

DeepPrivacy2: Towards Realistic Full-Body Anonymization (Supplementary)

Håkon Hukkelås Frank Lindseth
Norwegian University of Science and Technology
hakon.hukkelas@ntnu.no

Contents

A Flickr Diverse Humans Dataset	1
B The Updated Flickr Diverse Faces Dataset	1
C Experimental Details	1
D Random Anonymized Examples	2

A. Flickr Diverse Humans Dataset

The FDH dataset is generated by running the detection pipeline of DeepPrivacy2, where we only keep detections that are not filtered out by the set of criterias mentioned below. We filter out poor quality images in the FDH dataset through the following criterias:

- **Confidence thresholding:** All detections with confidence with less than 0.98.
- **Low resolution images:** All detections where the resulting image has an area lower than 144×80 .
- **Grayscale images:** All grayscale images.
- **Automatic image quality assessment:** We use an open-source implementation [6] of NIMA [12] and filter every detection with a score less than 3.
- **CSE Vertices per part:** We filter out detections with corrupted CSE detections by counting the number of vertices that belong each body part. Specifically, by finding pixel-to-vertex correspondences to the extended SMPL [9] body model used by CSE [10], we find the number of unique vertexes in the image that belong to each body part for the image (split into 26 regions). Then, if the average number of unique vertices per body part is less than 135, we filter the detection out.
- **Mask R-CNN and CSE IoU:** Detections where Mask R-CNN and CSE segmentations have an IoU threshold lower than 0.5.

- **Keypoint-to-CSE Correspondence:** We filter detections where the keypoints and CSE annotations do not match. Specifically, by using pixel-to-vertex correspondences we segment each CSE embedding into a semantic body part¹. Then, we count the number of keypoints that match to its corresponding body part (*e.g.* eyes should match to the body part "head"). We include any detection where at least 8 keypoints matches the corresponding body part. The keypoint annotations are predicted from a pre-trained Keypoint R-CNN [2] on COCO [8] from torchvision [11].

B. The Updated Flickr Diverse Faces Dataset

The Flickr Diverse Faces 256 (FDF256) dataset is a derivate from 1.08M images from the YFCC100M [13] dataset, following the dataset generation of the original FDF dataset [3]. The training dataset consists of 242,031 images and the validation set of 6533 images, where each face is up/downsampled to 256×256 . We filter out all faces where the original resolution is smaller than 64×64 . Each face is annotated with keypoints from a pre-trained keypoint R-CNN [2] R50-FPN from torchvision, and the bounding box is from the official implementation of DSFD [7].

C. Experimental Details

GAN Implementation Details We follow the implementation setup and hyperparameters of SG-GAN [4] for the optimizer and loss objective, unless stated otherwise. We train all networks with the Adam optimizer [5] with batch size of 32 and learning rate of 0.002. For both the FDF and FDH dataset, we use no data augmentation except random horizontal flip. The discriminator use the same architecture as SG-GAN [4], except that we remove the layers used for the feature pyramid network.

Market1501 Anonymization. The market1501 dataset consists of 19,732 test images and 3368 query images.

¹We use an open source semantic segmentation of the SMPL model, found here.

We anonymize the 19,732 test images and use the original 3368 query images to perform re-identification on the anonymized test images. We use a confidence threshold of 0.1 for Mask R-CNN, 0.5 for the face detector and 0.3 for the CSE detector. DSFD [7] and CSE [10] detects a large amount of false positives on the Market1501 dataset, which we believe is caused by the low resolution of the images. Out of all the detections, 45% are detected with CSE, 52% are detected by Mask R-CNN, and the remaining by DSFD. The detector fails to detect any individual for 10% of the test images. The mAP/R1 metric is calculated from all test examples, including cases where our detector fails to detect any individual.

D. Random Anonymized Examples

Cityscapes Randomly selected images from the Cityscapes [1] dataset are shown in Figure 1, Figure 2, and Figure 3.

COCO Randomly selected images from the COCO [8] dataset are shown in Figure 4, Figure 5, Figure 6.

