EPiC Series in Health Sciences

Volume 5, 2022, Pages 45–50

Proceedings of The 20th Annual Meeting of the International Society for Computer Assisted Orthopaedic Surgery

A benchmark for Deep Learning-based approaches for *In-vivo* segmentation of 2D images in Total Knee Arthroplasty

Baptiste Dehaine, Marion Decrouez, and Nicolas Loy Rodas

Ganymed Robotics, Paris, France

baptiste.dehaine@ganymedrobotics.com, nicolas.loyrodas@ganymedrobotics.com

Abstract

Progress in machine learning and artificial intelligence (AI) opens the way to the development of smart clinical-assistance systems and decision-support tools for the operating room (OR). Yet, before deploying these algorithms in the OR, assessment of their performances in real clinical conditions is necessary. Gathering intraoperative data for training and testing is hard, and robustness to the challenging conditions of the OR is not always demonstrated. In this paper we introduce a unique multi-patient dataset of images captured during Total Knee Arthroplasty (TKA) surgery. We use this dataset to compare five deep learning-based image segmentation approaches and provide quantitative and qualitative results. We hope that this work will help bringing light on the performances of AI in a real surgical environment.

1 Introduction

State-of-the-art AI approaches have been extensively evaluated on everyday non-clinical images, showing an increase in maturity and robustness. However, these are rarely demonstrated on clinical datasets and it remains hard to tell how well they can generalize to the OR environment [10]. Intraoperative data processing algorithms require curated, structured, and annotated data, and this hinders their development and evaluation.

With the aim of generating a reference clinical dataset, we organized a clinical trial[1] and gathered intraoperative data from 62 Total Knee Arthroplasty (TKA) surgeries. The study was approved by an ethics committee and took place over several months in France. This unique dataset presents the inherent challenges of the OR, such as illumination, clutter, occlusions, and patient morphology variations.

We hereby make use of a subset of our dataset to evaluate the performances of five deep learning-based approaches for medical image segmentation, applied to the segmentation of knee bones (femur and tibia) in RGB images. We chose this task since accurately localizing the bones in images from the exposed knee is crucial for enabling future marker-less registration and tracking systems [1]. Indeed, soft-tissues or surgical instruments surrounding the targeted

[1] Clinical trial NCT04912908

anatomy can bias the registration result. Therefore, a prior pre-processing step is typically applied to filter-out the target bones from the images and thus improve the robustness of any registration algorithm that follows [7, 6, 4, 1].

To the best of our knowledge, we are the first to benchmark AI approaches for *in-vivo* bone segmentation on a large multi-patient intraoperative dataset.

1.1 Related Works

[6, 4] and [1] both apply an adaptation of the U-Net[9] deep neural network for segmenting the knee bones on depth data, looking to improve marker-less registration in the context of surgical navigation. Images from a single cadaveric knee were captured in a controlled laboratory setup and used both for training and testing. Hence, the proposed approaches may suffer from overfitting and will not directly generalize well to other patient morphologies or different experimental conditions. Also, no quantitative evaluation is performed on real OR data. It is thus difficult to draw conclusions about the resilience of a U-Net based segmentation to the inherent complications of a clinical environment. While [1] also shows qualitative results on selected surgical videos found in the VuMedi library, these are hard to interpret and an evaluation on *in-vivo* data is still necessary.

In [8], Mask-RCNN [3] is applied to *in-vivo* data captured during five TKA surgeries. However, such a small set of patients do not provide enough bone morphology variations for a complete assessment of the method.

2 Method

2.1 *In-vivo* dataset description

Video sequences were recorded with a hand-held camera during 62 TKA procedures, capturing the femur and tibia before the bone resection phases. Of the 62 patients, 33 (53.2%) were females. The mean age of the population was 70.6 (CI 95% = [68.5; 72.7]) and the mean BMI was 28 kg/m^2 (CI 95% = [26.7; 29.3]). Both legs are equally represented (30 right TKA). All types of arthritis are also represented in the data since this has an influence on the bones' morphology: 33 (53.2%) patients suffered from internal femorotibial arthritis, 14 (22.6%) from external femorotibial arthritis, 4 (6.5%) from patellofemoral arthritis and 11 (17.7%) had several damaged knee compartments by arthritis. The recorded surgeries were performed by two different surgeons. For each patient, 12 images from varying points of view are extracted from the recording: 6 for the femur and 6 for tibia. This results in a total of 744 images with a high variability in shape, illumination conditions and viewpoint. The contour of the bones was manually annotated in all images by an expert.

2.2 Neural networks for bone segmentation

The following networks for 2D medical image segmentation were evaluated: U-Net[9], CE-Net[2], U-Net++[12], ResUNet[11], ResUNet++[5]. They each take as input an RGB image and output a prediction about the pixels corresponding either to femur or tibia.

