Interventional 2D/3D Registration with Contextual Pose Update

by

Wenhao Gu

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

Baltimore, Maryland
May, 2019

Abstract

Traditional intensity-based 2D/3D registration requires near-perfect initialization in order for image similarity metrics to yield meaningful gradient updates of X-ray pose. They depend on image appearance rather than content, and therefore, fail in revealing large pose offsets that substantially alter the appearance of the same structure. We complement traditional similarity metrics with a convolutional neural network-based (CNN-based) similarity function that captures large-range pose relations by extracting both local and contextual information, and proposes meaningful X-ray pose updates without the need for accurate initialization. Our CNN accepts a target X-ray image and a digitally reconstructed radiograph at the current pose estimate as input and iteratively outputs pose updates on the Riemannian Manifold. It integrates seamlessly with conventional image-based registration frameworks. Long-range relations are captured primarily by our CNN-based method while short-range offsets can be recovered accurately with an image similarity-based method. On both synthetic and real X-ray images of the pelvis, we demonstrate that the proposed method can successfully recover large rotational and translational offsets, irrespective of initialization.

Acknowledgments

First and foremost, I would like to thank Professor Mathias Unberath for his continuous support and guidance throughout the whole research process. This thesis would not be possible without his professional suggestions and encouragement.

In addition, I also would like to extend my sincere gratitude to my lab, Computer Aided Medical Procedures (CAMP) for providing the necessary hardware required in this thesis and Professor Nassir Navab and other lab mates for giving me constant feedback about my research in weekly lab meeting.

Last but not least, I would like to express my deep gratitude to my family for their financial and emotional support to my master degree program at Johns Hopkins University.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Intra-operative localization of patient anatomy is an integral part of navigation for computer-assisted surgical interventions. Traditional navigation systems use specialized optical or electromagnetic (EM) sensors and fiducial objects to recover the pose of patient anatomy (Yaniv, 2016). These systems often require large and invasive incisions to fixate rigid body objects to a patient's bones (Liu et al., 2014; Troelsen, Elmengaard, and Søballe, 2008). Furthermore, optical sensors are sensitive to occlusion, EM sensors are unreliable in the presence of metallic surgical tools, and both are not standard equipment in most operating rooms. Fluoroscopic imaging provides an alternative method for navigation. It is already widely used during surgery and is not sensitive to the limitations of optical and EM trackers.

A 2D/3D registration between the intra-operative 2D C-arm X-ray imaging system and a 3D CT volume may be used to perform navigation (Markelj et al., 2012). Example orthopedic applications involving hip surgery include control of a hip-replacement robot (Yao et al., 2000), guided cement injection into the femoral head (Otake et al., 2012), and localization of osteotomy bone

fragments (Grupp et al., 2019). Other applications include reduction of traumatic bone fractures (Gong, Stewart, and Abolmaesumi, 2011), spine surgery (Ketcha et al., 2017), and kinematic analysis of the wrist (Chen et al., 2013). The two main variants of 2D/3D registration are split between intensity-based and feature-based approaches. Feature-based approaches require manual or automated segmentation or feature extraction in both of the imaging modalities and optimized in point-to-point, curve-to-curve or surface to curve fashion. While feature extraction significantly reduce the amount of data making this method fast, its accuracy directly relies on the accuracy of feature extraction or segmentation. Intensity-based approaches directly use the information contained in pixels of 2D images and voxels of 3D volumes. The most commonly used method in literatures is iteratively optimizing the similarity measure of the simulated intra-operative X-ray images, digitally reconstructed radiographs (DRRs), with the real X-ray image. The optimization problem solved by registering a single object from a single view is described by (1.1).

$$\min_{\theta \in SE(3)} \mathcal{S}\left(I, \mathcal{P}\left(\theta; V\right)\right) + \mathcal{R}\left(\theta\right) \tag{1.1}$$

I defines the 2D fluoroscopic image, V the preoperative 3D model, θ the pose of the volume with respect to the projective coordinate frame, $\mathcal{P}$ the projection operator used to create DRRs, $\mathcal{S}$ the similarity metric used to compare DRRs and I, and $\mathcal{R}$ is a regularization over plausible poses. To be robust against the presence of surgical tools, the similarity may also be computed over local neighborhood patches and combined (Markelj et al., 2012). At the cost of increased computational complexity, state-of-the-art evolutionary search

strategies are adept at avoiding local minima of (1.1). Additionally, GPUs allow many objective functions to be evaluated simultaneously and thus, be used intraoperatively (Otake et al., 2012).

Despite advanced search strategies, a reasonable initial pose estimate is required for any intensity-based registration to find the true pose. A common technique used for initialization is to annotate corresponding anatomical landmarks in the 2D and 3D images and solve the PnP problem (Markelj et al., 2012; Bier et al., 2018). Another technique requires a user to manually adjust an object's pose and visually compare the estimated DRR to the intraoperative 2D image. These methods are time consuming and challenging for inexperienced users, making them impractical during surgery . Alternatively, some restrictions may be imposed on plausible poses to significantly reduce, or eliminate, the number of landmarks required for initialization (Markelj et al., 2012). In (Grupp et al., 2019), a single-landmark was used to initialize the registration of a 2D anterior-posterior (AP) view of the pelvis, and further views were initialized by restricting any additional C-Arm movement to orbital rotations. However, for certain applications, such as the chiseling of bone at near-lateral views, it is not feasible to impose such restrictions on the initial view or C-Arm movements.

We propose a convolutional neural network (CNN) approach that is capable of learning large scale pose updates when far away from ground truth, and finer pose updates when closer to the actual pose. The proposed network regresses a geodesic loss function over $SE(3)$ and was trained on simulated X-ray images from CT using an open-source tool (Unberath et al., 2018). For

large offsets, the network effectively learns the manual pose adjustment process that a human could conduct to initialize an intensity-based optimization. When close to the ground truth pose, updates will be dominated by a classic intensity-based method to make fine adjustments.

