EPiC Series in Health Sciences

Volume 4, 2020, Pages 183–187

CAOS 2020. The 20th Annual Meeting of the International Society for Computer Assisted Orthopaedic Surgery

In-vivo bone segmentation approach for Total Knee Arthroplasty

Nicolas Loy Rodas, Marion Decrouez, Blaise Bleunven, and Sophie Cahen

Ganymed Robotics, Paris, France
nicolas.loyrodas@ganymedrobotics.com

Abstract

Perceiving and making sense of the surgical scene during Total Knee Arthroplasty (TKA) surgery is crucial for building assistance and decision support systems for surgeons and their team. However, the need for large volumes of annotated and structured data for AI-based methods hinders the development of such tools. We hereby present a study on the use of transfer learning to train deep neural networks with scarce annotated data to automatically detect bony areas on live images. We provide quantitative evaluation results on *in-vivo* data, captured during several TKA procedures. We hope that this work will facilitate further developments of smart surgical assistance tools for orthopaedic surgery.

1 Introduction

AI's potential for healthcare applications such as medical diagnosis and big data analytics is the subject of much enthousiam and research [10, 2, 5]. In the context of Computer-Assisted Orthopaedic Surgery (CAOS), multi-modal data generated during a procedure can be used to develop smart clinical assistance tools using Machine/Deep Learning. However, this requires extensive, structured and high-quality labeled data, which can be hard to obtain.

In this paper we present a study on the use of deep neural networks to segment the bone surface on *in-vivo* images captured during TKA procedures. To cope with the data volume issue, we propose to apply transfer learning and leverage the use of large open-source *non-clinical* datasets to pre-train a neural network and thus alleviate the need of a large annotated *clinical* dataset. We provide promising quantitative and qualitative results on *in-vivo* images to illustrate the potential of transfer learning to support more frugal approaches to the complex problem of surgical scene perception and understanding.

1.1 Related Works

[3] shows the potential of using structured light scanners to capture 3D data of a surgically exposed knee to be applied as an intraoperative image modality. In a follow-up paper [4], the authors investigate the use of machine learning to classify various anatomical tissues using textural information from the scans. Their method is trained on 562 961 datapoints from scans of three frozen knee specimens, manually segmented by an expert.

F. Rodriguez Y Baena and F. Tatti (eds.), CAOS 2020 (EPiC Series in Health Sciences, vol. 4), pp. 183–187

Similarly, [9] proposes to use a deep learning network to segment the knee bone surface from depth images captured from a cadaveric knee. Their approach is trained on a dataset of 2 000 manually labeled images from a laboratory study.

The aforementioned works [3, 4, 9] are promising since they aim at obtaining 3D bone surface data without the need for an optically tracked probe. However, the proposed anatomy detection methodologies require large annotated datasets and have only been validated on cadaver/*ex-vivo* data. Images captured in operating rooms pose inherent challenges (illumination, occlusions, viewpoints...), hence *in-vivo* data is still needed to validate the applicability of such approaches in a real clinical setup.

1.2 Contributions

- We demonstrate the potential of transfer-learning to reduce the amount of labeled data needed to train a deep network for a CAOS application.

- To the best of our knowledge, this is the first quantitative study of a bone segmentation methodology evaluated on *in-vivo* color images captured during real TKA procedures in a real clinical setting (with its inherent challenges such as illumination, point-of-view and patient morphology variability).

2 Method

2.1 Femur detection model

We apply a state-of-the-art object detection and semantic segmentation model *Mask R-CNN* [7]. It combines a feature pyramid network backbone followed by a region proposal network for generating segmentation masks and class predictions at pixel-level, for a given query object within an input image.

Our implementation is based on [1] and relies on a ResNet-101 architecture as feature extractor backbone model. Training such a network end-to-end from scratch would require thousands of labeled images. Since this is not feasible in a clinical context, we propose to use weights pre-trained from an open-source dataset (COCO [8]) composed of 328k labeled images from everyday objects. We then retrain only the output layers with our clinical dataset to fine-tune the model to our application.

2.2 *In-vivo* Data

We recorded video sequences during five TKA procedures. Different points of view of the knee joint were captured between exposure and bone cutting (see figure 1). The sequences were randomly sub-sampled to obtain a database of 900 images (1280×720 px) that were manually annotated by a non-expert. We used an open-source annotation tool [6] to approximately draw a polygon around the femur on each image and then generate ground-truth masks. This annotation strategy was simple and fast (around 10 seconds per image), as opposed to [4, 9] where annotating 3D point clouds or depth images can be non-intuitive.

	Dataset size (images)	Epochs	Data	Mean DSC	Mean IoU
[9]	2000	250	Cadaver	92%	NA
Ours 1	900 (25% train-10% val.)	20	*In-vivo*	**94 ± 0.03** %	**89 ± 0.04%**
Ours 2	900 (20% train-Patient 1)	20	*In-vivo*	78 ± 2.1 %	73 ± 3%

Table 1: Quantitative results on our *in-vivo* dataset. For each approach, we provide the size of the labeled dataset, the number of training epochs, the nature of the data, the Dice Similarity Coefficient and the Intersection over Union. *Ours 1*: our test-set performance when training on 25% of our shuffled dataset. *Ours 2*: detection results when training on data from one patient and testing on the remaining four TKA sequences.

