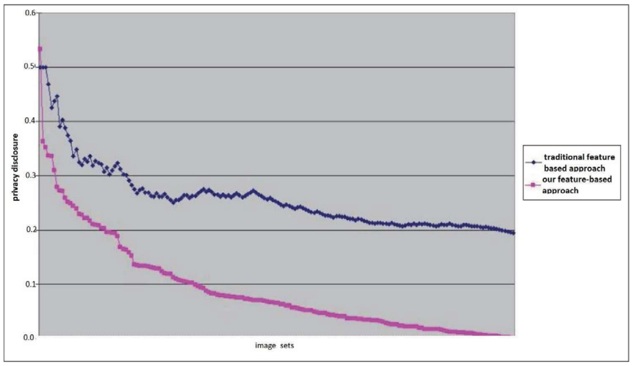


Fig. 16. The comparison on the effectiveness of our feature-based approach by using 1024-D deep features and the traditional one by using the 4096-D deep features extracted by the AlexNet, where the image sets are sorted according to their privacy disclosure.

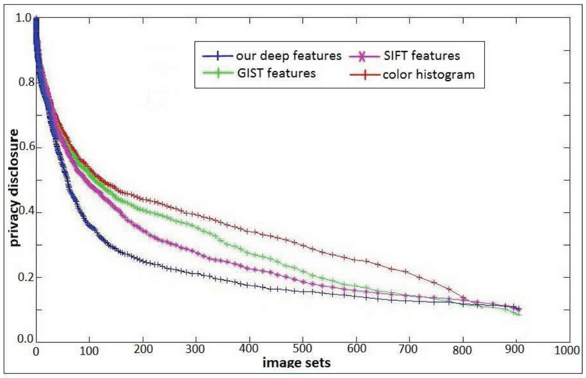


Fig. 17. The comparison on the effectiveness of the feature-based approach when different types of features are extracted for image sensitiveness representation, where the image sets are sorted according to their privacy disclosure.

for image content sensitiveness representation; (2) the traditional feature-based approach: the 4096-D deep features, which are extracted by the AlextNet [39]–[41] for recognizing 1000 atomic object classes, are arbitrarily used for image content sensitiveness representation.

As shown in Fig. 16, one can observe that our feature-based approach can significantly outperform the traditional feature-based approach. The reasons are two folds: (1) our feature-based approach can adapt the structure (FC6 and FC7) of our deep CNNs to learn more discriminative deep features for our new task (i.e., recognizing two categories or four categories for fine-grained privacy setting recommendation); and (2) our feature-based approach can incorporate the user-provided images to fine-tune the kernel weights according to our new task (i.e., recognizing two categories or four categories for fine-grained privacy setting recommendation rather than recognizing 1,000 atomic object classes for large-scale visual recognition application).

B. Comparison on Various Visual Features

It is also very interesting to evaluate whether using different types of visual features for classifier training may bring significant improvement on privacy setting recommendation. As shown in Fig. 17, we have compared the performance of our feature-based approach when different types of visual features are used for image content sensitiveness representation. From these comparison experiments, one can observe multiple interesting factors: (1) For most of 900 image subsets, the deep features can achieve the best performance as compared with other hand-crafted visual features such as SIFT, GIST and color histograms; (2) For some difficult image subsets, which may have low scores on the correlations between the visual features for image representation and the image content sensitiveness (privacy), all these features (including deep features) may not be able to achieve acceptable performance, e.g., all of them have large privacy disclosures because such visual features for image content representation may not be able to characterize the image content sensitiveness (privacy) effectively; (3) For some easy image subsets, which have high scores on the correlations between the visual features for image content representation and the image content sensitiveness, all these features (including hand-crafted visual features) can achieve good performance (resulting in small privacy disclosures).

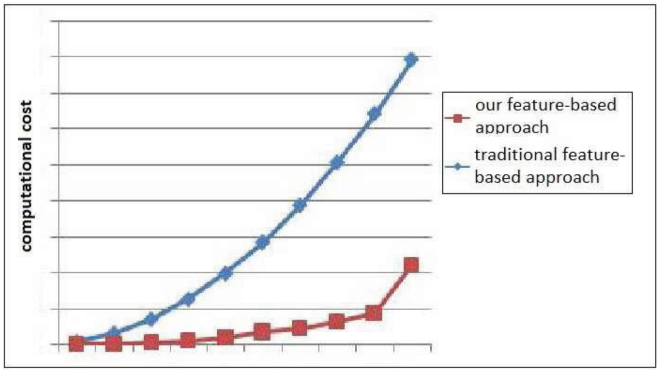


Fig. 18. The comparison on the computational cost for privacy setting recommendation between our feature-based approach and the traditional one by using the 4096-D deep features directly.

