3000788 Intro to Comp Molec Biol

Lecture 22: Microscopy data analysis

Fall 2025

Sira Sriswasdi, PhD

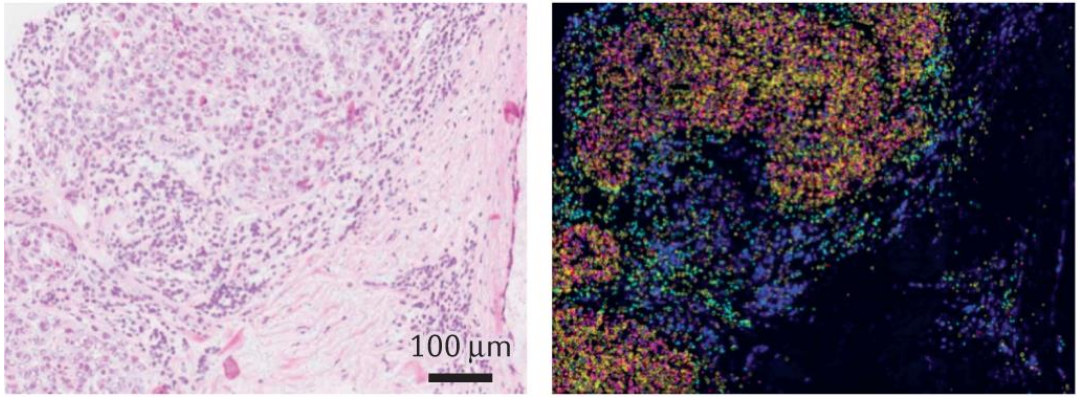
- Research Affairs
- Center of Excellence in Computational Molecular Biology (CMB)
- Center for Artificial Intelligence in Medicine (CU-AIM)

Today's agenda

- Roles of microscopy in biology
- Extraction of biological features from microscopy images
- Image processing concepts

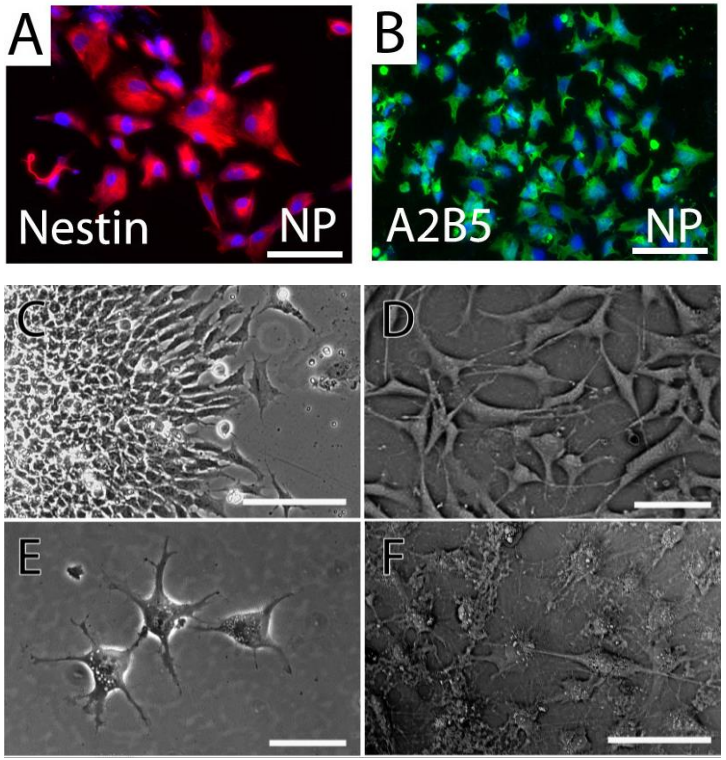
Roles of microscopy in biology

- Inexpensive molecular assays
- Cell morphology
- Tissue structure and microenvironment

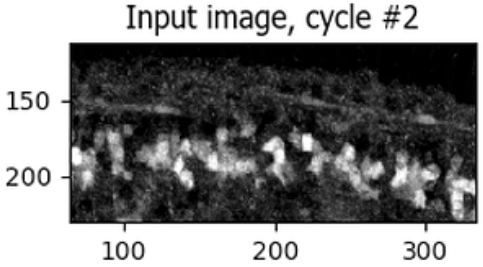


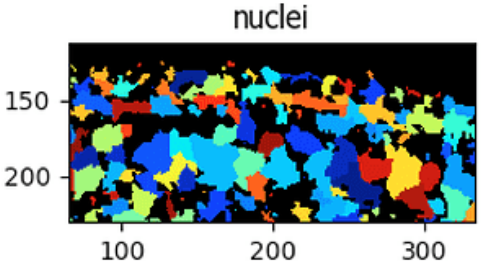
Longo et al. Nature Reviews Genetics 22:627-644 (2021)

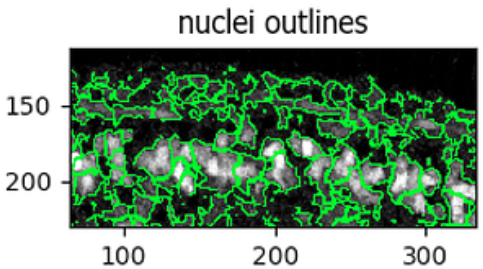
Ali, A.H. et al. PLoS ONE 10:e0116933 (2015)



Quantitative analysis of cells







# of accepted objects	538
10th pctile diameter	1.6 pixels
Median diameter	8.5 pixels
90th pctile diameter	19.5 pixels
Area covered by objects	24.4 %
Thresholding filter size	1.0
Threshold	0.1
Declumping smoothing filter size	5.4
Maxima suppression size	5.3

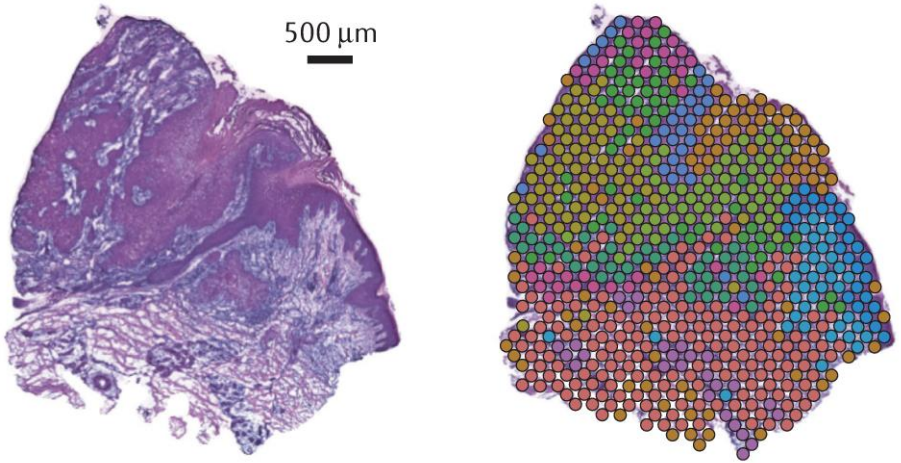
- Segment individual cells
- Quantify morphological characteristics of each cell
  - **Ex**: Size, nucleus/cytoplasm, roundness
- Quantify fluorescence staining signals in each cell / cellular component

https://forum.image.sc/t/counting-gilal-cells-in-tissue-using-cellprofiler/62332

Reconstruction of spatial transcriptomics

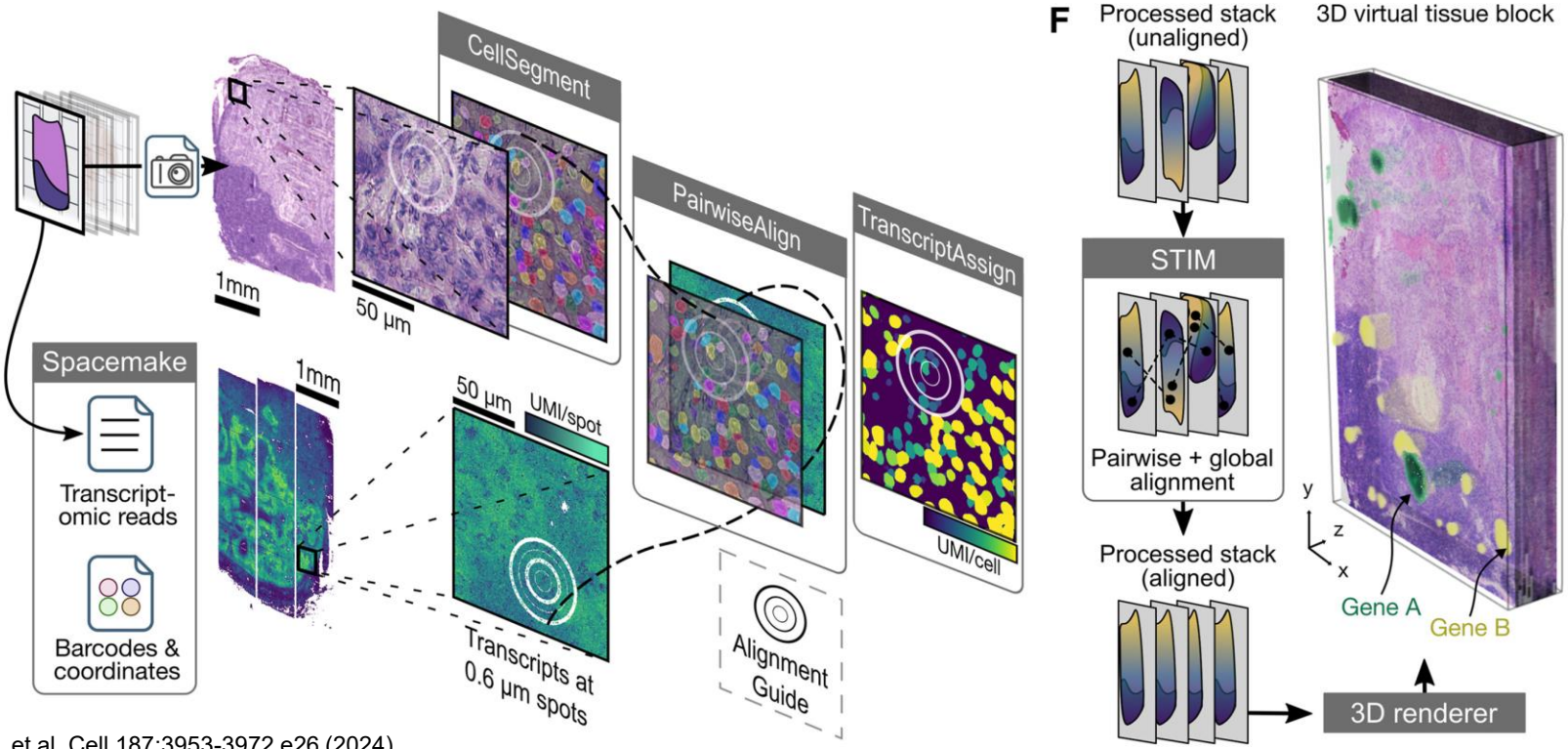
- Classify tissue regions in spatial transcriptomics
  - Transcriptomic similarity
  - Distance on the tissue
  - Visual similarity on the image

