Project Report
CSL7360: Computer Vision

Project Title:

Analog Clock Reader

Group Members:

Aakash Maheshwari (B21CS002)
Naman Goyal (B21CS048)
Yatharth Shukla (B21CS080)

Problem Statement:

This project aims to develop a computer vision system capable of accurately reading the time displayed on an analog clock face in an image. The primary focus is on reading time accurately from some synthetic datasets and then extending it to more real life scenarios.

Methodology:

For this project, we have explored 2 different methodologies

The first approach uses **Geometrical methods with classical CV techniques** to extract the time information

The second approach explored the potential of deep learning by utilizing **Convolutional Neural Networks (CNNs)** to learn the inherent patterns within clock images and their corresponding times.

1. Geometrical Approach:

We explored geometrical approaches to detect time from some synthetic datasets.

It follows a three-step procedure:

1. **Image Preprocessing:**
   The input images are resized to 224*224 in order to slightly reduce the computational complexity for the later steps while still preserving enough information to achieve good results.

2. **Clock Hand Detection:**
   We first **find a circle** in the image using **Hough Circle Transform**, which represents the clock body.
   Then the **Canny Edge Detection** algorithm is run to highlight the edges in the image and **Hough Line Transform** is used for finding the locations of straight edges.
   There can be many lines detected in the image by the Hough Transform. So, in order to filter the actually useful edges that can represent the clock hours, some **filtering criterion** are used, which involve:
   - Distance from the center of the circle
   - Removing lines with very similar slope values, as it could be multiple detections of a single edge

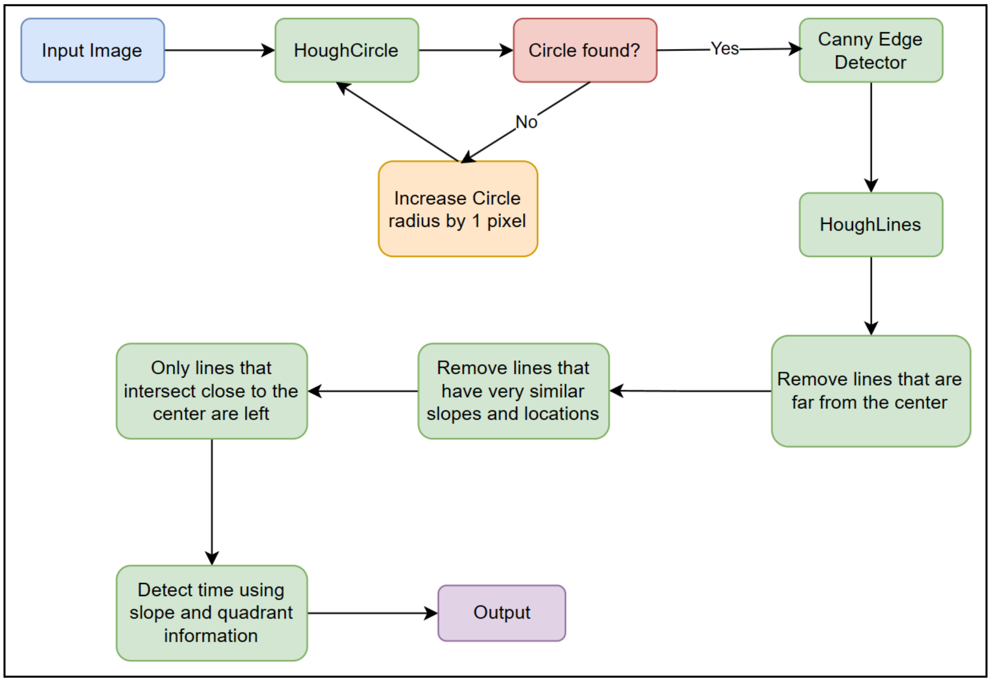
   The useful edges are then iterated through and the 2 lines closest to the center are selected to be treated as the hour and minute hands.

3. Now that we have the hands of the clock, we can determine the time. The angles of both the hands from the vertical are computed (it is assumed that the 12 o'clock position is in vertically upwards direction). Then using the angle values and the quadrant where the end of the hand lies in, we can calculate the time.

