Automatic Joint Detection Using Image Processing and Deep Learning

|  |  |  |
| --- | --- | --- |
| Dhruv Subhash Chamaria  *Computer Science*  *Pace University*  Jersey City, United States  dc05261n@pace.edu | Tarun Karambir Singh Dagar  *Computer Science*  *Pace University*  Jersey City, United States  td03603n@pace.edu | Manassa Middle Dharini  *Computer Science*  *Pace University*  Jersey City, United States  td03603n@pace.edu |

*Abstract*— Arthritis is an inflammatory disease that causes erosion in bones or narrowing of joint space in various joints of the body. The first symptom of this disease is seen in joints of hand finger and wrist joints thus making hand radiograph analysis extremely important. In this paper, an Image processing-based algorithm and a Deep learning algorithm Yolo are developed to yield solutions to major problems of joint detection. The Image processing algorithm is divided into the following steps, first image pre-processing is carried out using a Gaussian filter. The second-hand mask is extracted by separating the foreground and background by using Otsu’s binarization method. Third morphological thinning is applied to get thinned skeleton of the binarized image. Fourth To detect joint location in the original X-ray image Hough transform is used. Method 2 used yolov5 as a model which is used to predict the joint location as a bound box.

Keywords — Arthritis, X-ray, Joint detection, Yolov5, Image Processing

# Introduction

Arthritis is a musculoskeletal disease that frequently causes disability and losses functions of joints in fingers, wrists, knees, and feet. Conventional X-ray radiographs have been the most common standard method for identifying the progression of bone and joint damage in A. At the start of the disease, 90% of the symptom of Arthritis are seen in the hands, thus an automated and accurate hand radiograph analysis is required. Van’t Klooster developed a semi-automated method to measure all joint margins, JSW measurement was applied to both hands, and joints of a thumb are not considered for analysis [1]. Manual hand X-ray photo analysis is a very difficult and time-consuming task for the radiologist. Since bones in little fingers are thinner than other fingers, it is difficult to separate bones from the X-ray photo background [2]. For measurement and detection of joint margin, many algorithms are based on the active shape model ASM [3] and active appearance model AAM approach [4-5]. For medical image segmentation average edge magnitude and average, edge vector field models are used [6]. The existing method tends to emphasize semiautomatic joint localization [7]. Automatic joint location estimation is obtained by local linear mapping based on texture features. Bone contours are delineated by an active shape model comprised of a statistical model of bone shape and local texture [8].

The paper is organized as follows. In Section II, binarization and skeletonization are discussed. the automatic joint position detection and the accurate contour detection of bone are described. In Section III method 2 is discussed which is a representation of the gap between traditional computer vision and the new deep learning methods.

# Dataset

The given dataset of 3589 images in the DICOM format which is a medical imaging format that contains information related to x-rays such as patient name, type of image, and left hand or right hand.

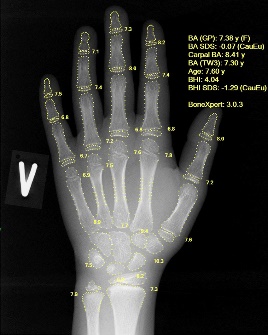


Fig 1: Dicom Image

Figure 1 shows how a DICOM image contains image data as well which is useful while processing. It helps to distinguish between left-hand, right-hand, and bilateral. The labeled text files were in the format of x, and y as the center of the joint and a rotation angle which was not used in this paper. The labels only contained data related to the right hand, which was mapped to left and bilateral-hand hand images as well. DICOM image was converted to jpg for processing purposes. For method 2 the image was transformed to a size of 640x640. The dataset was divided into the following categories of the first fifteen percent as a testing data set which comprised 533 images. Next fifteen percent i.e., 533 images as the validation set, and the last seventy percent i.e., 2489 images as the training set.

# Method 1

Image processing is the process of transforming an image into a digital form and performing certain operations to get some useful information from it. The image processing system usually treats all images as 2D signals when applying certain predetermined signal processing methods. This method proposed a computer vision algorithm using image processing that contains numerous steps to detect hand joint locations in an x-ray image. The steps include blurring of the image, morphological operations (such as dilation, erosion, etc.), feature detection, and edge detection. The following section will explain the process step by step and includes what is it and the purpose of the method used.

## Gray Scale Of the Image

A gray scale is applied to an image to minimize its size: For instance, RGB photos contain three colour channels and three dimensions, but grayscale shots are one-dimensional. lowering model complexity Consider neural network training on RGB pictures with a resolution of 10x10x3. The input layer will include 300 input nodes. The same neural network will only need 100 input nodes for grayscale images, though. Grayscale images are required for other ways to work. For instance, the OpenCV library's built-in Canny edge detection algorithm only works with grayscale pictures. Gray scaling on images is used to reduce dimension, and model complexity and to implement other computer vision algorithms like blurring, histogram equalization, edge detection, and Hough transform.

## Blurring Of the Image

The first blurring method utilized in the paper is averaging. An average filter replaces the centre pixel with the average by averaging all the pixels surrounding it, which is precisely what it could do.

By substituting the average of the region around a pixel with the value of that pixel's immediate vicinity, averaging smooths out its value. Because of this, the degree of detail and noise may be reduced by just utilizing the average.

