Tool Tracking in Orthopedics Surgery – Group 3
600.646 Computer Integrated Surgery II, Spring 2017

Seminar Paper Critical Review

Student: Jie Ying Wu

Background and goals

K-wires are long, smooth stainless steel pins that are widely used in orthopedics surgery to hold bones together. The pins are driven through the skin using a power or hand drill. Correct placement requires numerous intra-operative X-ray images, and often requires multiple attempts [1]. Misplacement of the K-wire could cause severe damage major structures [2]. This leads to multiple entry wounds on the patient and high X-ray exposure for the patient and clinicians. Recently, camera augmented solutions have been proposed to help surgeons with mental alignment of the patient, the X-ray scan, and the tool [5,6].

We propose a convolutional neural network (CNN) based tracking of the K-wire as the first step in an augmented reality guidance solution for the clinicians. Conventional computer vision solutions fails to track the K-wire because they are colour-based and the K-wire is too reflective. Optical trackers have been shown to be effective for similar tasks, but as the K-wire slides in and out of the drill, they may not track it well. Our hypothesis is that deep learning will learn more structural features and perform better.

Paper one - B. Diotte, N. Navab *et al.*, "Radiation-Free Drill Guidance in Interlocking of Intramedullary Nails," in *MICCAI*, 2012

Paper selection and back ground

"Radiation-Free Drill Guidance in Interlocking of Intramedullary Nails" [1] examines a similar problem where the physician must align a drill to a hole in a previously inserted nail. Although the marker-based tracker is not applicable to K-wires, I selected this paper since I will need to use similar methods to estimate the tip position after tracking instrument.

Intramedullary nails are often used to secure a fracture tibia. In this procedure, the surgeon first inserts in the nail along the length of the bone and then secure it with one or more screws perpendicular to the bone, as shown in Figure 1. Since careful alignment is required to place the screw correctly, this procedure averages 48 X-rays [2] and 13.7 min [3] to complete.

Figure 1: Illustration of nail insertion. Left image shows an intramedullary nail inserted into a dry bone phantom [1]. Right images show X-rays of the nail secured by screws inside a bone [2].

Technical approach

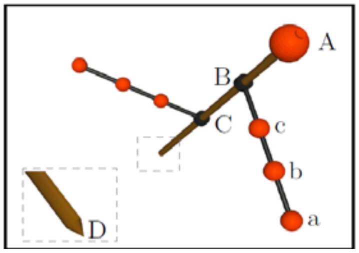
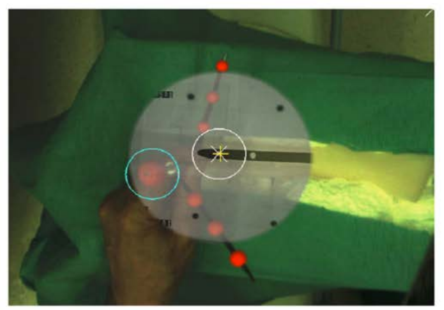


Figure 2: Proposed optical marker

Diotte *et al.* propose to attach an optical marker on the drill to eliminate the need for X-rays in steps 3-5. The marker is shown in Figure 2. The balls of the marker are different sizes and painted fluorescent so they can be easily distinguished by a camera mounted on the C-arm. OpenCV is used to segment the balls from RGB images. Least squares line fitting and least size variance are used to calculate the location of points B and C. Position of D, the tool tip, can then be estimated using cross ratios. The geometric relationship is as follows:

$$d_{AD} = \frac{(S \times d_{AB}) - d_{AC}}{S-1} \quad \text{where } S = \frac{crossratio \times d_{AC}}{d_{AB}} \quad \text{and } crossratio = \frac{AB \times CD}{AC \times BD}$$

Positions of B and C are similarly estimated from the branches. After solving for the position of the drill tip, the placement can be augmented onto the surgical area as shown in Figure 3. Once the blue and white circles, and the yellow and white X overlap, the drill should be in good alignment with the screw hole.



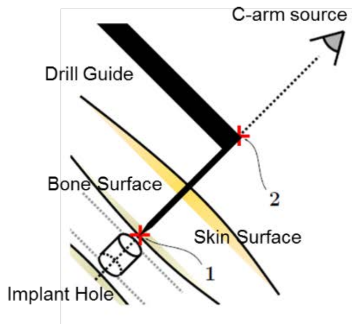


Figure 3: (Left) the augmented reality view proposed. The X-ray image is projected onto the phantom surface. The white circle with the X shows the screw hole, the blue circle the A of the optical marker, and the yellow X the estimated drill tip position. (Right) the desired alignment between the C-arm source, the drill guide, and the screw hole, achieved when the white circle aligns with the blue, and the white X with the yellow. [1]

Results

First, the authors validate their method in an artificial setup. They print a 5 mm circle and fix the drill tip to the centre. Using their algorithm, they predict the location of the drill tip in 200 trials sampling from 30° cone angle rotations. The mean error they observe is 1.72 ± 0.7 mm, with 57% falling below the mean, and 98% under 4 mm error. They identify that as the clinically relevant target since the screw hole is 5 mm.

