CELL NUCLEI SEGMENTATION: CLUSTERING

Project Proposal - DataScience SS20 - Project 2 Group 4

Julia Braun, Marika Högner, Stefanie Mantz

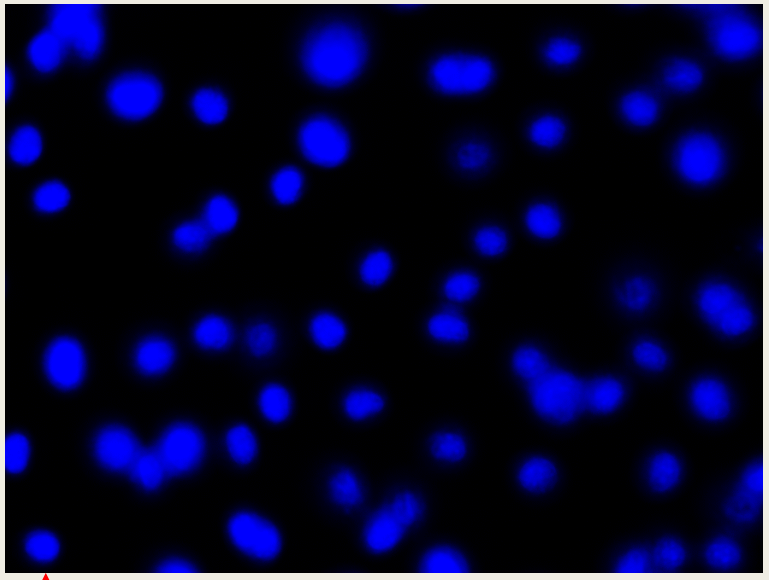
Project goals

- **Quantifying** cell nuclei manually entails
  - *Subjective decision*
  - *Low reproduceability*
  - *Time-consuming*

- **Solution:**
  - *Automated cell nuclei **segmentation based on clustering***
  - *Automated cell nuclei **quantification and data mining (e.g. counting)***

- **Literature:**
  - ***Diversity of clustering algorithms!*** *Compare Estivill-Castro, 2002. https://doi.org/10.1145/568574.568575*
  - ***Diversity of applications!*** *Compare Koyuncu et al. 2018 https://doi.org/10.1002/cyto.a.23594, Fouad et al., 2017 2018 https://doi.org/10.1371/journal.pone.0188717*

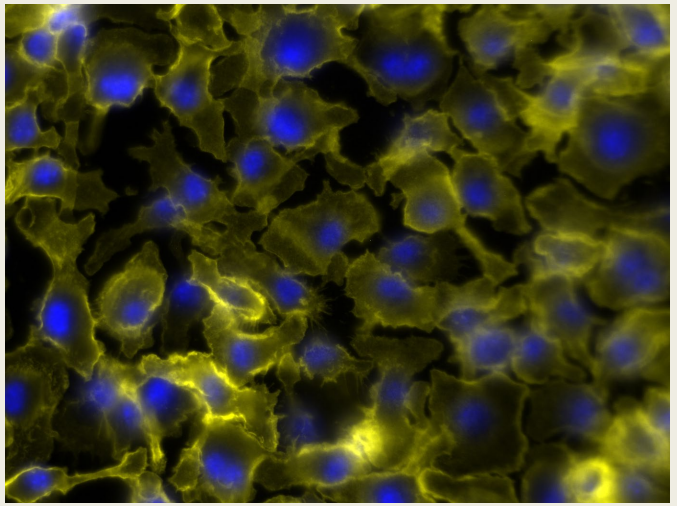
Fluorescent image, only nuclei.



Ground-truth image.

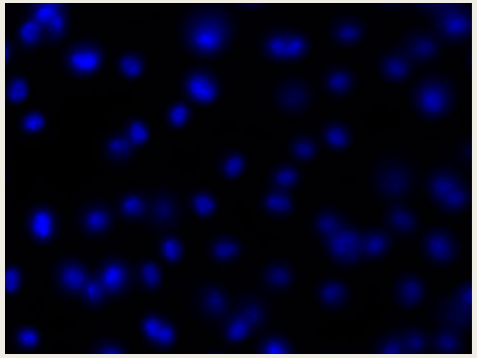
Dataset

Fluorescent image, entire cell.



