APP PHY 157 WFY-FX-2

LAB REPORT 7

Feature Extraction from Labeled Blobs

[Source code here!](https://github.com/lovelyandeo/AP157-Computational-Analysis-and-Modeling-in-Physics)

LOVELY L. ANDEO
2020-05405

https://github.com/lovelyandeo/AP157-Computational-Analysis-and-Modeling-in-Physics

Background

Feature extraction in image processing involves the extraction of significant and pertinent details from digital images. It encompasses the examination of image data to identify and capture specific attributes or features that hold value for subsequent analysis or classification purpose

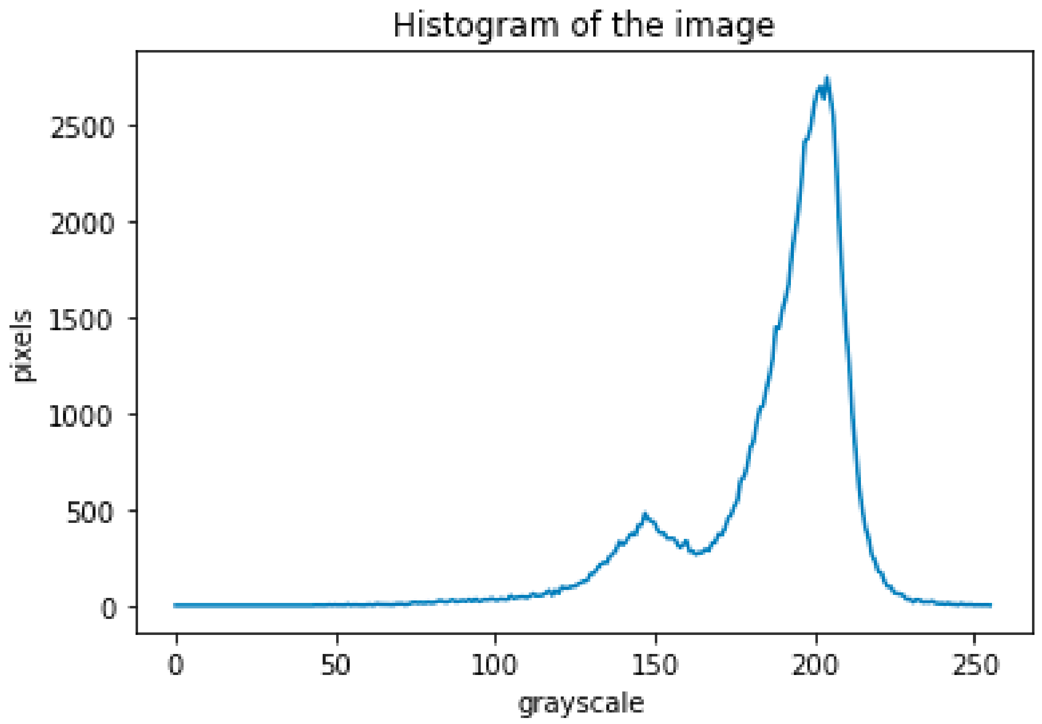
Objectives

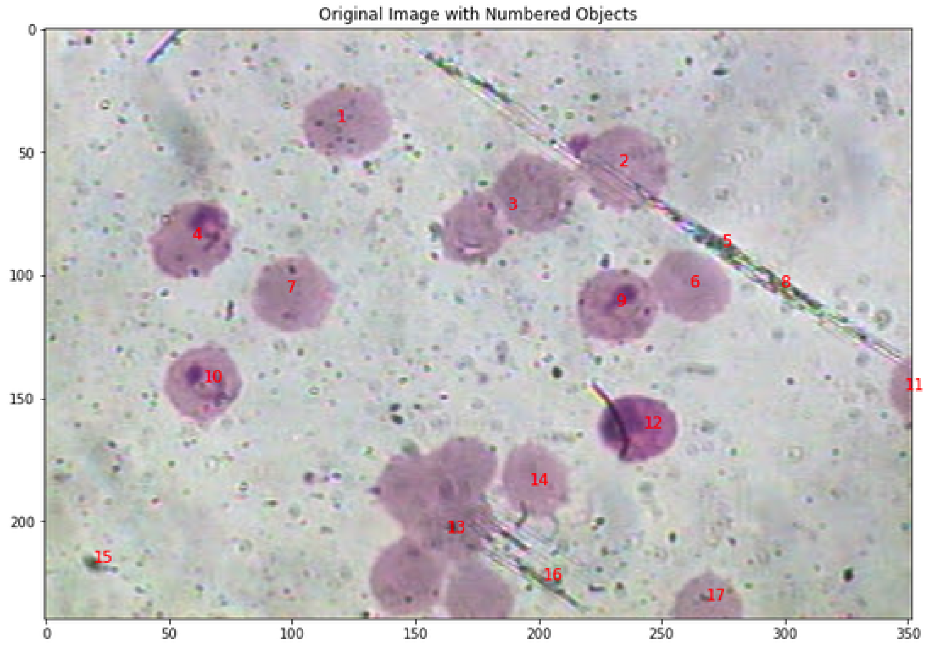
In this activity, we aimed to:

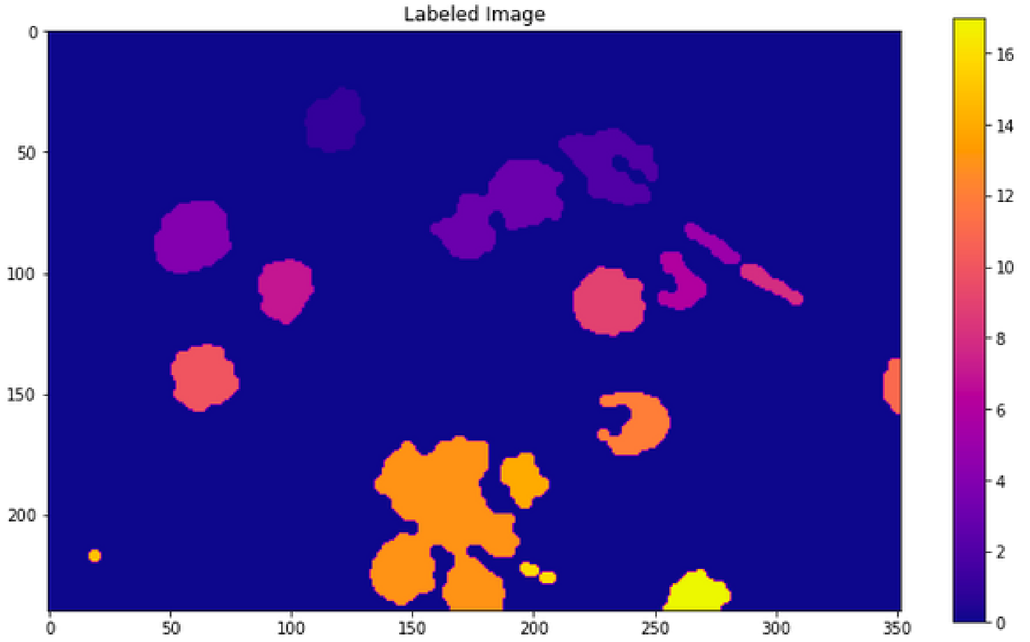
1. Automate feature extraction using *label* and *regionprops*
2. Generate insights from the feature extracted