C. Comparison on Computational Cost

We have compared the computational cost between two feature-based approaches: (a) our feature-based approach: the deep network is scaled down and its structure is adapted to our new task and only 1024-D deep features are used for image content sensitiveness representation; (b) the traditional feature-based approach: the 4096-D deep features, which are usually learned for recognizing 1000 atomic object classes, are arbitrarily used for image content sensitiveness representation. As shown in Fig. 18, one can observe that our feature-based approach can reduce the computational cost significantly.

D. Effectiveness of Privacy-Sensitive Object Classes

To evaluate the effectiveness of using privacy-sensitive object classes and events on fine-grained privacy setting recommendation, we have compared two approaches: (1) our object-based approach: 268 privacy-sensitive object classes and 12 privacy-sensitive image events are detected and they are used to generate a 280-D discriminative dictionary, thus 280-D bags of privacy-sensitive object classes and events are used for

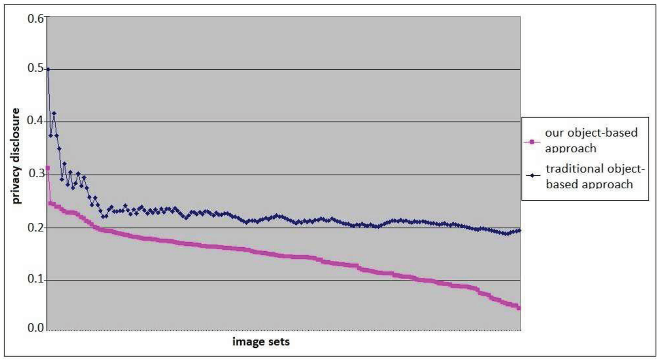


Fig. 19. The comparison on the effectiveness of our approach by using the privacy-sensitive object classes and events and the traditional one by using 1000 atomic object classes directly, where the image sets are sorted according to their privacy disclosure.

image content sensitiveness representation; (2) the traditional object-based approach: 1, 000 atomic object classes, which are originally detected by the AlexNet [39]–[41] for large-scale visual recognition application, are arbitrarily used for image content sensitiveness representation and 1000-D bags of atomic object classes are used for image content sensitiveness representation.

As shown in Fig. 19, one can observe that our object-based approach can significantly outperform the traditional object-based approach. The reason is that the privacy-sensitive object classes and events can characterize the image content sensitiveness effectively, on the other hand, 1000 atomic object classes [39]–[44], that are usually extracted for image semantics interpretation, may not be able to characterize the image content sensitiveness exactly, e.g., the appearances of such 1000 atomic object classes in the images do not exactly relate with the image privacy or cause privacy disclosure directly. Even detecting such 1000 atomic object classes can play important roles on image semantics interpretation, they may not be effective for characterizing the image content sensitiveness precisely.

E. Privacy-Aware Image Classification

In order to achieve more clear understanding of what kind of visual properties makes images to be private (not-share) or public (share), we have evaluated two approaches for privacy-aware image classification. As illustrated in Fig. 20, one can observe that whether the recommended privacy settings are appropriate for image sharing largely depends on whether the privacy-sensitive object classes are significant on giving some insights about the image sensitiveness (privacy). By detecting such privacy-sensitive object classes automatically, our object-based approach is able to achieve more effective solution for privacy setting recommendation, however, its performance is still not comparable with human beings and the reasons for this phenomenon are: (1) The set of privacy-sensitive object classes is not complete (only 268 privacy-sensitive object classes are used in our current work), which may not be able to characterize huge diversity of image privacy (sensitiveness) effectively and efficiently; (2) Because the training images are insufficient and deep learning scheme usually requires huge numbers of training images, the accuracy rates for detecting such privacy-sensitive object classes may not be high enough for us to harvest the advantages of leveraging them for image sensitiveness characterization; (3) Image sensitiveness (privacy) is a very subjective concept, it may largely depend on both the sensitivities of image content and users' personal conservativeness (i.e., different persons may have different privacy preferences). Based on these observations, it is very attractive to develop new algorithms for leveraging more information sources (such as users' personal preferences and keeping users in the loop to define their personalized privacy-sensitive object classes and events) to achieve more effective solutions for fine-grained privacy setting recommendation.