FDH Randomly selected images from the FDH [8] dataset are shown in Figure 7, Figure 8, and Figure 9

FDF256 Randomly selected images from the FDF256 [8] dataset are shown in Figure 10, Figure 11, and Figure 12

References

- [1] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3213–3223, 2016.
- [2] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask R-CNN. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988. IEEE, oct 2017.
- [3] Håkon Hukkelås, Rudolf Mester, and Frank Lindseth. DeepPrivacy: A Generative Adversarial Network for Face Anonymization. In George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Daniela Ushizima, Sek Chai, Shinjiro Sueda, Xin Lin, Aidong Lu, Daniel Thalmann, Chaoli Wang, and Panpan Xu, editors, *Advances in Visual Computing*, pages 565–578. Springer International Publishing, 2019.
- [4] Håkon Hukkelås, Morten Smebye, Rudolf Mester, and Frank Lindseth. Realistic full-body anonymization with surface-guided gans. *arXiv preprint arXiv:2201.02193*, 2022.
- [5] Diederik P Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations*, 2015.
- [6] Christopher Lennan, Hao Nguyen, and Dat Tran. Image Quality Assessment. <https://github.com/idealol/image-quality-assessment>, 2018.
- [7] Jian Li, Yabiao Wang, Changan Wang, Ying Tai, Jianjun Qian, Jian Yang, Chengjie Wang, Jilin Li, and Feiyue Huang. DSFD: Dual Shot Face Detector. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5055–5064, 2019.
- [8] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common Objects in Context. In *Computer Vision – ECCV 2014*, pages 740–755. Springer International Publishing, 2014.
- [9] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL. *ACM Transactions on Graphics*, 34(6):1–16, nov 2015.
- [10] Natalia Neverova, David Novotny, Vasil Khalidov, Marc Szafraniec, Patrick Labatut, and Andrea Vedaldi. Continuous Surface Embeddings. In *Advances in Neural Information Processing Systems*, pages 17258—17270. Curran Associates, Inc., 2020.
- [11] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, and Others. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.
- [12] Hossein Talebi and Peyman Milanfar. NIMA: Neural Image Assessment. *IEEE Transactions on Image Processing*, 27(8):3998–4011, aug 2018.
- [13] Bart Thomee, David A. Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. YFCC100M: The new data in multimedia research. *Commun. ACM*, 59(2):64–73, jan 2016.

