# FEATURE EXTRACTION FROM LABELED BLOBS

A presentation by **Hans Emmanuel Hernandez** (2020-11387)

# OBJECTIVES

In this activity, we perform feature extraction on a set of images. Specifically, we have the following objectives:

- Perform preliminary image segmentation
- Perform preliminary morphological operations
- Label blobs as distinct regions in an image
- Perform feature extraction
- Extract various properties and visualize them on a feature space
- Apply these techniques on multiple images
- Discuss the principles behind feature extraction and the apparent pitfalls of some methods

# 01 RESULTS AND ANALYSIS

# IMAGE PREPARATIONS

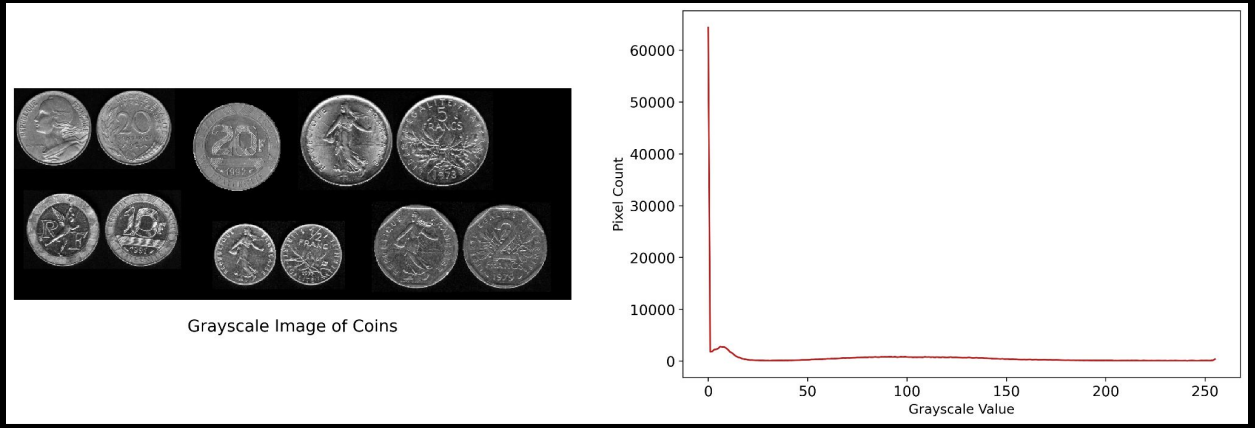



Figure 1. A sample image of coins and its corresponding grayscale histogram.

We choose an image of coins and compute for its corresponding grayscale histogram. **As we can see in Fig. 1, most of the pixels have a grayscale value of zero due to the black background.** To be safe, we choose a threshold value above zero that is close enough to isolate the features of the coins.

# IMAGE SEGMENTATION

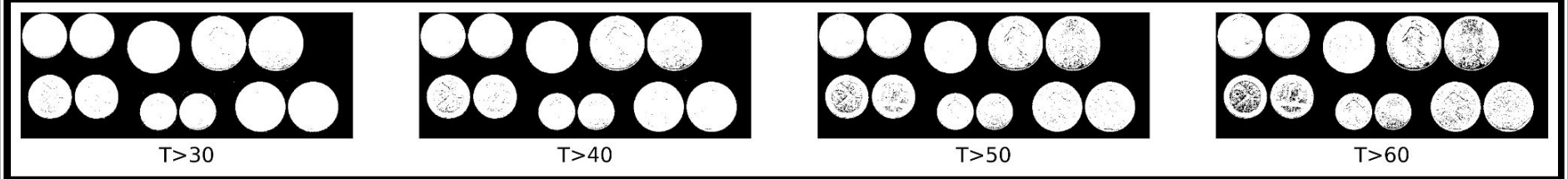



Figure 2. Various thresholding conditions on the image.