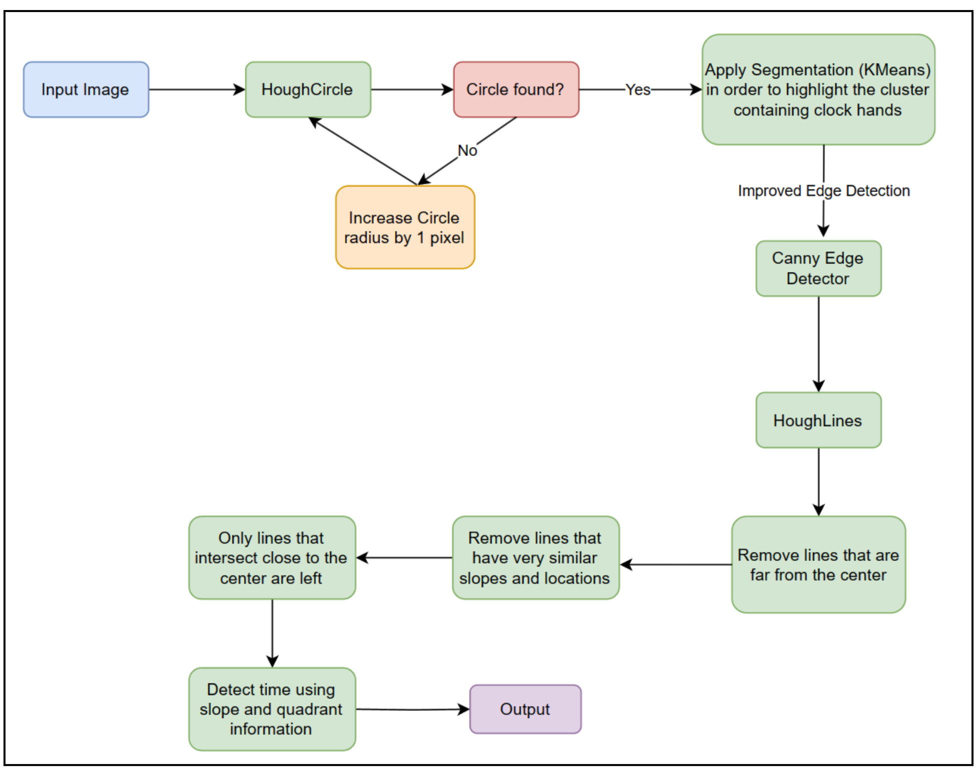
In some of the cases, the Canny Edge detector and Hough Line Transform were not able to detect the clock hands accurately as the colors of hands were similar to that of background, which led to poor gradient values and thus poor edge detection.

In order to address this problem, we explored image discretization to improve gradient values. Ultimately we finalized upon an approach to use KMeans clustering based image segmentation, which led to a **noticeable improvement** in the results.

The 2 variations for the geometrical approach are as follows:



First Approach



Improved Segmentation Approach

2. Learning based Approach:

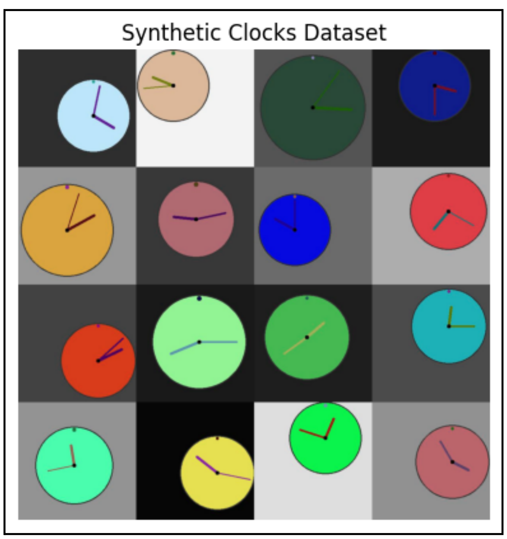
In this approach we tried making use of some commonly known CNN architectures and training them to learn the relationship between the clock patterns and the time. We tried to train ResNet18 on different datasets and have achieved decent results.

