DETECTION AND 3D LOCALIZATION OF SURGICAL INSTRUMENTS FOR IMAGE-GUIDED SURGERY

by

Irina Bataeva

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science in Engineering

Baltimore, Maryland

May 2021

Abstract

Placement of surgical instrumentation in pelvic trauma surgery is challenged by complex anatomy and narrow bone corridors, relying on intraoperative x-ray fluoroscopy for visualization and guidance. The rapid workflow and cost constraints of orthopaedic trauma surgery have largely prohibited widespread adoption of 3D surgical navigation. This thesis reports the development and evaluation of a method to achieve 3D guidance via automatic detection and localization of surgical instruments (specifically, Kirschner wires [K-wires]) in fluoroscopic images acquired within routine workflow. The detection method uses a neural network (Mask R-CNN) for segmentation and keypoint detection of K-wires in fluoroscopy, and correspondence of keypoints among multiple images is established by 3D backprojection and a rank-ordering of ray intersections. The accuracy of 3D K-wire localization was evaluated in a laboratory cadaver study as well as patient images drawn from an IRB-approved clinical study. The detection network successfully generalized from simulated training and validation images to cadaver and clinical images, achieving 87% recall and 98% precision. The geometric accuracy of K-wire tip location and direction in 2D fluoroscopy was 1.9 ± 1.6 mm and 1.8° ± 1.3°, respectively. Simulation studies demonstrated a corresponding mean error of 1.1 mm in 3D tip location and 2.3° in 3D direction. Cadaver and clinical studies demonstrated the feasibility of the approach in real data, although accuracy was reduced to with 1.7 ± 0.7 mm in 3D tip location and 6° ± 2° in 3D direction. Future studies aim to improve performance by increasing the volume and variety of images used in neural network training, particularly with respect to low-dose fluoroscopy (high noise levels) and complex fluoroscopic scenes with various types surgical instrumentation. Because the approach involves fast runtime and uses equipment (a mobile C-arm) and fluoroscopic images that are common in standard workflow, it may be suitable to broad utilization in orthopaedic trauma surgery.

Primary Reader and Advisor: Dr. Jeffrey H. Siewerdsen, John C. Malone Professor, Department of Biomedical Engineering, Johns Hopkins University.

Secondary Readers: Dr. Mehran Armand, Professor, Department of Mechanical Engineering, Johns Hopkins University; and Dr. Ali Uneri, Faculty Research Associate, Department of Biomedical Engineering, Johns Hopkins University.

Acknowledgements

Table of Contents

List of Tables

List of Figures

Chapter 1: Background and Motivation

1.1 Orthopaedic Surgery of Pelvic Fractures

Pelvic fractures account for 2-8% of all fractures in the United States and are observed across all demographics.[1] Incidence in younger individuals is mostly due to high-energy trauma such as traffic accidents, while the primary cause in older patients is injury due to falls. Osteoporosis predisposes an individual to a risk of pelvic fracture, and therefore the incidence increases with age and may be higher for women.[2] Although some pelvic fractures can be treated with external stabilization, the standard of care for treating unstable pelvic fractures involves surgical intervention and internal fixation. Pelvic fixation is achieved by first placing guidewires – specifically Kirchner wires (K-wires) – across the fracture line, followed by insertion of cannulated screws that follow the wire path. Minimally invasive surgery (MIS) is becoming an increasingly popular approach, as it is associated with decreased blood loss, risk of infection, and recovery time compared to open surgery.[3] The percutaneous insertion of K-wires in MIS, however, is challenging even for experienced surgeons due to the complex morphology of the pelvis with long and narrow bone corridors.

Surgical navigation[4] is an enabling approach to MIS that allows visualization of instruments within the patient as rendered virtually in a volumetric image – commonly computed tomography (CT) or cone-beam CT (CBCT). Navigation is frequently accomplished using surgical trackers (e.g., electromagnetic [EM] or optical [infrared or video] trackers) to provide accurate, real-time 3D localization of surgical instruments. EM tracking is challenged by the large operating fields of pelvic trauma surgery (exceeding the field of measurement for the tracker) and – especially – by geometric distortions induced by the presence of metal frames and instrumentation. Optical trackers are limited in application to flexible instruments due to rigidity assumption, and the long (~20-30 cm), thin K-wires used in orthopaedic trauma surgery

commonly bend within the patient during insertion. Moreover, ensuring a clear line-of-sight to the tracker camera can also be challenging for the wide variations in patient setup for pelvic surgery. For both EM and optical tracking, the workflow associated with the setup and registration of trackers has limited the adoption of surgical navigation in the fast-paced, dynamic workflow of orthopaedic trauma surgery.

The current standard of care for guiding orthopaedic instruments involves intraoperative fluoroscopy to visualize and cognitively / qualitatively assess the position of instruments relative to patient anatomy. However, x-ray fluoroscopy involves radiation dose to both the patient and operating staff, with up to 2 minutes of radiation exposure per screw presenting a justifiable concern.[5] Even with fluoroscopic guidance, achieving accurate K-wire insertion is challenging, as some bone corridors are just 1 cm wide[6]. As the cannulated screws follow the path of the inserted K-wires, screw malpositioning due to suboptimal K-wire placement is a common complication of pelvic fracture stabilization surgery: 20-30% of screw placements are rated as suboptimal,[7] and 6% breach the cortical bone,[8] which can require additional revision surgeries, result in neurological and vascular injuries, and can lead to long-term pelvic instability.[6]

1.2 3D Localization of Surgical Instruments

1.2.1 Image Registration for Surgical Guidance

Accurate knowledge of the location of surgical instruments relative to patient anatomy is critical to providing stable fixation and ensuring beneficial clinical outcomes. In minimally invasive trauma surgery, surgeons do not have direct visualization of the surgical target (as in open procedures), and surgical navigation (e.g., via trackers) is not widely adopted, as noted above; therefore, intraoperative imaging is an important tool to visualize and confirm the placement of instrumentation.

In some scenarios, intraoperative imaging involves 3D volumetric images – e.g., intraoperative magnetic resonance (MR) imaging, CT, or CBCT. With the possible exception of intraoperative CBCT, the availability

of intraoperative 3D imaging is not particularly prevalent in orthopaedic trauma settings. In this context, intraoperative CBCT can be used to directly localize the position of surgical instruments in 3D, recognizing that each 3D scan requires time and radiation dose. Moreover, 3D-3D image registration can be used to align information defined in preoperative 3D imaging (e.g., preoperative planning of K-wire trajectories) with the intraoperative 3D image. 3D-3D rigid or deformable registration of such images can be based on registration of point features, segmented surfaces, or image intensities (and gradients) solved via numerical optimization.[9–11]

Augmented reality (AR) approaches present an interesting alternative form of intraoperative guidance, presenting virtual representations of tools within a view of the operating field. The application of AR in orthopaedic surgery is emerging, and while such technology is still far from widespread utilization, a number of potentially valuable implementations and use cases have been identified.[12] For example, video markers attached to the skin and registered to preoperative CT have been used to superimpose a volumetric view of the patient's bone anatomy relative to the patient's skin via an AR camera-projector.[13] Similarly, the trajectory of a surgical needle in bone can be visualized relative to preoperative CT in a head-mounted display.[14] Such AR approaches offer new forms of visualization and interaction that may present advantages in terms of a more natural interface to complex 3D image data; however, they tend to suffer some of the same limitations mentioned above for conventional surgical navigation – e.g., insensitive to deformation of tissues during surgery – and some unique challenges remaining to be fully addressed – e.g., cumbersome head-mounted displays, sensitivity to motion of markers placed on the skin, and visual fatigue in long cases.

X-ray fluoroscopy is by far the most prevalent imaging modality for guidance of orthopaedic trauma surgery. Registration of information from other images (e.g., from preoperative CT) with the fluoroscopic scene can provide valuable augmentation of the image. For example, the optimal K-wire path as defined in preoperative CT can be registered to the fluoroscopic scene by 3D-2D registration.[15,16] Registration of

2D projection images to 3D volume images (or models) solves the (rigid) transformation between the projection of an object in 2D and its 3D model.[17] This solution naturally fits the workflow of MIS pelvic trauma procedures, using intraoperative fluoroscopy already within the standard of care and combining with information defined in preoperative CT, intraoperative CBCT, and/or 3D models of surgical instruments.

1.2.2 3D-2D Registration Using Physical Models

3D-2D image registration is an active area of research for performing 3D localization of surgical instruments from 2D intraoperative images like fluoroscopy[17] and ultrasound.[18] Many 3D-2D registration methods make use of optimization-based iterative approaches – for example, maximizing a maximizing an objective function (e.g., similarity in intensity and/or gradient information) by iteratively matching the projection of a 3D model with the fluoroscopic image.[19] For registering surgical instruments, the knowledge of the object shape is key to creating its projection. A large body of work demonstrated this approach using optimization-based 3D-2D registration of rigid, deformable, parametric, and exact models of the surgical instruments.[19–21]

For example, Goerres et al.[21] demonstrated the feasibility of 3D-2D registration of K-wires on pelvic radiographs using the deformable Known-Component Registration (dKC-Reg) algorithm. The accuracy achieved in K-wire localization with two or more fluoroscopic images was ~1 mm in tip and ~1° in direction, as demonstrated in cadaver studies. The algorithm runtime, however, was 1-3 minutes per K-wire, relied on a prior surgical plan for initialization, and was susceptible to other radiopaque instruments present in the radiographic scene.