- Quantify cell-cell communication
  - Proximity and composition of cell types in a region
  - Correlation of gene expression with spatial distance



Longo et al. Nature Reviews Genetics 22:627-644 (2021)

3D tissue visualization



Scott, M. et al. Cell 187:3953-3972.e26 (2024)

Cell morphology can be informative

https://sphweb.bumc.bu.edu/otlt/mph-modules/ph/ph709_cancer/ph709_cancer7.html

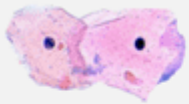
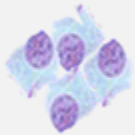
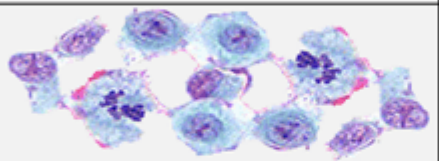
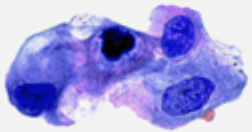
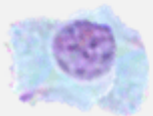
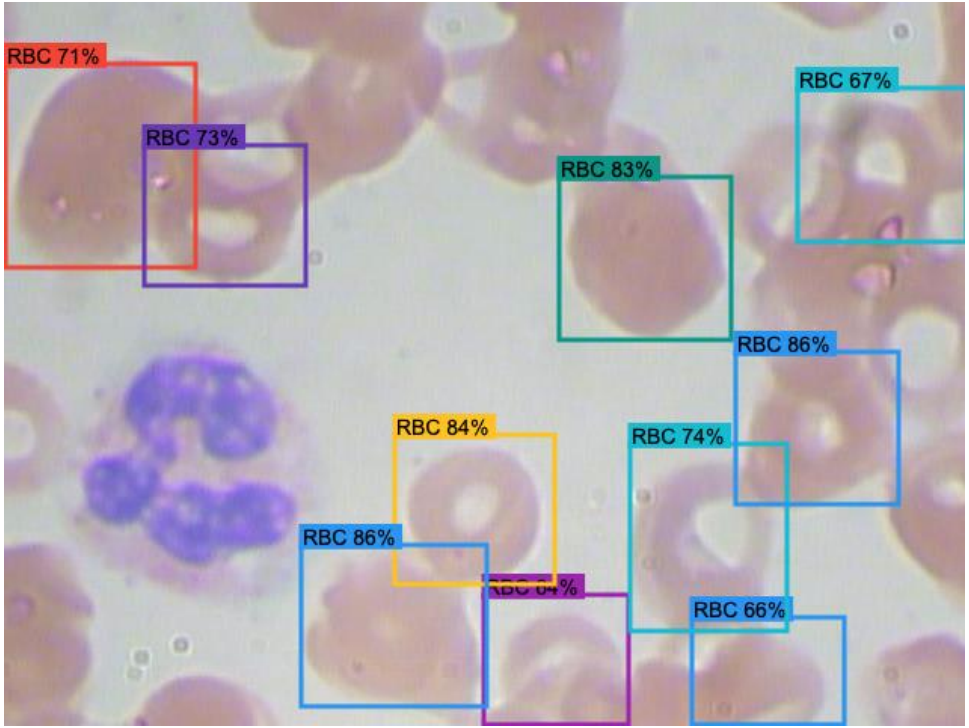
Normal	Cancer	
		Large, variably shaped nuclei
		Many dividing cells; Disorganized arrangement
		Variation in size and shape
		Loss of normal features

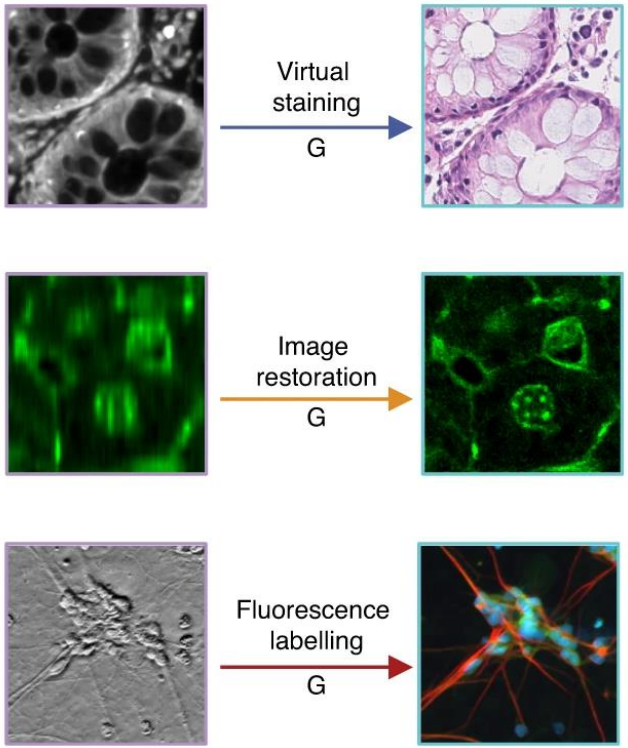
Image-based cell type prediction

- Cell morphology sometimes correlates with cell type and gene expression
- Train classification model that predict cell types from morphology (unlabeled images)
- **Cheap molecular assays**
  - Remove the need for omics



https://blog.intelec.ai/post/red-blood-cell-detector/

Virtual staining / *in silico* labeling / H&E 2.0

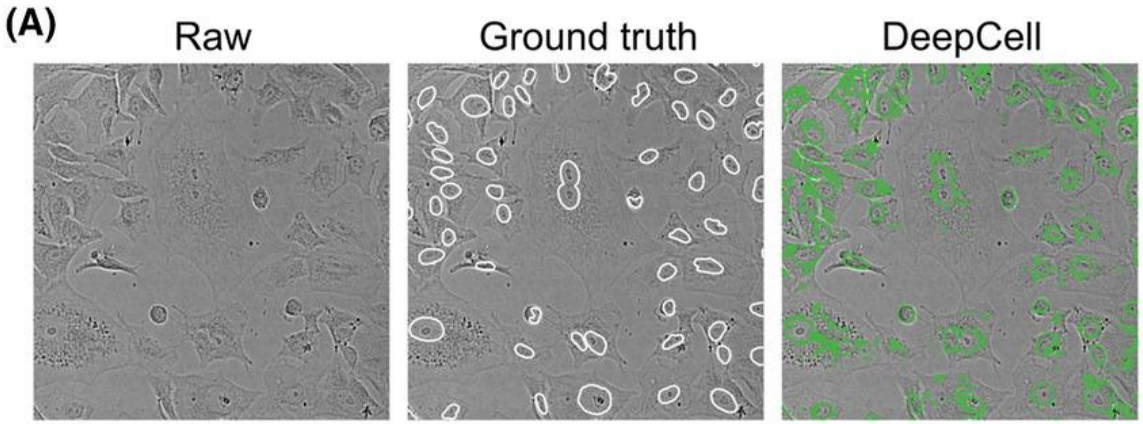


Li, X. et al. Light: Science and Applications10:44 (2021)

- Develop **generative AI** that can produce stained images from brightfield / unlabeled images
  - Remove the time and cost for staining
  - Accessible to labs with basic microscopes
  - Preliminary screening tool
- Also image sharpening and re-focusing

Extraction of biological features from microscopy images

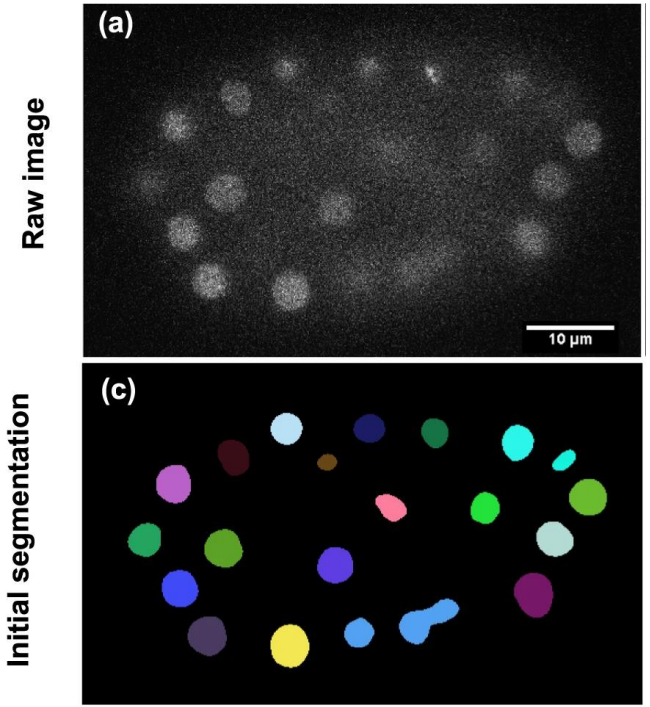
Nuclei segmentation



Fishman, D. et al. Journal of Microscopy 284:12-24 (2021)

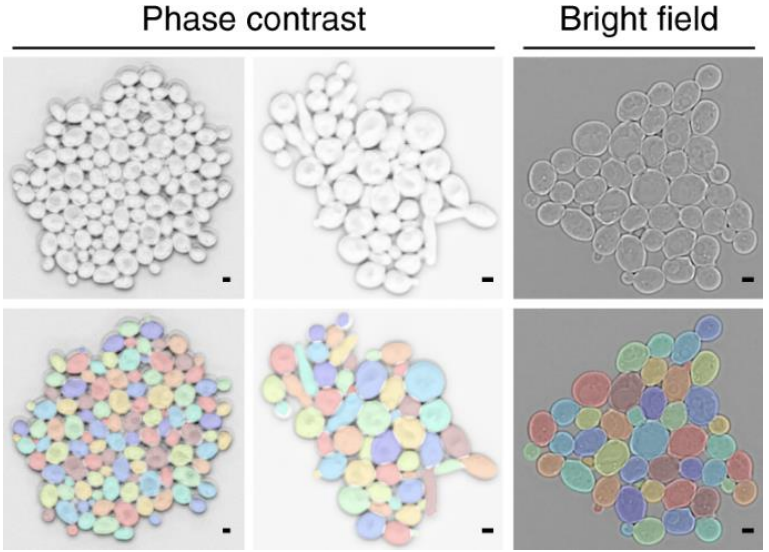
- Most distinctive cell features
  - Density in brightfield
  - Easy to stain
- Mark the position of all cells

