K-wire Tracking in 3D Camera Views

Jie Ying Wu,Athira Jane Jacob
Mentors: Mathias Unberath, Javad Fatouhi, Sing Chun Lee, Bernhard Fuerst

Abstract

K-wires are widely used instruments in orthopedic surgeries. Currently, K-wire insertion is a long, tedious procedure that requires multiple X-rays and mental alignment of patient, wire and X-rays by the doctor. We propose a deep learning based solution to track K-wire in RGB stereo images, which can be then used to detect and track the K-wire in 3D space. The goal is to estimate the orientation of the K-wire in 3D. Due to the shortage of real surgical scene data, we create our own artificial data for training by composing foreground (K-wire) and background separately. We then explore the performance of two different networks for K-wire segmentation, U-Net and HED. Finally, we validate the performance of the two different networks in 2D and 3D space with naturally acquired images and achieve $< 1^\circ$ and $< 5^\circ$ average error in each respectively.

1. Introduction

Kirshner wires or K-wires are long, smooth stainless steel pins that are widely used in orthopedics surgery [1] to fixate bones. The pins are driven through the bone using a power or hand drill. They can be used for temporary fixation before inserting screws or permanent fixation while the bones heal. Fig. 1 shows a surgeon placing a K-wire during surgery.

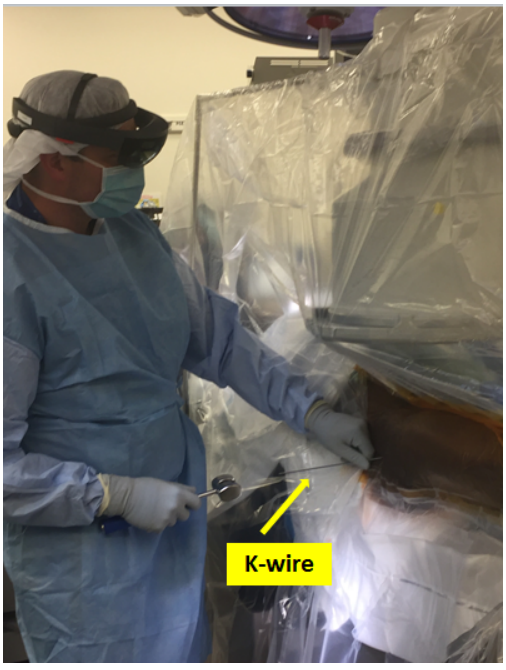

Figure 1: Inserting a K-wire during surgery. Arrow points to the K-wire.

K-wire and screw insertion is currently done with minimally invasive techniques [2], involving modern imaging technology and computer aided navigation systems. Correct placement requires numerous intra-operative X-ray images, and often requires multiple attempts before the surgeon achieves satisfactory placement and orientation [3]. A

sample X-ray is shown in Fig. 2. Misplacement of the K-wire could cause severe damage to nearby structures, for eg. the external iliac artery and vein & obturator nerve in pelvic surgery [4]. This leads to multiple entry wounds on the patient, high X-ray exposure for the patient and the surgical staff, increased OR time and frustration of the surgical team. A single K-wire insertion could as much as ten minutes [5].

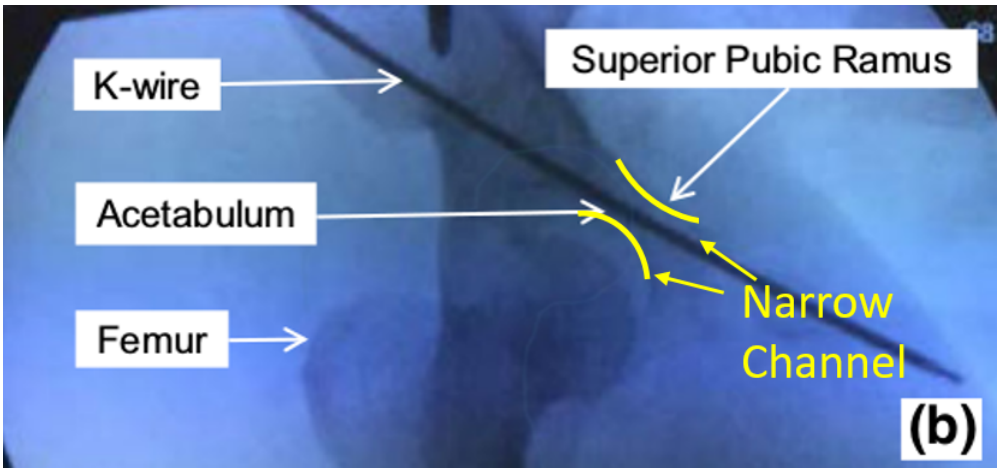

Figure 2: K-wire placement during a hip surgery in X-ray view. Notice the narrow channel it must pass through. Image from [6]

.