1.2.3 3D-2D Registration Using Deep Learning Approaches

Deep learning, specifically approaches using convolutional neural networks (CNNs), has seen unprecedented growth and adoption in recent years in computer vision and medical imaging.[22] Compared

to traditional approaches, CNNs trained on large datasets and running on modern graphical processing unit (GPU) hardware can solve complex problems with fast inference times that make them particularly suitable for intraoperative tasks. Supervised training of deep learning networks, however, requires large, annotated datasets, which are not commonly available in many surgical applications due to variations in procedure workflows, medical imaging equipment, surgical hardware, and the relatively resource-intensive process of acquiring and annotating clinical datasets.

The use of a CNN to segment guidewires on radiographs has been reported in previous literature. Martin et al.[23] and Gherardini et al.[24] demonstrated guidewire segmentation in cardiac interventions using U-Net type networks trained on simulated guidewires with real anatomical backgrounds. This type of guidewire detection was shown to operate well in radiographs for which only a single instrument is present. For more complex scenes that are commonly encountered in orthopaedic surgery, semantic segmentation in which the network outputs whether the pixel belongs to an object (but does not differentiate multiple objects) has limited utility, as it would not be able to distinguish between multiple instruments and would invite extensive post-processing to separate the segmentations.

In spinal neurosurgery, Esfandiari et al.[25] demonstrated that by using radiographs acquired from disparate view angles, a deep learning model could be applied for detection of the surgical implants (specifically, spinal pedicle screws) and subsequent post processing for establishing correspondence and 3D pose. The authors used a U-Net type network for semantic segmentation of the images trained on a dataset of simulated pedicle screws. The segmentations were post-processed to extract the screw centerline, the ends were projected onto the other view, and the distances to the projected epipolar lines were used to find the correspondent implants. The average screw tip error was 1.9 ± 0.6 mm with an average tunnel axis error of 1.9° ± 0.6° on studies in an animal specimen. This approach did not rely on a precise knowledge of the instrument geometry, which is an advantage for instruments like K-wires that exhibit numerous variations in shape based on the particular application. One disadvantage is that the method

does not differentiate between multiple pedicle screw detections and therefore requires post-processing to separate them, which can be a challenge in complex scenes. Nevertheless, the approach demonstrated the feasibility of instrument localization from multiple radiographs given the accurate segmentation of an instrument with a deep learning model.

For orthopedic applications, Kugler et al.[26] demonstrated that a 3D pose of a pedicle screw, a drill, and robot manipulator pose can be inferred from a single radiograph with a CNN (dubbed "i3PosNet") trained on a simulated dataset. The i3PosNet was trained to predict pseudo landmarks (keypoints) on radiographs of surgical hardware of a known shape from which it is possible to predict the 3D pose. On real images, the method produced high accuracy in position localization of instruments of 0.8 ± 0.8 mm and a forward angle error of 6.6° ± 10.4°. The relatively high angle error was attributed to end-on views with steep forward angle. The utility of i3PosNet would likely be challenged on instruments such as K-wires due to a lack of features, the thin shape of the wire, and deformability, which would likely exacerbate the forward angle error further.

Deep learning models for instance segmentation – multiple segmentations that are created for each object on the image separately – should, in principle, distinguish between multiple and even intersecting K-wires on pelvic radiographs and would therefore be preferable for use in the localization of surgical instruments. Mask R-CNN[27] is a popular instance segmentation deep learning model that has been increasingly tested in medical applications.[28,29] Moreover, Mask R-CNN is capable of keypoint detection by treating a keypoint as a single-pixel segmentation. Utilizing this feature would allow the network to be trained directly to detect the tip of the K-wire in a manner that is quick and accurate. We thus hypothesize that Mask R-CNN be trained for K-wire tip detection and segmentation. With accurate keypoint detections in two or more 2D fluoroscopy views, we further hypothesize that the 3D pose of the K-wires can be determined (i.e., the K-wires can be localized) by a backprojection method augmented with a sorting procedure to establish correspondence among multiple detections.

6

1.3 Thesis Overview

1.3.1 Thesis Statement

Fast 3D localization of instruments from 2D fluoroscopy can be achieved by detecting instrument tips and directions using a CNN, establishing correspondence on multiple views, and backprojecting to obtain 3D instrument pose. Such detection and localization could provide the basis for surgical guidance based on fluoroscopy alone, without the need for surgical trackers or 3D imaging.

1.3.2 Thesis Outline

The dissertation pursues the thesis in systematic development and evaluation as follows:

Chapter 2 details the method for detecting K-wires on x-ray radiographs and localizing their tips and directions in 3D.

Chapter 3 demonstrates the performance of the approach in experimental studies on cadaver and clinical datasets, as well as the results of the simulation studies.

Chapter 4 discusses the presented results and outlines potential directions for future research.

1.3.3 Scientific Communications

The research contained in this thesis was reported in part at the 2021 SPIE Medical Imaging Symposium, including the following conference proceedings paper and poster presentation:

> I. Bataeva et al., "Intraoperative guidance of orthopaedic instruments using 3D correspondence of 2D object instance segmentations", Proc. SPIE Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling, Vol 11598, DOI: 10.1117/12.2582239 (2021).

Figures appearing in the proceedings paper have been adapted and included in this thesis with permission from the publisher as cited at each instance.

Chapter 2: Algorithm and Experiments

The approach for 3D localization of tip and direction of pelvic guidewires is illustrated in Figure 1. Two fluoroscopic images that capture the target instruments are input to a pretrained Mask R-CNN model.[27] For each image, the network outputs predictions of 2D tip locations (keypoints) and instrument segmentations (masks). The direction vector for each guidewire is obtained via simple post-processing of its segmentation. Correspondence among keypoints on individual images is established using knowledge of the imaging system geometry to backproject, identify, and rank-order ray intersections. For each instrument, the 3D direction is calculated via backprojection of the corresponding 2D direction vectors. The estimated 3D guidewire tips and directions can then be overlaid on preoperative CT to provide surgical navigation.
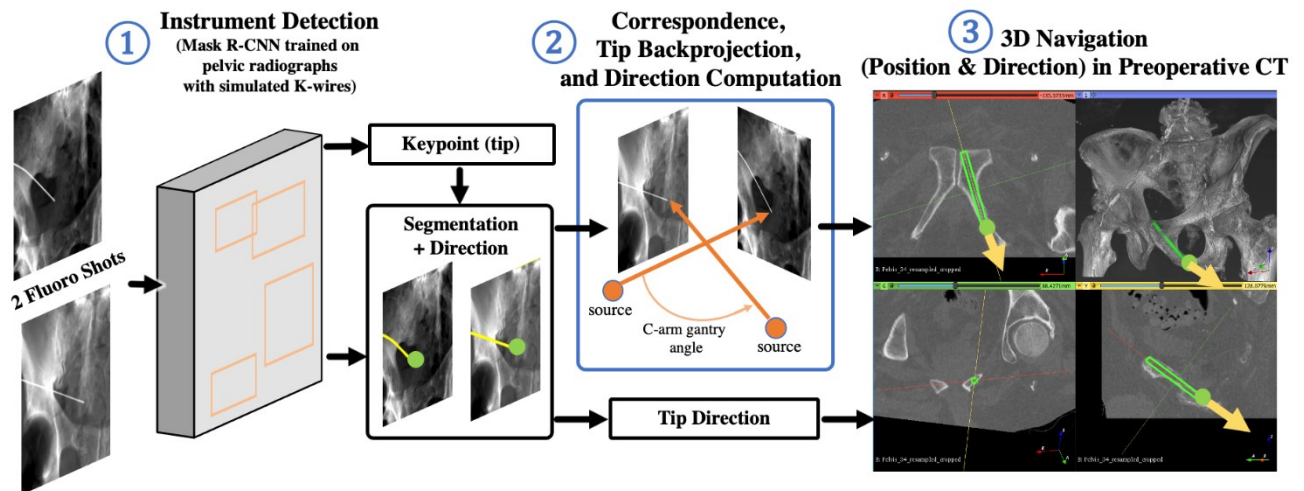
Figure 1. Algorithm flowchart depicting: (1) instrument detection using Mask R-CNN; (2) object correspondence and subsequent 3D instrument pose estimation; and (3) visualization of instrument pose on tri-planar views of a preoperative 3D CT image. Figure adapted from Bataeva et al. with permission from the publisher.

2.1 K-wire Detection

2.1.1 Mask R-CNN

Detection of K-wires used an extended version of Mask R-CNN[27] that was adapted for concurrent segmentation and keypoint detection within the bounding box for each object instance. The network uses ResNet-50, a 50-layer CNN that utilizes residual connections between convolutional layers. The architecture has demonstrated good performance in a variety of image classification tasks[30]. ResNet-50 was pretrained on the COCO dataset[31], a dataset of hundreds of thousands labeled photographic images with 91 classes of common objects and their segmentations, providing a feature extraction layer for Mask-RCNN. The subsequent layers propose regions, pool, and align them on the image grid to yield (regress) bounding box coordinates for each detected object (in this work, each surgical instrument). Two separate heads were used to output the instance segmentation and the single-pixel keypoint segmentations within each box. The output therefore includes the object bounding boxes, a segmentation mask for each box, a keypoint, and an associated detection score (0-100%).

The instrument tip location on the image is simply designated by the output keypoint. To obtain the instrument direction on the image, each segmentation is first clipped to eliminate pixels with <10% detection score, and a third-order weighted B-spline is then fit to the clipped segmentation using the scores as weights for each pixel. The first-order derivative at the B-spline endpoint (i.e., closest to the keypoint) determined the direction of the K-wire in detector coordinates.

