Anatomical Landmark Detection in Medical Applications Driven by Synthetic Data

Gernot Riegler[1] Martin Urschler[1,2] Mattias Rüther[1] Horst Bischof[1] Darko Stern[1,2]

[1]Graz University of Technology [2]Ludwig Boltzmann Institute for Clinical Forensic Imaging

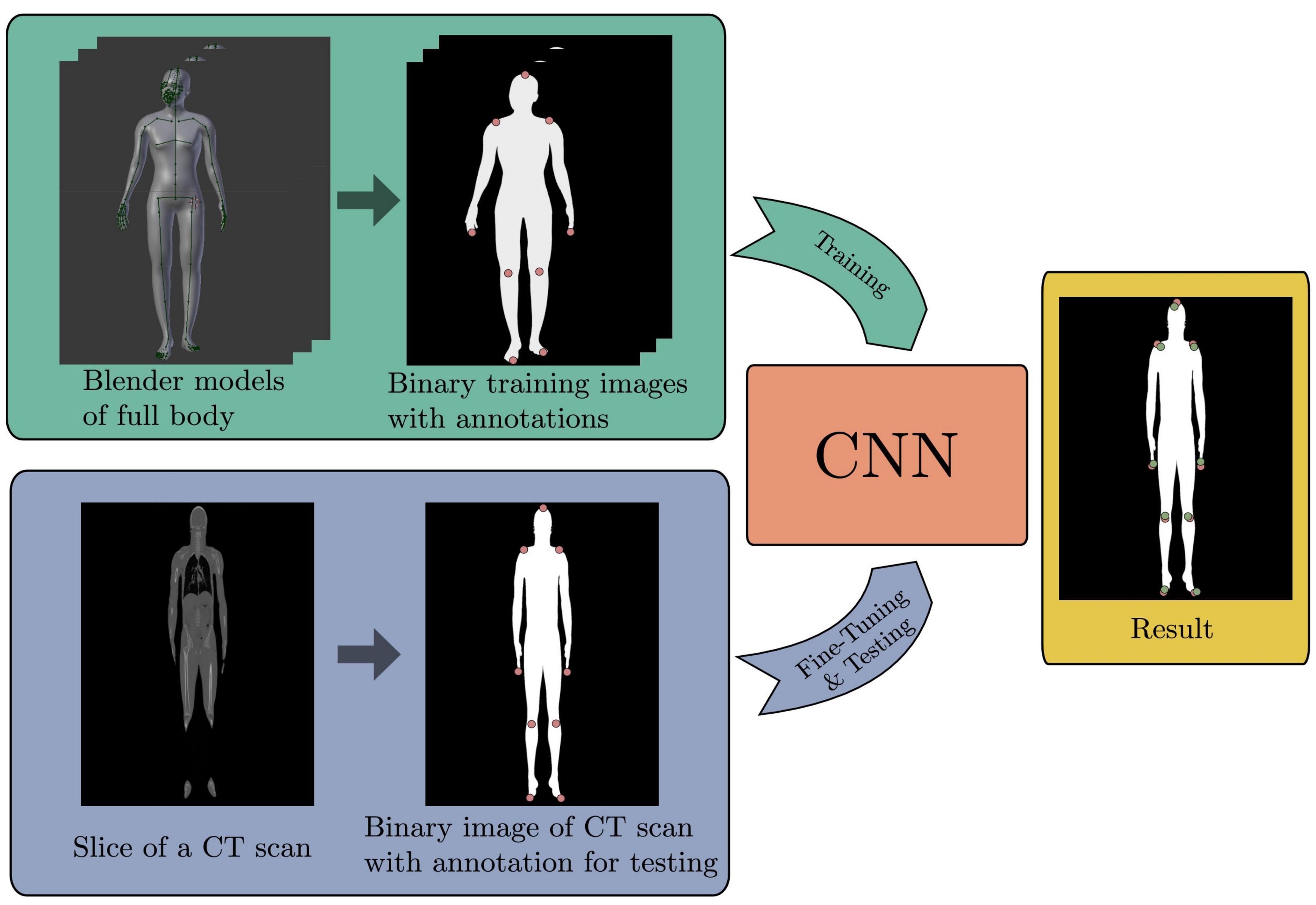
{riegler, ruether, bischof, stern}@icg.tugraz.at martin.urschler@cfi.lbg.ac.at

