Kansas State University | College of Engineering, Department of Computer Science

Identifying Private Content for Online Image Sharing

Ashwini Tonge
Kansas State University

Image Privacy Prediction & Prior Works

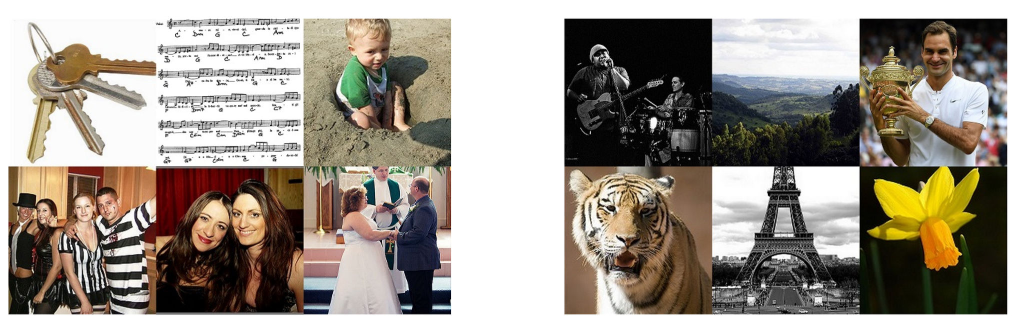
- An image Privacy Prediction system predicts the privacy setting for images and avoid a possible loss of users' privacy.
- Prior works explored models based on user tags and image content features such as SIFT (Scale Invariant Feature Transform) and RGB (Red Green Blue) [Zerr et al., 2012, Squicciarini et al., 2014] for privacy prediction.
- These studies found that users tags are informative and perform better than image content features such as SIFT.
- Recently, due to the success of object recognition from images using CNN [Krizhevsky et al., 2012], researchers started to investigate learning models of image privacy based on CNN [Tran et al., 2016, Tonge and Caragea, 2016].
- Tran et al. proposed privacy framework that combines features obtained from the two CNNs: one that extracts convolutional features, and another that derives object features.

My Contributions

- I aim to solve the problem of identifying private content for online image sharing.
- I derive features from the multi-modal information of the image that can adequately understand the image content and predict the prevalent privacy settings for uploaded images.
  - Since identifying sensitive content is inherently difficult because it requires the system to have an in-depth understanding of the visual content of the image.
- I propose to derive image tags, and visual content features by leveraging CNN architectures which are used in conjunction with machine learning classifiers to identify sensitive content accurately.
- I show empirically on a real world Flickr dataset that the deep features outperform:
  - Existing state-of-the-art models for image privacy prediction.
  - A rule-based learner that predicts an image as private if it contains people's faces.

Datasets

- I 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.
- I consider 32000 images randomly selected from PicAlert for the privacy prediction task.
- The public and private images are in the ratio of 3:1.



(a) Private (b) Public

Figure: Examples of private and public images from PicAlert dataset.

Features for Image Privacy Prediction

The features used in the classification are described below.

- **Deep features**
- Given the strengths of deeper CNN architectures for object recognition, features derived from the deep layers of the very deep CNNs provide finer clues for the image privacy prediction task.
- I employ very deep CNN architectures, i.e., ResNet, GoogLeNet, VGG and AlexNet to derive features from the various layers of these CNNs.
- **Semantic features**
- I believe that scene features can contribute along with object features 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.
- I employ two types of semantic features for privacy prediction: (1) objects features; and (2) scene features.
- **Privacy-aware User Tags**
- I propose privacy-aware tag recommendation algorithm that aims at improving the quality of user annotations while also preserving the images' original sharing settings.
- These improved set of tags can improve the privacy prediction performance.
- **Multimodal feature fusion**
- Finally, I propose an algorithm to combine the strengths of tags features, semantic (object and scene) features and privacy-specific features to improve privacy prediction. This work is currently under development.