F. User Study

For the same task of privacy setting recommendation, 31 students (16 females and 15 males) are invited to assess the interpret-ability of our object-based approach and our feature-based approach. In our user study, we ask 31 students to score the interpret-ability of our object-based approach and our feature-based approach in 7 levels (6 for the best one and 0 for the worst one). In order to help users understand the correspondences between the image privacy (sensitiveness) and the appearances of privacy-sensitive object classes, as shown in Fig. 21, the privacy-sensitive objects (human faces in this case) are identified and illustrated. Each user is asked to evaluate both our object-based approach and our feature-based approach over at least 20 image sets and each image set contains at most 100 images, and we average his/her scores for all these image sets. As shown in Fig. 22, one can observe that our object-based approach can significantly improve the interpret-ability because the appearances of privacy-sensitive object classes and events in the images have exact and explicit correlations with the image privacy (sensitiveness).

G. Experimental Results on Two Public Image Sets

We have also evaluated our proposed algorithms over two public image sets: PicAlert and Mirflickr [27], [28]. In these two public image sets, we have compared three approaches for privacy setting recommendation: (1) our feature-based approach; (2) our object-based approach; (c) the baseline method by Zerr *et al.* [27] and [28].

We first use the deep network learned from our image set to configure the structure of the deep networks for these two public image sets, and the images from these two public image sets are further used to fine-tune the kernel weights effectively. In addition, we partition the test images into 40 subsets and evaluate each subset independently. As shown in Fig. 23 and Fig. 24, one can observe that our object-based approach can achieve better performance than other two methods (i.e., it may cause lower privacy disclosure for image sharing). The reason for this phenomenon is that: Compared with the 1024-D or 4096-D deep features that are used in our feature-based approach and the baseline method [27], [28], the privacy-sensitive object classes and events (that are used in our object-based approach) have much stronger correlations with the image content sensitiveness (privacy), e.g., their

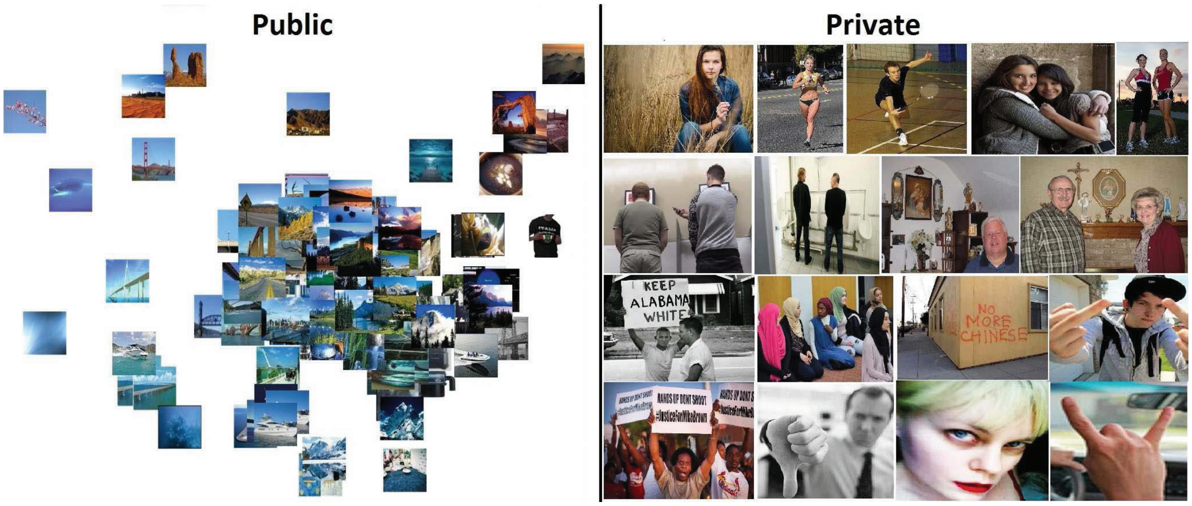


Fig. 20. Our experimental results on privacy-aware image classification, where the images in the public (share) category are visualized according to their visual similarities.

appearances in the images may cause privacy disclosure directly.

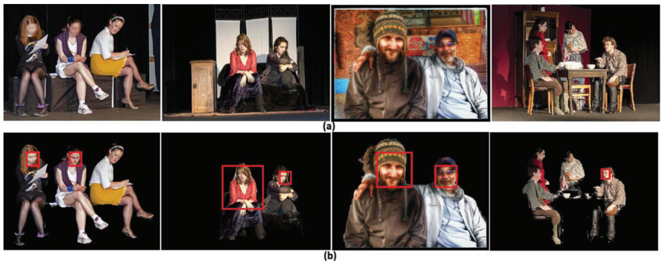


Fig. 21. The privacy-sensitive objects (in red boxes) that are identified from the images.

