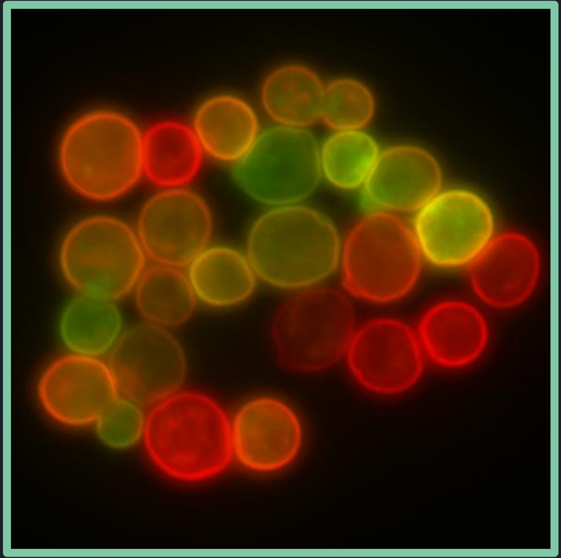
Image segmentation using K-Means clustering

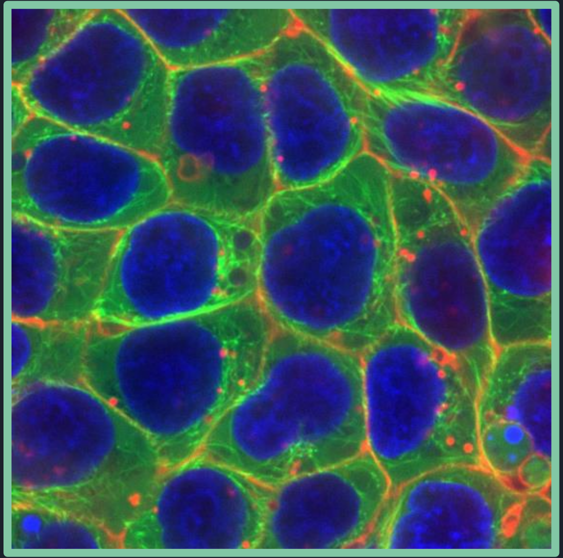
Project Proposal

Data Analysis Project, 4. FS MoBi
Gabriel Tulcan
Emily Locke
Melissa Ringeis
Cedric Leonhard Marquard
Supervisor: Prof. Dr. Karl Rohr