From Fig. 2, we set choose the best thresholding condition as T>30. **If the pixel values are within the range of T>30, they are set to True which corresponds to white in grayscale value.** The pixels that fall outside of this condition are set to False which corresponds to black in grayscale, **effectively preserving the black background while the coins are primarily set to white.**

# MORPHOLOGICAL OPERATIONS AND LABELING

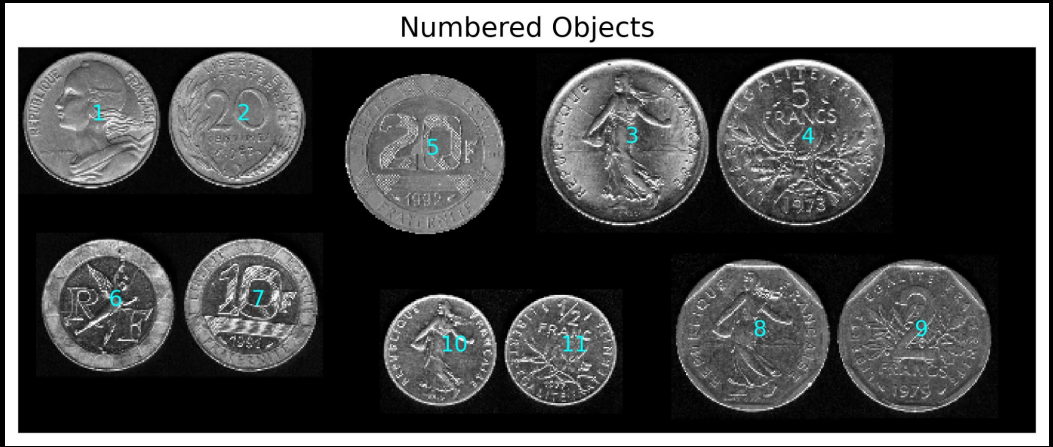



Figure 3. Numbered coins after morphological operations.

**To clean the objects, we perform morphological operations, specifically morphological closing followed by morphological opening.** As discussed previously, the closing operation dilates then erodes the image to fill in any small holes or gaps that we may have missed from segmentation. The opening operation erodes then dilates this image to preserve the edges and shapes of the larger objects. **Finally, we label these morphed objects and assign indices which we affix to their centroids.**

# MORE ON LABELING

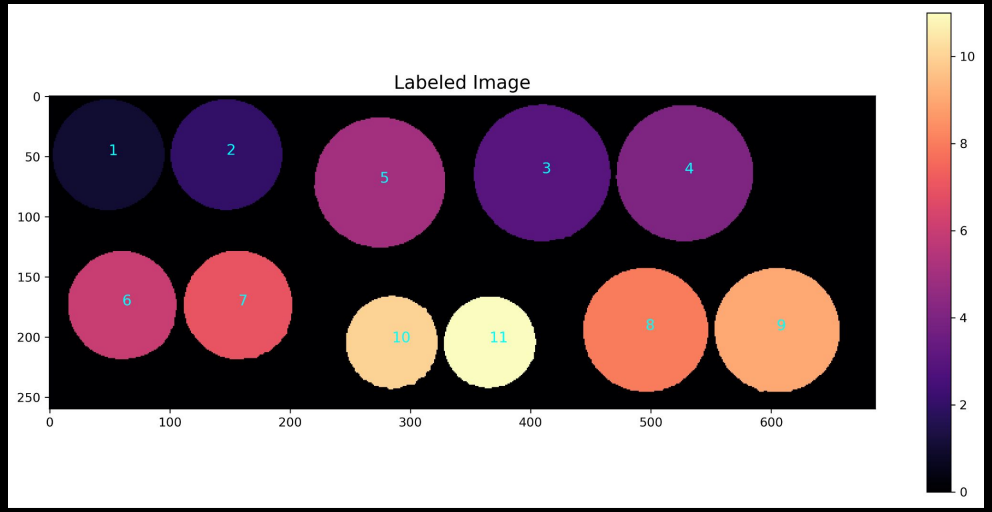



Figure 4. Labeled blobs of the original image sample of coins.

**By default, all pixels having a value of zero are set as the background of the image while the rest of the connected regions are labeled accordingly.** As we can see in Fig. 4, the package was able to separate **11 distinct blobs from the background.** We can also see that the blob edges are well preserved and did not merge nor break after previous morphological operations.

