Effects of stroke width on semi-automatic image segmentation

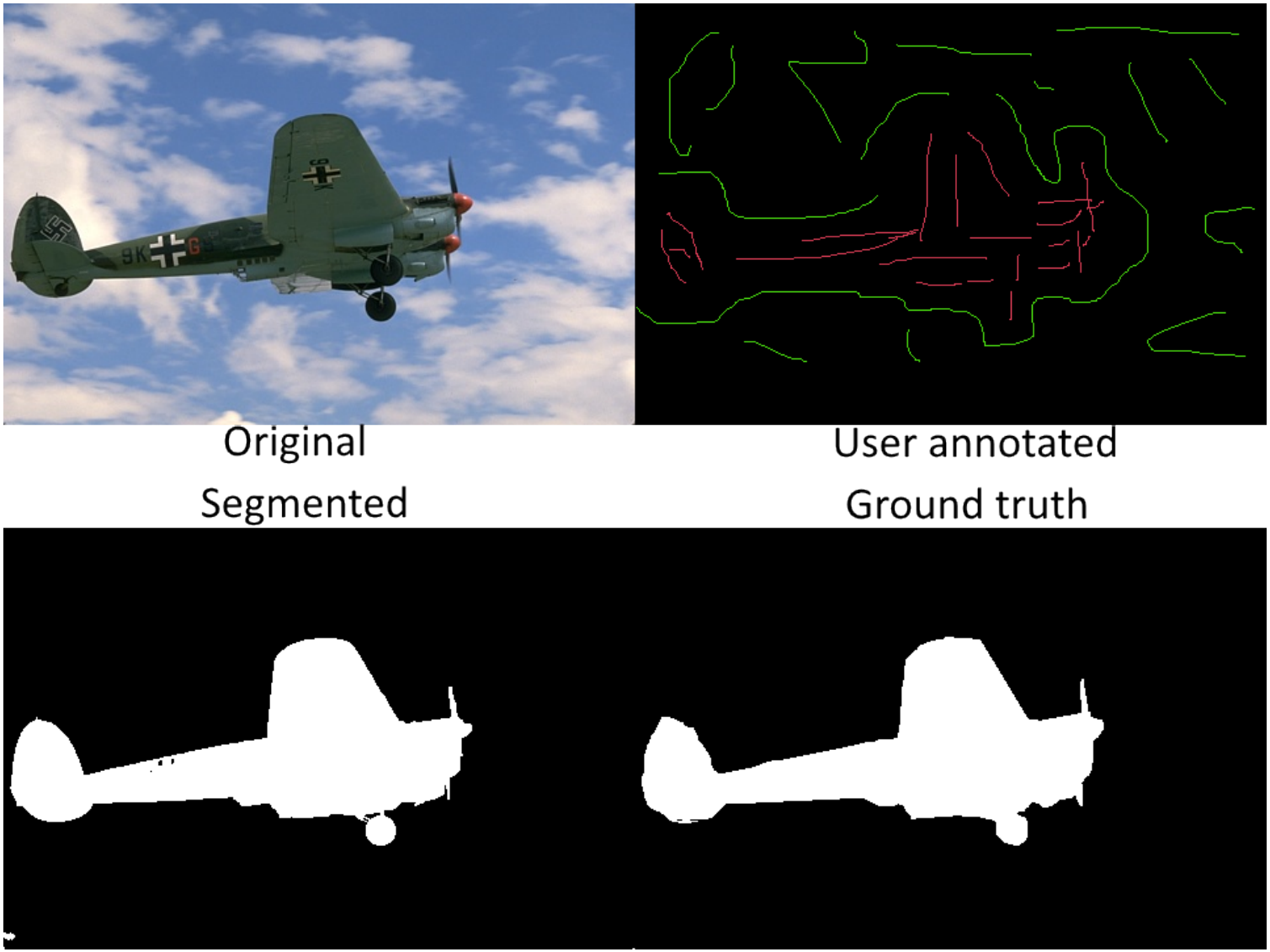
Katrina Hoffert
Under the supervision of Dr Mark Eramian

2016-2017

Semi-automatic image segmentation

- The goal of image segmentation is to portion an image into distinct segments.
- Semi-automatic image segmentation uses some degree of user interaction to assist in this segmentation.
- The method of user interaction used in our study is users annotating what is foreground and what is background.

