TOWARDS IMPROVING FACE DEIDENTIFICATION PIPELINE

Tadej Ciglarič, Žiga Emeršič, Peter Peer, Blaž Meden

Faculty of Computer and Information Science, University of Ljubljana

tc2922@student.uni-lj.si,
{ziga.emersic, peter.peer, blaz.meden}@fri.uni-lj.si

ABSTRACT: *Face deidentification is an important part of privacy and security domains. Deidentification methods that rely on image blurring, pixelization or black-boxes were replaced in recent years with approaches based on formal anonymity models that provide privacy guaranties and at the same time aim at retaining certain characteristics of the data. However, current state-of-the-art pipeline we developed in earlier work, still suffers from sometimes erroneous face detection and color discrepancies in the replaces faces. We made improvements to an existing deidentifications pipeline by replacing face detector and implementing color correction on replaced faces. We also changed blending of replacement face to the original image in order to reduce amount of visual artifacts. Resulting pipeline misses less faces and creates more natural looking face replacements.*

1. INTRODUCTION

Nowadays large mount of videos are recorded every day. Sharing videos databases may infringe personal rights - videos can contain personal information such as faces. In such cases personal information needs to be deidentified, while still retaining certain characteristics, e.g. person's pose and gender. This creates the need for automatic deidentification. To tackle this problem we improve on the recently introduced deidentifcation pipeline [1], namely with improvements to detecting faces and making face replacement artifacts less noticeable.

1.1 Deidentification Pipeline

The recently introduced pipeline [1] consists of multiple image processing modules implemented in Python and Keras deep learning framework [2]. The overview of the pipeline improvements is illustrated in Figure 1, together with added improvements presented in this paper.

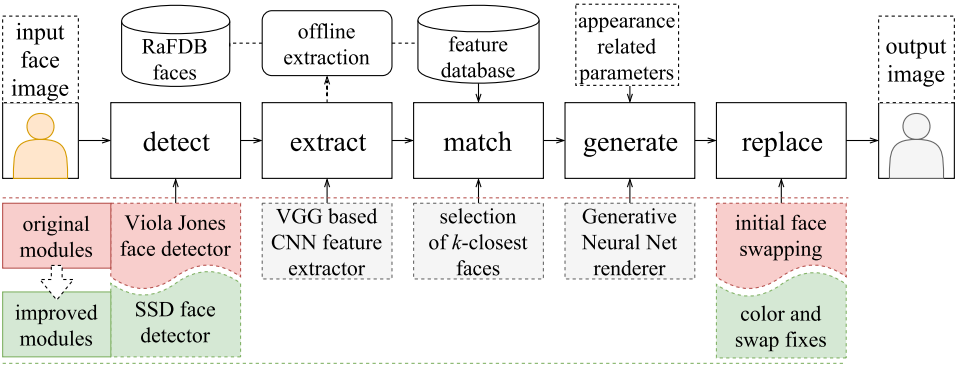


Figure 1: Overview of the deidentification pipeline from [1] with added module improvements. Viola-Jones face detector was replaced with CNN-based Single Shot Detector (SSD) trained for face detection. Face swapping module was updated with more advanced masking system with skin color corrections.

Pipeline begins with a face detection algorithm. The original pipeline uses Viola-Jones [3] face detector. Detected faces are then compared to database of predefined identities. Features, extracted by VGG [4] neural network are used for comparison. In the matching phase, k-closest identities are chosen to be combined to create a replacement face. This way the replaced face will be roughly similar to original one, but not enough to allow reliable identity recognition.

Data about chosen identities is fed into the generative neural network which generates replacement face, combining features of the chosen identities. The generating network has additional, appearance based inputs, which could control expression or any other visual aspects of the generated face to match the original one, thus preserving non-identity related attributes and enabling data utility in the deidentification process.

Finally original face must be replaced with generated one. On both original and generated face landmarks are detected. Detected points are used to estimate homography between original and generated face using RANSAC algorithm. Generated face is then warped according to estimated transformation. Gaussian mask is used to blend central part of the generated face to the original face – this is intended to suppress the sharp edges during face merging. The mask is originally combined with HSV threshold based skin segmentation in order to merge only facial regions, excluding hair and background.

