Automatic Segmentation of Anatomical Structures for Core Decompression Procedures

Background Reading and Critical Review

Student: Mingxu Liu
Mentors: Ping-Cheng Ku, Alejandro Martin-Gomez

Project Summary

Core decompression is a treatment for early-stage Osteonecrosis of the femoral head (ONFH). The core decompression procedures require preoperative MRI for surgical planning and intraoperative CT for surgical guidance. However, since the CT is not sensible to necrotic tissue, finding the transformation from MRI to CT is crucial to mapping the necrosis into CT coordinates during the surgery.

A pipeline to achieve such mapping is by: 1). segmenting the relative anatomies on both the MRI and CT volumes; 2). finding the transformation between MRI and CT by registering segmented bone anatomies; 3). using the change derived from step two, mapping the necrosis coordinates from MRI volume onto CT volume.

The goals of this project, corresponding to the procedures described above, are:

1. Realize the automatic segmentation of the anatomical structures, including femur, pelvis, and necrosis, in MRI volumes used for core decompression procedures.
2. Realize the automatic segmentation of femur and pelvis in CT volumes used for core decompression procedure.
3. Realize the registration between the segmented anatomical structures (femur and pelvis) in MRI (first goal) and CT (second goal). A further topic is integrating the segmentation and registration models to form an integrated system and realize its interaction with 3D Slicer.

Paper Selection

This project consists of multiple sub-tasks, including MRI segmentation, CT segmentation, and registration. Among them, MRI segmentation is the focus. Considering that there have been well-developed toolkits for registration, and this project is likely to use these toolkits in a black-box manner, the publication of related toolkits will not be discussed here.

Two papers are selected. For automatic MRI segmentation, the work by Deniz et al. [1]: *Segmentation of the Proximal Femur from MR Images using Deep Convolutional Neural Networks*, will be discussed. For CT segmentation, the work done by Krcah et al. [2]: *Fully Automatic and Fast Segmentation of the Femur Bone from 3D-CT Images with No Shape Prior* will be discussed. These two papers are contributive in their concentration, highly related to the project, and provide great insights into implementation.

Critical Review

Paper I: Segmentation of the Proximal Femur from MR Images using Deep Convolutional Neural Networks [1]

Summary of problem

MRI is a critical complementary method for osteoporosis assessment because of its advantages in distinguishing between bone marrow and trabecular bone tissue within the proximal femur [1]. Most recently proposed diagnosis practice involves investigating the mechanical properties of the entire femur. However, such an investigation required the segmentation of the whole of the proximal femur. Manual segmentation requires well-trained knowledge and is time-consuming, the quality of segmentation depends on expertise [1]. Deniz et al. introduced Convolutional Neural Networks to the task and achieved the automatic segmentation of the proximal femur. Moreover, they compared the performance of different network structures on this task.

Key Results

This work achieved the best performance with 3D CNN with dilated design, with a Dice Score (DSC) of 0.953 ± 0.016 and a precision of 0.954 ± 0.017. Five evaluation metrics, including recall rate, averaged symmetric surface distance (ASD), and maximum surface distance (MSD), are compared. A table summarizes the results are provided in Appendix Table A.

Deniz et al. have also summarized their key findings fourfold:

i. In general, statistically, 3D CNN outperforms 2D CNN on proximal segmentation tasks.
ii. Appropriate dilated design of the convolutional layers improves the model's performance.
iii. 3D CNN performs better on segmentation accuracy on the hypointense regions, with fewer false positives.

Significance of key results

Deniz et al. 's work, based on the existing development of CNN in medical image segmentation, realized and demonstrated the feasibility of using CNN to segment femurs in MRI images. At the same time, their work compared the performance of 2D CNN, unpadded 2D CNN, 3D CNN, and the dilated versions of these models under the task of femur segmentation with different network structure topology. It gave generalized design recommendations of network structure for similar work.

Background

Convolutional Neural Network (CNN) is a variation of a Deep Neural Network (DNN), which is designed to process spatial features in data using data patterns [3]. Concentrated on medical image segmentation, U-Net was proposed by Olaf et al. and is now regarded as the baseline model [4]. U-Net innovatively introduced the feature channels on up-sampling stages and the bottleneck design, which is well suited for segmentation tasks. Dedicated researchers have been working on developing novel segmentation networks based on the architecture of U-Net, including the 3D variation of U-net.

