X-ray-transform Invariant Anatomical Landmark Detection for Pelvic Trauma Surgery

Bastian Bier[1,2(✉)], Mathias Unberath[2], Jan-Nico Zaech[1,2], Javad Fotouhi[2], Mehran Armand[3], Greg Osgood[4], Nassir Navab[2], and Andreas Maier[1]

[1] Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
bastian.bier@fau.de
[2] Computer Aided Medical Procedures, Johns Hopkins University, Baltimore, USA
[3] Applied Physics Laboratory, Johns Hopkins University, Baltimore, USA
[4] Department of Orthopaedic Surgery, Johns Hopkins Hospital, Baltimore, USA

Abstract. X-ray image guidance enables percutaneous alternatives to complex procedures. Unfortunately, the indirect view onto the anatomy in addition to projective simplification substantially increase the task-load for the surgeon. Additional 3D information such as knowledge of anatomical landmarks can benefit surgical decision making in complicated scenarios. Automatic detection of these landmarks in transmission imaging is challenging since image-domain features characteristic to a certain landmark change substantially depending on the viewing direction. Consequently and to the best of our knowledge, the above problem has not yet been addressed. In this work, we present a method to automatically detect anatomical landmarks in X-ray images independent of the viewing direction. To this end, a sequential prediction framework based on convolutional layers is trained on synthetically generated data of the pelvic anatomy to predict 23 landmarks in single X-ray images. View independence is contingent on training conditions and, here, is achieved on a spherical segment covering $120^\circ \times 90^\circ$ in LAO/RAO and CRAN/CAUD, respectively, centered around AP. On synthetic data, the proposed approach achieves a mean prediction error of 5.6 ± 4.5 mm. We demonstrate that the proposed network is immediately applicable to clinically acquired data of the pelvis. In particular, we show that our intra-operative landmark detection together with pre-operative CT enables X-ray pose estimation which, ultimately, benefits initialization of image-based 2D/3D registration.

1 Introduction

X-ray image guidance during surgery has enabled percutaneous alternatives to complicated procedures reducing the risk and discomfort for the patient. This

B. Bier, M. Unberath, N. Navab and A. Maier—These authors have contributed equally and are listed in alphabetical order.

A. F. Frangi et al. (Eds.): MICCAI 2018, LNCS 11073, pp. 55–63, 2018.
https://doi.org/10.1007/978-3-030-00937-3_7

benefit for the patient comes at the cost of an increased task-load for the surgeon, who has no direct view on the anatomy but relies on indirect feedback through X-ray images. These suffer from the effects of projective simplification; particularly the absence of depth cues, and vanishing anatomical landmarks depending on the viewing direction. Therefore, many X-rays from different views are required to ensure correct tool trajectories [1]. Providing additional, "implicit 3D" information during these interventions can drastically ease the mental mapping from 2D images to 3D anatomy [2,3]. In this case, implicit 3D information refers to data that is not 3D as such but provides meaningful contextual information related to prior knowledge of the surgeon.

A promising candidate for implicit 3D information are the positions of anatomical landmarks in X-ray images. Landmark or key point detection is well understood in computer vision, where robust feature descriptors disambiguate correspondences between images, finally enabling purely image-based pose retrieval. Unfortunately, the above concept defined for reflection imaging does not translate directly to transmission imaging, since the appearance of the same anatomical landmark can vary substantially depending on the viewing direction. Consequently and to the best of our knowledge, X-ray-transform invariant anatomical landmark detection has not yet been investigated. However, successful translation of the above concept to X-ray imaging bears great potential to aid fluoroscopic guidance.

In this work, we propose an automatic, purely image-based method to detect anatomical landmarks in X-ray images independent of the viewing direction. Landmarks are detected using a sequential prediction framework [4] trained on synthetically generated images. Based on landmark knowledge, we can (a) identify corresponding regions between arbitrary views of the same anatomy and (b) estimate pose relative to a pre-procedurally acquired volume without the need for calibration. We evaluate our approach on synthetic validation data and demonstrate that it generalizes to unseen clinical X-rays of the pelvis without the need of re-training. Further, we argue that the accuracy of our detections in clinical X-rays may benefit the initialization of 2D/3D registration.