# FEATURE EXTRACTION

| index | perimeter | area | eccentricity | bbox-0 | bbox-1 | bbox-2 | bbox-3 |
|---|---|---|---|---|---|---|---|
| 1 | 302.73506 | 6698 | 0.127755943 | 3 | 3 | 95 | 96 |
| 2 | 302.73506 | 6689 | 0.139908256 | 3 | 101 | 95 | 194 |
| 3 | 374.0904 | 10134 | 0.068598768 | 7 | 353 | 121 | 467 |
| 4 | 373.26198 | 10103 | 0.106893577 | 8 | 471 | 121 | 585 |
| 5 | 356.19091 | 9221 | 0.100809012 | 18 | 220 | 126 | 329 |
| 6 | 296.83557 | 6397 | 0.156155693 | 129 | 16 | 219 | 106 |
| 7 | 296.24978 | 6434 | 0.170650768 | 129 | 112 | 219 | 202 |
| 8 | 339.70563 | 8408 | 0.118571409 | 143 | 444 | 246 | 548 |
| 9 | 341.11984 | 8441 | 0.118492863 | 143 | 553 | 246 | 657 |
| 10 | 255.96551 | 4621 | 0.121430292 | 166 | 247 | 244 | 323 |
| 11 | 250.45079 | 4563 | 0.167956244 | 167 | 328 | 243 | 405 |

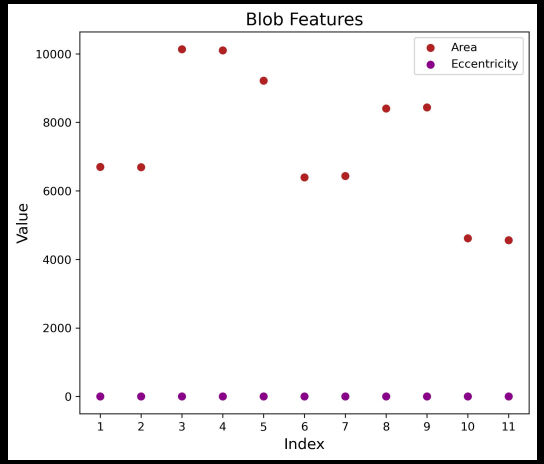



Figure 5. Extracted features from the labeled blobs and the scatter plot of area and eccentricity values.

Since we have already labeled each blob in the previous slide, **we simply call the regionprops function to extract the features of each blob.** By specifying the properties we want to measure, we can display various attributes as we can see in Fig. 5. **We have also scatter plotted the area and eccentricity values of the blobs to visualize them onto a 'feature space' separated by indices that indicate the blob identity.**

# 02 EXTRA DEMONSTRATIONS

# FEATURE EXTRACTION: WATERMELON SEEDS

Image of Watermelon Seeds

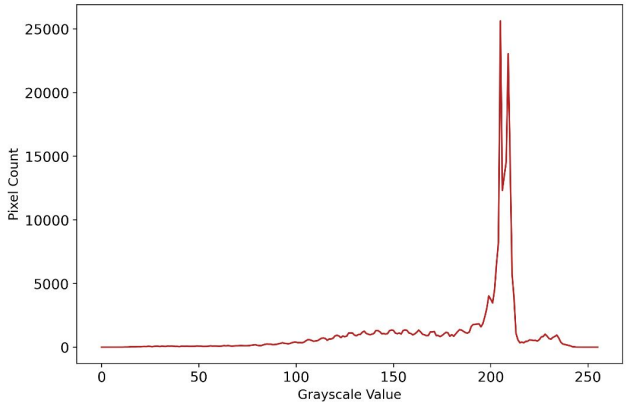



Figure 6. A sample image of watermelon slices and its corresponding grayscale histogram.