The main challenge during K-wire insertion has been identified as the mental alignment of patient, medical instruments, and intra-operative X-rays [7]. Recently, camera augmented solutions have been proposed to help surgeons in this mental alignment [8][9]. Multi-modal fusion between 3D surface from RGBD cameras and digitally reconstructed radiographs have been shown to considerably reduce the duration of surgeries, the number of X-rays, overall radiation dose, and the surgical workload [6].

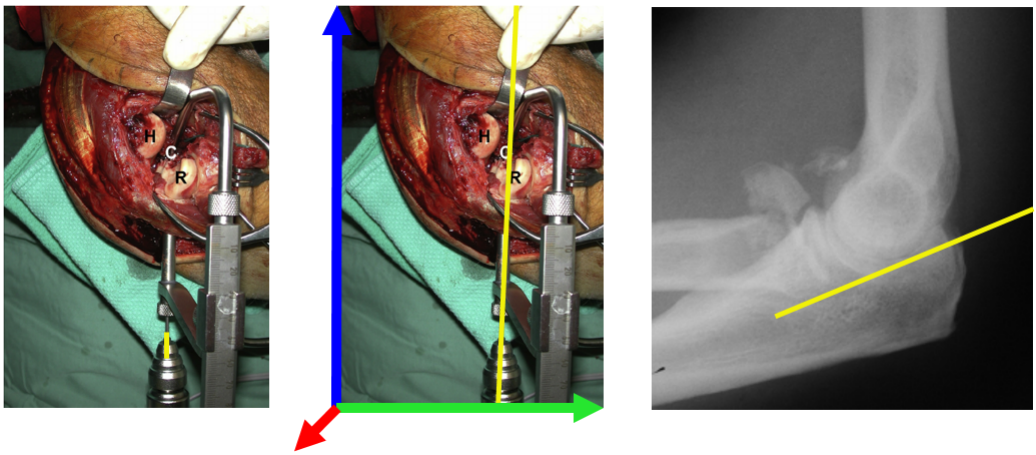

Figure 3: Potential pipeline to aid physicians in K-wire placement. The K-wire is first segmented in RGB image. Its orientation is extracted from stereo image pairs and the path is projected onto the intr-operative image. Images from [10].

Any computer assisted solution to assist surgeons in the mental alignment and localization, including augmented reality based solutions, will eventually require tracking of the K-wire. Conventional navigation systems for tool tracking are mainly based on tracking of optical markers and recovering the spatial transformation between the patient, medical images and the tool [11] [12]. Though such navigation systems offer submillimeter accuracy [13], they cannot be extended to K-wires due to their particular geometry. In addition, the K-wire can slide in and out of the drill, preventing accurate calibration. K-wires are also too thin to be tracked by depth camera and too reflective to be segmented from RGB using traditional computer vision techniques. Thus, we will explore deep learning tools to learn

structural features and K-wires in RGB images. As we use images from stereo-cameras, the segmentations are used to estimate the 3D orientation of the K-wire.

2. Methods

In this project we explore two deep learning architectures to segment the K-wire in 2D RGB images. A major challenge with such an approach is the lack of surgical scene data to train the architectures. Moreover, such data cannot be created easily due to the need for accurate and detailed annotations. Large, quality data is essential to training a successful neural network. Hence, we propose an innovative data creation technique to compose images that helps us to overcome this problem. The overall project is divided into three parts:

1. Data Creation: We capture the foreground (K-wire) and background (scene including drape, instruments etc.) separately and compose them in stages to generate data of varying complexity. After capturing the foreground separately on a plain background, we segment the K-wire and compose it onto the backgrounds. We define three levels of difficulty for the background, which are illustrated in Fig. 4:
   Level 0: K-wire on plain blue background
   Level 1: K-wire on blue background with gloved hands
   Level 2: K-wire on multiple drapes with hands and other instruments.