## Histogram Equalization

The paper demonstrates how to equalize photos using the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique which is an adaptive thresholding technique. CLAHE is an adaptation of Adaptive Histogram Equalization (AHE) that addresses the issue of contrast over-amplification. Instead of processing the entire image, CLAHE works with discrete sections of it called tiles. The false borders are then eliminated by combining the adjacent tiles using bilinear interpolation. You can use this algorithm to make photographs’ contrast better. CLAHE can also be used on colour photos; typically, this is done on the luminance channel, and the outcomes are considerably better for an HSV image after merely adjusting the luminance channel than they are for a BGR image after adjusting all the channels. CLAHE Histogram Equalization is used with a clip limit set as the threshold for restriction of contrast.

## Canny Edge Detection

The Canny edge detector is an edge detection operator that uses a multi-stage algorithm to detect a wide range of edges in images. It was developed by John F. Canny in 1986. Canny also produced a computational theory of edge detection explaining why the technique works. This method used canny edge detection to distinguish between the bones and the edges of the joint. Canny edge detection is a very powerful tool.

## Hough Transform

The task of finding the circles falls to the cv2.HoughCircles function. The circle detection technique—currently, OpenCV only supports the cv2.cv.HOUGH GRADIENT method, and it is likely to remain the sole method for some time—the accumulator value of 1.5, and finally a minDist of 100 pixels—are handed in as the first and second arguments, respectively. The two parameters are maximum value and minimum value Between the two, Param1 is the higher threshold.

The parameter 1/2 is set in the second one. The candidate-detected circles' accumulation threshold is parameter 2. By increasing this threshold number, we can make sure that only the best circles, which correspond to greater accumulator values, are returned. The minimal radius of a circle, minRadius. maxRadius is the circle's largest radius. Here, we utilized minRadius and maxRadius in addition to param1 and param2. After performing this function, we utilized the if loop to get the radius and circumference of the picture as well as to add a circle to it. Next, we displayed the image.

After Hough contour detection was used to decide what type of contours are present as circles, and rectangles to only show circles and then show the output image. The result was not satisfying.

# Method 2

YOLO is refreshingly straightforward. Multiple bounding boxes and class probabilities for those boxes are simultaneously projected by a single neural network for each box. YOLO directly optimizes detection performance while training on complete photos. Comparing this unified model to conventional object identification techniques provides several advantages.

First off, YOLO moves quickly. We don't require a sophisticated process as we define detection as a regression problem. To forecast detections, we only run our neural network on a fresh image at test time. On a Titan X GPU, our base network operates at 45 frames per second without batch processing, and a fast version operates at more than 150 frames per second. This indicates that we can process streaming video in real time with a latency of under 25 milliseconds. Additionally, YOLO achieves a mean average precision that is more than twice as high as that of other real-time systems.

## Network

Diagram, engineering drawing

Description automatically generatedFig. 2. Architecture of YoloV5.

It consists of three parts: (1) Backbone: CSPDarknet, (2) Neck: PANet, and (3) Head: Yolo Layer. The data are first input to CSPDarknet for feature extraction and then fed to PANet for feature fusion. Finally, Yolo Layer outputs detection results (class, score, location, size) [2].

# Results and Experiments

The dataset was divided into a total of 533 testing images which were used to get the results. For method 1, the accuracy was calculated on the bases of the distance of the predicted point with the ground truth points. It contained x, and y coordinated as the output which was compared with the ground truth labels.

A buffer of fifteen percent plus or minus was kept to get the accuracy. In this process, it was found that most of the images which were hard to correct using the various methods mentioned above had low accuracy and had 1 or two points detected correctly. In the image processing method, it was also observed that it was hard to distinguish between the thumb joints and finger joints which was easily overcome by Yolov5.

Method 2 used Yolov5-trained weights to get the results. The same 533 images were used in method 1 to get the accuracy and it was able to achieve an overall accuracy of 97 percent. This marked a clear distinction between the two methods and help to understand that without the help of training it is difficult to achieve very high accuracy. Accuracy was independent of the type of image it was given it could predict in all kinds of images whether it was a low contrast image or a high one we got very high accuracy.

| Images | mAP |
| --- | --- |
| 533 (Yolo) | 0.973 |
| 533 (Image Processing) | 0.65 |

Table 1. Result and experiment

# Conclusion

Our method is a fully automated system to detect three different types of joints in MCP, DIP, and SIP. The results are satisfying and reliable. This system fails to detect joints in which images are blacked out and have improper joint X–ray images. The major challenge is to accurately measure joint localization and apply proper image processing techniques to achieve higher accuracy in method 1. Though computer vision has advanced a lot since the beginning it still has a long way to go.

##### References

1. N. M. Krishna, R. Y. Reddy, M. S. C. Reddy, K. P. Madhav, and G. Sudham, "Object Detection and Tracking Using Yolo," 2021 Third International Conference on Inventive Research in Computing Applications (CIRCA), 2021, pp. 1-7, doi: 10.1109/ICIRCA51532.2021.95445982
2. Xu, Renjie & Lin, Haifeng & Lu, Kangjie & Cao, Lin & Liu, Yunfei. (2021). A Forest Fire Detection System Based on Ensemble Learning. Forests. 12. 217. 10.3390/f12020217.

.