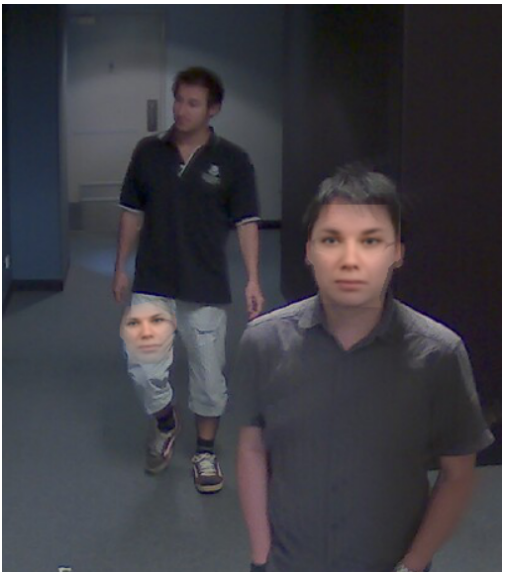


Figure 2: Failures of Viola-Jones face detector.

Figure 3: Visible face replacement artifacts.

2. PROBLEMS AND IMPROVEMENTS

Original pipeline implementation suffered from two major issues: face mis-detections and suboptimal face replacement. We addressed both of these problems and in this section we provide descriptions of the solutions to these problems and visual illustrations of some examples.

2.1 Improving Face Detector

Viola-Jones face detector, used in the original pipeline, is fast but suffer from mis-detections, when small or non-frontal faces are present in video sequences. In Figure 2 an example of the right person being correctly detected and deidentified, while face of the left person is not being detected. In this particular frame there is also a false detection in the knee area of the left person.

In order to avoid this kind of mis-detections we replaced the Viola-Jones-based face detector with a CNN-based approach.We selected Single Shot multibox Detector (SSD) proposed by Liu et al. [5]. The idea behind SSD is that it produces a fixed-size collection of bounding boxes and scores for the presence of object class instances in those boxes. This is followed by a non-maximum suppression step to produce the final detections. Detector uses VGG architecture [6] as a base model. Dense layers at the end of the original VGG are replaced with multi-scale convolutional layers, which predict the offsets to default boxes of different scales and aspect ratios and

their associated confidences.

We used MobileNet implementation from GitHub [7] which was adapted and trained for the task of face detection. It offers improved robustness compared to Viola-Jones. However faces are not always as well localized as Viola-Jones detections.

2.2 Improving Face Replacement Module

Next we have improved face replacement module. Using color-based skin segmentation and Gaussian kernel often creates more or less visible artifacts. Also there are visible sharp edges around replaced face. An extreme case can be seen in Figure 3.

Instead we try to estimate face area more precisely. We calculate convex hull of detected face landmarks on original face. Points of convex hull are used to render a polygon to use as a mask. It is than slightly blurred to prevent creation of sharp edges between original image and replaced face. Figure 4 shows the original replacement procedure and improved technique and visualises final deidentification result for both cases.

Generated faces often differ in color from originals. This is especially evident for different skin colors or in case of poor lighting conditions. Generated faces always have white skin and seem to be in bright environment. Examples are shown in Figure 5.

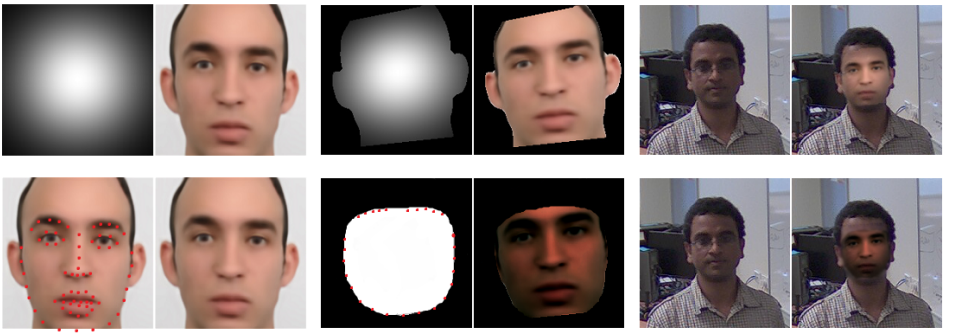


Figure 4: Top row shows the original pipeline with hard masks for face replacement [1]. The bottom line shows the improved face replacement, based on the detected face landmarks.