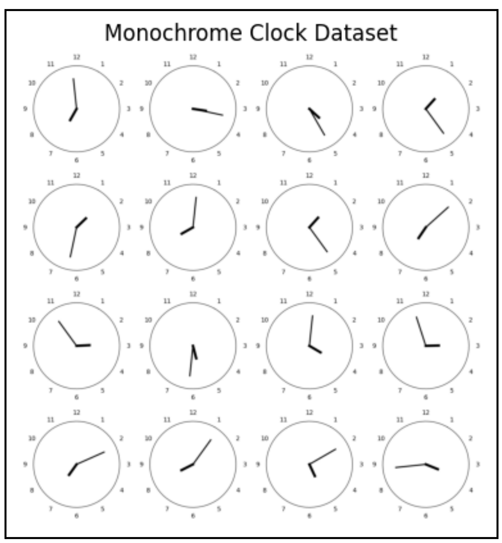
Training involved fine-tuning the pre-trained ResNet18 by replacing the final classification layers with new layers tailored to the specific task of predicting clock time. An appropriate loss function, such as CrossEntropyLoss was chosen to quantify the difference between the predicted time and the actual time displayed on the clock. The training process employed an optimizer like Adam to update the model weights based on the calculated loss.

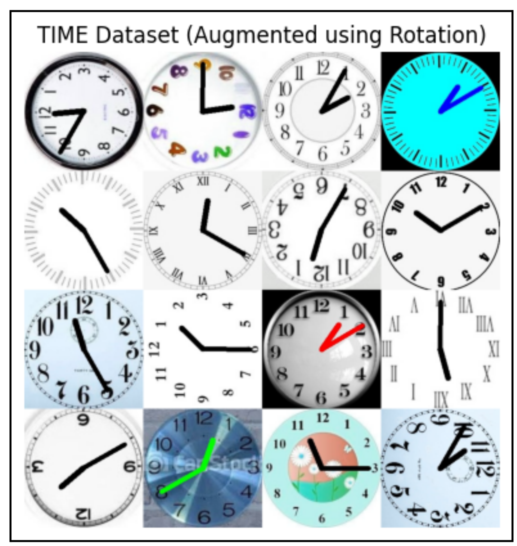
Results:

About the Datasets:

We used multiple different datasets in order to test the performance of our approaches.







For the first 2 datasets, the assumption of the 12 o'clock position being at vertically top is correct and thus we were able to run the geometric approach algorithm on those.

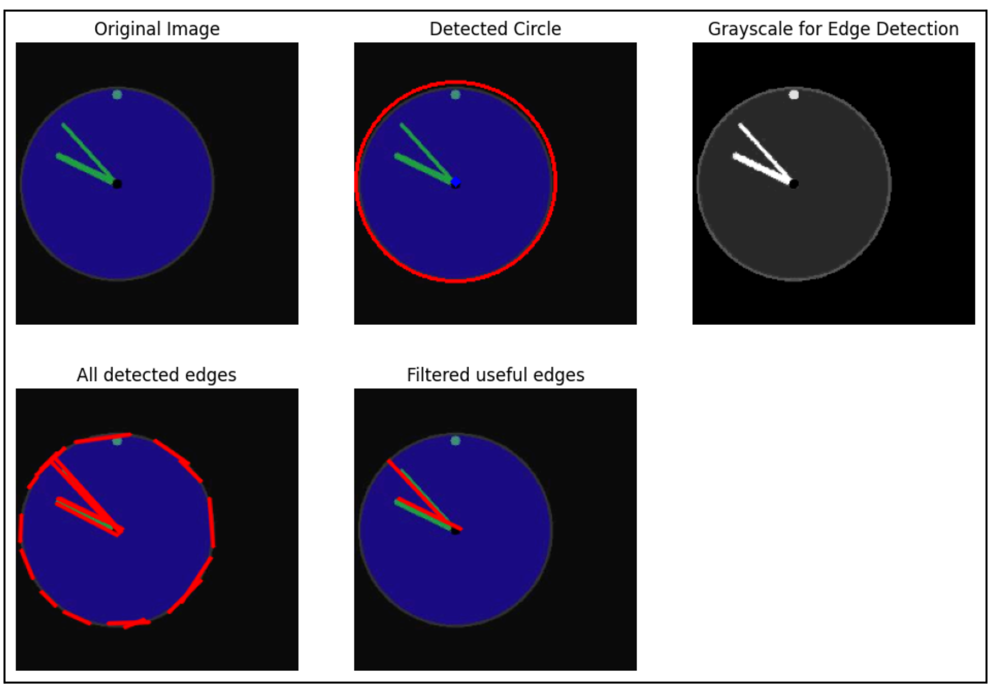
The 3rd dataset had images where the clocks were given some random rotations as well, which would be very complex to detect with only geometric models, thus we used CNNs in order to work on this dataset.