Deep Features

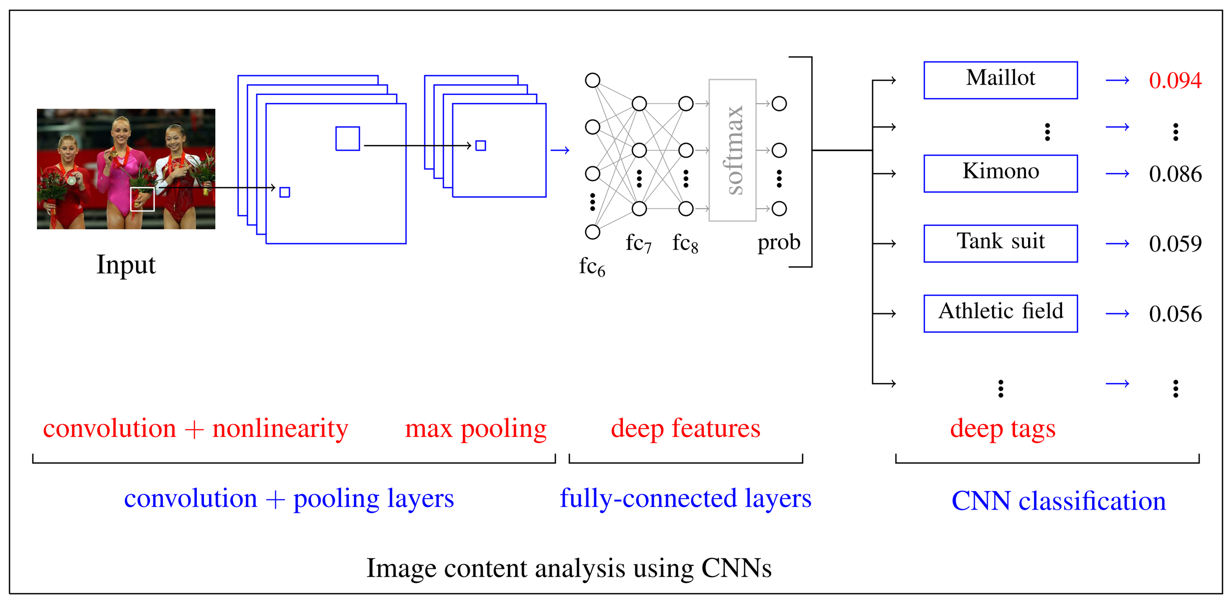


Figure: Deep Features: CNNs are used to extract deep visual features and deep image tags for input images.