Motivation

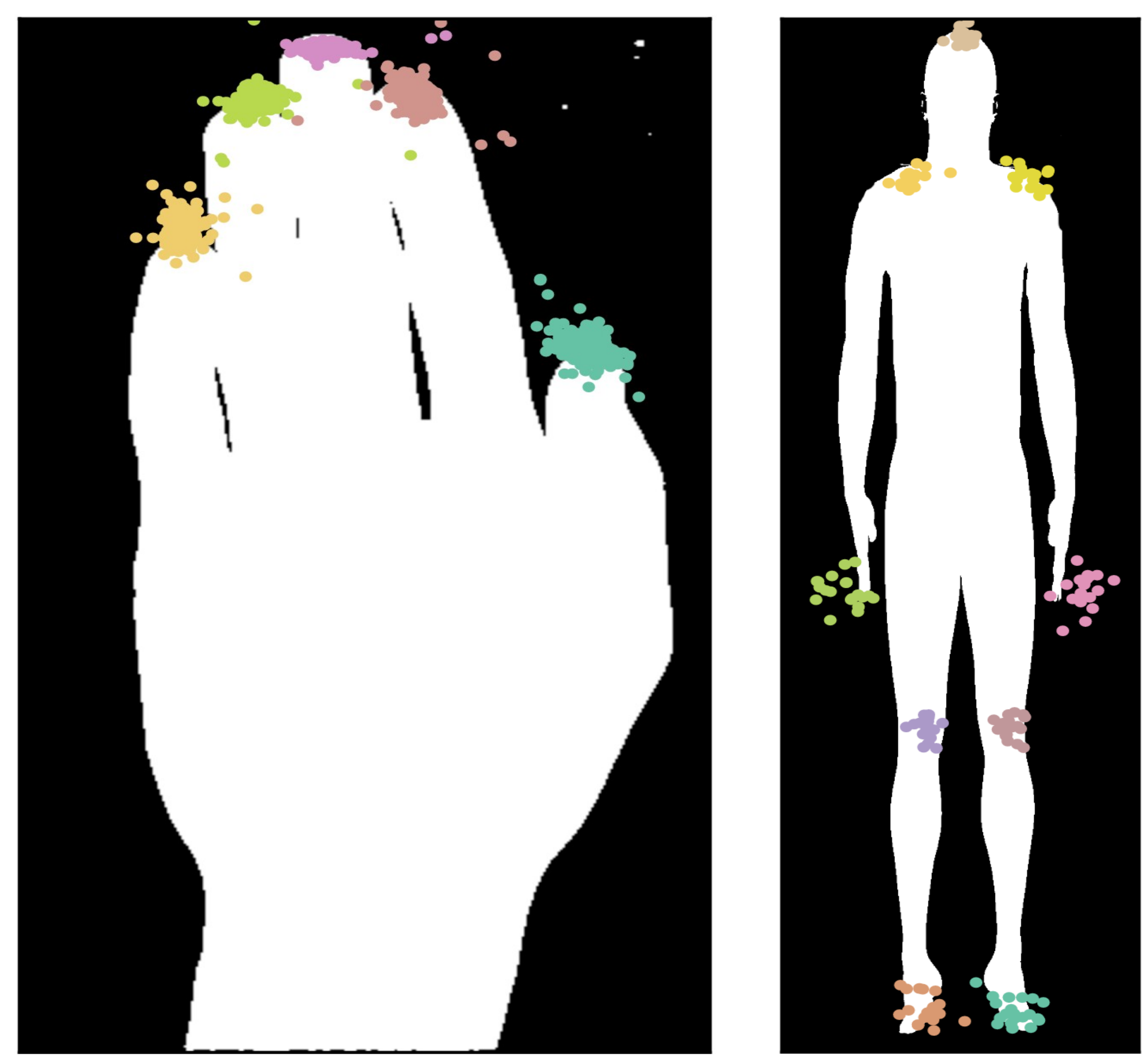
Modern machine learning techniques, e.g. convolutional neural networks, for object detection need a **large corpus of training data**, which is often an **unrealistic setting for medical datasets**. We investigate how to **adapt synthetic image datasets** from other computer vision tasks to overcome the under-representation of the anatomical pose and shape variations in medical image datasets.