Chapter 2

Methods

In this work, we present a large-range pose-estimation approach for estimating the spatial relation between a 2D X-ray image and corresponding 3D CT volume. To retrieve the relative pose, we employ an iterative strategy. In each iteration, a DRR is rendered from CT using the current pose estimate and compared with the input X-ray image with the trained network (Fig. 2.1). The network is trained to predict a relative pose transformation between two input images using an untangled representation of 3D location and 3D orientation. The iterative pose estimation pipeline still requires an initial guess for the starting pose. The AP view of the CT image is chosen, since it represents a view that is commonly used in clinical practice; however, any arbitrary view can, in principle, be used, since the CNN-based similarity metric trained with geodesic-based loss is globally convex (Sec. 2.1).

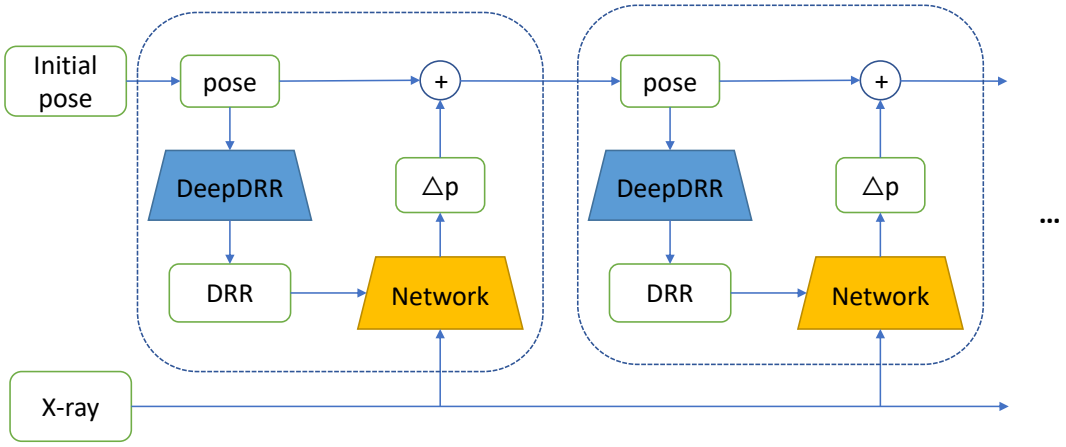

Figure 2.1: Schematic workflow of context-based registration in an iterative scheme.

2.1 Geodesic gradient loss

We would like to express 2D/3D registration in terms of the relative pose between the viewpoints. At the core of the proposed method is the question on how to properly model the similarity between two images. For 2D/3D registration purposes (and assuming that images will always show the same object), it would be appealing if we were able to express 2D image similarity in terms of the relative pose between the respective viewpoints.

A rigid-body pose is an element of $SE(3)$, the Special Euclidean group in 3D, which can be defined as:

$$\{(R,t)|R^TR = I, \det R = 1, t \in \mathbb{R}^3\}$$

where R is the rotation matrix and t represents the translational part of the pose.

This distance between two rigid-body poses $T = \exp(\hat{\xi}) \in SE(3)$ and $T' = \exp(\hat{\xi}') \in SE(3)$ can be defined as the gradient of the geodesic distance

on the Riemannian manifold as (Hou et al., 2018):

$$\nabla \text{dist}(\xi, \xi') = -2 \log_{\xi'}(\xi), \tag{2.1}$$

where $\log_{\xi'}(.)$ denotes the Riemannian Logarithm at ξ'; $\hat{\xi}$ and $\hat{\xi}'$ are the elements of the Lie algebra $se(3)$; $se(3)$ is the tangent space to the Lie group $SE(3)$; $(\xi, \xi') \in \mathbb{R}^6$ are the twist coordinates.

These geodesic gradients indicate the direction of update from one pose estimate to the other, considering the structure of $SE(3)$. It is the generalization of straight lines of Euclidean geometry in Riemannian manifolds. Detailed implementation is shown in (Miolane et al., 2018). While this metric cannot be computed analytically from two images, i. e. a target X-ray image and a DRR, it can be approximated with a CNN trained on a large structured dataset.

2.2 Datasets

To generate 2D simulated fluoroscopy images with ground truth viewpoint label for training, the open source tool, DeepDRR (Unberath et al., 2018), is used for dataset generation. DeepDRR takes into consideration the spectrum of X-ray imaging and uses neural network to simulate scattering effect and perform volume segmentation of different materials. It is shown in (Bier et al., 2018) and (Bier et al., 2019) that DeepDRR is able to generate more realistic simulated fluoroscopy that can generalize well onto real X-ray images in landmark detection.

Each simulated image is labeled with the position and orientation of the X-ray source with respect to the CT volume space that is used to generate

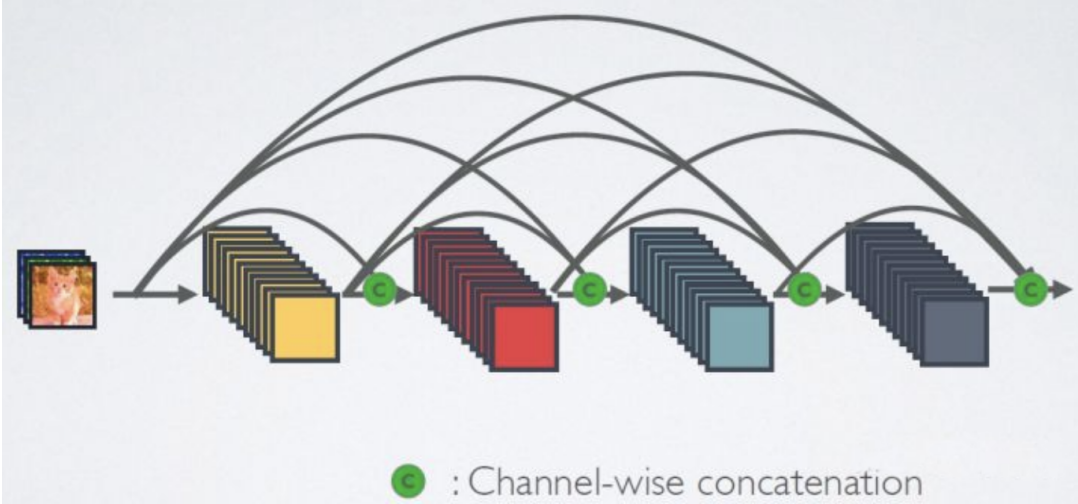

Figure 2.2: Overview of DenseNet structure (Huang et al., 2017)

the image. Two images generated from the same CT volume are randomly selected each time from all simulated images as the input image pair to the network. The output label of these two images are calculated by the open source tool geomstats package (Miolane et al., 2018) using the ground truth positions and orientations of the viewpoints of two input images. Each images are log-corrected and normalized into the range of $[-1, 1]$.

