UNT
UNIVERSITY OF NORTH TEXAS®

Image Privacy Prediction Using Deep Features

Ashwini Tonge and Cornelia Caragea
Computer Science and Engineering, University of North Texas

Why 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.
- Manually assigning privacy settings to each image every time can be cumbersome.
- Image Privacy Prediction predicts privacy setting for images and avoid a possible loss of users' privacy.

Previous Approaches to Image Privacy Prediction

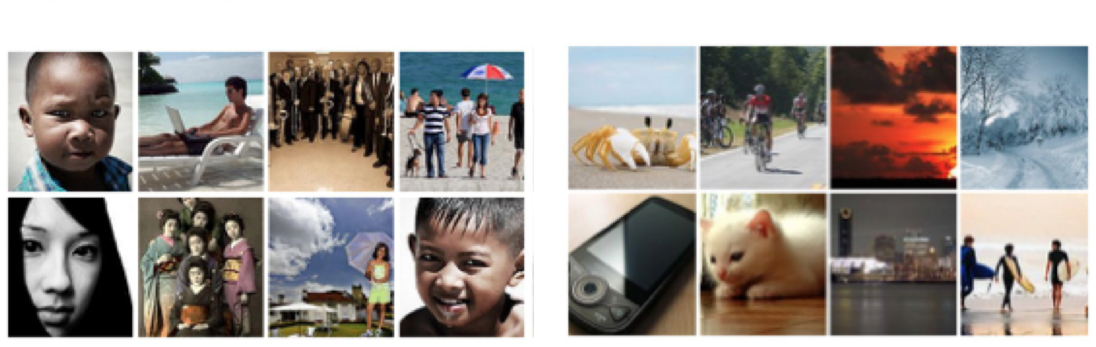
- Most existing privacy prediction techniques used user tags and image content features such as SIFT (or Scale Invariant Feature Transform) and RGB (or Red Green Blue) [Zerr et al., 2012, Squicciarini et al., 2014]
- Buschek et al. [Buschek et al., 2015] presented an approach to assigning privacy settings to shared images using metadata (location, time, shot details) and visual features (faces, colors, edges).
- Several works were conducted in the context of tag-based access control policies for images [Yeung et al., 2009, Klemperer et al., 2012, Vyas et al., 2009]
  - However, the scarcity of tags [Sundaram et al., 2012] precluded accurate analysis of images' sensitivity.
- We posit that, given large dataset of labeled images e.g., the ImageNet dataset [Russakovsky et al., 2015], user tags and SIFT features may not work well. However, deep neural networks are now able to learn powerful deep features [Jia et al., 2014] that go beyond SIFT and RGB, and have potential to improve privacy prediction.

Our Contributions

- In this study, we explore an approach to image privacy prediction based on **deep visual features** and **deep tags**.
- Empirically, deep features and deep tags outperforms baseline approaches SIFT, GIST, and user provided tags.
- Models trained on "SIFT" and "GIST" yield very low performance with respect to the private class.
- Combination of deep tags and user tags performs better than their individual performance.
- We evaluate our approach on Flickr images sampled from the PiCalert dataset [Zerr et al., 2012].
- Tag analysis can assist in understanding the characteristics of the private and public classes.

Datasets

- We evaluated our approach 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 selected 5,000 images from PiCalert randomly, out of which only 4,700 have user provided tags and these 4,700 images were used for our 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.

Private: Private image discloses sensitive information about a user. E.g., images with portraits, people on the beach, family photos, etc.

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

Proposed Approach: Privacy Prediction

- Feature Extraction
  We extracted visual features and tags for differentiating between private and public classes.
  - **Deep Visual Features**
  - In the convolutional neural network (CNN) architecture, features are extracted from images through each layer in a feed-forward fashion.
  - The architecture consists of eight layers; the first five layers are convolutional and the remaining three are fully-connected (FC).
  - The last two fully connected layers are referred as FC_7 and FC_8. and used as *deep visual features* for images.
  - The output layer "Prob" is obtained from the output of FC_8 via a softmax function, which produces a probability distribution over the 1000 object categories.
  - **Deep Tag Features**
  - for an image, we predict top k object categories from the probability distribution over categories, i.e., the "Prob" layer of the deep neural network.
  - The k predicted categories are used as tags to describe an image.
- Feature Classification
  Using above feature representations, we train maximum margin (SVM) classifiers and use them to predict the class of an image as *private* or *public*