Results and Analysis

Here, we'll be extracting features of distinct set of blobs:







The initial and crucial step in feature extraction is image segmentation. In the case of the malaria cell image, I employed histogram-based thresholding and applied closing and opening morphological operations to enhance the image's cleanliness. The resulting segmented image is displayed in the color bar below. Additionally, the blobs in the original image were labeled, albeit some noise or extraneous details remained. These segmented blobs can provide valuable features, allowing us to extract meaningful insights from them.

Overlapping blobs

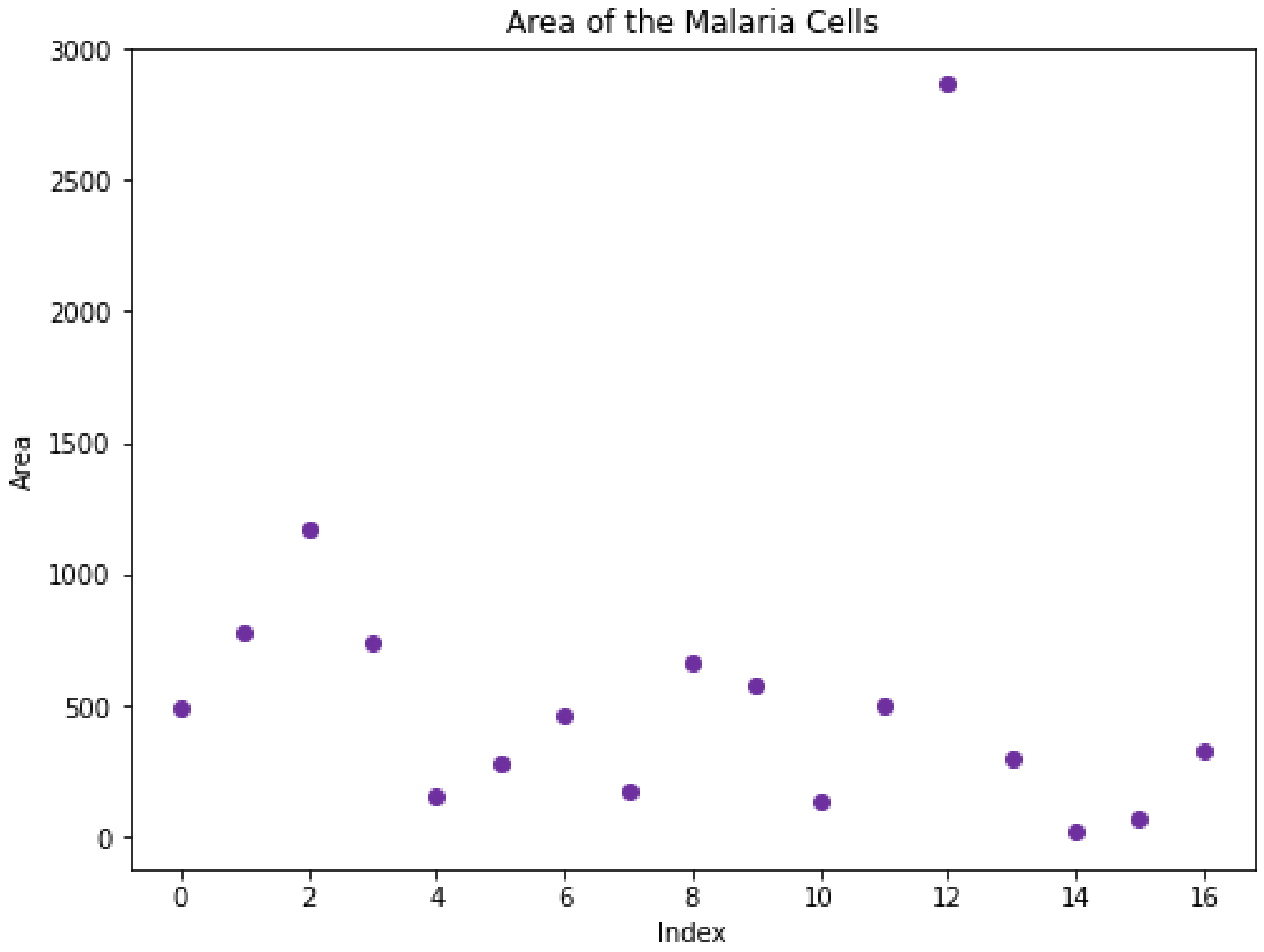
	area	convex_area	bbox_area	major_axis_length	minor_axis_length	eccentricity
0	490	533	675	28.102333	22.633277	0.592748
1	781	961	1312	41.883026	29.097612	0.719266
2	1168	1445	2255	62.116422	27.938644	0.893140
3	740	767	960	33.789982	28.067269	0.556812
4	152	176	391	29.183129	6.943041	0.971287
5	279	360	480	26.681029	18.006977	0.737911
6	458	483	621	26.618686	22.237049	0.549654
7	170	194	442	32.176504	7.090510	0.975418
8	661	698	870	30.268746	28.025181	0.377821
9	578	616	784	27.440047	27.226426	0.124537
10	131	136	161	23.435828	7.512313	0.947232
11	500	651	780	29.555802	28.100532	0.309922
12	2861	3783	4464	77.784286	58.606449	0.657507
13	296	325	460	22.750008	17.386934	0.644908
14	21	21	25	5.089672	5.089672	0.000000
15	69	86	135	18.243447	5.667019	0.950530
16	332	356	476	26.765692	16.706460	0.781285