Nasser, L. and Boudier, T. Scientific Reports 9:5654 (2019)

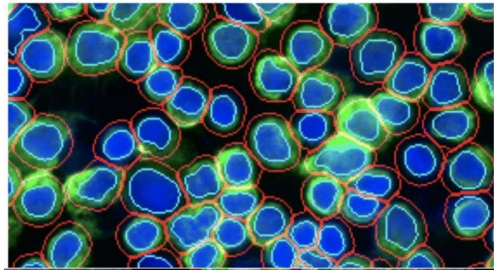


Cell segmentation

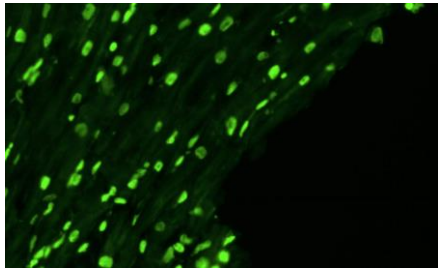
- Varying difficulty, depending on cell morphology and imaging quality
- Phase contrast enhances cell boundary
- Staining of cytoplasmic proteins

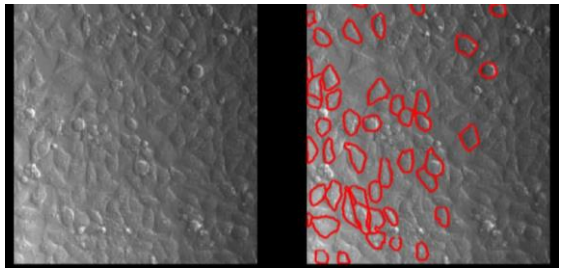


Dietler, N. et al. Nature Communications 11:5723 (2020)



Czech, E. et al. BMC Bioinformatics 20:448 (2019)

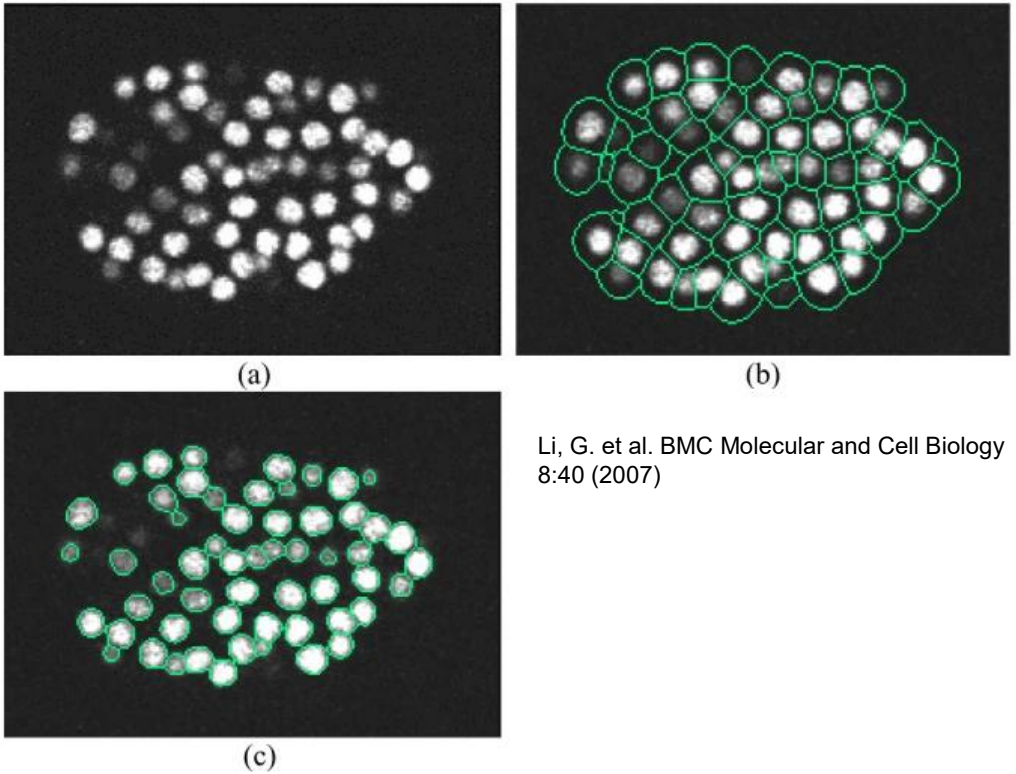




https://forum.image.sc/t/counting-cells-within-brightfield-images/75185/2

Approximated cell segmentation

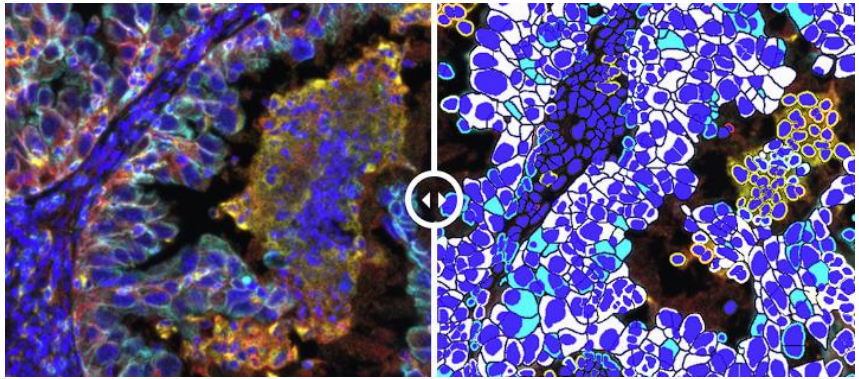
- **Assumptions**: Cells are spherical and equal in size
- Use mid-points between adjacent nuclei to define cell boundary
- Good enough for statistical analysis
  - Remove outliers
  - Population comparison



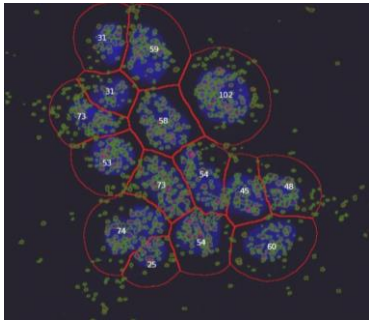
Li, G. et al. BMC Molecular and Cell Biology 8:40 (2007)

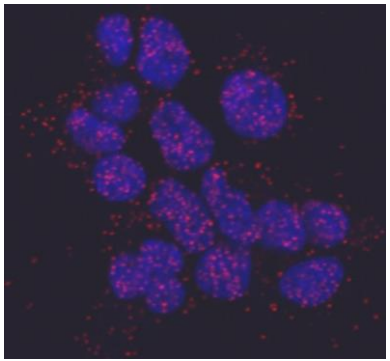
Quantifying signals in cellular regions

- Summarize signal intensity inside each segmented nucleus and cell
- Total signal intensity
  - Diffused proteins
- Number of spots
  - Protein aggregated, organelles
- Gradient of signal intensity
  - Polarized expression



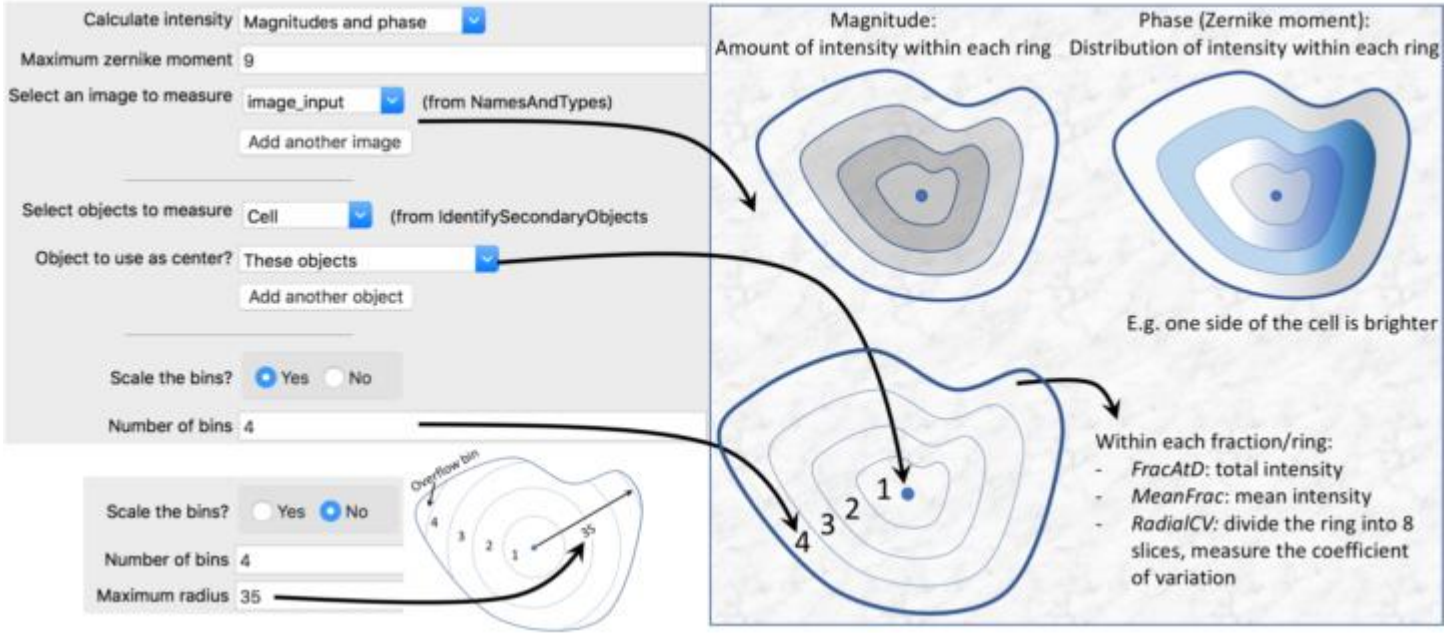
https://indicalab.com/halo/halo-modules/highplex-fl/