Proposed Approach: Feature Extraction

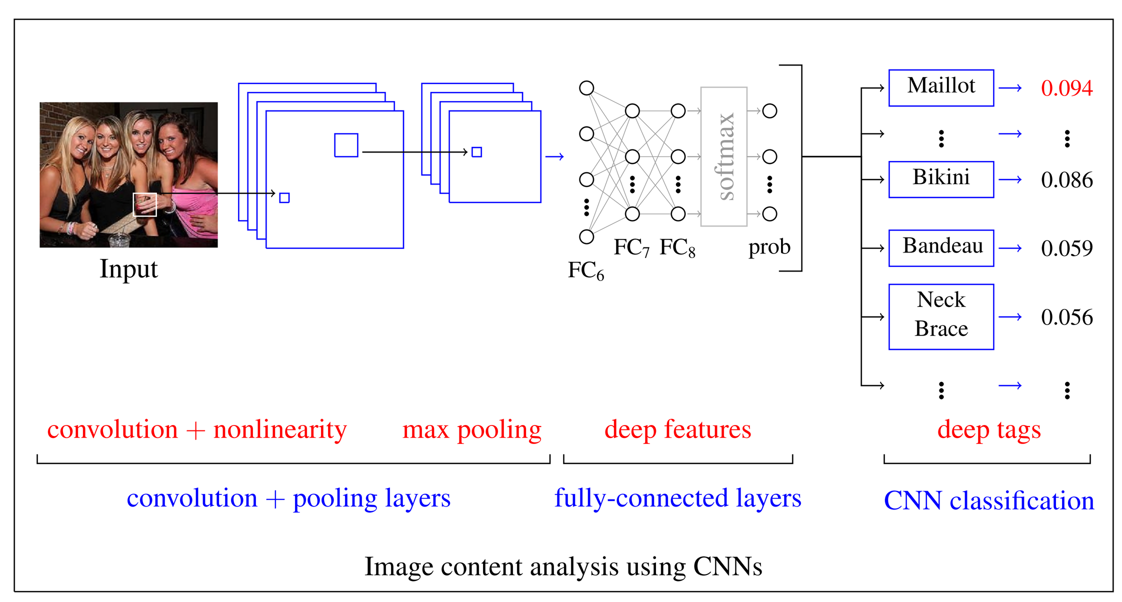


Figure: Proposed approach - Feature Extraction (Deep Features and Deep Tags): CNNs are used to extract deep visual features and deep image tags for input images.

