KANSAS STATE UNIVERSITY | College of Engineering
Department of Computer Science

Uncovering Scene Context for Predicting Privacy of Online Shared Images

Ashwini Tonge[1], Cornelia Caragea[1] and Anna Squicciarini[2]

[1]Kansas State University, [2]Pennsylvania State University

Image Privacy Prediction?

- Rapid increase in social media can cause threat to user's privacy
- Many users are quick to share private images without realizing the consequences of an unwanted disclosure of these images.
- Users rarely change default privacy settings, which could jeopardize their privacy [Zerr et al., 2012].
- Current social networking sites do not assist users in making privacy decisions for images that they share online.
- An image Privacy Prediction system predicts the privacy setting for images and avoid a possible loss of users' privacy.

Prior Works

- Recently, Squicciarini et al. [Squicciarini et al., 2014] and Zerr et al. [Zerr et al., 2012] found that user tags are informative for classifying images as *private* or *public*.
- However, since user tags are at the sole discretion of users, they typically tend to be noisy and incomplete.
- Tonge and Caragea [Tonge and Caragea, 2016] automatically derived object tags from images' content using Convolutional Neural Networks (CNN) and showed that the combination of object tags and user tags outperforms each set of tags individually.

Motivation

Our Contributions

- We propose the extraction of scene-centric tags to capture additional information from the visual content that is not captured by existing object-centric tags.
- We show that scene tags are able to learn privacy characteristics to make appropriate predictions.
- We explore the combination of user tags with object, scene and object-scene tags for privacy prediction.
- Our results show that the combination of all three types of tags (object, scene, and user) yields better performance compared with user tags and the combination of user tags with scene or object tags.

Datasets

- We evaluate the proposed features on a subset of Flickr images sampled from the PicAlert dataset [Zerr et al., 2012].
- PicAlert consists of Flickr images on various subjects, which are manually labeled as *public* or *private* by external viewers.
- We select 32000 images from PicAlert randomly, out of which 10000 images are used as Train set and 22000 images are used as Test set.
- The public and private images are in the ratio of 3:1.

Private: Private image discloses sensitive information about a user. E.g., images with self-portraits, family, friends, someone's home, etc.

Public: Public images generally depict scenery, objects, animals, etc., which do not provide any personal information about a user.

Important Links

https://goo.gl/bPruf6

Image privacy prediction using tag features.

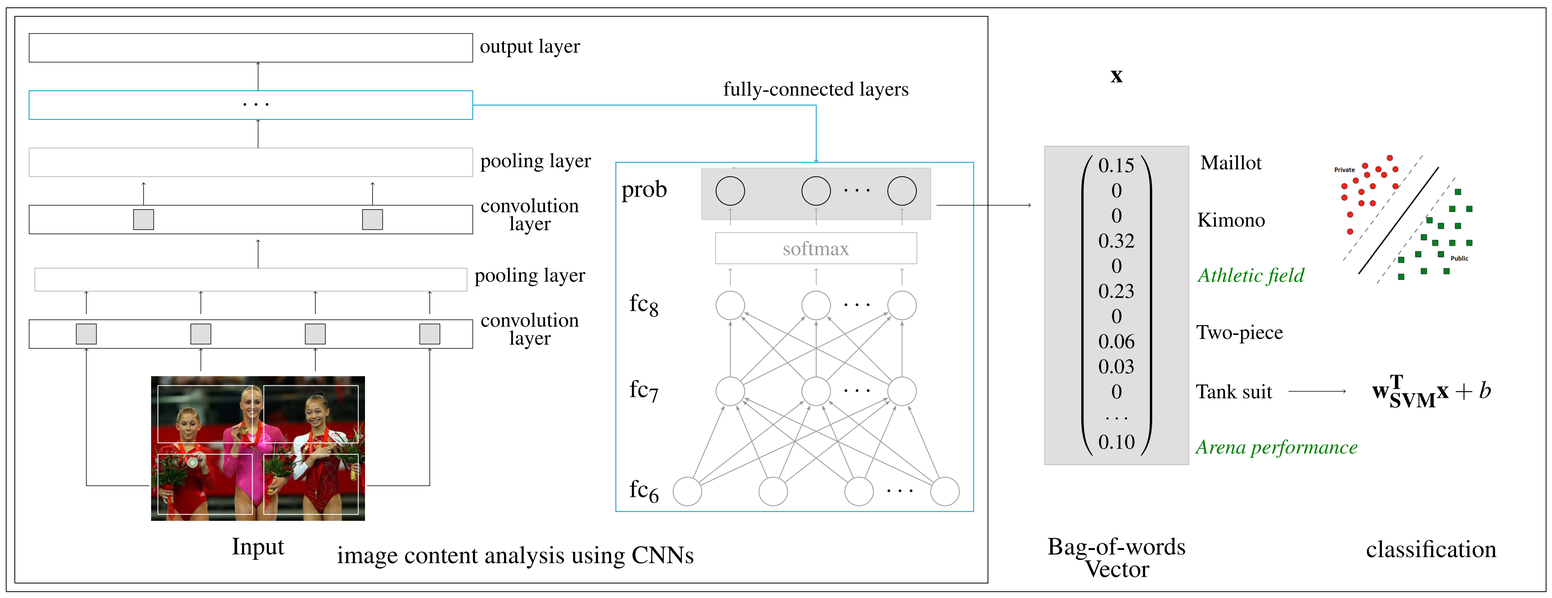

Figure: Image privacy prediction using tag features. 1. CNNs are used to extract object and scene tags (shown in green and italics) for input images. 2. The bag-of-word vectors for the tag features are used to predict the class of an image as public or private using Support Vector Machine (SVM) classifier.