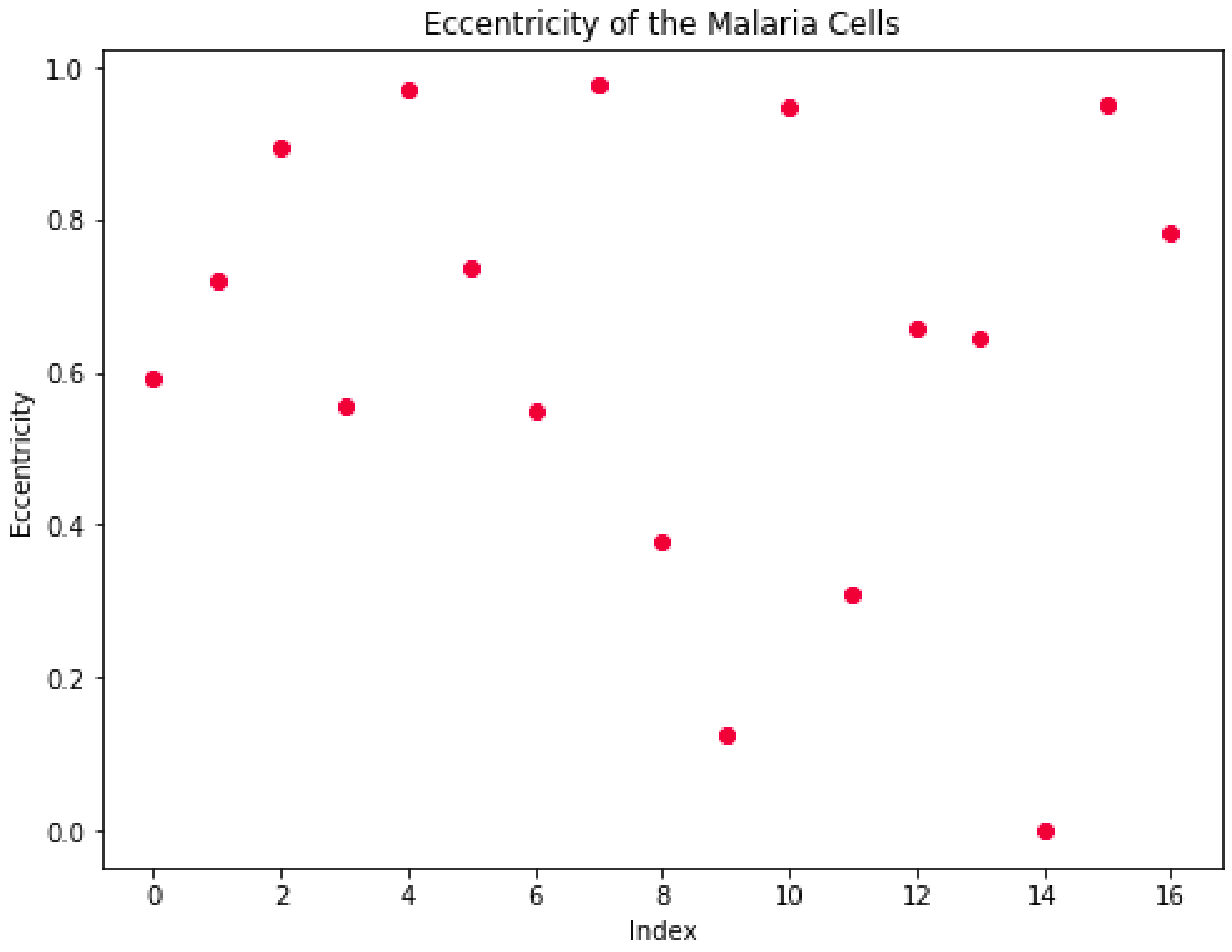
On the left side of the display, we have a table showcasing the extracted features from the labeled blobs in the malaria image. Each row in the table corresponds to an individual blob within the image. The listed features include properties related to the spatial size of the objects or blobs, such as their area, convex area, and the lengths of their minor and major axes. These features are particularly useful for distinguishing between small and large objects within the image.

Additionally, we have the eccentricity feature, which provides a measure of how stretched or elongated an object is in comparison to a perfect circle. The eccentricity is calculated based on the lengths of the major and minor axes. In the context of the malaria image, the eccentricity feature can be utilized to identify cells that are extensively infected or exhibit distinct characteristics. The eccentricity values range from 0 to 1, with 0 representing a perfect circle. As observed in the table, the cells in the image display varying degrees of elongation.

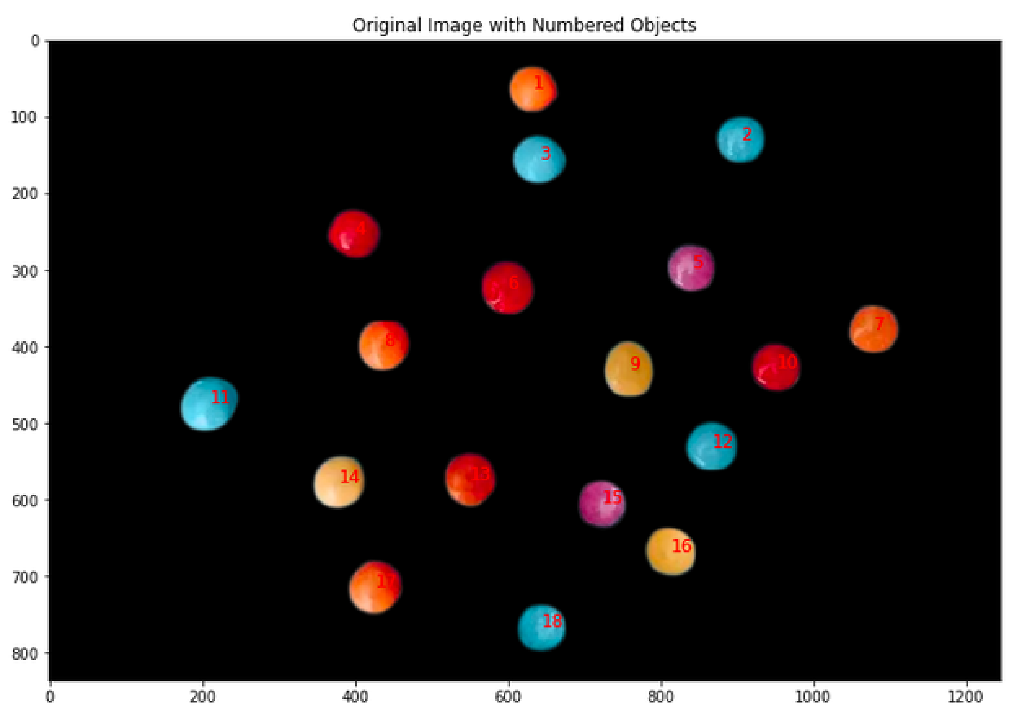
It's important to note that there is a discrepancy in indexing between the table and the labeled image. For example, the 0 index in the table corresponds to the label 1 in the labeled image, and so on.