2.3 Network structure

The neural network architecture we design accepts two input images of the same size as shown in Fig. 2.3 and predicts the gradient of the geodesic distance on the Riemannian manifold between them, as per Sec. 2.1. The images first pass through the convolutional part of a DenseNet-161 (Huang et al., 2017) that was pre-trained on ImageNet dataset (Deng et al., 2009). The structure of DenseNet is shown in Fig. 2.2. While other popular pre-trained

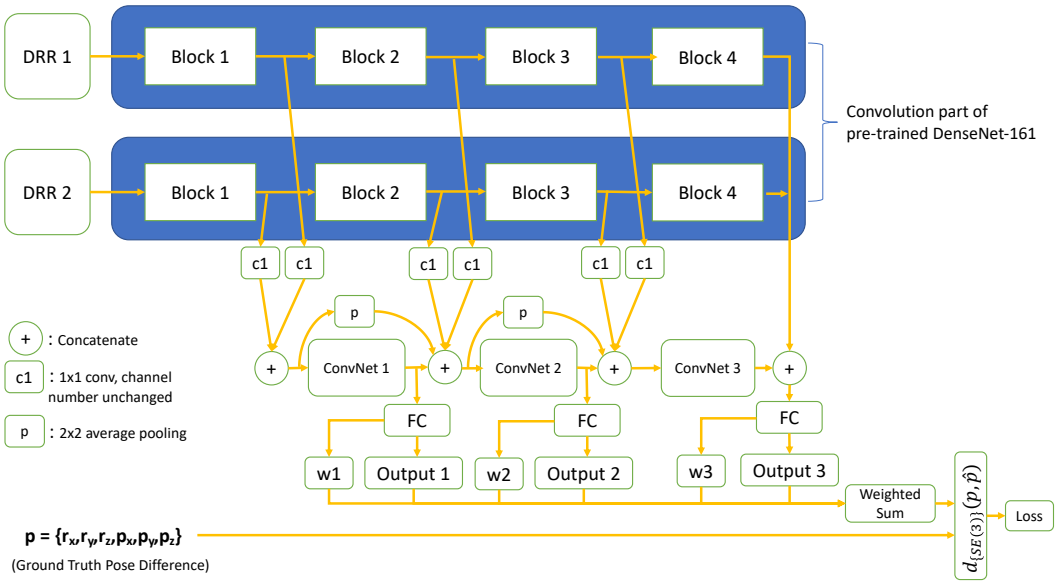

Figure 2.3: High-level overview of the proposed network architecture.

networks on ImageNet including VGG (Simonyan and Zisserman, 2015) and ResNet (He et al., 2016) are also tested in this architecture, DenseNet is able to produce the best result according to our experiment. This may be due to the fact that the design of densely connected blocks in DenseNet is able to most effectively reduce the gradient vanishing problem so that weights in the first few layers are still able to be updated properly. This architecture, shown in blue blocks in Fig. 2.3, is used to extract robust features from the two input images. Features from different depth are then fed into our custom architecture described in the remainder of this section. Note that the weights in pre-trained parts in the blue blocks are not updated in our training. Besides, as the DenseNet is designed to take three-channel RGB images as input, the weights of the first layer is averaged over the three channel so that it is able to fit the one channel intensity images.

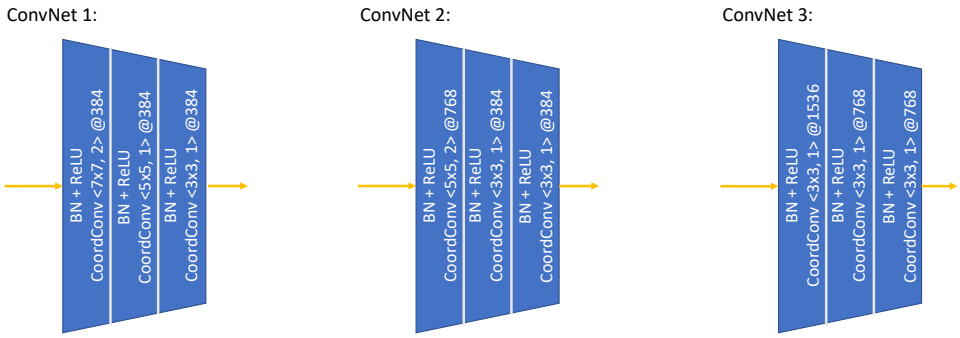

Figure 2.4: Detailed structure of each ConvNet.

While feature maps of deeper layers contain higher-level semantic information, most information on spatial configuration and local image appearance is lost; yet, such features are likely informative for predicting pose updates as relative image poses get closer. Assuming that feature maps from different depths of the pre-trained network represent different levels of semantic information, it is not guaranteed that those from deeper layers could always produce better result. Therefore, we 1) extract feature maps of both images at three different depths of DenseNet, 2) concatenate them, and 3) pass them through an additional CNN with fully connected layers to individually regress the geodesic gradient. Skip connection design is introduced in the additional CNN so that each CNN can get direct access to the lower level feature maps and the problem of gradient vanishing can be reduced. Average pooling layer is added to the skip connection to make the size of feature map

consistent. Since it is unclear which of these three estimates will be most appropriate at any given scenario, we delegate this decision to the CNN itself. This is realized by simultaneously regressing a weight corresponding to the geodesic gradient prediction at the respective depth. The weighted result, $\Delta p_{weighted} = \sum_{i=1}^{3} w_i \cdot \Delta p_i$, is the final output and is trained end-to-end with all the other three predictions.

While rotational pose changes were captured quite well in our experiments using this approach, purely translational displacements could not be predicted accurately. Following (Liu et al., 2018), we replace all convolution layers after feature extraction from DenseNet by CoordConv layers (Liu et al., 2018), which gives convolution filters access to the input image coordinate. As is suggested in the paper, all image coordinates are normalized into range [-1,1].

