ScaDS.AI
DRESDEN LEIPZIG
CENTER FOR SCALABLE DATA ANALYTICS
AND ARTIFICIAL INTELLIGENCE

Feature extraction

Robert Haase

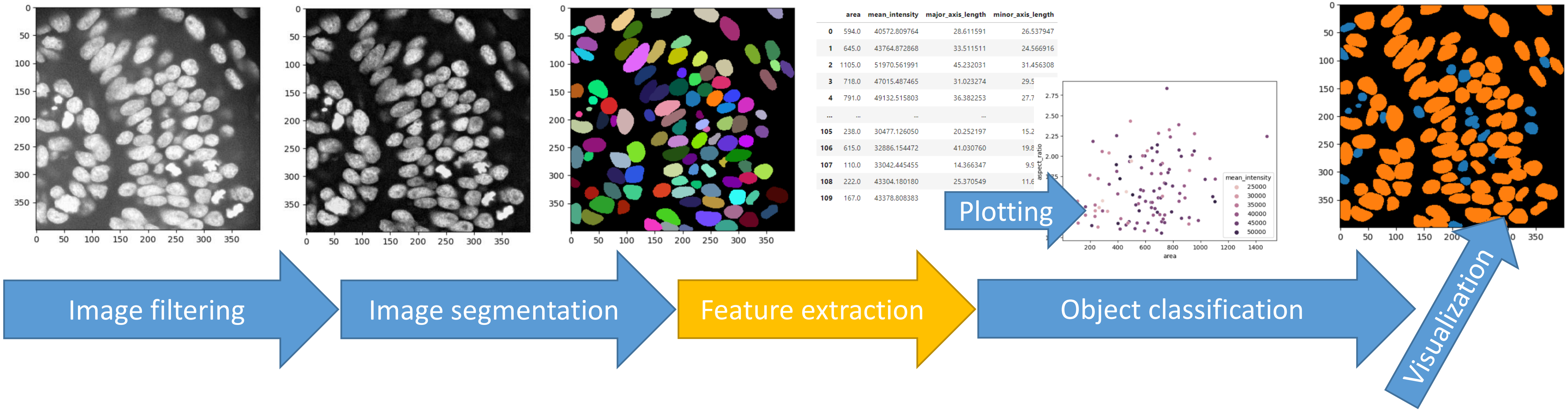
GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Lecture overview: Bio-image Analysis

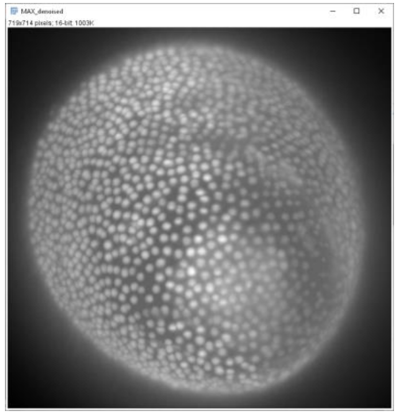
- Image Data Analysis workflows
- Goal: **Quantify observations, substantiate conclusions with numbers**



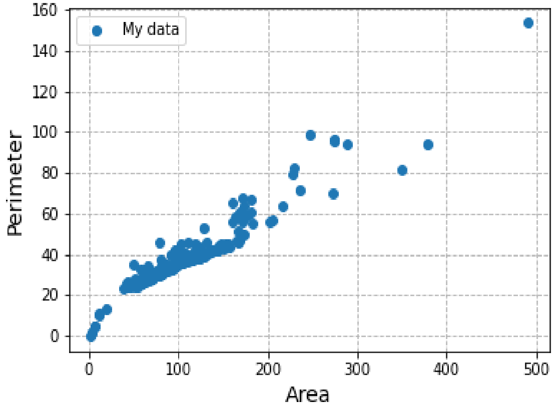
	area	mean_intensity	major_axis_length	minor_axis_length
0	594.0	40572.809764	28.611591	26.537947
1	645.0	43764.872868	33.511511	24.566916
2	1105.0	51970.561991	45.232031	31.456308
3	718.0	47015.487465	31.023274	29.5
4	791.0	49132.515803	36.382253	27.7
...	...	...	...	
105	238.0	30477.126050	20.252197	15.2
106	615.0	32886.154472	41.030760	19.8
107	110.0	33042.445455	14.366347	9.9
108	222.0	43304.180180	25.370549	11.6
109	167.0	43378.808383		

Feature extraction

- Feature extraction is a *late* processing step in image analysis.
- It can be used for images or



Feature Extraction



- or segmented/labelled images

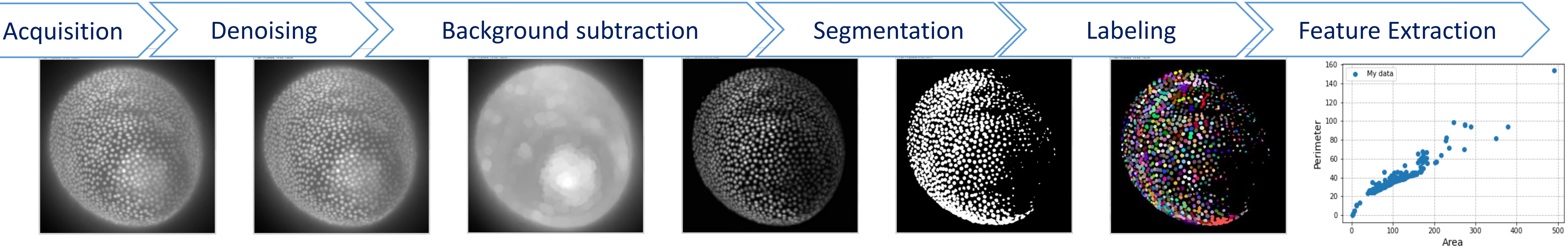


Image data source: Daniela Vorkel, Myers lab, MPI CBG

Feature extraction

- A *feature* is a countable or measurable property of an image or object.
- Goal of feature extraction is finding a minimal set of features to describe an object well enough to differentiate it from other objects.

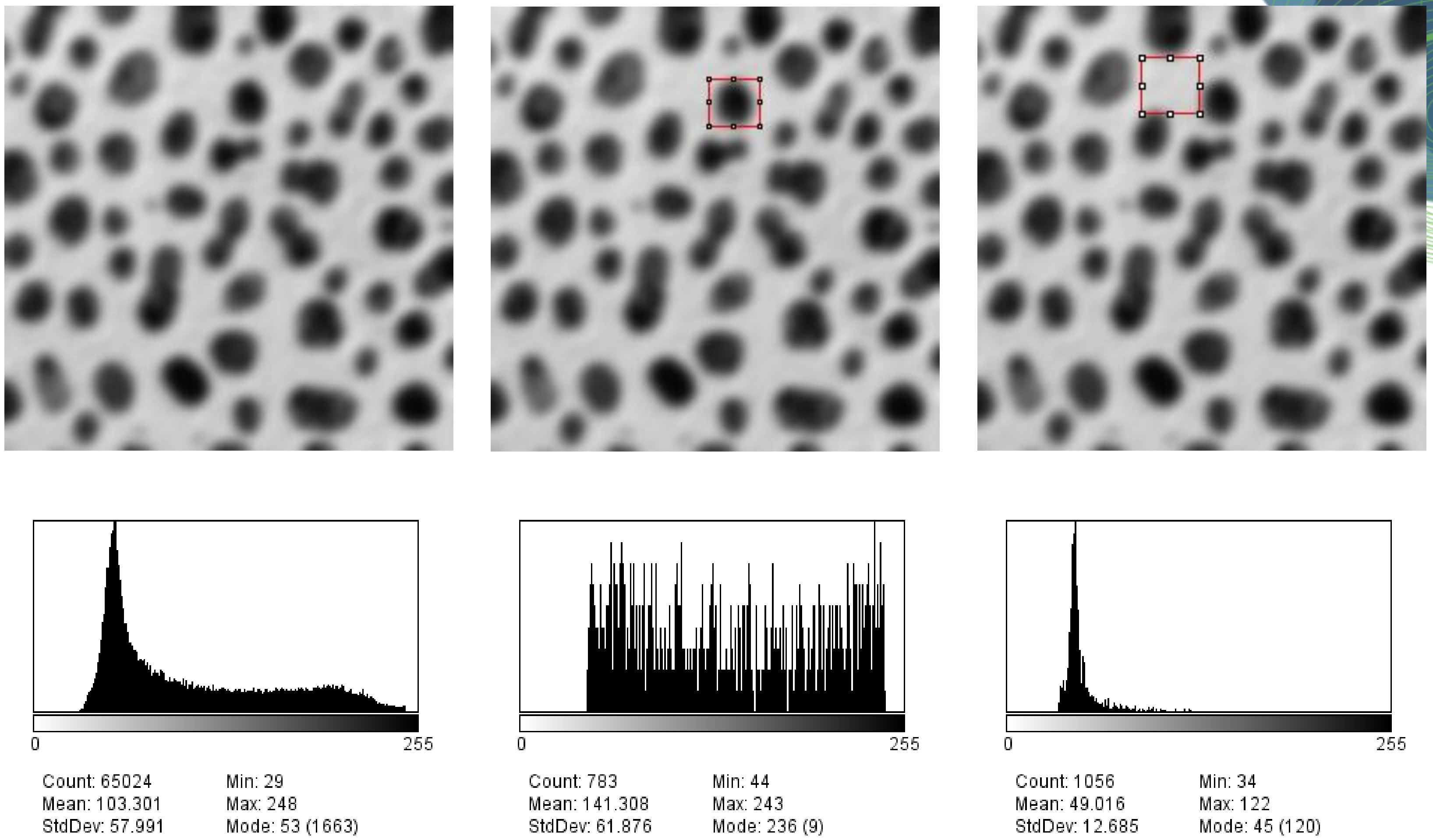
- **Intensity based**
  - Mean intensity
  - Standard deviation
  - Total intensity
  - Textures
- **Mixed features**
  - Center of mass
  - Local minima / maxima
  - Distance to neighbors
  - Average intensity in neighborhood
- **Shape based /spatial**
  - Area / Volume
  - Roundness
  - Solidity
  - Circularity / Sphericity
  - Elongation
  - Centroid
  - Bounding box
- **Spatio-temporal**
  - Displacement,
  - Speed,
  - Acceleration
- **Topological**
  - Number of neighbors
- **Others**
  - Overlap
  - Colocalization

