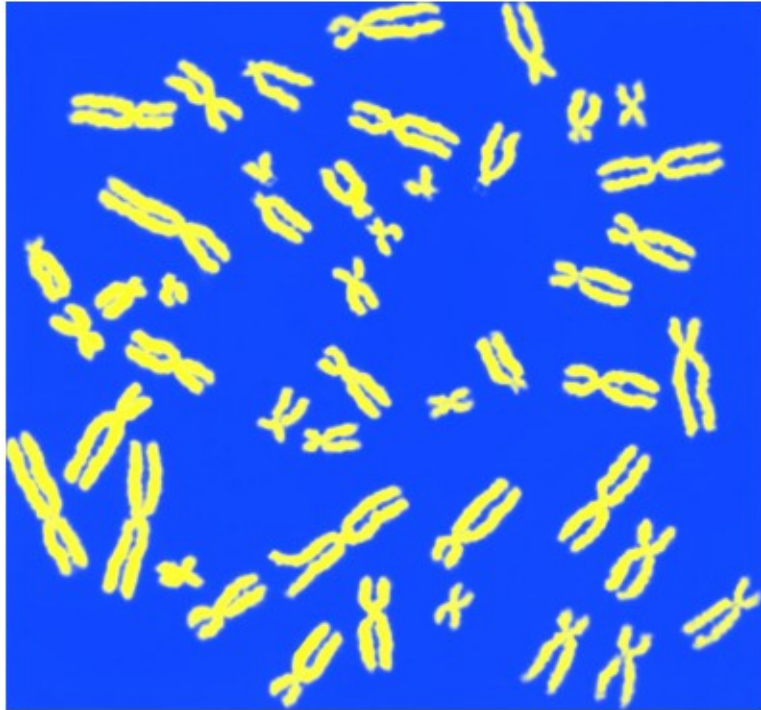Usman Akinyemi
LAB1 REPORT
MACHINE LEARNING AND PATTERN RECOGNITION

This is the image after reading the chromosomes.jpg → What I noticed after reading the image was that, when I tried to display it using matplotlib, the blue was showing Red background. So, I had to convert it from BGR to RGB before displaying it. My Observation from here will be that, when reading colored image using cv, it read it in BGR format.
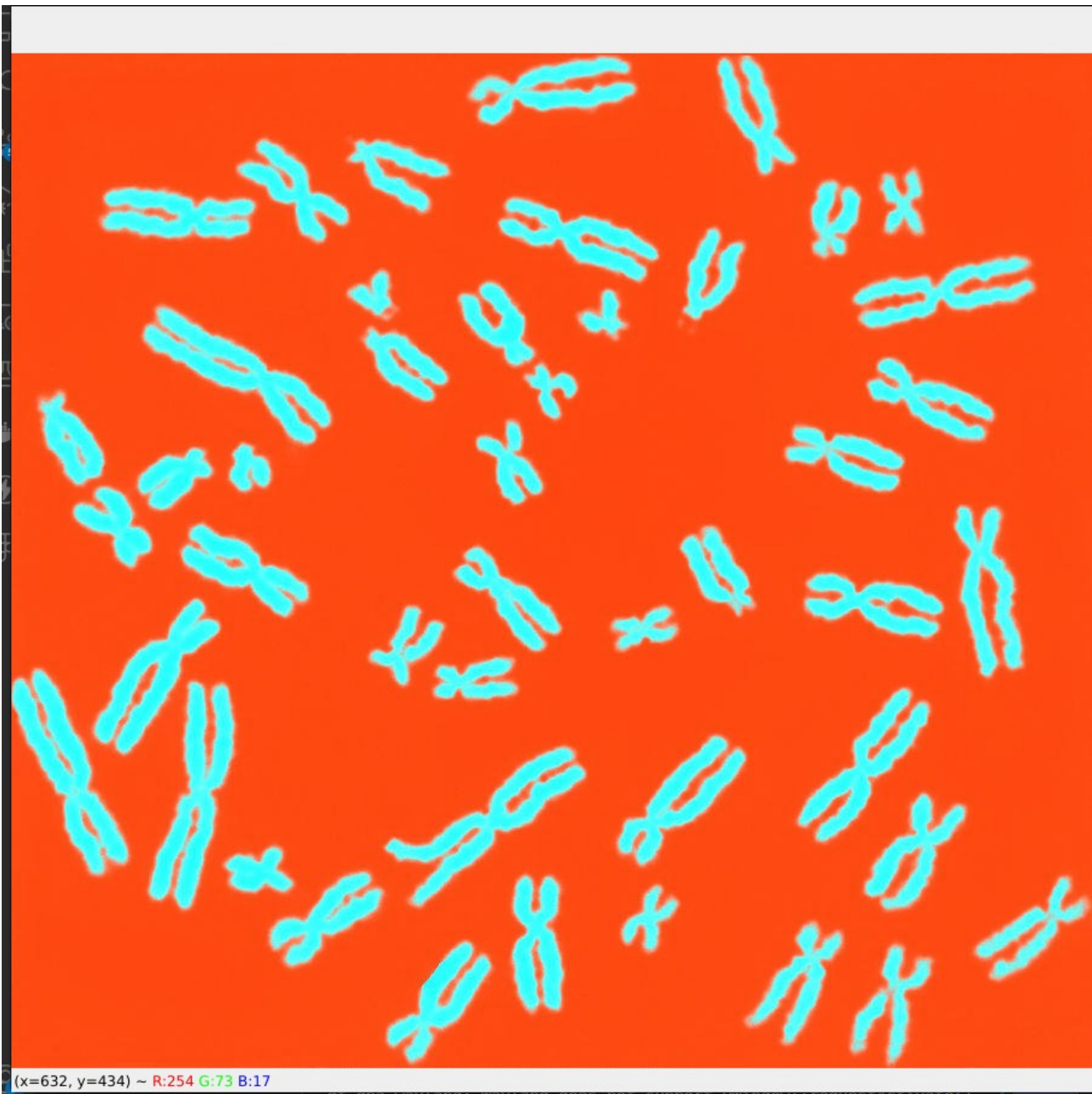
Image converted to Gray using `cv2.cvtColor(img, cv2.COLOR_RGB2GRAY)`
as the image was already in RGB format. I was expecting white and black image since it was supposed to be gray.

