Finger Joint Segmentation Using Machine Learning and Minimized Training Set

Author List

Yue Wang
Ming Zhang
Tino Cheung
Carmine Guida
Raymond Ren
Juan Shan

Background/Purpose

Hand Osteoarthritis (HOA) is a common form of OA disease. About half of all women and one-quarter of all men will experience the stiffness and pain of HOA by the time they are 85 years old. The 29 bones of hands and wrists come together to form many small joints that can be affected by HOA. One common HOA evaluation is to measure joint space width (JSW) on hand X-ray for each finger joint. This measurement usually requires manual labeling of finger joint bone edge, which is time-consuming. As machine learning gains its popularity in medical image segmentation, it could provide a new approach to tackle the problem of JSW measurement. However, a large training set with manual labeling is necessary to thoroughly train a machine learning model. To minimize the effort of manual delineation, we developed a novel iterative training strategy that requires only a small number of images to be manually delineated for the finger joint segmentation task.

Methods

In this study, we focused on Distal Interphalangeal (DIP) joints of four fingers (excluding the thumb). We will extend our study to Proximal Interphalangeal（PIP) and Metacarpophalangeal (MCP) joints shortly. 3557 hand X-ray images from the Osteoarthritis Initiative (OAI) were used in this study. 1561 DIP joints in square regions of interest (ROIs) were cut from some of the X-ray images, rotated to be aligned with the same direction, and normalized to the same image size 180 x 180.

We used an iterative training procedure to minimize the manual labeling. In round 1, we manually labeled the bone outline of 110 joints from the 1561 ROIs. Of those 110 manually labeled joint images, 90 were used for training samples and 20 for validation samples. These joints were used as the samples to train the first deep learning model (model 1), with U-net model as the backbone. In round 2, we applied model 1 to 500 joint images. These 500 joint images had not been shown to the model as training or validation samples. From the segmentation results of these 500 images, we selected 130 failure cases with the most inaccurate segmentation and manually segmented those 130 images. Then we used these 130 images plus the original labeled 110 images to train a new U-net model (model 2).

Results

To validate the performance of the final model (model2), we used another 951 PIP joint images as testing data. Table 1 lists the accuracy of our model on the testing set for different OA severities (Kellgren Lawrence (KL) grades 0-4). A joint segmentation is considered a success if the space between the top and bottom bones is shown in the same way as it is shown in the original X-ray image; otherwise, it is a failure. The accuracy is computed by dividing the number of success cases by the number of total cases. For joints with KL = 0, the model achieves an accuracy of 97.0%. The accuracy for KL = 1 and 2 are 78.5% and 77.0%, respectively. For more severe stages KL = 3 and 4, the accuracy increase again to 87.0% and 86.1%, since the bones start touching each other and the space between them is disappearing.

The average accuracy for all stages is 85.1%. Figure 1 shows a few examples of the model's segmentation results.

Conclusions

Acquiring a large medical image dataset is a challenge in many studies and manually labeling is often a labor-intensive procedure. In this work, we proposed an iterative strategy to minimize the number of training samples that need to be manually labeled. In a dataset of 1561 DIP joint images, we only manually delineated 240 images and our model could achieve 85.1% overall accuracy on the testing set with 951 DIP joints. This training set minimizing strategy has potential usage in other medical image processing problems to save the manual labeling effort (e.g., knee cartilage segmentation, bone marrow lesion segmentation, etc.). Our future work includes increasing the training samples from KL = 1 and 2 stages to improve the accuracy and extending the work to PIP and MCP joints for the whole dataset. The obtained U-net model(s) can serve as the first step for automatic JSW measurement which is the ultimate goal.

Table 1. Segmentation accuracy of the U-net model on the testing set

	Total cases	Failure cases	Accuracy
KL = 0	200	6	97.0%
KL = 1	200	43	78.5%
KL = 2	200	46	77.0%
KL = 3	200	26	87.0%
KL = 4	151	21	86.1%
Total	951	142	85.1%

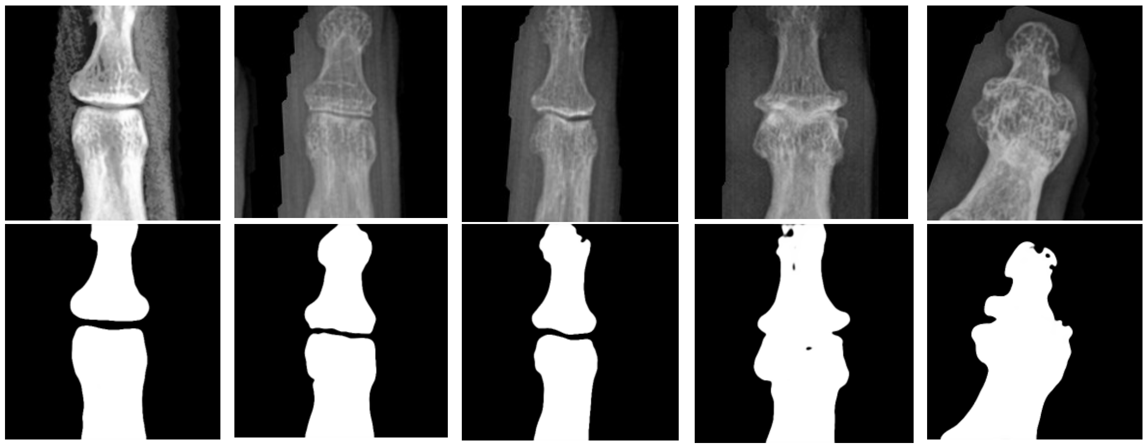

Figure 1. Joint ROI (top) and the corresponding segmentation result from the model (bottom) for each KL grade (0-4).