Detection without Recognition for Redaction

Shagan Sah[1], Ram Longman[1], Ameya Shringi[1], Robert Loce[2], Majid Rabbani[1], and Raymond Ptucha[1]

[1]Rochester Institute of Technology - 83 Lomb Memorial Drive, Rochester, NY USA, 14623
[2]Conduent, Conduent Labs - US, 800 Phillips Rd, MS128, Webster, NY USA, 14580
Email: *sxs4337@rit.edu*

Abstract

The prevalence of image and video capture along with public demand to view such recordings has made privacy protection a major concern. Redaction is used to obfuscate personally identifiable information such as faces, license plates, house numbers, and store fronts. Redaction aims to obfuscate targeted areas while maintaining scene context. We propose facial obfuscation methods that maintain context by allowing facial detection methods to find an obfuscated face, but prevent humans and facial recognition systems from identifying the same face.

1. Introduction

Advances in camera, data storage, and communications technology have made image and video capture ubiquitous. Dash cams capture accidents, surveillance cameras capture crimes, body cameras capture arrests, while consumer smart phones are ubiquitous. To protect the identity of individuals, it is becoming common practice to redact portions of imagery before releasing to the public. Redaction involves obscuring identifying information such as faces, license plates, identifying clothing, tattoos, house numbers, and computer screens.

Common redaction techniques mask private or Personally Identifiable Information (PII) by blurring, pixelizing, or blocking out PII pixels. Some approaches such as the JPEG encryption standard support APPn markers which allow selective scrambling of parts of the image using a public encryption and private decryption key. Recently, Corso et al. [2] presented a study with analysis on privacy protection in law enforcement cameras.

The three primary redaction steps include the localization of PII object(s), tracking of these objects over time, and their obfuscation. While these steps can be performed manually with video editing tools, current approaches are moving toward semi-automatic redaction with a manual review.

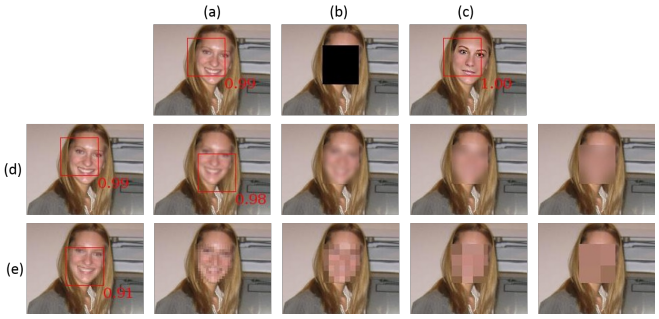

Figure 1. Example of a detection after redaction with different types and degrees of obfuscations (red box drawn of face detected with detection probability score). **(a)** Original image (174×135 pixels); **(b)** Detector fails when face obfuscated a mask; **(c)** Face detected when original face swapped with average female face; **(d)** Varying degree of blurring using Gaussian kernel (σ = 3, 5, 15, 25, 45); **(e)** Varying degree of pixelation (pixel block size = 2, 4, 8, 16, 32). Sample image from AFW dataset.

To ensure complete obfuscation, redaction methods are purposely aggressive. Unfortunately, aggressive obfuscation methods inadvertently alter important contextual meaning such as the identity of criminals or weapons used in a crime.

This research develops redaction methods that maximize PII obfuscation, while minimizing all other content changes. Specifically, with respect to faces, methods which blur or pixelize faces often change the tone or overall meaning of the scene. We find that methods which preserve face characteristics, but circumvent algorithmic or human facial recognition ensure minimal contextual alteration of a scene. Techniques which swap faces completely fool automatic recognition methods, but often don't fool first person observers. In particular, first person observers use additional cues such as hair and clothing to correctly identify an individual that has been obfuscated with facial swapping.

2. Methods

- **Face Detection**− Face detection localizes faces in an image. Recently, deep neural network based models have been proposed for the purpose of face alignment and detection. A review of traditional and recent face detection techniques is available in [7]. We use MT-CNN [8] to simultaneously detect faces and facial landmarks.
- **Face Recognition**− Recently, deep neural network based models such as FaceNet [5] and DeepFace [6] have been proposed for face recognition tasks. We use the OpenFace [1] implementation which uses a pre-trained FaceNet's Inception model for feature extraction. Each face is represented as a $128-$dimensional vector. A SVM classifier is used to identify the class (person) from this feature vector.