After applying morphological opening for background removal, I got this image. Which is a gray image.

Image after thresholding. I tried out different threshold_type until I got the `cv2.THRESH_OTSU` which give good result.
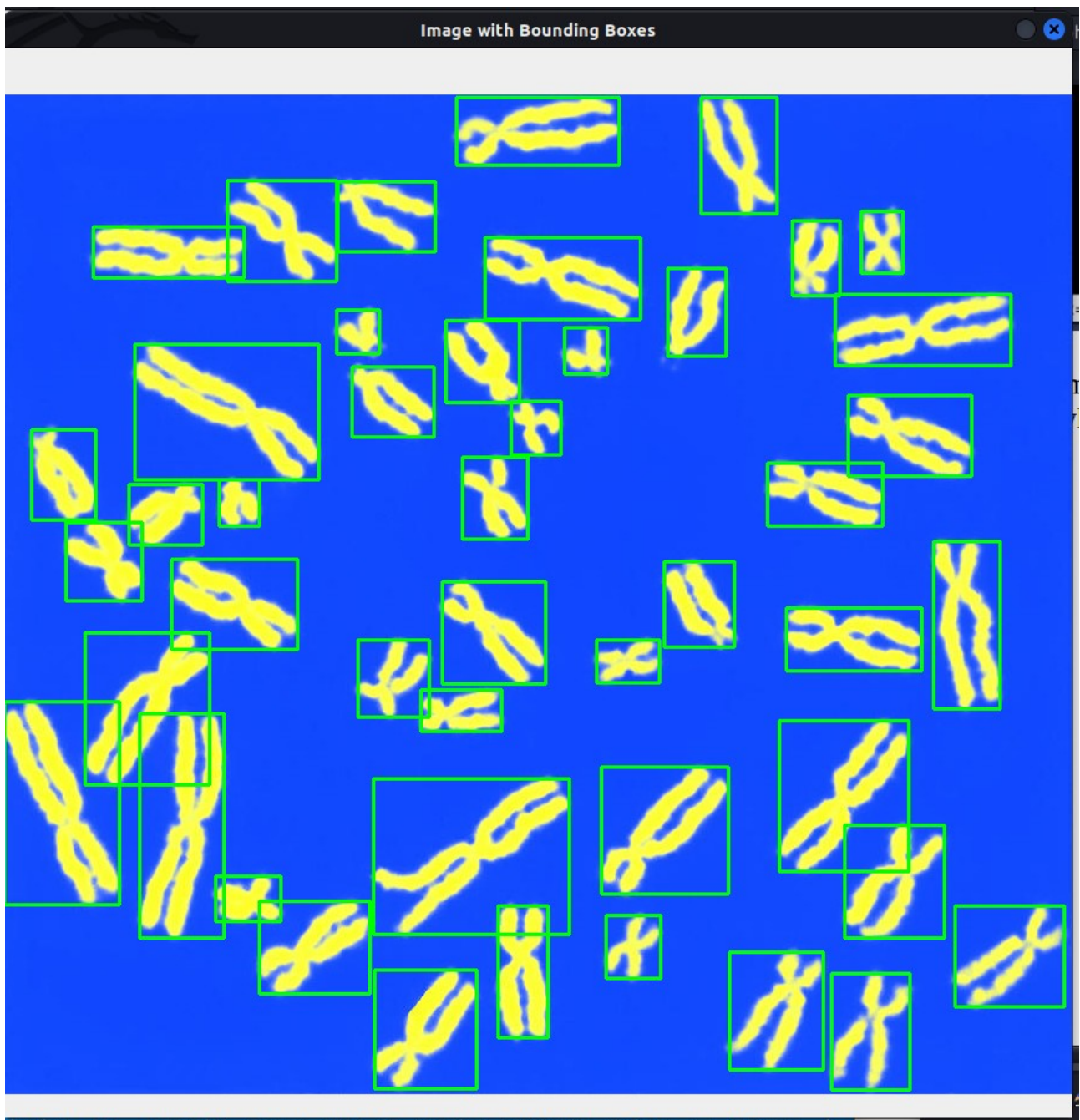
This is my image after finding the contour(which is a basically a continous line that can be used to indicate edges and seperate one type of object from another.).