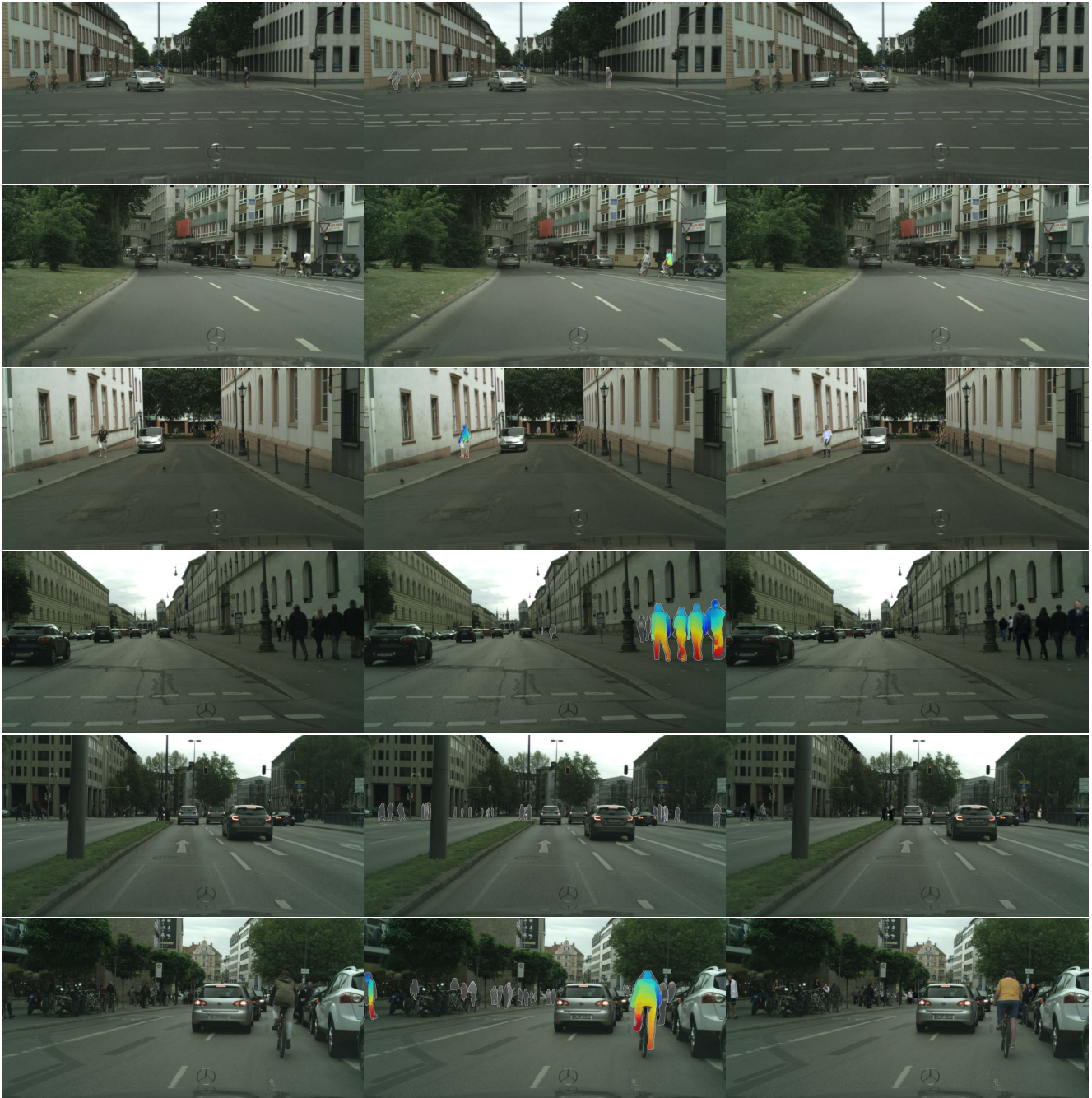


(a) Original

(b) Detections

(c) Anonymized

Figure 1: Randomly generated images from the Cityscapes [1] test set. (a) shows the original image (with blurred face to follow Cityscapes licensing), (b) the detections and (c) anonymized image



(a) Original

(b) Detections

(c) Anonymized

Figure 2: Randomly generated images from the Cityscapes [1] test set. (a) shows the original image (with blurred face to follow Cityscapes licensing), (b) the detections and (c) anonymized image