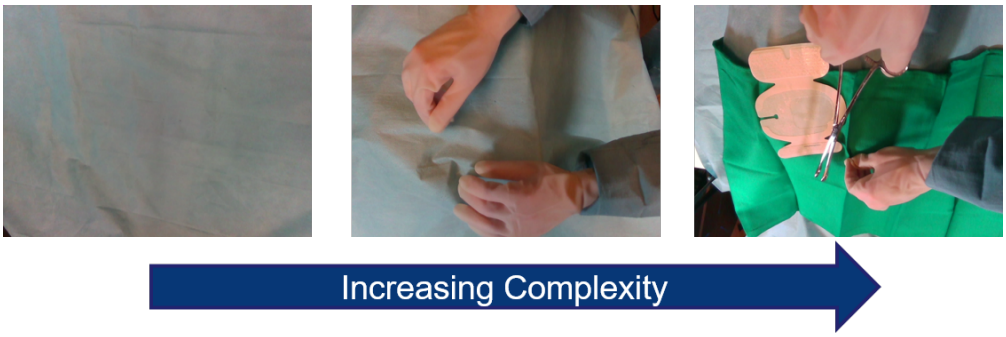

Figure 4: Sample background images from each of the three difficulty levels.

The images are also under varying lighting conditions to simulate real life conditions. Fig.5 shows sample foreground images captured in various conditions.

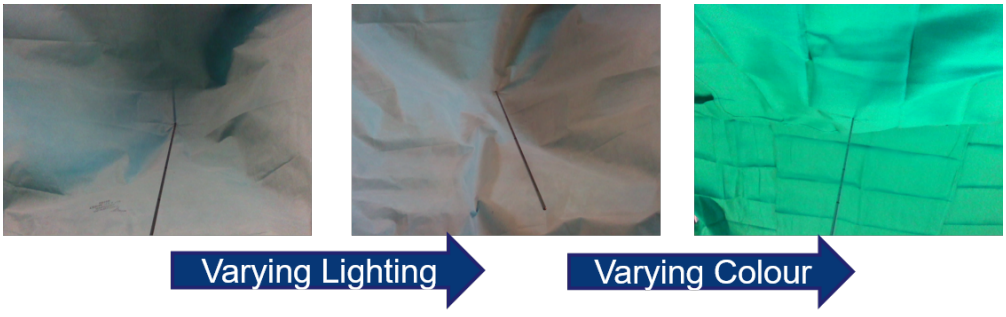

Figure 5: Sample foreground images with different conditionds.

The figures are then composed combinatorially to create a large dataset 6. We use Gaussian blurring outwards from the mask to smooth the transition and histogram matching to correct colour and lighting differences.

2. Network architecture: We explore two different kinds of network architecture, U-Net and HED. We train the network on the simplest level of difficulty (Level 0) and incrementally train it on more complex data.
3. Validation: Validation is done in two stages:
   2D validation: We compare orientations of the K-wire in the 2D images. We fit a line to the mask of the K-wire to estimate its angle and that is then compared with the angle derived from ground truth.