Methodologies and Experiments

In Deniz et al.'s work, based on the most common U-Net variations, including dilated/not dilated 2D CNN with 64 initial feature maps and three/four numbers of layers, dilated/not dilated 3D CNN with 16/32 initial feature maps and three/four numbers of layers, are trained and tested on 86 femur MRI volumes obtained from the New York University School of Medicine. Four-fold cross-validation is performed to assess the model's performance with fold separations of 21/21/22/22.

The network uses negative log-probability as a target loss function:

$$CE = -\frac{1}{N}\sum_{i=1}^{N}\left(\frac{N_b}{N} y_i \log\, p_i + \frac{N_p}{N}(1-y_i)\log\,(1-p_i)\right)$$

Where y_i is the ground truth of pixel i, and p_i is the corresponding output from the model. Other hyper-parameters, including learning rate, initialization methods, input size, etc., can be referred to from the paper. Early stopping is equipped for the task.

The workflow is illustrated as follows:

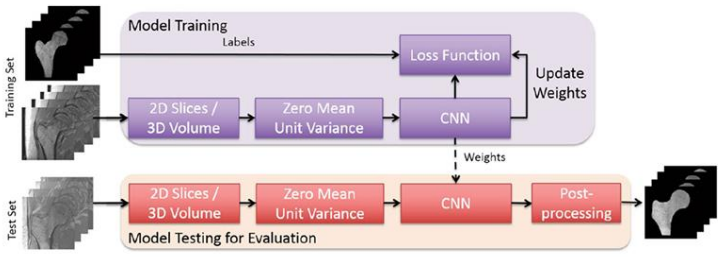


Figure 1 Workflow of Deniz et al. 's pipeline [1].

As for an illustration, the network architecture of 3D CNN used in this paper is shown in Appendix Figure A. Five evaluation metrics are introduced for the task, with each defined as:

$$Dice\ Score(DSC) = 2TP/(FP + 2TP + FN)$$

$$\text{sensitivity/recall } = TP/(TP + FN)$$
$$\text{precision } (PPV) = TP/(TP + FP)$$

And with $d(v, S) = \min_{s \in S} \| v - s \|$, S: segment surface; G: ground surface:

$$ASD = \frac{1}{N_S + N_G}\left(\sum_{x_S \varepsilon S} d(x_S, G) + \sum_{x_G \varepsilon G} d(x_G, S)\right)$$

$$MSD = max\left\{\max_{x_S \varepsilon S} d(x_S, G), \max_{x_G \varepsilon G} d(x_G, S)\right\}$$

Assessment

For the pros and significances of this work, the model finally obtained a dice score of 0.953 ± 0.016 and an ASD of 0.39 ± 0.20mm. With considerable accuracy, the time spent on femur segmentation was reduced from several hours for manual segmentation to 5 seconds for automatic segmentation, which has crucial implicational potential for the clinical diagnosis and treatment of osteonecrosis.

At the same time, Deniz et al. focused on using different details based on the original design of U-Net to compare the models horizontally, which summarizes the performance of various femur automatic segmentation models based on CNN to date. It evaluates the models from five different evaluation metrics, which are contributive work and systemic guidance for similar tasks.

On the other hand, cons are discovered. Deniz et al. 's work does not evaluate more excellent U-Net variants, which have already become the new baseline for medical segmentation models. For example, the U-Net which uses Residual blocks to replace conventional convolutional layers has outperformed the original U-Net on multiple tasks and is not considered by Deniz et al.

Therefore, the follow-up of this work, including those that might be realized in my project, mainly focuses on using more efficient U-Net models, including using residual blocks to replace traditional convolutional layers and using vision transformers to replace the convolutional down-sampling stage in U-Net's original design.

This work is highly related to my projects. As research related to the diagnosis and treatment of osteonecrosis, my project will use the 3D CNN segmentation method. The result of Deniz et al. also provided an essential reference for model selection, evaluation, and parameter adjustment.

Conclusion

In conclusion, for the task of proximal femur segmentation, Deniz et al. use 2D or 3D U-Net with different network topologies to realize the automatic segmentation of the femur and, at the same time, compares the potential of these other models for this task. Deniz's work provides constructive advice on model selection, parameter tuning, model evaluation, etc., for segmentation work related to osteonecrosis.

Paper II: Fully Automatic and Fast Segmentation of the Femur Bone from 3D-CT Images with No Shape Prior [2].