Segmentation example

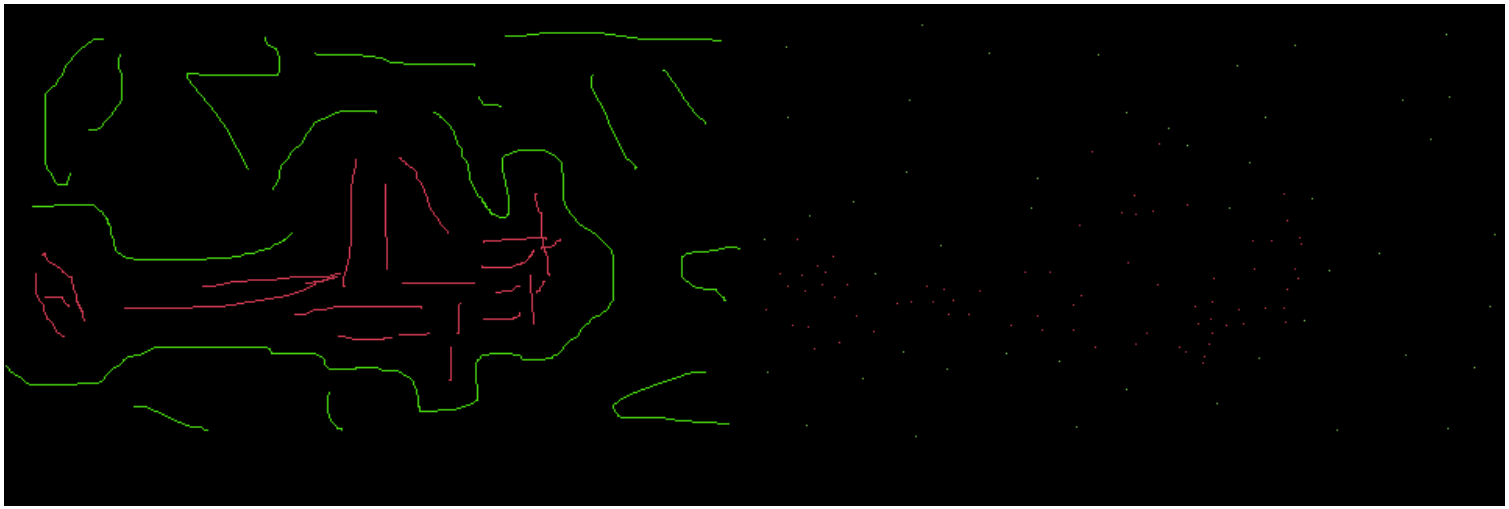


Data sources

- Images used and their ground truths all come from the Berkeley Segmentation Data Set and Benchmarks 500 (BSDS500).

Data sources

- All data used in this study came from two prior studies:
  - Steven Rau: Contrasted effectiveness of user annotations in points vs strokes.
  - Yuanxia Li: Contrasted effect of time limits on user (used only points)



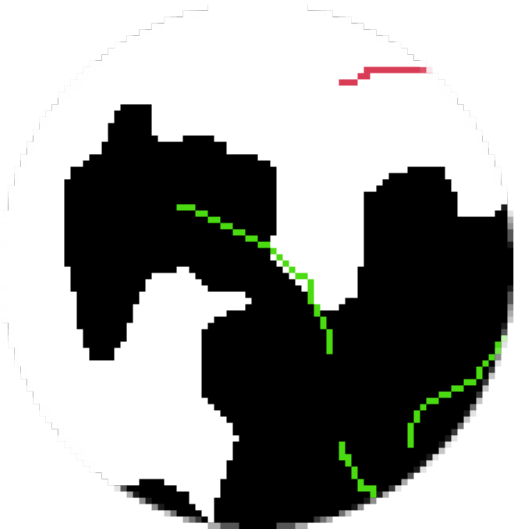
Problem

- All the data in these studies used strokes and points one pixel wide.
- **Question: does the width of these annotations significantly impact results?**
- Or in other words, can we get better results if we dilate all the annotations so that they're thicker?
- Supplies more data to the segmentation algorithm. But how valuable is that data? Can it introduce errors?