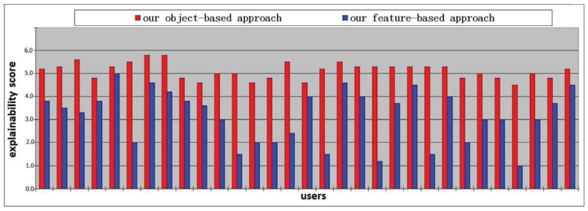


Fig. 22. User evaluation results on the interpret-ability of two approaches for privacy setting recommendation.

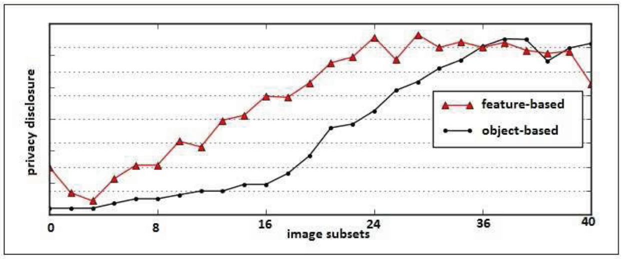


Fig. 23. The comparison between our feature-based approach and our object-based approach for privacy setting recommendation over PicAlert image set.

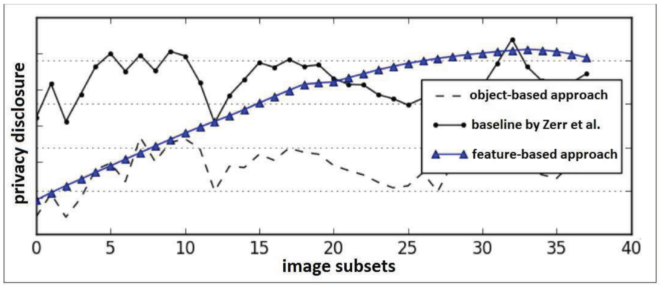


Fig. 24. The comparison between three approaches for privacy setting recommendation over Mirflickr image set: (a) our object-based approach; (b) baseline method by Zerr *et al.* [27] and [28]; (c) our feature-based approach, where 40 image subsets are sorted according to their privacy disclosure obtained by our feature-based approach.

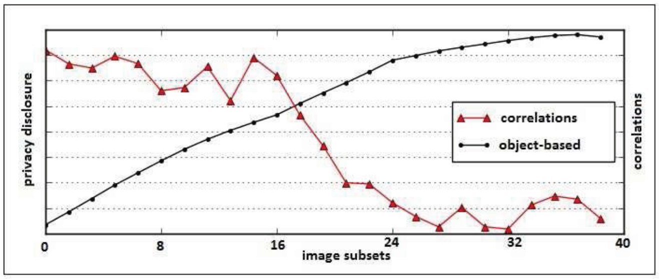


Fig. 25. The effectiveness of using the privacy-sensitive object classes and events for image sensitiveness representation in PicAlert image set: (a) the correlations curve between our 280-D bags of privacy-sensitive object classes and the image content sensitiveness (privacy); (b) the privacy disclosure curve induced by our object-based approach.

The effectiveness of our object-based approach (on recommending appropriate privacy settings for image sharing) largely depends on the correlations between the image privacy (sensitiveness) and the privacy-sensitive object classes and events that are extracted for image content sensitiveness representation. As shown in Fig. 25, when such correlations are low, our object-based approach may induce higher privacy disclosures. The reasons for this phenomenon are: (a) our small set of privacy-sensitive object classes and events could be incomplete because many others may also result in privacy disclosures but they are not detected in our current work; (b) our deep multiple instance learning

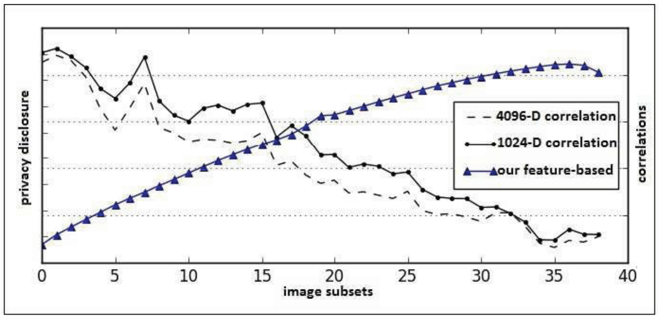


Fig. 26. The effectiveness of using various deep features for image content sensitiveness representation in Mirflickr image set: (a) the correlations between the traditional 4096-D deep features and the image content sensitiveness; (b) the correlations between our 1024-D deep features and the image content sensitiveness; (c) the privacy disclosure induced by our feature-based approach.

algorithm may fail to learn discriminative models to detect these privacy-sensitive object classes accurately; and (c) The privacy-sensitive object classes and events could be user-dependent and users should be involved in the loop to define their personalized privacy-sensitive object classes and events. As shown in Fig. 26, we have also demonstrated similar observation for the feature-based approach, when the correlations between the image privacy (sensitiveness) and the 1024-D or 4096-D deep features for image content sensitiveness representation are low, the feature-based approach may induce higher privacy disclosures. The reason for this phenomenon is that such 1024-D or 4096-D deep features for image content representation may not have exact correlations with the image privacy (sensitiveness).