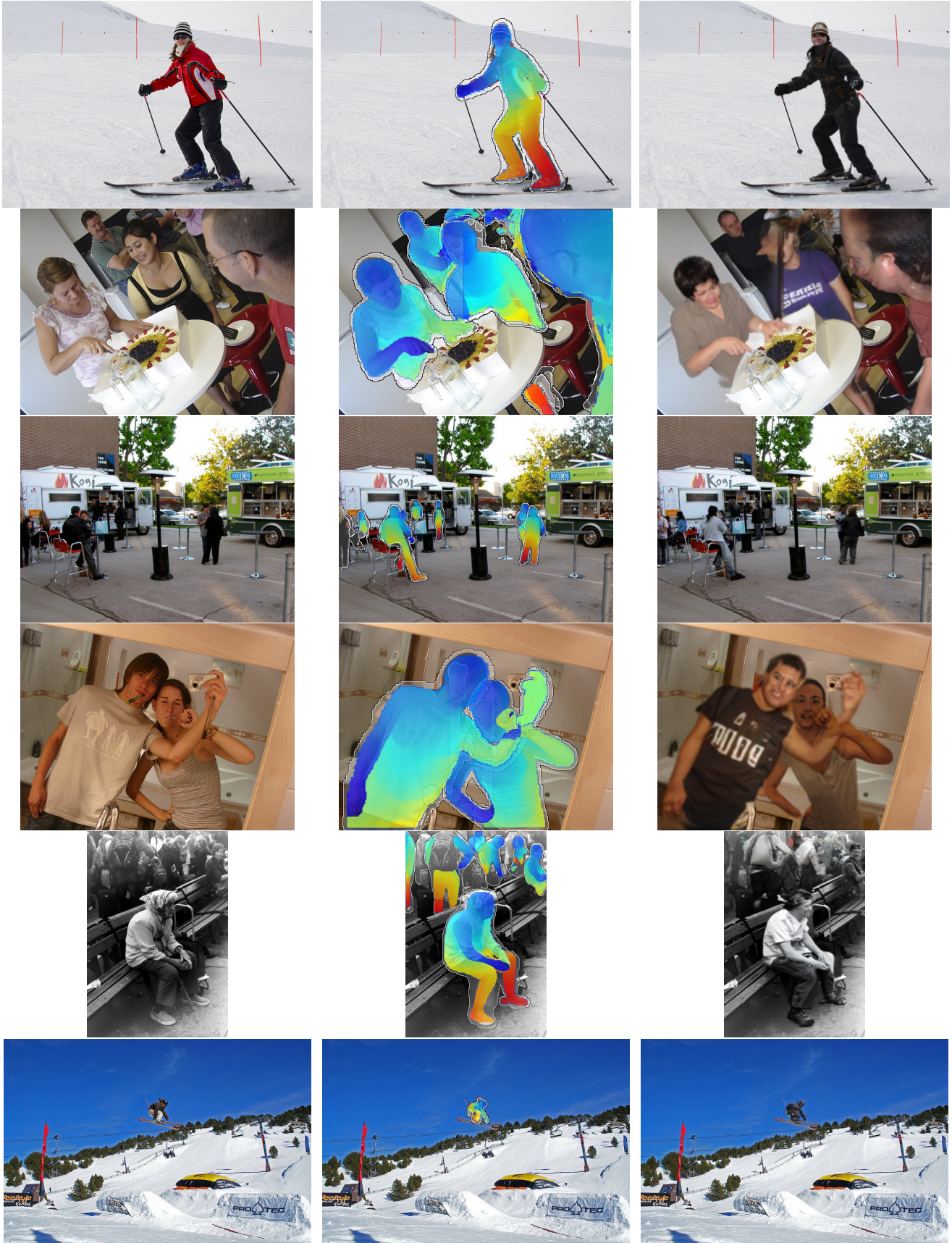


(a) Original

(b) Detections

(c) Anonymized

Figure 3: Randomly generated images from the Cityscapes [1] test set. (a) shows the original image (with blurred face to follow Cityscapes licensing), (b) the detections and (c) anonymized image



(a) Original

(b) Detections

(c) Anonymized

Figure 4: Randomly generated images from COCO [8] val2017. (a) shows the original image, (b) the detections and (c) anonymized image