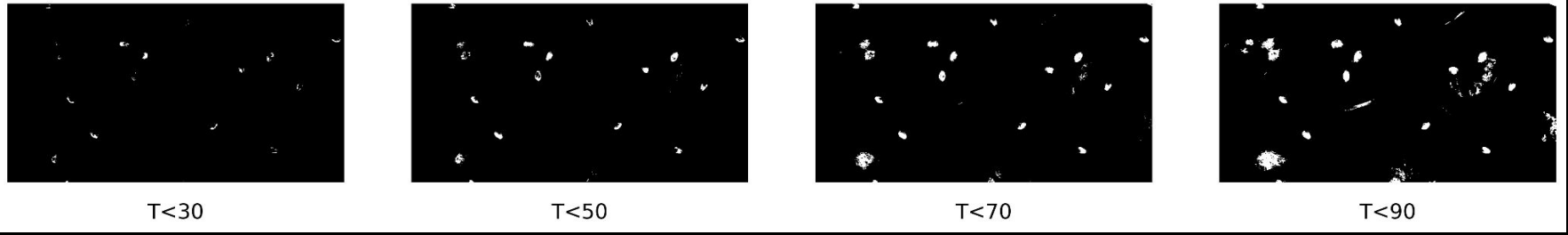
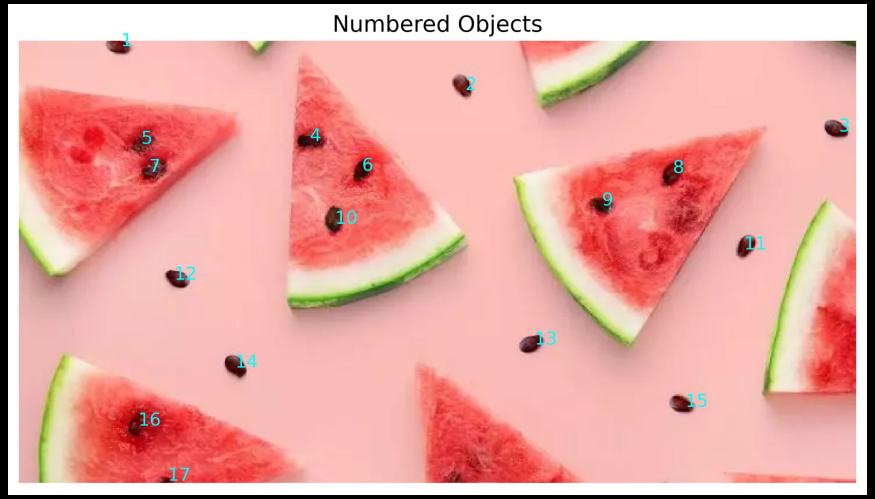



Figure 7. Various thresholding conditions for image segmentation. T<50 was chosen to avoid further sensitivities in morphological operations.

# FEATURE EXTRACTION: WATERMELON SEEDS



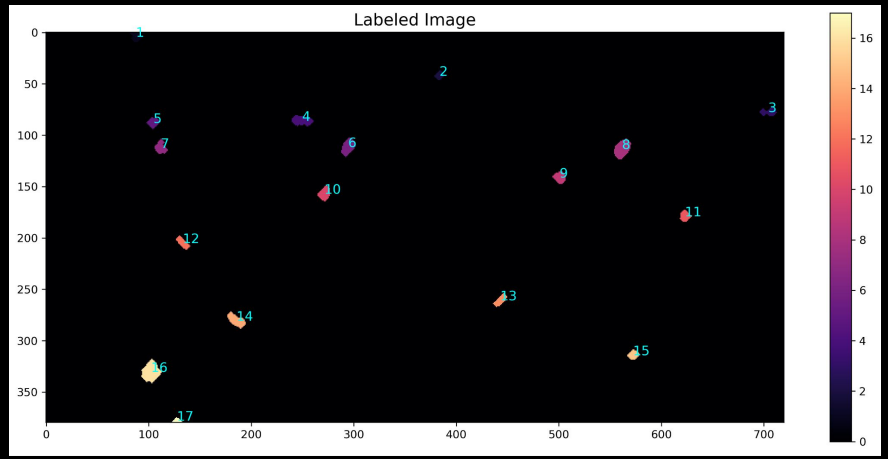




Figure 8. Labeled images after morphological operations.

Again, we repeat the same set of morphological operations as before: morphological closing then opening. **Surprisingly enough, the algorithm was able to pick out every single watermelon seed in the image; even those that are embedded inside the watermelon flesh.** We map these blobs out onto our 'feature space' for feature extraction.