2.4 Reference coordinate system for regressing

The network was trained to regress the pose of X-ray source with respect to the CT volume in our initial experiments. However, regressing the camera pose is not able to produce stable results when the network is used to do registration because the target anatomy would easily move beyond the field of view in intermediate DRR images. The illustration of why target object always moves beyond the field of view is shown in Fig 2.5. Although the deep network is able to predict the rough direction that X-ray source should be moved to in the next step, the accuracy of each prediction is not guaranteed. The numerical error of a single prediction of the network is shown in Sec. 3.2. Small error in the prediction of X-ray source pose update may lead to large

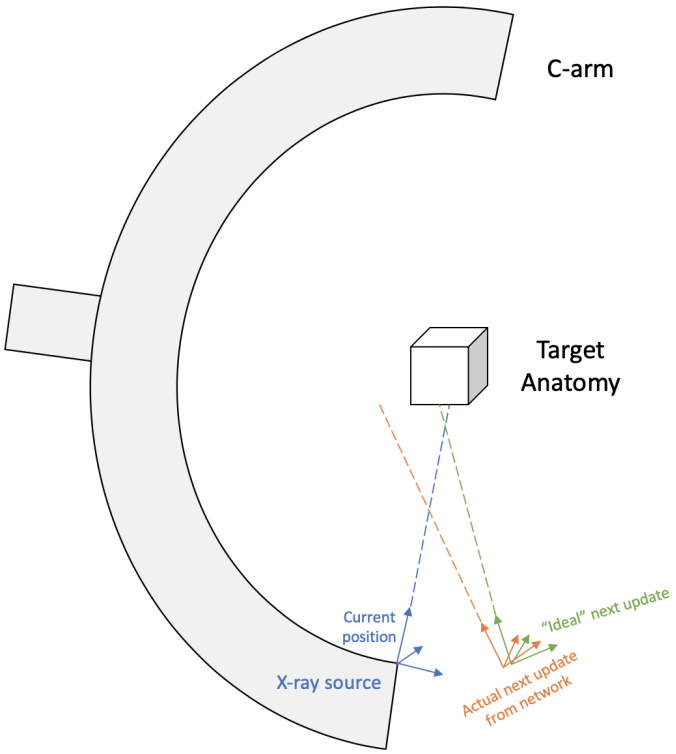

Figure 2.5: Comparison between "Ideal" X-ray source pose update and actual X-ray source pose update suggested by network

offset of the anatomical structure in the next simulated X-ray image. If the anatomical structure is mostly out of the field of view of detector in one of the intermediate DRRs, there is no way that the registration can proceed. To solve this problem, the reference coordinate frame is changed to the center of the target anatomy (in our test, the center of pelvis) and the X-ray source is assumed to be fixed. The network is then trained to regress the pose of CT volume relative to the new reference coordinate frame. As the network now predicts the pose update of CT volume relative to its own center, the target anatomy will not easily go out of the field of view if proper step size is applied (0.15 of geodesic gradient in our experiment).

Chapter 3

Experiments and Results

3.1 Experiment setup

We select five high-resolution CT volumes from the NIH Cancer Imaging Archive as the basis for our synthetic dataset and split data on the CT level (3 volumes for training, 1 for validation, 1 for testing). For each CT volume, a total of 4311 X-ray images were generated from different positions (randomly sampling poses with rotations $\phi \in [-40^\circ, 40^\circ]$, $\theta \in [-20, 20]$, and translations of ± 75 mm in all directions). Training was run on a single Nvidia Quadro P6000. Batch size used for training is 16; the learning rate starts from 1e-6, and decrease to 30% after every 30,000 iterations.

3.2 Network prediction accuracy

After training the network with simulated X-ray images which are generated from three different CT volumes, the performance of the network is tested on simulated X-ray images generated from the test CT volume. Fig. 3.1 shows the accuracy of 40 randomly sample simulated X-ray image pairs which are

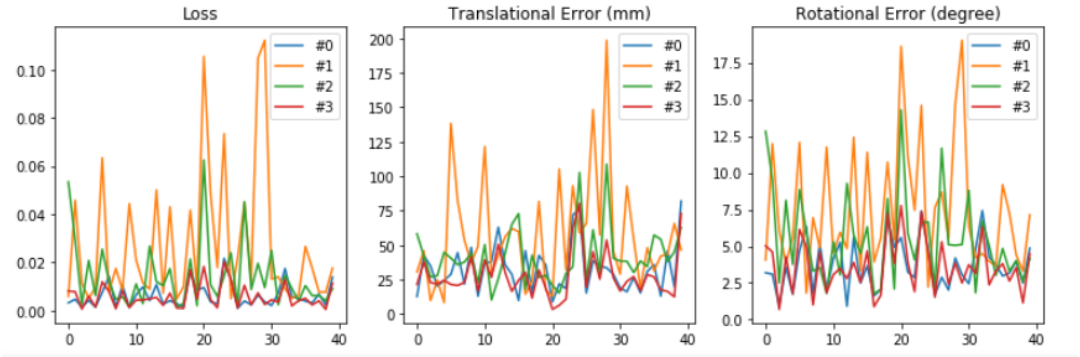

Figure 3.1: The rotational and translational accuracy of a single prediction by the network of 40 randomly sampled simulated X-ray image pairs from test CT volume. #0-2 indicates three individual predictions of three depths of network from shallow to deep. #3 is the weighted prediction.

generated from the test CT volume. Of the four results shown in the figure, #0-2 are predictions from three depths of network (the output 1-3 shown in Fig. 2.3), and the #3 is the final weighted prediction. It can be seen from the diagram that the weighted output is able to, in most cases, give the best result among all predictions. This result justifies the introduction of weights as a way to increase the capability of the network to make prediction on images with both large and small offset.

3.3 Registration results on simulated images

An exemplary case with intermediate DRRs is shown in Fig. 3.2 and Fig. 3.3.