Intensity based features

- Min / max
- Median
- Mean
- Mode
- Variance
- Standard deviation

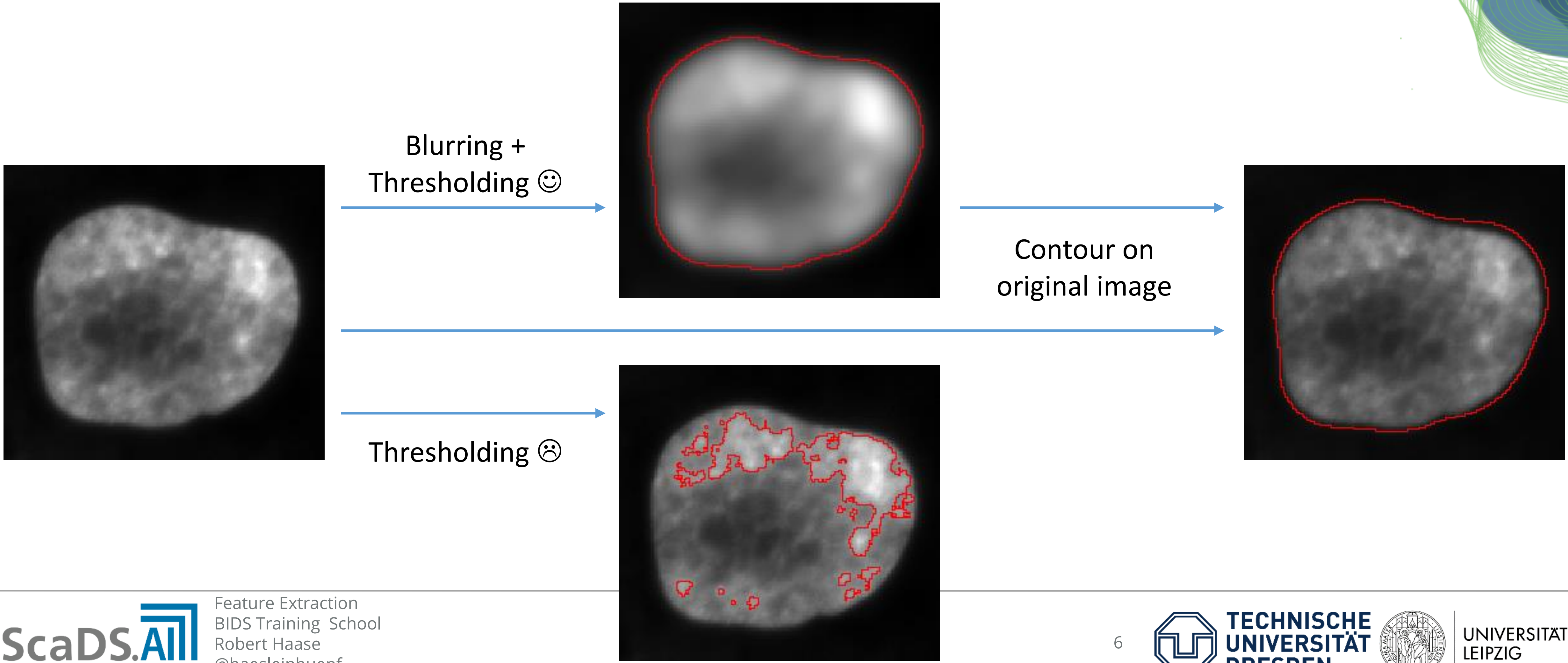
- Can be derived from pixel values
- Don’t take spatial relationship of pixels into account

- See also:
  - descriptive statistics
  - histogram



Reminder: Measure on raw data

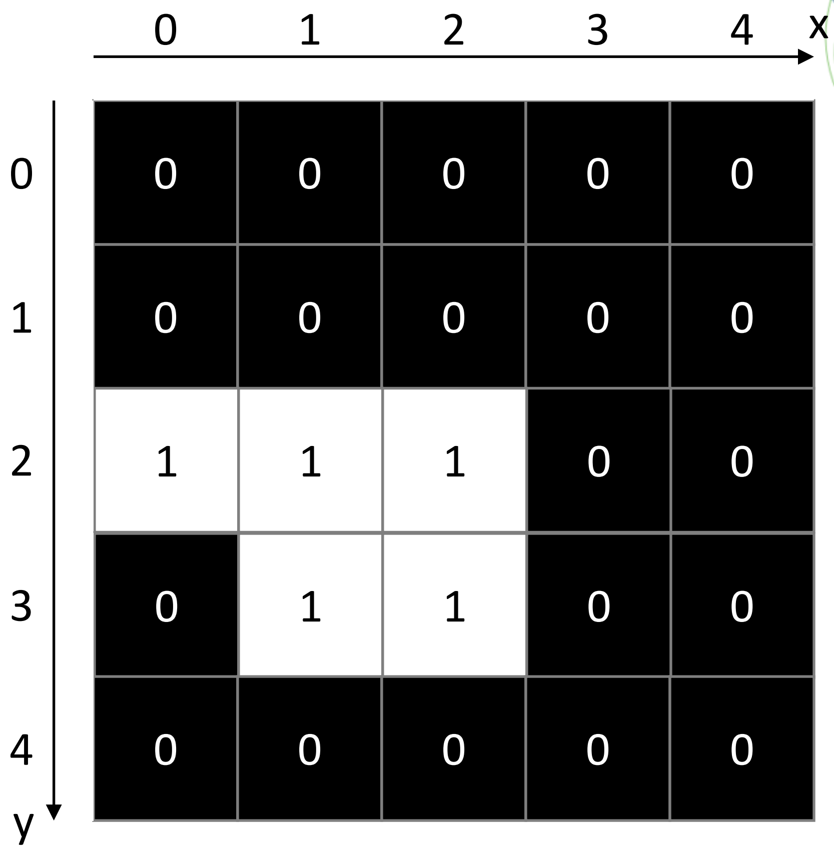
- In case thresholding algorithms outline the wrong structure, blurring in advance may help.
- However: **Do not** continue processing the blurred image, continue with the original!



Bounding rectangle / bounding box

- Position and size of the smallest rectangle containing all pixels of an object
  - x_b, y_b … position of the bounding box
  - w_b … width of the bounding box
  - h_b … height of the bounding box

variable	value
x_b	0
y_b	2
w_b	3
h_b	2



Center of mass

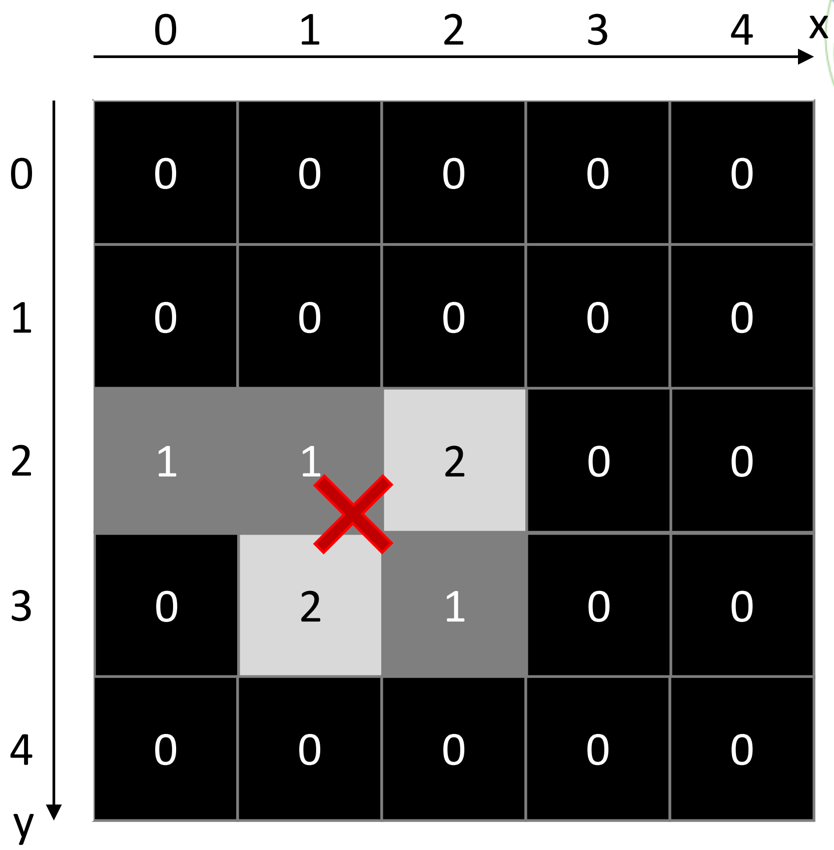
- Relative position in an image weighted by pixel intensities
  - x, y … pixel coordinates
  - w … image width
  - h … image height
  - µ … mean intensity
  - $g_{x,y}$ … pixel grey value
  - x_m, y_m … center of mass coordinates

$$\mu = \frac{1}{wh} \sum_{y=0}^{h-1} \sum_{x=0}^{w-1} g_{x,y}$$

$$x_m = \frac{1}{wh\mu} \sum_{y=0}^{h-1} \sum_{x=0}^{w-1} x\, g_{x,y}$$

“sum intensity”
“total intensity”

$$y_m = \frac{1}{wh\mu} \sum_{y=0}^{h-1} \sum_{x=0}^{w-1} y\, g_{x,y}$$



x_m = 1/7 (1·0 + 1·1 + 2·2 + 2·1 + 1·2) = 1.3

y_m = 1/7 (1·2 + 1·2 + 2·3 + 2·2 + 1·3) = 2.4

Center of geometry / centroid

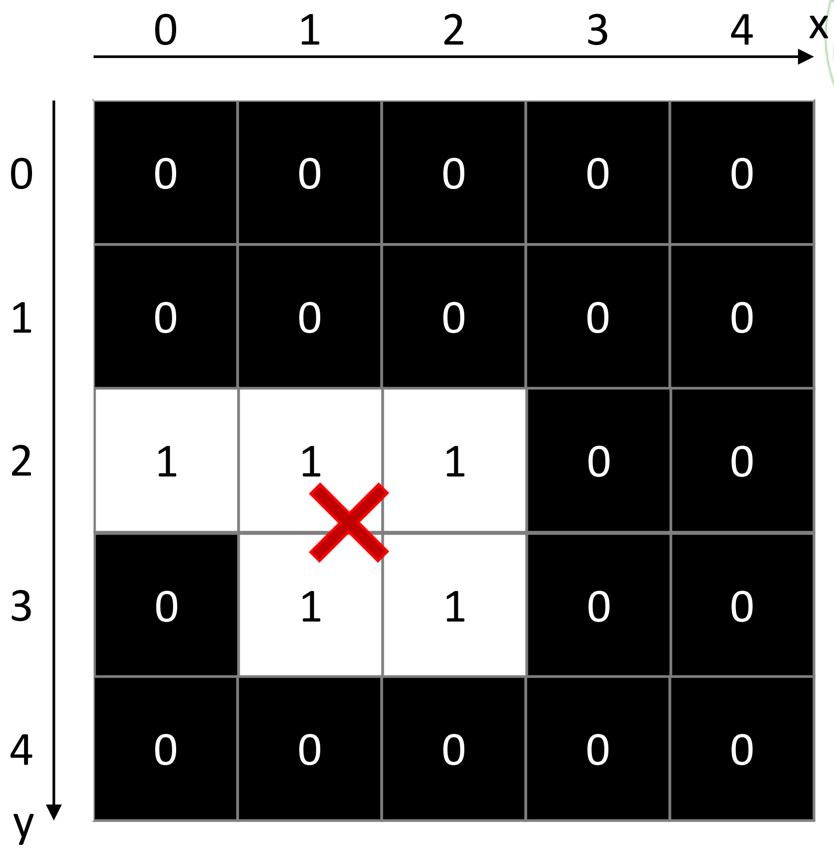
- Relative position in an image weighted by pixel intensities
- Special case of center of mass for binary images
  - x, y … pixel coordinates
  - w … image width
  - h … image height
  - µ … mean intensity
  - $g_{x,y}$ … pixel grey value, integer in range [0;1]
  - x_m, y_m … center of mass coordinates