The first 2 datasets were also later used to train the CNNs in order to compare the results against the geometrical approach.

Results from the Geometrical Approach:

- In order to test the accuracy, an error window of 3 mins was used i.e. if the calculated time was within 3 minutes of actual labeled time, then it was treated as correct, otherwise incorrect.
- For minute specific accuracy, the same window was used.
- For hour specific accuracy, a window of 0 was used i.e. if the hour was exactly correct, only then it was treated as correct.

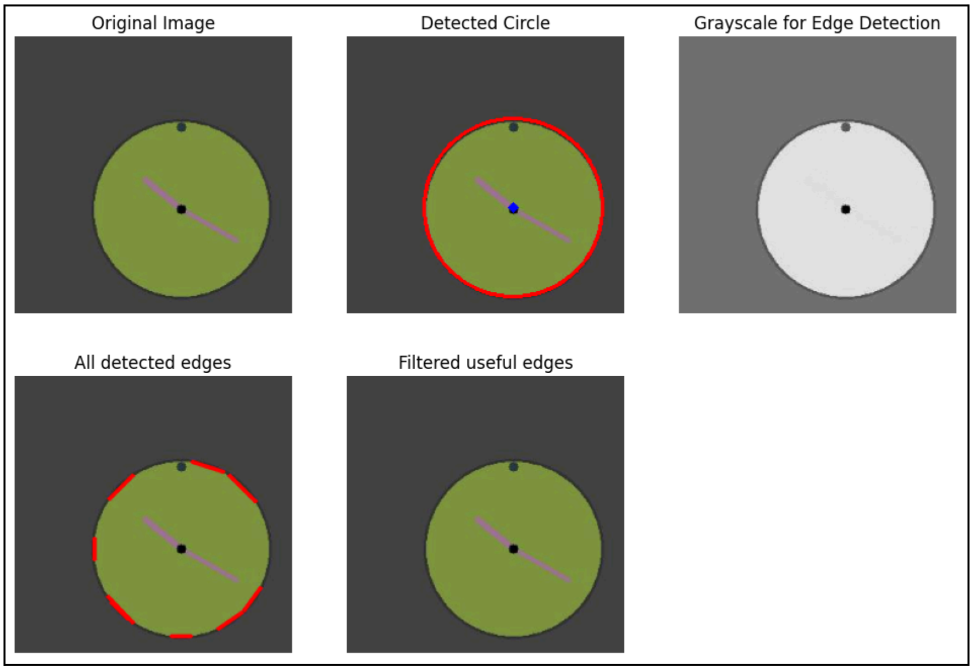
Some images showing the progression of each step in geometric algorithm are as follows:



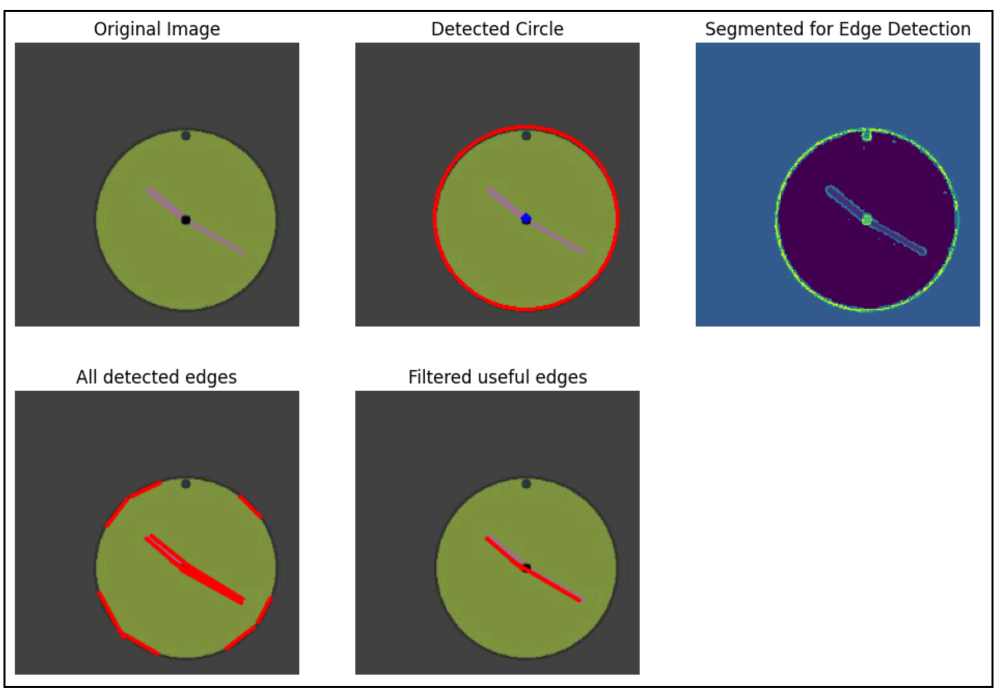

```
Time Calculated       : (9, 52)
Time Labelled (actual): (9, 53)
```

Improved Segmentation Approach:

Here is an example of the segmentation approach providing improved results:



Standard Approach



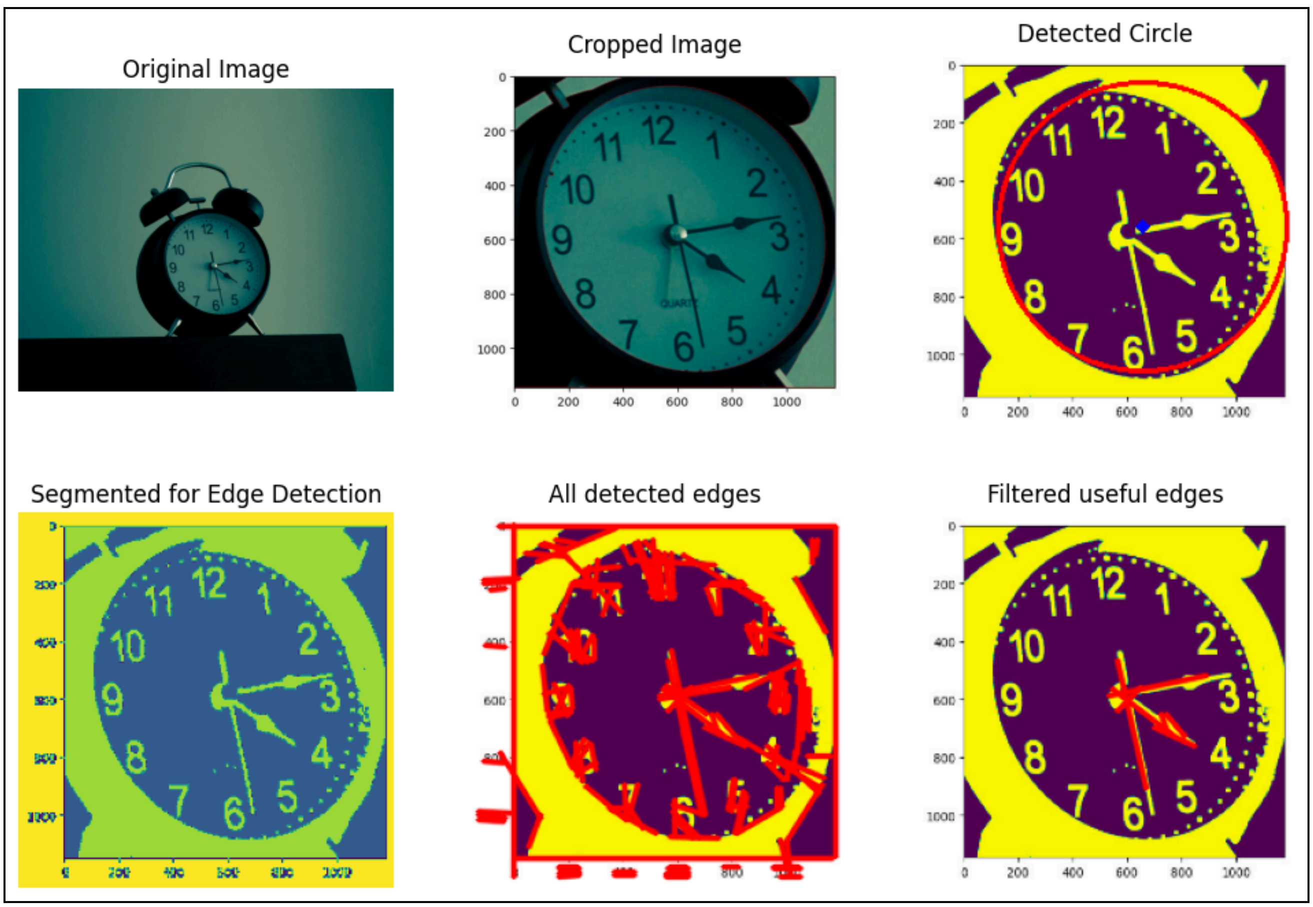
Improved Segmentation Approach

Testing with Real World Images:

Due to non-availability of sufficient labeled real-life data, we were unable to train any CNN based models for the said problem statement.
However, we have planned a pipeline to handle such data, which involves:
1. Clock localisation
2. Segmentation (for improved edge detection)
3. Time prediction using either Geometric approach or Learning approach.

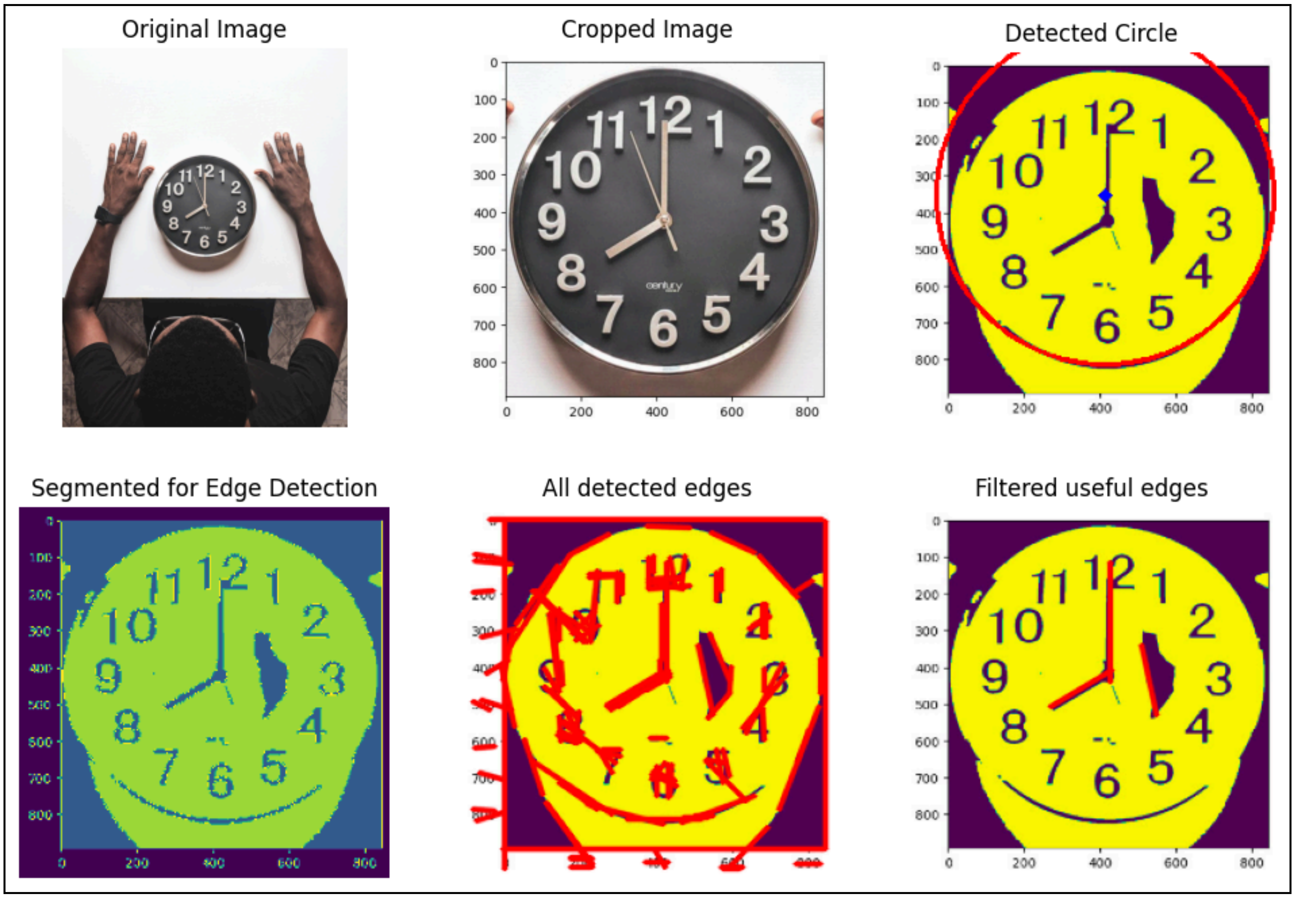
We tried to experiment with some images using manual labeling, results are as follows:

Image 1:




```
Time Calculated       : (4, 13)
Time Labelled (actual): (4, 13)
```

Image 2:




```
Time Calculated       : (8, 0)
Time Labelled (actual): (8, 0)
```

In order to locate the clock from the given image, we made use of a pre-trained YOLOv5 model, which is used for detection of objects of various sizes and outputs a vector containing the object category probability, confidence, and bounding box position information.

This approach **might not always yield good results** as real-life clocks can contain more than 2 hands, which can cause the algorithm to do flawed detection. However in these 2 cases presented above, we were able to get correct results.

The summary of results from Geometrical Approach is as follows:

We tested the algorithm on 1000 samples from each of the first 2 datasets.

	Synthetic Clocks Dataset	
	Standard Approach	Improved Segmentation Approach
Minutes Accuracy	60.3 %	**77.0 %**
Hour Accuracy	61.2 %	**83.6 %**
Overall Accuracy	51.4 %	**71.4 %**

	Monochrome Clocks Dataset	
	Standard Approach	Improved Segmentation Approach
Minutes Accuracy	95.9 %	95.3 %
Hour Accuracy	81.09 %	79.5 %
Overall Accuracy	78.7 %	77.4 %

Here we can see that the Segmentation based approach leads to significant improvements on the synthetic clocks dataset because of improved gradient detection. It doesn't have much effect on the Monochrome dataset as the images are already discretized into 2 color values, and thus there is no highly noticeable effect on the gradients.

The summary of results from Learning Based Approach is as follows:

	Minutes Accuracy	Hour Accuracy	Overall Accuracy
Synthetic Clock Dataset	96.43 %	99.76 %	NA
Monochrome Clock Dataset	NA	NA	72.9 %
TIME Dataset (augmented with rotation)	NA	NA	96.73 %

Observations and Conclusion:

We can observe that the CNNs perform better than the geometrical algorithm.
One reason for this is that CNNs can better adapt and generalize to give good performance on different kinds of input data, whereas an algorithm is more limited in this sense.
But they have a drawback as well, because of high computation complexity of training the model, and the memory usage while computing the outputs for test images. The geometric algorithm is relatively more efficient and runs in relatively lower time.

Overall, we conclude that it is possible for a computer to read the time of an analog clock, using both a geometric approach or a learning based approach. There is some scope for improvement for the geometric algorithm. We can choose between the two approaches depending on the system resource and time constraints.

References:

1. Research Paper: It's About Time: Analog Clock Reading in the Wild (paper)
2. Dataset: Synthetic Analog Clocks (synthetic)
3. Dataset: Monochrome Clocks (monochrome)
4. Dataset: TIME (augmented using rotation) (augmented)