Before GETTING THIS, I was actually getting small box inside of the bigger box(overlapping). I had to use filter the box by the size to remove the smaller.

```
           Area   Perimeter   Circularity   Height   Width
    0     2172.5  415.806130     0.157902       98      66
    1     3092.0  440.558436     0.200190      100      86
    2     2515.5  431.203098     0.170008       99      79
    3     1000.0  213.338093     0.276104       53      46
    4     2285.0  438.901582     0.149060       85      92
    5     3206.5  475.462984     0.178241      111      42
    6     2957.0  354.534051     0.295628       78      93
    7     1324.5  180.267025     0.512188       38      55
    8     2912.0  460.759447     0.172366       95      84
    9     5432.0  612.582822     0.181903      131     165
   10     3680.0  497.470124     0.186863      107     107
   11     3915.0  605.553382     0.134164      127     109
   12     5723.0  690.759446     0.150723      189      71
   13     6009.5  489.830515     0.314743      171      97
   14     1500.5  265.580734     0.267333       35      68
   15     1003.5  193.438599     0.337009       36      53
   16     1475.0  265.622364     0.262708       65      60
   17     4794.5  386.031525     0.404304      128     105
   18     3008.0  485.989894     0.160042       53     114
   19     2534.5  386.859951     0.212811       86      87
   20     2222.0  225.137082     0.550884       72      59
   21     3384.0  343.705624     0.359971       76     106
   22     3618.0  596.274166     0.127875      141      56
   23     1968.0  241.480228     0.424103       66      64
   24     1778.0  187.681239     0.634308       51      62
   25      892.0  128.568541     0.678118       39      34
   26     2492.0  397.220343     0.198470       53      97
   27     1560.5  270.208150     0.268582       69      55
   28     2189.5  214.551297     0.597713       76      54
   29      818.5  174.610171     0.337357       45      42
   30     2748.5  437.060963     0.180810       68     104
   31     2098.5  230.208150     0.497597       59      69
   32     5326.5  525.570629     0.242320      114     155
   33      709.5  133.982755     0.496666       39      36
   34     1986.5  288.149275     0.300651       69      62
   35      743.0  124.568541     0.601703       37      36
   36     3588.5  632.658942     0.112664       60     148
   37     1767.0  313.421353     0.226042       74      49
   38     3653.5  477.203098     0.201610       69     131
   39     3856.0  365.279218     0.363159       43     127
   40     1552.5  263.379723     0.281240       63      40
   41     1041.0  220.994947     0.267853       52      35
   42     1988.0  337.563488     0.219238       59      82
   43     2753.5  342.291411     0.295327       85      92
   44     2436.0  398.676187     0.192596       98      64
   45     3494.5  563.546243     0.138272       57     137
```

Features before standardization and normalization.