https://cellprofiler.org/previous-examples

Polarized protein expression

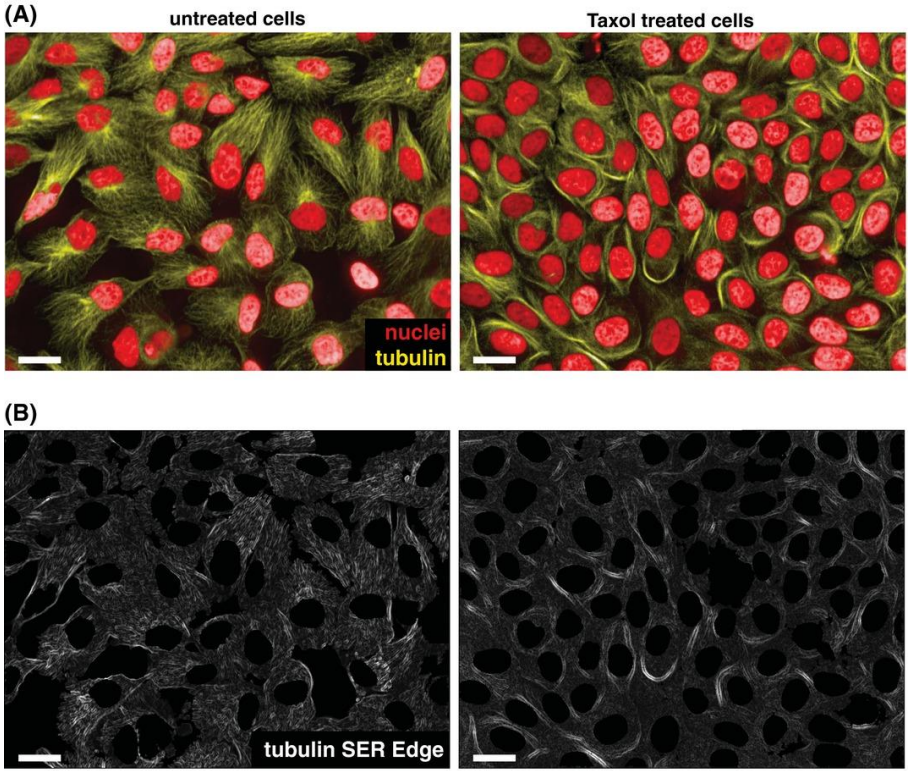


CellProfiler manual

- Find signal imbalance that centers around the nucleus

Texture characteristics

- Roughness = abrupt change in intensity across adjacent pixels
- Smoothness = small change in intensity across adjacent pixels
- **Texture** is described by histograms of intensity patterns across nearby pixels



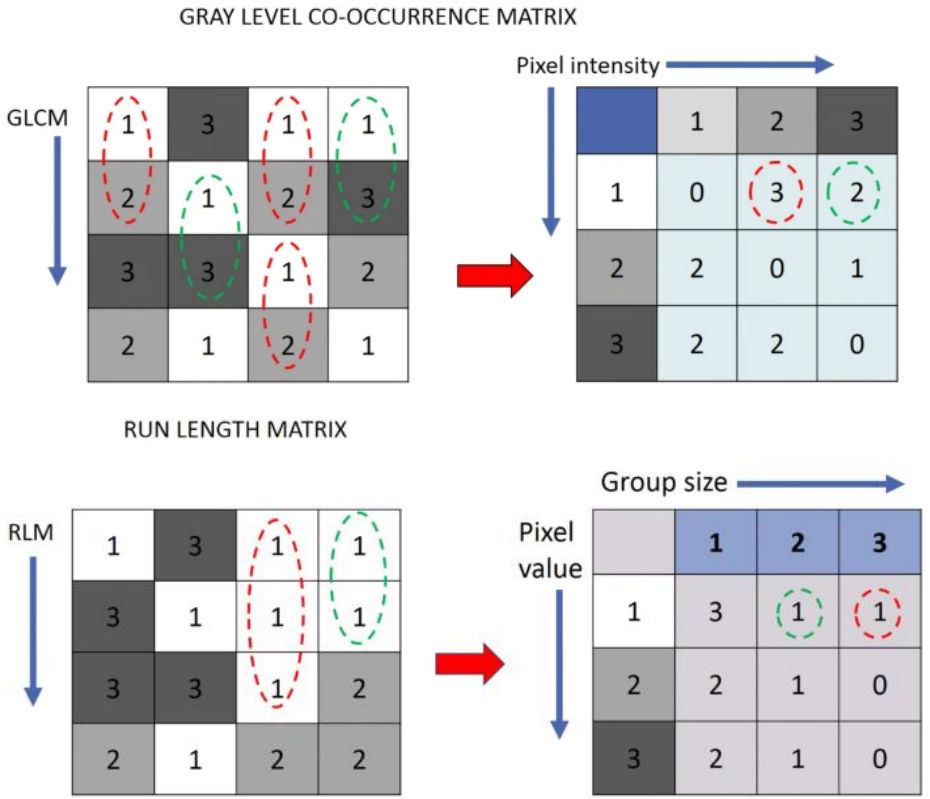
SER Edge = intensity in masked image:intensity in unmasked image

Evans, L. et al. Journal of Microscopy 252:1-7 (2013)

Example of texture features

- GLCM = histogram of paired intensity in adjacent pixels
  - Diagonal = no intensity change
  - Distance from diagonal reflects magnitude of intensity change

- RLM = histogram of consecutive pixels with the same intensity
  - Size of pixel patches with no intensity change



Thomas, R. et al. Abdominal Radiology 44:2501-2510 (2019)

Examples of image-based features

ImageNumber	ObjectNumber	AreaShape_Area	AreaShape_BoundingBoxArea	AreaShape_Center_X	AreaShape_Center_Y	AreaShape_Compactness	AreaShape_ConvexArea	AreaShape_Eccentricity	AreaShape_EquivalentDiameter
1	1	4668	6320	79.22836332	66.53663239	1.28539909	4974	0.332132659	77.09398287
1	2	4444	8099	634.5335284	77.38253825	2.780957604	5622	0.666699148	75.22151645
1	3	3640	7474	407.9239011	71.29615385	1.787572618	4275	0.710994905	68.07783738
1	4	3375	4810	748.8103704	112.6802963	1.185275098	3454	0.706339156	65.55290584
1	5	4323	6192	381.864446	125.8565811	1.216229044	4500	0.649281856	74.19039393
1	6	4811	6278	131.8667637	141.361879	1.406683646	5123	0.589402987	78.26592777
1	7	4261	5810	575.1783619	161.8955644	1.646662247	4778	0.598052282	73.65645729
1	8	6908	11948	1009.645049	151.0616676	1.995367921	8298	0.764000448	93.78453377
1	9	4873	7636	200.1875641	162.17874	1.572888841	5416	0.701271667	78.76862511
1	10	3872	5110	63.13197314	160.5599174	1.130655945	3965	0.497447565	70.21384135

- **Area and shape** = geometry of the nucleus and cell
- **Intensity** = summary statistics of staining signals
- **Neighbors** = relative positioning of cells with different signals
- **Radial distribution** = polarity of staining signals
- **Texture** = summary of changes in signals across adjacent pixels
- **Correlations** = correlation of features (across multiple proteins or cells)

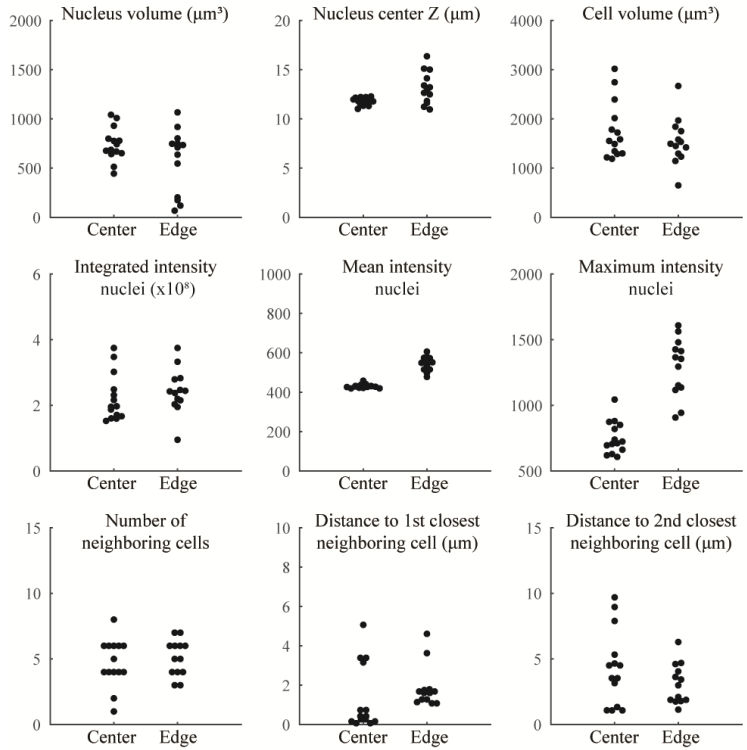
Screening for useful image-based features

Shotgun feature extraction approach

- Microscopy image analysis software can produce >1,000 image-based features for each cell
- Same principle as how omics techniques produce measurement for all SNPs, genes, proteins, metabolites, etc.
- High sample sizes, noisy measurements
  - Similar to single-cell techniques

Univariate feature selection

- **Goal**: identify features that change significantly across cell types or experimental conditions
- **Non-parametric test of means**: Mann-Whitney (Wilcoxon) or Kruskal-Wallis
  - Unknown value distribution
- Correct for multiple testing

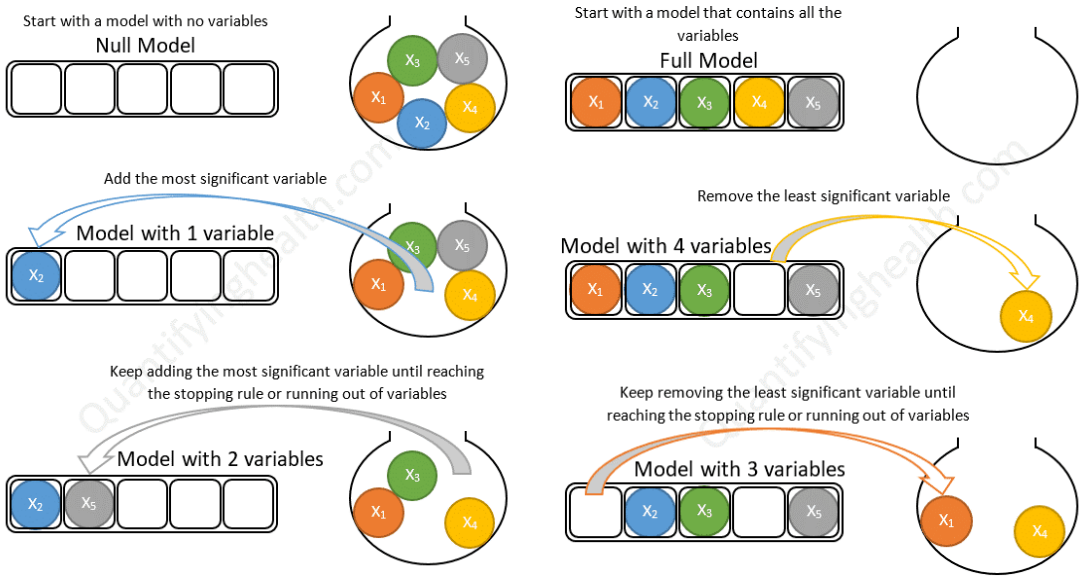


McQuin, C. et al. PLoS Biology 16:e2005970 (2018)