$$\mu = \frac{1}{wh} \sum_{y=0}^{h-1} \sum_{x=0}^{w-1} g_{x,y}$$

$$x_m = \frac{1}{wh\mu} \sum_{y=0}^{h-1} \sum_{x=0}^{w-1} x \, g_{x,y}$$

$$y_m = \frac{1}{wh\mu} \sum_{y=0}^{h-1} \sum_{x=0}^{w-1} y \, g_{x,y}$$

Number of white pixels

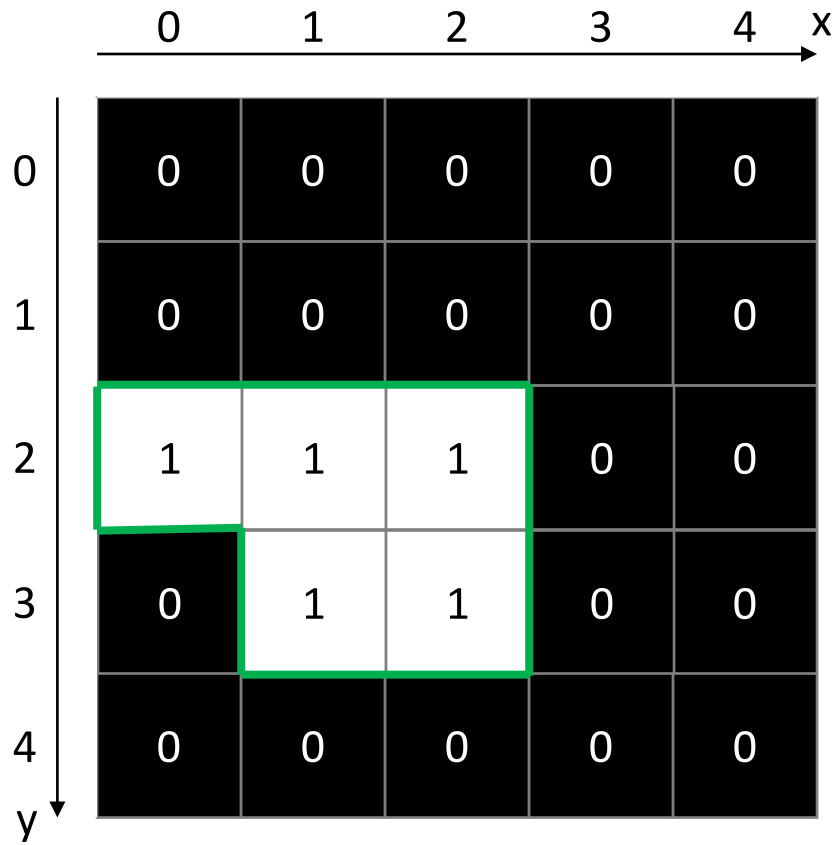


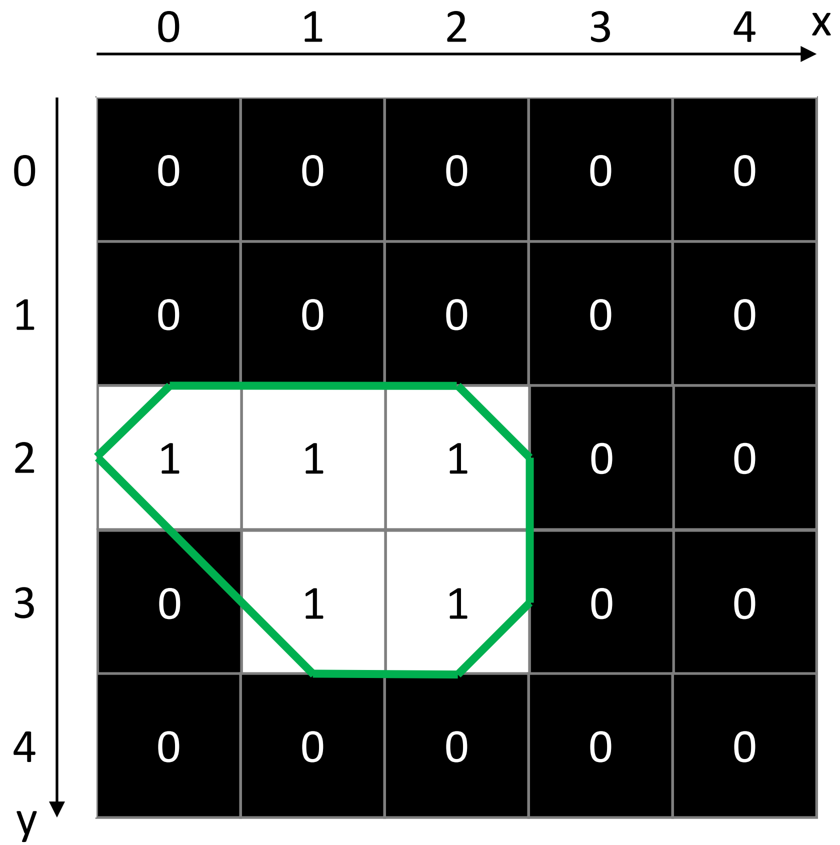
x_m = 1/5 (1·0 + 1·1 + 1·2 + 1·1 + 1·2) = 1.2

y_m = 1/5 (1·2 + 1·2 + 1·3 + 1·2 + 1·3) = 2.4

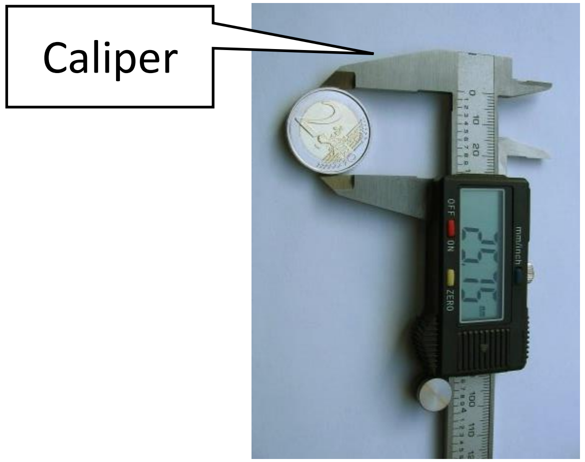
Perimeter

- Length of the outline around an object
- Depends on the actual implementation

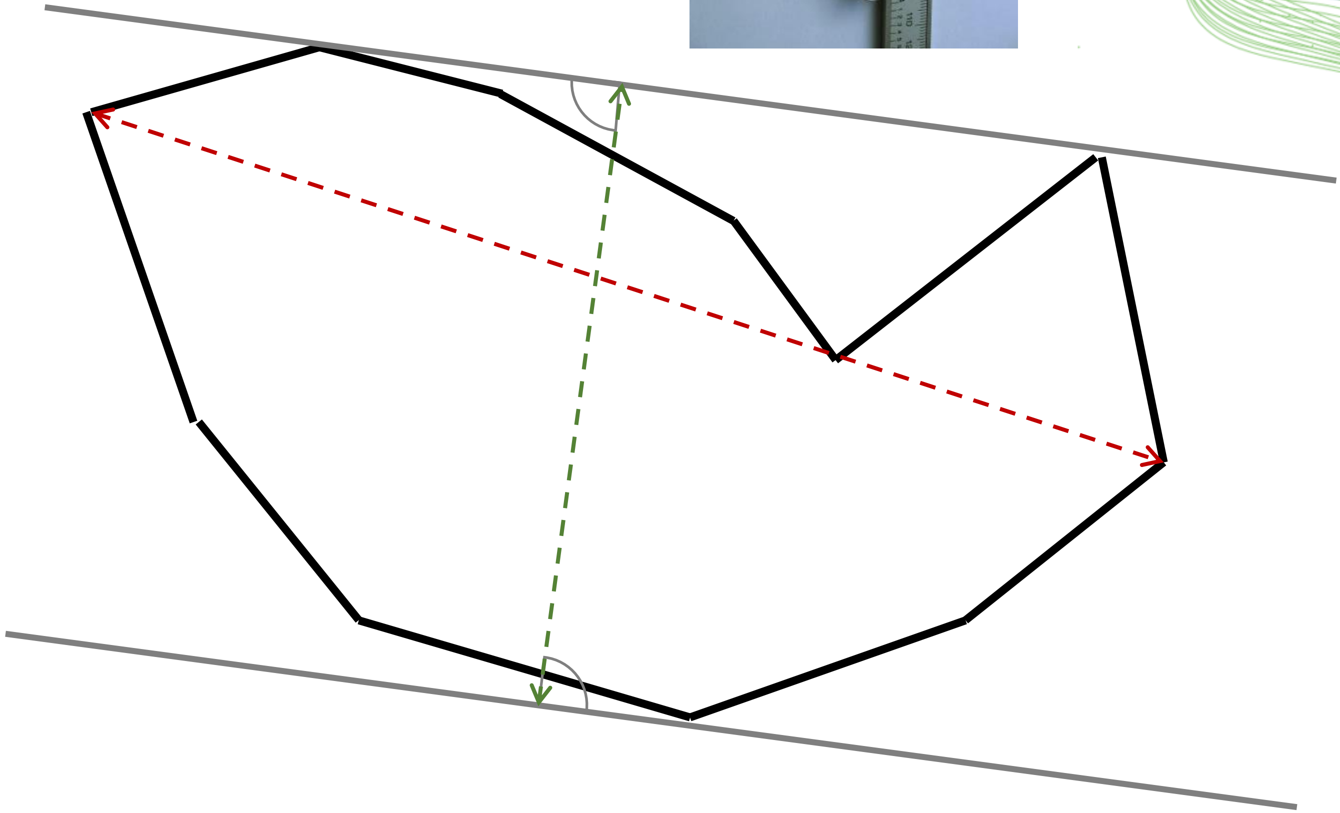




Feret's diameter



- **Feret's diameter** describes the maximum distance between any two points of an outline.
- The **minimum caliper** ("Minimum Feret") describes the shortest distance, the object would fit through.
- Feret and Minimum Feret do not need to be perpendicular to each other!



https://en.wikipedia.org/wiki/Feret_diameter#/media/File:DigitalCaliperEuro.jpg Image license: Public domain

Feret's diameter

- Feret's diameter (L.R. Feret, 1931) is often cited, but impossible to read online …
- The term "Feret's Diameter" was established in the 1970s

LA GROSSEUR DES GRAINS DES MATIÈRES PULVÉRULENTES

par

L. R. FERET

Ancien Elève de l'Ecole Polytechnique,
Chef du Laboratoire des Ponts et Chaussées de Boulogne-sur-Mer
BOULOGNE-SUR-MER (France)

