# Automatic Detection of Knee Joints in Radiographic Images

Deep Patel
Senior Thesis
Undergraduate Biomedical Engineering Student
University of Florida

Lin Yang, Ph.D.
Principal Investigator, Biomedical Image Computing and Imaging Informatics Lab
Department of Biomedical Engineering
University of Florida

Thesis Supervisory Committee
Ruogu Fang, Ph.D.
Kyle Allen, Ph.D.
Alina Zare, Ph.D.

Oral defense date: April 3rd, 2018

# Abstract

Knee osteoarthritis (OA) is a highly prevalent, disabling joint disease affecting people worldwide. It is currently diagnosed by evaluating symptoms as well as plain radiographic images. However, existing diagnostic methods suffer from subjective assessment, hindering early diagnosis of the disease to enable effective treatment. A plausible approach is the use of computer-aided image analysis to provide an objective evaluation of knee OA severity. A critical step in this approach involves detection of knee joint region to enable higher-level analysis, a task with significant room for improvement. In this preliminary study, the use of transfer learning to automatically detect knee joint is explored and YOLOv2, a state-of-the-art real-time object detection system, is trained to perform the task of detecting knee joint in X-ray images. The model is trained and tested using data from Osteoarthritis Initiative (OAI), a multi-center, observational study of knee osteoarthritis in men and women between the age of 45 and 79. Once trained, the proposed model accurately detected and localized knee joint region with mean intersection-over-union of 0.907 and an average time of 0.0153 seconds to analyze a single high-resolution radiographic image. This model can be used in OA research as many of the existing studies attempting to quantify knee OA severity rely on knee joint area detection. Furthermore, the pipeline can potentially be extended to automatically evaluate knee joint features during the early stages of the disease to predict knee OA progression. Such tool for an objective assessment of knee OA can support clinicians in their decision-making.


# Introduction

Osteoarthritis (OA), also known as degenerative joint disease, is the most prevalent form of arthritis. It is a chronic disease characterized by pain, loss of articular cartilage, remodeling of subchondral bone (any bone distal to calcified cartilage), osteophyte formation, and synovial inflammation [1]. It is a leading cause of global disability associated with significant societal and economic burden [1] [2]. Typically, OA affects older individuals, however, recent increasing trend in sports injuries raise concerns of efflux in OA incidence among younger individuals [3]. Moreover, OA is expected to pose a major problem for the health systems globally due to the combination of an aging population and increase in obesity throughout the world [2]. While OA affects several different joints including those of hands and hip, OA of knee contributes the most to OA burden accounting for more than 80% of the disease's total burden [4]. It is a slowly progressing disease with symptoms including pain, stiffness, and joint deformities. Due to the lack of an effective cure for OA, the disease can eventually lead to joint failure requiring joint replacement surgery, especially for a weight bearing joint such as knee.

Knee OA is commonly categorized as either clinical or radiographic [5]. Clinical knee OA diagnosis is primarily based on symptoms, such as pain and stiffness, whereas, radiographic diagnosis is based on the structural features of the joint, including formation of osteophytes, sclerosis, and joint space loss [1] [5]. Clinical knee OA is commonly assessed using Western Ontario and McMaster Universities Osteoarthritis Index, a standardized questionnaire to evaluate pain, stiffness, and physical functioning of the joints [6]. Radiographic knee OA is typically evaluated using Kellgren and Lawrence (K&L) classification system [6]. Studies suggest that clinical and radiographic knee OA are rather weakly associated [5]. That is, the extent of radiographic change in individuals with clinical knee OA varies considerably, and individuals with radiographic evidence of knee OA often experience no symptoms. Consequently, there may be many individuals with subclinical knee OA as screening for the disease is often not prescribed until an individual reports severe symptoms. The discordance between clinical and radiographic knee OA is in part due to the significant variation in pain definition, OA grading, and radiographic image view [5] [6].