# FEATURE EXTRACTION: WATERMELON SEEDS

| index | perimeter | area | eccentricity | bbox-0 | bbox-1 | bbox-2 | bbox-3 |
|---|---|---|---|---|---|---|---|
| 1 | 36.041631 | 83 | 0.62444014 | 0 | 81 | 10 | 94 |
| 2 | 25.79899 | 57 | 0.56500718 | 37 | 379 | 47 | 388 |
| 3 | 43.941125 | 95 | 0.89074778 | 73 | 696 | 82 | 713 |
| 4 | 53.941125 | 164 | 0.89059977 | 81 | 240 | 92 | 261 |
| 5 | 35.941125 | 92 | 0.55432971 | 83 | 98 | 95 | 112 |
| 6 | 47.941125 | 164 | 0.79981232 | 103 | 288 | 122 | 302 |
| 7 | 43.355339 | 115 | 0.61945705 | 104 | 106 | 119 | 119 |
| 8 | 55.355339 | 224 | 0.76395245 | 104 | 554 | 124 | 571 |
| 9 | 36.627417 | 111 | 0.47188427 | 136 | 494 | 148 | 507 |
| 10 | 40.284271 | 123 | 0.68476623 | 149 | 265 | 165 | 277 |
| 11 | 33.455844 | 88 | 0.42713462 | 173 | 618 | 185 | 630 |
| 12 | 37.941125 | 90 | 0.90776545 | 198 | 127 | 212 | 141 |
| 13 | 37.941125 | 89 | 0.90705142 | 254 | 436 | 268 | 450 |
| 14 | 52.183766 | 161 | 0.87368792 | 272 | 177 | 289 | 195 |
| 15 | 30.627417 | 78 | 0.57973527 | 309 | 567 | 319 | 579 |
| 16 | 65.740115 | 284 | 0.57253025 | 318 | 93 | 342 | 113 |
| 17 | 21.313708 | 29 | 0.85625505 | 375 | 123 | 380 | 133 |

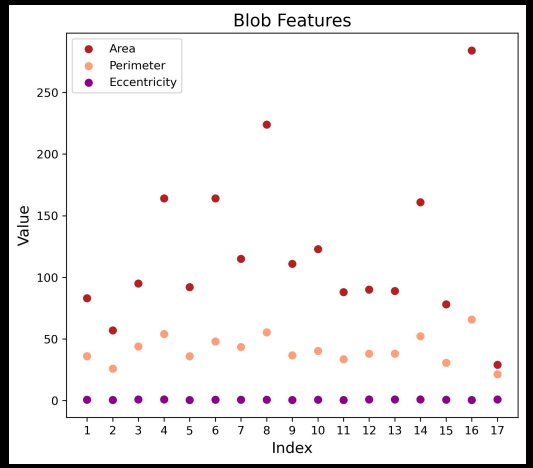



Figure 9. Extracted features from the labeled watermelon seeds and the scatter plot of area, perimeter, and eccentricity values.

We have successfully extracted the following features from the watermelon seeds of the image: **perimeter, area, eccentricity, and the bounding box coordinates.** As we can see on the scatter plot, **watermelon seed # 16 has the highest values for both the area and perimeter while watermelon # 17 had the lowest for both.** However, if we look back in Fig. 8, we will find that **this is not necessarily true** because the surrounding flesh of seed # 16 was superimposed due to segmentation sensitivities which in turn could have been enhanced further by subsequent morphological operations (i.e. in choosing the structuring element). The low values for seed # 17 are also attributed mainly to the fact that half of the seed is out of the picture and thus, **its measurement values did not manifest accurately in feature extraction.** Therefore, it is important to always know the context at which we are extracting our features from.

# FEATURE EXTRACTION: SEASHELLS

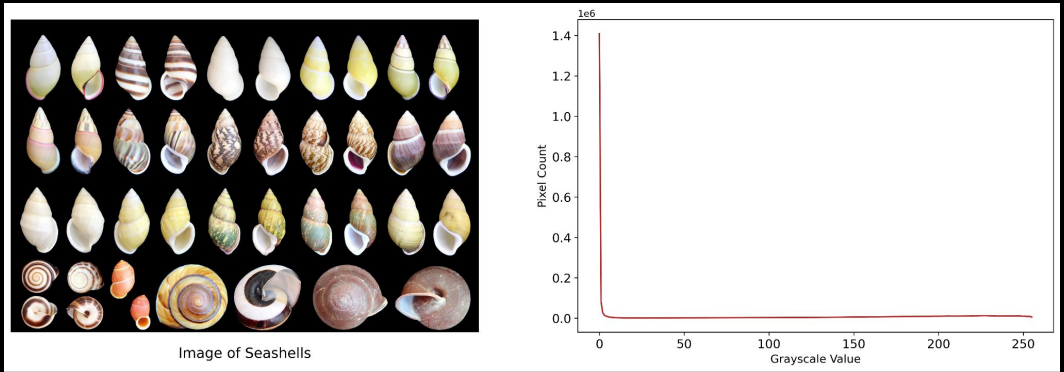



Figure 10. A sample image of seashells and its corresponding grayscale histogram.

