Predicting Segmentation Accuracy for Biological Cell Images

Adele Peskin[1], Alden Dima[2], Joe Chalfoun[2], John Elliott[2]
[1]NIST, Boulder, CO [2]NIST, Gaithersburg, MD

Abstract - By comparing segmentation results from approximately 40000 cell images of two mammalian cell lines with a variety of segmentation techniques, we find a linear relationship between the highest segmentation accuracy seen for a given cell and the fraction of pixels in the neighborhood of the edge of that cell. This fraction of pixels is at greatest risk for error when cells are segmented. We call the ratio of the size of this pixel fraction to the size of the cell the extended edge neighborhood and suggest that this metric can predict segmentation accuracy of any isolated cell within our image set.

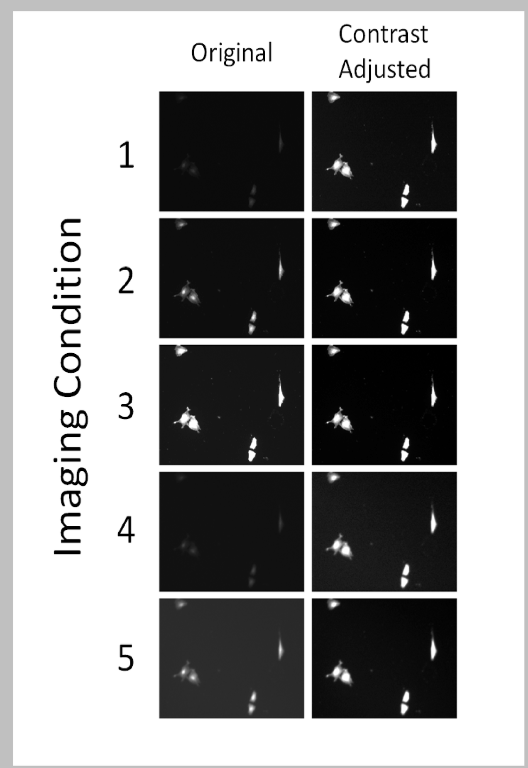
Data Description

2 mammalian cell lines: A10 & NIH 3T3

5 different image settings

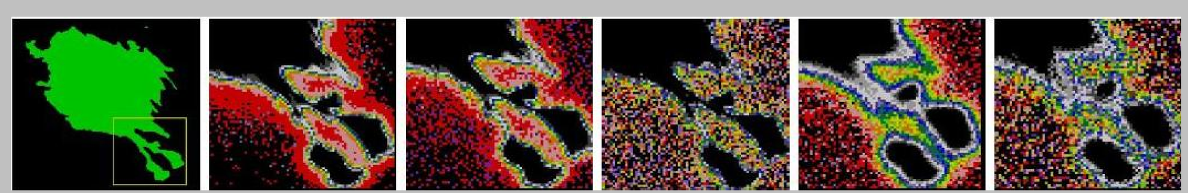
Reference data automatically generated:

allows us to scale up



Edge Quality: Quality Index Calculation

1. Find: 3-component intensity distribution: background, edge, cell.
2. Find: Average gradient magnitude as f(intensity) at the cell edge.
3. Smooth the gradient in this region to fill in any gaps, G(Intensity).
4. Find the physical thickness of the edge in the image.