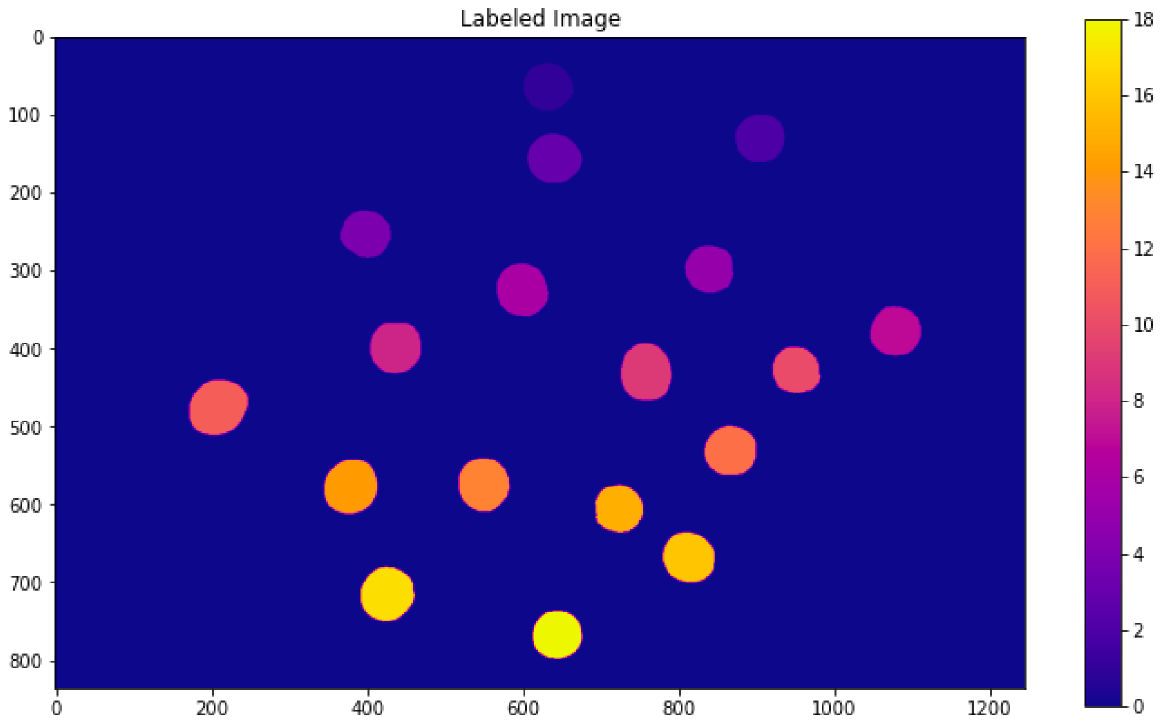
To facilitate better interpretation, these features are often visualized through graphs. In the case of the area feature, plotting the data allows us to easily grasp the distribution of blob sizes. Notably, we can observe an outlier with an exceptionally large area, which corresponds to the clump of blobs or cells with an index of 12. In the eccentricity feature, we can also infer that a lot of the blobs' shapes have deviated from the perfect circle.





When conducting data analysis, the presence of such a clump can pose challenges. One approach is to exclude it from the analysis entirely. Alternatively, we may consider averaging the area over the number of cells within the clump, depending on the specific goals of the analysis. Making this decision will be influenced by the particular insights or patterns we seek to derive from the data.

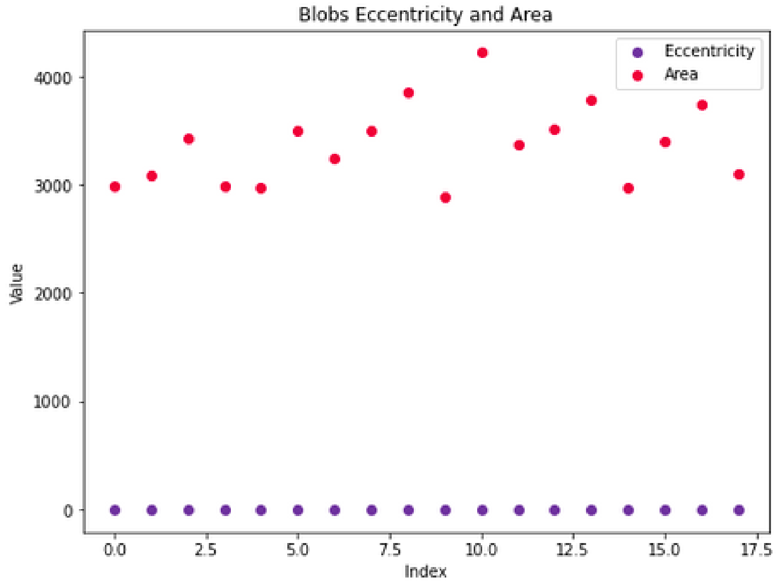




	bbox-0	bbox-1	bbox-2	bbox-3	area	eccentricity
0	34	602	94	666	2994	0.283975
1	100	874	160	937	3086	0.272741
2	124	607	187	677	3426	0.412869
3	223	367	283	431	2988	0.360770
4	268	809	328	871	2979	0.221802
5	291	568	358	634	3500	0.336253
6	346	1047	409	1112	3239	0.284058
7	367	405	432	470	3499	0.214518
8	393	727	466	792	3851	0.476631
9	398	922	457	983	2892	0.354445
10	440	173	511	248	4221	0.505860
11	499	834	562	900	3373	0.310114
12	541	519	609	584	3521	0.322450
13	543	346	612	414	3793	0.388686
14	575	694	636	755	2970	0.228329
15	636	781	700	847	3399	0.381885
16	680	393	750	462	3745	0.314622
17	737	614	798	677	3096	0.267292