3. Experiments and Discussion

3.1. Face Detection after Redaction

Table 1 reports the detection Intersection-over-Union (IoU) scores on the AFLW [3] dataset after obfuscation. The face detector uses a pre-trained multi-task cascaded convolutional network [8] model. The mean-IoU decreases from 0.68 to 0.36 for images that are blurred. The overall context of the images can be lost with full blurring as the surrounding objects are often obfuscated. Similarly, the IoU scores decrease with increase in degree of pixelation. The detector was unable to detect any face when the faces were masked out, while swapping the face with an average face increase the detection scores.

Table 1. Face detection results on the AFLW [3] dataset.

Input		Detection (IoU)
Original image		0.688
Obfuscation	Full blur ($\sigma = 25$)	0.368
	Blur face ($\sigma = 5$)	0.713
	Blur face ($\sigma = 15$)	0.609
	Blur face ($\sigma = 25$)	0.421
	Blur face ($\sigma = 45$)	0.093
	Pixelate face ($p = 2$)	0.708
	Pixelate face ($p = 4$)	0.645
	Pixelate face ($p = 8$)	0.171
	Pixelate face ($p = 16$)	0.0
	Mask face	0.0
	Face swap	0.544

3.2. Face Recognition after Redaction

Table 2 reports the detection IoU scores and recognition accuracies on the FaceScrub [4] dataset using the OpenFace [1] implementation. In the FaceScrub dataset, we use 70% of faces for all celebrities as training data and the remaining 30% as test data. These experiments involve two stages of detection− the output of a first stage of face detection is used to localize various obfuscation techniques, and a second stage applied on obfuscated images. Faces detected in the second stage are passed onto the face recognition. As expected, as detection scores decrease, the recognition accuracy also decreases. However, the recognition accuracies with the face swapping technique were low despite having higher detection scores.

Table 2. Face detection and recognition on FaceScrub [4] dataset.

Input	Det. (IoU)	Recog. (% acc.)
Original image	0.901	98.29
Full blur ($\sigma = 25$)	0.570	48.80
Blur face ($\sigma = 5$)	0.898	97.53
Blur face ($\sigma = 15$)	0.893	91.11
Blur face ($\sigma = 25$)	0.769	68.32
Blur face ($\sigma = 45$)	0.228	7.66
Pixelate face ($p = 2$)	0.899	97.87
Pixelate face ($p = 4$)	0.889	96.10
Pixelate face ($p = 8$)	0.450	22.27
Pixelate face ($p = 16$)	0.000	0.02
Mask face	0.000	0.00
Face swap	0.903	34.73

4. Conclusion & Future Work

With the rising popularity of surveillance, body, car, and smart phone camera devices, image and video redaction or obfuscation of personal information for privacy protection is becoming very important. We conducted experiments to demonstrate the effect of various obfuscation techniques on face detection. Moreover, we also discuss the trade-off between face detection and recognition accuracies with varying degree of obfuscation. Experiments show that face swapping minimizes alteration of the scene, and are only partially effective at preventing recognition techniques from identifying individuals. We are currently performing human evaluations to see how well facial swapping techniques work at concealing the identity of individuals. Tentative results find face swapping works well for third person observers, but not as well for first person observers. Future work will investigate how well humans can identify individuals when full head and clothing swapping is performed.

References

[1] B. Amos et al. Openface. Technical report, Technical report, CMU School of Computer Science, 2016.

[2] J. J. Corso et al. Video analysis for body-worn cameras in law enforcement. *arXiv preprint arXiv:1604.03130*, 2016.

[3] M. Koestinger et al. Annotated facial landmarks in the wild. In *IEEE International Workshop on Benchmarking Facial Image Analysis Technologies*, 2011.

[4] H.-W. Ng and S. Winkler. A data-driven approach to cleaning large face datasets. In *ICIP*, 2014.

[5] F. Schroff et al. Facenet: A unified embedding for face recognition and clustering. In *CVPR*, 2015.

[6] Y. Taigman et al. Deepface: Closing the gap to human-level performance in face verification. In *CVPR*, 2014.

[7] S. Zafeiriou et al. A survey on face detection in the wild: past, present and future. *CVIU*, 2015.

[8] K. Zhang et al. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 2016.