Effect of Face Obfuscation Methods on Pose-Based Action Recognition

Muhammad Ahmed Raza Chris Lochhead Robert B. Fisher

School of Informatics, The University of Edinburgh, Edinburgh, UK
{m.a.raza, christopher.lochhead, r.b.fisher}@ed.ac.uk

Abstract

Previous research has shown that pose estimation deteriorates with common face obfuscation methods. This paper demonstrates that modern obfuscation methods do not necessarily affect the performance of the downstream task of pose-based action recognition.

Introduction and Background

Vision-based at-home health monitoring is a growing and effective modality for unobtrusively and inexpensively tracking the health of vulnerable people without the kind of sacrifice of autonomy that accompanies residing in a hospital or a care home. The vision modality comes with its own set of unique challenges, principally the preservation of privacy. To address the problem of both people's identities and homes being collected as data, the common solution is to use only second-order data extracted from the original image data. This is typically in the form of either silhouettes, point clouds [7], or skeletal graph data [6]. The issues with this solution are: (1) real images must be collected before the privacy-preserving processing, and (2) releasing this downstream data limits the opportunity to apply new pre-processing methods or experiment with different data representations.

Figure 1: Images generated by the three face obfuscation strategies i.e., face blur (left), deep-privacy1 (middle), and deep-privacy2 (right). These show the case when the eating utensil is near the subject's mouth.

To address this, researchers have been using machine learning-powered identity protection algorithms, which are effective at preserving privacy without the need for transforming data into these second-order representations, meaning much of the original image data can be kept and released for research. While traditional methods of obfuscation of the body and face have detrimental effects on performance across several tasks (pose estimation, image segmentation, etc.) [3], machine learning-powered "smart" methods of obfuscation are effective at minimizing this. Researchers [5] also determined that, for pose estimation, traditional blurring if done weakly enough has only a low drop in performance.

The question then, is "do identity protection algorithms significantly impact the accuracy of downstream tasks after pose estimation?". **This paper demonstrates, using three face obfuscation methods, that pose-based action recognition is essentially unaffected by privacy-preserving routines.**

Experiments and Discussions

The experiments used RGBD videos from the EatSense dataset [8], an action recognition dataset comprising 16 common sub-actions performed during eating, involving 27 subjects and a total of 135 videos. The videos were obfuscated using three strategies: face blur, deep-privacy1 [4], and deep-privacy2 [2]. Next, a two-step approach to 3D pose estimation was employed. Firstly, 2D joint positions (pose) were estimated using either HigherHRNet [1] or ViPNAS [9]. Secondly, the 2D joint positions were projected into 3D space using depth maps. The performance of pose-based action recognition was evaluated using ST-GCN as the action recognition framework, with top-1 accuracy as the performance metric. The evaluation only used EatSense videos where the subjects were not wearing any weights and with one video for each subject, resulting in $N = 27$ videos. To obtain a robust estimate of performance, 3-fold cross-validation was conducted using 18 and 9 videos for training and testing.

Table 1: The table shows the mean (μ) and standard deviation (σ) of top-1 accuracy estimated using 3-fold cross-validation on action recognition results with ST-GCN based on two 2D pose estimators and three strategies for face obfuscation.

	None		**Blur**		**deep-privacy1** [4]		**deep-privacy2** [2]	
	μ	σ	μ	σ	μ	σ	μ	σ
HigherHRNet [1]	70.33	5.57	59.11	9.95	72.08	6.41	66.38	5.40
ViPNAS [9]	67.17	8.23	57.37	10.50	66.10	7.42	64.66	5.92

Visual analysis of the output videos revealed that while face-blurring obscured the subject's identity to a considerable extent, it noticeably degraded the performance of 2D pose estimation, particularly in the facial region. Additionally, for actions such as 'move hand to mouth,' 'eat it', and 'move hand away from mouth', where the face is partially obscured, both deep-privacy1 and deep-privacy2 methods in-paint a realistic face mask over the eating utensils (*e.g.* forks and spoons), resulting in unrealistic deformed masks. Figure 1 shows a sample frame with all three face-obfuscation strategies, where the middle and right images show instances of unrealistic/deformed face obfuscation.

Table 1 presents the mean and standard deviations of the top-1 action recognition accuracies for two action recognition algorithms. Face-blur results in inferior performance compared to the scenario where no obfuscation is applied ('None' in the table), primarily due to the loss of critical information. Conversely, despite both deep-privacy1 and deep-privacy2 generating unrealistic masks, their performance is comparable to the 'None' condition. Thus, although the accuracy of the 2D pose- estimators may decrease with face-obfuscation [3], pose-based action recognition with obfuscated data achieves a similar level of accuracy to the action recognition with non-obfuscated data on the EatSense dataset. McNemar's test of statistical significance showed a p-value of greater than 0.90 meaning that there is statistically no difference between the 'None' and the two deep-privacy conditions.

Conclusion

This research, through experiments on (only) the EatSense dataset with various configurations, demonstrated that face-obfuscation strategies that pseudonymize facial features do not necessarily affect the performance of downstream tasks such as pose-based action recognition. Future research could explore additional action datasets and action recognition algorithms to corroborate these results.

References

[1] Cheng, B., Xiao, B., Wang, J., Shi, H., Huang, T.S., Zhang, L.: Higherhrnet: Scale-aware representation learning for bottom-up human pose estimation. In: Proc. IEEE/CVF conf. on computer vision and pattern recognition. pp. 5386–5395 (2020)

[2] Hukkelås, H., Lindseth, F.: Deepprivacy2: Towards realistic full-body anonymization. In: Proc IEEE/CVF Winter Conf on Applications of Computer Vision. pp. 1329–1338 (2023)

[3] Hukkelås, H., Lindseth, F.: Does image anonymization impact computer vision training? In: Proc. IEEE/CVF Conf on Computer Vision and Pattern Recognition. pp. 140–150 (2023)

[4] Hukkelås, H., Mester, R., Lindseth, F.: Deepprivacy: A generative adversarial network for face anonymization. In: Int Symp on visual computing. pp. 565–578. Springer (2019)

[5] Jiang, J., Skalli, W., Siadat, A., Gajny, L.: Effect of face blurring on human pose estimation: Ensuring subject privacy for medical and occupational health applications. Sensors **22**(23), 9376 (2022)

[6] Jun, K., Lee, Y., Lee, S., Lee, D.W., Kim, M.S.: Pathological gait classification using kinect v2 and gated recurrent neural networks. Ieee Access **8**, 139881–139891 (2020)

[7] Ortells, J., Herrero-Ezquerro, M.T., Mollineda, R.A.: Vision-based gait impairment analysis for aided diagnosis. Medical & biological engineering & computing **56**, 1553–1564 (2018)

[8] Raza, M.A., Chen, L., Nanbo, L., Fisher, R.B.: Eatsense: human centric, action recognition and localization dataset for understanding eating behaviors and quality of motion assessment. Image and Vision Computing **137**, 104762 (2023)

[9] Xu, L., Guan, Y., Jin, S., Liu, W., Qian, C., Luo, P., Ouyang, W., Wang, X.: Vipnas: Efficient video pose estimation via neural architecture search. In: Proc IEEE/CVF Conf on computer vision and pattern recognition. pp. 16072–16081 (2021)