Current treatments for knee OA are cost-intensive and deal primarily with symptom relief and mitigating the effects of joint disabilities, instead of targeting the disease progression [7] [8]. This is in part due to the unclear understanding of the underlying causes of the disease, which has motivated the efforts to abate or, possibly, reverse the structural degradation of joints to avert the incidence of advanced knee OA and alleviate socio-economic burden [8] [9]. These efforts have led to the development of several disease-modifying therapies for both clinical and radiographic knee OA [9]. However, in order for these therapies to be effective, it is crucial to diagnose the disease at an early stage and intervene with appropriate therapy. While it is clinically possible to diagnose knee OA at an early stage, the progression of the disease in a particular individual is difficult to predict, hindering health care providers for creating an effective long-term management plan.

The most widely used diagnostic tool in identifying knee OA in patients and making clinical decision is plain X-ray imaging. While clinicians may assess the severity of pain using standardized questionnaire or make use of magnetic resonance imaging (MRI) for a better diagnosis of the disease, plain film X-rays remain the most useful tool in the evaluation of knee OA due to their accessibility, cost-effectiveness, and speed [10]. More importantly, they provide critical information, including osteophytes formation, joint misalignment, and joint space narrowing (JSN), which are good indications of knee OA [10] [11]. However, such changes present themselves in more advanced stages of the disease. In addition, plain film X-ray images provide an insensitive measure of knee OA, since a more effective measure of disease is the deterioration of articular cartilage and X-ray imaging is limited in its ability to directly visualize non-osseous tissues [11]. Furthermore, interpreting the images to extract useful information requires a clinician with

considerable experience, which at best is still subjective evaluation of the disease. These limitations of current approaches in the diagnosis of knee OA pose challenges to effective early diagnosis of the disease and necessitate an objective mean of evaluating knee OA to help clinicians in their decision-making.

A reasonable approach is the use of computer-aided image analysis to provide support to practitioners in their decisions. Aside from the lack of clinical bias, the major advantage of this method is the ability to detect subtle changes in image properties. It could potentially allow for the identification of the presence of knee OA in early stages of the disease when radiographic changes are minimal and, possibly, make accurate predictions concerning the disease progression. Indeed, such methods have gained increasing attention in recent years to analyze changes in the structure of knee joint for early diagnosis of OA [12] [13] [14]. In particular, the use of convolutional neural networks (CNN) has demonstrated innovative results in a variety of tasks associated with the diagnosis of knee osteoarthritis [15] [16]. An essential step in computer-aided knee OA evaluation involves automatic detection of knee joint from X-ray images to serve as an input for higher level analysis. The accuracy of results at this step has a considerable impact on subsequent analysis. Studies that have used detection of knee joint to extract appropriate region of interest from X-ray images prior to quantifying knee OA severity have fared better compared to those that have attempted to use whole images [15] [16] [17]. While previous studies have yielded promising results in automatically detecting knee joints from radiographic images, there is significant room for improvement.

# Methods

This work investigates the use of transfer learning approach in automatically detecting knee joint in X-ray images and compares the results with other existing approaches, including fully convolutional networks (FCN) and support vector machine (SVM). The primary contribution of this work is a robust method which maximizes the amount of image data to be transferred onto the subsequent higher-level analysis for computer-aided diagnosis of knee OA. This is important since the amount of data available to be used in these studies is limited and deep learning approaches such as CNN require a large amount of data to perform well.

### Data

Data used in this work is from Osteoarthritis Initiative (OAI), a multi-center, prospective observational study of knee osteoarthritis in men and women aimed at developing research resources for use in scientific evaluation of biomarkers for OA [18]. The study followed the participants, aged 45-79 from February 2004 to May 2006, to record changes

in the status of their knees. The radiographic knee image data used in this investigation consists of a total 4,130 images containing the left and the right knees of individuals who either have or are at high risk of developing knee osteoarthritis. The view used for obtaining the X-ray images is bilateral posteroanterior fixed flexion. In training and testing the model, only the images collected at baseline are used due to the interest in utilizing the same dataset for predicting disease progression in future investigations.