While automatic approaches to detect anatomical landmarks are not unknown in literature, all previous work either focuses on 3D image volumes [5] or 2D X-ray images acquired from *a single predefined* pose [6,7]. In contrast to the proposed approach that restricts itself to implicit 3D information, several approaches exist that introduce explicit 3D information. These solutions rely on external markers to track the tools or the patient in 3D [8], consistency conditions to estimate relative pose between X-ray images [9], or 2D/3D registration of pre-operative CT to intra-operative X-ray to render multiple views simultaneously [8,10]. While these approaches have proven helpful, they are not widely accepted in clinical practice. The primary reasons are disruptions to the surgical workflow [3], susceptibility to both truncation and initialization due to the low capture range of the optimization target [11].

2 Method

Background: Recently, sequential prediction frameworks proved effective in estimating human pose from RGB images [4]. The architecture of such network is shown in Fig. 1. Given a single image, the network predicts belief maps b_t^p for each anatomical landmark $p \in \{1, \ldots, P\}$. The core idea of the network is that belief maps are predicted in stages $t \in \{1, \ldots, T\}$ using both local image information and long-range contextual dependencies of landmark distributions given the belief of the previous stage. In the first stage of the network, the initial belief b_1^p is predicted only based on local image evidence x using a block of convolutional and pooling layers. In subsequent stages $t \geqslant 2$, belief maps are obtained using local image features x' and the belief maps of the preceding stage. Over all stages, the weights of x' are shared. The cost function is defined as the sum of L2-distances between the ground truth $b_*^p(z)$ and the predicted belief maps accumulated over all stages. The ground truth belief maps are normal distributions centered around the ground truth location of that landmark. The network design results in the following properties: (1) In each stage, the predicted belief maps of the previous stage can resolve ambiguities that appear due to locality of image features. The network can learn that certain landmarks appear in characteristic configurations only. (2) To further leverage this effect, each output pixel exhibits a large receptive field on the input image of 160×160. This enables learning of implicit spatial dependencies between landmarks over long distances. (3) Accumulating the loss over the predicted belief in multiple stages diminishes the effect of vanishing gradients that complicates learning in large networks.

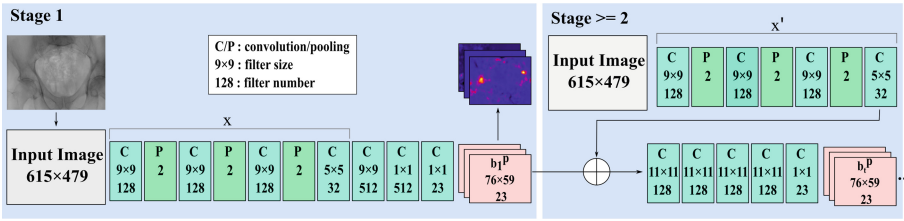

Fig. 1. Network architecture: In subsequent stages, landmarks are predicted from belief maps of the previous stage and image features of the input image. Adapted from [4].

X-ray Transform Invariant Landmark Detection: We exploit the aforementioned advantages of sequential prediction frameworks for the detection of anatomical landmarks in X-ray images independent of their viewing direction. Our assumption is that anatomical landmarks exhibit strong constraints and thus characteristic patterns even in presence of arbitrary viewing angles. In fact, this assumption may be even stronger compared to human pose estimation if limited anatomy, such as the pelvis, is considered due to rigidity. Within this paper

and as a first proof-of-concept, we study anatomical landmarks on the pelvis. We devise a network adapted from [4] with six stages to simultaneously predict 23 belief maps per X-ray image that are used for landmark location extraction (Fig. 1). Implementation was done in tensorflow, with a learning rate of 0.00001, and a batchsize of one. Optimization was performed using Adam over 30 epochs (convergence reached).