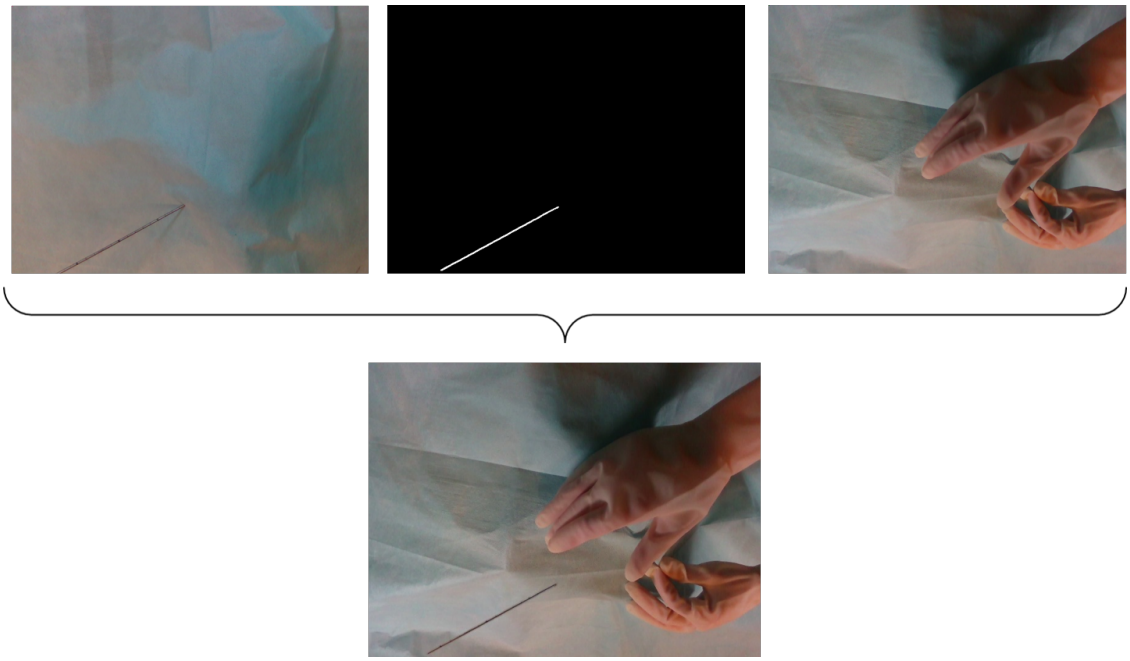

Figure 6: Sample composition of images. Top from left to right: The foreground image, the corresponding mask, and the background image. The bottom image shows the composition results.

3D validation: We collect validation data using AR toolkit[14] marker attached to the K-wire. This gives us ground truth for the pose of the K-wire in 3D. The predictions made by the network are translated into 3D space using calibration parameters of the stereo-camera. All validation is done on natural images, and not composed images.

2.1. Network Architecture

Deep learning has shown remarkable successes in the recent past [6], mainly due to deeper networks, larger datasets and better optimization techniques. Segmentation is a typical task that deep learning algorithms generally excel at. Traditionally, pixel wise segmentation is done through a patch based approach, where the image is divided into many patches, each surrounding a pixel. This method, however, is slow and inefficient, as forward passes are performed for each patch and typically a single image will have thousands of over-lapping patches. In addition, there is a tradeoff between context and localization accuracy. Fully convolutional networks (FCN) offer to solve some of these problems [7]. In such networks, the fully connected layers are replaced by convolutional layers, hence retaining spatial context. In addition, this allows end-to-end training, with any input size. The contracting path is supplemented by an up-sampling path, that up-samples the images to the size required. This gives an efficient, fast way of training a segmentation network.

2.1.1. U-Net

U-Net [15] is a modification of the FCN that has given state of the art results in the domain of biomedical image segmentation. This task faces some challenges similar to our task and thus makes it a potential candidate for solving our problem of segmenting the K-wire.

The usual contracting network is supplemented by successive up-sampling layers. Hence, these layers increase the resolution of the output. In order to improve localization, high resolution features from the contracting path are combined with the upsampled output via skip-ahead connections. The upsampling part has a large number of feature channels, which allows the network to propagate context information to higher resolution layers. As a consequence, the expansive path is more or less symmetric to the contracting path, yielding a U-shaped architecture. The network does not have any fully connected layers and only uses the valid part of each convolution, without padding.

We use extensive data augmentation of both the foreground and the background images separately to compensate for the relatively small size of the data. In addition we also use weight balanced cross entropy loss to account for

the fact that, the K-wire constitutes only around one percent of the total number of pixels. The cross entropy loss function is computed using a pixel-wise softmax over the final feature map. The soft-max is defined as $p_k(x) = \exp(a_k(x))/(\sum_{k=1}^{K} \exp(a_k(x))$ where $a_k(x)$ denotes the activation in feature channel k at the pixel position $x \in \Omega$ with $\Omega \subset \mathbb{Z}^2$. K is the number of classes and $p_k(x)$ is the approximated maximum-function, i.e. $p_k(x) \approx 1$ for the k that has the maximum activation $a_k(x)$ and $p_k(x) \approx 0$ for all other k. The cross entropy then penalizes, at each position, the deviation of $p_{l(x)}(x)$ from 1 using,

$$E = \sum_{x \in \Omega} w(x) \log(p_{l(x)}(x)) \tag{1}$$

where $l : \Omega \rightarrow \{1, \ldots, K\}$ is the true label of each pixel and $w : \Omega \rightarrow \mathbb{R}$ is the introduced weight map. Adam optimizer with momentum is used to optimize the loss function.

2.1.2. Holistically-nested Edge Detector (HED)

Holistically-nested edge detection (HED) [16], performs image-to-image prediction that leverages FCN networks and deeply-supervised nets. It was originally developed for edge detection. HED introduces the original images at various levels so that details such as boundaries are preserved throughout the net. Thus, boundaries can be detected at various scales. This feature is important for our project since we aim to detect a very thin structure.