We randomly split our dataset into 65% for training, 15% for validation and 20% for testing. Images from the same patient are present in a single split only, hence no training and testing is done on data from the same surgery. Every image is resized to 256×256. We apply a binary cross entropy loss with an Adam optimizer for training, a batch size of 2, a learning rate

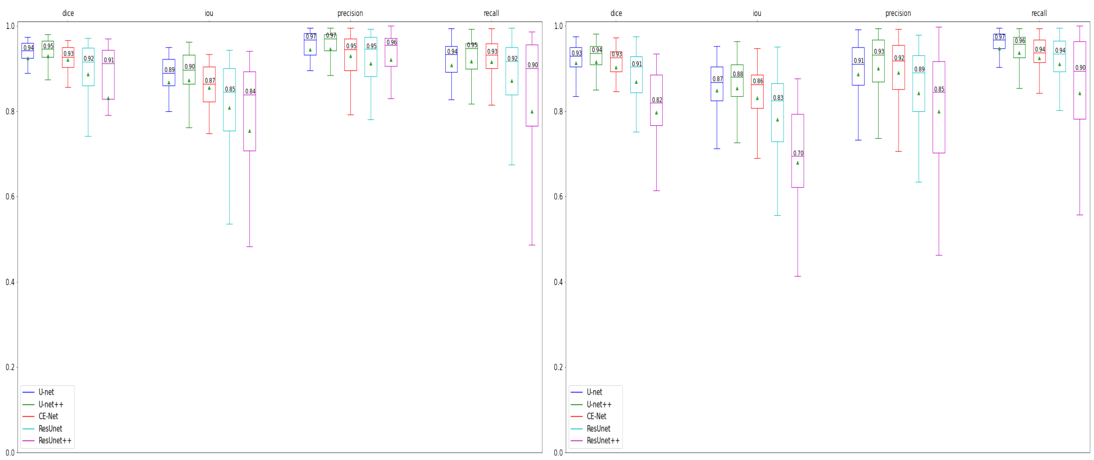

Figure 1: Evaluation results (Dice similarity score, IoU, Precision and Recall) for five deep learning-based segmentation approaches on an *in-vivo* dataset: Left for femur and right for tibia.

starting at 10^{-4} and early stopping. Standard data augmentation steps are applied: random horizontal flipping and rotation, along with changes in image brightness, contrast, and saturation. Furthermore, among the patients included in the training set, only one image per patient is randomly selected and used for each training epoch[2]. By doing so, we exploit the multiple point of views included in our dataset and reduce any possible overfitting to the same patient.

3 Evaluation results

Overall, all evaluated networks yield good results on our dataset (see boxplots 1). The best performance is achieved by U-Net++. ResUNet and ResUNet++ yield the lowest precision and recall, therefore seem less adapted for an OR environment. The performances for tibia segmentation are almost equivalent to femur, even if the tibia is partially occluded by surrounding tissues in the images (see figure 2 for qualitative results). Our training strategy allows all networks to generalize properly and avoid overfitting, as segmentation is robust to variations in bone morphology and set-up.

Further work will focus on including data from additional hospitals to add more variations to the images and also adding an inter-annotator variations' study.

4 Conclusions

In this paper, we present comparative results of five state-of-the-art deep-learning approaches applied to the segmentation of bony structures in *in-vivo* images during TKA surgery. Previous works [7, 6, 4, 1] provide quantitative results from *ex-vivo* data only and thus models that would not directly generalize to intraoperative data. Hence, we introduced a unique multi-patient dataset of RGB images captured during 62 surgeries and used it for this benchmark study.

[2]Training and testing was performed on a Dell XPS 15 with a GTX 1650 Ti GPU.

Data-driven approaches capable of interpreting and extracting knowledge from real-life intraoperative data are crucial for future computer-assisted systems. These can be applied for suggesting patient-specific intervention strategies, quantify bone resections or blood loss, help prevent medical errors, and overall increase both quality of care and physician experience. Yet, there still is a need to better understand how AI-based algorithms behave in real-life environment. This work provides useful insights about the performances of state-of-the-art approaches in challenging OR conditions.