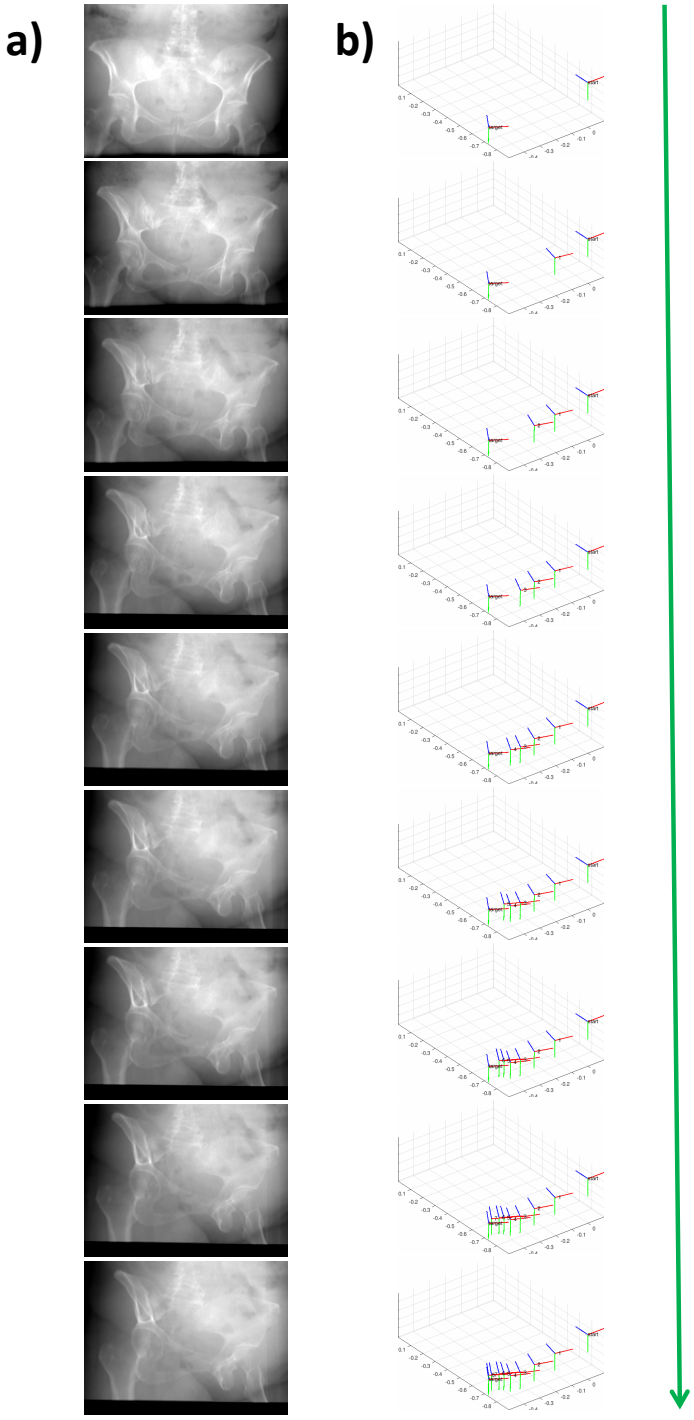

Figure 3.2: Synthetic data example: **(a)** DRRs as the network converges to the final pose. **(b)** The X-ray source pose correspond to the DRR on the left.

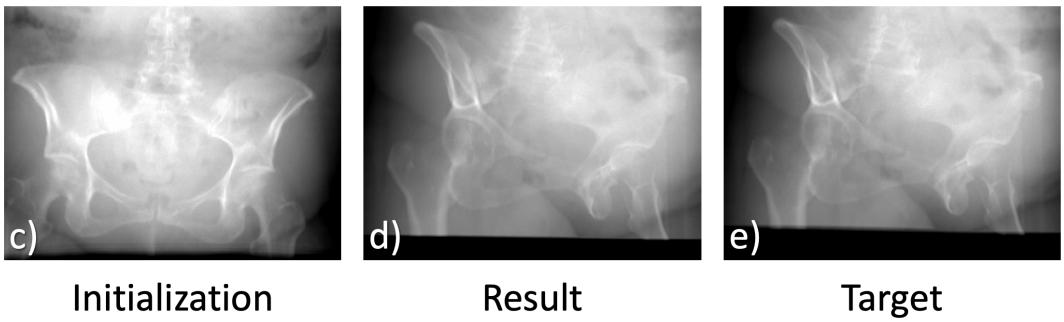

Figure 3.3: Synthetic data example (cont.): In **(c-e)** we show a DRR in AP view at initialization, the final DRR view, and the target image, respectively.

3.4 Registration results on real X-ray images

In Fig. 3.4 and Fig. 3.5, the network is used to predict the 2D/3D registration pose on a CT and two X-ray images from a cadaveric specimen. The X-ray images were acquired from a Siemens CIOS Fusion C-arm. The X-ray image was cropped for the registration because of the collimator trace on the boundaries of the image. Median filter is applied to the real X-ray images before they are used for prediction. While the registration result in Fig. 3.4 seems to be acceptable, that in Fig. 3.5 is rather off from the target. The results of both cases are compromised compared with simulated images. This indicates the fact that the simulated X-ray images is not in the same "domain" as real X-ray images. Directly training this network with simulated images without any pre-processing cannot generalize the model very well to real data. It is somehow counterintuitive because one of the advantage of neural network which researchers generally believe is that neural network is able to find out the most suitable kernel during training without the need of hand-crafting

features by human. Any kinds of pre-processing, if needed, should already be considered in the first few layers in ConvNet. This behavior, however, may only be true when there is enough variation in training dataset that will lead the network to form convolution kernel for that purpose. In our experiment, all training images are generated using the same model (DeepDRR). This, to certain extent can explain the compromised effect when applying the model on real X-ray images.