2.1.2 Model Training

Training dataset

Sixteen CBCT scans of different regions of two cadaver pelves (without implants) were acquired using a 3D-capable mobile C-arm (Cios Spin, Siemens Healthineers, Forchheim, Germany). The raw projections underlying the scans were randomly selected, log-transformed to enhance the dynamic range of the

image, and image intensities were inverted such that highly attenuated regions (e.g., K-wires) have higher pixel intensities. The edges of the fluoroscopy images were cropped by 8 pixels to remove the high-contrast borders (collimator edges), yielding images with 960 × 960 pixels and 0.304 × 0.304 mm^2 pixel spacing. Affine transformations were applied to increase the statistical variation of the background anatomy, consisting of ±2.5° rotations, ±3 mm translations, and ±20% scaling – all sampled from uniform distributions. Images were padded following affine transformation as necessary by mirroring edge pixels. The resulting 4,400 images present the background anatomical context for simulated K-wires, with 4,000 images used for training and 400 images used for validation of the network (10:1 split).

Instrument modeling and forward projection

Pelvic K-wires were modeled as 3D triangular meshes, each consisting of a cylinder with a principal axis defined by a third-order B-spline curve with a conical tip attached to one end representing the K-wire tip. The location of the B-spline middle control point was uniformly sampled between 1/3 and 2/3 of the cylinder length, with its deviation from the principal axis uniformly sampled between 0 and 30 mm along the x and y axes. Other model variations included (Figure 2a): 1-1.9 mm radius of the cylinder (r); 20-40 cm cylinder length (l_{wire}); 1-4 mm length for the conical tip (l_{tip}), all of which were sampled according to a uniform distribution. The K-wire models were additively forward projected[1] onto radiographs (Figure 2b-c) using the geometric calibration of the imaging system detailed in §2.2.1. The intensity of the projections was randomized according to a uniform distribution to account for different imaging protocols and implants of different material compositions.

One-to-six K-wires were simulated on each image. The current work was limited to wires with one end (tip) captured in the image, excluding cases with no ends or both ends present in the image. The resultant

---

[1] The forward projection algorithm was developed by Dr. Ali Uneri (Department of Biomedical Engineering, Johns Hopkins University).

instrument forward projections were used to define the segmentations of the K-wires, from which the corresponding bounding boxes were created by taking the lowest left corner (v_x, v_y) and the highest right corner of the segmentation (v_{x1}, v_{y1}) to define a bounding box $v = (v_x, v_y, v_{x1} - v_x, v_{y1} - v_y) = (v_x, v_y, v_w, v_h)$. The endpoint on the conical tip was separately projected onto the image to define the location of the keypoint (Figure 2c).
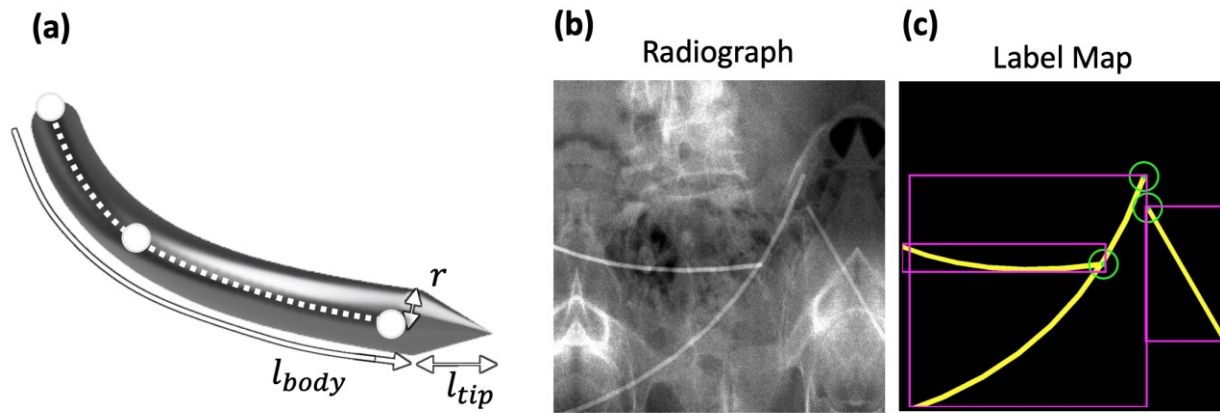
Figure 2. K-wire shape model used for generating annotated training data. (a) 3D mesh model of the guidewire annotated with pertinent parameters, including: (b) Sample radiograph from the training dataset with (c) corresponding bounding boxes (magenta), guidewire segmentations (yellow), and tips as keypoints (green). Figure adapted from Bataeva et al. with permission from the publisher.

Network training

The network was trained on the 4,000 images of real pelvic radiographs augmented with simulated K-wires, thus leveraging a transfer learning approach to generalize to real images of K-wires. Data augmentation was used during the training process to ensure that no image was seen by the network more than once, specifically at each epoch: Gaussian noise was added with 0-20% standard deviation of image pixel intensities; and images were flipped left-right with a 50% chance. The model was trained for 75 epochs using the Adam optimizer with a learning rate of 10^{-4} and batch size of 20.

The loss function of the model was a multi-task loss[27]:

$$L = L_{\text{cls}} + L_{\text{bb}} + L_{\text{seg}} + L_{\text{keypoint}} \tag{1}$$

where class loss is a log loss for the correct class u:

$$L_{\text{cls}} = -\log p_u \tag{2}$$

with $p = (p_o, p_1)$ being the probability distribution of class 0 (background) and 1 (K-wire) per region.

Bounding box loss was defined only for the K-wire class:

$$L_{bb} = \sum_{i \in \{x,y,w,h\}} \text{smooth}_{L_1}(t_i - v_i) \tag{3}$$

where $t = (t_x, t_y, t_w, t_h)$ is the predicted bounding box, $v = (v_x, v_y, v_w, v_h)$ is the target, and

$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 \\ |x| - 0.5 \end{cases}$ is a robust L1 loss. Segmentation loss was:

$$L_{\text{seg}} = \frac{1}{N} \sum_{i \in 1} y_i * \log p(y_i) + (1 - y_i) * \log(1 - p(y_i)) \tag{4}$$

where N is the number of segmentation pixels, y_i is the pixel label (0 for background and 1 for the instrument). Keypoint loss, L_{keypoint}, was identical to L_{seg} with $N = 1$.

2.2 3D Localization

2.2.1 Backprojection

Each K-wire tip from each view was backprojected according to the pre-calibrated imaging system geometry, to obtain the ray that originates at the radiation source, passes through the tip of the K-wire, and hits the detector. Specifically, each ray was defined by a source location s_i in 3D coordinates for the

view i and a 3D direction $\vec{v}_{i,j}$ vector pointing from the source to the jth detected tip on the radiograph according to the 3 × 4 projection matrix P_i[2] for view i:

$$s_i = P_i^{(1...3)} P_i^{(4)} \tag{5a}$$

$$\vec{v}_{i,j} = \left(P_i^{(1...3)}\right)^{-1} [k_{i,j}, 1]^T \tag{5b}$$

where $P_i^{(k)}$ defines the kth column of matrix P_i. The keypoint output, $k_{i,j}$, was redefined relative to the image center and converted to millimeters via the pixel spacing.

2.2.2 Ray Intersection

Assuming a perfect detection of the K-wire tips and a perfectly reproducible imaging system geometry, a 3D tip location would be defined by the intersection of corresponding backprojected rays from two or more fluoroscopic views. Due to random and/or systematic errors expected in the detection and C-arm calibration, however, the rays are not expected to form a perfect intersection. Instead, the 3D tip location (t_j) was defined in a least-squares sense by solving for the point that minimizes the sum of squared distance from the point to the rays:

$$t_j = \underset{t}{\operatorname{argmin}} \sum_{i=1}^{V} (s_i - t)^T \left(I - \vec{v}_{i,j} \vec{v}_{i,j}^T\right)(s_i - t) \tag{6}$$

where V is the number of views.

Equation (6) can be rearranged to define A and b terms as follows:

$$A = \sum_{i=1}^{V} (I - \vec{v}_{i,j}\vec{v}_{i,j}^T) \tag{7a}$$

$$b = \sum_{i=1}^{V} (I - \vec{v}_{i,j}\vec{v}_{i,j}^T)s_i \tag{7b}$$

which form the familiar $Ax = b$ relation:

$$At_j = b \tag{8}$$

The 3D tip is then:

$$t_j = A^+ b \tag{9}$$

where A^+ is a Moore-Penrose pseudo inverse of matrix A. For each t_j, a metric Δ_i is defined as a sum of

distances from the point to the V rays forming the intersection:

$$\Delta_i = \sum_{i=1}^{V} \frac{\left\| \overrightarrow{d_{i,J}} \times (s_i - \vec{t}) \right\|}{\left\| \overrightarrow{d_{i,J}} \right\|} \tag{10}$$

The metric Δ_i was used to establish correspondence between multiple detections on multiple views as

described in §2.2.3.

2.2.3 Identification of Corresponding Keypoints

The correspondence of multiple detections from multiple radiographs was solved by choosing the

permutation of intersections of rays from different views that minimizes the sum of total distances to the

rays. As outlined in the pseudocode in Table 1, all rays from both views are taken in a combinatorial

fashion to compute 3D ray intersections in the least-squares sense, i.e., minimizing the sum of squared

distances. A nominal threshold of $\Delta_i = 10$ mm was enforced on the sum of distances to filter gross outliers.

The predicted 3D points (i.e., intersections and their associated rays) are identified as the set of points

that use distinct rays (i.e., no single ray is used for multiple intersections) and that results in the smallest total sum of distances.

$$
\begin{aligned}
&\text{for each } k_{1,j}, k_{2,l} \text{ in } K_1 \times K_2 \\
&\quad t_i \leftarrow \text{intersect } v_{1,j}, v_{2,l} \\
&\quad d_i \leftarrow \text{dist } t_i, v_{1,j} + \text{dist } t_i, v_{2,l} \\
&\quad T \leftarrow \{t_i : \Delta_i < 10 \text{ mm}\}. \\
&\quad \{t_i\} \leftarrow \underset{\Sigma^n \Delta_i}{\arg\min} \binom{T}{n}.
\end{aligned}
$$

Table 1. Algorithm pseudocode for establishing 3D point correspondences.