H. Discussions

Blurring faces may protect image privacy at certain level but it may also raise speculations. Thus one of our future researches is to use GANs [70], [71] to generate perceptually-similar but privacy-free image patches to replace the privacy-sensitive objects in the images be shared while maintaining their local smoothness among various neighboring image components, so that we can protect image privacy effectively while we may not raise speculations.

The privacy-sensitive object classes and events and their definitions are user-dependent and context-dependent. Based on this understanding, the privacy-sensitive object classes and events can be partitioned into two categories: (a) common ones; and (b) user-dependent ones or personalized ones. Our current work (presented in this paper) focuses on the common ones and thus one of our future researches is to involve users in the loop to define their personalized privacy-sensitive object classes and events, and we can also leverage personalized information sources (such as users' personal preferences and user-dependent privacy-sensitive object classes) to recommend fine-grained privacy settings for social image sharing.

The user trustworthiness characterization also plays an important role in supporting fine-grained privacy setting recommendation, thus one of our future researches is to develop new algorithms for achieving more accurate characterization of user trustworthiness: (a) achieving multi-level characterization of user trustworthiness to achieve more accurate assignments of fine-grained privacy settings for social image sharing; (b) using a large number of categories for fine-grained privacy settings and training more discriminative classifiers to achieve better assignments between the images and the users (image-user pairs).

VIII. Conclusions

This paper has developed a new approach to recommend fine-grained privacy settings for social image sharing, where both the image content sensitiveness and the user trustworthiness are simultaneously considered and integrated to train more discriminative tree classifier. Our experimental studies have demonstrated both efficiency and effectiveness of our proposed algorithms.

References

[1] K. Liu and E. Terzi, "A framework for computing the privacy scores of users in online social networks," in *Proc. IEEE ICDM*, Dec. 2009, pp. 228–297.

[2] A. C. Squicciarini, H. Xu, and X. L. Zhang, "CoPE: Enabling collaborative privacy management in online social networks," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 62, no. 3, pp. 521–534, 2011.

[3] N. A. Van House, "Collocated photo sharing, story-telling, and the performance of self," *Int. J. Hum.-Comput. Stud.*, vol. 67, pp. 1073–1086, Dec. 2009.

[4] N. Wang, H. Xu, and J. Grossklags, "Third-party apps on Facebook: Privacy and the illusion of control," in *Proc. ACM CHIMIT*, 2011, Art. no. 4.

[5] H. Hu and G.-J. Ahn, "Multiparty authorization framework for data sharing in online social networks," in *Proc. DBSec*, 2011, pp. 29–43.

[6] A. Besmer and H. R. Lipford, "Moving beyond untagging: Photo privacy in a tagged world," in *Proc. CHI*, 2011, pp. 1563–1572.

[7] O. Nov, M. Naaman, and C. Ye, "Motivational, structural and tenure factors that impact online community photo sharing," in *Proc. ICWSM*, 2009, pp. 138–145.

[8] A. Vahdat and G. Mori, "Handling uncertain tags in visual recognition," in *Proc. IEEE ICCV*, Dec. 2013, pp. 737–744.

[9] L. Chen, D. Xu, I. W. Tsang, and J. Luo, "Tag-based Web photo retrieval improved by batch mode re-tagging," in *Proc. IEEE CVPR*, Jun. 2010, pp. 3440–3446.

[10] K. Q. Weinberger, M. Slaney, and R. Van Zwol, "Resolving tag ambiguity," in *Proc. ACM MM*, 2008, pp. 111–120.

[11] N. Vyas, A. C. Squicciarini, C.-C. Chang, and D. Yao, "Towards automatic privacy management in Web 2.0 with semantic analysis on annotations," in *Proc. IEEE CollaborateCom*, Nov. 2009, pp. 1–10.

[12] A. C. Squicciarini, S. Karumanchi, D. Lin, and N. DeSisto, "Automatic social group organization and privacy management," in *Proc. CollaborateCom*, Oct. 2012, pp. 89–96.

[13] A. C. Squicciarini, D. Lin, S. Sundareswaran, and J. Wede, "Privacy policy inference of user-uploaded images on content sharing sites," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 1, pp. 193–206, Jan. 2014.

