Unsupervised Magnification of Posture Deviations Across Subjects
Supplementary Material

Michael Dorkenwald* Uta Büchler* Björn Ommer
HCI / IWR, Heidelberg University, Germany

A. Videos

In the main paper we provided exemplary generations of our magnification approach for all three datasets. Now we additionally provide videos to further demonstrate the benefit of our magnifications. The videos are also available on our project page[1].

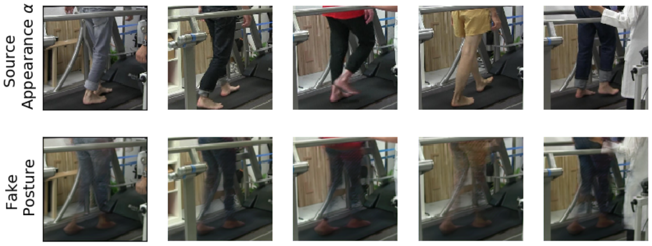
A.1. Magnification

We provide four magnification videos for the Golf Swing (*golf_swing_example*.mp4*) and the HG2DB (*HG2DB_example*.mp4*) dataset. The videos display the change in posture while the magnification factor λ is slowly increased. We also show the results achieved with [4], and for HG2DB we additionally include the magnifications generated by our model without $\mathcal{L}_{\text{mag}}$ and/or $\mathcal{L}_{\text{dis}}$. We summarize the content of the provided videos in Suppl. Tab. 1. The videos demonstrate that our model successfully magnifies the posture deviations while the approach by Oh *et al.* [4] generates mostly blurry images without considering the posture differences between reference and query.

In CUEye_same_appearance.mp4 and CUEye_transfer_appearance.mp4 we show the magnification of the pupil's movement (last 5 seconds of every video with no deliberate movements) and compare with the results generated by Oh *et al.* [4]. In the former video, we only magnify the differences within the same video as conducted for motion magnification. The latter video transfers the posture deviation of a specific subject (1st column) to five different subjects with different appearances (2nd column) as in Fig. 6 of the main submission. The videos show, that our approach is able to successfully address both settings. We achieve similar results as [4] if the magnification is performed in the same video, but clearly outperform [4] when transferring posture deviations to different subjects. CUEye_transfer_appearance.mp4 shows, that [4] is not able to properly disentangle appearance from posture and therefore generates images with the same appearance for all subjects.

File	Difference to Reference
golf_swing_example1.mp4	arms are higher & right knee is twisted inside
golf_swing_example2.mp4	arms are more on the left
golf_swing_example3.mp4	arms are lower
golf_swing_example4.mp4	arms are higher
HG2DB_example1.mp4	feet are closer to each other
HG2DB_example2.mp4	feet are more parallel
HG2DB_example3.mp4	left foot is lower & feet are slightly more parallel
HG2DB_example4.mp4	feet are further apart & left foot is slightly higher

Suppl. Table 1. Overview of the provided videos for HG2DB and Golf Swing showing the change in posture while the magnification factor λ is slowly increased.



Suppl. Figure 1. Fake Posture Images. The images are generated using the frames in the first row as input to the appearance encoder and random Gaussian noise as posture encoding. The resulting generations (2nd row) only contain appearance information (e.g. background or color of trousers) and display an average posture for all subjects.

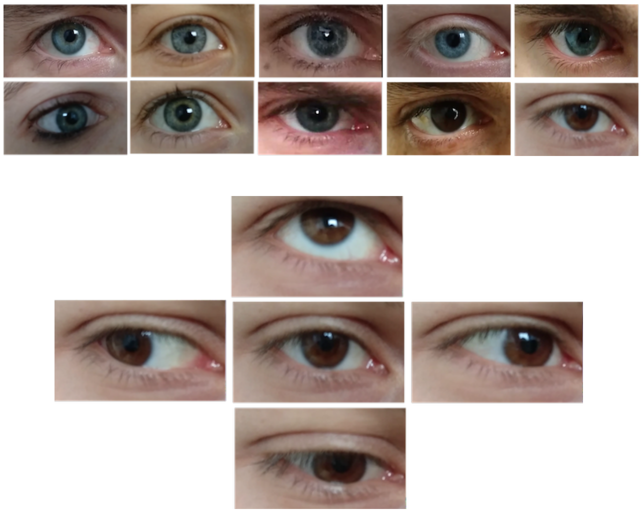
A.2. Disentanglement

As indicated in the main paper, the magnification of posture deviations across subjects requires the posture and appearance encoders to be disentangled. To provide further insight into our disentanglement approach we show the reconstructions with either fake appearance $\hat{x}^{q}_{\pi}$ or fake posture $\hat{x}^{q'}_{\alpha}$. In Disentanglement_HG2DB_fake_appearance.mp4 we present images generated with fake appearance by replacing the appearance encoding with random Gaussian