**Preprocessing Raw Images**

The images, in their raw state, vary in their physical resolution and dimensions since they were produced at different study centers. These inconsistencies can result in poor model performance, thus necessitating the need for pre-processing prior to training the neural network. Consequently, all images were resized to ensure the same physical resolution. First, the physical resolution of all images in the dataset was determined and a value of 0.14 mm/pixel, which is close to the median physical resolution of the images in the dataset, was chosen as the final resolution for all images. Next, the images were cropped to same size in order to facilitate the subsequent processing. In order to make certain that both knee joints, for left and right leg, are contained in the image, the height and the width of the cropped image was chosen to be 2048 pixels by 2560 pixels.

Next, the preprocessed images were labeled with a bounding box containing that region of interest (ROI) to generate the ground-truth for the dataset. Figure 1 shows the region of interest chosen to be detected by the algorithm. The region was chosen to ensure that it captures the features of interest considered in assessing knee OA, which are mainly in the knee joint region. It is important to note that the size of the bounding box was intentionally chosen to exclude the left and right edges of the knee joint, which can potentially interfere with the detection of osteophytes, due to interest in automatic computation of joint-space width in works to follow. Also, the process of resizing the bounding box can be automated to achieve bounding boxes of desired height and width, if necessary. The images in the prepossessed dataset were randomly split for training and testing with 4:1 ratio. From the training set, 20% of the images were used for validation. Table 1 gives a detailed description of the different sets of images used in this study.

There are several deep neural network architectures currently available for object detection, including Faster R-CNN [19], SSD [20], YOLOv2 [21] and Mask R-CNN [22], which have demonstrated good performance in natural object detection tasks. In this preliminary work, YOLOv2 architecture is customized for detecting knee joint in radiographic images. It is an improved version of YOLO, a state-of-the-art object detection system which frames object detection as a regression problem to spatially separate

bounding boxes and the associated class probabilities [23]. It was chosen over the other object detection architectures for its simplicity as well as its flexibility in designing the bounding boxes.

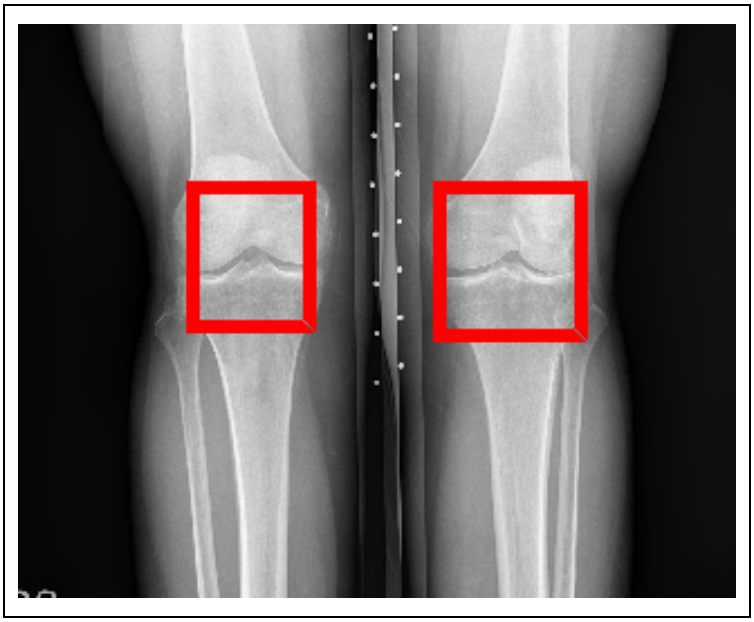

**Figure 1:** Knee X-ray with the region of interest, knee joint, enclosed in bounding box.

| Group | Images |
|---|---|
| Training | 3,302 |
| Validation | 660 |
| Testing | 828 |

**Table 1:** Description of the different set of images used in this study. The values in the table indicate the number of images used in each group. Each image contains two knee joints. The images in training set were used to train the model, those in validation set were used to adjust hyperparameters of the model, and those in testing set were used to evaluate model performance.