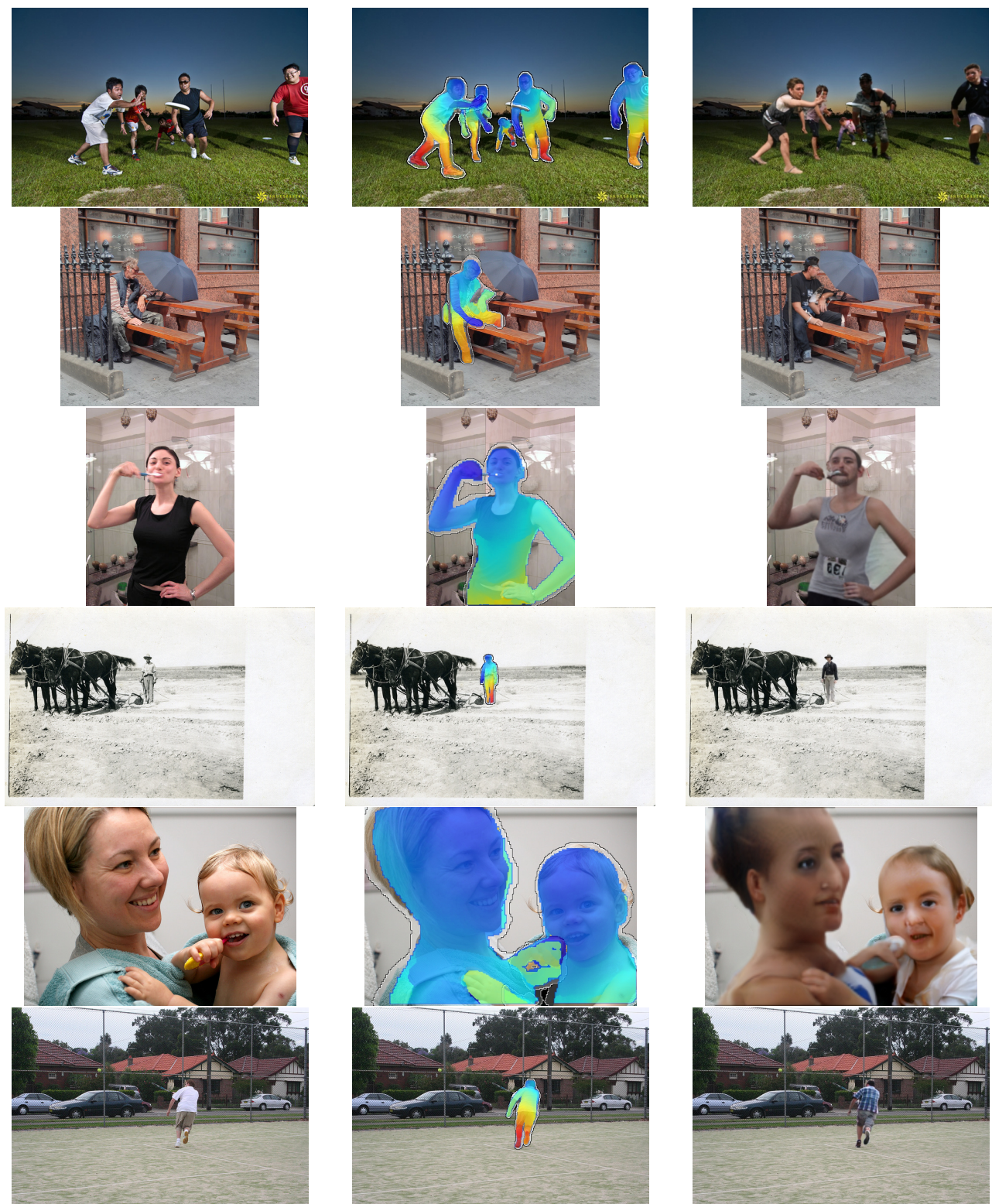


(a) Original

(b) Detections

(c) Anonymized

Figure 5: Randomly generated images from COCO [8] val2017. (a) shows the original image, (b) the detections and (c) anonymized image



(a) Original

(b) Detections

(c) Anonymized

Figure 6: Randomly generated images from COCO [8] val2017. (a) shows the original image, (b) the detections and (c) anonymized image