SOMMAIRE

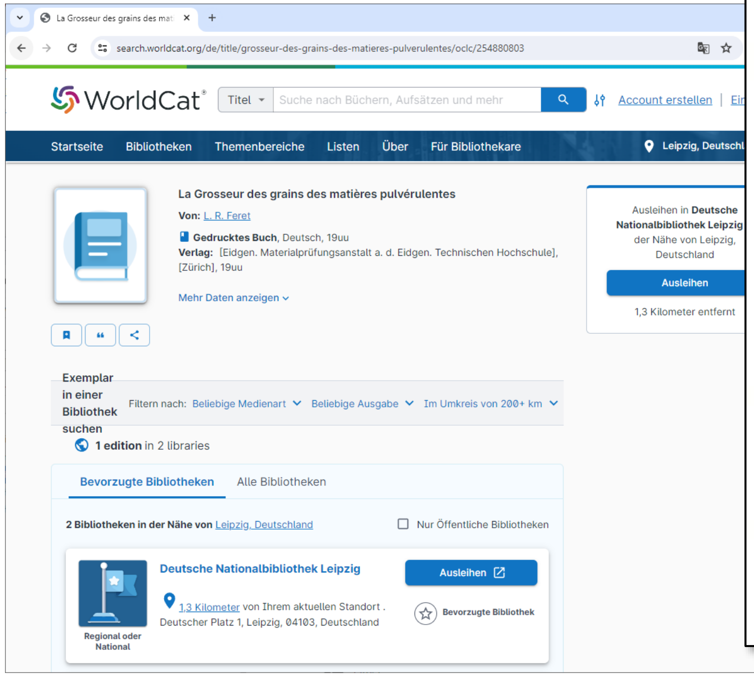
AUSZUG

DIE KORNGRÖSSE PULVERFÖRMIGER STOFFE

Zur Kennzeichnung der linearen Grösse von Körnern einer bestimmten Kornfraktion, unabhängig von der Grössenordnung und dem zur Abscheidung benutzten Verfahren, scheint am geeignetsten das Mittel aus einer genügenden Anzahl von Messungen des Abstandes je zweier an entgegengesetzten Seiten des Umrisses der Körner gelegter Tangenten, die parallel zu einer beliebigen, aber für alle Messungen gleichen Richtung verlaufen. Die Messung geschieht unbahängig von der Lage der Körner zu der gewählten Richtung der Tangenten.

Auf Grund des so erhaltenen Mittelwertes, der als *mittlere Kornbreite* bezeichnet wird, baut Verfasser mittelst geometrischer Progressionen, die auf der Normalreihe von *Renard* beruhen, eine Einteilung nach Kornbreiten für das ganze Gebiet der gekörnten und staubförmigen Materialien auf. Die verschiedenen Kornklassen sind gekennzeichnet durch die Grenzwerte der entsprechenden *mittleren Kornbreiten* und ausserdem durch Namen, die so ausgewählt wurden, dass sie leicht in alle Sprachen eingeführt werden können.

Diese Einteilung wird vervollständigt durch eine Definition der *Kornzusammensetzung* unter Hinweis auf die Bestimmung der letzteren, entweder, ob diese Bestimmung in strenger Uebereinstimmung mit der allgemeinen Einteilung oder auf einfachere Weise im Hinblick auf gewisse gebräuchliche Anwendungen geschieht.

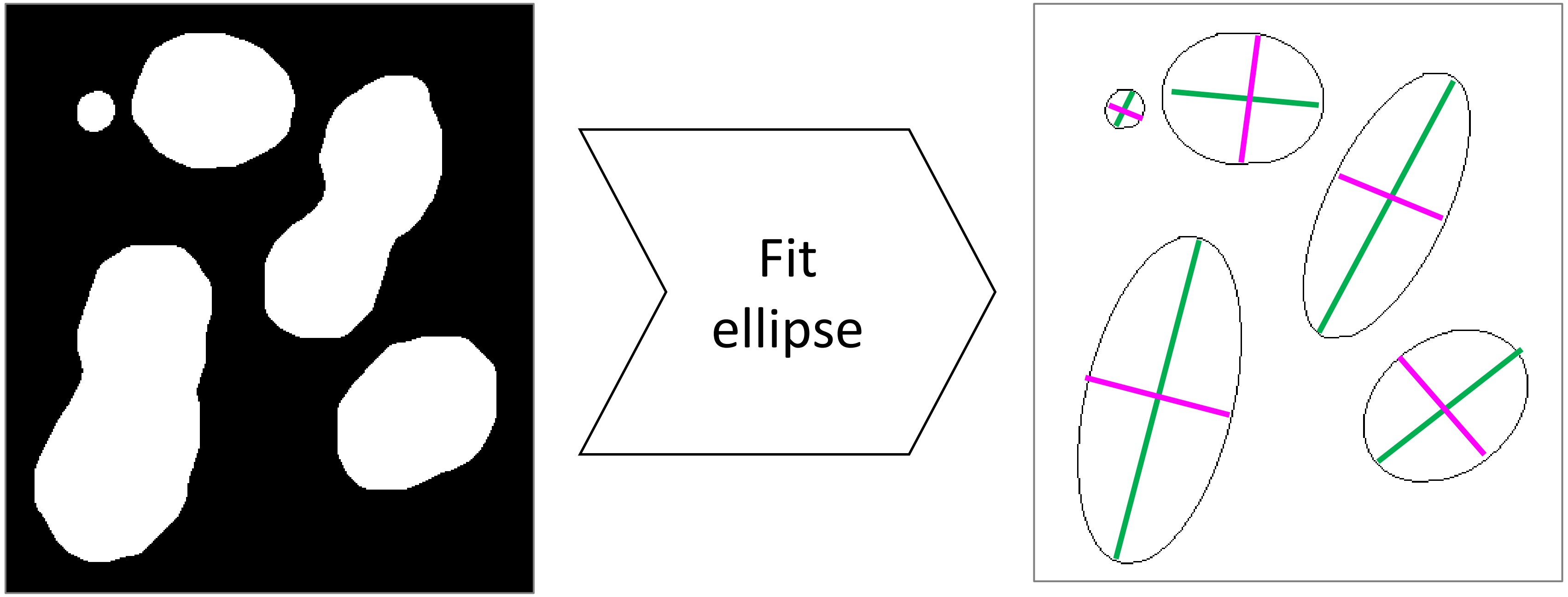


https://search.worldcat.org/de/title/grosseur-des-grains-des-matieres-pulverulentes/oclc/254880803

Minor / major axis

- For every object, find the optimal ellipse simplifying the object.
- Major axis … long diameter
- Minor axis … short diameter

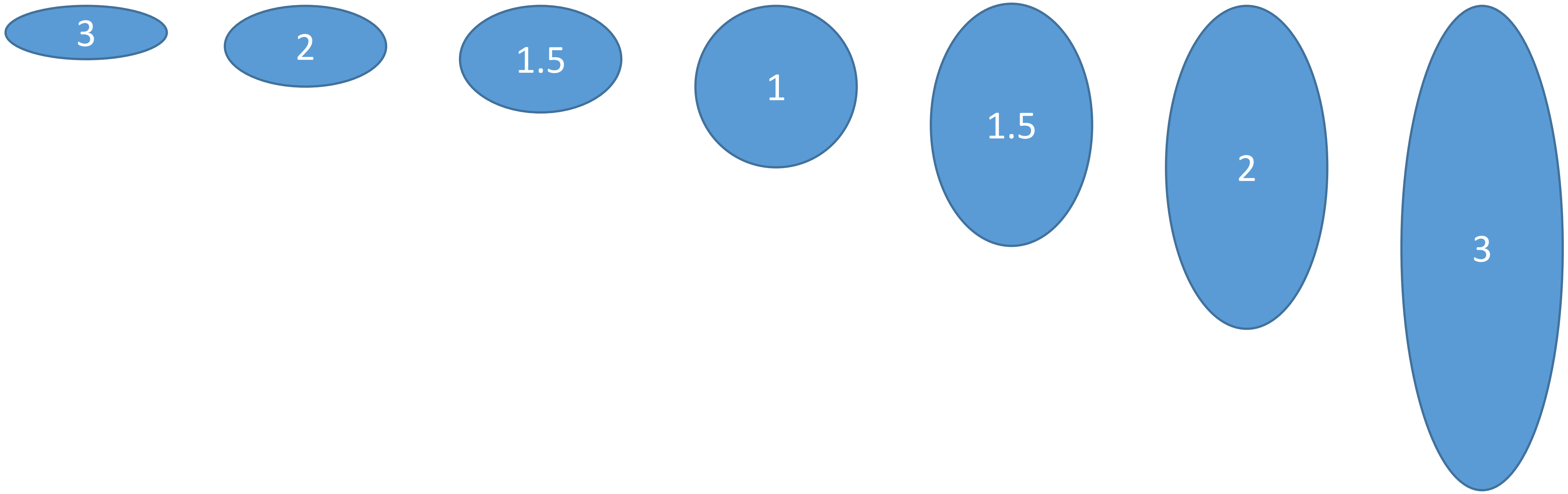
- Major and minor axis are perpendicular to each other



Aspect ratio

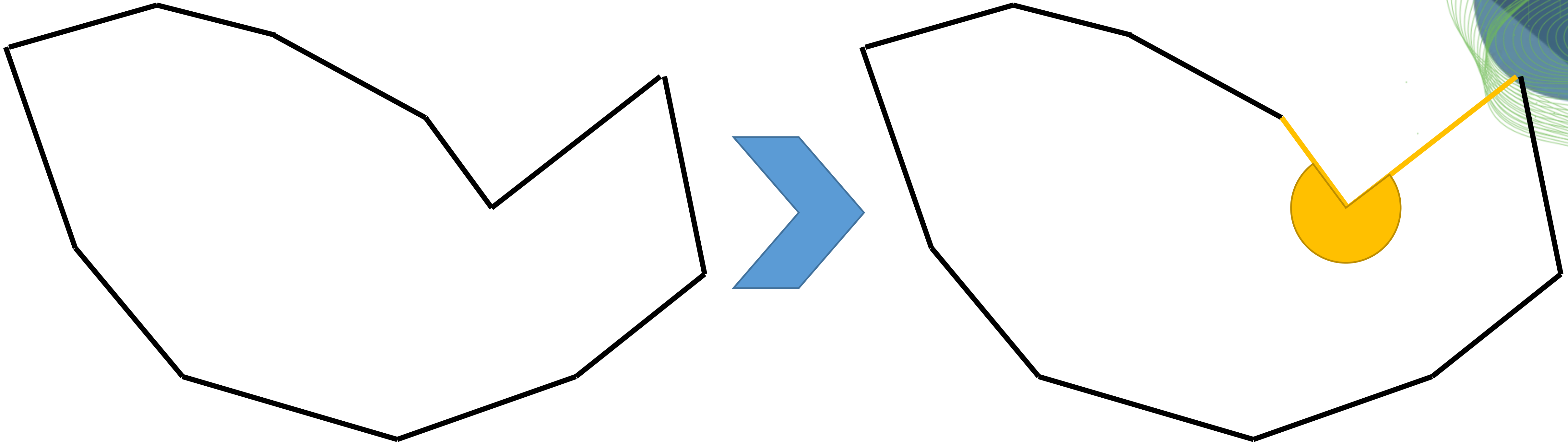
- The aspect ratio describes the elongation of an object.