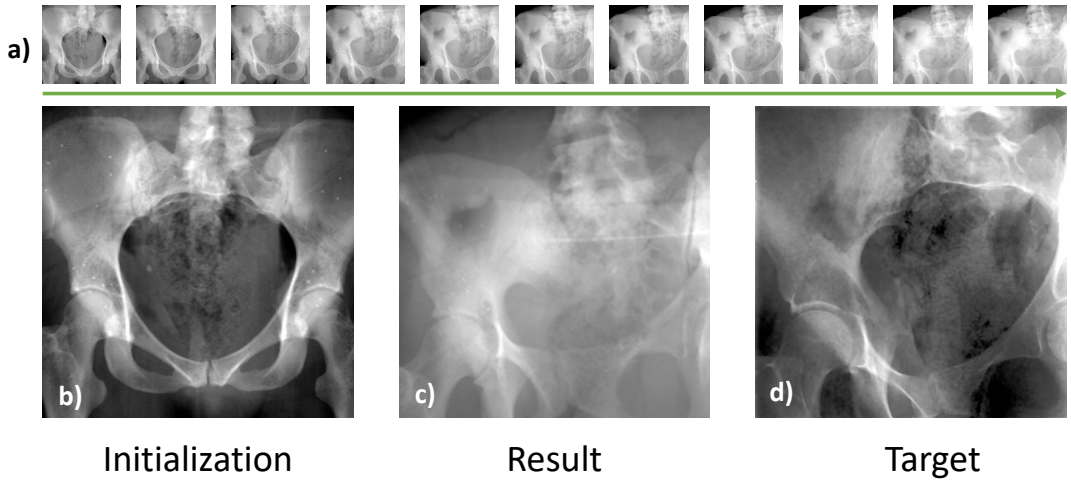

Figure 3.4: Real data example: DRRs are rendered from the CT every iteration (**a**). We also show a DRR in AP view at the first iteration, the final DRR view, and the real target image in (**(b-d)**), respectively.

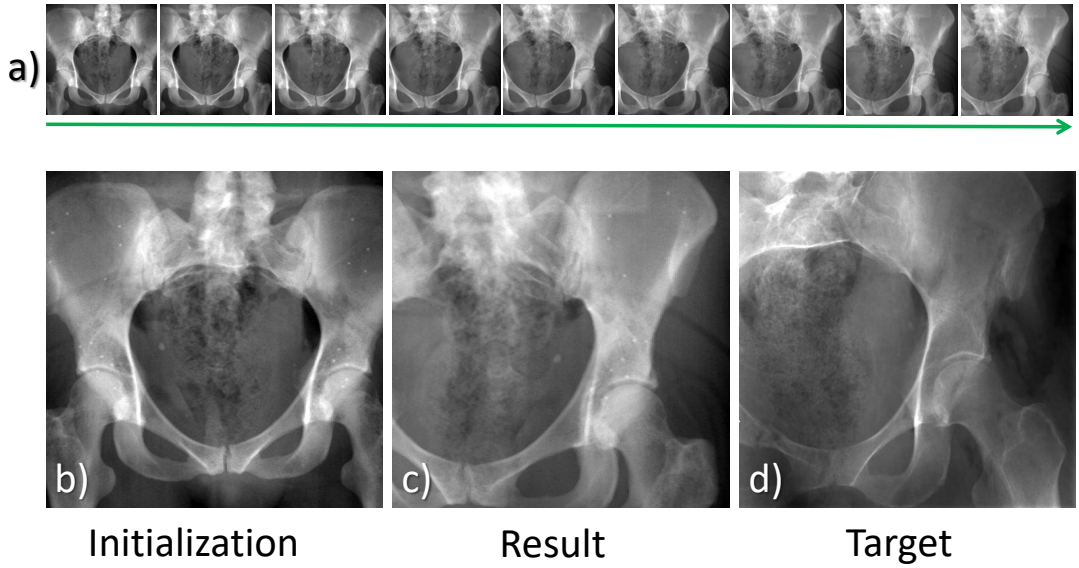

Figure 3.5: Failed real data example: DRRs are rendered from the CT every iteration (**a**). We also show a DRR in AP view at the first iteration, the final DRR view, and the real target image in (**(b-d)**), respectively.

3.5 Registration results compared with intensity-based registration on simulated images

In Fig. 3.6 we present the rotational and translational misalignment errors for ten synthetic test cases where ground truth pose is perfectly known. We compare the final pose of our contextual registration to intensity-based registration using covariance matrix adaptation evolution strategy (CMA-ES); both initialized at AP. The numeric comparison between the two approaches is presented in Table 3.1. Since the target image in our testing mostly has large offset from AP view (initialization), image-based registration method fails in all tests and results are trapped at local minimum close to initialization.

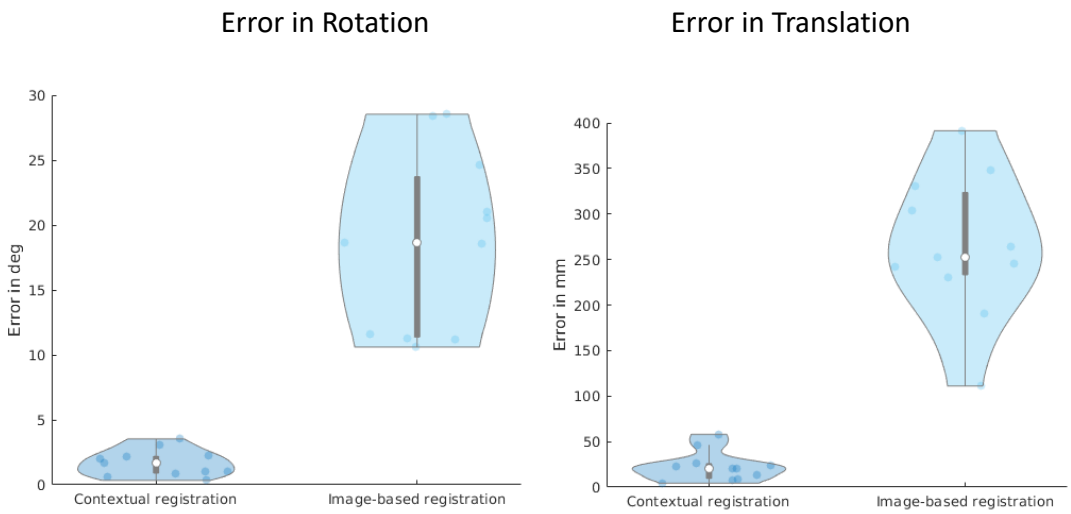

Figure 3.6: Violin plots showing the error distribution for rotational and translational misalignment. The plots compare the outcomes of contextual registration with classic intensity-based registration.

Table 3.1: Comparison of the proposed contextual registration with standard image-based registration, both initialized at AP view.