At each of its 5 scales, the network produces an output and that layer is optimized independently based on the loss on that output. Since edge pixels are generally scarce in an image, HED uses a naturally balancing loss function

$$\mathcal{L}_{side}^{(m)}(\boldsymbol{W}, \boldsymbol{w}^{(m)}) = -\beta \Sigma_{j \in Y_+} \log Pr(y_j = 1 | X; \boldsymbol{W}, \boldsymbol{w}^{(x)}) - (1 - \beta) \Sigma j \in Y_- \log Pr(y_j = 0 | X; \boldsymbol{W}, \boldsymbol{w}^{(m)}) \tag{2}$$

where $\beta = |Y_-|/|Y|$ and $1 - \beta = |Y_+|/|Y|$. $|Y|$ is the sample size while $|Y_+|$ and $|Y_-|$ denote the edge and non-edge label set sizes respectively. The probability is given by a sigmoid function with the activation value at pixel j. All side-outputs are fused as follows to create an overall prediction map.

$$L(\boldsymbol{W}, \boldsymbol{w}, h) = Dist(Y, \hat{Y}_{fuse}) \tag{3}$$

The overall objective function is then:

$$(\boldsymbol{W}, \boldsymbol{w}, h)* = (\mathcal{L}_{side}(\boldsymbol{W}, \boldsymbol{w}) + \mathcal{L}_{fuse}(\boldsymbol{W}, \boldsymbol{w}, h)) \tag{4}$$

This ensures that edges are detected at all levels. Since K-wires are similar in appearance to edges, we start with pre-trained weights for the network. We then refine it using our training images.

2.2. Line Extraction and Pose Validation

After obtaining the foreground and background probabilities from the networks, we threshold the foreground probability to create a binary mask. Then, we identify the strongest and longest line from the mask using Hough transform. In 2D, the slope of the line is compared with that of the line extracted from the ground truth mask and the mean absolute error in the angle between them is calculated; however, the 3D errors are much more relevant to how well our technique can be used in the operating room. To evaluate in 3D, we identify corresponding points on the lines in stereo image pairs from the ground truth mask, and the segmentation results from each of the networks. We use triangulation, based on epipolar geometry [17], to reconstruct the points in 3D space in camera frame. Using two points, we identify the vector connecting them and normalize the vector to represent the orientation of the line. Then we take the inner product between the vectors from each of the networks and the ground truth and report the angle between them as the quality of the orientation estimation. Though this does not give an estimate of errors in 3D position, it gives us a convenient way of comparing orientations.

3. Results

Fig. 7 shows qualitative segmentation results from all three levels, from both networks, with a line fitted to each mask. As mentioned earlier, the validation is performed on natural images.

Both U-Net and HED perform well with Level 0 and Level 1 images, qualitatively. Level 2 images show some difficulties to both the networks, due to the presence of instruments like the scissors (Fig. 7d). The K-wire segmentation is considered to be successful if both the θ and ρ parameters of the Hough line are within a threshold (5 degrees