Multivariate feature selection

https://quantifyinghealth.com/stepwise-selection/

Forward:
Iteratively add the most predictive feature (best accuracy) to the model

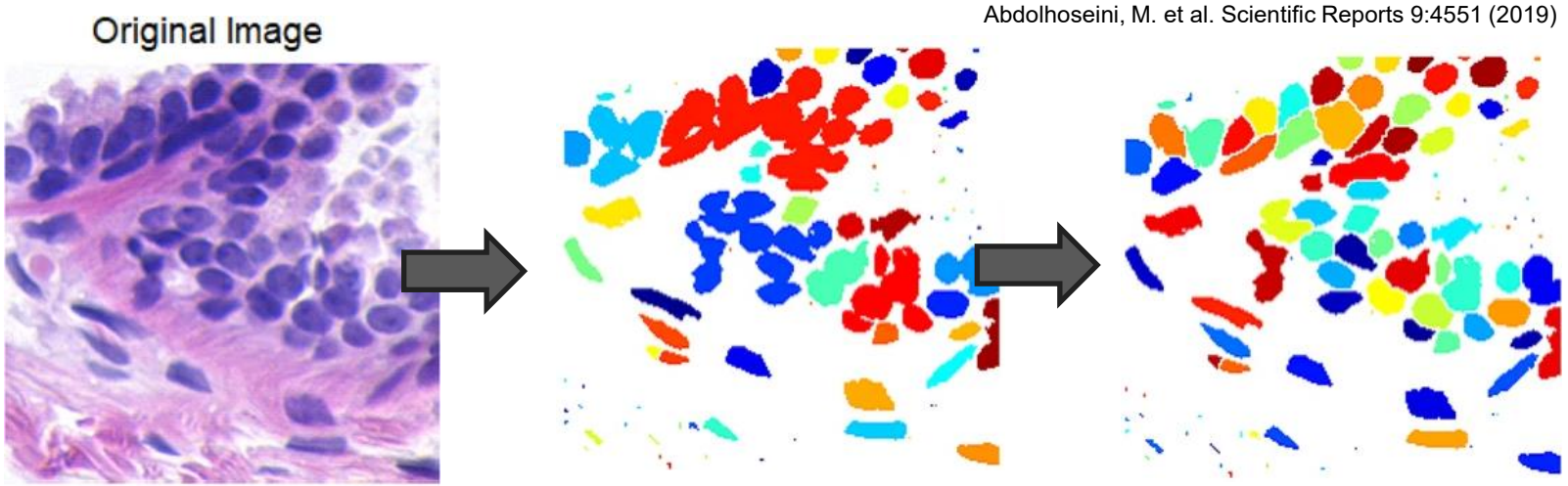


Backward:
Iteratively remove the worst feature (smallest coefficient) from the model

- Identify panel of features that predict cell types or experimental conditions
- Iterative, trial-and-error procedure

Image processing

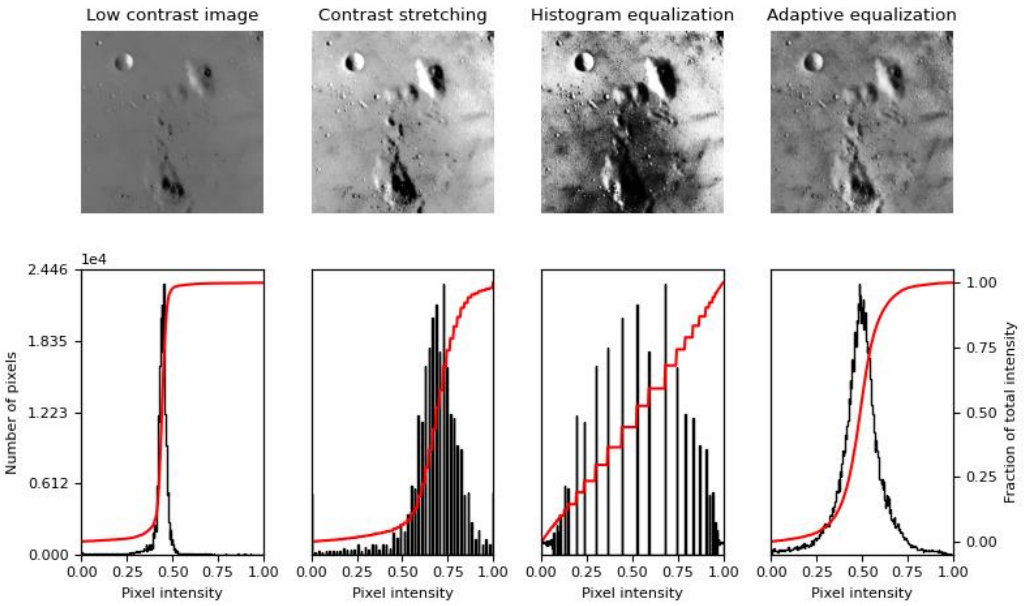
Nuclei detection process



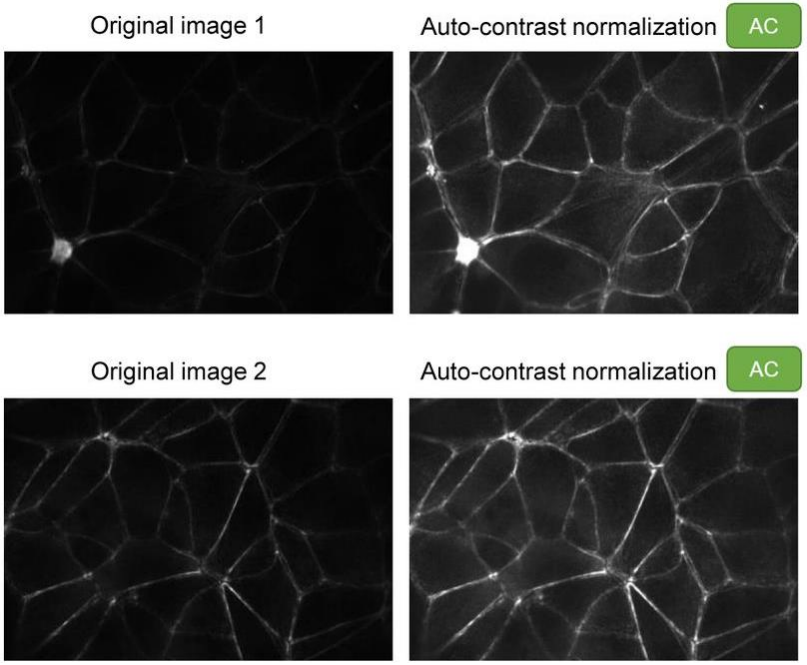
Abdolhoseini, M. et al. Scientific Reports 9:4551 (2019)

- Step 1: Thresholding on nuclei-specific signals (DAPI channel, color hue)
- Step 2: Dividing of “touching nuclei” (clumped/occluded object)
- Step 3: Filter object by sizes

Auto-contrast for visualization



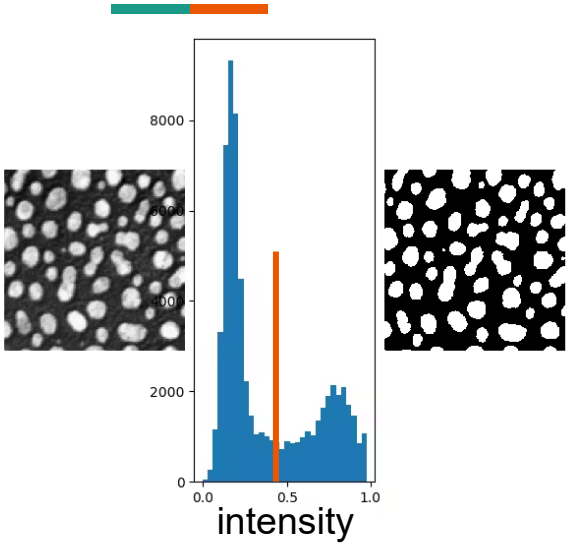
https://scikit-image.org/docs/0.25.x/auto_examples/color_exposure/plot_equalize.html

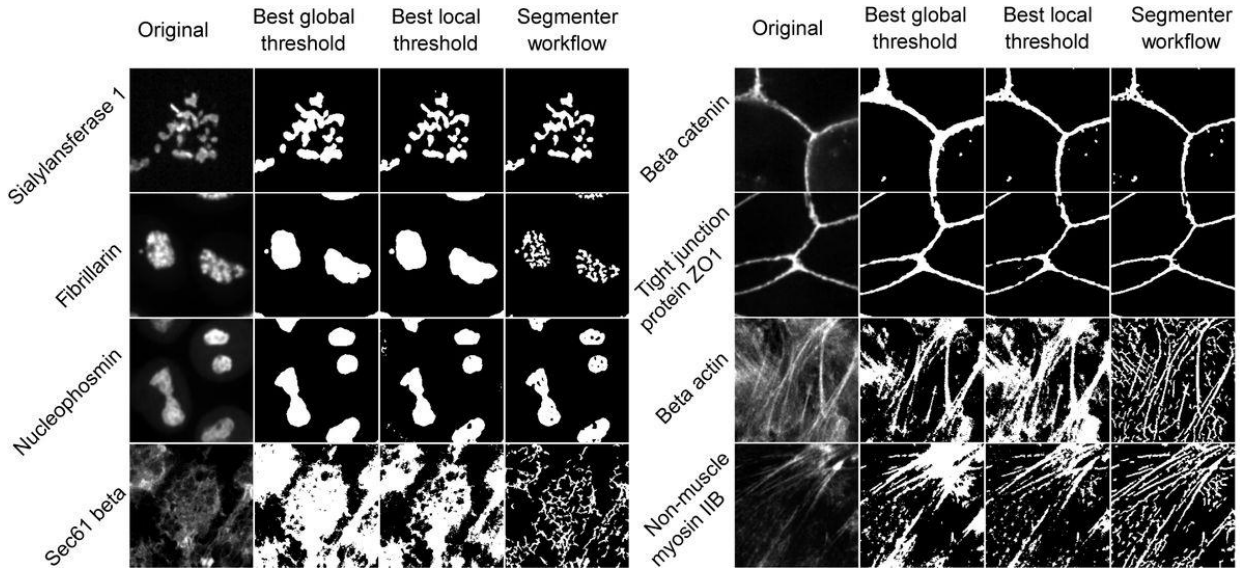


Chen, J. et al. https://www.biorxiv.org/content/10.1101/491035v1.full (2018)

