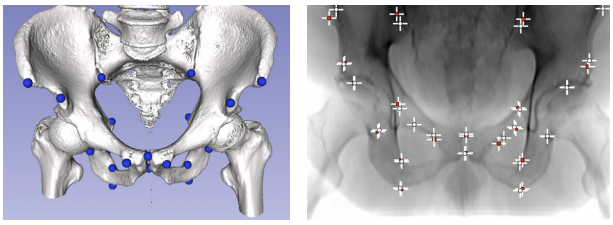
Critical Review: Learning to Detect Anatomical Landmarks of the Pelvis in X-rays from Arbitrary Views

Benjamin Killeen
killeen@jhu.edu



(a) Simulated Model (b) Simulated X-ray

Fig. 1. Anatomical landmarks of the hip anatomy. (a) shows a simulated 3D model of the hip from [1]. (b) shows the simulated X-ray of that model, with the same anatomical landmarks. Images from [1]. Figures from [2].

I. Introduction

Minimally invasive surgery presents a compelling alternative to traditional operating procedure. The reduced incision size results in a lower risk of infection, less blood loss, and cosmetically favorable outcomes for the patient. However, operating through a small incision comes with its own set of challenges for the surgeon. Percutaneous navigation requires intraoperative imaging of anatomical structures. Fluoroscopy is a popular tool for intraoperative imaging, allowing the surgeon to view anatomical structures on a monitor in the operating room. Although surgeons are skilled at operating in this context, the mental burden of aligning the anatomy with a 2D image exacts an exhaustive toll. As such, recent approaches aim to convert the 2D information obtained from fluoroscopic X-ray images into 3D visualizations, possibly registered with a preoperative plan based on a 3D computerized tomography (CT) scan, in order to mitigate the surgeon's mental effort.

A useful and convenient method for performing this 2D/3D registration employs the notion of anatomical landmarks, such as those shown in Fig. 1. Once these landmarks have been accurately located in both the preoperative CT volume as well as the intraoperative X-ray, a basis transformation can be easily derived by solving a system of linearly independent equations. Thus, accurately localizing as many landmarks as possible is highly desirable. Moreover, this localization should be performed as quickly as possible, in order to minimize the interruption to surgical procedure. For these two reasons, recent work has sought to automate anatomical landmark detection.

Paper selection: Here, we review "Learning to Detect Anatomical Landmarks of the Pelvis in X-rays from Arbitrary Views," which presents the first scheme for automated landmark detection suitable for intraoperative imaging [3]. In [3], a stage-based deep neural network (DNN) is used to predict belief maps for each of 23 anatomical landmarks, an approach which has shown remarkable success in the similar problem of human pose estimation. This approach was particularly relevant to our ongoing work, "Improved Generalization of Pelvis X-ray Landmark Detection," which aims to address shortcomings in [3] as it performs on real data.

Key contributions: The key contributions of [3] are (1) a view-invariant data augmentation method using simulated X-ray images, (2) a fully trained DNN architecture for anatomical landmark detection of the hip, achieving an average detection error of 5.6 ± 4.5mm, and (3) successful initialization of 2D/3D registration on real X-ray images. The work is the first known investigation of anatomical landmark detection within the context of view invariance, a property which is required for analysis of fluoroscopic images due to the spatial constraints of the operating room.

II. Background

Anatomical landmarks consist of meaningful and uniquely identifiable points in the anatomy. Establishing correspondences between accurately localized anatomical landmarks is useful for a variety of applications, but [3] is primarily concerned with the 2D/3D registration of a fluoroscopic X-ray image with a 3D preoperative CT scan. This is in order to provide intraoperative feedback to the surgeon in an intraoperative manner. To this end, it is important to clarify that [3] identifies not just locations of anatomical landmarks on an X-ray image but actually the location of specific, predefined anatomical landmarks. This allows a point-matching correspondence to be established between the intraoperative X-ray and the CT. As [3] discusses, this enables the computation of the projection matrix $\mathbf{P} \in \mathbb{R}^{3\times4}$ between the two bases. If we denote the ordered set of 2D detections as homogeneous points $\{\mathbf{d}_n \in \mathbb{R}^3 | n \in [1,\dots,N]\}$ and the corresponding ordered set of homogeneous vectors $\{\mathbf{r}_n \in \mathbb{R}^3 | n \in [1,\dots,N]\}$, then we can establish the following set of linearly independent