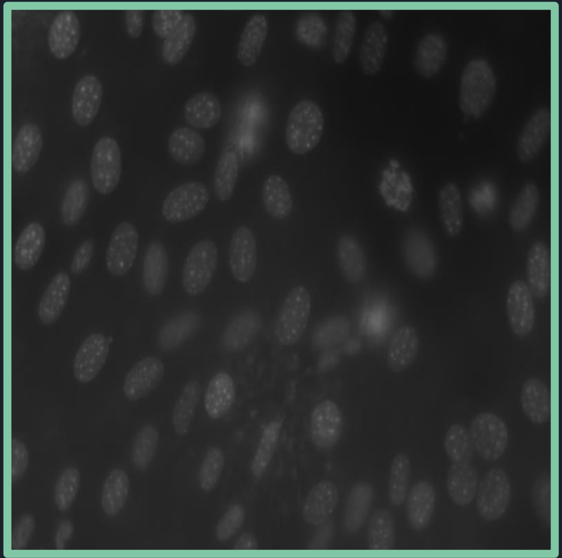
Data



Yeast Cells

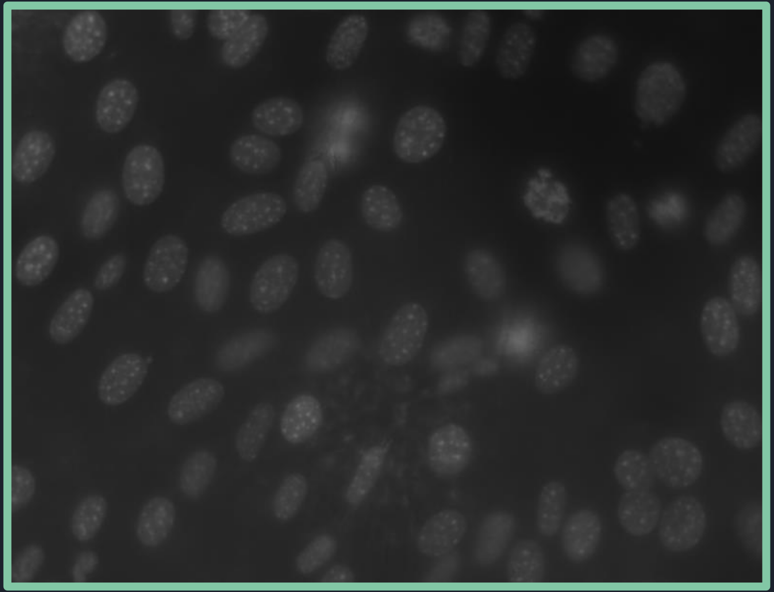


Cell Nuclei

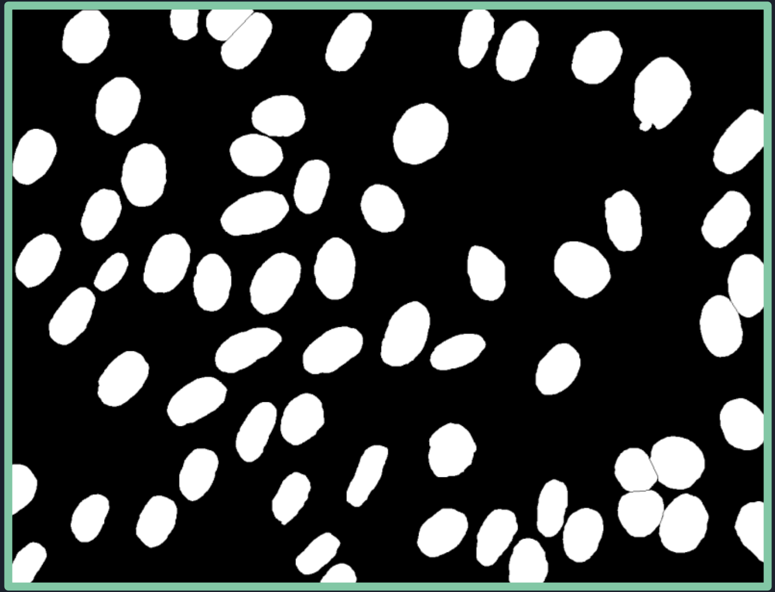


NIH3T3

Data



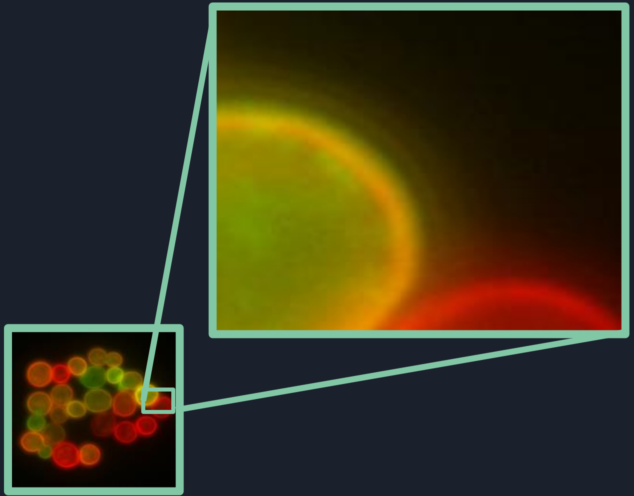
Original image



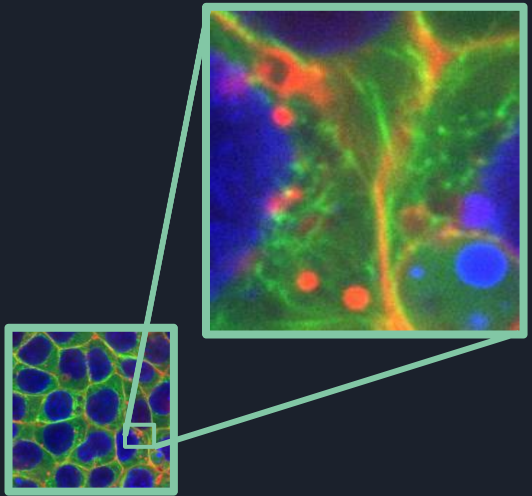
Ground truth image

Challenges

Blurry transitions

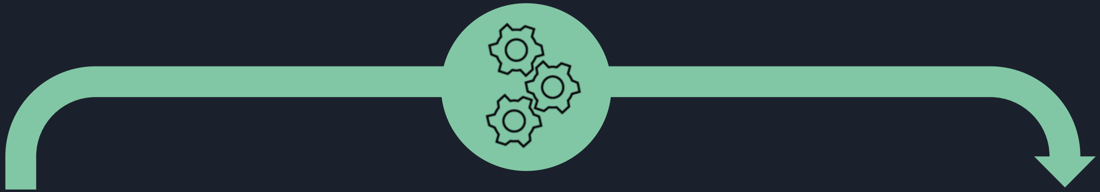


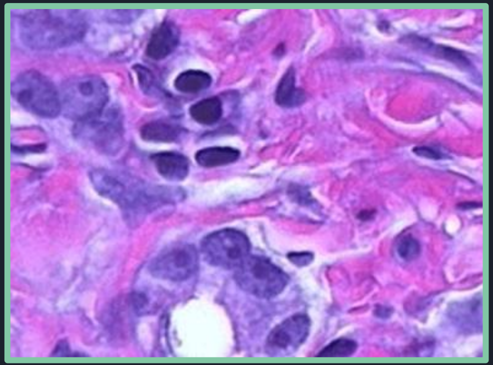
Red spots



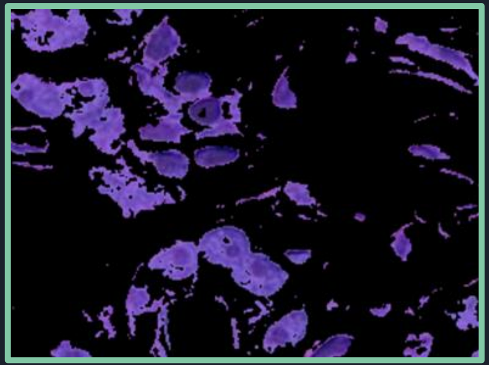
Reflections

Our Goal





Input



Output

https://www.mathworks.com/help/images/color-based-segmentation-using-k-means-clustering.html

Milestones

Clustering Method

Dice Score

Color Models