[14] A. C. Squicciarini, D. Lin, S. Sundareswaran, and J. Wede, "A3P: Adaptive policy prediction for shared images over popular content sharing sites," in *Proc. 22nd ACM Conf. Hypertext Hypermedia*, 2011, pp. 261–270.

[15] M. de Choudhury, H. Sundaram, Y.-R. Lin, A. John, and D. D. Seligmann, "Connecting content to community in social media via image content, user tags and user communication," in *Proc. IEEE ICME*, Jun. 2009, pp. 1238–1241.

[16] P. Klemperer *et al.*, "Tag, you can see it!: Using tags for access control in photo sharing," in *Proc. CHI*, 2012, pp. 377–386.

[17] R. Ravichandran, M. Benisch, P. G. Kelley, and N. M. Sadeh, "Capturing social networking privacy preferences," in *Proc. SOUPS*, 2009, pp. 1–18.

[18] C.-M. Yeung, L. Kagal, N. Gibbins, and N. Shadbolt, "Providing access control to online albums based on tags and linked data," in *Proc. AAAI Symp.*, 2009, pp. 9–14.

[19] G. Denezis, "Inferring privacy policies for social networking services," in *Proc. 2nd Workshop Secur. Artif. Intell.*, 2009, pp. 5–10.

[20] F. Adu-Opong, C. K. Gardiner, A. Kapadia, and P. P. Tsang, "Social circles: Tackling privacy in social networks," in *Proc. Symp. Usable Privacy Secur. (SOUPS)*, 2008, pp. 1–2.

[21] L. Fang and K. LeFevre, “Privacy wizards for social networking sites,” in *Proc. ACM WWW*, 2010, pp. 351–360.

[22] B. Krishnamurphy and C. E. Wills, “Characterizing privacy in online social networks,” in *Proc. 1st Workshop Online Soc. Netw.*, 2008, pp. 37–42.

[23] A. Simpson, “On the need for user-defined fine-grained access control policies for social networking applications,” in *Proc. Workshop Secur. Opportunities Soc. Netw.*, 2008, Art. no. 1.

[24] K. Ghazinour, S. Matwin, and M. Sokolova, “Monitoring and recommending privacy settings in social networks,” in *Proc. EDBT/ICDT*, 2013, pp. 164–168.

[25] J. McAuley and J. Leskovec, “Learning to discover social circles in ego networks,” in *Proc. NIPDS*, 2012, pp. 539–547.

[26] G. Petkos, S. Papadopoulos, and Y. Kompatsiaris, “Social circle discovery in ego-networks by mining the latent structure of user connections and profile attributes,” in *Proc. IEEE/ACM Int. Conf. Adv. Soc. Netw. Anal. Mining (ASONAM)*, Aug. 2015, pp. 880–887.

[27] S. Zerr, S. Siersdorfer, J. Hare, and E. Demidova, “Privacy-aware image classification and search,” in *Proc. ACM SIGIR*, 2012, pp. 35–44.

[28] S. Zerr, S. Siersdorfer, and J. Hare, “PicAlert!: A system for privacy-aware image classification and retrieval,” in *Proc. ACM CIKM*, 2012, pp. 2710–2712.

[29] A. C. Squicciarini, C. Caragea, and R. Balakavi, “Analyzing images’ privacy for the modern Web,” in *Proc. ACM Hypertext*, 2014, pp. 136–147.

[30] A. Tonge and C. Caragea. (2015). “Privacy prediction of images shared on social media sites using deep features.” [Online]. Available: https://arxiv.org/abs/1510.08583

[31] D. Buschek, M. Bader, E. van Zezschwitz, and A. de Luca, “Automatic privacy classification of personal photos,” in *Proc. LNCS*, vol. 9297. 2015, pp. 428–435.

[32] E. Spyromitros-Xioufis, S. Papadopoulos, A. Popescu, and Y. Kompatsiaris, “Personalized privacy-aware image classification,” in *Proc. ACM ICMR*, 2016, pp. 71–78.

[33] J. Yu, B. Zhang, Z. Kuang, D. Lin, and J. Fan, “iPrivacy: Image privacy protection by identifying sensitive objects via deep multi-task learning,” *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 5, pp. 1005–1016, May 2017.

[34] F. Dufaux and T. Ebrahimi, “Scrambling for privacy protection in video surveillance systems,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 8, pp. 1168–1174, Aug. 2008.

[35] M.-R. Ra, R. Govindan, and A. Ortega, “P3: Toward privacy-preserving photo sharing,” in *Proc. 10th USENIX Symp. Netw. Design Implement.*, 2013, pp. 515–528.