Summary of problem

Segmentation of CT is a core topic in orthopedics since it is essential in intraoperative guidance, reconstructive surgery, etc. [2]. Due to the high contrast of bone anatomy and soft tissues in CT volumes, the segmentation of CT is well developed. However, challenges still exist in segmenting the joint epiphysis areas [2].

Key Results

Krcah et al.'s work outperforms the previous approaches with a result precision (True Positive rate > 0.85, False Positive rate < 0.001, and Hausdorff Distance smaller than 8mm) in 81% of the tested cases. A figure

summarizes the results are provided in Appendix Figure B.

Significance of key results

Krcah et al.'s work outperformed the existing work on automatic CT segmentation, significantly reducing the False-Positive rate and Hausdorff distance. The model is fast to process but has a higher consumption of memory. Krcah et al.'s model is well generalized to segment any other bone anatomies in CT volumes.

Background

Graph-cut framework segments images by minimizing energy function [5]. Generally, such function usually includes two terms: a regional term ($R(A)$) and a boundary term ($B(A)$) [5]:

$$E(A) = \lambda \cdot R(A) + B(A)$$

$$where \quad R(A) = \sum_{p \in \mathcal{P}} R_p(A_p)$$

$$B(A) = \sum_{\{p,q\} \in \mathcal{N}} B_{p,q} \cdot \delta_{A_p \neq A_q}$$

The regional term quantifies the penalties of assigning a pixel I_P to label x. A common definition is [5]:

$$R_p(\text{ "obj" }) = -\ln \Pr\left(I_p \mid \text{ "obj" }\right)$$

$$R_p(\text{ "bkg" }) = -\ln \Pr\left(I_p \mid \text{"bkg"}\right)$$

Where two labels "obj" and "bkg" indicates objective and background, respectively.
The boundary term penalizes on the discontinuities of boundaries by assigning a large value to $B_{p,q}$ if p and q are similar, and vice versa.

Methodologies and Experiments

Krcah et al's work is based on the graph-cut framework, with self-defined thresholds and modified energy functions. All images are firstly augmented with a unsharp masking mask that: $\mathcal{I}^U = \mathcal{I} + k(\mathcal{I} - \mathcal{I} \star G_s)$, $\star$ denotes convolution and G_s is a Gaussian kernel with s^2 as the variance, k is a constant.

Then, Krcah et al. define the energy function as [2]:

$$E(A) = \sum_{p \in \Omega} R_p(A_p) + \lambda \sum_{(p,q) \in \mathcal{N}} \delta(A_p, A_q) B(p,q)$$

With boundary term that:

$$B(p,q) \propto \begin{cases} \exp\left\{-\dfrac{|S(p) - S(q)|}{\sigma_s}\right\} & \text{for } S(p) \geq S(q) \\ 1 & \text{otherwise} \end{cases}$$

Where:

$$S_\sigma(x) = -\text{sgn}\,(\lambda_3)\exp\left\{-\frac{R_{\text{sheet}}^2}{\alpha^2}\right\}\exp\left\{-\frac{R_{\text{tube}}^2}{\beta^2}\right\}\left(1 - \exp\left\{-\frac{R_{\text{noise}}^2}{\gamma^2}\right\}\right)$$

With $R_{\text{sheet}} = |\lambda_2|/|\lambda_3|, R_{\text{tube}} = |\lambda_1|/(|\lambda_2||\lambda_3|), R_{\text{noise}} = (|\lambda_1| + |\lambda_2| + |\lambda_3|)/T$. T denotes the average trace of Hessian at each voxel. λ_1, λ_2 and λ_3 are the eigenvalues of Hessian matrix and $|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$.

And the per-pixel (regional) term that:

$$R_p(A_p) \propto \begin{cases} 1 & \text{if } A_p = \text{ "bone" and } p \in E_{\neg \text{ bone}} \\ 1 & \text{if } A_p = \text{ "bkg" and } p \in E_{\neg bkg} \\ 0 & \text{otherwise.} \end{cases}$$

Where:

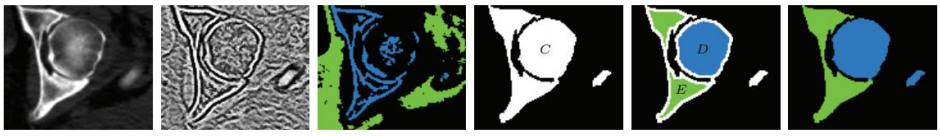
$$E_{\neg bkg} = \{x \in \Omega \mid \mathcal{I}(x) \geq 400\text{HU} \wedge S(x) > 0\}$$
$$E_{\neg bone} = \text{lcc}\left(\{x \in \Omega \mid \mathcal{I}(x) < -50\text{HU}\}\right)$$