Discussion and Conclusion

Data domains are transformed to a common one in such a way that a convolutional neural network (CNN) can be **trained on the larger synthetic image dataset** and **fine-tuned on the smaller medical image dataset**. Our evaluations on data of MR hand and whole body CT images demonstrate that this approach improves the detection results compared to training only on the medical data.



Adaptation of body pose estimation for anatomical landmark detection in medical images. Both image domains are transformed to a common domain, i.e. a **binarization of the data**. The larger, synthetic image dataset is used to optimize the parameters of a CNN, while fine-tuning and evaluation of the method is done on binarized MR/CT data.



The estimated anatomical landmark localizations projected into one common images for binarized MR hand and CT whole body images.

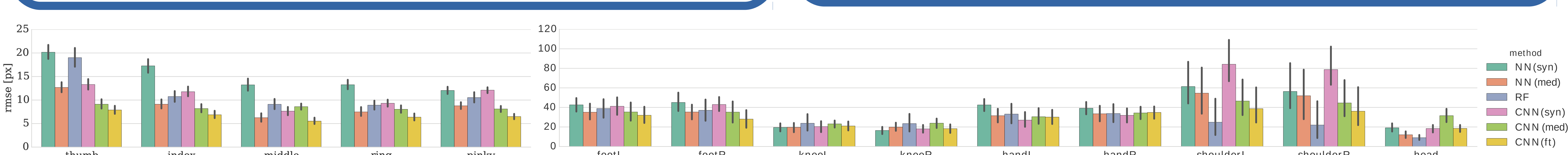
Method

We formulated the localization problem as a regression task. The binarized inputs and landmark locations are **transformed to a unit size**. We train a **6-layer CNN** on this data **minimizing the MSE** between estimated and ground-truth anatomical landmark locations in 2D. The variation in depth is neglectable. The network is first trained on the synthetic data to learn a rich **feature representation** and then fine-tuned on medical data to re-parameterize the output space.

Experimental Set-Up

Dataset: 132 left hand T1-weighted 3D gradient echo MR images; 20 whole body CT images; 2.2M left hand and 0.6M whole body synthetic images generated with Blender utilizing MakeHuman models.

Experiment: The network is trained for 30 epochs on the synthetic data with a mini-batch size of 100 and an initial learning rate of 0.01. For fine-tuning, we perform a cross-validation (three-fold for hand and leave-one-out for full-body) on the medical data.



Anatomical landmark detection results on MR hand data (left) and CT scans of whole body (right) reported as mean and standard deviation of the root mean squared error (rmse) between estimated and ground-truth anatomical landmark location in pixel (px). The proposed convolution neural network CNN method trained on synthetic images and fine-tuned on medical images CNN(ft) is compared with: CNN trained only on the synthetic images CNN(syn) or only on the medical images CNN(med), random forest trained on the medical images RF and with nearest neighbor search performed on the synthetic NN(syn) or medical image NN(med) data set.

Acknowledgment: This work was supported by the Austrian Science Found (FWF): P 28078-N33 (FAME)