[36] W. Zhang, S. S. Cheung, and M. Chen, “Hiding privacy information in video surveillance system,” in *Proc. IEEE ICIP*, Sep. 2005, pp. II-868–II-871.

[37] G. Friedland and R. Sommer, “Cybercasing the Joint: On the privacy implications of geo-tagging,” in *Proc. USENIX Workshop Hot Topics Secur. (HotSec)*, 2010, pp. 1–6.

[38] Y. Nakashima, N. Babaguchi, and J. Fan, “Privacy protection for social video via background estimation and CRF-based videographer’s intention modeling,” *IEICE Trans. Inf. Syst.*, vol. E99-D, no. 4, pp. 1221–1233, 2016.

[39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Proc. NIPS*, 2012, pp. 1097–1105.

[40] Y. Jia *et al.*, “Caffe: Convolutional architecture for fast feature embedding,” in *Proc. ACM MM*, 2014, pp. 675–678.

[41] J. Donahue *et al.*, “DeCAF: A deep convolutional activation feature for generic visual recognition,” in *Proc. ICML*, 2014, pp. I-647–I-655.

[42] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *Proc. ICLR*, 2015, pp. 1–29.

[43] C. Szegedy *et al.*, “Going deeper with convolutions,” in *Proc. IEEE CVPR*, Jun. 2015, pp. 1–9.

[44] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification,” in *Proc. ICCV*, Dec. 2015, pp. 1026–1034.

[45] L.-C. Chen, G. Papandreou, I. Kokkions, K. Murphy, and A. L. Yuille, “Semantic image segmentation with deep convolutional nets and fully connected CRFs,” in *Proc. ICLR*, 2015, pp. 1–14. [Online]. Avalilable: https://arxiv.org/pdf/1412.7062

[46] B. Hariharan, P. Arbelaez, R. Girshick, and J. Malik, “Hypercolumns for object segmentation and fine-grained localization,” in *Proc. IEEE CVPR*, Jun. 2015, pp. 447–456.

[47] P. O. Pinheiro and R. Collobert, “From image-level to pixel-level labeling with convolutional networks,” in *Proc. IEEE CVPR*, Jun. 2015, pp. 1713–1721.

[48] S. Zheng *et al.* (Feb. 2015). “Conditional random fields as recurrent neural networks.” [Online]. Available: https://arxiv.org/abs/1502.03240

[49] J. Long, E. Shelhamer, and T. Darrel, “Fully convolutional networks for semantic segmentation,” in *Proc. IEEE CVPR*, Jun. 2015, pp. 3431–3440.

[50] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” in *Proc. ICML*, 2001, pp. 282–289.

[51] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[52] L.-J. Li, H. Su, L. Fei-Fei, and E. P. Xing, “Object bank: A high-level image representation for scene classification & semantic feature sparsification,” in *Proc. NIPS*, 2010, pp. 1378–1386.

[53] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[54] D. Bau, B. Zhou, A. Khosla, A. Oliva, and A. Torralba, “Network dissection: Quantifying interpretability of deep visual representations,” in *Proc. IEEE CVPR*, Jul. 2017, pp. 3319–3327.

[55] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez, “Solving the multiple instance problem with axis-parallel rectangles,” *Artif. Intell.*, vol. 89, nos. 1–2, pp. 31–71, 1997.

[56] D. Kotzias, M. Denil, N. de Freitas, and P. Smyth, “From group to individual labels using deep features,” in *Proc. ACM SIGKDD*, 2015, pp. 597–606.

[57] J. Wu, Y. Yu, C. Huang, and K. Yu, “Deep multiple instance learning for image classification and auto-annotation,” in *Proc. IEEE CVPR*, Jun. 2015, pp. 3460–3469.

[58] D. Pathak, E. Shelhamer, J. Long, and T. Darrell, “Fully convolutional multi-class multiple instance learning,” in *Proc. ICLR*, 2015, pp. 1–4.

[59] *Walmart Wants to Monitor Shoppers’ Facial Expressions*. Accessed: 2017. [Online]. Available: https://www.usatoday.com/story/money/2017/08/08/

[60] L. Jedrzejczyk, B. A. Price, A. K. Bandara, and B. Nuseibeh, “I know what you did last summer: Risks of location data leakage in mobile and social computing,” Open Univ., Milton Keynes, U.K., Tech. Rep. TR2009-11, 2009.

[61] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang, “Large-scale visual sentiment ontology and detectors using adjective noun pairs,” in *Proc. 21st ACM Int. Conf. Multimedia*, 2013, pp. 223–232.

[62] Y. Li, D. J. Crandall, and D. P. Huttenlocher, “Landmark classification in large-scale image collections,” in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 1957–1964.