AR = major / minor



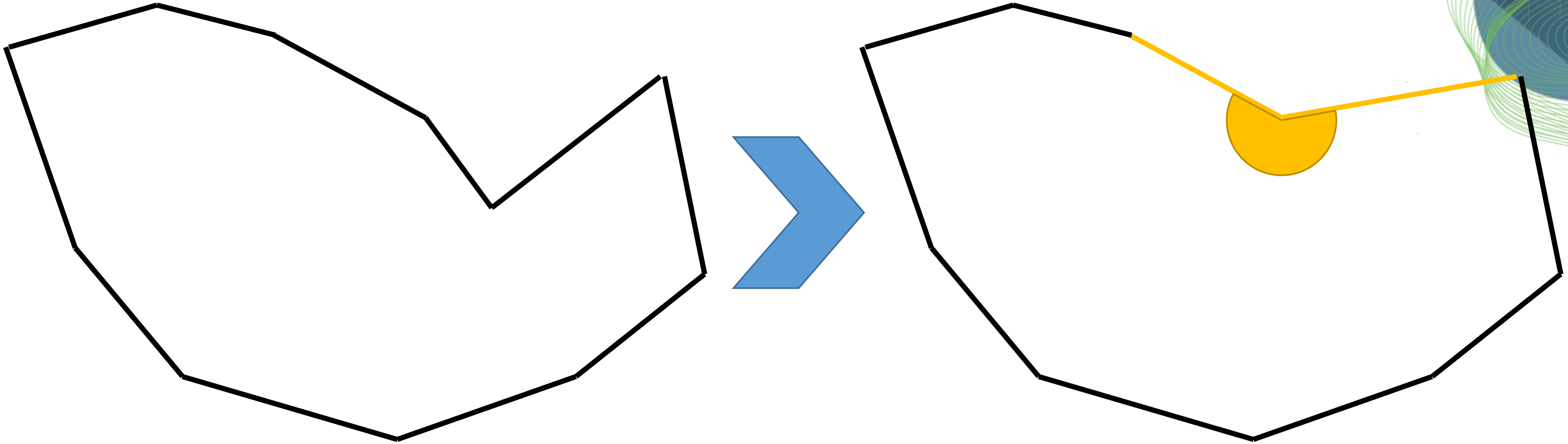
Convex hull

- By removing all concave corners of an object, we retrieve its convex hull.



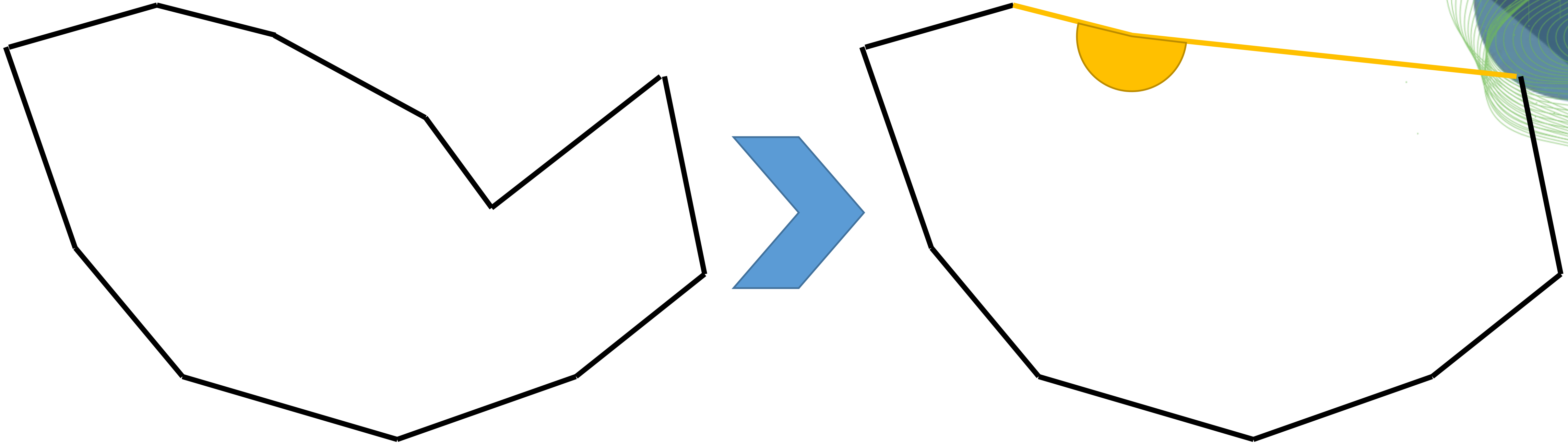
Convex hull

- By removing all concave corners of an object, we retrieve its convex hull.



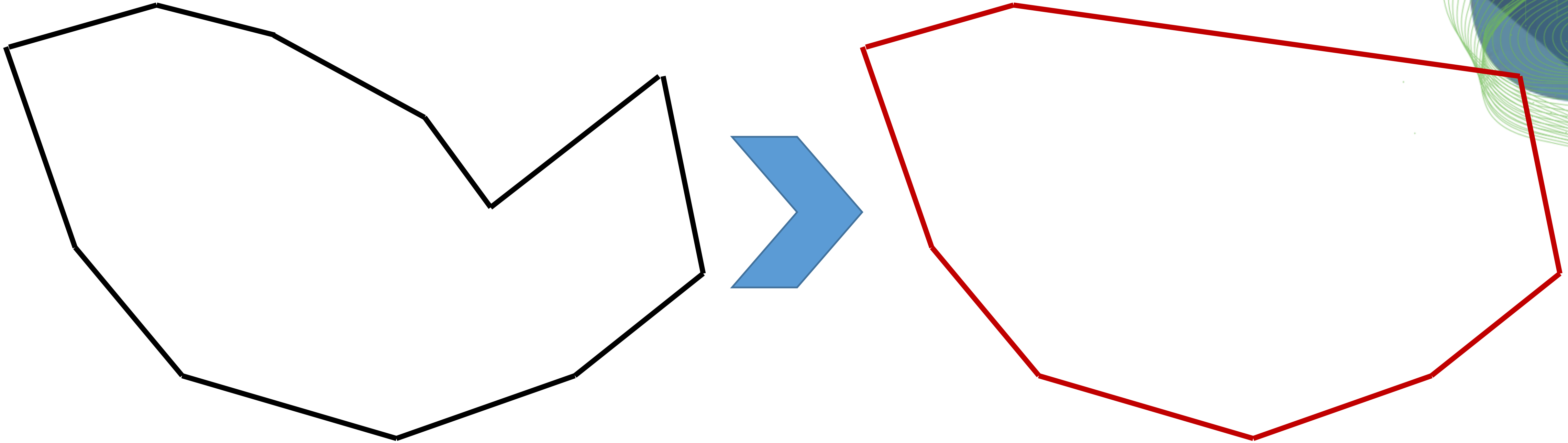
Convex hull

- By removing all concave corners of an object, we retrieve its convex hull.



Convex hull

- By removing all concave corners of an object, we retrieve its convex hull.

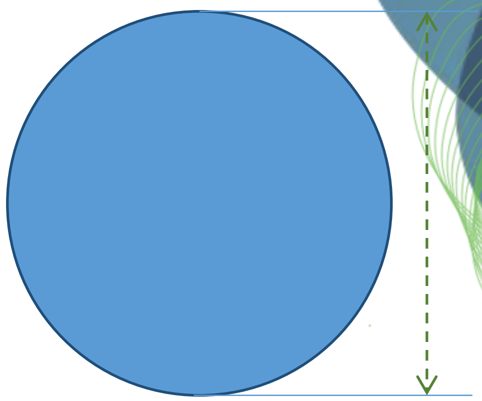


$$solidity = \frac{A}{A_{convexHull}}$$

Roundness and circularity

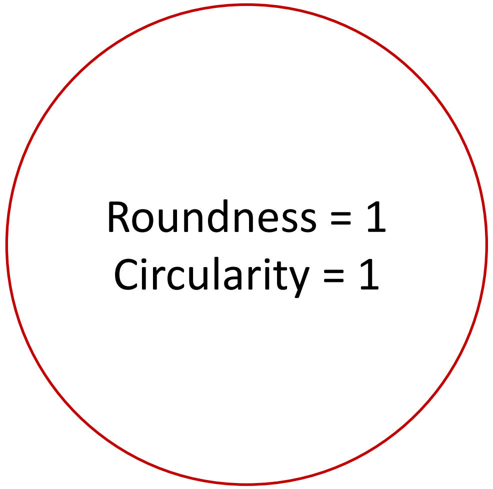
- The definition of a circle leads us to measurements of circularity and roundness.
- In case you use these measures, define them correctly. They are not standardized!

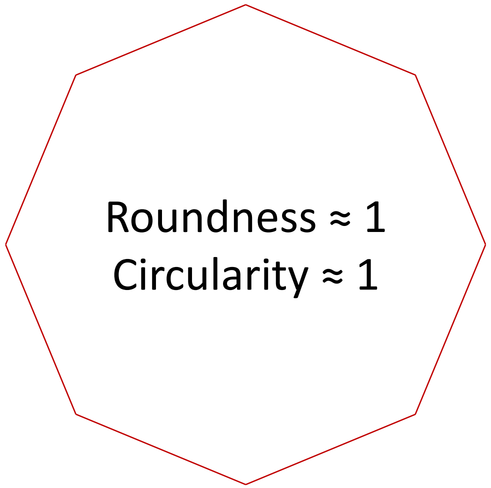
Diameter	d
Circumference	$C = \pi d$
Area	$A = \frac{\pi d^2}{4}$

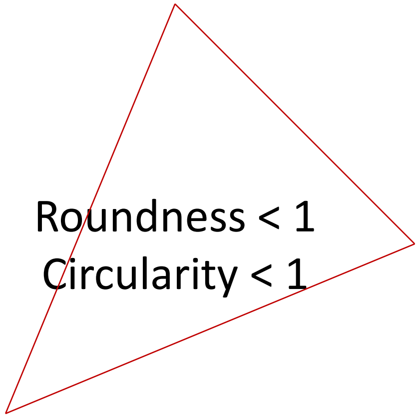


$$roundness = \frac{4 \ * A}{\pi \ major^2}$$

$$circularity = \frac{4\pi \ * A}{perimeter^2}$$





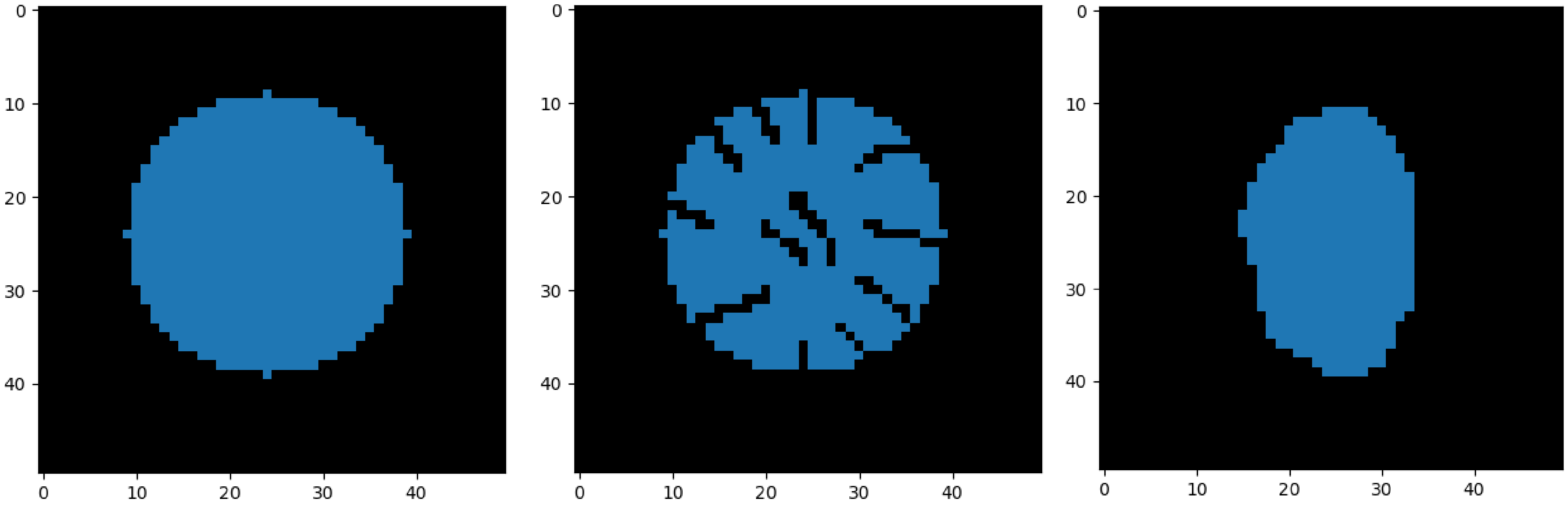


Roundness versus circularity

- To decide which metric to use, consider drawing example object, which reflect the phenotype you are studying.