Next they perform a phantom study by fixing a dry bone to a box and inserting an intramedullary nail through it. Then they have 3 surgeons, two experts and one resident, place nails with only optical guidance. Only two X-rays were taken throughout the procedure, a pre-drill one to find the screw hole pose, and one to confirm screw placement. In 93% of the cases, 56 out of 60, the surgeons were able to successfully place the screw. The average time to completion is 2 minutes. In the remaining 4, one was attributed to poorly designed experimental setup and the other three to the resident's inexperience.

Review

Diotte *et al.* show promising results on optical marker-based guidance for the placing of intramedullary screws. The need for a large marker limits its application to other cases though, such as tracking the K-wire.

They clearly show how they construct the augmented reality view using cross ratios, what it looks like, and the augmentations seems intuitive; however, they could have explained the algorithms better. There was no details on how what registration algorithm was used to co-register X-ray, phantom, and augmented reality space, nor the method to track the balls. These could also contribute to errors. Lastly, they could have given more details on the results, exactly how many cases did the experts do compared to the resident, and how timing varied across the groups. This would help with understanding the impact of their impact.

While the marker is not directly applicable to K-wires, their technique to estimate the tip position using cross ratios is one that we will explore in our project. We will use markings on the K-wire itself rather than an external rigging to calculate the cross ratios. In addition, the paper's experimental setup is a potentially applicable to our project and could be a good way to validate our results. Overall, they show the need and potential for augmented reality in the surgical room.

Paper two - K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," CVPR, 2016 (Microsoft research)

Paper selection and background

The second paper, "Deep Residual Learning for Image Recognition" [4], is more technically focused on convolutional neural networks (CNN). It discusses a recent innovation in deep network structure: skip-ahead layers. This technique reduces training complexity, which is crucial for our project as we will be collecting our own data set. Their results show good performance even on a dataset of just 50k training images. We plan on collecting an order of magnitude more data than that.

CNNs have been shown to learn well but requires a lot of data to train. It has been shown in literature that deeper plain networks perform better up to a point but quickly degrades after. Figure 4 shows the training and testing errors of different sized networks. Unlike traditional learning frameworks, this degradation does not seem to be from overfitting. The training error also worsens as the networks get deeper.

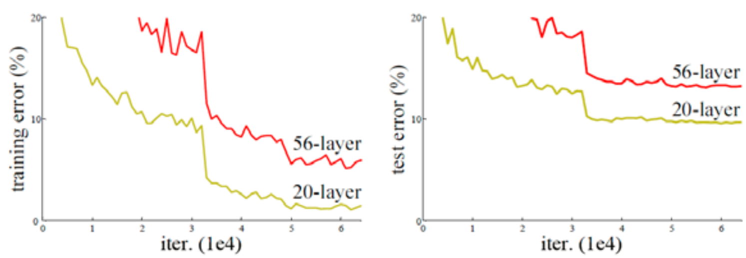


Figure 4: Graphs from [4] showing training (left) and testing (right) errors on plain networks of different depth trained on CIFAR-10 dataset [5].

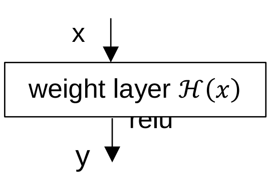


Figure 5: One convolutional layer

Figure 4 shows a general building block in a CNN. The authors were inspired by the fact that they can construct an artificially deeper network by padding a shallow network with identity layers – where the input, x, is equal to the output, y. This artificially deep network performed better than any network of the same length their solvers were able to find. From this result, they observed that it is hard to learn convolutions that produce identity mapping. Instead, they hypothesized that learning the residual of the filter is easier.

Technical approach

Suppose we have the desired mapping, $\mathcal{H}(x)$. Instead of learning it directly, the paper proposes to learn the residual of it, defined as

$$\mathcal{F}(x) \coloneqq \mathcal{H}(x) - x$$

So to reconstruct $\mathcal{H}(x)$, we add the original input back.

$$\mathcal{H}(x) = \mathcal{F}(x) + x$$

This is illustrated by the curved skip-ahead path on the right in Figure 6. In this block, learning the identity mapping is trivial. It is simply

$$\mathcal{F}(x) \to 0$$

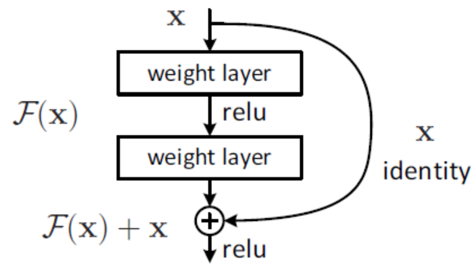


Figure 6: Convolutional layers with a skip-ahead path added. The input is added to the output of the weight layers to reconstruct the desired mapping.

The approach is neat since it does not introduce additional weights. The additional computation of a single addition per skip-ahead path is negligible. It does however, introduce the limitation that the output of $\mathcal{F}(x)$ must be the same size as input. To circumvent this, an projection could be used to rescale x, but this may introduce more weights.