- Make the intensity histogram broader and more uniform

Thresholding



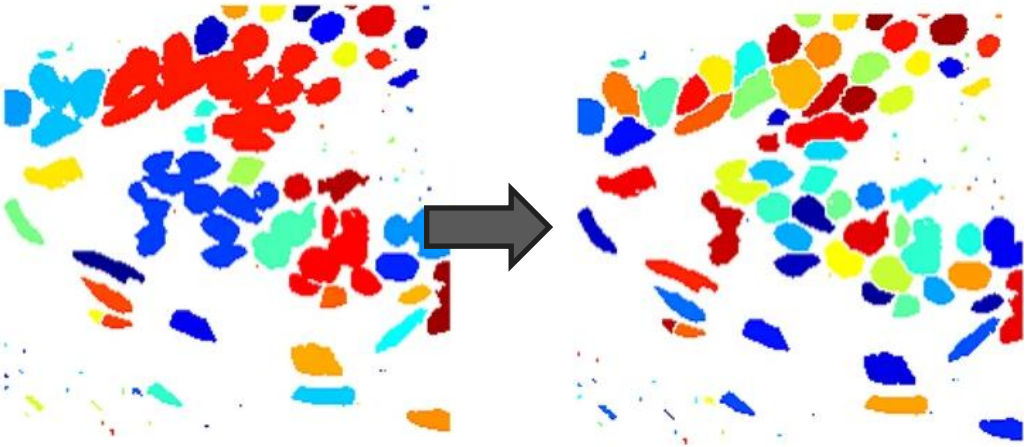


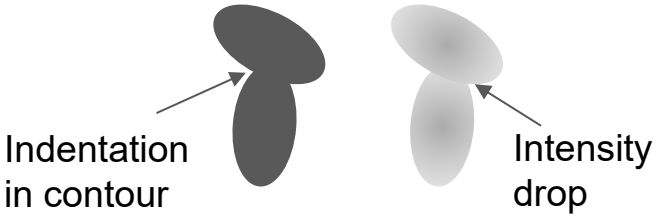
https://encord.com/blog/image-thresholding-image-processing/

Chen, J. et al. https://www.biorxiv.org/content/10.1101/491035v1.full (2018)

- **Assumption**: Foreground objects are brighter than background
- Local thresholding = adaptive cutoff (objects can vary in signal)

Declumping approaches

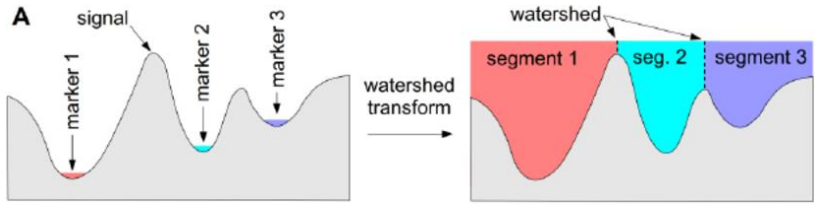




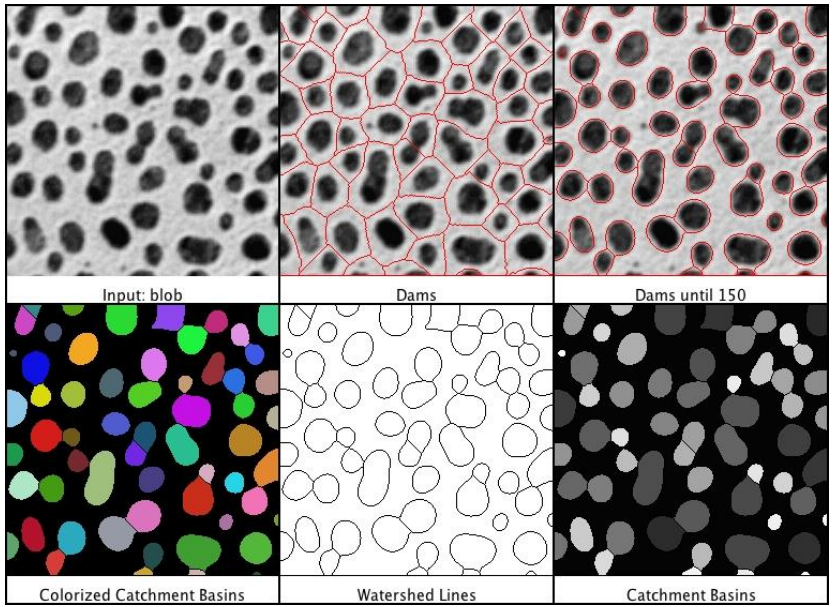
- **Assumption**: when two objects overlap, there will be an obvious change in the geometry of the contour and sometimes a change in intensity gradient

Watershed algorithm

Fisher, A. Remote Sensing 6:776-800 (2014)

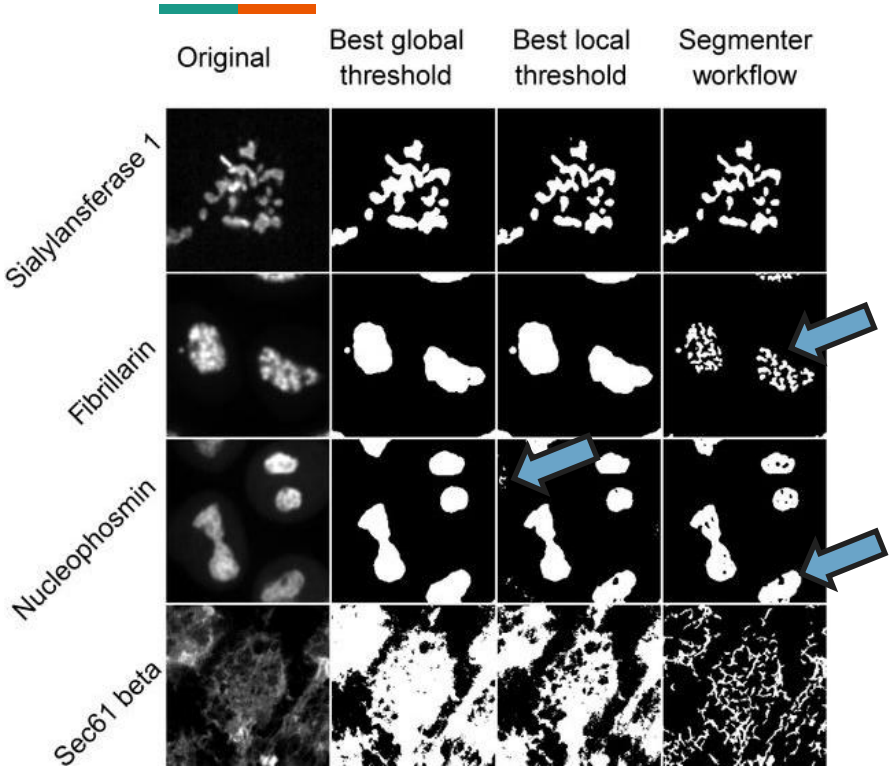


- Model pixel intensities as the water level (invert so that object is dark)
- Each object = bottom of a reservoir
- Fill in water until everything meet
  - Or up to a certain water level
  - Or up to a certain radius



https://bigwww.epfl.ch/sage/soft/watershed/

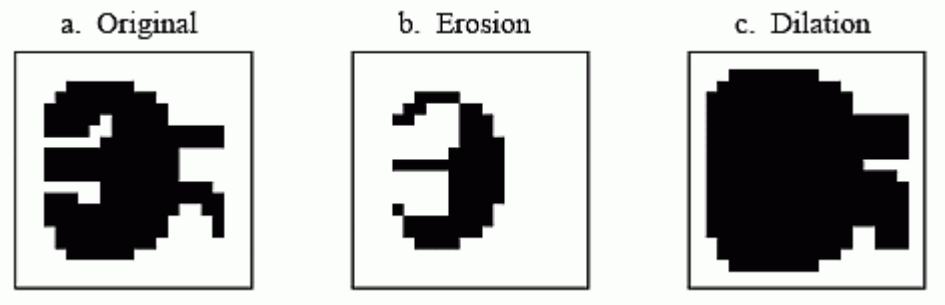
Cleaning up imperfect segmentation



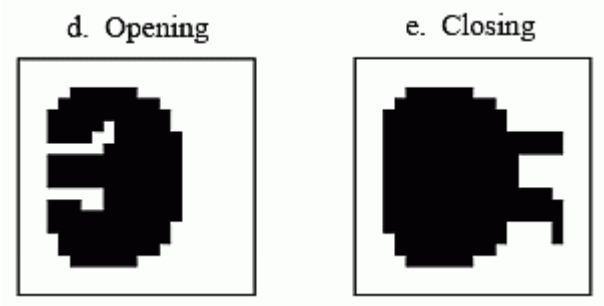
- How to connect fragmented objects?
- How to remove small objects? (without knowing a good size cutoff)
- How to fill in holes, cavities?
- **Morphological operations**
  - Erosion / dilation
  - Opening / closing

Chen, J. et al. https://www.biorxiv.org/content/10.1101/491035v1.full (2018)