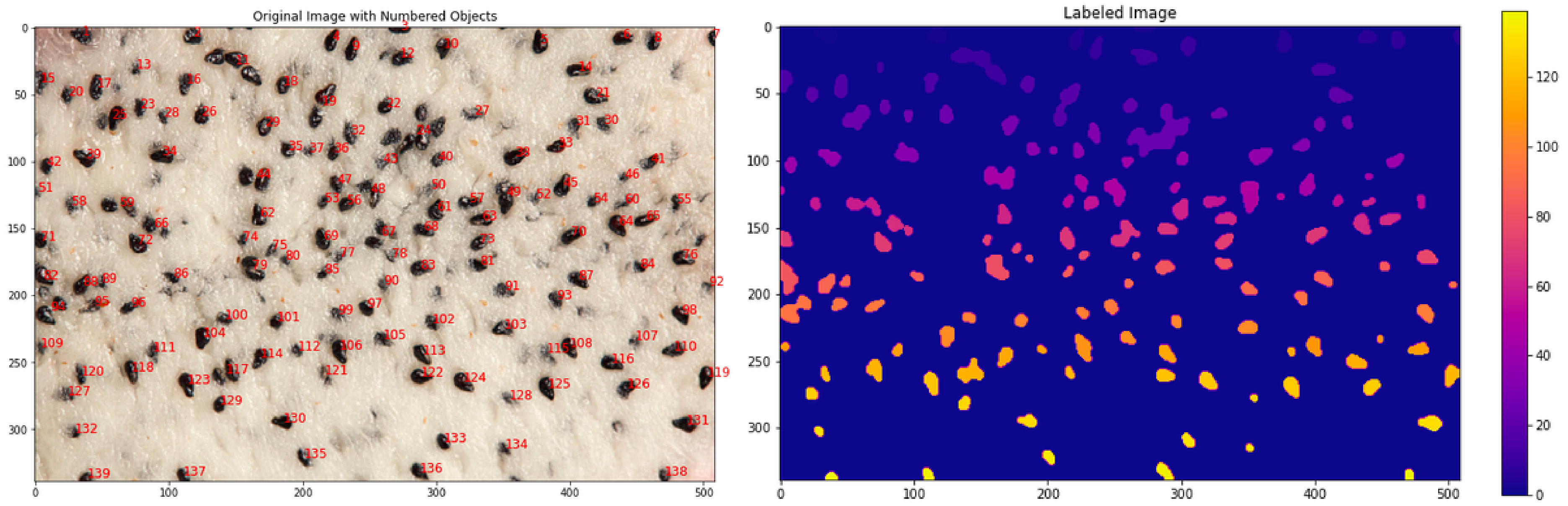
In this particular case, I utilized an image featuring candies that are separate and non-overlapping. Analyzing the eccentricity of these objects reveals values close to 0, indicating minimal elongation in their shapes, as visually observed. Furthermore, I extracted the area of each candy within the image, and the graph depicts a distribution where most candies exhibit a similar size. In contrast to the previous image, it is evident that there are no noticeable outliers in the current image. This can be attributed to the fact that the objects, in this case, the candies, are well-separated, maintaining independent and distinct shapes within the image.

Understanding the area and eccentricity of candies in an image can have various applications beyond visual analysis. For instance, companies might also find it useful as they decide and discuss quality control, packaging and pricing, and sorting of their products such as this.



Separated blobs

Granulated blobs

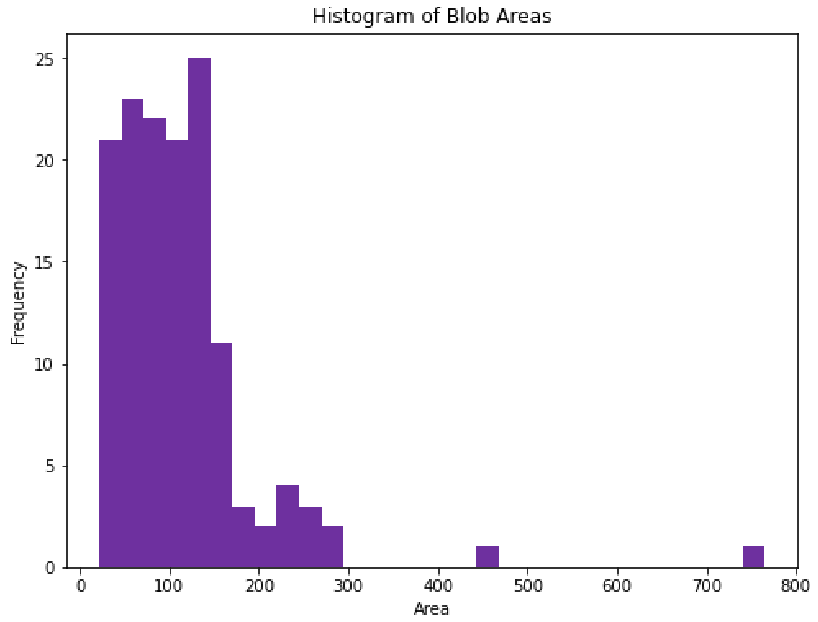


Driven by my fondness for dragonfruit, I have intentionally chosen this particular image to demonstrate the efficacy of feature extraction on small and granulated objects or blobs. Examining the labeled image reveals that there are instances of overlaps and seeds that may appear small due to being partially concealed within the fruit's flesh. Hence, this image serves as an excellent test case to evaluate the capabilities of this feature extraction technique.

One of the important features that I extracted is the area of the seeds. This particular feature holds significant importance as it provides valuable insights into the size and distribution of the seeds within the dragonfruit. By plotting this information in a histogram, we observe a notable concentration of values, enabling us to estimate the average area of the seeds. Knowing this can also aid in quality control, grading, or selection processes.

	area	bbox-0	bbox-1	bbox-2	bbox-3	convex_area	bbox_area	eccentricity
0	174	0	26	13	44	180	234	0.755887
1	139	0	111	13	125	146	182	0.394001
2	73	0	265	5	282	75	85	0.954068
3	121	2	217	18	228	129	176	0.856020
4	154	3	372	20	384	161	204	0.751619
...	...	...	...	...	...	...	...	...
134	75	317	197	328	206	76	99	0.774570
135	106	326	282	339	295	113	169	0.855519
136	65	330	107	339	116	67	81	0.687796
137	52	331	468	339	475	52	56	0.502379
138	48	333	34	339	44	48	60	0.788709