### Detection of Knee Joint Region

In general, YOLO runs a single convolutional network on an image and predicts multiple bounding boxes and probability of those boxes containing a specific class of object [23]. First, the algorithm divides the input image into an *M x N* grid based on the size of the input image, and each of these grid cells predicts *B* bounding boxes. Each bounding box

is defined by height, width, and center coordinates. The height and the width are predicted relative to the entire image, whereas, the center coordinates are predicted relative to the bounds of the grid cell. Additionally, the grid cells predict confidence scores for each bounding box to provide a quantitative measure of the model's certainty that the predicted bounding box contains an object as well as the accuracy of the box size in enclosing the object. In the context of this work, the confidence score is defined as

$$P(ROI) * IOU_{predicted}^{truth} \quad (1)$$

where $P(ROI)$ is the probability that the region of interest exists in that cell and $IOU_{predicted}^{truth}$ is intersection-over-union (IoU) between the predicted bounding box and the ground truth box. If the region of interest is not in a box, then the confidence score for that box is very close to zero. Otherwise, the confidence score is equal to the IoU between the predicted and the ground-truth box.

For each bounding box, the grid cells also predict one set of class probabilities per grid cell conditioned on the grid cell containing an object [23]. This study only considers two class, namely knee joint and not knee joint, hence the model predicts two class probabilities. In the context of this work, the confidence score for the bounding box and the class prediction are combined as follows,

$$P(Knee\ joint|ROI) * P(ROI) * IOU_{predcited}^{truth} = P(Knee\ joint) * IOU_{predicted}^{truth} \quad (2)$$

to obtain a final score that specifies both the probability of knee joint, $P(Knee\ joint)$, appearing in a specific box as well as how well the predicted box encloses the knee joint.

YOLOv2 addresses various shortcomings of YOLO to improve the recall and localization accuracy of model [21]. The use of batch normalization, anchor boxes to predict bounding boxes, and direct location prediction of center coordinates of the bounding boxes are three of the improvements adopted in detecting the knee joint region. Batch normalization refers to normalization of data for each training mini-batch, or the subset of images from the training dataset provided as input during a single pass through the network. It is a component of neural networks that operates between the layers of a network and addresses the challenges of a phenomenon called internal covariate shift [24]. Internal covariate shift, defined as changes in the distribution of data being inputted to each layer during training, occurs as the data passes through the network and the weights are updated [24]. Specifically, the inputs and the outputs of the intermediate layers of the network gradually deviate from zero mean, unit variance, and decorrelation conditions. This breaks the identical and independently distributed assumption used for the distribution of data features in various parts of training and testing the network. Batch

normalization maintains this assumption by taking the output of a particular layer and normalizing it before passing it onto the subsequent layer. This allows for data comparison across features as data passes through the layers of the network. There are several advantages to using batch normalization. It eliminates the need for careful initialization of weights and allows for much higher learning rate, hence faster convergence and less time to obtain a trained model [24]. In addition, it has regularizing effects on the model, discouraging complexity of the model which may compromise its ability to accurately detect knee joint in the images.

Another improvement in YOLOv2 that was incorporate in this investigation is the use of anchor boxes. Anchor boxes function as base-shape of objects, helping to identify objects of different shapes by matching the relative height-to-width ratio of a particular object class being detected [19]. With this approach, the model predicts the bounding boxes with prior size through dimension clustering. In this work, *k*-means clustering is applied to cluster knee size in the training dataset to generate three prior anchor boxes. Unlike some of the existing anchor box based object detection architectures [20], YOLOv2 does not assume the shape of the boxes. Instead, it allows for the anchor boxes to be designed particularly for a given dataset. Also, it uses IoU as the distance metric, which minimizes the localization loss and results in higher IoU between the predicted and ground truth box [21]. The third improved feature in YOLOv2 adopted for this inquiry is direct location prediction of the center coordinates of the bounding boxes. This approach predicts the location of center coordinates relative to the grid cell, restraining the object center to be located within the preset grid cells. This constraint on the center location prediction prevents the center coordinates of the bounding boxes from ending up too far from the grid cell that predicted the box, improving model stability [21].