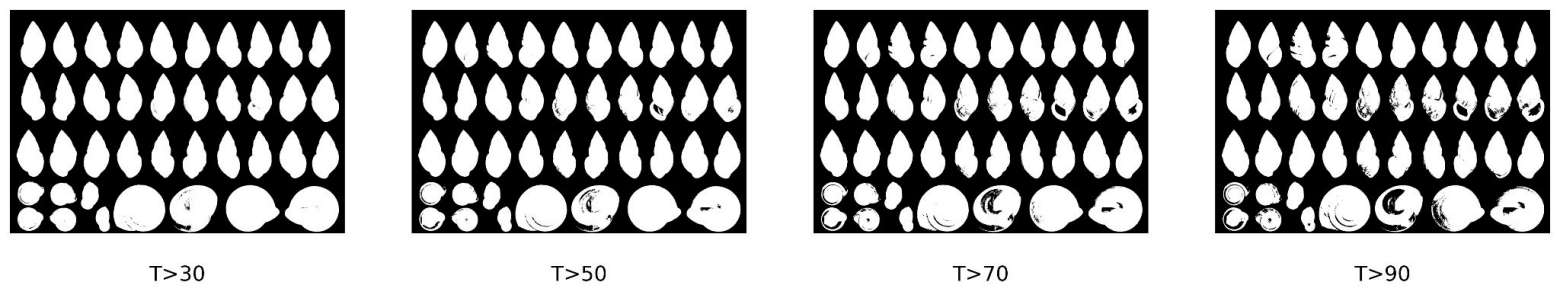



Figure 11. Various thresholding conditions for image segmentation. T>30 was chosen to avoid further sensitivities in morphological operations.

# FEATURE EXTRACTION: SEASHELLS

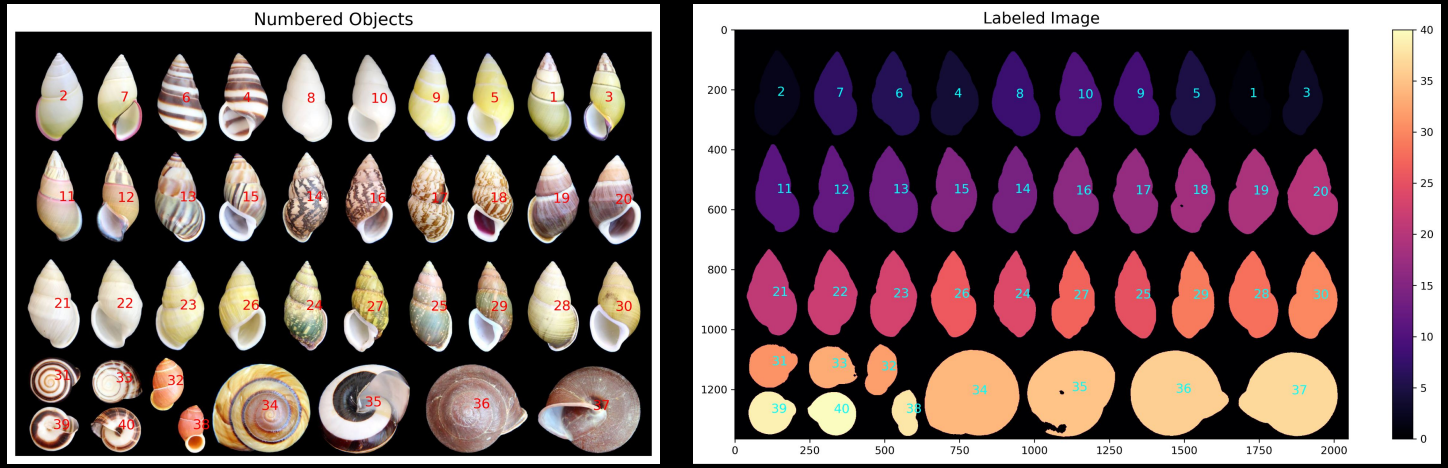



Figure 12. Labeled images after morphological operations.