The proposed approach is susceptible to false correspondences when using two views if multiple detections lie on the epipolar plane of the C-arm. The impact of this effect is demonstrated with respect to projection distance error (PDE) as defined in the studies below. In clinical practice, the potential risk of this effect can be easily identified following keypoint detection and mitigated by instructing the user to acquire a third view.
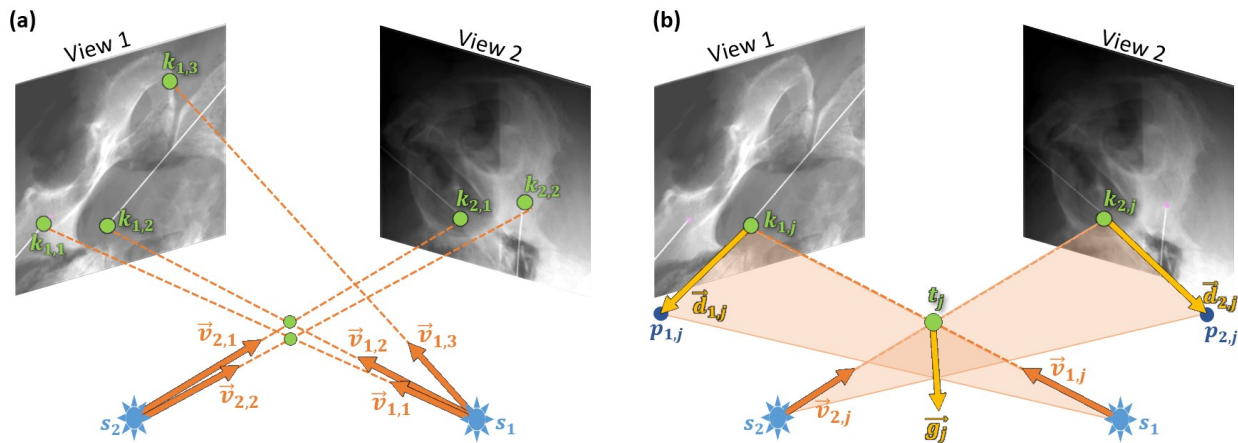
Figure 3. 3D correspondence, tip localization, and direction computation. (a) An example case of correspondence problem with two instrument detections per view. The intersections formed by $(\vec{v}_{1,1}, \vec{v}_{2,2})$ and $(\vec{v}_{1,2}, \vec{v}_{2,1})$ have the smallest distance sum Δ and are therefore chosen as the corresponding tips. (b) Example case of 3D direction calculation, where $k_{1,j}$ and $k_{2,j}$ are correspondent tips. Figure adapted from Bataeva et al. with permission from the publisher.

2.2.4 3D Direction

Direction vectors ($\vec{d}_{i,j}$) identified on the fluoroscopy image are backprojected to obtain the 3D orientation of the instrument. The 3D direction of the instrument tip is found by the intersection of the planes formed by backprojection of the directions on the detector plane. For each detection j on view i, a 3D plane is defined using 2 vectors: the direction vector $\vec{d}_{i,j}$, a 3D vector of the detected K-wire direction on the radiograph in imaging system coordinates, and the ray $\vec{v}_{i,j}$. The normal to this plane can be expressed as $\vec{v}_{i,j} \times \vec{d}_{i,j}$. The cross product of the normal of these planes using Equation (11) defines the gradient $\vec{g}_j$ of the line formed at the intersection of planes as illustrated in Figure 3b.

$$\vec{g}_j = (\vec{v}_{1,j} \times \vec{d}_{1,j}) \times (\vec{v}_{2,j} \times \vec{d}_{2,j}) \tag{11}$$

Note, however, that $\vec{g}_j$ is a non-directional solution for the K-wire direction; it defines the direction of the intersection line and not the actual direction of the K-wire. It therefore has a chance of pointing in the opposing direction. To determine the correct orientation, a 3D point is picked along $\vec{g}_j$ at a nominal distance from $\vec{t}_j$ and is projected onto the 2 radiographs to obtain $p_{i,j}$. We then confirm that point $p_{i,j}$ lies in the same direction from $k_{1,j}$ as the original $\vec{v}_{i,j}$ vectors. If not, $\vec{g}_j$ is multiplied by -1 to reverse its direction. The resulting $\vec{g}_j$ is the 3D direction of the K-wire.

2.3 Experimental Methods

2.3.1 Evaluation of K-Wire Detection

Detection was evaluated on fluoroscopic images of a cadaver (24 images) and patients (17 images) from an IRB-approved clinical study. Images were selected to contain only views with one end of the K-wire showing. All test images exhibited a variety of pelvic views and variation in size, shape, and number (1-3) of K-wires. For each image, the 2D tips and directions of the K-wires were manually segmented as ground truth definition for error analysis. The process for determining the direction involved segmenting the K-

wire close to or including the tip (~1 cm length), medial line extraction, and a linear fit to the extracted skeleton.

A correct detection was defined based on the intersection over the union (IOU):

$$\text{IOU} = \frac{(A \cap B)}{(A \cup B)} \tag{12}$$

of the predicted (A) and true (B) bounding boxes, such that IOU > 50%. For 2 or more predictions, the box with the higher IOU was selected. Recall and precision were defined as the ratio of correct detections to all instruments and as the ratio of correct detections to all detected instruments, respectively:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{13}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{14}$$

where TP is a true positive, FN is a false negative, and FP is a false positive detection. The tip detection was evaluated based on the Euclidean distance between the predicted and ground truth coordinates of the tips from the true positive detections only. The direction error was calculated by measuring an absolute angle difference between the ground truth and that calculated from segmentation tip direction.

2.3.2 Evaluation of 3D Localization in Simulation Studies

A simulation study was conducted to estimate the expected errors in 3D tip localization, where both 2D points and 2D direction vectors were perturbed by the median detection accuracy observed in the clinical and cadaver datasets. Random 3D points were created to represent the instrument tips and reprojected according to the calibrated projection matrices, making sure they were contained within the detector extent. The projected tips were randomly shuffled prior to establishing correspondence. The accuracy of correspondence was defined in terms of true positive rate (TPR) and was analyzed with respect to the geometric accuracy in identifying the true tip location (viz., PDE).

By simulating the PDE expected from the tip detection by Mask R-CNN, we analyzed the distribution of Δ_i values that are expected for correspondent ray intersections. This distribution can therefore be used to define a threshold for Δ_i above which the correspondence between rays would not be assigned. This threshold thus allows filtering out false positive detections and increasing the speed of the algorithm by decreasing the number of combinations that must be analyzed.

As noted above, the correspondence method is susceptible to errors from detections along the same epipolar planes. We therefore used the simulation study to quantify the relationship and to create a threshold ϵ – the distance between keypoints on a detector along the C-arm rotation plane. If all detections fall above this threshold, then correspondence between points is established with a high degree of confidence. If any detections fall below the threshold ϵ, an additional third fluoroscopic view of the scene can be acquired to decrease the chance of false correspondence. Finally, we examined the effects of addition of a third view on the accuracy of correspondence and if it can be used in conjunction with an ϵ threshold.

2.3.3 Evaluation of System Performance: Single K-wire

The full pipeline for obtaining 3D K-wire tip localization from a pair of 2D fluoroscopy images was evaluated on 2D projection images acquired in CBCT scans. The images were first cropped by 8 pixels (collimator edges), log-transformed, resized to 960 × 960 pixels, and input to Mask R-CNN. The outputs were then processed as described in previous sections to obtain a 3D tip pose of the K-wire. The CBCT reconstruction was used to define the 3D ground truth for the scene using 3D Slicer.[32] The true 3D tip location was defined by marking a 3D point closest to the tip of the K-wire as observed on the resliced image. The ground truth for 3D direction was obtained by matching a cylinder of 1 cm length to the tip of the K-wire and calculating a unit vector along the cylinder axis.

18

The Euclidean distance between the ground truth and predicted tip was defined as the tip error δ_x:

$$\delta_x = \|\hat{t} - t\| \tag{15}$$

where $\hat{t}$ is the estimated 3D tip position and t is the ground truth tip position.

The absolute angle between 3D predicted and ground truth direction was calculated as the direction error (δ_ϕ):

$$\delta_\phi = cos^{-1}\frac{\hat{v} \cdot v}{\|\hat{v}\|\|v\|} \tag{16}$$

where $\hat{v}$ is the estimated direction vector and v is the ground truth direction vector.

Four CBCT scans (two pairs) with a single K-wire from a cadaver study were used for quantitative analysis of the algorithm performance. Each pair of CBCT scans showed the same anatomical background with the difference between scans being in the progression of the K-wire. All 4 scans were acquired from the same cadaver with the same K-wire used in each case. Each scan contained 400 projections with 976 × 976 pixels with 0.304 × 0.304 mm² pixel size. System performance was evaluated on 200 pairs images with 60° view separation randomly chosen from the scans. On each image, keypoints and directions were detected, and 3D localization pipeline was executed. Then, the difference between predictions and ground truth was calculated.