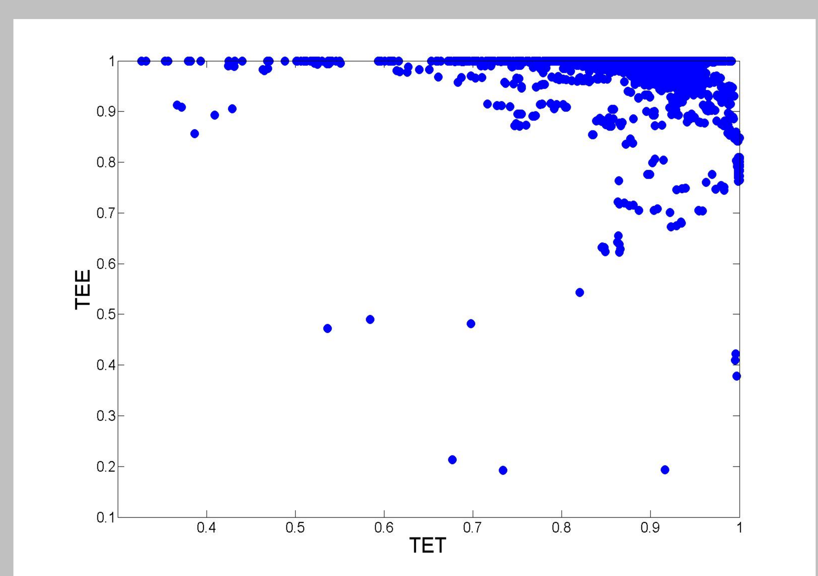


The mask for cell 5 of set 2, and enlarged pictures of the lower right hand marked section, in order of decreasing quality index, imaging conditions 3,2,1,5,4. , the quality index says that the edge covers about 2, 2, 3, 4, and 4.5 pixel lengths in the images above.

Segmentation Distance

To compare the reference pixel set T, with a segmentation mask E, we define:

$S = \frac{\lvert T \cap E \rvert}{\lvert T \cup E \rvert}$	$0 \le S \le 1$	Similarity Index
$TET = \frac{\lvert T \cap E \rvert}{\lvert T \rvert}$	$0 \le TET \le 1$	Bivariate Metric
$TEE = \frac{\lvert T \cap E \rvert}{\lvert E \rvert}$	$0 \le TEE \le 1$	
$d_{seg} = \sqrt{(1-TET)^2 + (1-TEE)^2}$	$0 \le d_{seg} \le \sqrt{2}$	Segmentation Distance

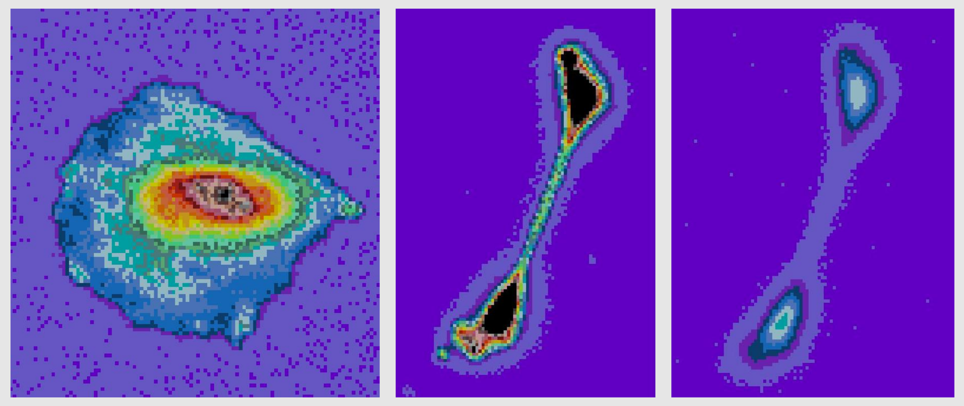


Plot of TET vs. TEE for 3,000 A10 cells; Segmentation distance = distance to (1,1).

Extended Edge Neighborhood

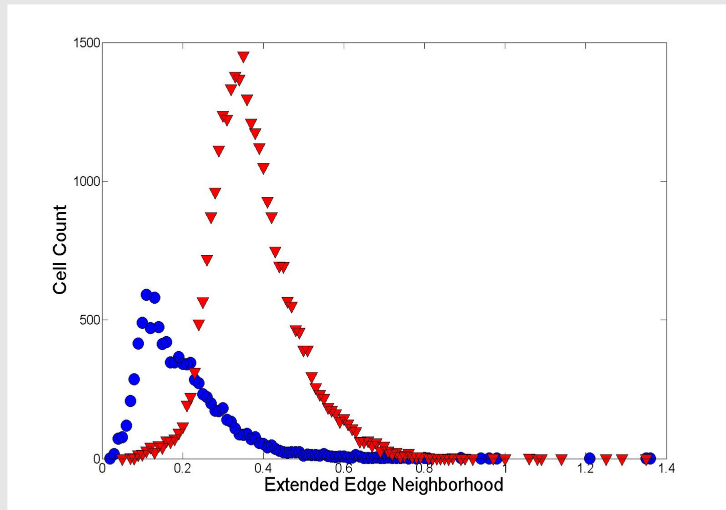
The extended edge neighborhood represents the fraction of a cell image at the cell edge. The edge quality gives the physical pixel size of the edge and the cell geometry defines how much of the image is near the cell boundary.