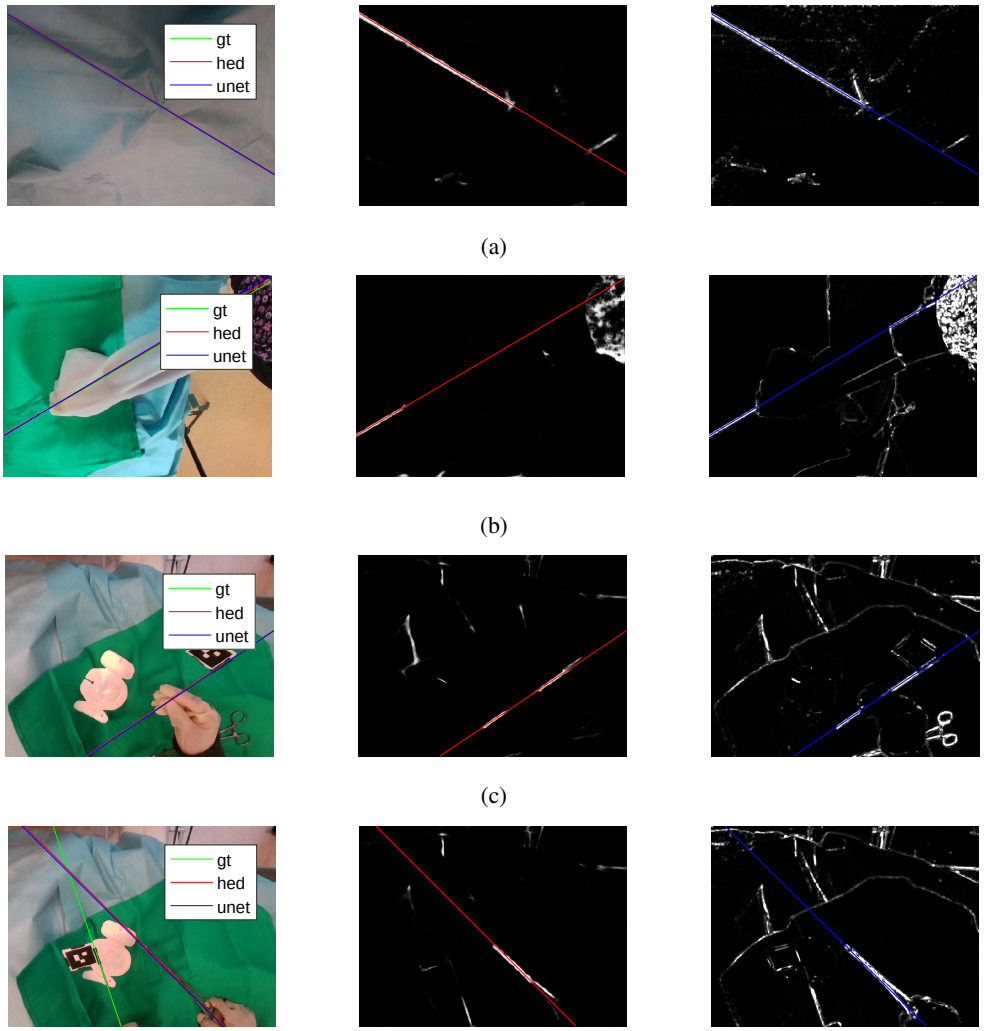

Figure 7: Left: Original images, centre: HED output, right: U-Net output. These figures show sample line fittings from Level 0 (a), Level 1 (b) and Level 2 (c & d) images. (a)-(c) are successful in detecting the K-wire but both networks detect the scissors instead of the K-wire in (d).

and 20 pixels, for θ and ρ respectively) of the ground truth. When detected successfully, the K-wire is segmented with less than 1° deviation from the ground truth in 2D. Table. 1 shows percentage of detection as well as average angle deviation in 2D.

Table 1: Validation Results (2D).

Difficulty Level	No. of Images	No. of Correct Detections		Error (deg)			
		HED	U-Net	HED		U-Net	
				Mean	Variance	Mean	Variance
0	10	10	10	0.33	0.25	0.40	0.27
1	10	10	10	0.55	1.00	0.50	0.37
2	20	18	13	0.83	1.32	0.77	1.19

The same images are reconstructed in 3D for level 2 images to evaluate how the error in 2D affects accuracy in 3D. Detections are considered correct only if the K-wire was identified successfully in both images of a stereo pair. Both the HED and the U-Net generally fall below 5°; however, errors in 2D are greatly magnified in 3D. There is one case,

when the predictions throw an error greater than 15° after reconstruction.

Table 2: Validation Results (3D). Mean and variance are calculated only over successful detections.

HED	Success	1	1	1	1	1	0	1	0	1	1	Mean	Variance
	Error	1.5	3.2	20.1	0.6	1.0	17.0	1.1	80.5	4.2	0.2	4.0	44.5
U-Net	Success	1	0	0	1	0	0	1	1	0	0	Mean	Variance
	Error	16.4	34.2	65.9	1.2	15.5	27.1	2	4.6	29.7	27.2	6.0	49.9

4. Discussion

Although our results are not yet accurate enough for clinical use, they show that artificially composed images can be used to train deep networks that generalize to natural images. This could greatly augment datasets where samples are scarce. Our networks performed well in cases with simple backgrounds, even in the presence of occlusions.

We observe that the errors are generally low when the K-wire is successfully detected; however, any error in 2D is magnified in 3D so that the maximum error even for a successful detection can be prohibitively high. More post-processing may help in rejecting these cases as well as false positives where another tool is detected. Reprojection error may be helpful to check whether the 3D line maps back to the 2D segments.

5. Future Work

A major area of future work could be to improve the accuracy of K-wire detection in the presence of other tools. We observed that the network often detects the K-wire in one stereo image, but not in the other. Often, this is because of high responses from other instruments, such as scissors. We believe the detection accuracy can be improved by searching for the K-wire in the corresponding region, as defined by epipolar geometry, of the other image, when it is detected only in one. This may lower false positives when the network erroneously detects structures but requires a confidence score for how sure it is that it has identified the K-wire.