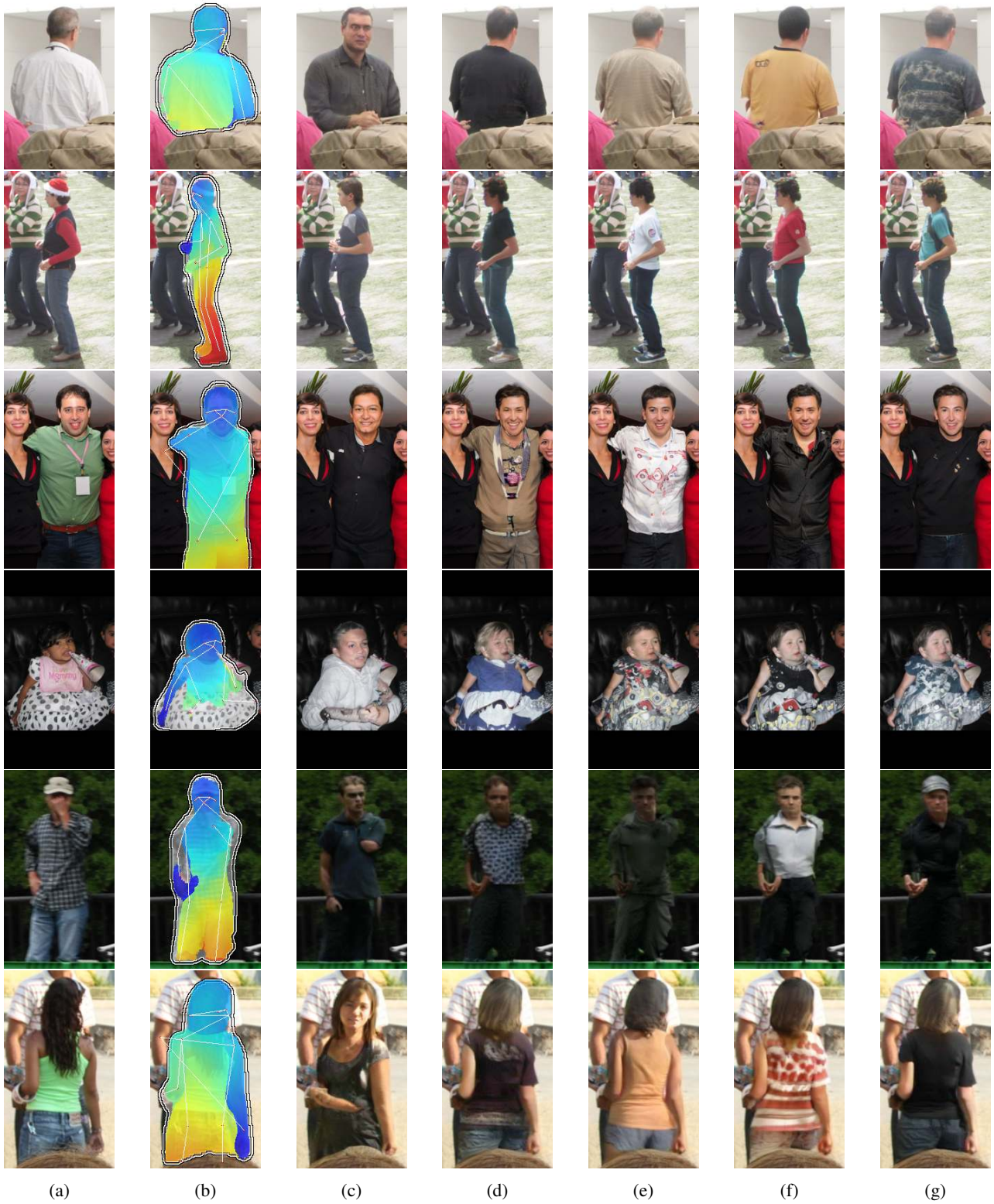


Figure 7: Randomly generated images from the FDH dataset. (a) shows the original image, (b) is the conditional input, (c) is the generated result for the unconditional generator, and (c-g) are from the CSE-guided generator.



Figure 8: Randomly generated images from the FDH dataset. (a) shows the original image, (b) is the conditional input, (c) is the generated result for the unconditional generator, and (c-g) are from the CSE-guided generator.

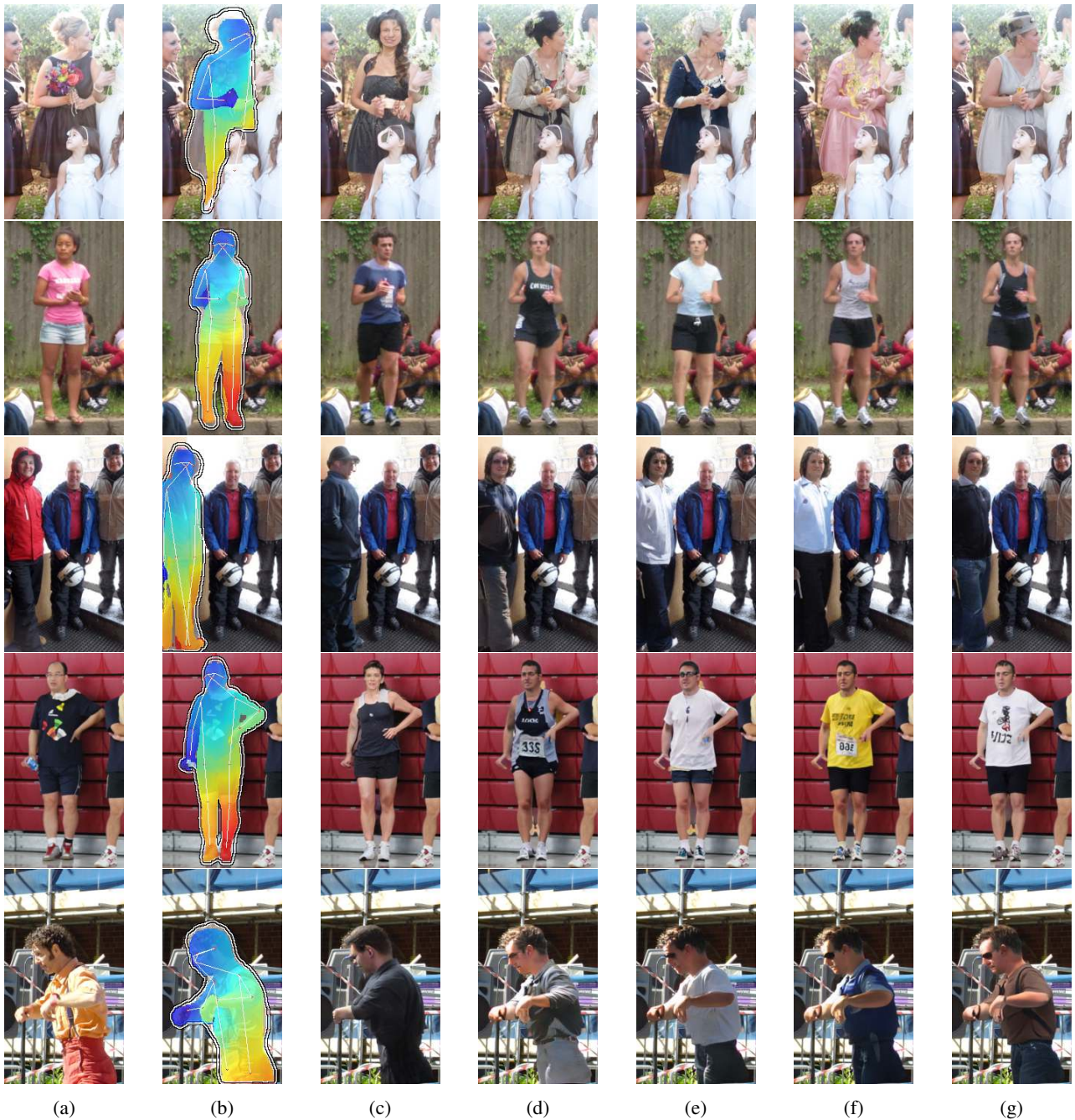


Figure 9: Randomly generated images from the FDH dataset. (a) shows the original image, (b) is the conditional input, (c) is the generated result for the unconditional generator, and (c-g) are from the CSE-guided generator.