Pre-Processing

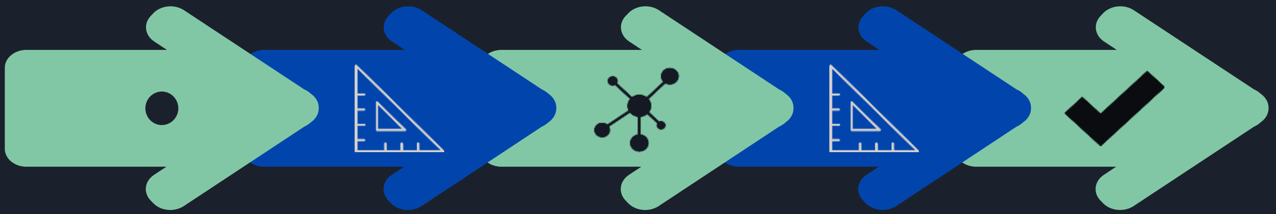
Other Possibilities

K-Means Clustering

Repeat until mean remains unchanged or until maximum number of iterations is reached

Step 2
Calculate distances

Step 4
Recalculate centroids



Step 1
Randomly place centroids

Step 3
Assign clusters

Finished segmentation

K-Means Clustering

Euclidean distance:

$$d_{Euclidean}(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Assignment:

$$\hat{k}^{(n)} = argmin_k d\{m^{(k)} x^{(n)}\}$$

Responsibility:

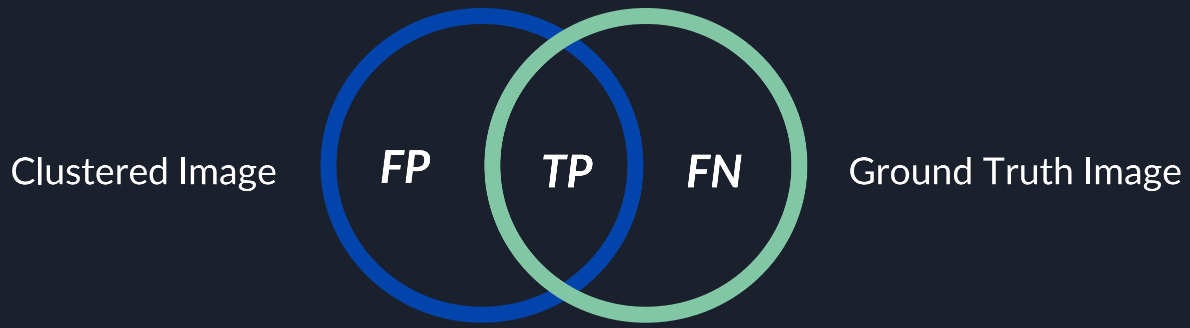
$$r^{(n;k)} = \begin{cases} 1, if\ k = \hat{k}^{(n)} \\ 0, if\ k \neq \hat{k}^{(n)} \end{cases}$$