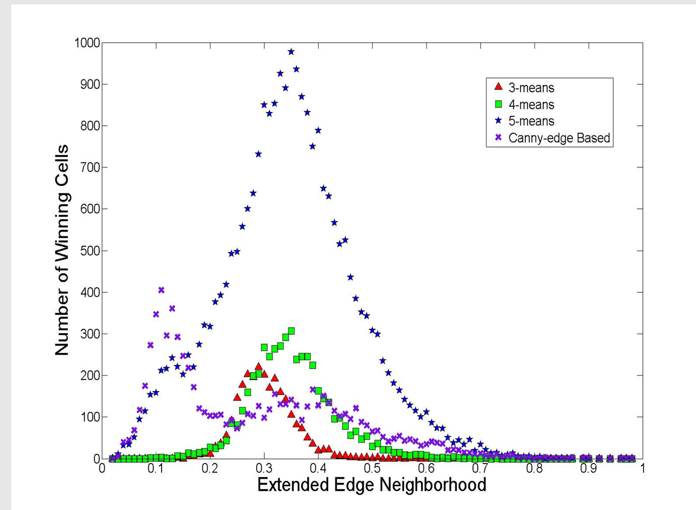
EEN = edge thickness * perimeter / area.



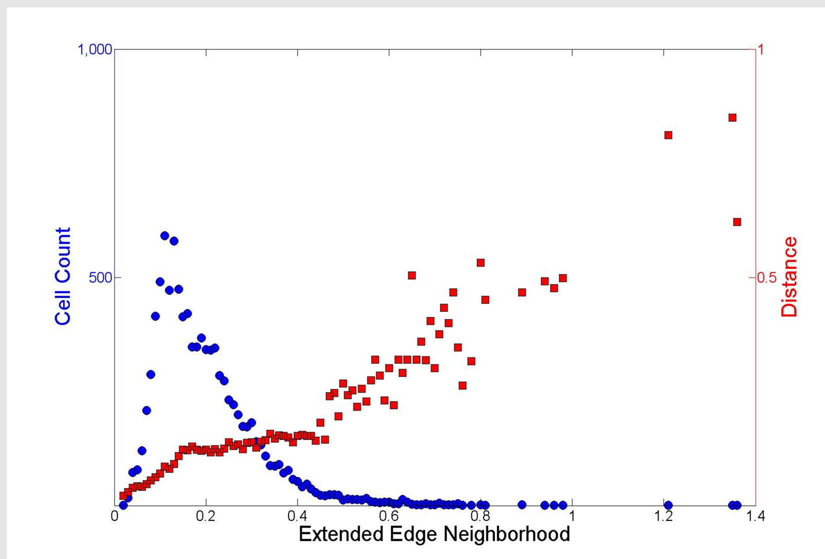
A large, round cell with low extended edge neighborhood; a thin cell with higher extended edge neighborhood, even though the edges are sharp; the same cell but a blurrier image, and the extended edge neighborhood is greater 1.0.