2.3.4 Evaluation of System Performance: Multiple K-wires

A second set of studies assessed the end-to-end performance on images containing multiple K-wires by utilizing two CBCT scans. Scan #1 was obtained from a cadaver pelvis specimen containing 3 guidewires and 2 pedicle screws. The K-wires were not inserted into the pelvic bone but instead either inserted into the soft tissue or were on the surface of the cadaver. Scan #2 originated from the same IRB-approved clinical study as the fluoroscopy images used for evaluation of K-wire detection. The patient scan contained 2 K-wires and 1 pelvic screw connected to a K-wire. Both scans contained 200 projections dimensioned 976 × 976 pixels with 0.304 × 0.304 mm² pixel size. The ground truth for K-wire localization

19

and the method for processing the CBCT projections for the input into Mask R-CNN was defined using the

same procedure described in section §2.3.3. For full pipeline evaluation, 2 random projections separated

by at least 45° were selected from the scans, and the detection, correspondence, and calculation of

instrument tips and directions were performed, and the 3D tip and direction errors were calculated.

Chapter 3: System Performance Evaluation

3.1 K-wire Detection with Mask R-CNN

3.1.1 Model Training

The loss curves for network training computed from the training and validation datasets are shown in Figure 4. The training was performed for 150 epochs during which the loss curves exhibit the largest descent until 20th epoch – after which the training loss continued to decrease, while the validation loss is roughly unchanged. Accordingly, the model from the 75th epoch was chosen.
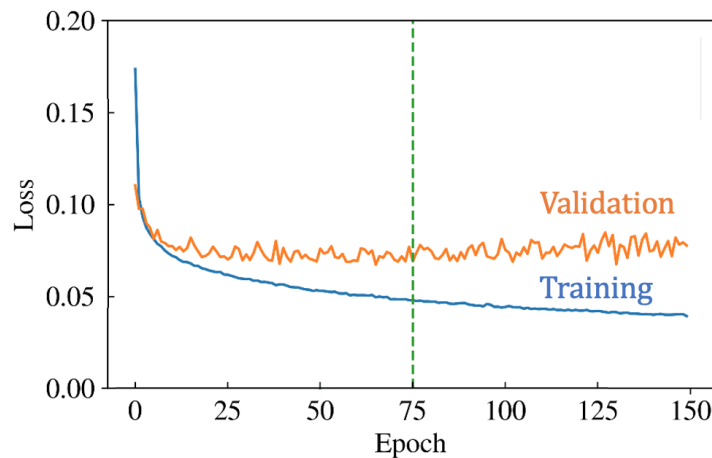
Figure 4. Loss curves for training and validation datasets. The model from the 75th epoch was chosen to avoid overfitting, as observed by the lack of change in validation loss.

3.1.2 K-Wire Detection Accuracy

The accuracy of guidewire detection is summarized in Table 2. Examples of detections in validation, cadaver, and clinical images are shown in Figure 5a-b, illustrating the predicted tip locations and their directions. Detection score and IOU thresholds used to classify the prediction as true positive were selected as 75% and 50%, respectively, based on recall and precision trends observed in the validation

dataset. The mean error in tip location for combined cadaver and clinical datasets was 1.8 ± 1.6 mm, and the mean error in 2D direction was 1.9° ± 1.3°. The average runtime on a GeForce GTX TITAN (Nvidia, Santa Clara CA, USA) was 0.24 s per image.

Dataset	Recall	Precision	Tip Error	Direction Error
Validation	97%	98%	1.4 ± 1.4 mm	2.4° ± 3.2°
Cadaver	95%	90%	1.7 ± 1.2 mm	1.5° ± 1.2°
Clinical	76%	100%	2.5 ± 2.1 mm	2.5° ± 1.4°

Table 2. Summary of detection accuracy for various datasets with a score threshold of 25% and IOU threshold of 50%.
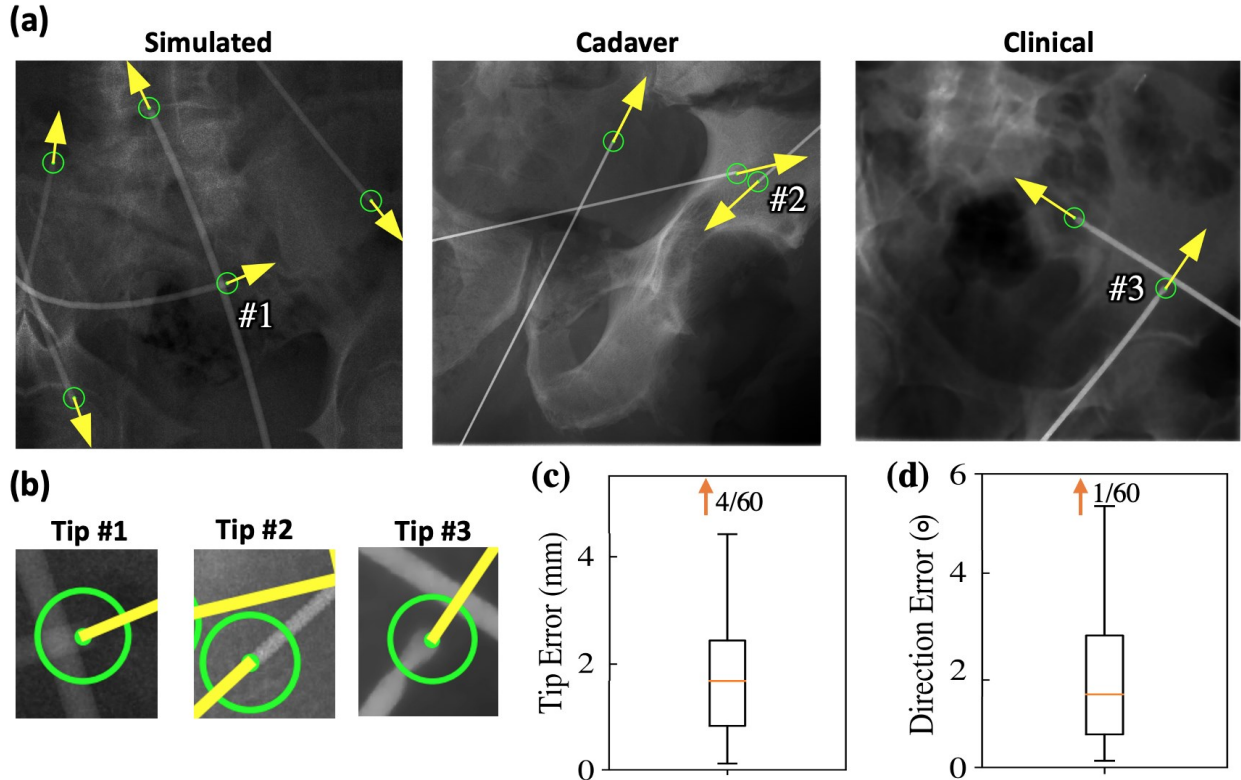
Figure 5. Instrument tip and direction. (a) Visualization of predicted guidewire tip locations and directions on sample fluoroscopic images from simulated, cadaver, and clinical datasets. (b) Magnified tips #1, #2, and #3 illustrate the typical accuracy of predictions from the 3 datasets. (c) Tip detection achieved 1.7 mm median error in combined cadaver and clinical images containing a total of 60 K-wires. (d) Direction detection achieved 1.5° median error over the same combined dataset. Figure adapted from Bataeva et al. with permission from the publisher.

3.2 3D Localization Evaluation in Simulation Studies

3.2.1 Accuracy of 3D Tip Location and Direction

The relationship between 2D detection errors and 3D localization errors analyzed in 1000 simulation studies is shown in Figure 6. As observed in Figure 6a, the relationship between the PDE and median 3D tip error was observed to be linear, with PDE = 1 mm resulting in 0.6 mm 3D tip error, which is related to the magnification factor of the imaging system. Thus, for the 1.6 mm median PDE obtained in fluoroscopic images, a 3D tip error of 1.0 mm is expected with an interquartile range (IQR) 0.85-1.07 mm. Figure 6b demonstrates a linear relationship of the 2D direction error and the 3D direction error with a slope of 3D vs 2D being 1.3. The median 2D direction error of 1.5° from the cadaver and clinical dataset is therefore expected to result in 2.0° 3D direction error with IQR 1.6°-2.4°.
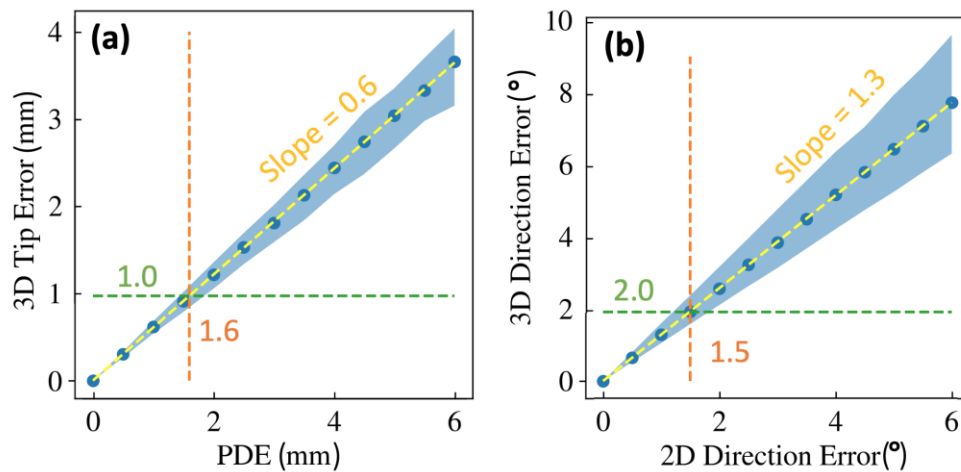
Figure 6. Simulation study describing the effect of 2D errors on K-wire 3D localization accuracy. (a) The 3D tip error has a linear relationship with PDE proportional to the magnification factor of the imaging system. (b) 3D Direction errors also follow a linear relationship with the 2D direction error with a slope equal to 1.3.