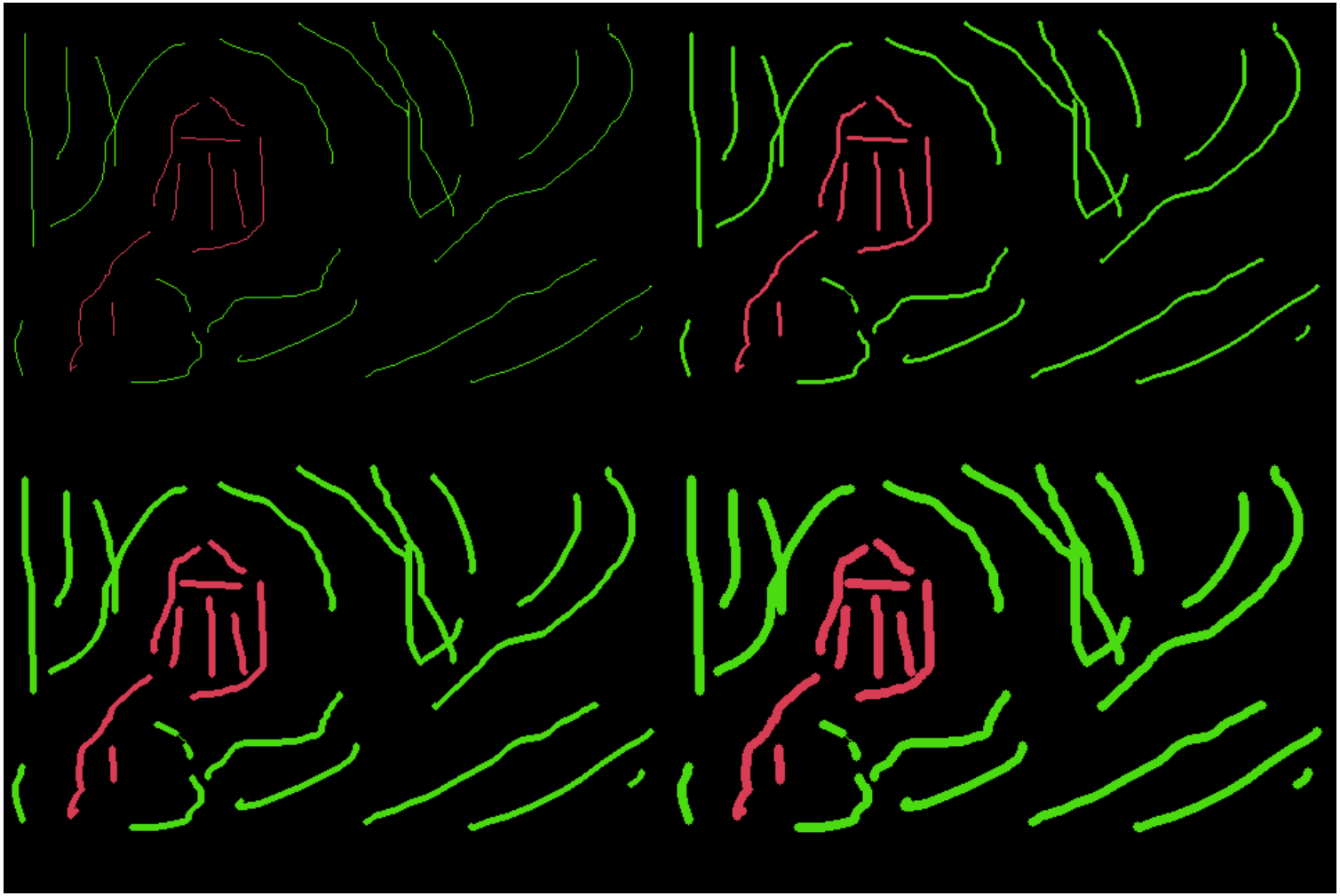
The experiment

- There was the worry of introducing errors that users wouldn't have made unless they were aware of the thicker stroke radius, so our dilation method only dilated if it wouldn't introduce *new* errors in doing so.
- If an area is background (which we know because we have the ground truth), then a foreground pixel that was in the foreground wouldn't be dilated into the background.
- But if a foreground pixel was in the background, it would be dilated as normal.