Moreover, in order to separate the bones, again with the graph-cut framework, Krcah et al. introduced a postprocess that:

$$\forall p \in C: R_p(A_p) = \begin{cases} \infty & \text{if } A_p = \text{"D" and } p \in E \\ \infty & \text{if } A_p = \text{"E" and } p \in D \\ 0 & \text{otherwise.} \end{cases}$$

And boundary term that: $B(p,q) = B(q,p) = 1$.

Based on the proposed functions, Krcah et al.'s work can be defined to the stages illustrated as follows:



(a) Input (b) Bone boundaries enhanced. (c) Initialization of bone and background exclusion regions. (d) All bone voxels segmented. (e) Bones identified with morphological erosion. (f) Final result

Figure 2 illustrates the segmentation pipeline of the work from Krcah et al.

97 CT volumes which has been cropped around the femur are used for experiment. Three evaluation metrics are used to describe the performance: i). True-Positive rate: $\text{TPR} = \text{TP}/(\text{TP} + \text{FN})$, ii). False-Positive rate: $\text{FPR} = \text{FP}/(\text{FP} + \text{TN})$, and iii). Hausdorff distance between prediction and ground truth.

Assessment

For the pros and significance, Krcah et al.'s work achieves automatic segmentation of femurs without prior knowledge (given model). Using the graph-cut architecture and the threshold based on the empirical data of CT segmentation, the algorithm can efficiently segment and separate the CT bone structure. Algorithms outperform previous algorithms on multiple evaluation metrics.

Cons still exist. A common defect of such algorithms that cannot perform multi-label segmentation and provide semantic information. For CT volumes containing multiple bone structures, Krcah et al.'s work can only provide binary segmentation results such as bone and background. However, for the diagnosis and treatment of ONFH, we need to segment and distinguish various bone structures, including femur and pelvis.

In response to this problem, a classifier can be introduced as its successor pipeline to solve this problem. Another more subversive way is to achieve pixel-level segmentation and classification at the same time through deep learning, i.e., semantic segmentation. The Krcah et al. algorithm will play a key role in the annotation: CT volume has dozens of times the volume of MRI, and it will be even more time-consuming to complete manual annotation. The algorithm of Krcah et al. will help expertise to do most of the work.

The work of Krcah et al. provides a good baseline model for the CT segmentation part of my project, which can be used as part of the pipeline to achieve automatic segmentations and semi-automatic registration. At the same time, it will help me immensely when (if needed) developing deep learning based solutions.

Conclusion

In conclusion, the work of Krcah et al. achieves a femur automatic segmentation on CT volume that is not based on prior knowledge. The algorithm is based on the graph-cut framework and incorporates empirical parameters for CT segmentation. Their model yields performance on multiple evaluation metrics due to previous algorithms. The algorithm can segment CT volume well, but cannot provide classification or semantic information. It provides a good baseline for the segmentation of CT in my project.

Reference

[1]. Deniz, C.M., Xiang, S., Hallyburton, R.S. et al. Segmentation of the Proximal Femur from MR Images using Deep Convolutional Neural Networks. *Sci Rep* **8**, 16485 (2018). https://doi.org/10.1038/s41598-018-34817-6

[2]. Krcah, Marcel et al. “Fully automatic and fast segmentation of the femur bone from 3D-CT images with no shape prior.” *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (2011): 2087-2090.

[3]. Yamashita, R., Nishio, M., Do, R.K.G. et al. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* **9**, 611–629 (2018). https://doi.org/10.1007/s13244-018-0639-9

[4]. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical image computing and computer-assisted intervention.* Springer, Cham, 2015.

[5]. Boykov, Y., Funka-Lea, G. Graph Cuts and Efficient N-D Image Segmentation. *Int J Comput Vision* **70**, 109–131 (2006). https://doi.org/10.1007/s11263-006-7934-5

Appendix

Table A Deniz et al. experimented with ten different network architectures [1].