3.2.2 Accuracy of Correspondence

The distribution of Δ_i for intersections formed with PDE of 5.8 mm is shown in Figure 7a. The median Δ_i was 2.22 mm, and the maximum was 6.90 mm (obtained from 10000 simulations). A conservative threshold of Δ_i < 10 mm was chosen, as it would encompass the correct K-wire tip detections we expect to obtain with the trained Mask R-CNN on the real data. Simulation studies of correspondence accuracy showed a positive correlation between the fiducial localization errors from the detection stage and false positives in correspondence between the instrument tips. Most errors in correspondence occurred when the projected tips appeared in the rotation plane of the C-arm. Figure 7b illustrates that controlling for the distance ϵ between the projected points on the plane increased the true positive rate (TPR) of correspondence. Based on Figure 7b, we chose ϵ < 10 mm threshold for acquiring a third view, as this threshold was shown to produce >99% correspondence accuracy with two views. For such cases, introducing a third view (as suggested in §2.2.3) resolved most false positives (TPR increased from 96.4% to 99.6%) in the expected range of PDE values (<3 mm), as shown in Figure 7c.
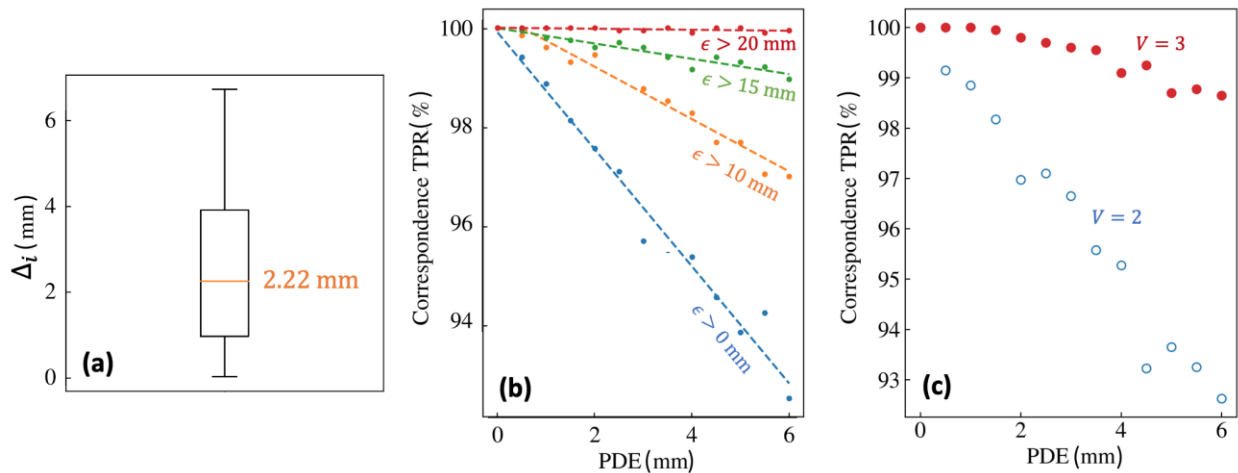
Figure 7. Simulation studies establishing thresholds for the correspondence method. (a) Distribution of the Δ_i for the simulated 5.8 mm (95th percentile) PDE observed on cadaver and clinical dataset. (b) Correspondence accuracy vs tip PDE for a range of ϵ distance constraints applied. The larger the distance between points on the detector along the rotation plane, the higher the correspondence accuracy. (c) Correspondence accuracy vs tip PDE for $V = 2$ (0°, 90°) and $V = 3$ (0°, 45°, 90°) views used to establish correspondence.

3.3 Evaluation in Laboratory and Clinical Studies

3.3.1 Detection and Localization of a Single K-wire

Figures 8a-b shows the pooled results for K-wire localization from 146 K-wires that were detected from 200 image pairs with single K-wires. The median 3D tip error was 1.7 ± 0.7 mm, and median direction error was 6° ± 2°. Figures 8c-d demonstrate an example of detection with errors close to the median errors of the dataset (1.65 mm 3D tip error and 6° 3D direction error) on a CBCT reconstruction slice along the predicted direction axis.
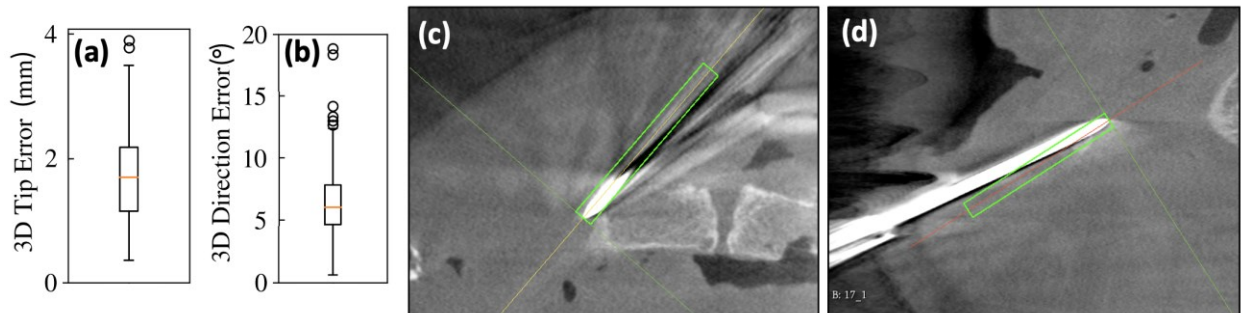
Figure 8. Demonstration of K-wire localization using pairs of views of a cadaver pelvis with a single K-wire. (a-b) Error metrics from 146 image pairs each with 60° angle separation from 4 CBCT scans. (c-d) Visualization of the K-wire localization result at the tip resliced along the predicted axis with accuracy similar to the median observed in these studies: 1.7 mm 3D tip error and 6° 3D direction error.

3.3.2 Detection and Localization of Multiple K-wires

The full K-wire detection and localization pipeline was further evaluated in projections containing multiple K-wires. Figure 9 shows 2 K-wires localized from 2 radiographs (separated by 70°) of a cadaver pelvis. The accuracy of 3D tip locations for both was <1 mm, and the direction errors were 9°- 12°. The postoperative CBCT image was resliced along the predicted instrument orientation to help visualize the tip and direction errors: it can be observed that the tip locations are well matched; however, there are fairly large errors in alignment of the predicted and actual guidewire (9°-12° error), requiring further investigation.
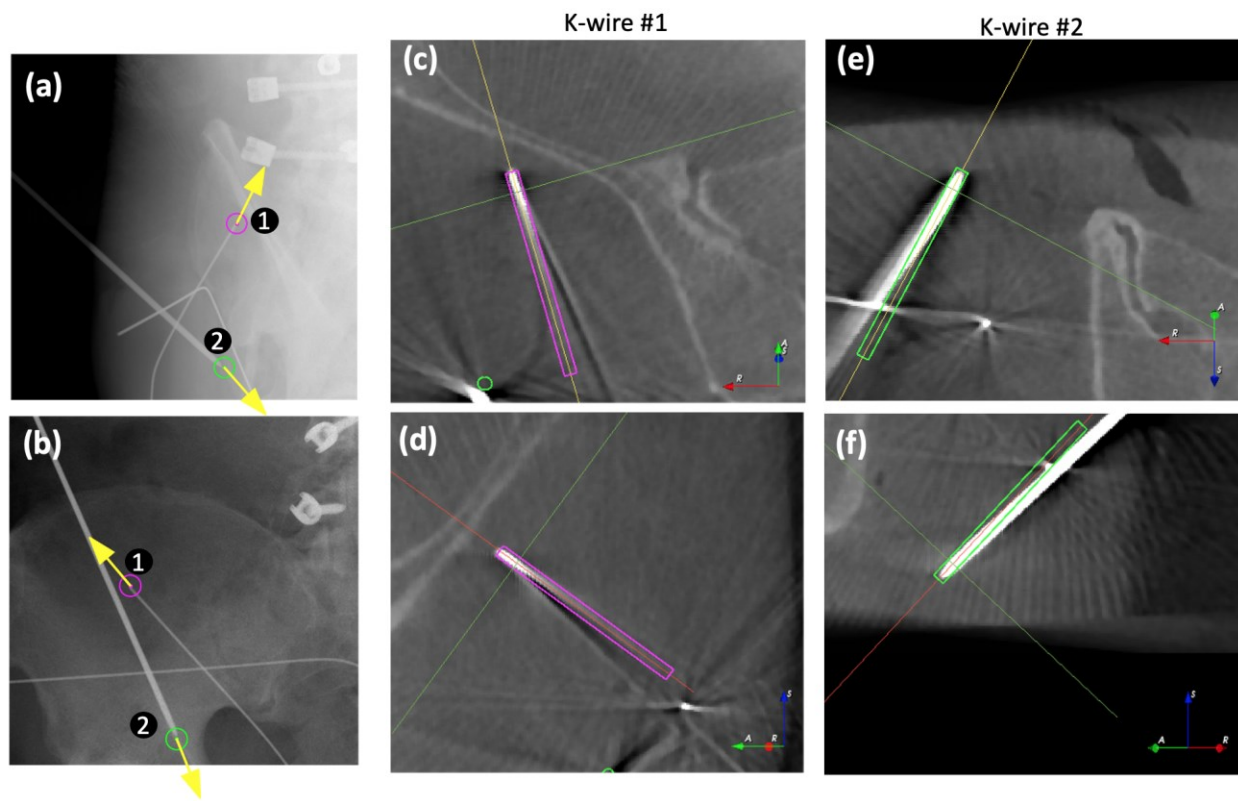
Figure 9. Demonstration of the pipeline using random views of a cadaveric pelvis. (a-b) Detection of 2 K-wire tips and directions in fluoroscopy. (c-f) Visualization of the K-wire directions at the tip resliced along the predicted axis. Figure adapted from Bataeva et al. with permission from the publisher.