Morphological operations



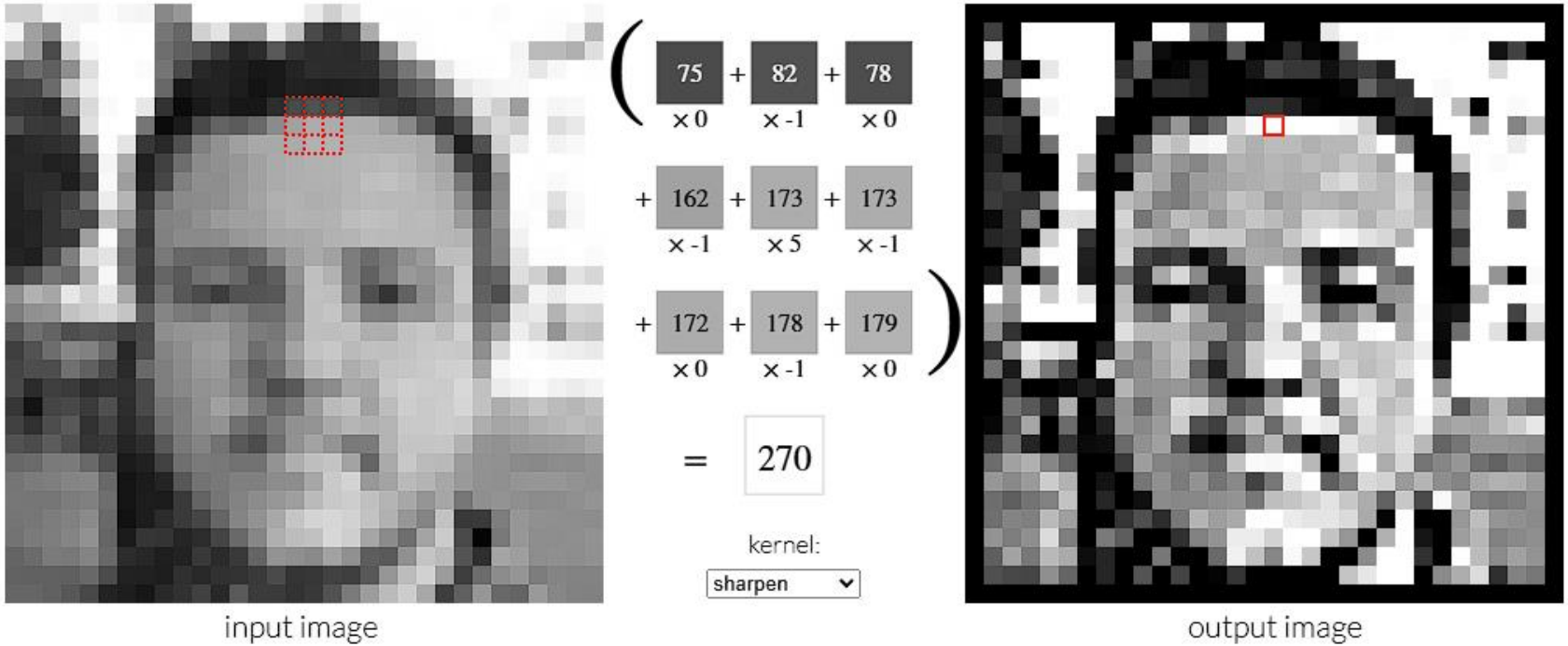
https://www.dspguide.com/ch25/4.htm



- Iterative thinning or expanding by 1 pixel
- Erosion removes thin lines
- Dilation fills in holes
- Opening = erosion + dilation
- Closing = dilation + erosion

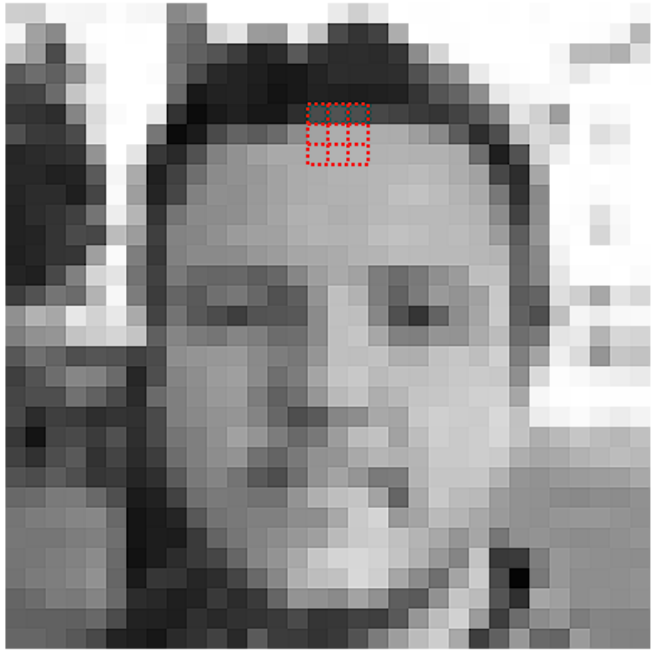
Image processing with kernel (convolutional operation)

Image processing with kernel (filter)



- **Meaningful transformation of image** with the matrix dot product

Feature extraction with kernel (filter)



input image

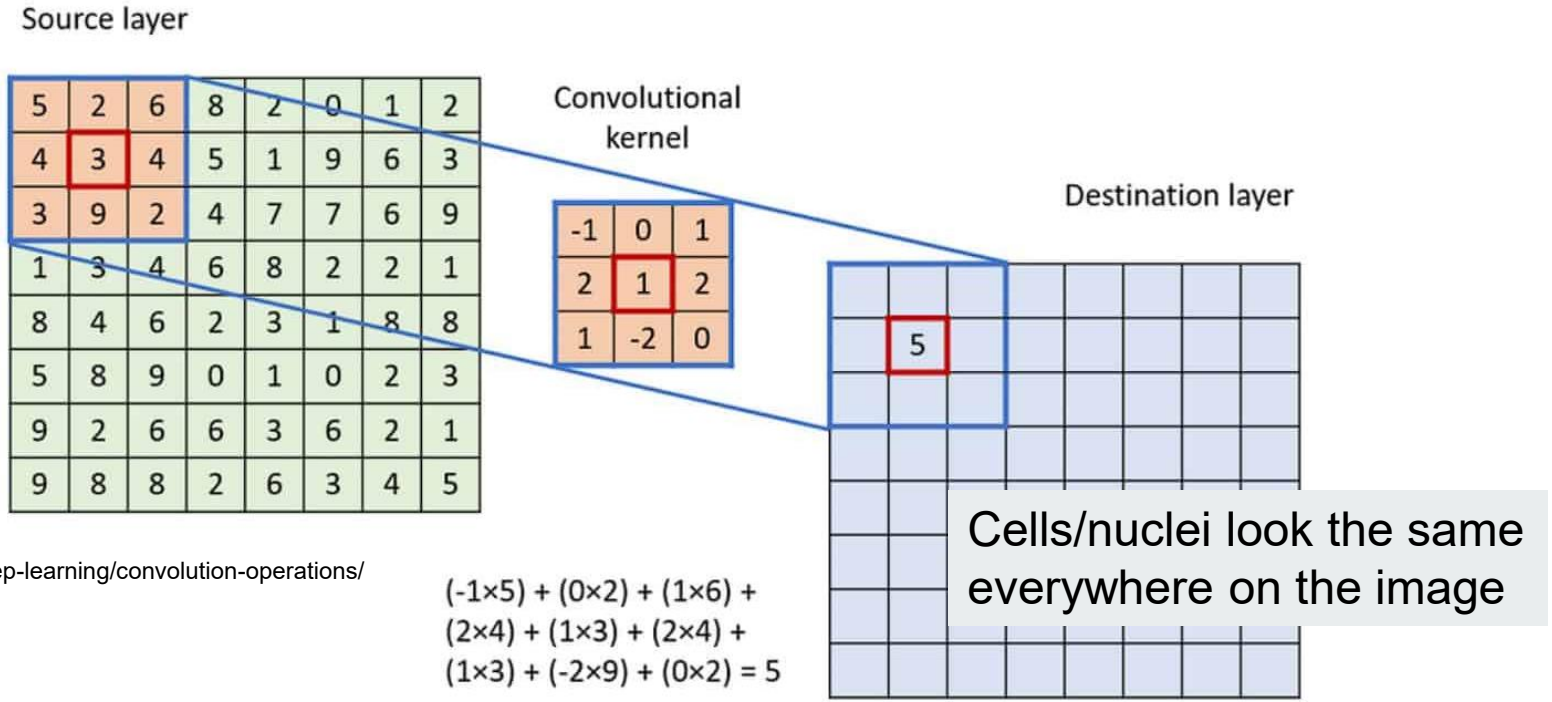
$$\Big(\ 75\times -1 + 82\times -1 + 78\times -1 + 162\times -1 + 173\times 8 + 173\times -1 + 172\times -1 + 178\times -1 + 179\times -1\ \Big) = 285$$

kernel: outline

output image

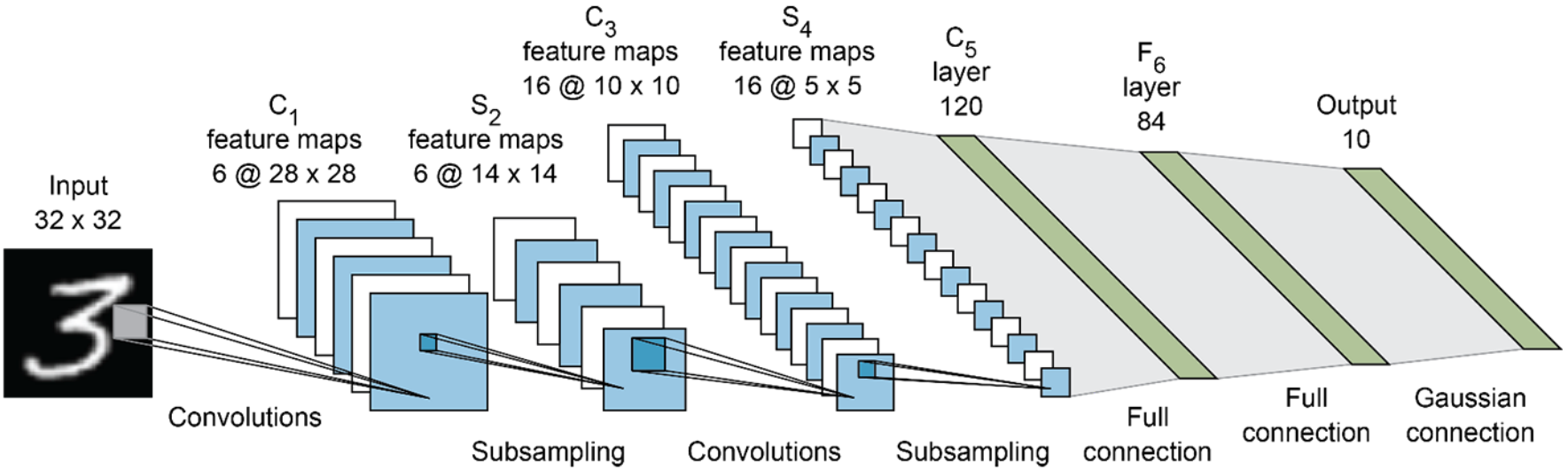
- **Presence of edge** can be extracted with the right kernel