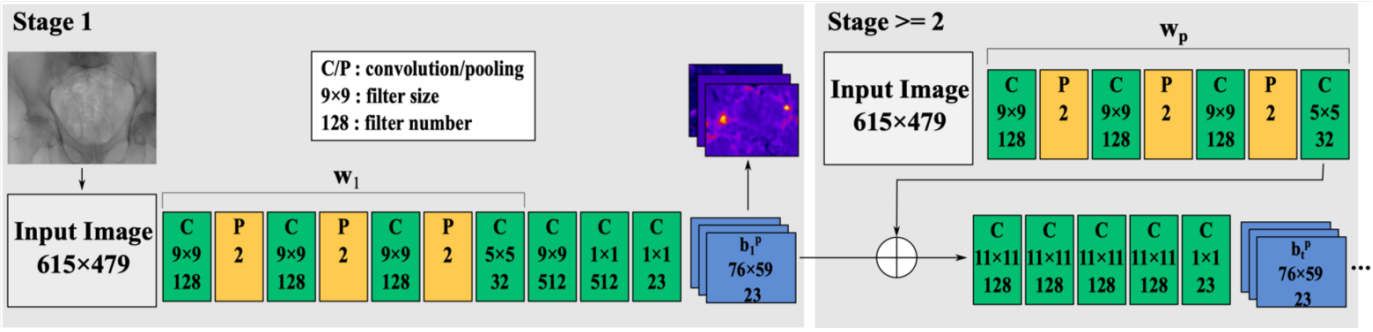


Fig. 2. The stage-based DNN, from [3], for predicting and refining belief maps for each anatomical landmark. Each channel p in the output image corresponds to the belief map for that landmark. Figure from [3].

equations:

$$\begin{bmatrix} 0^T & -w_i\mathbf{r}_i^T & y_i\mathbf{r}_i^T \\ w_i\mathbf{r}_i^T & 0^T & -x_i\mathbf{r}_i^T \end{bmatrix} \begin{pmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \mathbf{p}_3 \end{pmatrix} = \mathbf{0} \tag{1}$$

where $\mathbf{d}_n = (x_n, y_n, w_n)$, and N is the number of corresponding landmarks.

Thus, accurately localizing the points $\mathbf{d}_n$ is of the utmost importance. Manually localizing these points is undesirable, both because of the potential for error during an operation as well as the time required, which interrupts the surgical procedure.

Automated landmark detection, which involves the application of computer vision algorithms to identify and localize each landmark, has received some attention from recent work. [3] reviews these efforts, and we shall refer to their treatment of the subject for a more complete survey. In brief, several approaches predict the landmark positions directly on the X-ray image, possibly refining these estimates with a secondary model-fitting step [3]. Another approach uses a generative decision tree model to identify landmarks in the hand, refining predictions in an iterative manner [4]. Some approaches [5], [6] even use DNN architectures, such as the U-net [7], to predict belief maps, which we describe in Sec. III, for anatomical landmarks on the chest and spine, respectively.

All of these methods, however, focus on anatomical landmark detection from a single view. That is, they assume each X-ray image is taken with the same orientation with respect to the anatomy. For some applications, this is a reasonable assumption, but in general—and especially for fluoroscopic imaging—it imposes a severe constraint on the imaging procedure [3]. This is especially undesirable for intraoperative procedures, where the possible view angles may be constricted by surgical tools and other apparatus. As such, [3] introduces significant data augmentation, leveraging CT scans, that results in view-invariant landmark detection. We discuss this method below in Sec. III.

III. Method

Following prior work on human pose estimation, [3] uses a multi-stage DNN architecture to predict and refine belief maps. A belief map ^{p}, where $p \in [1, \ldots, P]$ indexes the landmarks, is a pixel-wise likelihood model giving the "belief" that landmark p is present in a given location. Each stage of the DNN, shown in Fig.2, predicts a $76 \times 59 \times 23$ image containing the belief maps for each of the landmarks. In [3], the output resolution is 76×59, and $P = 23$. As can be seen in Fig. 2, the belief maps in the initial stage are not very exact. Subsequent stages refine this initial guess by continually incorporating global information from the large receptive field of the previous layer with local features directly from the initial image.