	Contextual Registration		**Image-based Registration**	
	Rotation	Translation	Rotation	Translation
Mean	1.68	22.8	18.6	264
Standard deviation	1.08	17.3	7.16	81.8
Median	1.53	20.1	19.6	249
Minimum	0.33	4.41	10.6	111
Maximum	3.54	58.0	28.5	391

Chapter 4

Discussion and Conclusion

4.1 Discussion and outlook

On synthetic and real X-ray images of the pelvis, we demonstrate that regressing geodesic gradients between target and current X-ray pose from the respective images enables the recovery of large pose differences in 2D/3D registration. These learned updates focus on context and content to overcome accurate initialization, a major challenge in intensity-based 2D/3D registration. While our results are promising, we have identified several limitations and directions for future work.

Currently, our contextual registration pipeline combines pose update proposals obtained from three learning-based sub-networks. Upon convergence, the recovered poses are reasonably close to the desired target pose, however, state of the art intensity-based registration methods (Grupp et al., 2019) can achieve even better performance if the parameters are well tuned. However, it may not be necessary to train a network that is able to produce similar level of prediction accuracy as intensity-based image registration when two images

are close enough. Our learning-based architecture is, in principle, capable of integrating pose updates provided by an intensity-based registration algorithm, the corresponding weight of which could be learned end-to-end. However, since such algorithm cannot provide an estimate of the geodesic gradient, careful design of the overall loss function is necessary.

When we try to use the trained model on real X-ray images, the results are quite compromising as is mentioned in Sec. 3.4. While texture information seems to be a nice feature in simulated-to-simulated image registration, it may not help as much in real-to-simulated image registration because it is not easy to normalize both input images in such a way that the intensity value at corresponding anatomical features are close to each other in similar viewpoint as it is in simulated-to-simulated registration. On the other hand, geometric information is a more robust feature than texture. One possible strategy would be pre-processing the X-ray images to emphasize more on geometric information and less on image texture before prediction. Possible pre-processing methods include applying Sobel filter and Gaussian filter to highlight the geometric contours and blur texture information. Using transfer-learning to stylized the raw image as a way of data augmentation is also an alternative to eliminate the effect of texture (Geirhos et al., 2018). A prospective cadaver study would allow implantation of radiopaque fiducial markers that can provide accurate ground truth, enabling these investigations and retraining of our CNN on real data.

4.2 Conclusion

In this work, we have shown that CNN is capable of learning large scale pose updates even when two images are far away from each other since the CNN-based similarity metric is globally convex. The proposed network regresses a geodesic loss function over $SE(3)$ and the results tested on simulated X-ray images are promising. However, compromised performance is observed when applying the method to real data that may further deteriorate as tools and implants are introduced during surgery. As a clear next step, we plan on quantifying the performance of our method on clinical data. Evaluating this behavior, however, is not trivial since ground truth poses for clinically acquired X-ray images are difficult to obtain, and strategies to improve generalization ability of CNNs are highly sought after. Some possible pre-processing methods mentioned in Sec. 4.1 can be applied to simulated images before used for training to improve the performance and generalization ability of the network on real data.

References

Yaniv, Ziv (2016). "Registration for orthopaedic interventions". In: *Computational radiology for orthopaedic interventions*. Springer, pp. 41–70.

Liu, Li, Timo Ecker, Steffen Schumann, Klaus Siebenrock, Lutz Nolte, and Guoyan Zheng (2014). "Computer assisted planning and navigation of periacetabular osteotomy with range of motion optimization". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 643–650.

Troelsen, Anders, B Elmengaard, and K Søballe (2008). "A new minimally invasive transsartorial approach for periacetabular osteotomy". In: *JBJS* 90.3, pp. 493–498.

Markelj, Primoz, Dejan Tomaževič, Bostjan Likar, and Franjo Pernuš (2012). "A review of 3D/2D registration methods for image-guided interventions". In: *Medical image analysis* 16.3, pp. 642–661.

Yao, Jianhua, Russell H Taylor, Randal P Goldberg, Rajesh Kumar, Andrew Bzostek, Robert Van Vorhis, Peter Kazanzides, and Andre Gueziec (2000). "AC-arm fluoroscopy-guided progressive cut refinement strategy using a surgical robot". In: *Computer Aided Surgery: Official Journal of the International Society for Computer Aided Surgery (ISCAS)* 5.6, pp. 373–390.

Otake, Yoshito, Mehran Armand, Robert S Armiger, Michael D Kutzer, Ehsan Basafa, Peter Kazanzides, and Russell H Taylor (2012). "Intraoperative image-based multiview 2D/3D registration for image-guided orthopaedic surgery: incorporation of fiducial-based C-arm tracking and GPU-acceleration". In: *IEEE transactions on medical imaging* 31.4, pp. 948–962.

Grupp, Robert B, Rachel A Hegeman, Ryan J Murphy, Clayton P Alexander, Yoshito Otake, Benjamin A McArthur, Mehran Armand, and Russell H Taylor (2019). "Pose Estimation of Periacetabular Osteotomy Fragments with Intraoperative X-Ray Navigation". In: *arXiv preprint arXiv:1903.09339*.

Gong, Ren Hui, James Stewart, and Purang Abolmaesumi (2011). "Multiple-object 2-D–3-D registration for noninvasive pose identification of fracture

fragments". In: *IEEE Transactions on Biomedical Engineering* 58.6, pp. 1592–1601.

Ketcha, MD, T De Silva, A Uneri, MW Jacobson, J Goerres, G Kleinszig, S Vogt, JP Wolinsky, and JH Siewerdsen (2017). "Multi-stage 3D–2D registration for correction of anatomical deformation in image-guided spine surgery". In: *Physics in Medicine & Biology* 62.11, p. 4604.

Chen, Xin, Jim Graham, Charles Hutchinson, and Lindsay Muir (2013). "Automatic inference and measurement of 3D carpal bone kinematics from single view fluoroscopic sequences". In: *IEEE transactions on medical imaging* 32.2, pp. 317–328.

Bier, Bastian, Mathias Unberath, Jan-Nico Zaech, Javad Fotouhi, Mehran Armand, Greg Osgood, Nassir Navab, and Andreas Maier (2018). "X-ray-transform invariant anatomical landmark detection for pelvic trauma surgery". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 55–63.