Network	DSC ↑	Precision ↑	Recall ↑	ASD [mm] ↓	MSD [mm] ↓
2D CNN*, F:64, L:3	0.886 ± 0.055	0.890 ± 0.080	0.889 ± 0.056	6.15 ± 3.61	65.78 ± 5.78
2D CNN*, F:64, L:4	0.864 ± 0.044	0.872 ± 0.061	0.860 ± 0.060	6.82 ± 3.00	64.89 ± 6.36
2D CNN, F:64, L:3	0.924 ± 0.032	0.920 ± 0.041	0.930 ± 0.045	3.13 ± 1.76	54.40 ± 6.72
2D CNN, F:64, L:4	0.937 ± 0.026	0.932 ± 0.037	0.943 ± 0.036	2.13 ± 1.23	42.22 ± 5.52
2D CNN-dilated †, F:64, L:4	0.946 ± 0.022	0.948 ± 0.024	0.944 ± 0.034	1.75 ± 1.24	40.03 ± 8.37
3D CNN, F:16, L:3	0.927 ± 0.032	0.931 ± 0.029	0.927 ± 0.063	0.66 ± 0.32	10.62 ± 6.85
3D CNN, F:16, L:4	0.935 ± 0.028	0.938 ± 0.026	0.936 ± 0.053	0.59 ± 0.39	9.75 ± 6.56
3D CNN, F:32, L:3	0.942 ± 0.026	0.944 ± 0.022	0.942 ± 0.052	0.50 ± 0.25	11.97 ± 7.57
3D CNN, F:32, L:4	0.945 ± 0.029	0.948 ± 0.023	0.944 ± 0.052	0.45 ± 0.25	13.44 ± 13.14
3D CNN-dilated†, F:32, L:4	**0.953 ± 0.016**	**0.954 ± 0.017**	**0.953 ± 0.030**	**0.39 ± 0.20**	**7.88 ± 4.33**

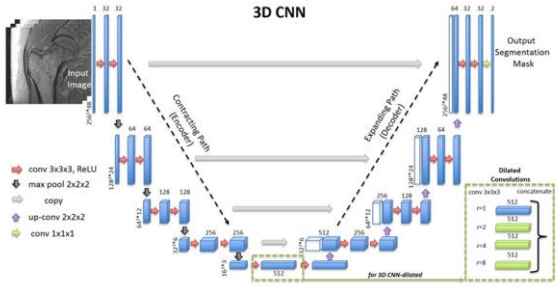


Figure A Illustration of one network structure [1].

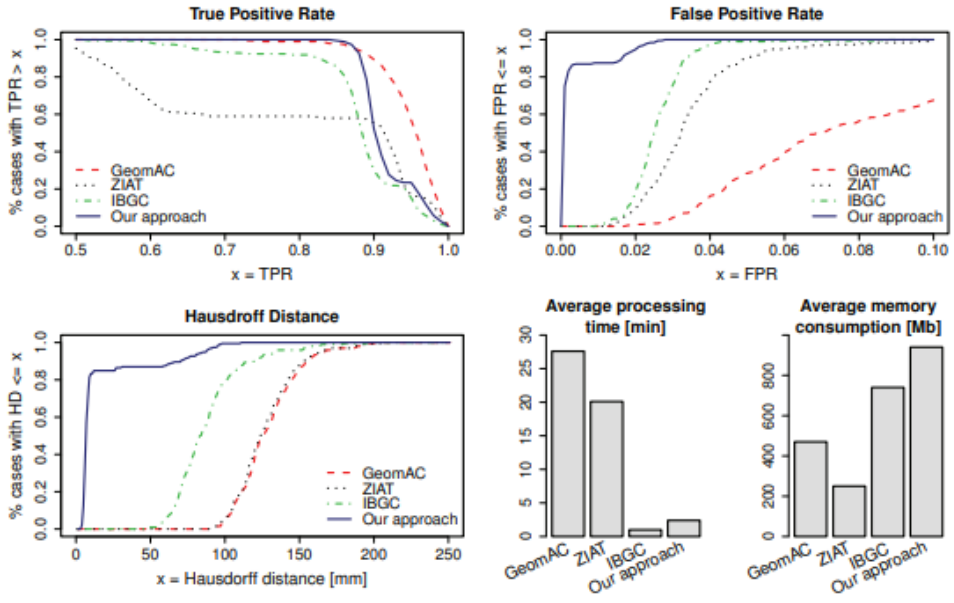


Figure B Krcah et al.'s work compared to previous contributive work.