The experiment

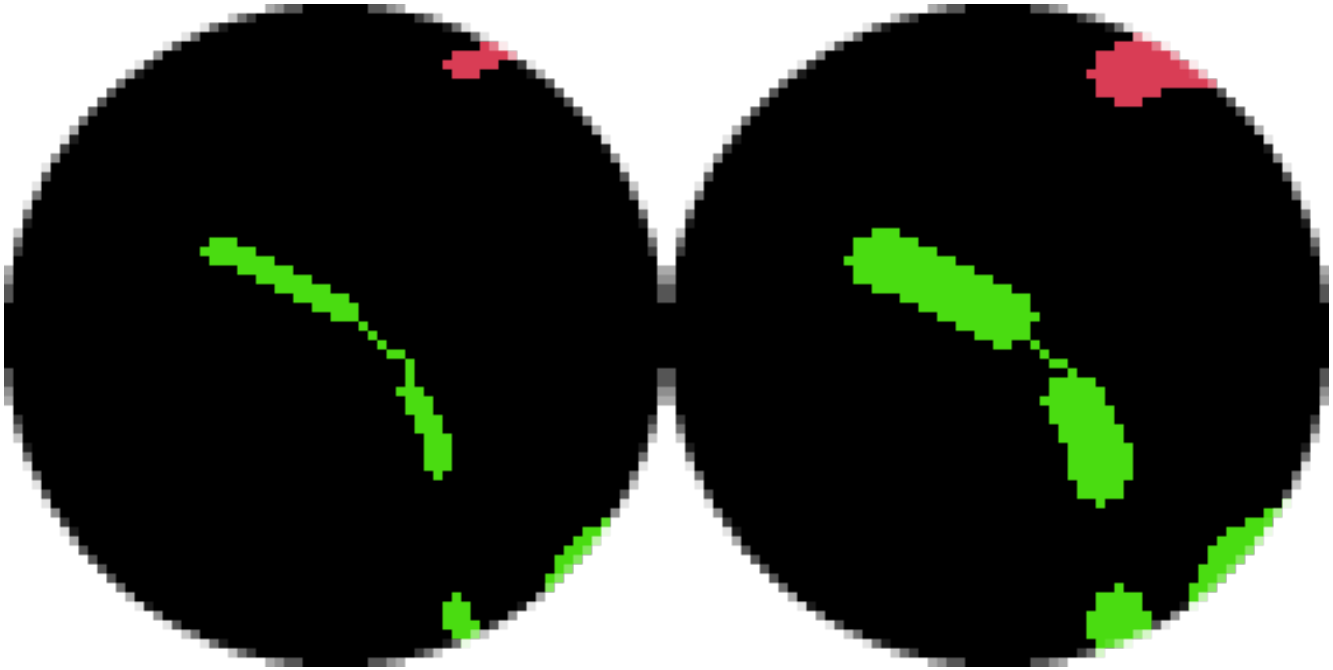


The experiment



The experiment

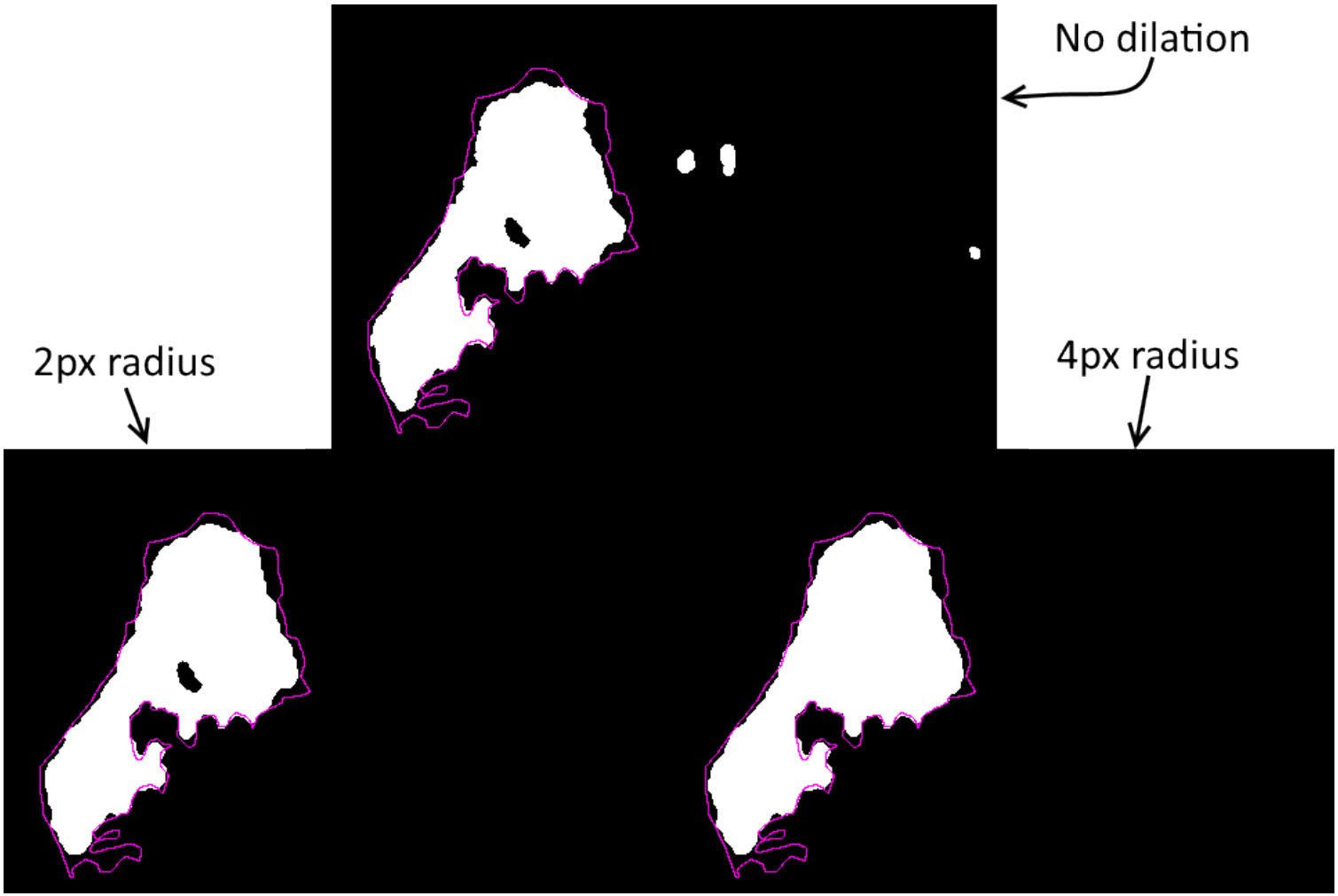
- The idea is that the strokes/points represent the center of where the user actually intended the stroke/point to be.
- Ultimately doesn't seem to make a big difference with the study data and dilation radii we used.



The experiment

- With everything dilated with radii 0 through 4, we performed segmentation with the Boykov graph cut segmentation algorithm.
- This was performed on all images. There was 2075 label images, with 5 different radii of dilation each for a total of 10375 images to segment.
- Analysis is then run on all the segmented images to determine the effects of the dilation.