139 rows × 8 columns



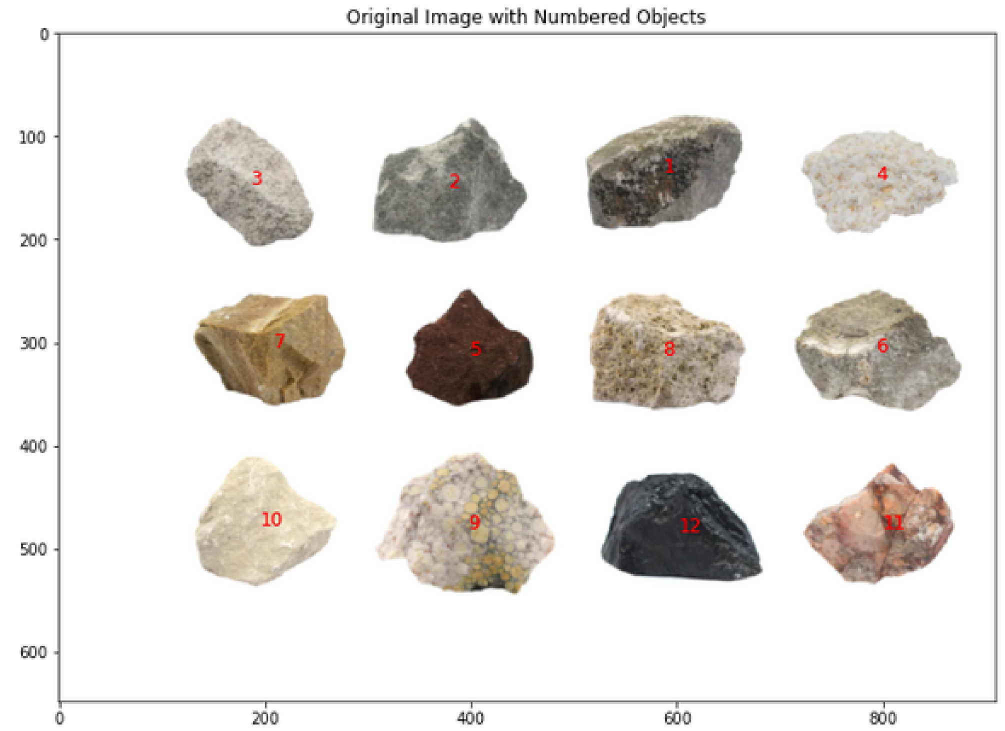
Irregularly-shaped blobs

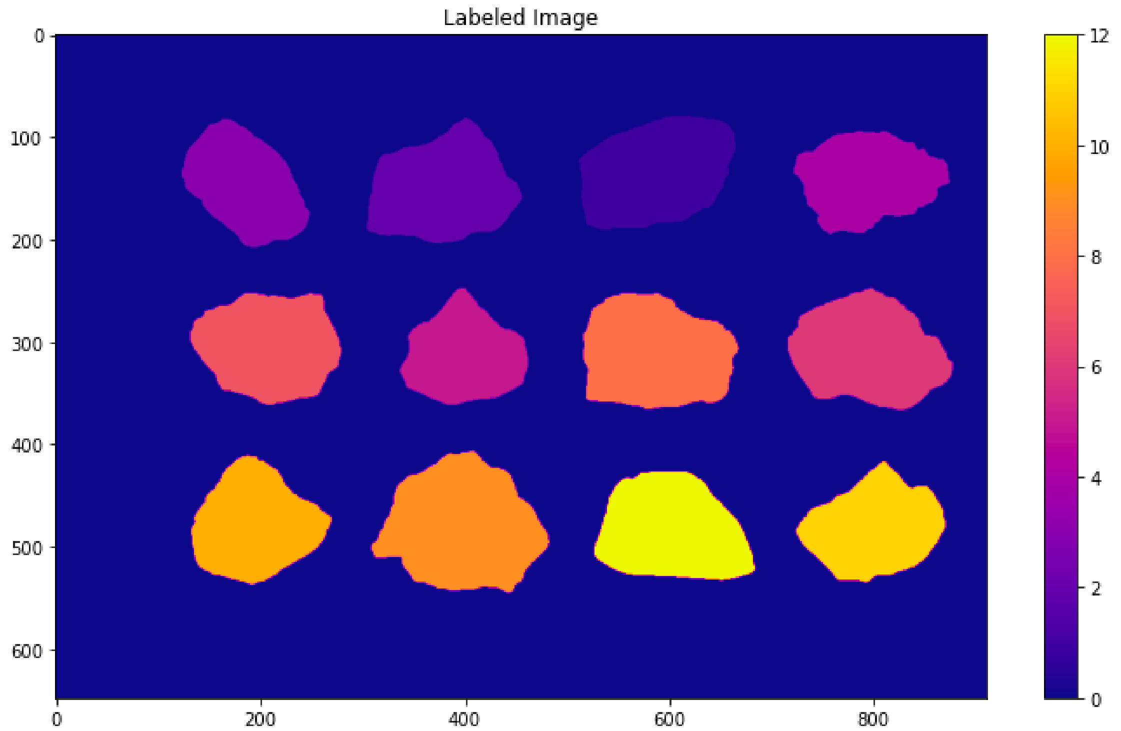
In this particular example, I utilized an image comprising various samples of rocks that exhibit irregular shapes. The segmentation process yielded satisfactory results, as there were no instances of overlapping colors in the original image.

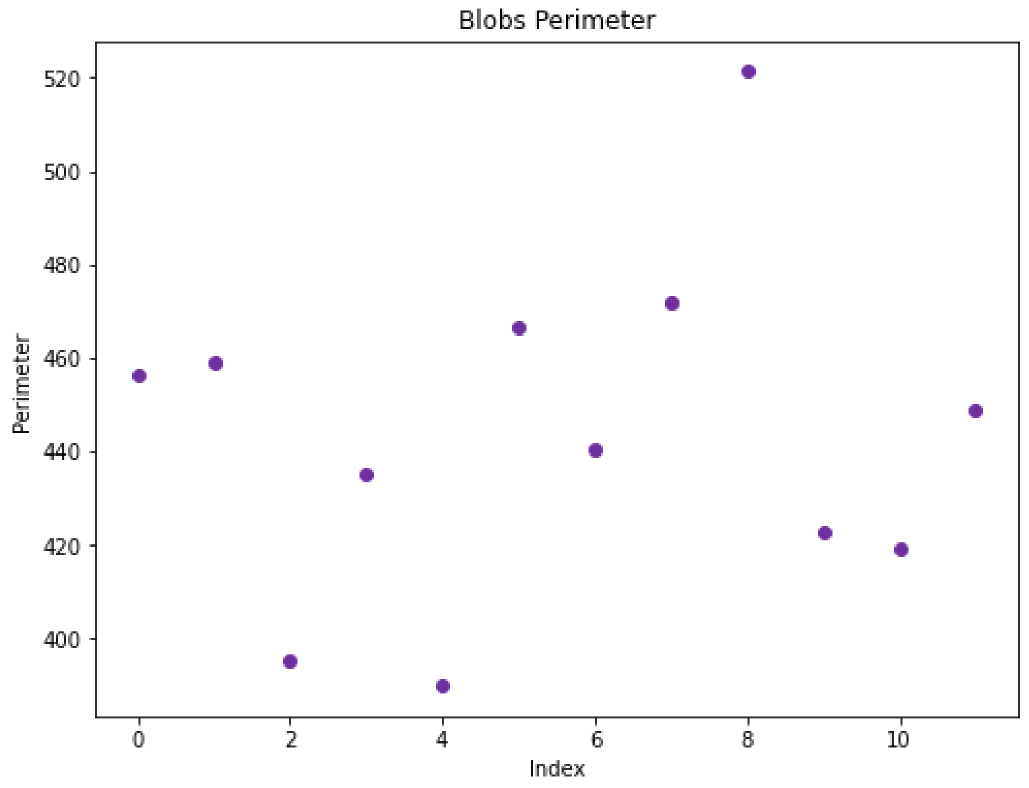
By analyzing the perimeter, we can differentiate between rocks with different shapes, such as round, elongated, irregular, or jagged. This information aids in classifying the rocks and understanding their distinct characteristics. Furthermore, by measuring the length of the rock boundaries, we can estimate their overall sizes, which is valuable for geological studies, rock volume assessment, or other relevant applications.