Another goal is to continue working towards our original maximum deliverable by reading the lines on the K-wire and using cross ratios to estimate the tip position. Lastly, temporal information may be incorporated to maintain estimation consistency across frames. This could greatly aid differentiation between K-wire and other surgical instruments.

6. Management Summary

We met a few times per week to make sure that our goals were aligned. After accomplishing each task, we would discuss what the next steps were and split the work depending on each teammate's experience and interest. For example, Athira implemented image augmentation while Jie Ying worked on composing the foreground and background. After creating the dataset, we each explored a different network to compare their performance. We had a shared folder on Thin6 to share datasets and a Github repository to share tools as needed. Weekly meetings with our mentors gave us timely feedback and helped us overcome challenges when we were unsure how to proceed. For data analysis, Jie Ying wrote scripts to extract lines while Athira worked on getting the angles from lines. We worked together to analyze the data.

6.1. Planned vs Accomplished

We accomplished up to parts of our maximum deliverable. We were able to extract the orientation of the K-wire in 3D with 85% success rate and performed validation in 3D using triangulation in MATLAB. Although we had originally planned to use AR Toolkit, we switched to MATLAB after difficulties with finding the correct libraries to build it with visualization tools. We were unable to finish locating the tool tip in 3D but plan to continue to work on this project next year.

One addition to our deliverable list was to explore the efficacy of using artificial datasets to train the network to extract the K-wire in 3D and to publish it if it proved effective. The dataset we created seems to give reasonably good performance, even on naturally acquired images, as well as artificially composed images. We showed that artificially created datasets is a viable approach in cases where data is scarce and plan on working with our mentor to publish our dataset and the tools to create it in the coming semester.

6.2. Learning Outcomes

Before this course, neither of us realized the difficulties of tracking thin, reflective instruments. We had some exposure to deep learning but this course allowed us to further explore what it can do as opposed to traditional computer vision techniques. Specifically, we gained software experience in TensorFlow, knowledge of fully convolutional neural networks, and experimented with the effects of hyper-parameter tuning.

Additionally, we saw the importance of various techniques learned in other classes such as stereo-camera calibration, transformations between camera spaces, tool space, and AR toolkit marker space, and epipolar geometry. Lastly, the scope of this project made us more aware of the importance of time and resource management, and how to coordinate with different people to get mentoring and access to resources.

7. Acknowledgement

We would like to thank all of the CAMP group, especially Mathias for taking over mid-project as our main mentor, Sing Chun for technical support in data collection, Daniil in helping with various aspects of deep learning and using Thin6, and Javad for explaining 3D geometry in vision and presentation feedback. We would also like to thank Dr. Bernhard Fuerst for initially setting up the project, and Dr. Alex Johnson and Dr. Greg Osgood for kindly arranging for us to observe surgeries and providing us with clinical pictures.

8. Technical Appendix

All our code have been uploaded to the CAMP LCSR git repository.

9. References

[1] Dr Matt A. Morgan et. al. K wire.

[2] AJ Starr, AL Jones, CM Reinert, and DS Borer. Preliminary results and complications following limited open reduction and percutaneous screw fixation of displaced fractures of the actabulum. *Injury*, 32:45–50, 2001.

[3] Ulrich Stckle, Klaus Schaser, and Benjamin Knig. Image guidance in pelvic and acetabular surgery–expectations, success and limitations. *Injury*, 38(4):450–462, April 2007.

[4] Pierre Guy, Mohammad Al-Otaibi, Edward J. Harvey, and Nader Helmy. The 'safe zone' for extra-articular screw placement during intra-pelvic acetabular surgery. *Journal of Orthopaedic Trauma*, 24(5):279–283, May 2010.

[5] A. J. Starr, C. M. Reinert, and A. L. Jones. Percutaneous fixation of the columns of the acetabulum: a new technique. *Journal of Orthopaedic Trauma*, 12(1):51–58, January 1998.

[6] Marius Fischer, Bernhard Fuerst, Sing Chun Lee, Javad Fotouhi, Severine Habert, Simon Weidert, Ekkehard Euler, Greg Osgood, and Nassir Navab. Preclinical usability study of multiple augmented reality concepts for K-wire placement. *International Journal of Computer Assisted Radiology and Surgery*, 11(6):1007–1014, June 2016.

[7] A. J. Starr, A. L. Jones, C. M. Reinert, and D. S. Borer. Preliminary results and complications following limited open reduction and percutaneous screw fixation of displaced fractures of the acetabulum. *Injury*, 32 Suppl 1:SA45–50, May 2001.