Predicted belief maps b_t^p are averaged over all stages prior to estimating the position of the landmarks yielding the averaged belief map b^p. We define the landmark position l_p as the position with the highest response in b^p. Landmarks with responses $b^p < 0.4$ are discarded since they may be outside the field of view or not reliably recognized.

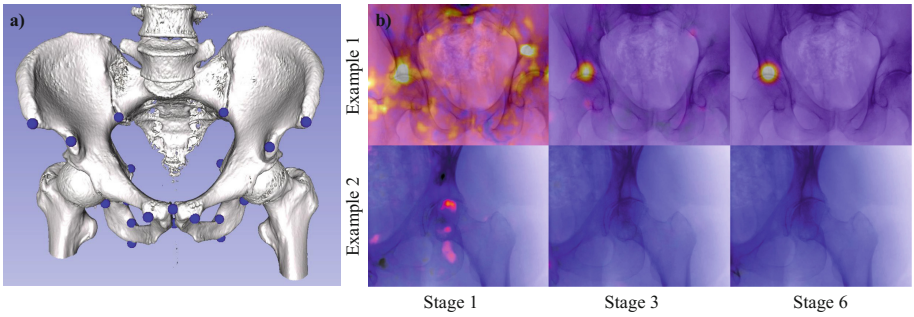

Fig. 2. Uncertainty in the local feature detection after stage 1 is resolved in subsequent stages. This is shown for a symmetric response (Example 1, anterior inferior iliac spine) and for a landmark not in the FOV (Example 2, tip of femoral head).

Training data is synthetically generated from full body CTs from the NIH Cancer Imaging Archive [12]. In total, 20 CTs (male and female patients) were cropped to an ROI around the pelvis and 23 anatomical landmark positions were annotated manually in 3D. Landmarks were selected to be clinically meaningful and clearly identifiable in 3D; see Fig. 2(a). From these volumes and 3D points, projection images and projected 2D positions were created, respectively. X-rays had 615×479 pixels with an isotropic pixel size of 0.616 mm. The belief maps were downsampled eight times. During projection generation augmentation was applied: We used random translations in all three axes, variation of the source-to-isocenter distance, and horizontal flipping of the projections. Further and most importantly, we sampled images on a spherical segment with a range of 120° in LAO/RAO and 90° in CRAN/CAUD centered around AP, which approximates the range of variation in X-ray images during surgical procedures on the pelvis [13]. The forward projector computes material-dependent line integral images, which are then converted to synthetic X-rays [14]. A total of 20.000 X-rays with corresponding ground truth belief maps were generated. Data was split $18 \times 1 \times 1$-fold in training, testing, and validation. We ensured that images of one patient are not shared among sets.

3 Experiments and Results

3.1 Synthetic X-Rays

Experiment: For evaluation, we uniformly sampled projections from our testing volume on a spherical segment covering the same angular range used in training. The angular increment between samples was 5°, soure-to-isocenter distance was 750 mm, and source-to-detector distance was 1200 mm.

Confidence Development: In Fig. 2 the refinement of the belief maps is shown for two examples. After the first stage, several areas in the X-ray image have a high response due to the locality of feature detection. With increasing stages, the belief in the correct landmark location is increased.