Results

The authors show on CIFAR-10 [5] and ImageNet [6] datasets that the residual network does not lose accuracy when going deeper.

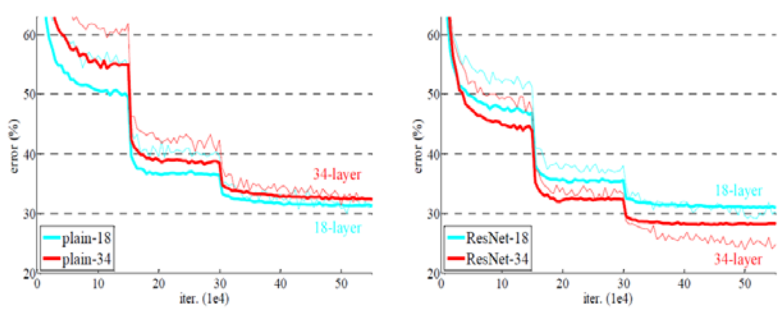


Figure 7: Comparing plain (left) and residual (right) networks of different depths trained and tested on ImageNet[1] . Thin lines show training error and bolded lines show test error. Note the deeper net performs worse in the traditional network, but better in the residual.

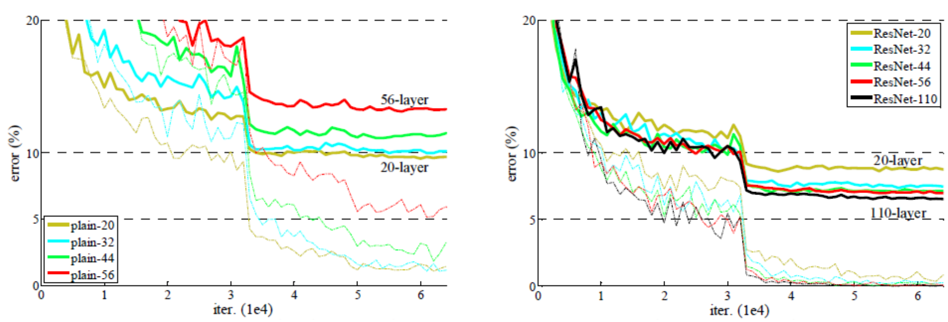


Figure 8: Training (dashed) and test (solid) errors on CIFAR-10 [1] dataset. We see that deeper residuals networks do better, while deeper plain nets do not. The advantage of additional layers diminish quickly though.

Network trained on both datasets support the result that the residual network outperforms traditional networks and can go much deeper; however, there is a limit. The paper also presents results for a 1202 layer network trained on the CIFAR-10 dataset [5]. It does have higher testing error than the 110 layer network, but this appear to be because of overfitting. The training error remains low.

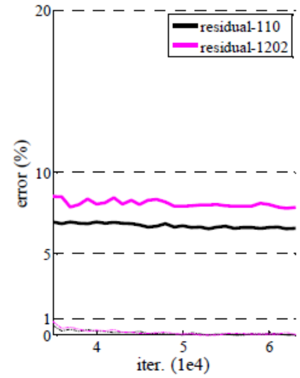


Figure 9: A 1202 layer residual network overfits the data and has higher test error

Review

The skip-ahead paths presented in this paper simplifies training and allows us to train more complex models. Its results are compelling that the residual networks outperform plain ones. The main drawback is that the authors do not discuss why this work and therefore, what sort of guarantees this method provides. Also, they do not give any training time so it is unclear how much training is simplified. Both plain and residual networks seem to converge at a similar rate, although the residual networks find a better solution. Lastly, the paper uses very plain networks for baseline comparison. Other techniques such as drop out and batch normalization have been proposed to simplify training and it would have been interesting to see the residual networks compared with those, and whether multiple technique can be used together.

Although the authors here look at the problem of object classification, the skip-ahead paths can be generalized and applied to fully convolutional networks (FCNs) that can provide pixel-level labelling. Pixel-level labelling will be necessary for segmenting the K-wire from images. This paper shows that the scale of the dataset we would need to learn a good network is within our data collection plan.

References

[1] B. Diotte *et al.*, “Radiation-Free Drill Guidance in Interlocking of Intramedullary Nails,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, 2012, pp. 18–25.
[2] R. Rohilla, R. Singh, N. K. Magu, A. Devgan, R. Siwach, and S. S. Sangwan, “Simultaneous Use of Cannulated Reamer and Schanz Screw for Closed Intramedullary Femoral Nailing,” *ISRN Surg.*, vol. 2011, 2011.
[3] N. Suhm, P. Messmer, I. Zuna, L. A. Jacob, and P. Regazzoni, “Fluoroscopic guidance versus surgical navigation for distal locking of intramedullary implants,” *Injury*, vol. 35, no. 6, pp. 567–574, Jun. 2004.
[4] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[5] A. Krizhevsky and G. Hinton, “Learning multiple layers of features from tiny images,” 2009.
[6] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.