Convolutional operation = repeated kernel application



- **Assumption**: Pattern of signal does not depend on locations

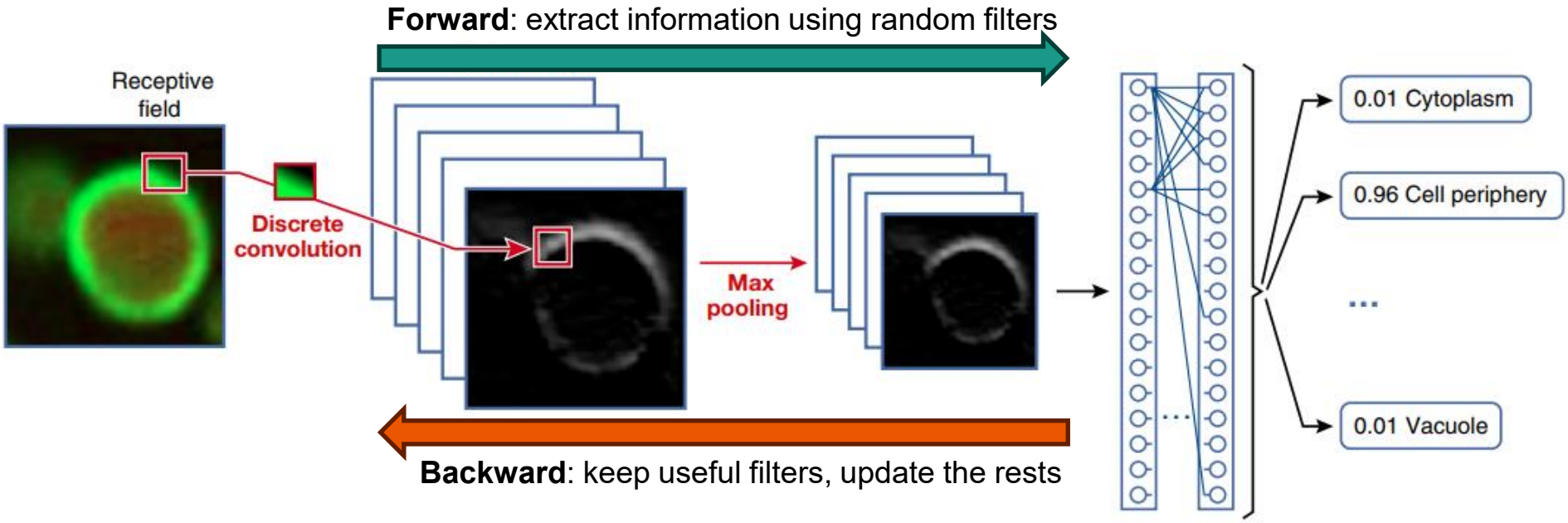
Convolutional neural network



https://www.superannotate.com/blog/guide-to-convolutional-neural-networks

- Stacked layers of convolutional operations to process raw images
- Connected to classifier or regressor to make prediction

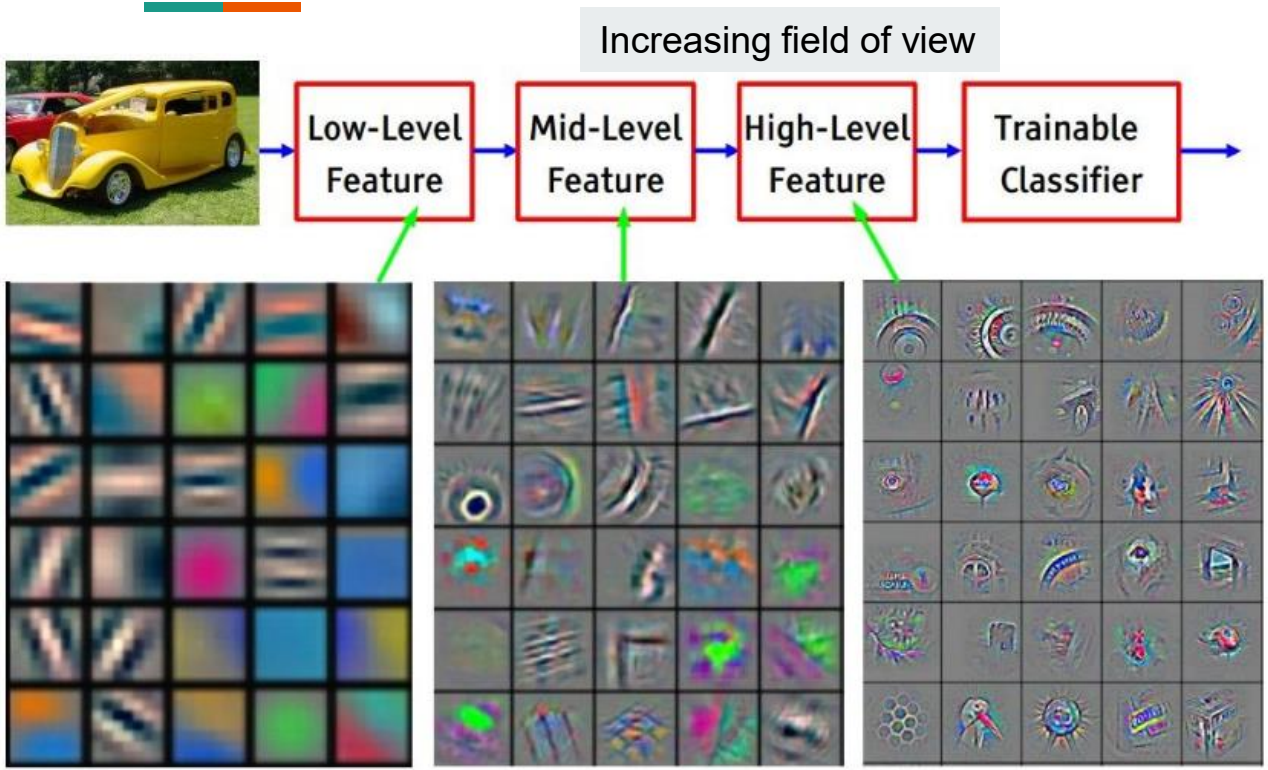
Data-driven optimization of kernels



- CNN learns appropriate kernels from the data
  - But need a lot of data to learn from

Angermueller, C. et al. Molecular Systems Biology 12:878 (2016)

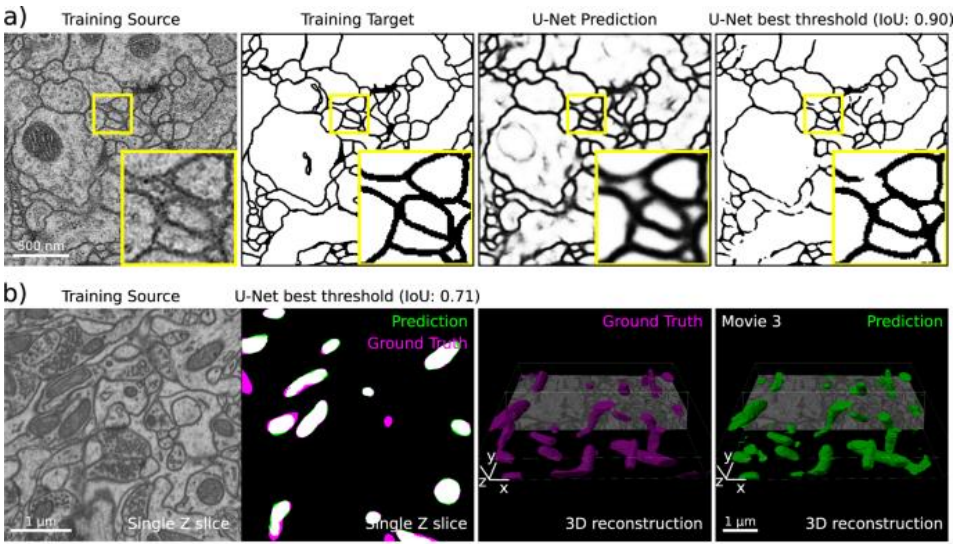
Hierarchical reconstruction in CNN



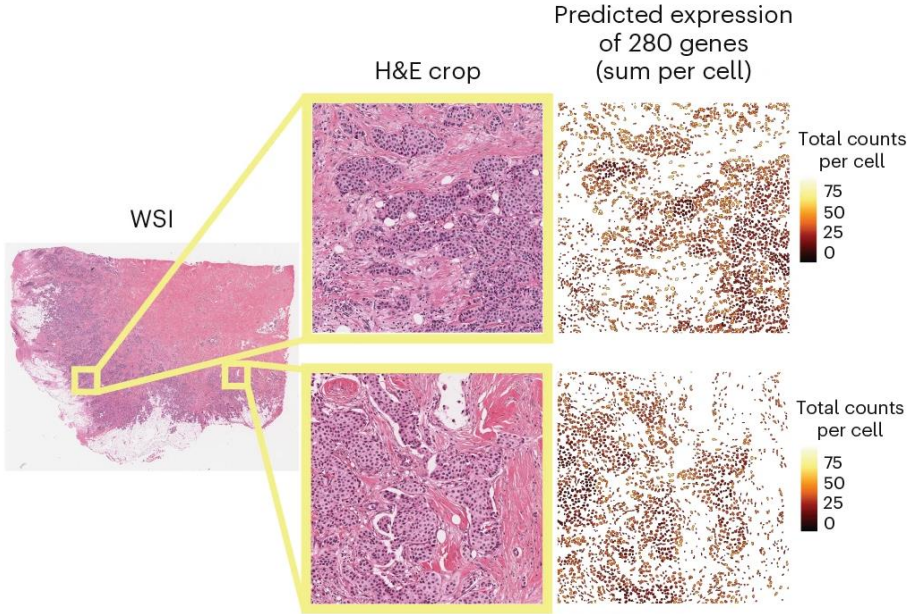
- Early layers capture simple patterns with narrow field of view
- Middle layers combine simple patterns into shapes
- Last layers produce recognizable objects

http://cs231n.stanford.edu/slides/winter1516_lecture7.pdf

Maturity of computer vision in biology



von Chamier, L. et al. Nature Communications 12:2276 (2021)



Fu, X. et al. Nature Methods 22:1900-1920 (2025)

- AI models for detecting cells, classifying cells, and staining the images are becoming mainstream, some can even adapt to your data

Any question?

- See you next time