The preprocessed knee X-ray images of size 2048x2560 pixels were resized because they were too large to be feed directly to YOLOv2 model. All images in the training dataset were resized to be 256x320 pixels and the knee joint annotation bounding boxes were resized accordingly. The grid size for the input images was determined by the model to be 8x10 based the size of the input image. The model was trained using a NVIDIA GTX 1080 graphics processing unit (GPU). A batch size of 8 was selected and stochastic gradient descent was used to optimize the weights with a momentum of 0.9 and a starting learning rate of 2.0E-4. The network was trained for 160 epochs with the learning rate decaying in the $60^{th}$ and $120^{th}$ epochs by a factor of 0.8. The aforementioned parameters were adjusted and selected by running the model numerous times on the validation set following training. In testing the model, bounding boxes with confidence score smaller than 0.12 were removed. Following, non-maximum suppression, which merges all detections belonging to the same object, was applied with an intersection-over-union threshold of 0.7 to suppress the remaining ones. Of those still remaining, the bounding

box with the highest confidence score was deemed the final detection result. Ultimately, the detected bounding box was mapped to a higher resolution for evaluation of detection accuracy.

## Results

The following are results obtained after training YOLOv2 model and testing it on the test dataset. Also presented are the performance results of two state-of-the-art methods, fully convolutional network (FCN) and support vector machine (SVM), for detecting knee joints in plain X-ray images.

| Method | $\mathrm{FNR}_{J \geq 0.75}$ | $\mathrm{Recall}_{J \geq 0.75}$ | Mean IoU | Average Time [s] |
|---|---|---|---|---|
| FCN [16] | - | 0.892 | 0.830 | - |
| HOG-SVM [15] | - | - | 0.840 | 0.169 |
| Proposed | 0.006 | 0.994 | 0.907 | 0.0153 |

**Table 2:** Knee joint region detection performance of two state-of-the- art methods and the proposed model based on YOLOv2 architecture.

In Table 2, the false negative rate (FNR) indicates the probability of the model not detecting a knee joint when it is present in an image. Recall indicates the probability of the model correctly detecting knee joints in an image. Mean IoU specifies the average intersection-over-union between the bounding box predicted by the model and the ground-truth. Average time denotes the time taken by the different models to analyze a single image on average during testing. The model based on FCN approach was trained and tested using two different datasets, Osteoarthritis Initiative and Multicenter Osteoarthritis Study [17]. The model based on SVM approach was trained, validated, and tested using images from Multicenter Osteoarthritis Study [16]. The different studies considered in this work did not use all the metrics used to evaluate the proposed model, as such, not all values are available for comparison of performance.

Figure 2 displays some example images demonstrating the performance of the proposed model in detecting knee joint region. In the image, the red box represents the ground-truth bounding box enclosing the region of interest, and the green box represents the region detected by the model as knee joint. The number above the boxes indicates the IoU value for the two boxes. The radiographic images in the dataset used in this study varied in characteristic, such as contrast and intensity. In addition, some of the patients

had implants. The sample images shown in Figure 2 represent the different varieties of images in the dataset and the performance of the proposed model on those images.