The compactness of a rock, as inferred from its perimeter, can provide insights into its density or how closely packed the rock material is. A rock with a smaller perimeter relative to its area may indicate a more compact structure, while a rock with a larger perimeter may suggest a more porous or fragmented composition.

	area	convex_area	bbox_area	major_axis_length	minor_axis_length	eccentricity	perimeter
0	13058	13375	16983	164.536086	102.014672	0.784591	456.534055
1	12489	13245	18422	150.446144	109.388133	0.686540	459.060967
2	9875	10146	15625	140.555391	90.426853	0.765569	395.244733
3	10501	11214	15200	145.195815	94.650734	0.758319	434.918831
4	9327	9904	14250	120.540562	102.263759	0.529394	389.889394
5	13174	13777	19159	157.493462	108.437021	0.725220	466.759451
6	12079	12622	16132	144.793096	107.780561	0.667760	440.575685
7	13596	14204	17176	157.495732	112.614429	0.699092	472.048773
8	16077	17090	24012	159.140695	131.333989	0.564739	521.286363
9	11145	11579	17262	127.872516	115.337852	0.431788	422.658946
10	10772	11415	17228	140.353621	100.028965	0.701477	419.043723
11	12521	12780	16642	155.690686	106.591224	0.728886	449.019336

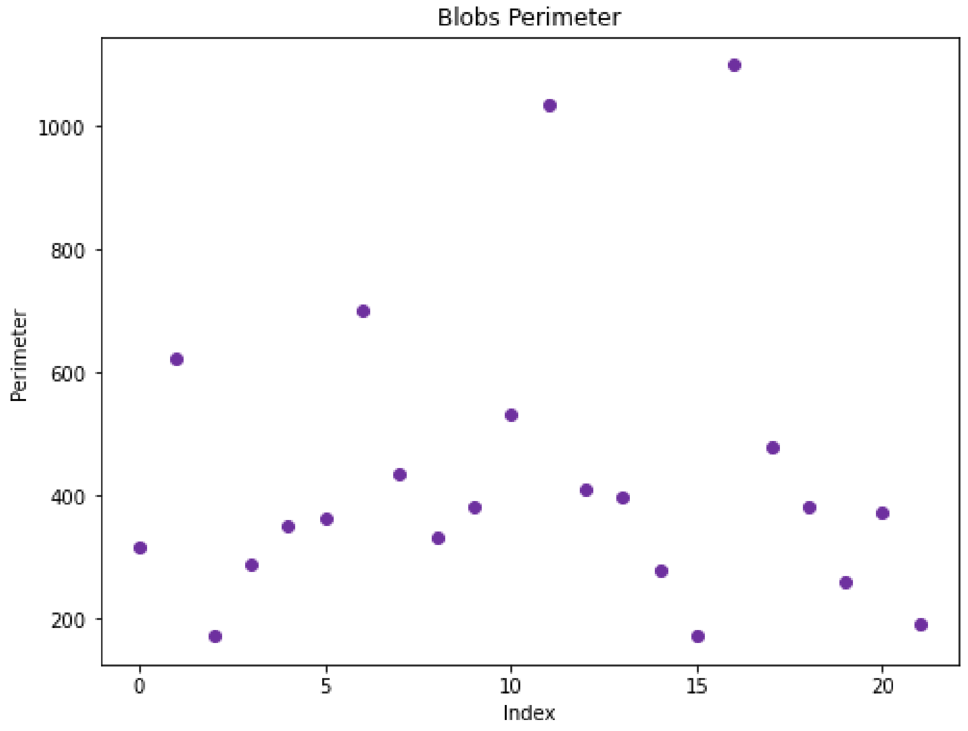


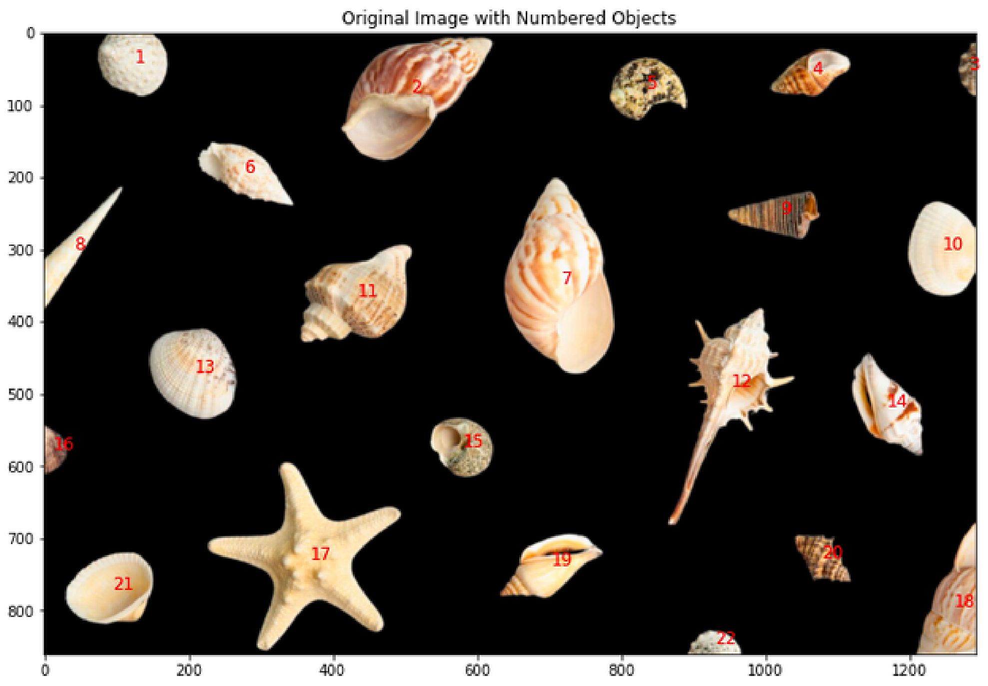


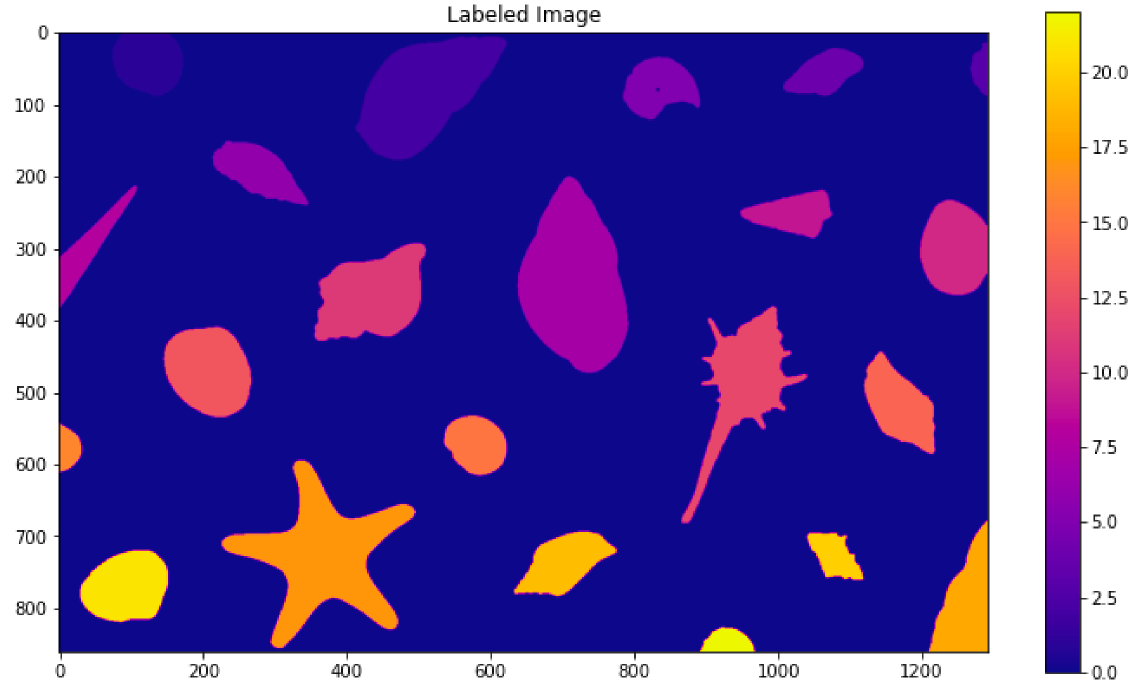


Different-shaped blobs

	area	bbox-0	bbox-1	bbox-2	bbox-3	convex_area	bbox_area	major_axis_length	minor_axis_length	perimeter
0	6980	0	75	88	173	7083	8624	102.045684	87.813917	317.036580
1	21501	6	413	177	623	22024	35910	232.215398	122.474732	622.825469
2	1239	14	1268	88	1292	1320	1776	70.325609	24.740962	172.083261
3	4551	22	1008	88	1119	4745	7326	110.356325	53.710705	287.563492
4	6216	34	785	121	892	7005	9309	106.442582	79.763927	351.019336
5	6456	151	214	240	347	6767	11837	139.182731	60.849627	363.504617
6	27967	200	638	473	792	28482	42042	257.893535	141.678903	699.109740
7	4446	213	0	383	109	4723	18530	186.961347	35.942050	433.327994
8	5159	219	948	286	1076	5415	8576	125.326117	57.488968	331.521861
9	10027	233	1198	365	1292	10126	12408	133.269617	96.108598	381.220346
10	14805	293	356	429	511	16106	21080	175.441640	113.415947	530.859956
11	15453	381	866	682	1041	28138	52675	261.796197	108.827600	1036.489465
12	11829	409	146	535	268	11969	15372	136.042695	111.207365	410.575685
13	8166	444	1120	586	1219	8695	14058	146.579974	73.107685	396.960461
14	5437	533	536	616	623	5539	7221	92.021054	75.518328	276.835570
15	1596	543	0	612	32	1624	2208	64.380256	34.160527	172.225397
16	25762	595	227	856	496	49258	70209	220.116963	216.869234	1099.584920
17	8809	679	1210	861	1292	9139	14924	183.537376	70.895105	477.722871
18	7327	694	633	784	776	7918	12870	138.889526	69.634066	381.546248