---

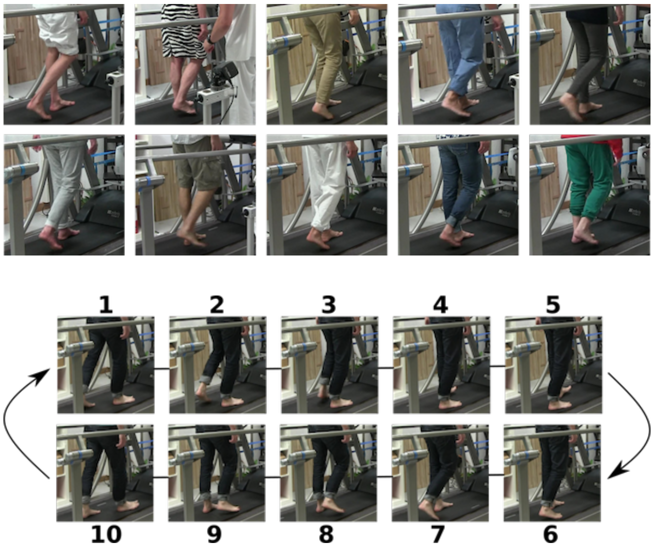
*Indicates equal contribution

[1] https://compvis.github.io/magnify-posture-deviations/



Suppl. Figure 2. Overview of CUEye. *Top*: One image represents one subject. *Bottom*: Exemplary postures available in CUEye.

noise for several walking cycles. In Suppl. Fig. 1 we display images generated with random Gaussian noise as posture encoding (fake posture). As intended, the images shown in the video only represent the posture and do not contain any appearance information; while the images in Suppl. Fig. 1 contain appearance information such as the background or color of clothes but display all subjects with the same average posture indicating the lack of posture information.



Suppl. Figure 3. Overview of HG2DB. *Top*: Subset of subjects in HG2DB. *Bottom*: One Walking cycle represented by 10 linearly spaced frames.

Suppl. Figure 4. Overview of Golf Swing.

B. Datasets

In this section, we provide a more detailed description and examples of our three datasets CUEye, HG2DB and Golf Swing.

CUEye: CUEye contains close-up videos of eyes of 10 different subjects with 1 video per subject. The videos are recorded with a standard HD camera with 30 frames per second and an average length of 25 seconds. For the experiments, we cropped the videos so that only the eye is shown which results in frames with an average size of 500×350 (width×height). The dataset contains 3 different eye colors: green, blue and brown. Suppl. Fig. 2 shows one exemplary frame per video. The subjects were asked to first move their eyes left, right, up and down, followed by 5 seconds without any deliberate movement (looking straight). The frames with movement are used to train the generative model, and the last 5 seconds to evaluate the model on the magnification abilities.

HG2DB: The dataset contains 229 videos of human subjects walking on a treadmill, where 59 of the 69 unique subjects are affected by diseases debilitating their walking skills. The remaining 10 subjects are healthy and used as reference. The videos have been recorded by clinicians from University Hospital Zurich with an HD camera and a frame rate of 25. For our experiments, we cropped the frames so that the subjects are placed in the middle and only shown from the waist down (for anonymity reasons), resulting in frames with a size of 400×400. Suppl. Fig. 3 shows the variety of our dataset; the different subjects wear different types of trousers in different colors. We additionally depict in Suppl. Fig. 3 (bottom) one walking cycle linearly spaced in 10 frames. These postures represent the poses we used for the quantitative experiment (Tab. 1) in the main submission.

Golf Swing: Golf Swing contains 48 videos of various people performing a golf swing on different tournaments. We collected the slow-motion videos with 120 fps from YouTube[1]. The videos are recorded with a static camera. For our experiments, we cropped the frames so that the person-of-interest is located in the center which results in frames with an average size of 600x600. Suppl. Fig. 4 shows a subset of available subjects in Golf Swing and various types of postures.

---

[1] https://www.youtube.com/user/GolfswingHD/

C. Implementation Details

We use a network architecture similar to the model proposed by Esser *et al*. [1] with 6 Resnet blocks [2] per encoder/decoder and skip connections between the appearance encoder and the decoder. For downsampling we use a convolutional layer with a kernelsize of 3, a stride of 2 and padding of 1. For upsampling we employ NN upsampling followed by a 1x1 convolutional layer. Our discriminator C follows the structure from DCGAN[5]. The weights are Xavier initialized and we train our model using the Adam optimizer [3] with a learning rate of 1×10^{-4}. We train on a single Titan Xp with a batchsize of 8 and an image size of 128.

The standard deviation σ for the Gaussian distribution used in $\mathcal{L}_{\text{dis}}$ is set to 0.01. We did not see that as a critical parameter; slight changes did not affect the results.

To train our model with $\mathcal{L}_{\text{mag}}$ we require the nearest neighbors of every query sample x^q. Therefore, we first train our model without $\mathcal{L}_{\text{mag}}$ and then employ the trained posture encoding for obtaining the NNs. Afterwards, we train our model with all losses. We train in total for 300 epochs for Golf Swing and 100 Epochs for HG2DB and CUEye.

References

[1] Patrick Esser, Ekaterina Sutter, and Björn Ommer. A variational u-net for conditional appearance and shape generation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8857–8866, 2018.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[3] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[4] Tae-Hyun Oh, Ronnachai Jaroensri, Changil Kim, Mohamed Elgharib, Fr'edo Durand, William T Freeman, and Wojciech Matusik. Learning-based video motion magnification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 633–648, 2018.

[5] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.