Again, we repeat the same set of morphological operations as before: morphological closing then opening. **The algorithm seemed to work very solidly for the seashells since most of the blobs do not contain holes.**

# FEATURE EXTRACTION: SEASHELLS

| index | perimeter | area | eccentricity | bbox-0 | bbox-1 | bbox-2 | bbox-3 |
|---|---|---|---|---|---|---|---|
| 1 | 709.31075 | 25553 | 0.887987 | 67 | 1648 | 352 | 1790 |
| 2 | 719.49452 | 29213 | 0.85093503 | 68 | 65 | 353 | 219 |
| 3 | 698.82547 | 23992 | 0.89821695 | 68 | 1828 | 352 | 1963 |
| 4 | 733.15137 | 29909 | 0.84085025 | 69 | 654 | 353 | 814 |
| 5 | 720.52395 | 27758 | 0.86431673 | 69 | 1454 | 353 | 1605 |
| 6 | 725.59502 | 28399 | 0.85740477 | 71 | 457 | 353 | 617 |
| 7 | 700.8671 | 26664 | 0.86487524 | 72 | 265 | 352 | 409 |
| 8 | 733.00923 | 30087 | 0.84793048 | 72 | 858 | 356 | 1020 |
| 9 | 711.69553 | 27518 | 0.86070412 | 73 | 1265 | 352 | 1417 |
| 10 | 723.69553 | 29125 | 0.84967294 | 74 | 1070 | 355 | 1225 |
| 11 | 732.72496 | 27177 | 0.88886122 | 382 | 69 | 675 | 216 |
| 12 | 720.58283 | 25416 | 0.89498898 | 386 | 262 | 677 | 399 |
| 13 | 742.22244 | 30815 | 0.85579466 | 387 | 448 | 677 | 612 |
| 14 | 723.83766 | 28218 | 0.86976342 | 387 | 860 | 679 | 1009 |
| 15 | 727.35238 | 29603 | 0.85209509 | 392 | 655 | 675 | 809 |
| 16 | 730.0803 | 28716 | 0.86282922 | 392 | 1062 | 682 | 1215 |
| 17 | 718.90873 | 28173 | 0.8652202 | 395 | 1262 | 680 | 1413 |
| 18 | 750.12193 | 27581 | 0.86704722 | 395 | 1455 | 678 | 1603 |
| 19 | 730.84986 | 31056 | 0.83045164 | 397 | 1648 | 682 | 1815 |
| 20 | 747.193 | 32093 | 0.82261578 | 397 | 1844 | 685 | 2013 |
| 21 | 739.29351 | 30982 | 0.83669602 | 732 | 40 | 1020 | 208 |
| 22 | 737.29351 | 31323 | 0.82035578 | 733 | 244 | 1019 | 410 |
| 23 | 730.56558 | 30143 | 0.84915449 | 736 | 452 | 1025 | 609 |
| 24 | 714.18081 | 26366 | 0.87775988 | 736 | 864 | 1026 | 1005 |
| 25 | 722.72496 | 27092 | 0.89186469 | 736 | 1268 | 1030 | 1406 |
| 26 | 726.0803 | 28586 | 0.86314533 | 737 | 654 | 1026 | 807 |
| 27 | 729.69553 | 27174 | 0.87938013 | 737 | 1057 | 1031 | 1202 |
| 28 | 742.80822 | 31401 | 0.83809172 | 738 | 1649 | 1029 | 1814 |
| 29 | 723.21024 | 27364 | 0.8779638 | 739 | 1459 | 1030 | 1601 |
| 30 | 743.05087 | 31110 | 0.84229646 | 740 | 1848 | 1031 | 2012 |
| 31 | 511.91378 | 17452 | 0.49072184 | 1049 | 48 | 1194 | 210 |
| 32 | 463.00209 | 13523 | 0.8017823 | 1049 | 434 | 1219 | 543 |
| 33 | 535.08535 | 16895 | 0.48290732 | 1056 | 248 | 1196 | 410 |
| 34 | 1016.9331 | 72521 | 0.4538145 | 1068 | 634 | 1354 | 950 |
| 35 | 1188.6804 | 61744 | 0.47600843 | 1070 | 976 | 1358 | 1269 |
| 36 | 995.54329 | 67355 | 0.45274376 | 1070 | 1321 | 1355 | 1649 |
| 37 | 999.30065 | 67485 | 0.51252181 | 1076 | 1681 | 1356 | 2012 |
| 38 | 410.0904 | 10119 | 0.84058826 | 1201 | 524 | 1355 | 612 |
| 39 | 493.38687 | 16893 | 0.40524598 | 1207 | 46 | 1350 | 205 |
| 40 | 496.25693 | 17053 | 0.48899465 | 1209 | 243 | 1354 | 405 |

Figure 13. Extracted features from the labeled seashells.