[8] N. Navab, S. M. Heining, and J. Traub. Camera Augmented Mobile C-Arm (CAMC): Calibration, Accuracy Study, and Clinical Applications. *IEEE Transactions on Medical Imaging*, 29(7):1412–1423, July 2010.

[9] S. Habert, J. Gardiazabal, P. Fallavollita, and N. Navab. RGBDX: First Design and Experimental Validation of a Mirror-Based RGBD X-ray Imaging System. In *2015 IEEE International Symposium on Mixed and Augmented Reality*, pages 13–18, September 2015.

[10] David Kovacevic, Laura A. Vogel, and William N. Levine. Complex Elbow Instability: Radial Head and Coronoid. *Hand Clinics*, 31(4):547–556, November 2015.

[11] Florian Gras, Ivan Marintschev, Kajetan Klos, Thomas Mückley, Gunther O Hofmann, and David M Kahler. Screw placement for acetabular fractures: which navigation modality (2-dimensional vs. 3-dimensional) should be used? an experimental study. *Journal of orthopaedic trauma*, 26(8):466–473, 2012.

[12] Florian T Gebhard, Michael D Kraus, Eugen Schneider, Ulrich C Liener, Lothar Kinzl, and Markus Arand. Does computer-assisted spine surgery reduce intraoperative radiation doses? *Spine*, 31(17):2024–2027, 2006.

[13] Li Liu, Timo Ecker, Steffen Schumann, Klaus Siebenrock, Lutz Nolte, and Guoyan Zheng. Computer Assisted Planning and Navigation of Periacetabular Osteotomy with Range of Motion Optimization. In *Medical Image Computing and Computer-Assisted Intervention MICCAI 2014*, pages 643–650. Springer, Cham, September 2014.

[14] Open Source Augmented Reality SDK | ARToolKit.org.

[15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention MICCAI 2015*, pages 234–241. Springer, Cham, October 2015.

[16] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1395–1403, 2015.

[17] Zhengyou Zhang. Determining the epipolar geometry and its uncertainty: A review. *International journal of computer vision*, 27(2):161–195, 1998.

Questionnaire – Project # 03, Athira Jane Jacob and JieYing Wu

10/10 Overall project and progress

- Were you satisfied with the overall technical progress made in the course of the semester?
  Yes
- Was the total accomplishment appropriate for the number and level (undergrad/graduate) of students on the project?
  Yes
- Will the results be useful to you in the future?
  Yes
- Do you see a prospect for patents or publication to result?
  We are considering to invest some more work to refine the results. After this, the results may translate into a publication

Training of neural nets is not always straight-forward. Difficulties in training the U-Net resulted in small delays that slightly set back the progress, but was a good learning experience for future work on deep learning.

9.5/10 Report (which the students should have shared with you)

- Does the project report accurately reflect the scope and accomplishment of the project?
  Yes
- Were you given an adequate opportunity to review the report?
  Yes
- Does the report and its appendices, together with the web site, provide sufficient information that subsequent groups can make effective use of the project results.
  Yes
- In particular, are any project designs or code adequately documented.
  The Code I have seen so far, yes. Particularly given that some development was done in Matlab, many functions have good documentation.

10/10 Web site

- Does the web site reflect the scope and accomplishment of the project?
  Yes. At the time I checked, the final report and poster were not yet uploaded but they assured me to do it on time.
- Do you wish the web site to remain password protected after May 30? If so, for how long?
  Removing password protection of the website should be fine.

10/10 Management

- Were the students fully engaged in the project?

Absolutely, they worked long-hours and re-did experiments to achieve acceptable quality

- How often did they meet with you? Was this enough?
  We met regularly once a week and irregularly based on need (informal discussions and emails). It was enough.
- Were the “deliverables” and “dependencies” realistic?
  In general yes. As mentioned earlier, we faced problems in getting the U-Net to learn meaningful features, a problem that, unfortunately, cannot be accelerated. We came up with a solution (pre-trained networks) so this worked well and Athira and Jie Ying did a good job in the transition.
- Was the plan realistic? Were unmet dependencies approached in an effective manner?
  As mentioned above, yes.

Other comments or suggestions

- Do you have any other comments or suggestions, either about the specific project or about the overall structure of the course for next year.

  a) It would be great to spend more time on the subject of the project to refine the results.

  b) Having to hand in this sheet on the same day that the reports are due is probably not optimal, considering that the students (if about to hit the maximum deliverable) are working hard on both the experiments as well as the report in the last days. This may lead to the supervisors reading unfinished versions of the report for the grading.