$$roundness = \frac{4 * A}{\pi \ major^2}$$

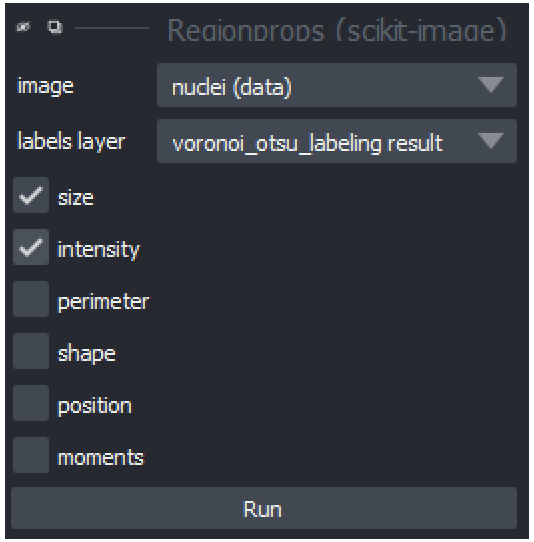
$$circularity = \frac{4\pi * A}{perimeter^2}$$

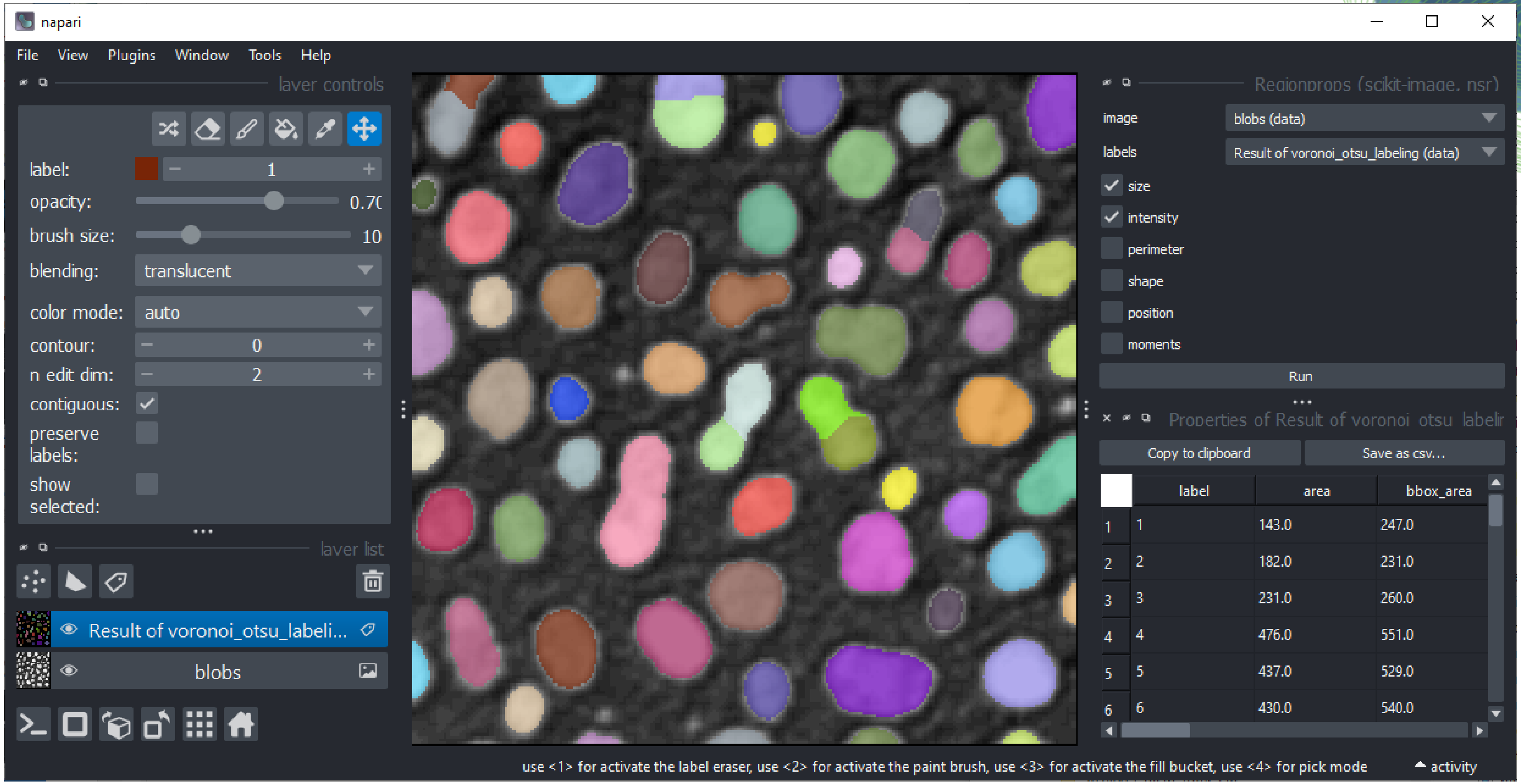


roundness: 1.00	roundness: 0.86	roundness: 0.64
circularity: 0.91	circularity: 0.14	circularity: 0.87

Feature extraction in Napari

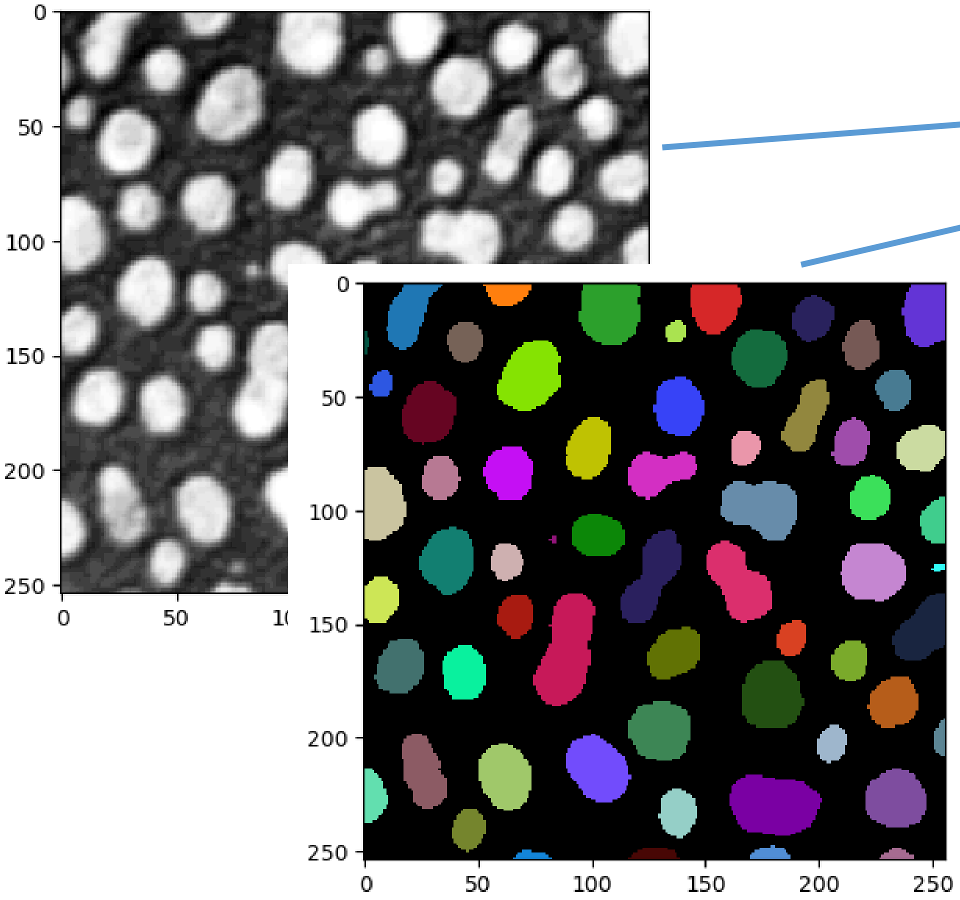
- In Napari: Menu Tools > Measurements tables > Regionprops





Feature extraction in Python

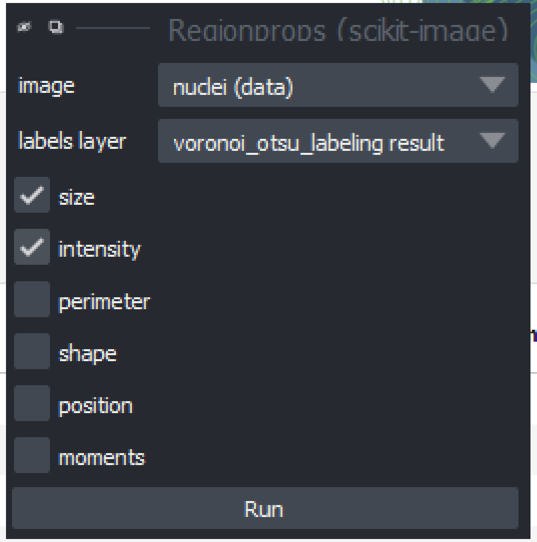
- The transition from image data to tabular data / pandas DataFrames



```
[4]: df = pd.DataFrame(regionprops_table(image , label_image,
                                        perimeter = True,
                                        shape = True,
                                        position=True,
                                        moments=True))
     df
```