Feature Classification

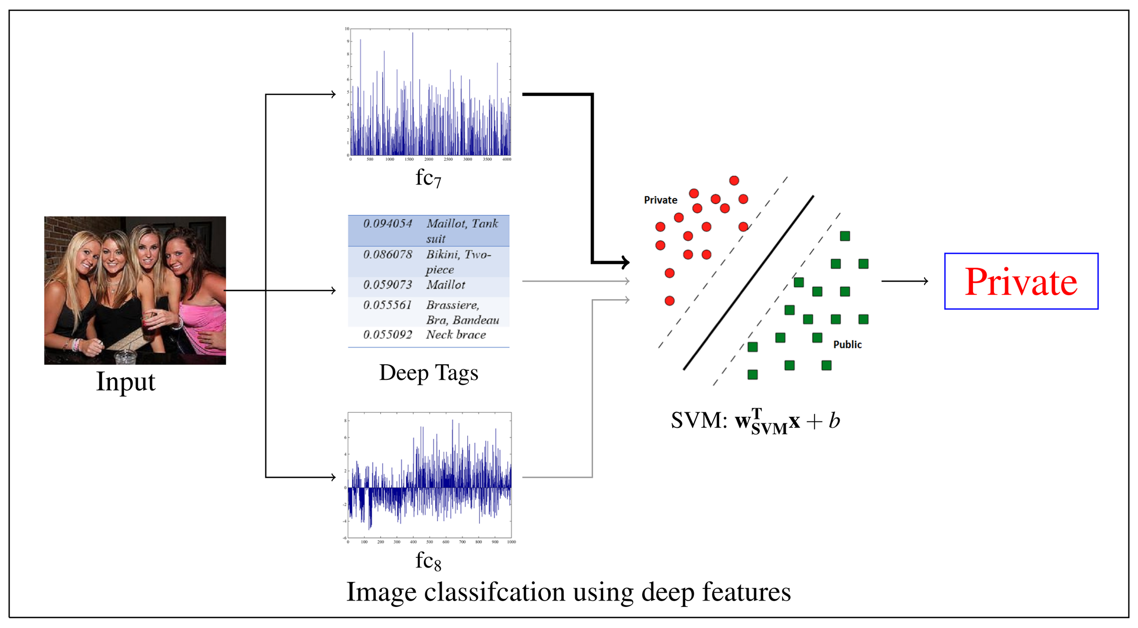


Figure: Feature Classification (Deep Features and Deep Tags): The features from the fully-connected (fc) layers and deep tags are used to predict the class of an image as public or private using SVM.

Semantic features

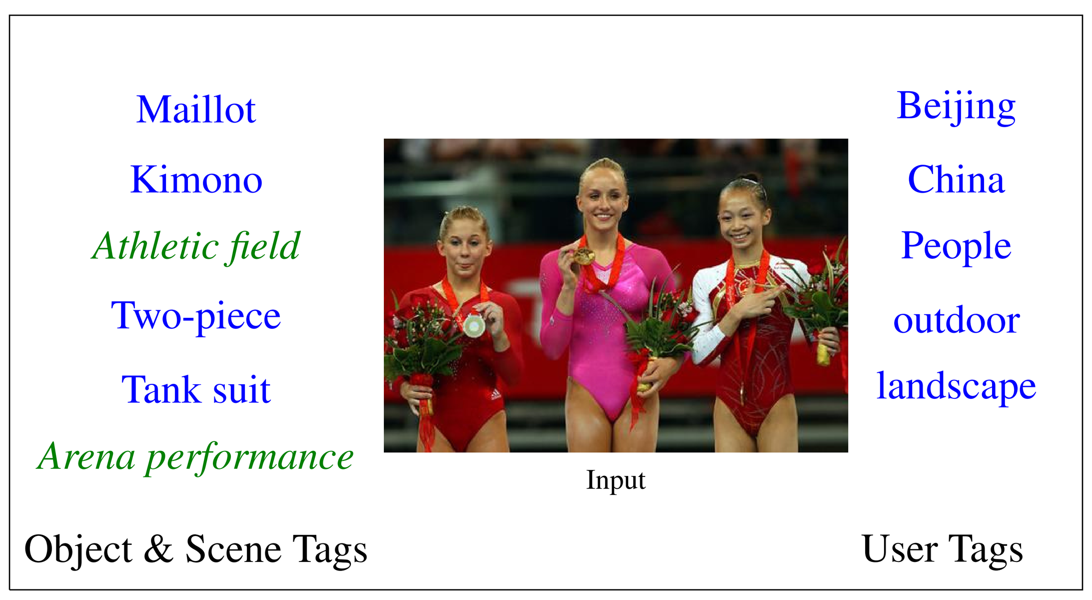


Figure: Object, Scene and User tags for the input image.

Privacy-aware User Tags

- I posit that visually similar images can possess very different sets of user-input tags if these images have different privacy orientations.
- Intuitively, user-input tags provide users' intention behind sharing the image which can vary based on whether the image to be shared with everyone on the web or not.
  - Yet, prior image tagging systems failed to consider the privacy aspect of an image.
- I present a collaborative filtering based approach to privacy-aware image tagging.