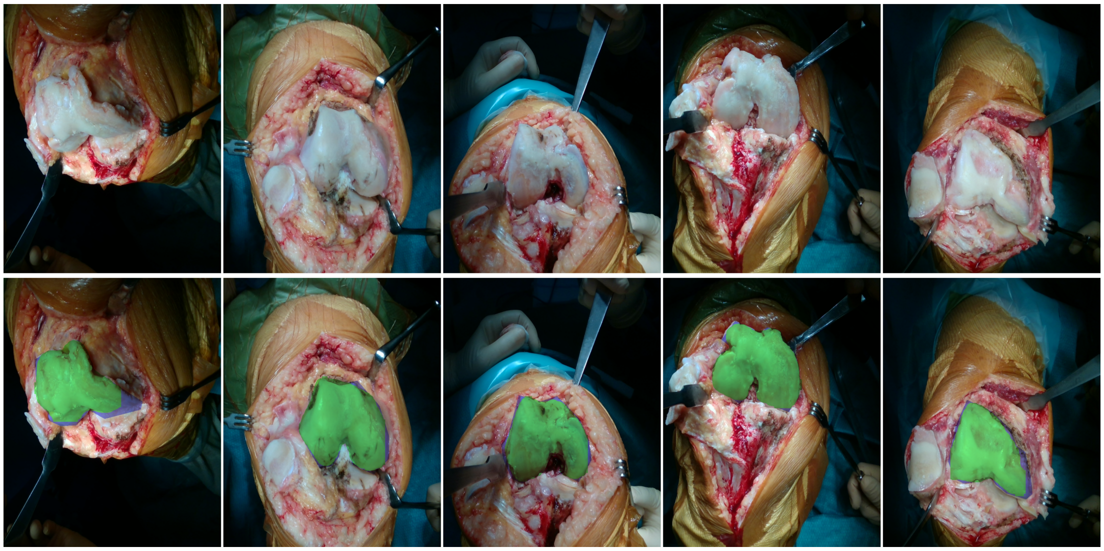

Figure 1: *First row:* Examples of *in-vivo* images from five TKA procedures from our dataset. *Second row:* Results of our femur detection approach in images from the testing dataset. The prediction is shown in green and the ground-truth in purple.

3 Evaluation

3.1 Results

To prevent biased model fitting, we split our shuffled dataset into training, validation and testing sets. We compare our results to [9] in table 1 since their segmentation task is the same as ours (on depth images). We also provide results when training only in images from a single patient to assess the method's generalization performance to morphology variations. We obtain a 94% DSC (89% IoU) on our testing set (images unseen by the network) when using only 25% of our labeled images for training. Also, when training with images from one TKA only (one patient), our method yields correct segmentation results on the rest of the data (78% DSC).

We have deployed our model into a C++ application to perform inference directly in the live stream of our camera. With no code optimization, our inference runs at 7 fps on a laptop[1].

[1] Dell XPS 15 with a GeForce 1050 Ti GPU.

3.2 Discussions

Our approach reaches a comparable DSC with significantly less annotated data and less training time (only 20 epochs needed thanks to our transfer learning strategy), when compared to [9]. [4] performs a multi-class detection on 3D point-clouds, therefore methods are not directly comparable. Yet, since *Mask R-CNN* [7, 1] can be easily applied for multi-class detection, our approach can be extended to detect multiple anatomical structures while requiring less annotated images (versus 300k for [4]).

This is a preliminary study with room for further experimentation (e.g. occlusions, larger testing set, comparison to other detection approaches). Still, the fact that using a training dataset with few clinical images yields good anatomy detection results is encouraging.

4 Conclusions

In this paper we explore the feasibility of applying transfer learning to train a surgical scene recognition model for CAOS. We provide promising results on *in-vivo* data from real clinical conditions. We hope this work can contribute to the development of smart surgical assistance tools despite the difficulty of obtaining high-quality data inherent to the field. Furthermore, as CAOS expands from surgical guidance to more complex topics such as workflow recognition and tool tracking, novel approaches for sensing and understanding the surgical scene will be needed.

Acknowledgments

We would like to thank Dr. Michel Bonnin and Dr. Tarik Ait Si Selmi for helping us capturing the data for this study.

References

[1] Waleed Abdulla. Mask r-cnn for object detection and instance segmentation on keras and tensorflow. https://github.com/matterport/Mask_RCNN, 2017.

[2] Ahmed Alsinan, Michael Vives, Vishal Patel, and Ilker Hacihaliloglu. Spine surface segmentation from ultrasound using multi-feature guided cnn. In *CAOS 2019. The 19th Annual Meeting of the International Society for Computer Assisted Orthopaedic Surgery*.

[3] Brandon Chan, Jason Auyeung, John F. Rudan, Randy E. Ellis, and Manuela Kunz. Intraoperative application of hand-held structured light scanning: a feasibility study. *International Journal of Computer Assisted Radiology and Surgery*, 2016.

[4] Brandon Chan, Jason Auyeung, John F. Rudan, Parvin Mousavi, and Manuela Kunz. Tissue classification using machine learning to aid in intraoperative registration: a pilot study. In *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*. International Society for Optics and Photonics, SPIE.

[5] David Chen, Sijia Liu, Paul Kingsbury, Sunghwan Sohn, Curtis B. Storlie, Elizabeth B. Habermann, James M. Naessens, David W. Larson, and Hongfang Liu. Deep learning and alternative learning strategies for retrospective real-world clinical data. *npg Digital Medicine*, 2019.

[6] Abhishek Dutta and Andrew Zisserman. The VIA annotation software for images, audio and video. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, New York, NY, USA, 2019. ACM.

[7] 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.

[8] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, 2014.

[9] 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.

[10] Xiaoxuan Liu *et. al.* A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The Lancet Digital Health*, 1(6), 2019.