19	3174	696	1041	761	1119	3541	5070	89.138155	48.023610	259.580736
20	9462	720	30	820	152	9622	12200	126.363677	96.256496	371.462987
21	1882	827	892	861	968	1936	2584	72.473691	35.550622	190.225397







Lastly, I intentionally used shells with diverse shapes to extract their features. By examining their perimeters, we observed that some shells exhibited larger values, indicating long and large shells within the image.

Analyzing the perimeters of shells aids in species identification, as different mollusk species have distinct shell shapes. Understanding changes in shell shape and size over time also helps researchers infer evolutionary trends and relationships between species. Additionally, variations in shell perimeters provide valuable insights into the environmental conditions that influenced mollusks, including habitat characteristics and resource availability. By analyzing shell perimeters, we can gain a deeper understanding of mollusk species, their evolution, and the environmental contexts in which they thrived.

Key takeaways

- Feature extraction is a valuable technique for extracting meaningful insights from objects within an image.
- The accuracy and usefulness of the extracted features are highly dependent on the preciseness and execution of the object segmentation.

Reflection

Definitely one of the coolest activities so far! As I previously mentioned, while I have worked with features and analyzed them in the past, I had never personally derived those features myself. Engaging in this activity has provided me with a valuable opportunity to experience the process firsthand. Initially, there was a considerable amount of noise present in the extracted features. However, I employed some filtering to refine the results and focused solely on the features corresponding to the labeled blobs or objects of interest. The outcome turned out great! This success has further fueled my enthusiasm to leverage these features for further analysis using machine learning techniques in upcoming activities :)

Self-evaluation

100/100

+ 5 bonus points

I believe I was able to deliver what was required for this lab report. And I also added a few examples to examine feature extraction with different blobs.

References

Here are the materials I used as guide to accomplish this activity:

Soriano, M. (2023). Activity 7- Feature Extraction From Labeled Blobs. https://uvle.upd.edu.ph/pluginfile.php/872182/mod_resource/content/1/Activity%205%20-%20Feature%20Extraction%20%20Image%20Segmentation%20%20%28Part%201%20of%203%29.pdf

Caubalejo, R. (2021). Image Processing - Blob Detection. https://towardsdatascience.com/image-processing-blob-detection-204dc6428dd