Figure 10 similarly illustrates the results of localization of 2 K-wires from 2 radiographs (also separated by 70°) from a CBCT scan acquired in the clinical study. The first K-wire had a 3D tip accuracy of 1.2 mm with a direction error of 3.8°, and the second showed tip accuracy of 1.9 mm and direction error of 12.9°. The average runtime requirement was 1.55 s per pair of images on a GeForce GTX TITAN, which included the time to load the DICOM images from disk. As noted above, a third projection view could reduce the error somewhat, but in cases for which the underlying error is in a failed 2D detection (rather than slight error in keypoint location or K-wire direction), a third view would not be helpful.
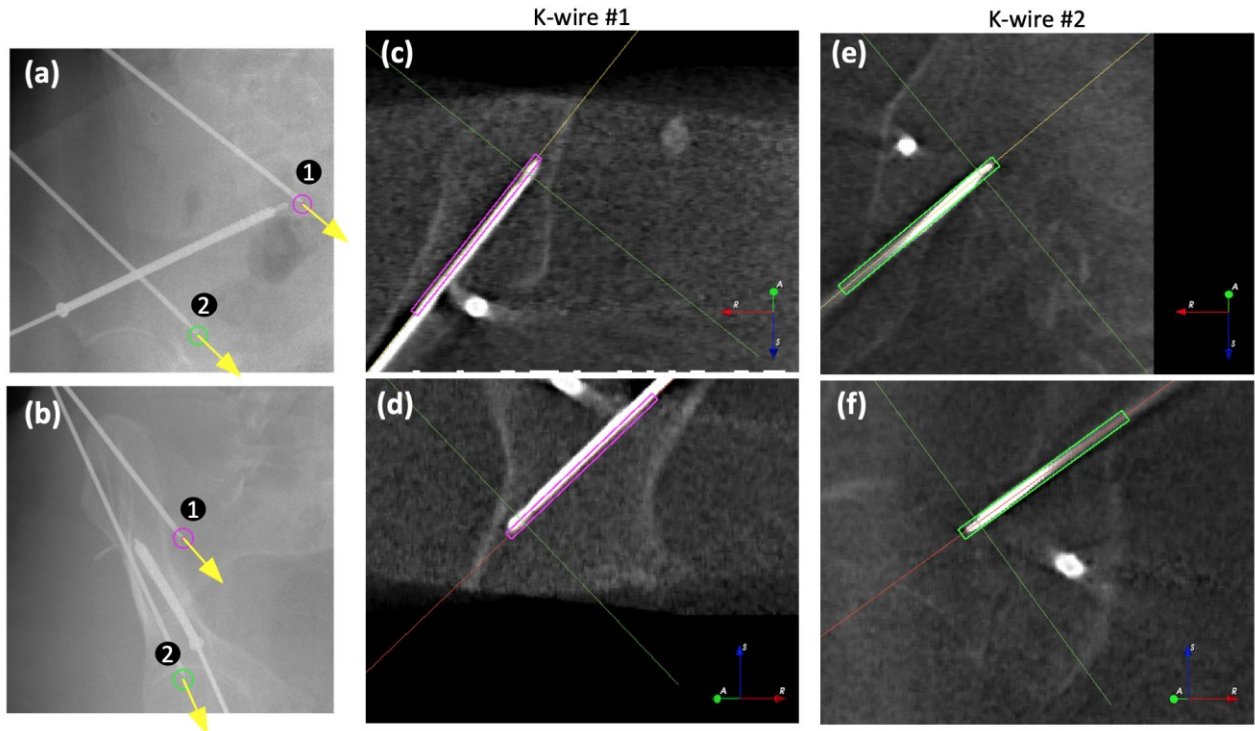
Figure 10. Demonstration of K-wire detection and 3D localization in clinical CBCT (patient undergoing surgery for pelvic fracture fixation). (a-b) Detection of 2 K-wire tips and their directions on 2 fluoroscopic views. (c-f) Visualization of the K-wire directions at the tip resliced along the predicted direction axis. Figures adapted from Bataeva et al. with permission from the publisher.

Chapter 4: Conclusions

4.1 Summary of Findings

The approach presented in this dissertation combines a deep learning method with knowledge of imaging system geometry to provide accurate and fast 3D localization of pelvic K-wires. The feasibility and accuracy of the solution were demonstrated in studies involving cadaver specimens as well as first patient images from a clinical study.

K-wire detection using Mask R-CNN achieved 87% recall and 98% precision in cadaver and clinical radiographs. The median geometric accuracies for detected K-wires in radiographs were 1.7 ± 1.6 mm for instrument tip and 1.5° ± 1.3° for direction. Simulation studies demonstrated TPR = 99% for correspondence for scenes with up to 4 K-wires. Simulation studies that modeled 3D accuracy based on the magnitude of 2D detection errors suggested median 3D tip and direction errors of ~1 mm and ~2°, respectively. The accuracy of K-wire detection could degrade in instances presenting multiple K-wires, due to challenges posed by K-wire intersections (in the projected image), which was observed to interfere with segmentation accuracy (and therefore direction estimation).

The end-to-end performance of detection and correspondence was quantitatively evaluated in CBCT scans of cadaver pelves, each containing a single K-wire. The median errors in 3D localization in this study were higher than indicated from the simulation tests, with 1.7 ± 0.7 mm tip and 6° ± 2° direction error. The extent to which this level of accuracy may be suitable to particular clinical scenarios is discussed below.

The performance of the method was also demonstrated in CBCT scans of a cadaver (in laboratory studies) and a patient (in an IRB-approved clinical study). The complexity of such fluoroscopic scenes under realistic clinical conditions with multiple K-wires and other instruments and implants present on the scans

resulted in higher errors and point to future work involving a great volume and variety of training data that reflect such complex scenes.

4.2 Discussion and Conclusions

While the method performed well in laboratory cadaver studies for both detection and 3D localization, a loss in performance was observed for detections in clinical images – most notably in the estimation of K-wire direction. Within the current work, the method performed at a level of accuracy (with 1.7 ± 0.7 mm tip error and 6° ± 2° direction error) that is within the desired margins of surgical targeting for many pertinent applications – e.g., K-wire placement through relatively large bone corridors (e.g., the shaft of the femur or tibia for long-bone nailing) and/or relatively short trajectories (e.g., the sacroiliac (SI) joint for SI screw placement). However, the accuracy of approach was challenged in several realistic scenarios discussed below (e.g., fluoroscopic scenes presenting numerous complex instrumentation and/or poor image quality), and further research is warranted to see the method toward clinical application.

 One of the main challenges to performance likely stems from the complexity of the fluoroscopic scene (e.g., other instruments, clutter) and variability in acquisition protocols (i.e., images with higher noise than those used in training). Future work will aim to address these factors by further expanding the training dataset with low-dose simulation (i.e., injecting noise)[33] to help the network to generalize to a wider variety of imaging protocols. The susceptibility to misclassify other surgical instruments and implants (e.g., pedicle screws) as K-wires will be addressed by using images with real surgical clutter (as background images) or additional simulation (forward projection) of a variety of types of instrumentation (or even arbitrary shapes, such as letters)[34] in the training dataset to improve the ability to differentiate between K-wires and other radiopaque objects.

Further improvements to K-wire detection could likely be achieved through customization of the Mask R-CNN architecture and training pipeline. The model used for this work was pretrained on a COCO dataset containing photographic images. With x-ray fluoroscopic images, retraining the model completely or fine-tuning the network on images with real K-wires could improve the performance of the detection task. Further optimization of hyperparameters within the current pipeline may also yield performance improvements. For example, a higher learning rate or an aggressive gradient clipping could lead to a better-performing Mask R-CNN model. Furthermore, alternative backbones to ResNet-50 could be tested, and the number of layers in the Mask and Keypoint prediction heads could be modified to improve K-wire detection.

We also observed suboptimal performance in estimating K-wire direction based on segmentations in the projection images. Failure modes suggest possible causes for relatively large errors in direction estimation: incomplete segmentations of the tip; network confusion at K-wire intersections; and segmentation of K-wire holders (e.g., a robotic manipulator). These errors in segmentation affected the B-spline fit to the segmentation and therefore affected the direction of the identified K-wire.

Future work could similarly aim to improve the accuracy of direction prediction. A potentially simple solution within the existing architecture would be to train the network with the medial lines as target segmentations (rather than the full K-wire segmentation). Alternatively, it may be possible to change the architecture of Mask R-CNN to add a head with a direction vector output. Also, changes to the training dataset, such as the inclusion of K-wires with alternative tips (e.g., a threaded drill tip), could yield more accurate segmentation. Simulating instruments used for K-wire insertion (such as a manipulator observed on CBCT scans with single K-wires) could further improve the segmentation accuracy. Modifications to the post-processing of the mask output of the network could also be of benefit, such as adding edge detection or integrating knowledge of the K-wire shape, to improve the accuracy of direction estimation.

Two components of the proposed 3D localization method are dependent on the K-wire tip relative to the imaging system epipolar geometry. The 3D direction calculation method cannot be used if the K-wire lies at the image center and points along the rotation plane, thus making the normal of the backprojected planes of the 2D directions parallel, which would cause their cross product to go to zero. This failure case can be detected and accounted for, however, by acquiring a third view at a different angular offset. Correspondence is also susceptible to the same failure mode when detections lie along the same epipolar plane. As demonstrated in the simulation study, these failures can be predicted directly from the location of the detections, and most of the incorrect predictions are resolved if a third view is acquired. The addition of a third view, however, requires potential disruption to surgical workflow. The feasibility of such workflow should be investigated in future work with close consultation with stakeholders, including surgeons and fluoroscopy technologists. Further improvements to establishing correspondence among K-wire detections could be accomplished with additional information, such as the K-wire direction as defined in a surgical plan. Such knowledge would decrease the search space for the correspondent K-wires by providing new constraints.