Training this network requires significant labeled data. Since manually labeled data is expensive and also contains the possibility for human error—due to the variability of landmark appearance from multiple viewpoints— [3] synthetically generates training data from full body CTs of the NIH Cancer Imaging Archive [8]. In total, they manually label 23 landmarks in each of 20 full-body CT images from male and female patients. This allows the authors to forward project an X-ray image from the CT volume, including the landmark positions, resulting in a labeled X-ray image of the hip anatomy with 615×479 pixels and an isotropic pixel spacing of 0.616mm [3]. The corresponding ground truth belief maps contain normal distributions centered on these landmarks. In total, [3] generate 20,000 X-rays, and they responsibly divide these images into training, validation, and test data in order to validate their results.

IV. Experiments

[3] test their approach on both simulated and real data. On simulated data, they achieve a mean detection error of 5.6 ± 4.5mm. They also provide detection accuracy curves for successive stages, showing that this stage-based refinement of initial guesses does indeed improve detection, with diminishing returns after a few stages. They also provide a visualization of the detection results for X-rays viewed at various angles. They find that, for simulated data at least, their method performs best for X-rays taken away from the edges of their training data. That is, the viewpoint angles which are at the edge of the angle set that they've chosen are those on

which detection fared poorly. From this, one might infer that the DNN benefited not just from each viewpoint on its own but also viewpoints in the immediate vicinity of angle-space. The authors suggest extending training data beyond the angle-space they used, then validating on a narrower set of viewpoints [3].

On real X-rays, [3] use their DNN-based landmark detection to initialize a traditional 2D-3D registration, the results of which are shown in Fig. 3. As is shown there, their method generalizes well to the real X-rays, but they still require a secondary algorithm to refine the registration. Furthermore, the DNN struggles to adapt to unseen situations, such as surgical tool occlusion or anatomical anomaly caused by fracture. Nevertheless, as a proof-of-concept, their method stands out as a first-of-its-kind approach for automated landmark detection in a mult-view manner suitable to intraoperative imaging procedure.

V. Assessment

We find that the results in [3] are an impressive foundation on which to build. In particular, the successful generalization of their DNN trained on simulation data to real X-rays serves as a significant baseline which our effort aims to improve. We note that the shortcomings discussed in Sec. IV are a significant obstacle to implementation in a clinical setting. Surgical tool occlusion is certainly a given for fluoroscopic images taken during a minimally invasive procedure.

Relevance to ongoing work: The inconvenience of removing surgical tools for every registration far outweighs any potential advantage gained from 3D visualizations of the patient anatomy with respect to a preoperative plan. Our method, which we discussed in our project proposal, involves local patch-based normalization which allows for global receptive field to inform local behavior without corrupting local features. In this way, a surgical tool in one part of the image may not corrupt the detection of landmarks in another part, as it does in Fig. 3.

VI. Conclusion

We have provided a critical review of [3], "Learning to Detect Anatomical Landmarks of the Pelvis in X-rays from Arbitrary Views." We have briefly discussed their method, which involves a stage-based DNN that predicts and refines belief maps for each anatomical landmark, as well as their unique data augmentation strategy, which simulates X-ray images from multiple views based on CT volumes. We also point out shortcomings in their method and aim to improve on their transfer from simulated X-rays to real images, perhaps foregoing the need for an additional registration algorithm initialized by our procedure.

References

[1] M. Unberath, J.-N. Zaech, C. Gao, B. Bier, F. Goldmann, S. C. Lee, J. Fotouhi, R. Taylor, M. Armand, and N. Navab, "Enabling machine learning in X-ray-based procedures via realistic simulation of image formation," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 9, pp. 1517–1528, Sep. 2019.

[2] B. Bier, M. Unberath, J.-N. Zaech, J. Fotouhi, M. Armand, G. Osgood, N. Navab, and A. Maier, "X-ray-transform Invariant Anatomical Landmark Detection for Pelvic Trauma Surgery," *arXiv:1803.08608 [cs]*, Mar. 2018.

[3] B. Bier, F. Goldmann, J.-N. Zaech, J. Fotouhi, R. Hegeman, R. Grupp, M. Armand, G. Osgood, N. Navab, A. Maier, and M. Unberath, "Learning to detect anatomical landmarks of the pelvis in X-rays from arbitrary views," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 9, pp. 1463–1473, Sep. 2019.

[4] M. Urschler, T. Ebner, and D. Štern, "Integrating geometric configuration and appearance information into a unified framework for anatomical landmark localization," *Medical Image Analysis*, vol. 43, pp. 23–36, Jan. 2018.