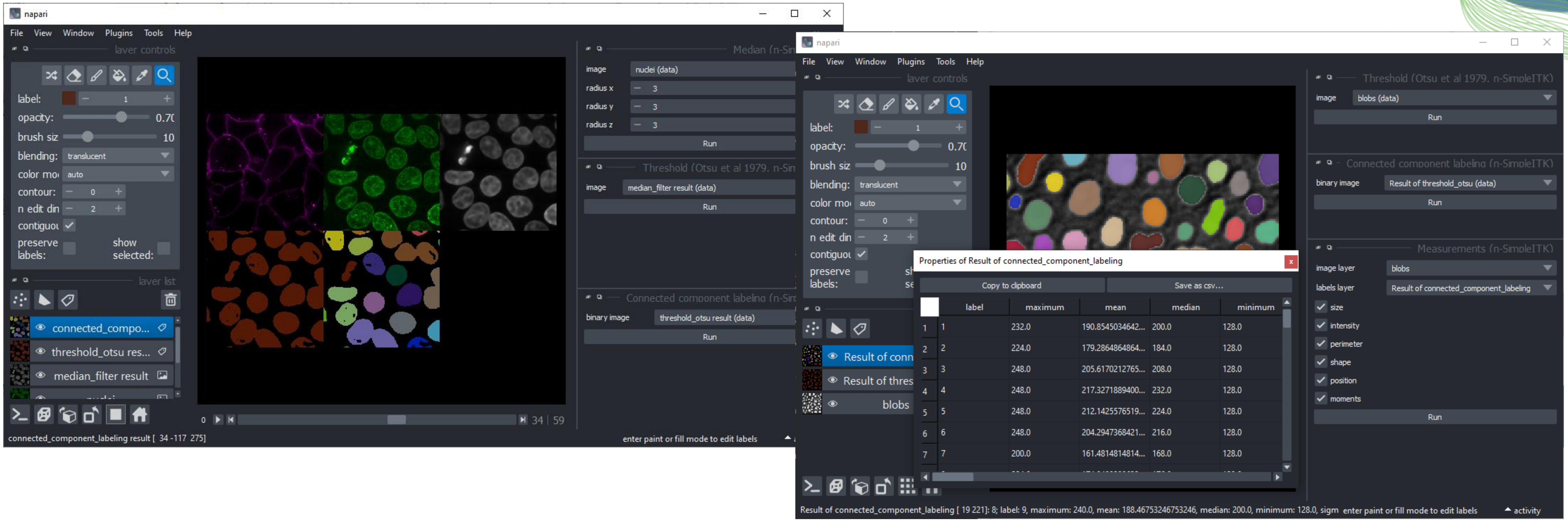
[4]:

	label	area	bbox_area	equivalent_diameter	convex_area	max_intensity			
0	1	429.0	750.0	23.371345	479.0	232.0			
1	2	183.0	231.0	15.264430	190.0	224.0			
2	3	658.0	756.0	28.944630	673.0	248.0			
3	4	433.0	529.0	23.480049	445.0	248.0	217.515012	120.0	77.455844
4	5	472.0	551.0	24.514670	486.0	248.0	213.033898	128.0	83.798990
...	...	...	...	...	...	...	...	...	...
57	58	213.0	285.0	16.468152	221.0	224.0	184.525822	120.0	52.284271
58	59	79.0	108.0	10.029253	84.0	248.0	184.810127	128.0	39.313708
59	60	88.0	110.0	10.585135	92.0	216.0	182.727273	128.0	45.692388
60	61	52.0	75.0	8.136858	56.0	248.0	189.538462	128.0	30.692388
61	62	48.0	68.0	7.817640	53.0	224.0	173.833333	128.0	33.071068



SimpleITK

- Recommended for 3D-measurements, based on the ITK-project
- In Napari: Menu Tools > Measurements tables > Measurements (SimpleITK)



SimpleITK

- Many Napari plugins for feature extraction can also be called from Python

```
statistics = label_statistics(blobs, labels,
                              intensity=True,
                              size=True,
                              shape=True,
                              perimeter=True,
                              position=True,
                              moments=True)

df = pd.DataFrame(statistics)
df
```

	label	maximum	mean	median	minimum	sigma	sum	variance	bbox_0	bbox_1
0	1	224.0	137.526132	136.0	112.0	13.360739	157880.0	178.509343	0	0
1	2	232.0	193.014354	200.0	128.0	28.559077	80680.0	815.620897	11	0
2	3	224.0	179.846995	184.0	128.0	21.328889	32912.0	454.921516	53	0
3	4	248.0	207.082171	216.0	120.0	27.772832	133568.0	771.330194	95	0
4	5	248.0	223.146402	232.0	128.0	30.246515	89928.0	914.851647	144	0
5	6	248.0	214.906725	224.0	128.0	26.386796	99072.0	696.263020	238	0
6	7	248.0	211.565891	224.0	136.0	30.197236	54584.0	911.873073	189	7
7	8	200.0	166.171429	168.0	136.0	16.466894	11632.0	271.158592	133	17

Basic descriptive statistics

- Pandas DataFrames allow basic statistics (more tomorrow)

- Overview:

```
[8]: df_selection.describe()
```

[8]:

	label	area	extent	aspect_ratio	roundness	circularity
count	62.000000	62.000000	62.000000	62.000000	62.000000	62.000000
mean	31.500000	355.370968	0.761363	1.637991	0.692418	0.894101
std	18.041619	211.367385	0.065208	0.794366	0.210973	0.183024
min	1.000000	7.000000	0.541102	1.048053	0.213334	0.529669
25%	16.250000	194.750000	0.744329	1.168451	0.538616	0.805774
50%	31.500000	366.000000	0.781076	1.316003	0.757485	0.925560
75%	46.750000	500.750000	0.799519	1.769976	0.851463	0.966037
max	62.000000	896.000000	0.870370	4.417297	0.974824	1.886542

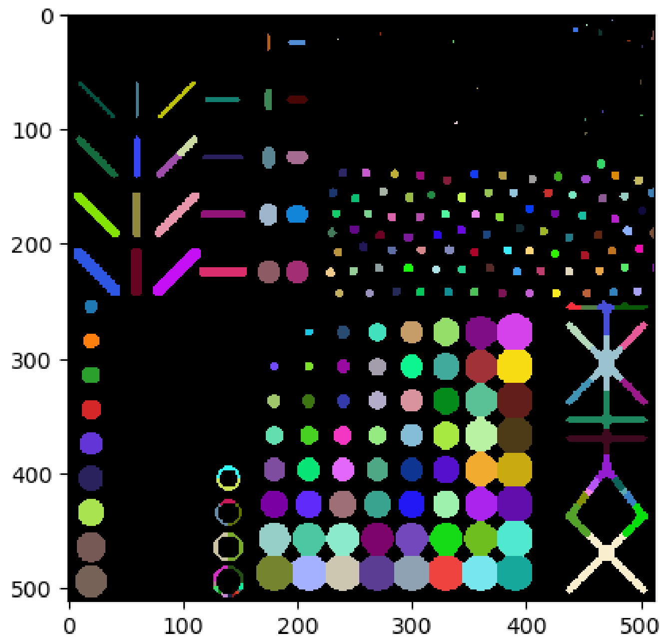
- Specifics:

```
[9]: df_selection['area'].mean()
```