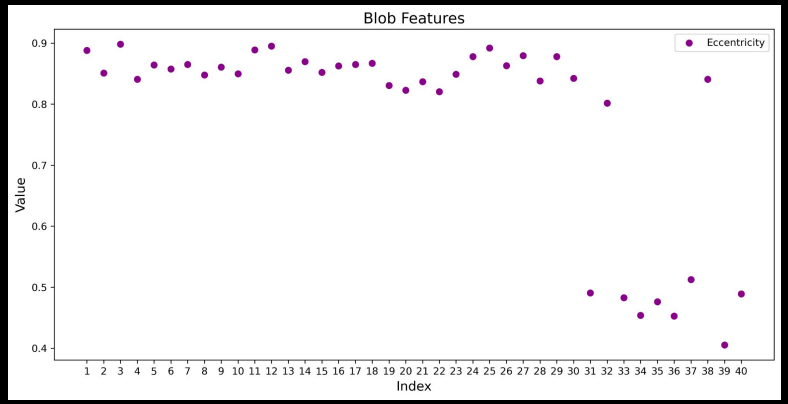



Figure 14. Scatter plot of eccentricity values of the labeled seashells.

Again, we have successfully extracted the following features from the watermelon seeds of the image: **perimeter, area, eccentricity, and the bounding box coordinates.** However, since we are now more interested in the eccentricity values of the seashells, we scatter plot them in Fig. 14. **As we can see, the more globular or rounded seashells (index nos. 31, 33, 34, 35, 36, 37, 39, 40) have a lower eccentricity value which is accurate to how eccentricity values work: the closer it is to zero, the rounder the object is.** The values that are closer to one are more parabolic in terms of curves.

# 03 REFLECTION

# REFLECTION

Overall, this was a great activity!

Combining image segmentation and morphological operation techniques to extract image features has got to be one of the most satisfying computational activities yet. In this activity, I was able to holistically relate each previous techniques and reconcile them into a singular algorithm. I also learned how to label objects of interest and extract the features from each region accordingly.

Besides that, I believe that my results are accurate and stayed true to the topic, with additional cross references and analyses. I also cross-validated my results with my peers. However, if I was given more time, I would implement these morphological operations on more complex images like satellite data and space data.

# SELF-GRADE

**Technical Correctness:** 35/35

I believe that my results are correct through math, research, and through validation with my peers and with my instructors.

**Quality of Presentation:** 35/35

I believe that the quality of my powerpoint is up to par with the course expectations. I constructed the figures as instructed, and exported my data accordingly.

**Self-Reflection:** 30/30

I believe that I have acknowledged and reflected upon the activity well enough. I also have complete citation on the next slide.

**Initiative:** 10/10

I went above and beyond with my data presentation, and included extra analyses for the activities.

# REFERENCES


[1] Chatterjee, S. (2022, July 26). *What is Feature Extraction? Feature Extraction in Image Processing | Great Learning*. Great Learning Blog: Free Resources What Matters to Shape Your Career! https://www.mygreatlearning.com/blog/feature-extraction-in-image-processing/

[2] *How to use the skimage.measure.regionprops function in skimage | Snyk*. (n.d.). Snyk Advisor. https://snyk.io/advisor/python/skimage/functions/skimage.measure.regionprops

[3] *Module: measure — skimage v0.13.1 docs*. (n.d.). https://www.devdoc.net/python/scikit-image-doc-0.13.1/api/skimage.measure.html

[4] *Segmentation — Bioimage analysis fundamentals in Python*. (n.d.). https://jni.github.io/i2k-skimage-napari/lectures/2_segmentation_and_regionprops.html

[5] Singh, A. (2023). 3 Beginner-Friendly Techniques to Extract Features from Image Data using Python. *Analytics Vidhya*. https://www.analyticsvidhya.com/blog/2019/08/3-techniques-extract-features-from-image-data-machine-learning-python/