(a) *Private*, Stylish, Elegant Corporate, Style, Pretty Fashion, Girl, Woman (b) *Public*, Parisi, Sabrina News, Celebrity, Woman Famous, Girl, Hollywood

Figure: Anecdotal evidence for visually similar images with privacy-aware user tags.

Important Links

https://goo.gl/HFRmwU

Experiments and Results

What is the impact of the network architecture on the privacy prediction?

Features	Acc %	F1	Prec	Re
AlexNet				
fc_6	82.29	0.82	0.819	0.823
fc_7	82.97	0.827	0.825	0.83
fc_8	85.51	0.849	0.849	0.855
prob-A	82.76	0.815	0.816	0.828
GoogLeNet				
$pool_5$	86.41	0.861	0.86	0.864
$loss_3$	86.42	0.861	0.86	0.864
prob-G	82.66	0.815	0.816	0.827
VGG				
fc_6-V	83.85	0.837	0.836	0.839
fc_7-V	84.43	0.843	0.842	0.844
fc_8-V	86.72	0.864	0.863	0.867
prob-V	81.72	0.801	0.804	0.817
ResNet				
fc-R	**87.58**	**0.872**	**0.872**	**0.876**
prob-R	80.6	0.784	0.789	0.806

Table: Comparison of pre-trained architectures AlexNet, GoogLeNet, VGG and ResNet.

How do DeepPrivate features perform as compared to baselines?

Features	Acc %	F1	Prec	Re
Deep features				
fc-R	**87.58**	**0.872**	**0.872**	**0.876**
Hierarchical Deep Features [Tran et al., 2016]				
PCNH	83.13	0.824	0.823	0.831
AlexNet Deep Features [Tonge and Caragea, 2016]				
fc_8	85.51	0.849	0.849	0.855
SIFT/GIST [Zerr et al., 2012, Squicciarini et al., 2014]				
SIFT+GIST	72.67	0.704	0.691	0.727
Rule-based models				
Rule-1	77.35	0.683	0.694	0.672
Rule-2	77.93	0.673	0.704	0.644

Table: Deep features vs. Baselines.

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.

Tag analysis

Rank 1-10	Rank 11-20	Rank 21-30	Rank 31-40	Rank 41-50
people	pyjama	maillot	**promontory**	jersey
wig	jammies	**girl**	t-shirt	mole
portrait	sweatshirt	suit of clothes	foreland	groin
bow-tie	**outdoor**	ice lolly	**headland**	bulwark
neck brace	**lakeside**	suit	bandeau	seawall
groom	**lakeshore**	lollipop	miniskirt	**seacoast**
bridegroom	sun blocker	two-piece	breakwater	**indoor**
laboratory coat	sunscreen	tank suit	**vale**	stethoscope
hair spray	sunglasses	bikini	hand blower	**valley**
shower cap	military uniform	swimming cap	**jetty**	**head**

Table: Top 50 highly informative tags.

Conclusions

- I employ deep features depicting multimodal information of an image derived through CNN networks to understand the images' content in-depth for image privacy classification.
- The results show the remarkable improvements in performance of image privacy prediction when using deep features as compared to baselines.
- In future, with the help of these features, it would be interesting to explore learning models for personalized image privacy prediction with varying degree of sensitivity.

References

- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *NIPS'12*, pages 1097–1105.
- 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.
- Tran, L., Kong, D., Jin, H., and Liu, J. (2016). Privacy-cnh: A framework to detect photo privacy with convolutional neural network using hierarchical features. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, pages 1317–1323. AAAI Press.
- Zerr, S., Siersdorfer, S., Hare, J., and Demidova, E. (2012). Privacy-aware image classification and search. In *ACM SIGIR' 12*.

Acknowledgements: I would like to thank my advisor Dr. Cornelia Caragea for guidance and Dr. Anna Squicciarini for helpful discussions. This research is supported by NSF.

LinkedIn: ashwini-tonge-66993516