Unberath, Mathias, Jan-Nico Zaech, Sing Chun Lee, Bastian Bier, Javad Fotouhi, Mehran Armand, and Nassir Navab (2018). "Deepdrr–a catalyst for machine learning in fluoroscopy-guided procedures". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 98–106.

Hou, Benjamin, Nina Miolane, Bishesh Khanal, Matthew Lee, Amir Alansary, Steven McDonagh, Joseph Hajnal, Daniel Rueckert, Ben Glocker, and Bernhard Kainz (2018). "Deep Pose Estimation for Image-Based Registration". In:

Miolane, Nina, Johan Mathe, Claire Donnat, Mikael Jorda, and Xavier Pennec (2018). "geomstats: a Python Package for Riemannian Geometry in Machine Learning". In:

Bier, Bastian, Florian Goldmann, Jan-Nico Zaech, Javad Fotouhi, Rachel Hegeman, Robert Grupp, Mehran Armand, Greg Osgood, Nassir Navab, Andreas Maier, and Mathias Unberath (2019). "Learning to detect anatomical landmarks of the pelvis in X-rays from arbitrary views". In: *International journal of computer assisted radiology and surgery*, pp. 1–11.

Huang, Gao, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger (2017). "Densely connected convolutional networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708.

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei (2009). "Imagenet: A large-scale hierarchical image database". In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee, pp. 248–255.

Simonyan, K. and A. Zisserman (2015). "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *International Conference on Learning Representations*.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun (2016). "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pp. 770–778. DOI: 10.1109/CVPR.2016.90. URL: https://doi.org/10.1109/CVPR.2016.90.

Liu, Rosanne, Joel Lehman, Piero Molino, Felipe Petroski Such, Eric Frank, Alex Sergeev, and Jason Yosinski (2018). "An intriguing failing of convolutional neural networks and the coordconv solution". In: *Advances in Neural Information Processing Systems*, pp. 9628–9639.

Geirhos, Robert, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix Wichmann, and Wieland Brendel (2018). "ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness". In:

WENHAO GU

(+1) 443 388 2823 ⋄ petergu684@gmail.com
500 W University Pkwy, Apt 9K2
Baltimore, MD 21210

SUMMARY

A master student in robotics enthusiastic about integrating Augmented Reality (AR), Robotics and Deep Learning with traditional procedures.

EDUCATION

Johns Hopkins University *Sep 2017 - Present*
MSE in Robotics (Medical Robotics Track) GPA: 3.9/4

The Hong Kong Polytechnic University *Sep 2013 - Jun 2017*
BEng in Mechanical Engineering GPA: 3.86/4

TECHNICAL STRENGTHS

Programming Languages	C/C++, MATLAB, Python
Software & Tools	Solidworks, ROS, Pytorch, OpenCV, PCL, Arduino, Unity(C#)
Language	English, Chinese

WORKING EXPERIENCE

Course Assistant (Augmented Reality, by Nassir Navab) Feb 2019 - May 2019
Johns Hopkins University

· Correct written and programming assignments and hold office hours to answer students' questions.
· Supervise Augmented Reality class projects.

Research Internship (PI: Nassir Navab) Jun 2018 - Aug 2018
Computer Aided Medical Procedure (CAMP), Johns Hopkins University

· Develop Augmented Reality app on HoloLens using Unity to change the teaching experience of Orthopedic Trauma Surgery in Johns Hopkins Medical Institute.
· Setup the interactive learning environment by establishing the sharing service between two HoloLens.
· Evaluate and optimize the user interaction by conducting preliminary user study.

Research Assistant Mar 2016 - May 2017
The Hong Kong Polytechnic University

· Assist a PhD research group to do investigation on the temperature dependence of elastic modulus of advanced metallic materials.
· Use Matlab to analyze the data obtained and give suggestions for error reductions.

Research Assistant May 2016 - Aug 2016
Chinese University of Hong Kong (Shenzhen)

· Develop gesture recognition algorithm on a webcam of elder care robot with OpenCV.
· Realize face tracking with OpenCV on the robot.
· Setup human-robot interaction strategy from the vision perspective.

PROJECTS

3D/2D registration from CT to X-ray (master thesis) Aug 2018 - Present
Python, Pytorch

- Train a network to estimate pose difference between two digitally reconstructed radiograph (DRR).
- Develop a registration pipeline with the above-mentioned model to register a intra-operative 2D X-ray to a pre-operative 3D CT volume.

Segmentation of pelvis in X-ray images Nov 2018 - Dec 2018
Python, Pytorch

- Prepare dataset consisting of digitally reconstructed radiograph (DRR) images and segmentation labels for training.
- Train a convolutional neural network (U-Net) to segment out pelvis from X-ray images.

Tracking the surgical tool in Periacetabular Osteotomy Feb 2018 - May 2018
C++, PCL

- Process point cloud data from depth camera (RealSense) and do 3D background segmentation and pose estimation on a surgical tool with known model.

Measuring distance from virtual to real in AR environment Apr 2018 - May 2018
C#, Unity

- This project is to develop an AR app using Unity that can measure 6DOF (translational and rotational) distance between virtual and real object (markerless).
- I am responsible for processing 3D point cloud data obtained from HMD (Meta 2) and come up with a reasonable pose estimation.

Provide solar power solutions in rural villages in Cambodia Jun 2015

- Build mini solar power charging stations for two rural villages and install customized electrical appliances in six families.
- Teach local residents how to use electrical appliances and recharge the batteries. Write a troubleshooting documentation that can be understood by residents who do not know English.

Manufacturing of Robotic Beetles Sep 2014 - Apr 2015

- Fabricate movable robotic metal beetles by using conventional machining, CNC machining, metal casting, stamping, surface finishing and QC measurement in the industry on campus.

RELEVANT COURSES

Computer Integrated Surgery I & II
Robot Devices, Kinematics, Dynamics, and Control
Robot System Programming
Medical Image Analysis
Algorithms for Sensor-Based Robotics
Augmented Reality
Machine Learning: Deep Learning

PUBLICATIONS

J. Fotouhi*, M. Unberath*, T. Song*, **W. Gu**, A. Johnson, M. Armand, G. Osgood, N. Navab. "Interactive Flying Frustums (IFFs): Spatially-Aware Surgical Data Visualization", *The 10th International Conference on Information Processing in Computer-Assisted Interventions*, 2019