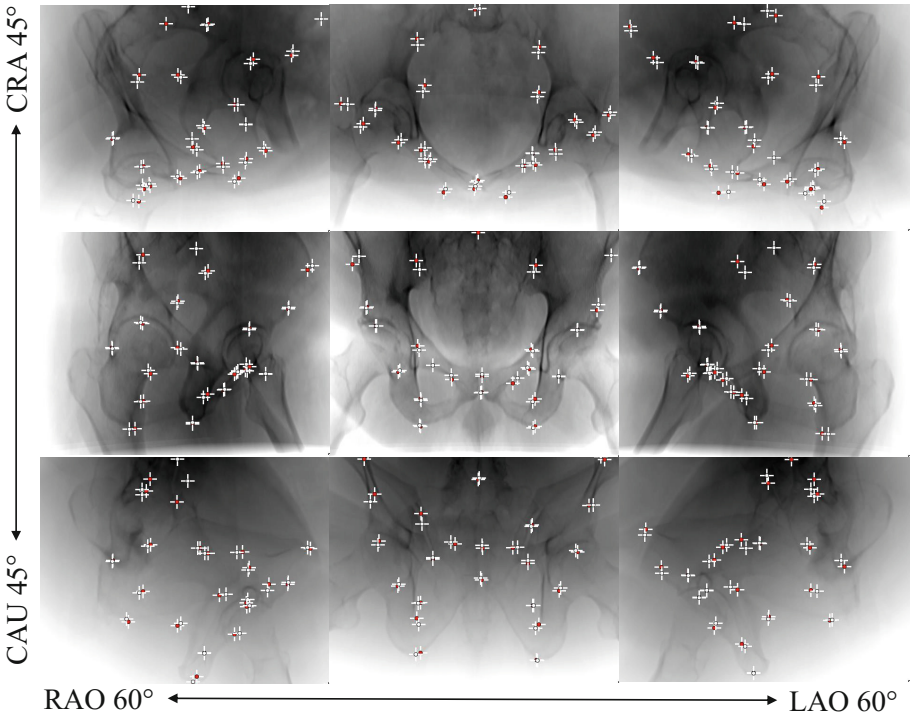

Fig. 3. Detection results over the sampled sphere. White and red marker positions indicate ground truth and predicted landmark location, respectively.

Qualitative Results: In Fig. 3, example X-rays covering the whole angular range are shown. Visually, one notices very good overall agreement between the predicted and true landmark locations[1].

[1] https://camp.lcsr.jhu.edu/miccai-2018-demonstration-videos/.

Belief Map Response and Prediction Error: The maximum belief is an indicator for the quality of a detected landmark. Figure 4 shows the correlation between belief map response and prediction error. As motivated previously, we define a landmark as *detected* if the maximum belief is ≥ 0.4. Then, the mean prediction error with respect to ground truth is 9.1 ± 7.4 pixels (5.6 ± 4.5 mm).

View Invariance: The view invariance of landmark detection is illustrated in the spherical plot in Fig. 4. We define accuracy as the ratio of landmarks with an error < 15 pixels to all detected landmarks in that view. The plot indicates that detection is slightly superior in AP compared to lateral views. To provide intuition on this observation, we visualize the maximum belief of two representative landmarks as a function of viewing direction in Fig. 5. While the first landmark is robust to changes in viewing direction, the second landmark is more reliably detected in CAUD/RAO views.

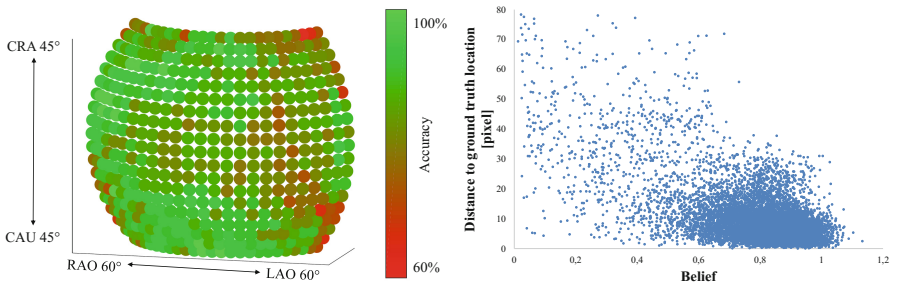

Fig. 4. Left: detection accuracy in dependence of the viewing direction. Right: correlation between belief and prediction error.