Centroid Recalculation:

$$m^{(k)} = \frac{\sum_n r^{(n;k)} x^{(n)}}{\sum_n r^{(n;k)}}$$

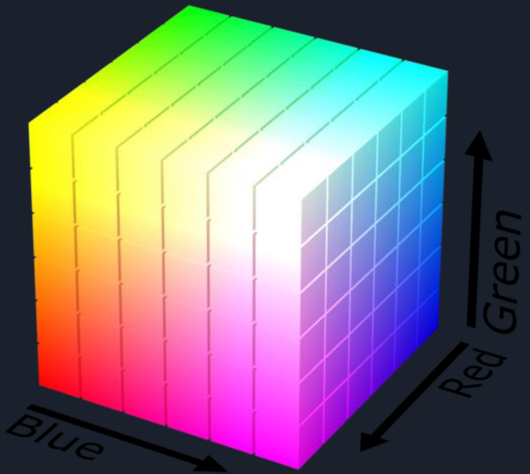
Applications

Dice Score

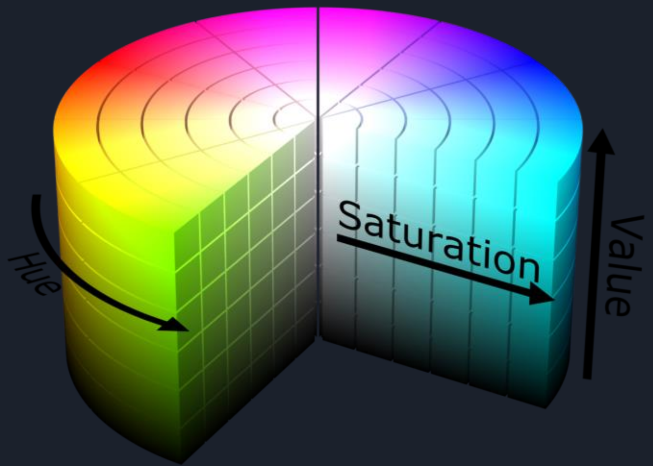


$$DSC = \frac{2TP}{2TP + FP + FN}$$

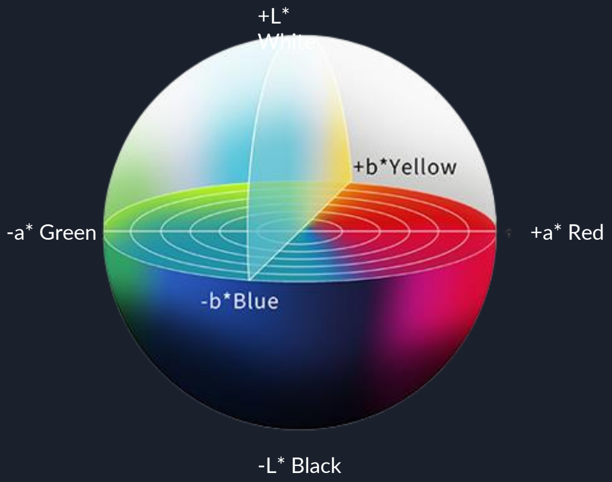
Color Spaces



RGB



HSV



L*a*b

https://en.wikipedia.org/wiki/HSL_and_HSV (adapted)
https://www.linshangtech.com/tech/color-space-tech1439.html (adapted)