Observations

What type of privacy-aware information do user tags contain?

Rank	Tags				
1-5	people	portrait	**outdoor**	girl	woman
6-10	**landscape**	smile	architecture	plant	building
11-15	**nature**	boy	sky	hair	**wedding**
16-20	family	bride	bird	animal	flower
21-25	eyes	child	happy	**snow**	**indoor**
26-30	cloud	**party**	illustration	drawing	**winter**

Table: Tags having high information gain.

Proposed Features: Privacy Prediction

- Feature Extraction

  We believe that scene tags can contribute along with object tags to learn privacy characteristics of a given image, as they can help provide clues into what the image owners intended to show through the photo. Therefore, we employ two types of semantic features for privacy prediction based on:

  - **Object-centric Tags**
  - We use CNN pre-trained on a large scale object dataset (ImageNet) [Russakovsky et al., 2015], to capture the objects depicted in the image.
  - We use the probability distribution over 1000 object categories for the input image obtained by applying the softmax function over the last fully-connected layer of the AlexNet CNN [Krizhevsky et al., 2012].
  - We consider the top k objects of highest probabilities as *object tags*.

  - **Scene-centric Tags**
  - We use CNN pre-trained on a large scale scene dataset (Places2) [Zhou et al., 2016], to obtain the scene context of the image.
  - we obtain the top k scenes derived from the probability distribution over 365 scene categories of the AlexNet CNN.
  - We refer to the top k predicted scenes as *scene tags*.

- Feature Representations for Classification
  - To encode the scene and object tags, we use the probability of the tag obtained from the softmax layer of the corresponding CNN.
  - We also consider user tags and encode them using a binary representation.
  - Using these feature representations, we train maximum margin (SVM) classifiers to predict the class of an image as *private* or *public*

Object Tags, Scene Tags & User Tags

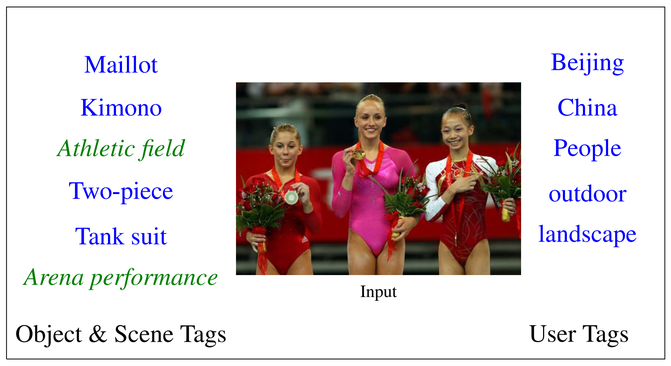

Figure: Object, Scene and User tags for the input image.

Experiments and Results

Would scene-centric tags obtained from the visual content bring additional information to improve privacy prediction?

Features	Acc %	F1	Precision	Recall	#IncPred
UT	81.73	0.789	0.803	0.817	-
$k = 2$					
UT+ST	82.26	0.797	0.81	0.823	293
UT+OT	83.09	0.812	0.819	0.831	477
UT+ST+OT	**83.59**	**0.819**	**0.825**	**0.836**	**587**
$k = 10$					
UT+ST	83.21	0.814	0.821	0.832	503
UT+OT	84.35	0.833	0.834	0.843	755
UT+ST+OT	**84.80**	**0.841**	**0.84**	**0.848**	**854**

Table: Object Tags vs. Scene Tags. The best performance is shown in bold.

Results

Features			
UT	public	*public*	public
UT+OT	*private*	private	public
UT+ST	public	private	*private*
UT+ST+OT	*private*	*public*	*private*

Figure: Results obtained using tag features.

Conclusions

- We proposed the use of scene-centric tags (along with user tags and object tags) and showed that they can improve image privacy prediction.
- The results show that adding scene tags to user tags improves the performance over user tags alone.
- The best performance is achieved when we consider the combination of user, scene, and object tags.
- We conclude that scene and object tags complement each other and help boost the performance.
- These automatically derived tags can provide the relevant cues for privacy-aware image retrieval and can become an essential tool for surfacing the hidden content of the deep web without exposing sensitive details.
- In the future, more sophisticated methods to combine objects and scenes can be explored.

References

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012).
Imagenet classification with deep convolutional neural networks.
In *NIPS'12*, pages 1097–1105.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015).
ImageNet Large Scale Visual Recognition Challenge.
IJCV, pages 1–42.

Squicciarini, A. C., Caragea, C., and Balakavi, R. (2014).
Analyzing images' privacy for the modern web.
HT '14, pages 136–147.

Tonge, A. K. and Caragea, C. (2016).
Image privacy prediction using deep features.
In *AAAI' 16*, pages 4266–4267.

Zerr, S., Siersdorfer, S., Hare, J., and Demidova, E. (2012).
Privacy-aware image classification and search.
In *ACM SIGIR' 12*.

Zhou, B., Khosla, A., Lapedriza, A., Torralba, A., and Oliva, A. (2016).
Places: An image database for deep scene understanding.
arXiv preprint arXiv:1610.02055.

Acknowledgements: This research is supported by a grant from NSF to Cornelia Caragea and Anna Squicciarini.

WWW: people.cs.ksu.edu/~ccaragea