Features mean and standard deviation after standardzation.

```
45   3494.5   563.546243       0.138272       57    137
Standardized Features:
Area          1.01105
Perimeter     1.01105
Circularity   1.01105
Height        1.01105
Width         1.01105
dtype: float64
Area          -1.134358e-16
Perimeter     -1.399846e-16
Circularity    2.896234e-17
Height        -9.895466e-17
Width         -1.641199e-16
dtype: float64
```

Standardized Features.

```
Standardized Features:
        Area  Perimeter  Circularity    Height     Width
0   -0.346347   0.355871    -0.914526  0.553956 -0.414485
1    0.342358   0.524524    -0.630160  0.611283  0.182270
2   -0.089440   0.460780    -0.833117  0.582619 -0.026595
3   -1.224549  -1.023674    -0.119671 -0.735907 -1.011241
4   -0.262085   0.513235    -0.973984  0.181329  0.361296
5    0.428118   0.762351    -0.777756  0.926583 -1.130592
6    0.241243  -0.061615     0.011616 -0.019317  0.391134
7   -0.981499  -1.249009     1.467883 -1.165862 -0.742701
8    0.207538   0.662166    -0.817260  0.467965  0.122594
9    2.095016   1.696637    -0.753131  1.499855  2.539453
10   0.782769   0.912300    -0.719776  0.811928  0.808863
11   0.958784   1.648741    -1.074153  1.385201  0.868538
12   2.312975   2.229304    -0.962802  3.162345 -0.265297
13   2.527564   0.860246     0.140159  2.646400  0.510485
14  -0.849675  -0.667712    -0.178653 -1.251852 -0.354810
15  -1.221927  -1.159262     0.289883 -1.223189 -0.802376
16  -0.868774  -0.667428    -0.209757 -0.391944 -0.593512
17   1.617529   0.152997     0.742411  1.413864  0.749187
18   0.279442   0.834078    -0.900138 -0.735907  1.017727
19  -0.075209   0.158642    -0.545287  0.209992  0.212108
20  -0.309272  -0.943280     1.728095 -0.191298 -0.623350
21   0.561066  -0.135396     0.444292 -0.076644  0.779025
22   0.736332   1.585515    -1.116443  1.786491 -0.712863
23  -0.499517  -0.831924     0.875556 -0.363280 -0.474161
24  -0.641827  -1.198491     2.289083 -0.793234 -0.533836
25  -1.305441  -1.601264     2.583688 -1.137198 -1.369294
26  -0.107042   0.229234    -0.641726 -0.735907  0.510485
27  -0.804735  -0.636182    -0.170255 -0.277289 -0.742701
28  -0.333614  -1.015408     2.042999 -0.076644 -0.772538
29  -1.360492  -1.287552     0.292225 -0.965216 -1.130592
30   0.085077   0.500693    -0.760484 -0.305953  0.719349
31  -0.401773  -0.908728     1.369764 -0.563926 -0.324972
32   2.015997   1.103767    -0.346853  1.012573  2.241075
33  -1.442133  -1.564373     1.363506 -1.137198 -1.309618
34  -0.485661  -0.513938     0.045398 -0.277289 -0.533836
35  -1.417042  -1.628518     2.069830 -1.194525 -1.309618
36   0.714236   1.833428    -1.218735 -0.535262  2.032211
37  -0.650066  -0.341743    -0.456316 -0.133971 -0.921727
38   0.762921   0.774208    -0.620609 -0.277289  1.524969
39   0.914593   0.011598     0.465733 -1.022543  1.405618
40  -0.810727  -0.682708    -0.085137 -0.449271 -1.190267
41  -1.193840  -0.971503    -0.175159 -0.764571 -1.339456
42  -0.484537  -0.177247    -0.502072 -0.563926  0.062919
43   0.088822  -0.145032     0.009592  0.181329  0.361296
44  -0.148986   0.239153    -0.681228  0.553956 -0.474161
45   0.643830   1.362519    -1.046526 -0.621253  1.703995
```

Observation :- I observe, after standardization, the value of each of the features come to the same scale around -3 to 3 which differs from the original features in which each of the features have different scale. I also observe that each of the feature have mean of approximately zero and std of 1.