[5] A. O. Mader, J. von Berg, A. Fabritz, C. Lorenz, and C. Meyer, "Localization and Labeling of Posterior Ribs in Chest Radiographs Using a CRF-regularized FCN with Local Refinement," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, ser. Lecture Notes in Computer Science, A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, Eds. Cham: Springer International Publishing, 2018, pp. 562–570.

[6] C.-W. Wang, C.-T. Huang, M.-C. Hsieh, C.-H. Li, S.-W. Chang, W.-C. Li, R. Vandaele, R. Marée, S. Jodogne, P. Geurts, C. Chen, G. Zheng, C. Chu, H. Mirzaalian, G. Hamarneh, T. Vrtovec, and B. Ibragimov, "Evaluation and Comparison of Anatomical Landmark Detection Methods for Cephalometric X-Ray Images: A Grand Challenge," *IEEE transactions on medical imaging*, vol. 34, no. 9, pp. 1890–1900, Sep. 2015.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, ser. Lecture Notes in Computer Science, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[8] A. Seff, L. Lu, A. Barbu, H. Roth, H.-C. Shin, and R. M. Summers, "Leveraging Mid-Level Semantic Boundary Cues for Automated Lymph Node Detection," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. Frangi, Eds. Cham: Springer International Publishing, 2015, vol. 9350, pp. 53–61, series Title: Lecture Notes in Computer Science.

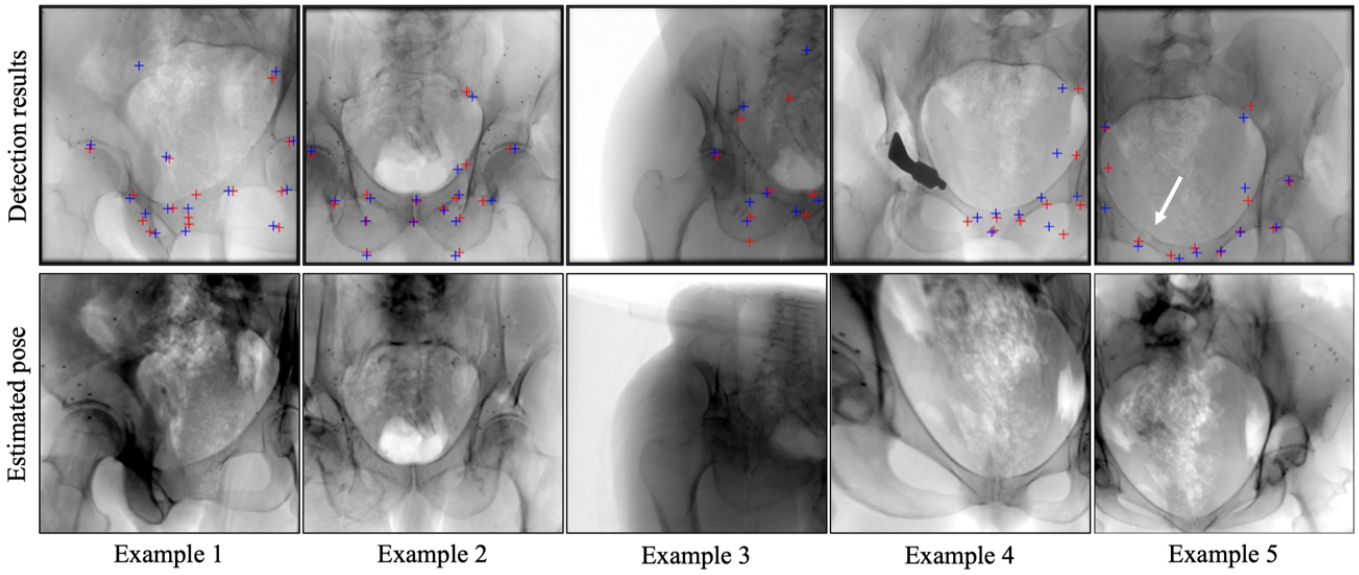


Fig. 3. Results of 2D/3D registration using the DNN-based anatomical landmark detections as initialization for a traditional registration algorithm, in [3]. As can be seen, the method struggles in the presence of unseen situations, such as occlusions by surgical tools (Example 4) and anatomical anomalies such as fracture (Example 5). Figure from [3].