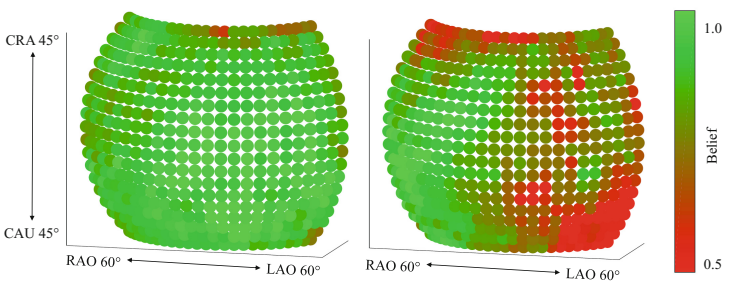

Fig. 5. Belief distribution of two single landmarks. Each landmark has it own detection belief distribution over the sphere.

3.2 Real X-Rays

Landmark Detection: Real X-ray images of a cadaver study were used to test generality of our model *without* re-training. Sample images of our landmark

detection are shown in Fig. 6, top row. Visually, the achieved predictions are in very good agreement with the expected outcome. Even in presence of truncation, landmarks on the visible anatomy are still predicted accurately, see Fig. 6(c). A failure case of the network is shown in Fig. 6(d), where a surgical tool in the field of view impedes landmark prediction.

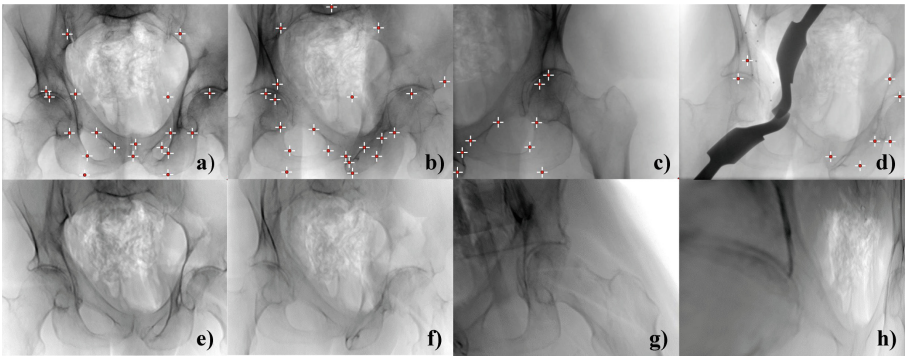

Fig. 6. Top: landmark predictions on clinical X-rays. Bottom: generated projection images after X-ray pose was retrieved using the detections.

Applications in 2D/3D Registration: As candidate clinical application, we study initialization of 2D/3D registration between pre-operative CT and intra-operative X-ray of cadaver data based on landmarks. Anatomical landmarks are manually labeled in 3D CT and automatically extracted from 2D X-ray images using the proposed method. Since correspondences between 2D detections and 3D references are known, the X-ray pose yielding the current view can be estimated in closed form [15]. To increase robustness of the estimation and because belief of a landmark may depend on viewing direction, only landmarks with a belief above 0.7 (but at least 6) are used. The estimated projection matrix is verified via forward projection of the volume in that geometry (Fig. 6, bottom row). While initialization performs well in standard cases where most landmarks are visible and detected, performance deteriorates slightly in presence of truncation due to the lower amount of reliable landmarks, and exhibits poor performance if landmark detection is challenged by previously unseen scenarios, such as tools in the image.

4 Discussion and Conclusions

We presented an approach to automatically detect anatomical landmarks in X-rays invariant of their viewing direction to benefit orthopedic surgeries by providing implicit 3D information. Our results are very promising but some limitations remain. (1) As shown in Fig. 6(d), the performance of our method is susceptible to scenarios not included in training, such as surgical tools in the image.