3. QUALITATIVE RESULTS

Evaluation of existing and modified pipeline was done on Chokepoint dataset [8]. It contains images from surveillance-like scenarios. Cameras are placed at "chokepoint" locations, where many people pass by.

(a) (b)

Figure 5: Two examples of color artifacts, present in the original implementation. (a) Generated face is not adapted to dark environment. (b) Deviating skin color effect of replaced face.

Replacing the Viola-Jones with SSD results in less missed face detections and therefore more complete deidentification, however the face localization is not that accurate than in Viola-Jones, which can result in face landmark detection failures in the final step of the pipeline. Modified face replacement module creates more natural looking faces as shown in Figure 6 (a) and (b).

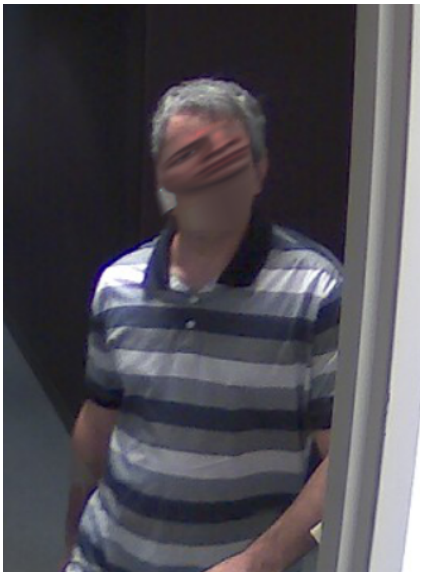
SSD face detector we used does not seem to localize faces as well as Viola-Jones. Often this causes landmarking to fail. In such case the replacement face is not aligned with the head. Two examples of this are shown in Figure 6 (c) and (d). Nevertheless, the overall results are significantly better compared to using Viola-Jones.

4. CONCLUSION

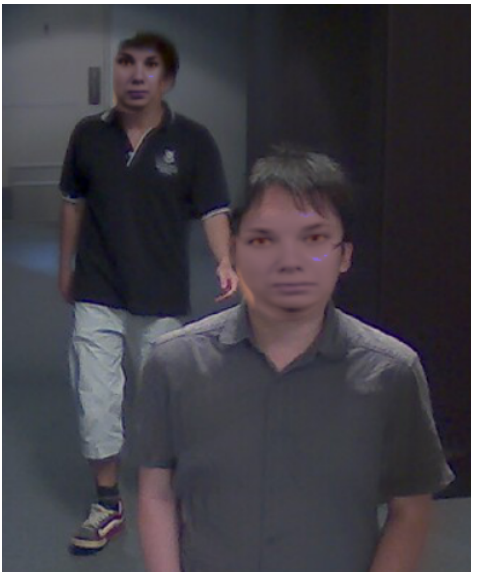
We made two significant improvements to an existing deidentification pipeline. Significantly less faces are missed and replacements look more natural. Landmarking failures could be addressed by picking another face detector or landmarking model. Another possibility would be to use detections from SSD to retrain the landmarker. Future work will also include hair segmentation and forehead deidentification.

(a)

(a)



(c)



(d)

Figure 6: Four deidentified examples using improved pipeline. In top row (a) and (b) show a good and the most common examples of the deidentification. In bottom two failures are shown. (c) An example of misaligned face. (d) An example of one well deidentified face and and one bad alignment in the back.

REFERENCES

[1] B. Meden, R. C. Mallı, S. Fabijan, H. K. Ekenel, V. Štruc, and P. Peer, "Face deidentification with generative deep neural networks," *IET Signal Processing*, vol. 11, no. 9, pp. 1046–1054, 2017.

[2] F. Chollet *et al.*, "Keras," https://github.com/fchollet/keras, 2015.

[3] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[4] O. M. Parkhi, A. Vedaldi, A. Zisserman *et al.*, "Deep face recognition." in *BMVC*, vol. 1, no. 3, 2015, p. 6.

[5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C. Fu, and A. C. Berg, "SSD: single shot multibox detector," *CoRR*, vol. abs/1512.02325, 2015. [Online]. Available: http://arxiv.org/abs/1512.02325

[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

[7] "A mobilenet ssd face detector," https://github.com/yeephycho/tensorflow-face-detection.

[8] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell, "Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition," in *IEEE Biometrics Workshop, Computer Vision and Pattern Recognition (CVPR) Workshops*. IEEE, June 2011, pp. 81–88.