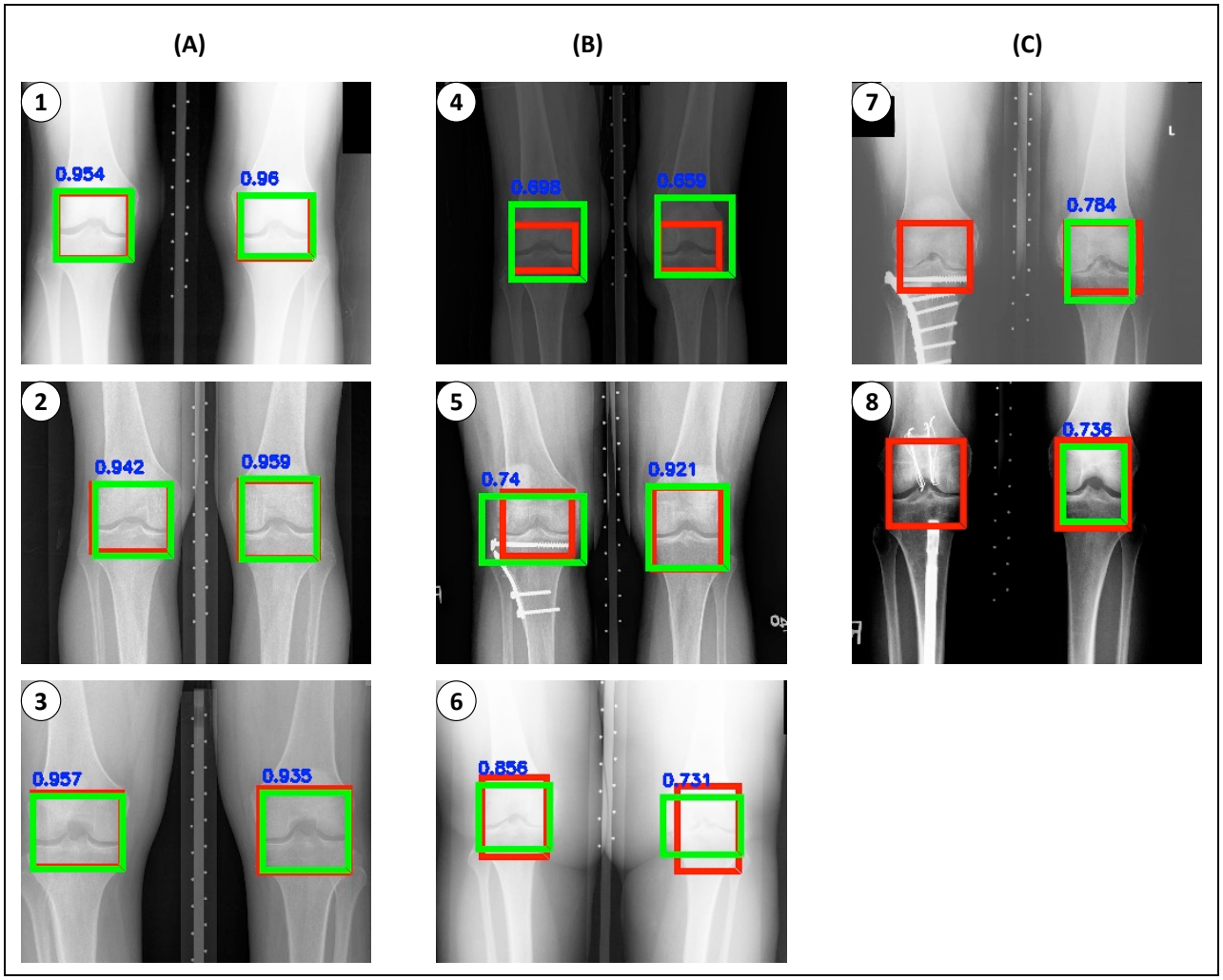


**Figure 2:** Example images demonstrating the preformace of the proposed model. Red box represents the ground truth containig the region of interest, and the green box indicates the region detected by the model as knee joint. The number above the boxes indicates the IoU value.