[63] Z. Stone, T. Zickler, and T. Darrell, “Autotagging Facebook: Social network context improves photo annotation,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2008, pp. 1–8.

[64] Y.-T. Zheng *et al.*, “Tour the world: Building a Web-scale landmark recognition engine,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 1085–1092.

[65] B. C. Becker and E. G. Ortiz, “Evaluation of face recognition techniques for application to facebook,” in *Proc. 8th IEEE Int. Conf. Auto. Face Gesture Recognit. (FG)*, Sep. 2008, pp. 1–6.

[66] O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition,” in *Proc. BMVC*, Sep. 2015, vol. 1. no. 3, pp. 1–12.

[67] R. Jenkins and A. M. Burton, “100% accuracy in automatic face recognition,” *Science*, vol. 319, no. 5862, p. 435, 2008.

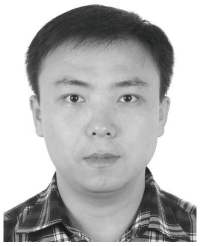
[68] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, “DeepFace: Closing the gap to human-level performance in face verification,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1701–1708.

[69] D. Pathak, P. Kruahenbuhl, J. Donahue, T. Darrell, and A. Efros. (Apr. 2016). “Context encoders: Feature learning by inpainting.” [Online]. Available: https://arxiv.org/abs/1604.07379

[70] I. J. Goodfellow *et al.* (Jun. 2014). “Generative adversarial networks.” [Online]. Available: https://arxiv.org/abs/1406.2661

[71] P. Isola, J.-Z. Zhu, T. Zhou, and A. A. Efros. (Nov. 2016). “Image-to-image translation with conditional adversarial networks.” [Online]. Available: https://arxiv.org/abs/1611.07004

Jun Yu (M'3) received the B.Eng. and Ph.D. degrees from Zhejiang University, Zhejiang, China. He was an Associate Professor with the School of Information Science and Technology, Xiamen University. From 2009 to 2011, he was with Nanyang Technological University, Singapore. From 2012 to 2013, he was a Visiting Researcher with Microsoft Research Asia. He was a short-term Visiting Scholar with UNC-Charlotte. He is currently a Professor with the School of Computer Science and Technology, Hangzhou Dianzi University. He has authored or co-authored over 50 scientific articles. His research interests include multimedia analysis, machine learning, image processing, and image privacy protection. He is a Professional Member of the ACM and CCF. He has (co-)chaired several special sessions, invited sessions, and workshops. He served as a program committee member or reviewer for top conferences and prestigious journals.



Zhenzhong Kuang received the B.S. and Ph.D. degrees in computer science from the China University of Petroleum, Tsingdao, China, in 2012 and 2017, respectively. He was a Visiting Student with UNC-Charlotte. He is currently an Assistant Professor with the School of Computer Science and Technology, Hangzhou Dianzi University. His research interests include computer vision, machine learning, and image privacy protection.

Baopeng Zhang received the Ph.D. degree in computer science from Tsinghua University in 2008. He was a Visiting Scholar with UNC-Charlotte from 2015 to 2016. He is currently an Associate Professor with the School of Computer and Information Technology, Beijing Jiaotong University, China. His research interests include semantic image/video classification and retrieval, statistical machine learning, large-scale semantic data management and analysis, and image privacy protection.

Wei Zhang received the B.A. and M.A. degrees in economics and the Ph.D. degree in computer science from Fudan University, Shanghai, China, in 2000, 2003, and 2008, respectively. He is currently an Associate Professor with the School of Computer Science, Fudan University, and also a Visiting Scholar with the Computer Science Department, UNC-Charlotte. His current research interests include machine learning, computer vision, and deep neural network.

Dan Lin received the Ph.D. degree in computer science from the National University of Singapore in 2007. She was a Post-Doctoral Research Associate with Purdue University for two years. She is currently an Associate Professor and the Director of Cybersecurity Laboratory, Missouri University of Science and Technology. Her main research interests cover many areas in the fields of database systems and information security.

Jianping Fan received the M.S. degree in theory physics from Northwestern University, Xian, China, in 1994, and the Ph.D. degree in optical storage and computer science from the Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai, China, in 1997. He was a Post-Doctoral Researcher with Fudan University, Shanghai, from 1997 to 1998. From 1998 to 1999, he was a Researcher with the Department of Information System Engineering, Japan Society of Promotion of Science, Osaka University, Osaka, Japan. From 1999 to 2001, he was a Post-Doctoral Researcher with the Department of Computer Science, Purdue University, West Lafayette, IN, USA. He is currently a Professor with UNC-Charlotte. His research interests include image/video privacy protection, automatic image/video understanding, and large-scale deep learning.