Figure 10: Randomly generated images from the FDF dataset. (a) shows the original image, (b) is the conditional input, and (b-f) are generated images. No latent truncation is applied for the generated images.

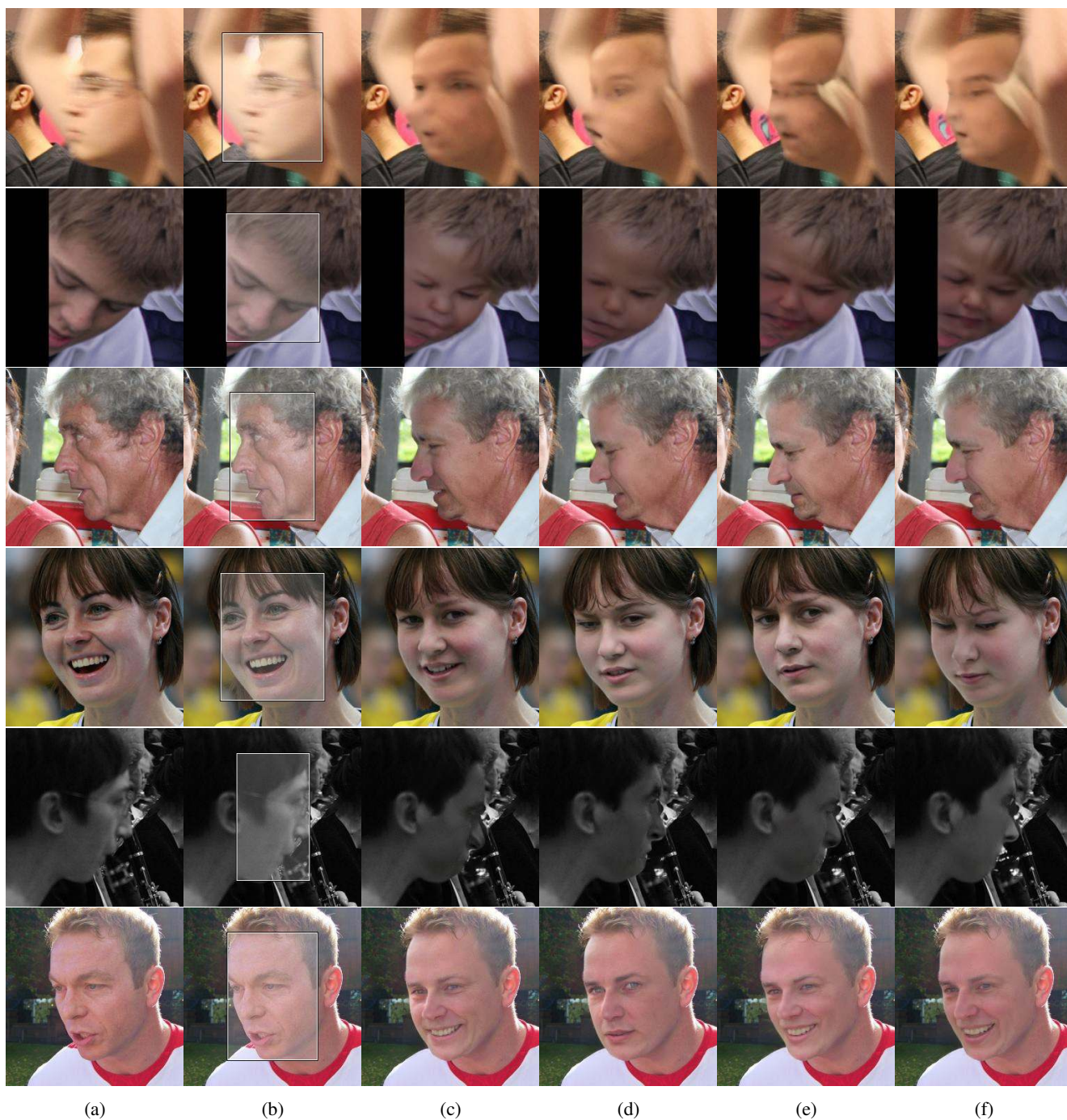


Figure 11: Randomly generated images from the FDF dataset. (a) shows the original image, (b) is the conditional input, and (b-f) are generated images. No latent truncation is applied for the generated images.