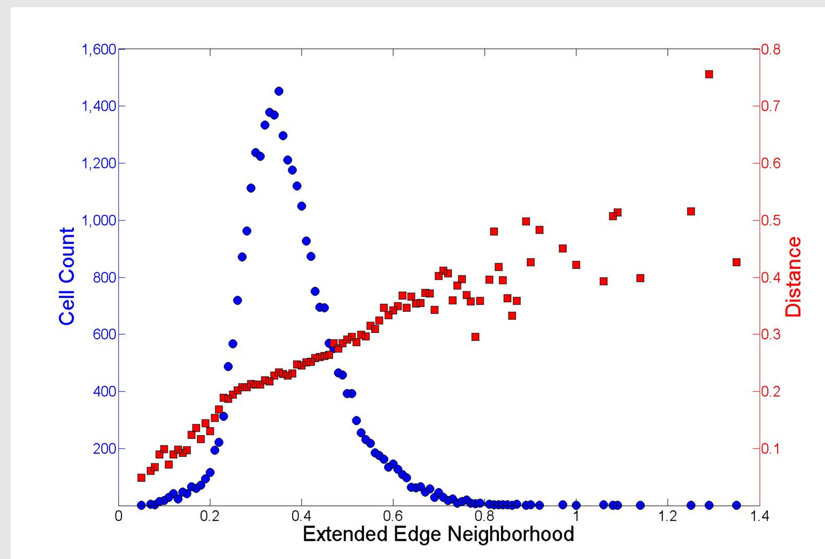
Results

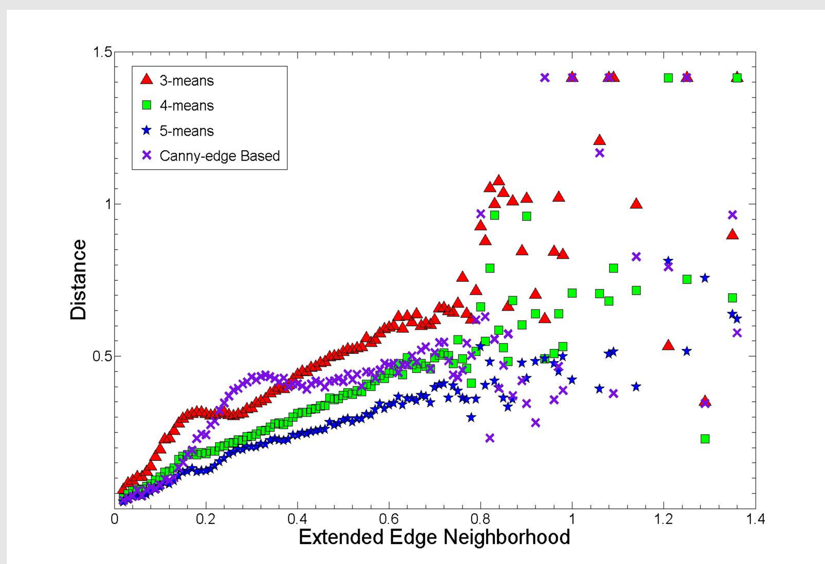


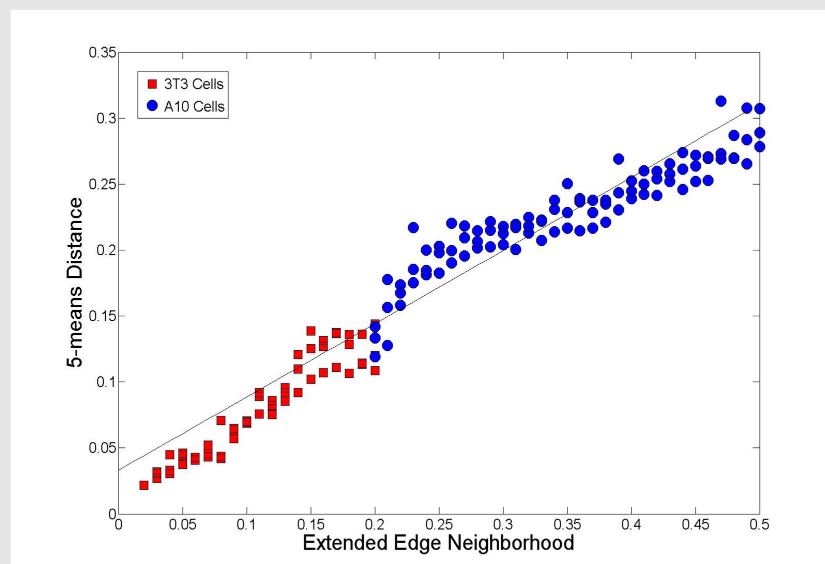


Left: Numbers of cells as a function of extended edge neighborhood for the A10 cells in red, and the 3T3 cells in blue. Right: Number of cells for which a technique worked best: 3-means clustering (red), 4-clustering (green), 5-means clustering (blue), Canny edge (purple), as a function of extended edge neighborhood.









Top row: Segmentation distance and cell count for A10 cells (left), 3T3 cells (right) as a function of extended edge neighborhood. Bottom Left: 3-means clustering (red), 4-means clustering (green), 5-means clustering (blue), and Canny edge (purple), for the whole data set. Bottom Right: Segmentation distance for both cell lines fit to a line: distance = 0.051 + EEN × 0.477, with a correlation coefficient of 0.9815.

Conclusions

From this large scale test, we define a method to pre-process images and determine their vulnerability to segmentation error. The accuracy that is possible from the segmentation techniques that we have considered is directly proportional to the extended edge neighborhood of each individual cell within an image.

NIST National Institute of Standards and Technology U.S. Department of Commerce