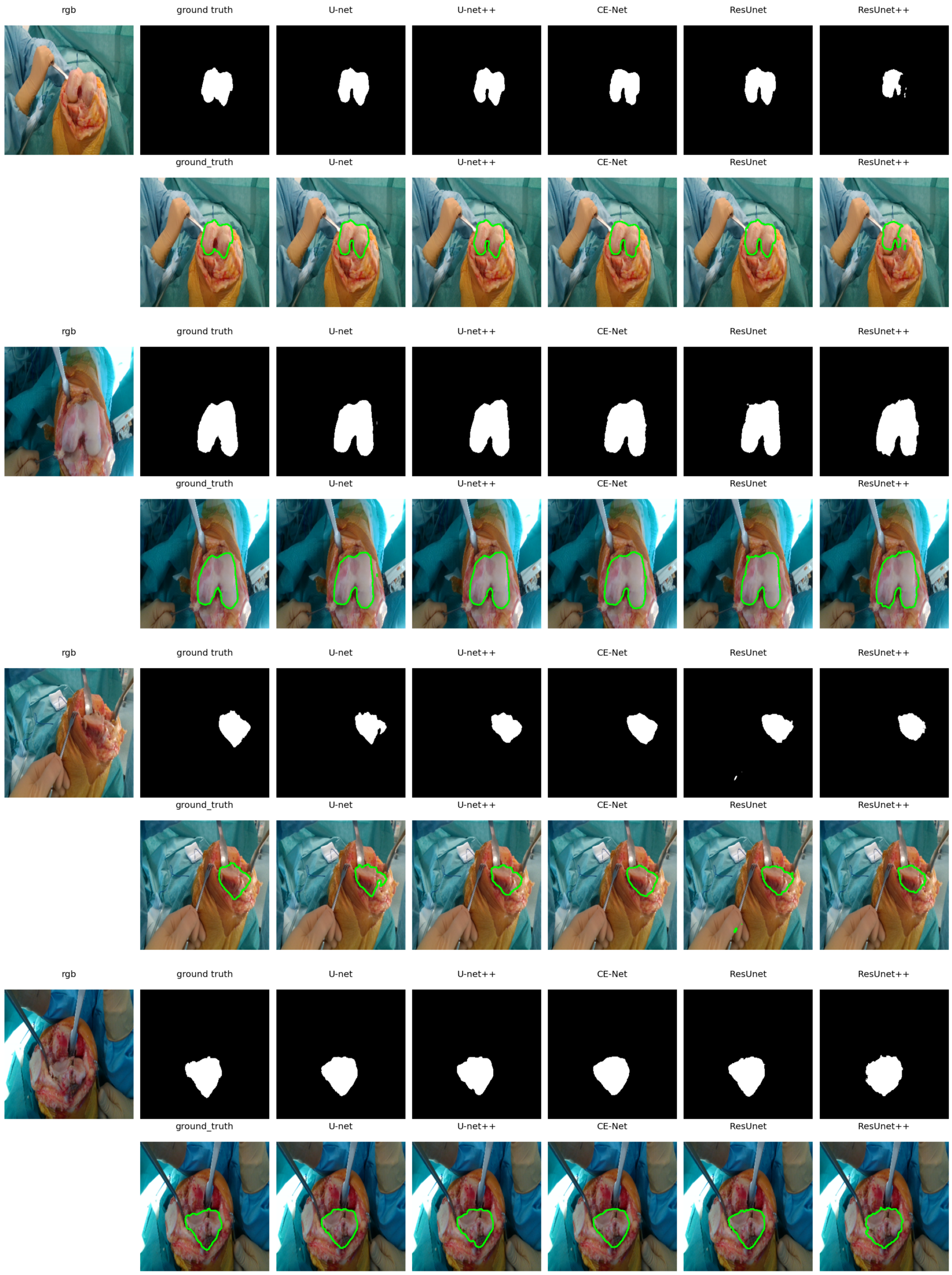

Figure 2: Qualitative results for all five evaluated approaches for femur and tibia segmentation on intraoperative RGB images.

References

[1] Inês Félix, Carolina Raposo, Michel Antunes, Pedro Rodrigues, and João P. Barreto. Towards markerless computer-aided surgery combining deep segmentation and geometric pose estimation: application in total knee arthroplasty. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 9(3):271–278, 2021.

[2] Zaiwang Gu, Jun Cheng, Huazhu Fu, Kang Zhou, Huaying Hao, Yitian Zhao, Tianyang Zhang, Shenghua Gao, and Jiang Liu. Ce-net: Context encoder network for 2d medical image segmentation. *IEEE Transactions on Medical Imaging*, 38(10):2281–2292, 2019.

[3] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, pages 2980–2988, 2017.

[4] Xue Hu, He Liu, and Ferdinando Rodriguez Y. Baena. Markerless navigation system for orthopaedic knee surgery: A proof of concept study. *IEEE Access*, 9:64708–64718, 2021.

[5] Debesh Jha, Pia H. Smedsrud, Michael A. Riegler, Dag Johansen, Thomas De Lange, Pål Halvorsen, and Håvard D. Johansen. Resunet++: An advanced architecture for medical image segmentation. In *2019 IEEE International Symposium on Multimedia (ISM)*, pages 225–2255, 2019.

[6] He Liu and Ferdinando Rodriguez Y Baena. Automatic markerless registration and tracking of the bone for computer-assisted orthopaedic surgery. *IEEE Access*, 8:42010–42020, 2020.

[7] He Liu and Ferdinando Rodriguez Y Baena. Automatic Bone Extraction from Depth Images in Robotic Assisted Knee Replacement. *Proceedings of the 12th Hamlyn Symposium on Medical Robotics*, 2019.

[8] Nicolas Loy Rodas, Marion Decrouez, Blaise Bleunven, and Sophie Cahen. In-vivo bone segmentation approach for total knee arthroplasty. In Ferdinando Rodriguez Y Baena and Fabio Tatti, editors, *CAOS 2020. The 20th Annual Meeting of the International Society for Computer Assisted Orthopaedic Surgery*, volume 4 of *EPiC Series in Health Sciences*, pages 183–187. EasyChair, 2020.

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[10] Tom Vercauteren, Mathias Unberath, Nicolas Padoy, and Nassir Navab. Cai4cai: The rise of contextual artificial intelligence in computer-assisted interventions. *Proceedings of the IEEE*, 108(1):198–214, 2020.

[11] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018.

[12] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 39(6):1856–1867, 2020.