This dissertation demonstrates a new method for K-wire detection and localization that is potentially suitable for guidance of instrument placement in orthopaedic trauma surgery. The work demonstrates significant progress toward a robust K-wire detection process that leverages deep learning to achieve fast and fairly accurate localization in these early results. The method could be used to provide fast initialization for the dKC-Reg method describes by Goerres et al.,[21] which was shown to yield a highly accurate K-wire localization (if well initialized).

Further work is required to achieve the desired levels of accuracy for clinical use and to evaluate workflow integration. The work established in this dissertation forms a prototype from which a system more suitable for clinical use could potentially be developed. The method for 3D localization from 2D fluoroscopy images could also be integrated with a navigation system and with robotic systems for K-wire

insertion.[35] Since such systems currently do not account for K-wire deformation, the method shown here could be leveraged to further improve the accuracy of instrument placement in image-guided, robot-assisted surgery.

References

1. Buller LT, Best MJ, Quinnan SM. A Nationwide Analysis of Pelvic Ring Fractures: Incidence and Trends in Treatment, Length of Stay, and Mortality. *Geriatr Orthop Surg Rehabil*. 2016;7(1):9-17. doi:10.1177/2151458515616250

2. Vu CPCL, Runner RP, Reisman WM, Schenker ML. The frail fail: Increased mortality and post-operative complications in orthopaedic trauma patients. *Injury*. 2017;48(11):2443-2450. doi:10.1016/j.injury.2017.08.026

3. Schweitzer D, Zylberberg A, Córdova M, Gonzalez J. Closed reduction and iliosacral percutaneous fixation of unstable pelvic ring fractures. *Injury*. 2008;39(8):869-874. doi:10.1016/j.injury.2008.03.024

4. Peters BS, Armijo PR, Krause C, Choudhury SA, Oleynikov D. Review of emerging surgical robotic technology. *Surg Endosc*. 2018;32(4):1636-1655. doi:10.1007/s00464-018-6079-2

5. Goerres J, Uneri A, Jacobson M, et al. Planning, guidance, and quality assurance of pelvic screw placement using deformable image registration. *Phys Med Biol*. 2017;62(23):9018-9038. doi:10.1088/1361-6560/aa954f

6. Wong JML, Bucknill A. Fractures of the pelvic ring. *Injury*. 2017;48(4):795-802. doi:10.1016/j.injury.2013.11.021

7. Džupa V, Chmelová J, Pavelka T, et al. Multicentric study of patients with pelvic injury: basic analysis of the study group. *Acta Chir Orthop Traumatol Cech*. 2009;76(5):404-409. Accessed August 24, 2020. https://europepmc.org/article/med/19912705

8. Poole G V., Ward EF, Griswold JA, Muakkassa FF, Hsu HSH. Complications of pelvic fractures from blunt trauma. In: *American Surgeon*. Vol 58. ; 1992:225-231. Accessed April 28, 2021.

https://europepmc.org/article/med/1586080

9. Besl PJ, McKay ND. A Method for Registration of 3-D Shapes. *IEEE Trans Pattern Anal Mach Intell*. 1992;14(2):239-256. doi:10.1109/34.121791

10. Rusinkiewicz S, Levoy M. Efficient variants of the ICP algorithm. *Proc Int Conf 3-D Digit Imaging Model 3DIM*. Published online 2001:145-152. doi:10.1109/IM.2001.924423

11. Arun KS, Huang TS, Blostein SD. Least-Squares Fitting of Two 3-D Point Sets. *IEEE Trans Pattern Anal Mach Intell*. 1987;PAMI-9(5):698-700. doi:10.1109/TPAMI.1987.4767965

12. Jud L, Fotouhi J, Andronic O, et al. Applicability of augmented reality in orthopedic surgery - A systematic review. *BMC Musculoskelet Disord*. 2020;21(1):1-13. doi:10.1186/s12891-020-3110-2

13. Wu JR, Wang ML, Liu KC, Hu MH, Lee PY. Real-time advanced spinal surgery via visible patient model and augmented reality system. *Comput Methods Programs Biomed*. 2014;113(3):869-881. doi:10.1016/j.cmpb.2013.12.021

14. Abe Y, Sato S, Kato K, et al. A novel 3D guidance system using augmented reality for percutaneous vertebroplasty. *J Neurosurg Spine*. 2013;19(4):492-501. doi:10.3171/2013.7.SPINE12917

15. Han R, Uneri A, Ketcha M, et al. Multi-body 3D-2D registration for image-guided reduction of pelvic dislocation in orthopaedic trauma surgery. *Phys Med Biol*. 2020;65(13):135009. doi:10.1088/1361-6560/ab843c

16. Han R, Uneri A, De Silva T, et al. Atlas-based automatic planning and 3D-2D fluoroscopic guidance in pelvic trauma surgery. *Phys Med Biol*. 2019;64(9):17. doi:10.1088/1361-6560/ab1456

17. Markelj P, Tomaževič D, Likar B, Pernuš F. A review of 3D/2D registration methods for image-guided interventions. *Med Image Anal*. 2012;16(3):642-661. doi:10.1016/j.media.2010.03.005

18. Che C, Mathai TS, Galeotti J. Ultrasound registration: A review. *Methods*. 2017;115:128-143. doi:10.1016/j.ymeth.2016.12.006

19. Uneri A, De Silva T, Stayman JW, et al. Known-component 3D-2D registration for quality assurance of spine surgery pedicle screw placement. *Phys Med Biol*. 2015;60(20):8007-8024. doi:10.1088/0031-9155/60/20/8007

20. Uneri A, Otake Y, Wang AS, et al. 3D-2D registration for surgical guidance: Effect of projection view angles on registration accuracy. *Phys Med Biol*. 2014;59(2):271-287. doi:10.1088/0031-9155/59/2/271

21. Goerres J, Uneri A, Jacobson M, et al. Planning, guidance, and quality assurance of pelvic screw placement using deformable image registration. *Phys Med Biol*. 2017;62(23):9018-9038. doi:10.1088/1361-6560/aa954f

22. Haskins G, Kruger U, Yan P. Deep learning in medical image registration: a survey. *Mach Vis Appl*. 2020;31(1):1-18. doi:10.1007/s00138-020-01060-x

23. Wagner MG, Laeseke P, Speidel MA. Deep learning based guidewire segmentation in x-ray images. In: Bosmans H, Chen G-H, Gilat Schmidt T, eds. *Medical Imaging 2019: Physics of Medical Imaging*. Vol 10948. SPIE; 2019:150. doi:10.1117/12.2512820

24. Gherardini M, Mazomenos E, Menciassi A, Stoyanov D. Catheter segmentation in X-ray fluoroscopy using synthetic data and transfer learning with light U-nets. *Comput Methods Programs Biomed*. 2020;192:105420. doi:10.1016/j.cmpb.2020.105420

25. Esfandiari H, Newell R, Anglin C, Street J, Hodgson AJ. A deep learning framework for segmentation and pose estimation of pedicle screw implants based on C-arm fluoroscopy. *Int J Comput Assist Radiol Surg*. 2018;13:1269-1282. doi:10.1007/s11548-018-1776-9

26. Kügler D, Sehring J, Stefanov A, et al. i3PosNet: instrument pose estimation from X-ray in temporal bone surgery. *Int J Comput Assist Radiol Surg*. 2020;15:1137-1145. doi:10.1007/s11548-020-02157-4

27. He K, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. *IEEE Trans Pattern Anal Mach Intell*. 2020;42(2):386-397. doi:10.1109/TPAMI.2018.2844175

28. Jader G, Fontineli J, Ruiz M, Abdalla K, Pithon M, Oliveira L. Deep Instance Segmentation of Teeth in Panoramic X-Ray Images. In: *Proceedings - 31st Conference on Graphics, Patterns and Images, SIBGRAPI 2018*. Institute of Electrical and Electronics Engineers Inc.; 2019:400-407. doi:10.1109/SIBGRAPI.2018.00058

29. Fox M, Taschwer M, Schoeffmann K. Pixel-based tool segmentation in cataract surgery videos with mask R-CNN. In: *Proceedings - IEEE Symposium on Computer-Based Medical Systems*. Vol 2020-July. Institute of Electrical and Electronics Engineers Inc.; 2020:565-568. doi:10.1109/CBMS49503.2020.00112

30. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol 2016-Decem. IEEE Computer Society; 2016:770-778. doi:10.1109/CVPR.2016.90

31. Lin TY, Maire M, Belongie S, et al. Microsoft COCO: Common objects in context. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 2014;8693 LNCS(PART 5):740-755. doi:10.1007/978-3-319-10602-1_48

32. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging*. 2012;30(9):1323-1341. doi:10.1016/j.mri.2012.05.001

33.   Wang AS, Stayman JW, Otake Y, et al. Low-dose preview for patient-specific, task-specific technique selection in cone-beam CT. *Med Phys*. 2014;41(7):071915. doi:10.1118/1.4884039

34.   Kordon F, Maier A, Swartman B, Kunze H. Font augmentation: Implant and surgical tool simulation for x-ray image processing. In: *Informatik Aktuell*. Springer; 2020:176-182. doi:10.1007/978-3-658-29267-6_36

35.   Vijayan RC, Han R, Wu P, et al. Fluoroscopic guidance of a surgical robot: pre-clinical evaluation in pelvic guidewire placement. In: Linte CA, Siewerdsen JH, eds. *Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling*. Vol 11598. SPIE-Intl Soc Optical Eng; 2021:49. doi:10.1117/12.2582188