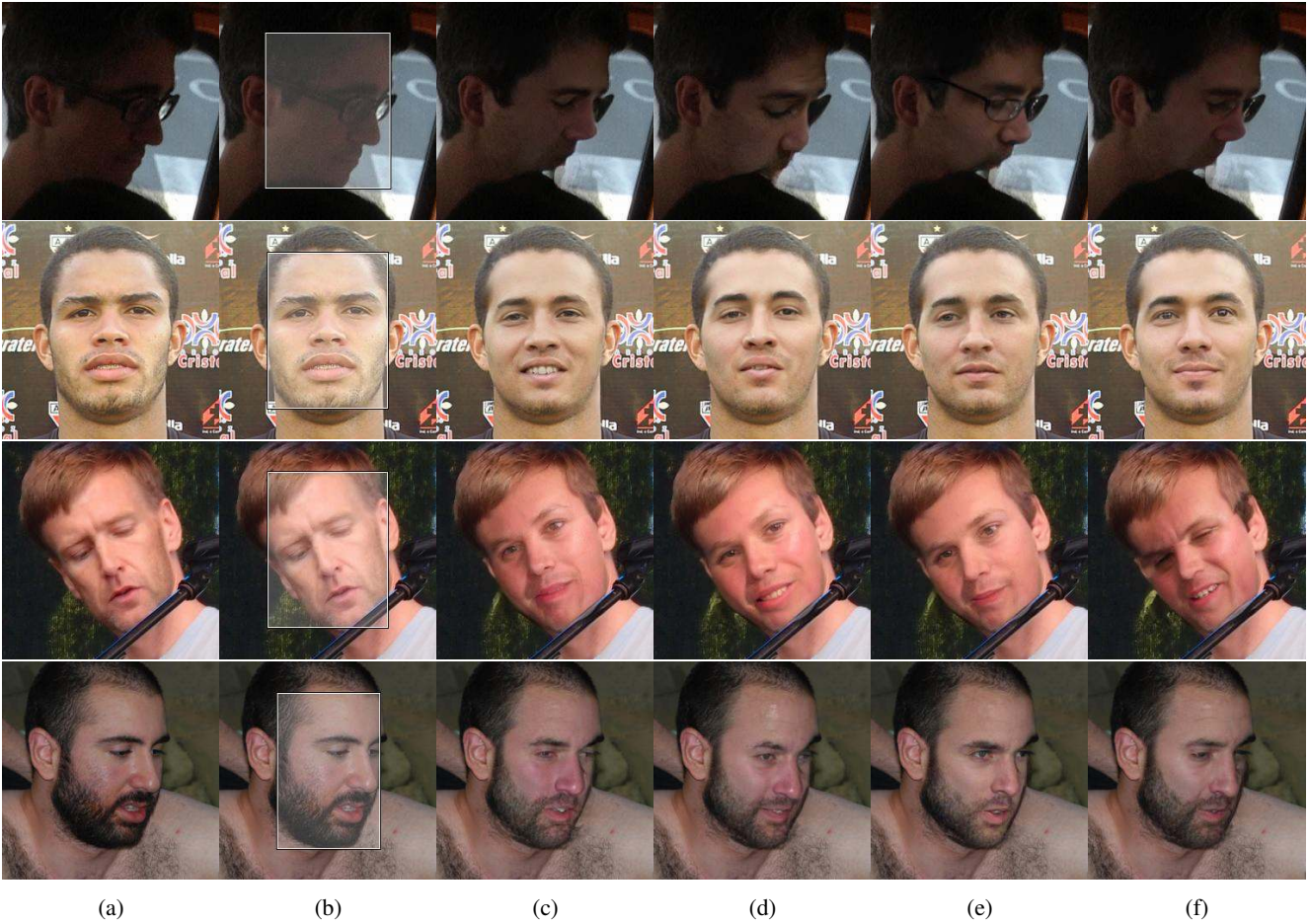


Figure 12: Randomly generated images from the FDF dataset. (a) shows the original image, (b) is the conditional input, and (b-f) are generated images. No latent truncation is applied for the generated images.