a

```
Normalized Features:
        Area   Perimeter  Circularity    Height     Width
0    0.276038   0.514381     0.080004  0.409091  0.244275
1    0.449528   0.558098     0.154789  0.422078  0.396947
2    0.340755   0.541575     0.101414  0.415584  0.343511
3    0.054811   0.156784     0.289043  0.116883  0.091603
4    0.297264   0.555171     0.064367  0.324675  0.442748
5    0.471132   0.619746     0.115973  0.493506  0.061069
6    0.424057   0.406162     0.323570  0.279221  0.450382
7    0.116038   0.098374     0.706554  0.019481  0.160305
8    0.415566   0.593777     0.105584  0.389610  0.381679
9    0.891038   0.861925     0.122449  0.623377  1.000000
10   0.560472   0.658615     0.131221  0.467532  0.557252
11   0.604811   0.849510     0.038024  0.597403  0.572519
12   0.945943   1.000000     0.067308  1.000000  0.282443
13   1.000000   0.645122     0.357376  0.883117  0.480916
14   0.149245   0.249054     0.273531  0.000000  0.259542
15   0.055472   0.121638     0.396752  0.006494  0.145038
16   0.144434   0.249128     0.265351  0.194805  0.198473
17   0.770755   0.461793     0.515762  0.603896  0.541985
18   0.433679   0.638338     0.083788  0.116883  0.610687
19   0.344340   0.463256     0.177110  0.331169  0.404580
20   0.285377   0.177623     0.774987  0.240260  0.190840
21   0.504623   0.387037     0.437360  0.266234  0.549618
22   0.548774   0.833121     0.026902  0.688312  0.167939
23   0.237453   0.206488     0.550778  0.201299  0.229008
24   0.201604   0.111469     0.922522  0.103896  0.213740
25   0.034434   0.007065     1.000000  0.025974  0.000000
26   0.336321   0.481555     0.151748  0.116883  0.480916
27   0.160566   0.257227     0.275740  0.220779  0.160305
28   0.279245   0.158927     0.857804  0.266234  0.152672
29   0.020566   0.088383     0.397368  0.064935  0.061069
30   0.384717   0.551921     0.120516  0.214286  0.534351
31   0.262075   0.186579     0.680750  0.155844  0.267176
32   0.871132   0.708245     0.229296  0.512987  0.923664
33   0.000000   0.016627     0.679104  0.025974  0.015267
34   0.240943   0.288914     0.332455  0.220779  0.213740
35   0.006321   0.000000     0.864860  0.012987  0.015267
36   0.543208   0.897384     0.000000  0.162338  0.870229
37   0.199528   0.333550     0.200509  0.253247  0.114504
38   0.555472   0.622819     0.157301  0.220779  0.740458
39   0.593679   0.425140     0.442999  0.051948  0.709924
40   0.159057   0.245167     0.298125  0.181818  0.045802
41   0.062547   0.170307     0.274450  0.110390  0.007634
42   0.241226   0.376189     0.188475  0.155844  0.366412
43   0.385660   0.384540     0.323038  0.324675  0.442748
44   0.325755   0.484126     0.141359  0.409091  0.229008
45   0.525472   0.775317     0.045289  0.142857  0.786260
```

Normalized Features:- After Normalization, I noticed that the value of each of the features use the same scale from 0 – 1.

Q1. Since the contours is a boundary or continous mark along an object in an image obtained through techniques like edge detection, we can employ this to segement one object from another in an image. And also based on the shapes of the contour marks, object of different shapes can be categorized.

Q2. - Standardizations make our model to converge faster and yield better performance since it mitigate the impact of extreme values. Standardization also ensure all the features uses the same scale which make it easier to compare features and show their relative importance.

Q3. There are many method that can be use when one of the value of he width is missing, we can fill the missing value using regression or interpolation, another one is by filling it with mean or median of the column.

Q4. Normalization ensure that data are scaled between 0 and 1 and it good for non-Guassian data. It also ensures that all the features use similar scale making it easier to comapre and interpret their contribiton to the model.

I observe for nomalization, the features all have positive columns and ranges from 0 – 1 while for the standardization there are some negative value and it have std of 1 and mean of zero.

Q5. Actually for this question, in the first instance when I was working on the problems, I got the problem of overlapping, the way I resolve was to defined a minimum area for the box for it to render. After this, I was able to solve the problem.