- 25 images (1388×1040)
  - *Fluorescence microscopy (DAPI stain)*
  - *Murine bone-marrow derived macrophages*
  - *30–50 cell nuclei per image*
- Broad Bioimage Benchmark Collection (BBBC)
  - *Original images: BBBC020_v1_images*
  - *Ground-truth: BBBC020_v1_outlines_nuclei*
- Benchmark: *compare* algorithms
  - *Ground-truth as a common reference point!*

Logical algorithm flow



- Input: Image
- Filtering and pre-processing
- Segmentation – K means clustering
- Data mining: Count nuclei
- Output: Segmented image, amount of nuclei …

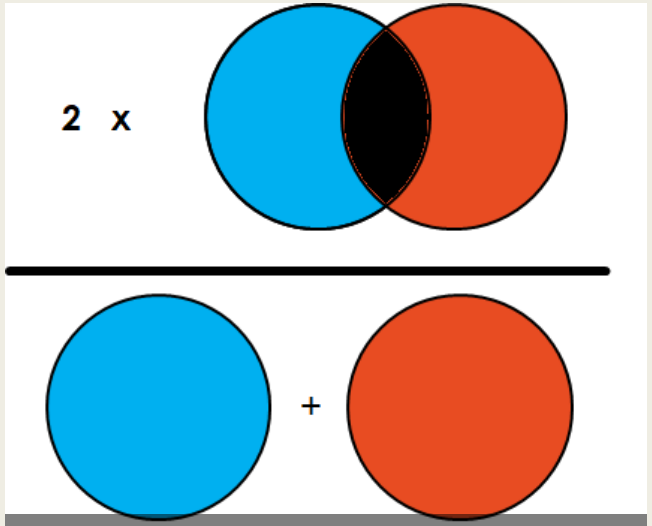
Performance evaluation: Dice score

1. Milestone: Implement evaluation measure

- **Delivers:** measure to compare two images
  - *tests the performance of our segmentation method*
  - *compares the segmentation with the control*

- **Planned analysis steps:**
  - *Write a Dice-score-function on our own*
  - *unit testing: tests the performance of the Dice-score-function using mock objects ⟶ synthetic images*
  - *Write a function that generates synthetic images*

Dice Score



https://towardsdatascience.com/metrics-to-evaluate-your-semantic-segmentation-model-6bcb99639aa2, accessed 18.05.2020

$$2 * \frac{|X \cap Y|}{|X| + |Y|}$$

https://forums.fast.ai/t/understanding-the-dice-coefficient/5838 accessed 18.05.2020

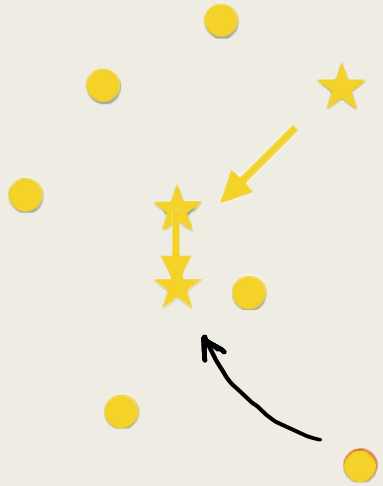
- Statistical method to evaluate the similarity of two samples
- Compares output with the ‘ground truth’
- In our case: segmented image ↔ control image
- Output: Value between 0 and 1
  - *0 → no match*
  - *1 → complete match*

2. Milestone: Implement image segmentation

- **Delivers** image with labelled pixels
  - *Should result in two labels: ‘cell nuclei’ and ‘background’*

- **Planned analysis steps:**
  - *Write a K-means-function on our own*
  - *Test the performance with the Dice-score-function, using the control images*

K means clustering



Source: Lecture Computer Science Dr. Carl Hermann

1. Choose k random points → centers
2. Assign each datapoint to the nearest center → cluster
3. Update centers of each cluster → mean value of cluster
4. Assign each datapoint to the nearest new center → new cluster
5. Repeat steps 3. & 4. until:
   a. *Predefined maximum number of iterations is reached*
   b. *No datapoint changed its cluster*

3. Milestone: Implement pre-processing methods

- **Delivers:** images, that yield better clustering results
  - *Denoise the image*
  - *Separate nuclei, that appear fused due to the segmentation method*

- **Planned analysis steps:**
  - *Write gaussian-filter-function to reduce noise*
  - *Find a method to separate nuclei that appear fused after segmentation (e.g. Opening, Watershed)*

4. Milestone: Counting the nuclei after segmentation

- **Delivers:** amount of cell nuclei in the image

- **Planned analysis steps:**
  - *contouring after segmentation → counting the amount of contours (each nuclei has one circular line enclosing it = contour)*

Timeline


Dice score
21.05. – 03.06.
K means
04.06. – 17.06.
Pre-Processing
18.06. – 01.07.
Counting
02.07. – 15.07.

Thank you for your attention!