(2) Lateral views of the pelvis exhibit slightly worse prediction performance compared to AP-like views. We attribute this behavior to more drastic overlap of the anatomy and lower amount of training samples seen by the network. We are confident that this effect can be compensated by increasing the angular range during training while limiting validation to the current range. Since some landmarks are equally well predicted over the complete angular range, the concept of maximum belief is powerful in selecting reliable landmarks for further processing. (3) Downsampling of ground truth belief map limits the accuracy of the detection despite efforts to increase accuracy, e. g. sub-pixel maximum detection. Detecting anatomical landmarks proved essential in automatic image parsing in diagnostic imaging, but may receive considerable attention in image-guided interventions as new approaches, such as this one, strive for clinically acceptable performance. In addition to 2D/3D registration, we anticipate applications for the proposed approach in clinical tasks that inherently involve X-ray images from multiple orientations, in particular K-wire placement.

Acknowledgments. The authors gratefully acknowledge funding support from NIH 5R01AR065248-03.

References

1. Stöckle, U., Schaser, K., König, B.: Image guidance in pelvic and acetabular surgery-expectations, success and limitations. Injury **38**(4), 450–462 (2007)
2. Starr, R., Jones, A., Reinert, C., Borer, D.: Preliminary results and complications following limited open reduction and percutaneous screw fixation of displaced fractures of the acetabulum. Injury **32**, SA45–50 (2001)
3. Härtl, R., Lam, K.S., Wang, J., Korge, A., Audigé, F.K.L.: Worldwide survey on the use of navigation in spine surgery. World Neurosurg. **379**(1), 162–172 (2013)
4. Wei, S.E., Ramakrishna, V., Kanade, T., Sheikh, Y.: Convolutional pose machines. In: CVPR, pp. 4724–4732 (2016)
5. Ghesu, F.C., Georgescu, B., Mansi, T., Neumann, D., Hornegger, J., Comaniciu, D.: An artificial agent for anatomical landmark detection in medical images. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9902, pp. 229–237. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46726-9_27
6. Wang, C.W., Huang, C.T., Hsieh, M.C.: Evaluation and comparison of anatomical landmark detection methods for cephalometric x-ray images: a grand challenge. Trans. Med. Imaging **34**(9), 1890–1900 (2015)
7. Chen, C., Xie, W., Franke, J., Grutzner, P., Nolte, L.P., Zheng, G.: Automatic x-ray landmark detection and shape segmentation via data-driven joint estimation of image displacements. Med. Image Anal. **18**(3), 487–499 (2014)
8. Markelj, P., Tomaževič, D., Likar, B., Pernuš, F.: A review of 3D/2D registration methods for image-guided interventions. Med. Image Anal. **16**(3), 642–661 (2012)
9. Aichert, A., Berger, M., Wang, J., Maass, N., Doerfler, A., Hornegger, J., Maier, A.K.: Epipolar consistency in transmission imaging. IEEE Trans. Med. Imag. **34**(11), 2205–2219 (2015)
10. Tucker, E., et al.: Towards clinical translation of augmented orthopedic surgery: from pre-op CT to intra-op x-ray via RGBD sensing. In: SPIE Medical Imaging (2018)

11. Hou, B., et al.: Predicting slice-to-volume transformation in presence of arbitrary subject motion. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10434, pp. 296–304. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66185-8_34
12. Roth, H., et al.: A new 2.5D representation for lymph node detection in CT. The Cancer Imaging Archive (2015)
13. Khurana, B., Sheehan, S.E., Sodickson, A.D., Weaver, M.J.: Pelvic ring fractures: what the orthopedic surgeon wants to know. Radiographics **34**(5), 1317–1333 (2014)
14. Unberath, M., et al.: DeepDRR-a catalyst for machine learning in fluoroscopy-guided procedures. In: Frangi, A.F., et al. (eds.) MICCAI 2018. LNCS, vol. 11073, pp. 98–106. Springer, Heidelberg (2018)
15. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2004). ISBN 0521540518