## Discussion

The standard metric for evaluating the performance of an object detection model is intersection-over-union (IoU), also known as Jaccard Index [25]. It is defined as the size of the intersection divided by the union of two regions. In this case, it measures the similarity between the predicted bounding box and the ground-truth bounding box. In order to assess the performance of the trained model in detecting knee joint, the following four metrics are employed: (1) false negative rate (FNR) with Jaccard index greater than 0.75, (2) recall with Jaccard index greater than 0.75, mean IoU, and time expense. A

comparison between the performance of the proposed method and two state-of-the-art methods are shown in Table 2. Based on the data presented in the table, it is evident that the proposed model based on YOLOv2 framework has the best performance in terms of the considered evaluation metrics. While the two methods being compared with the proposed model did not use the same datasets or the same number of images, all three approaches randomly divided their respective datasets into training, validation, and testing set. As such, it is reasonable to compare the three approaches. In absolute terms, the trained model was able to detect 1654 knees out of 1656 knees, of which 1646 had an IoU value greater than 0.75 in the testing dataset. The set of X-ray knee images for which the detected knee joint areas with Jaccard index is greater than 0.75, the mean IoU was 0.907. This high mean IoU indicates that the knee joint region detected by the model is very close to the actual knee joint region.

While the proposed model preforms well in detecting knee joints in radiographic images, it has a few limitations. Figure 2 displays several example images resulting from this experiment. Images shown in column A are typical among the training dataset, hence the model is able to accurately detect knee joints in such images with high a Jaccard index. Columns B and C present the types of images for which the model is unable to accurately detect knee joint. It is due to low image contrast or knee implants. Such images are not represented very well in the training dataset, therefore the model treats them as outlier cases. It may be possible to train the model with a large number of low contrast images to further improve its detection accuracy.

## Future Work

This preliminary investigation demonstrated a transfer learning approach in detection of knee joints in radiographic images. The proposed model yielded promising results and improved detection accuracy relative to previous works. However, further work is necessary to enable computer-aided diagnostic of knee OA by quantifying the disease severity using features present in the knee joint region detected in the X-ray images. Potential features to quantify the severity of knee OA include joint-space reduction, osteophytes formation, and sclerosis. The model presented in this work was specifically designed to isolate the knee joint region to enable automatic computation of joint-space width. Thus, future work on this project will deal with constructing a tool that can objectively measure joint-space width and, possibly, predict changes in it to truly enable clinical utility of computer-aided diagnosis for knee OA.

## References


[1] Litwic, A., Edwards, M., Dennison, E. and Cooper, C. (2013). Epidemiology and burden of osteoarthritis. *British Medical Bulletin*, 105(1), pp.185-199.

[2] Cross, M., Smith, E., Hoy, D., Nolte, S., Ackerman, I., Fransen, M., Bridgett, L., Williams, S., Guillemin, F., Hill, C., Laslett, L., Jones, G., Cicuttini, F., Osborne, R., Vos, T., Buchbinder, R., Woolf, A. and March, L. (2014). The global burden of hip and knee osteoarthritis: estimates from the Global Burden of Disease 2010 study. *Annals of the Rheumatic Diseases*, 73(7), pp.1323-1330.

[3] Ackerman, I., Kemp, J., Crossley, K., Culvenor, A. and Hinman, R. (2017). Hip and Knee Osteoarthritis Affects Younger People, Too. *Journal of Orthopedic & Sports Physical Therapy*, 47(2), pp.67-79.

[4] Wallace, I., Worthington, S., Felson, D., Jurmain, R., Wren, K., Maijanen, H., Woods, R. and Lieberman, D. (2017). Knee osteoarthritis has doubled in prevalence since the mid-20th century. *Proceedings of the National Academy of Sciences*, 114(35), pp.9332-9336.

[5] Bedson, J. and Croft, P. (2008). The discordance between clinical and radiographic knee osteoarthritis: A systematic search and summary of the literature. *BMC Musculoskeletal Disorders*, 9(1).

