PrivAttNet: Predicting Privacy Risks in Images Using Visual Attention

Chen Zhang[1]; Thivya Kandappu[2]; Vigneshwaran Subbaraju[1]
[1]IHPC, A*STAR, [2]Singapore Management University

ABSTRACT

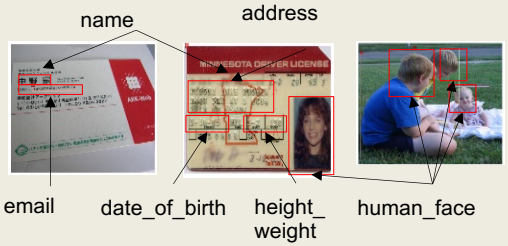
Visual privacy concerns associated with image sharing is a critical issue that need to be addressed to enable safe and lawful use of online social platforms. Given the recent success of visual attention based deep learning methods in measuring abstract phenomena like image memorability, we are motivated to investigate whether visual attention-based methods could be useful in measuring psychophysical phenomena like "privacy sensitivity". In this paper, we propose PrivAttNet – a visual attention-based approach, that can be trained end-to-end to estimate the privacy sensitivity of images without explicitly detecting sensitive objects and attributes present in the image. We show that our PrivAttNet model outperforms various SOTA and baseline strategies – a 1.6-fold reduction in L1 – error over SOTA and 7%–10% improvement in Spearman-rank correlation between the predicted and ground truth sensitivity scores. Additionally, the attention maps from PrivAttNet are found to be useful in directing the users to the regions that are responsible for generating the privacy risk score.

CONTACT

Thivya Kandappu
School of Information Systems
Singapore Management University
Email: thivyak@smu.edu.sg

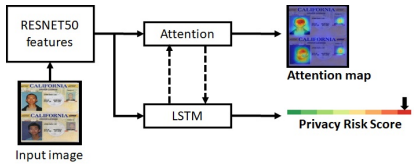
MOTIVATION

- Visual privacy concerns in sharing images
- Need of vision technique that addresses the following characteristics of images:
  - Multiple sensitive attributes present in the same image with one/more labels
  - Multi-label correlation or inter-dependencies
  - Multi labels can lie anywhere → different parts of images may have varying significance



OBJECTIVES

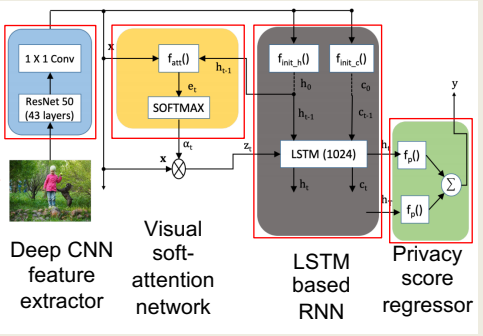
- Measure psycho-physical phenomena "privacy sensitivity" of images
- Localise the privacy sensitive attributes



APPROACH

Visual attention-based hybrid CNN-RNN approach
→ Sensitivity of an image may arise due to salient /non-salient artefacts present in the image

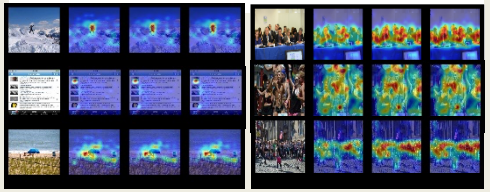
- CNN based feature extractor/encoder
- A visual soft-attention [1] network
- LSTM based RNN to preserve memory
- Privacy score regressor (multi-layer perceptron)



RESULTS

Dataset: VISPR [2]
- publicly available 22k Flickr images
- 68 privacy attributes
- Attribute-level privacy scores by human annotators (i.e., ground truth)

Performance Evaluation
- L1 error
- Correlation coefficients



Heat maps for small (left) and cluttered (right) objects

Method	L1-Error	Correlation ρ_p	Correlation ρ_s
AP-PR [19]	0.656	–	–
PR-CNN [19]	0.637	–	–
PrivAttNet	0.40	0.87	0.84
PrivAttNet$_{MLC}$	0.44	0.83	0.76
PrivNet	0.43	0.83	0.78

↓ ~ 59% improvement in L1 error

↑ Better correlation with human provided scores

CONCLUSIONS

To automatically estimate the privacy risk of an image and inform the user about the potentially sensitive parts of the image:
- Visual attention-based network
- End-to-end trainable, by-passing the need to explicitly detect the presence of attributes
- Localise the sensitive regions using heat maps

REFERENCES

[1] J. Fajtl, V. Argyriou, D. Monekosso, and P. Remagnino, "Amnet: Memorability estimation with attention," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6363–6372.

[2] T. Orekondy, B. Schiele, and M. Fritz, "Towards a visual privacy advisor: Understanding and predicting privacy risks in images," in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 3686–3695.