Proposed Approach: Feature Classification

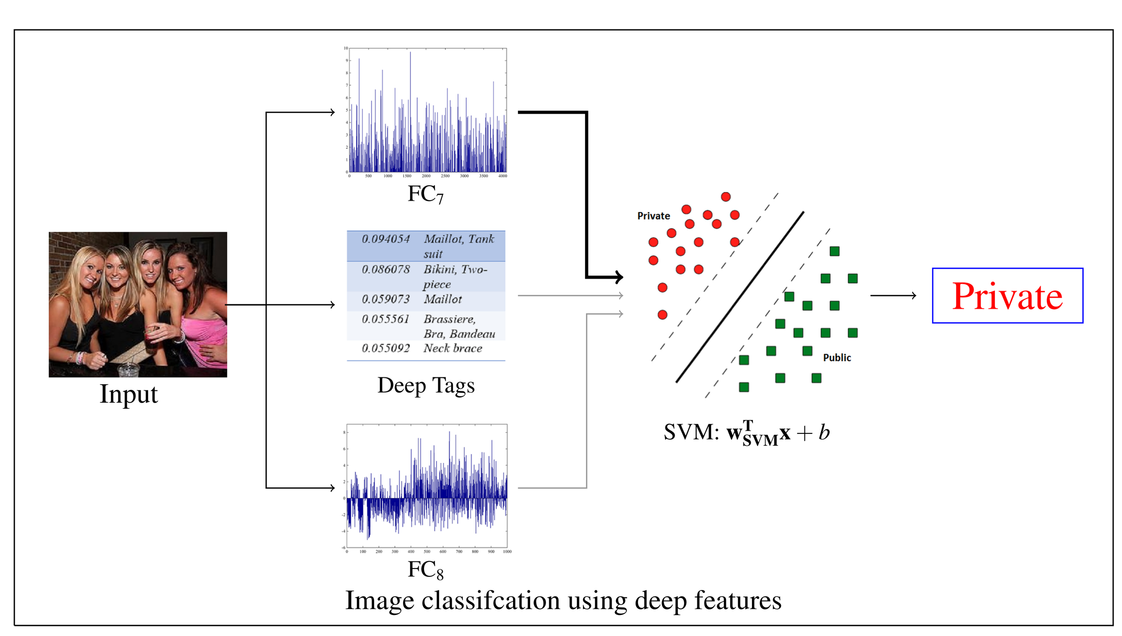


Figure: Proposed approach - Feature Classification (Deep Features and Deep Tags): The features from the fully-connected (fc) layers and deep tags are used to predict the class of an image as public or private using SVM.

Deep Tags vs. User Tags

Figure: Deep Tags vs. User Tags. For deep tags, we consider top $K = 5$ object labels as tags.

Important Links

- Extended Abstract: `http://www.cse.unt.edu/~ccaragea/posters/aaai16.pdf`
- Dataset: `https://www.dropbox.com/s/ydfpu51dec51krh/idsAndPrivacy.csv?dl=0`
- Full-length Paper: `http://arxiv.org/abs/1510.08583`

Experiments and Results

How do deep visual features compare with other existing state-of-the-art methods SIFT and GIST?

Features	Accuracy	F1-Measure	Precision	Recall
Test (*PiCalert*$_{783}$)				
FC_7	81.23%	0.805	0.804	0.812
FC_8	**82.63%**	**0.823**	**0.822**	**0.826**
SIFT + GIST	72.67%	0.661	0.672	0.727

Table: Performance of SVM using deep features in comparison with the combination of SIFT and GIST, on **Test**. For SIFT, we constructed a vocabulary of 128 visual words. For GIST, we considered feature vector of 512 (16 averaged value ×32 gabor filters) length.

How do tag features perform on the privacy prediction task?

Features	Accuracy	F1-Measure	Precision	Recall
Test (*PiCalert*$_{783}$)				
User Tags	79.82%	0.782	0.786	0.798
Deep Tags	80.59%	0.801	0.799	0.806
User + Deep Tags	**83.14%**	**0.827**	**0.826**	**0.831**

Table: Results obtained on tag features. For deep tags, we consider top $K = 10$ object labels as tags.

How do deep features perform for private class compared to SIFT and GIST?

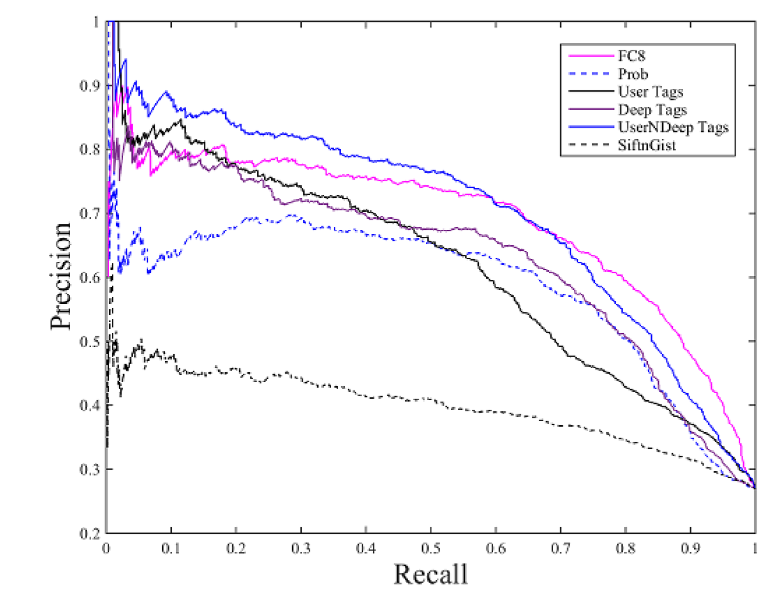
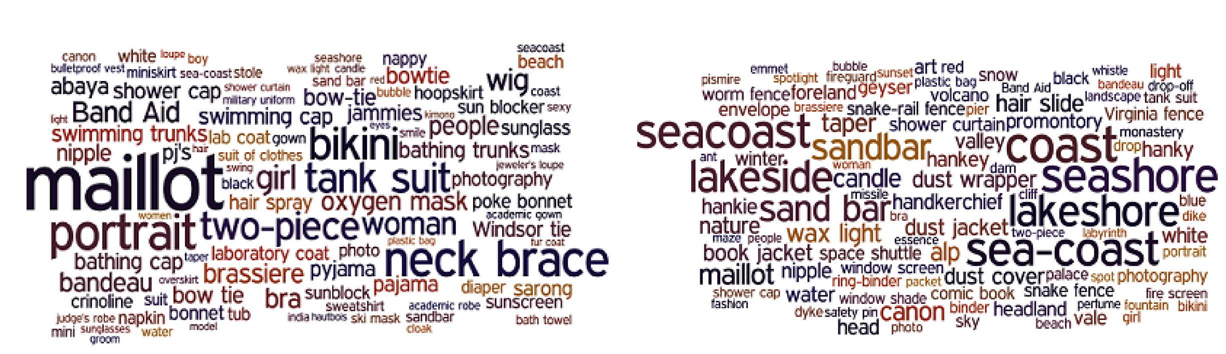