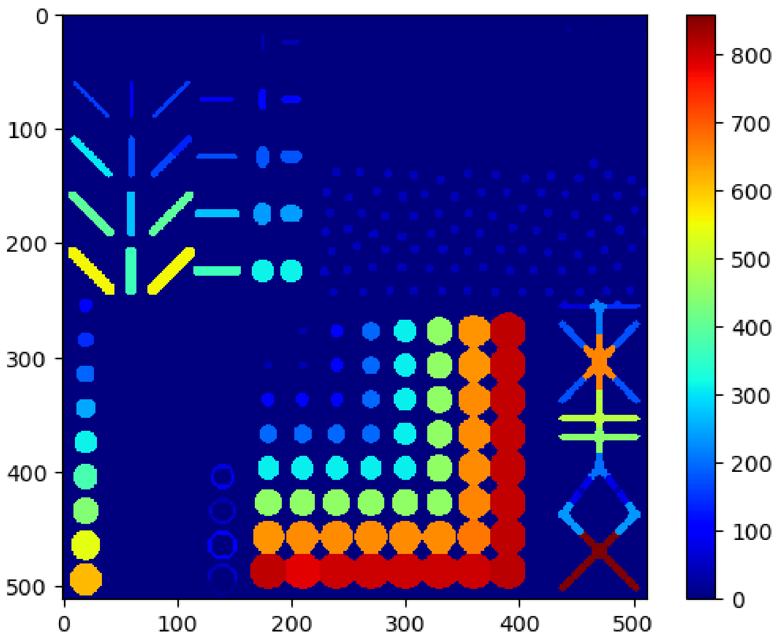
```
[9]: 355.3709677419355
```

Parametric images

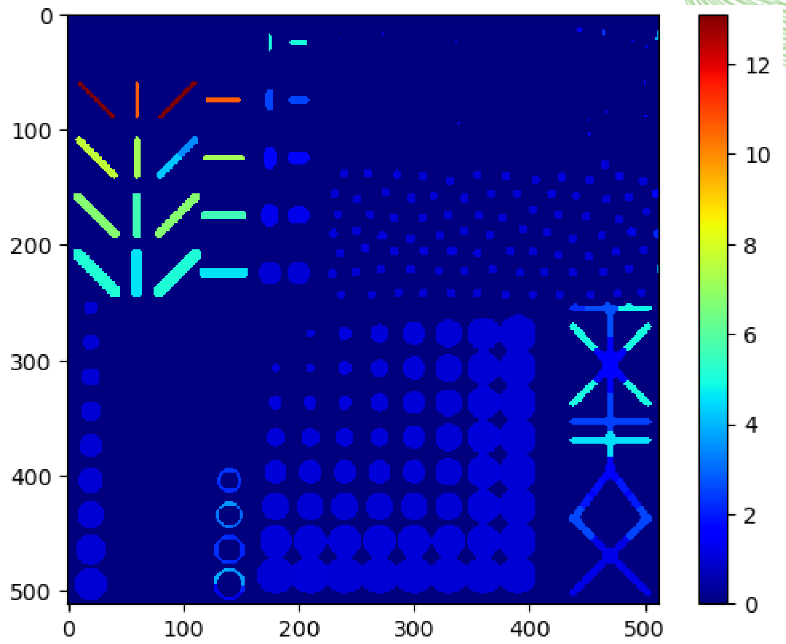
- Visualizing quantitative measurements



Label image



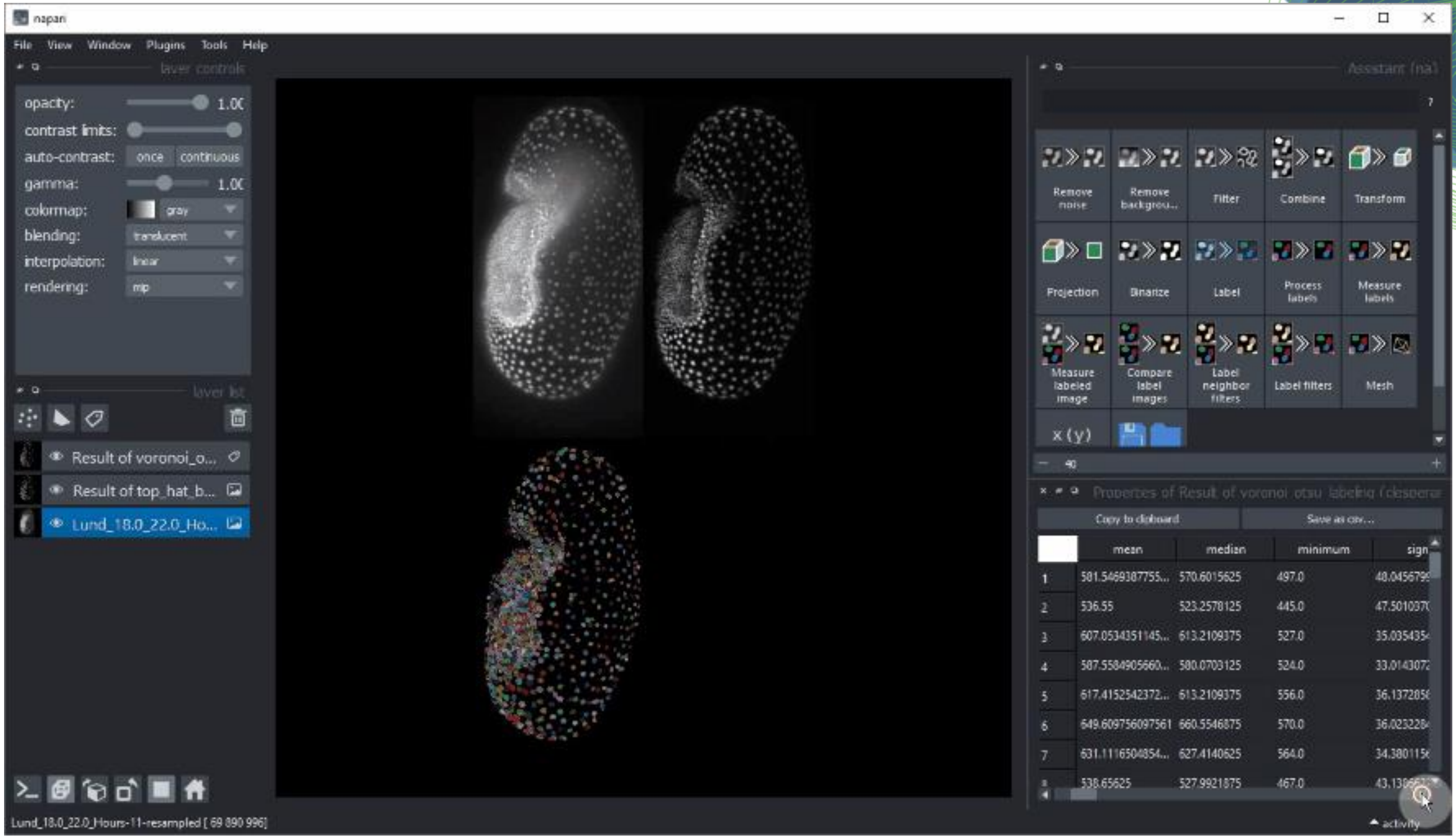
Pixel count image



Aspect ratio image

Exploring features in Napari

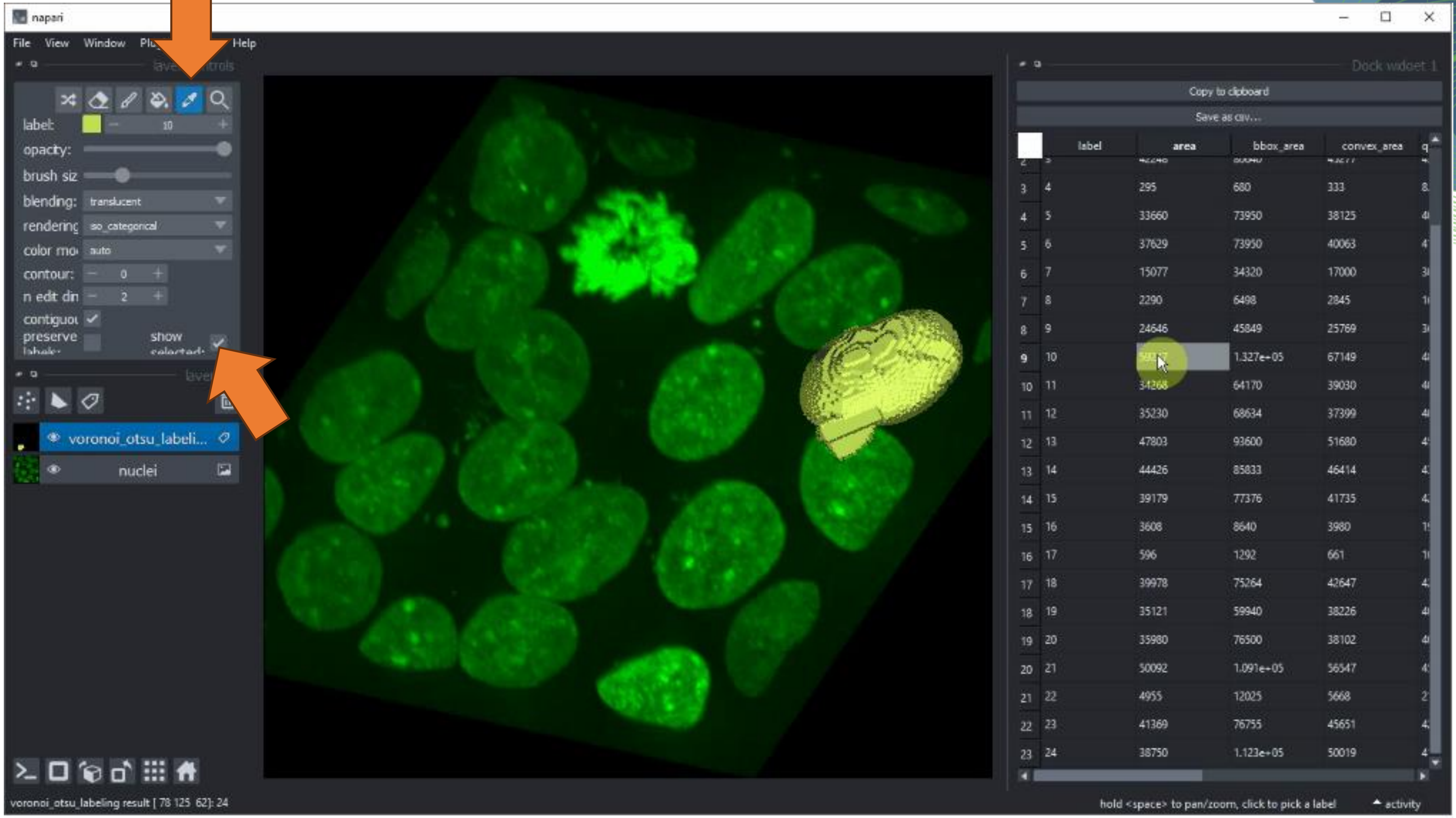
- Double-click on table column to retrieve a parametric map image



https://www.napari-hub.org/plugins/napari-skimage-regionprops

Exploring features in Napari

- Select table rows and view corresponding object in 2D/3D space



ScaDS.AI DRESDEN LEIPZIG
CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

Exercises

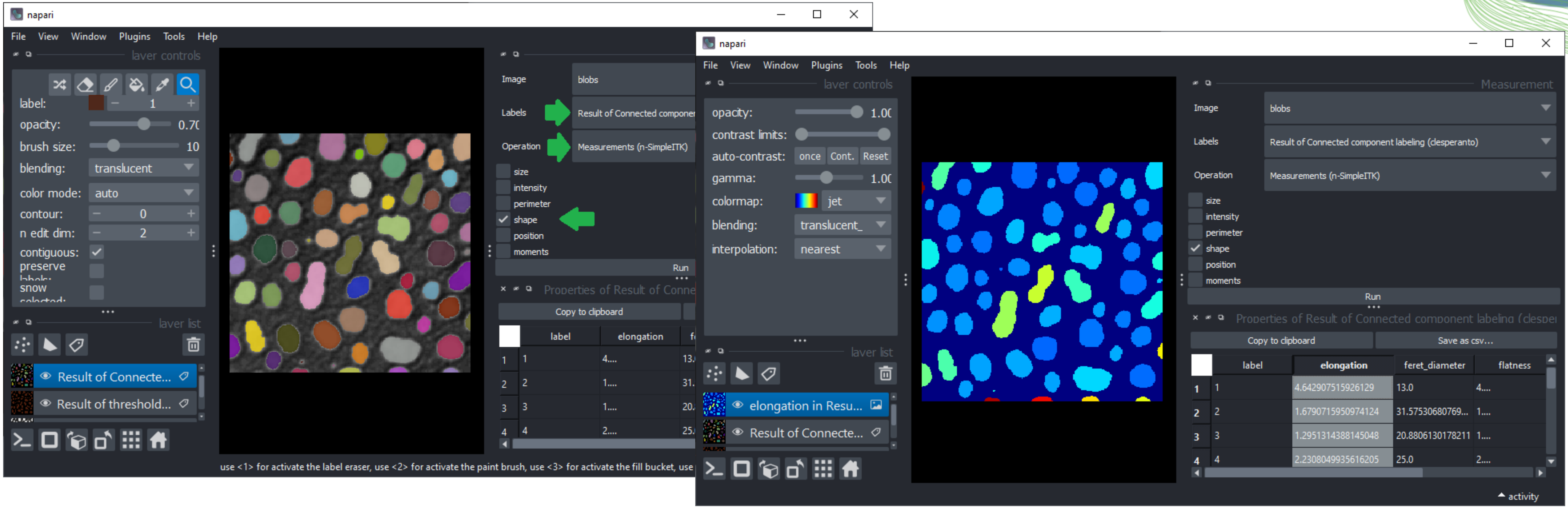
Robert Haase

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

SACHSEN
Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Exercise: Parametric maps

- Produce a parametric map representing 'elongation' in Napari.
- Reproduce the same map using Python



Exercise: Quantitative measurements

- Use the given feature extraction notebook to apply some basic statistics to measurements

```
[5]: df = pd.DataFrame(regionprops_table(image , label_image,
                                         perimeter = True,
                                         shape = True,
                                         position=True,
                                         moments=True))
     df
```

[5]:

	label	area	bbox_area	equivalent_diameter	convex_area	max_intensity	mean_intensity	min_intensity	perimeter	perimete
0	1	429.0	750.0	23.371345	479.0	232.0	191.440559	128.0	89.012193	
1	2	183.0	231.0	15.264430	190.0	224.0	179.846995	128.0	53.556349	
2	3	658.0	756.0	28.944630	673.0	248.0	205.604863	120.0	95.698485	
3	4	433.0	529.0	23.480049	445.0	248.0	217.515012	120.0	77.455844	
4	5	472.0	551.0	24.514670	486.0	248.0	213.033898	128.0	83.798990	
...	...	...	...	...	...	...	...	...	...	
57	58	213.0	285.0	16.468152	221.0	224.0	184.525822	120.0	52.284271	
58	59	79.0	108.0	10.029253	84.0	248.0	184.810127	128.0	39.313708	
59	60	88.0	110.0	10.585135	92.0	216.0	182.727273	128.0	45.692388	
60	61	52.0	75.0	8.136858	56.0	248.0	189.538462	128.0	30.692388	
61	62	48.0	68.0	7.817640	53.0	224.0	173.833333	128.0	33.071068	

62 rows × 86 columns

Exercises

Make a table with only `area`, `mean_intensity`, `standard_deviation_intensity` and `label`.

[ ]:

How many object are in the dataframe?

[ ]:

How large is the largest object?

[ ]:

What is the mean intensity of the brightest object?

[ ]:

What are mean and standard deviation intensity of the image?

[ ]:

Exercise 2D versus 3D:

- Scikit-image is good for 2D measurements, SimpleITK for 3D. Compare both!