Pre-Processing

FILTERS

INTENSITY CLIPPING

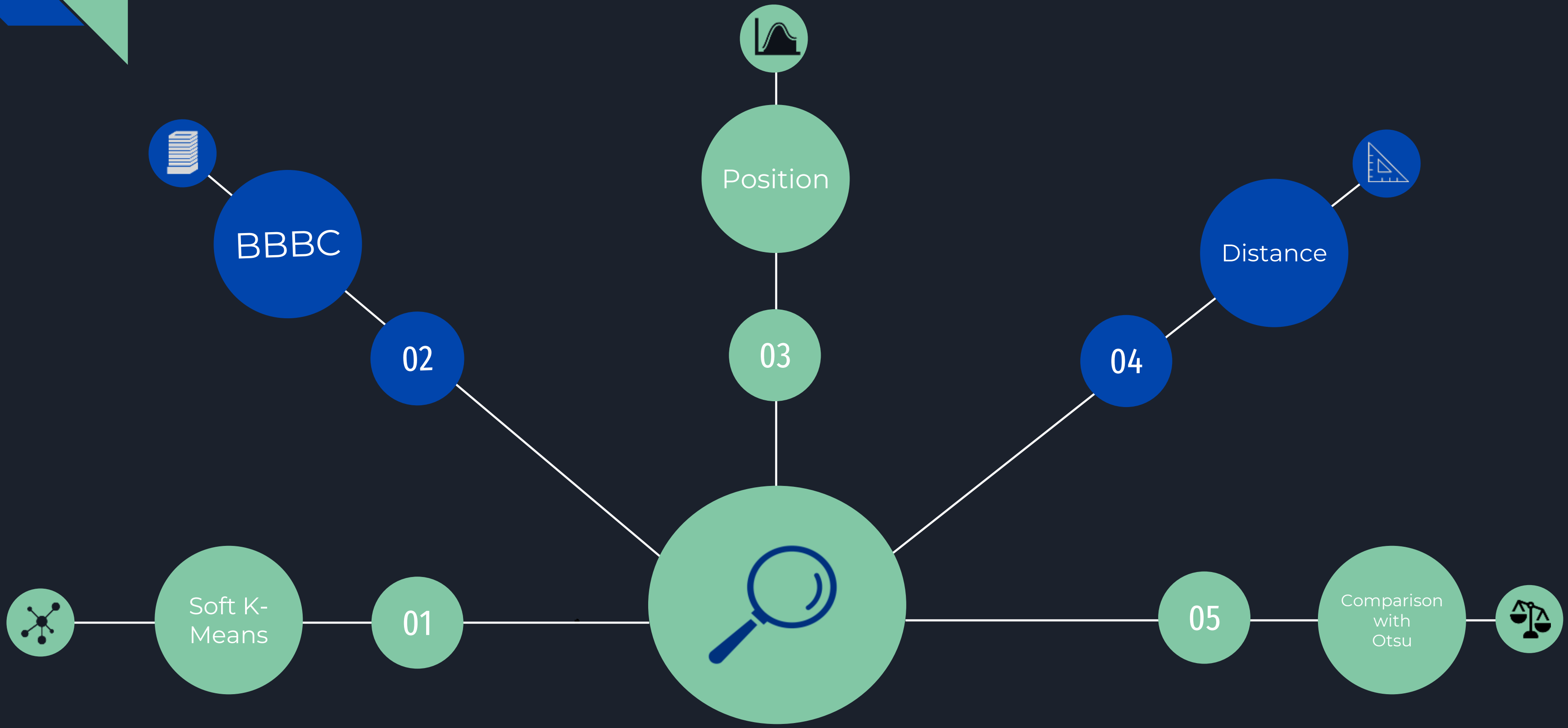
EDGE ENHANCEMENT

01 Apply different filters: Mean, Median, Gauss

02 Reduce the intensity spectrum

03 Improve the acuteness of the image

Other possibilities



Team Management

Teams of two
- working in parallel

Weekly meetings
- brainstorm ideas
- present progress of previous week
- solve problems
- plan further steps

Mentoring
- ask questions
- show progress

VisualStudioCode & GitHub
- simultaneous programming
- share links and ideas

Timeline

K-Means Clustering
19.05 - 29.05

Dice Score
30.05 - 12.06

Color space conversion
13.06 - 19.06

Pre-Processing
20.06 - 26.06

Other possibilities
27.06 - 20.07

Thank you for your attention!

Distances

Euclidean distance:

$$d_{Euclidean}(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Manhattan distance:

$$d_{Manhattan}(x,y) = \sum_{i=1}^{n}|x_i - y_i|$$

Correlation distance:

$$d_{corr}(x,y) = \frac{1}{2}(1 - r)$$

Soft K-Means

Responsibility:

$$r^{(n;k)} = \frac{e^{-\beta d\{m^{(k)};x^{(n)}\}}}{\sum_{k'} e^{-\beta d\{m^{(k')};x^{(n)}\}}}$$

Centroid Recalculation:

$$m^{(k)} = \frac{\sum_n r^{(n;k)} x^{(n)}}{\sum_n r^{(n;k)}}$$

K-Means Clustering - Centroids

Silhouette Method

$$s^{(n)} = \frac{b_n - a_n}{\max\{a_n; b_n\}}$$

$$a^{(n)} = \frac{\sum_{ň} d\{x_n; x_{ň}\} * r^{(ň;k̂^{(n)})}}{\sum_{ň} r^{(ň;k̂^{(n)})}}$$

$$b^{(ň)} = min_{k \backslash k̂^{(ň)}} \frac{\sum_n d\{x_{ň}; x_n\} * r^{(n;k)}}{\sum_n r^{(n;k)}}$$