Figure: Precision and recall curves of different features for private class.

Which user tags and deep tags are useful for privacy prediction task?

Rank 1-5	Rank 6-10	Rank 11-15
Portrait	Maillot	Bathing Cap
Neck Brace	Wig	Swimming Cap
Two-piece	Bow-tie	Oxygen Mask
Bikini	**Girl**	Swimming Trunks
Tank Suit	**Woman**	Band Aid

Table: Tags with high information gain calculated using 5-fold cross validation. Bold words indicate user provided tags, while the others are deep tags.



(a) Private (b) Public

Figure: Tag clouds contains top 100 high frequency tags with respect to private and public images. High frequency tags represents frequently occurring tags to describe images for a particular privacy setting.

Conclusions

- We proposed an approach based on deep features and tags for privacy prediction.
- Deep features are explored at various network layers and also used top layer (probability) for auto-annotation mechanism.
- We examined user annotated tags and deep tag features.
- Our experiments shows that proposed method outperforms all baseline approaches.
- Future directions.
  - Refine user tags by using keyword extraction mechanism.
  - Combine visual features and tag features to get improved results.

References

Buschek, D., Bader, M., von Zezschwitz, E., and De Luca, A. (2015).
Automatic privacy classification of personal photos.
In Abascal, J., Barbosa, S., Fetter, M., Gross, T., Palanque, P., and Winckler, M., editors, *Human-Computer Interaction INTERACT 2015*, volume 9297 of *Lecture Notes in Computer Science*, pages 428–435. Springer International Publishing.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T. (2014).
Caffe: Convolutional architecture for fast feature embedding.
In *Proceedings of the ACM International Conference on Multimedia*, MM '14, pages 675–678, New York, NY, USA. ACM.

Klemperer, P. F., Liang, Y., Mazurek, M. L., Sleeper, M., Ur, B., Bauer, L., Cranor, L. F., Gupta, N., and Reiter, M. K. (2012).
Tag, you can see it! Using tags for access control in photo sharing.
In *CHI 2012: Conference on Human Factors in Computing Systems*. ACM.

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.
International Journal of Computer Vision (IJCV), pages 1–42.

Squicciarini, A. C., Caragea, C., and Balakavi, R. (2014).
Analyzing images' privacy for the modern web.
In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, HT '14, pages 136–147, New York, NY, USA. ACM.

Sundaram, H., Xie, L., De Choudhury, M., Lin, Y., and Natsev, A. (2012).
Multimedia semantics: Interactions between content and community.
Proceedings of the IEEE, 100(9):2737–2758.

Vyas, N., Squicciarini, A. C., Chang, C.-C., and Yao, D. (2009).
Towards automatic privacy management in web 2.0 with semantic analysis on annotations.
In *CollaborateCom*, pages 1–10.

Yeung, C., Kagal, L., Gibbins, N., and Shadbolt, N. (2009).
Providing access control to online photo albums based on tags and linked data.
Social Semantic Web: Where Web, 2.

Zerr, S., Siersdorfer, S., Hare, J., and Demidova, E. (2012).
Privacy-aware image classification and search.
In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, NY, USA. ACM.

Acknowledgements: This research is supported by a grant from NSF to Cornelia Caragea. We thank Adrian Silvescu and Raymond J. Mooney for helpful discussions. WWW: http://www.cse.unt.edu/~ccaragea