[6] Kinds, M., Welsing, P., Vignon, E., Bijlsma, J., Viergever, M., Marijnissen, A. and Lafeber, F. (2011). A systematic review of the association between radiographic and clinical osteoarthritis of hip and knee. *Osteoarthritis and Cartilage*, 19(7), pp.768-778.

[7] Bhatia, D., Bejarano, T. and Novo, M. (2013). Current interventions in the management of knee osteoarthritis. *Journal of Pharmacy and Bioallied Sciences*, 5(1), p.30.

[8] Huang, Z., Ding, C., Li, T. and Yu, S. (2017). Current status and future prospects for disease modification in osteoarthritis. *Rheumatology*.

[9] Zhang, W., Ouyang, H., Dass, C. and Xu, J. (2016). Current research on pharmacologic and regenerative therapies for osteoarthritis. *Bone Research*, 4(1).

[10] Braun, H. and Gold, G. (2012). Diagnosis of osteoarthritis: Imaging. *Bone*, 51(2), pp.278-288.

[11] Peterfy, C. and Kothari, M. (2006). Imaging osteoarthritis: Magnetic resonance imaging versus x-ray. *Current Rheumatology Reports*, 8(1), pp.16-21.

[12] Wu, Y., Yang, R., Jia, S., Li, Z., Zhou, Z. and Lou, T. (2014). Computer-aided diagnosis of early knee osteoarthritis based on MRI T2 mapping. *Bio-medical Materials and Engineering*, 24(6), pp.3379-3388.

[13] Thomson, J., O'Neill, T., Felson, D. and Cootes, T. (2015). Automated Shape and Texture Analysis for Detection of Osteoarthritis from Radiographs of the Knee. *Lecture Notes in Computer Science*, pp.127-134.

[14] Shamir, L., Ling, S., Scott, W., Hochberg, M., Ferrucci, L. and Goldberg, I. (2009). Early detection of radiographic knee osteoarthritis using computer-aided analysis. *Osteoarthritis and Cartilage*, 17(10), pp.1307-1312.

[15] Tiulpin, A., Thevenot, J., Rahtu, E. and Saarakkala, S. (2017). A Novel Method for Automatic Localization of Joint Area on Knee Plain Radiographs. *Image Analysis*, pp.290-301.

[16] Antony, J., McGuinness, K., Moran, K. and O'Connor, N. (2017). Automatic Detection of Knee Joints and Quantification of Knee Osteoarthritis Severity Using Convolutional Neural Networks. *Machine Learning and Data Mining in Pattern Recognition*, pp.376-390.

[17] Antony, J., McGuinness, K., Connor, N. and Moran, K. (2016) Quantifying Radiographic Knee Osteoarthritis Severity using Deep Convolutional Neural Networks. *ICPR 2016 Proceedings.*

[18] Nevitt, M., Felson, D. and Lester, G. (2006). *The Osteoarthritis Initiative: Protocol for the Cohort Study*. [PDF] National Institute of Arthritis, Musculoskeletal and Skin Diseases, pp.6-8. Available at: https://oai.epi-ucsf.org/datarelease/docs/studydesignprotocol.pdf [Accessed Aug. 2017].

[19] Ren, S., He, K., Girshick, R. and Sun, J. (2016). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), pp.1137-1149.

[20] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. and Berg, A. (2016). SSD: Single Shot MultiBox Detector. *Computer Vision – ECCV 2016*, pp.21-37.

[21] Redmon, J. and Farhadi, A. (2017). YOLO9000: Better, Faster, Stronger. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[22] He, K., Gkioxari, G., Dollar, P. and Girshick, R. (2017). Mask R-CNN. *2017 IEEE International Conference on Computer Vision (ICCV)*.

[23] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[24] Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *ICML*. 37, pp.448-456.

[25] Rahman, M. and Wang, Y. (2016). Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation. *Advances in Visual Computing*, pp.234-244.