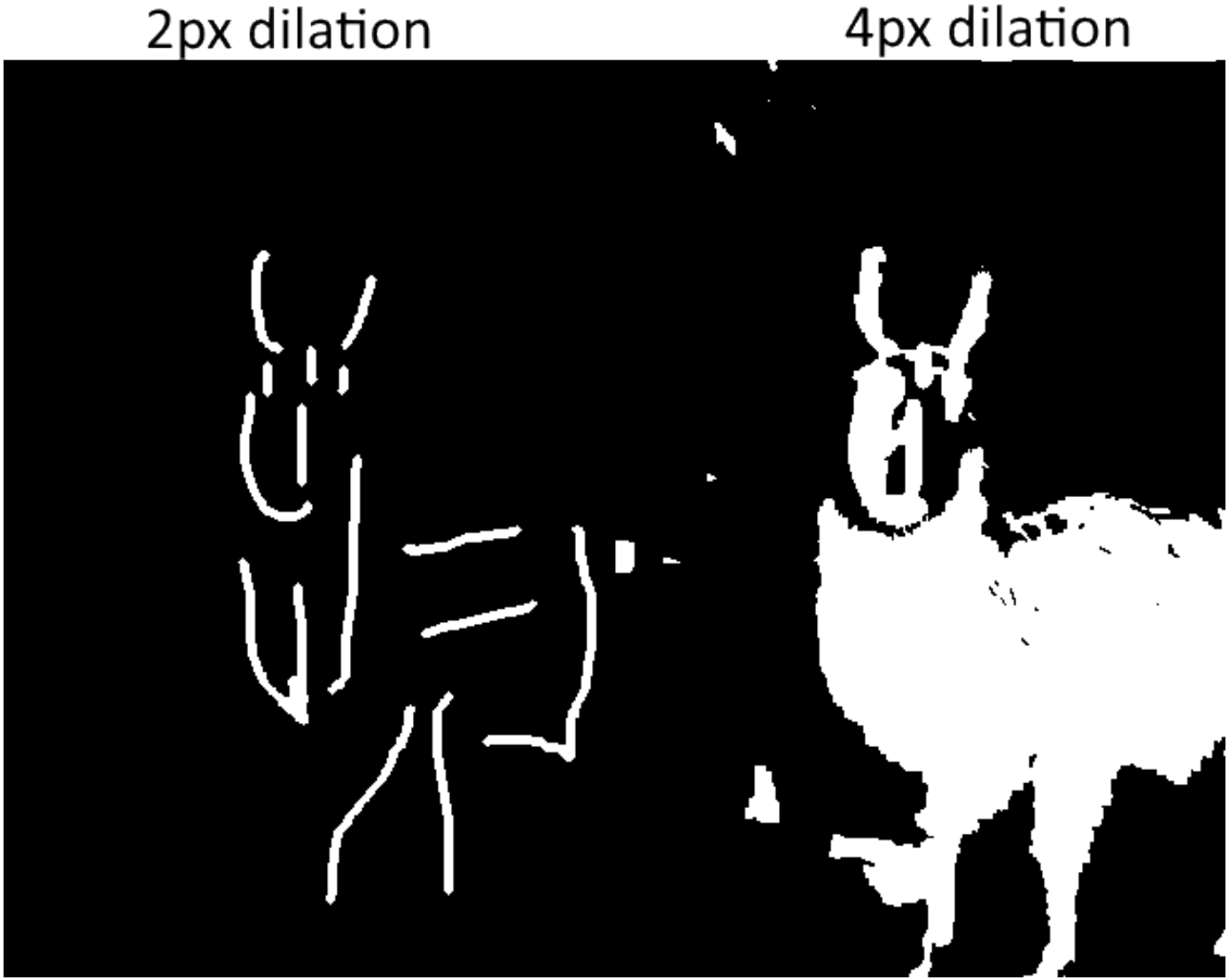
The experiment



Segmentation programs

- We also tried the experiment on a different segmentation program to contrast the differences. The OneCut algorithm was chosen for this.
- Adapted an implementation from Lena Gorelick to fit into our pipeline.
- The OneCut algorithm seems to be much more sensitive to the amount of user input than the Boykov graph cut algorithm.
- Works best where users provided a lot of annotations.

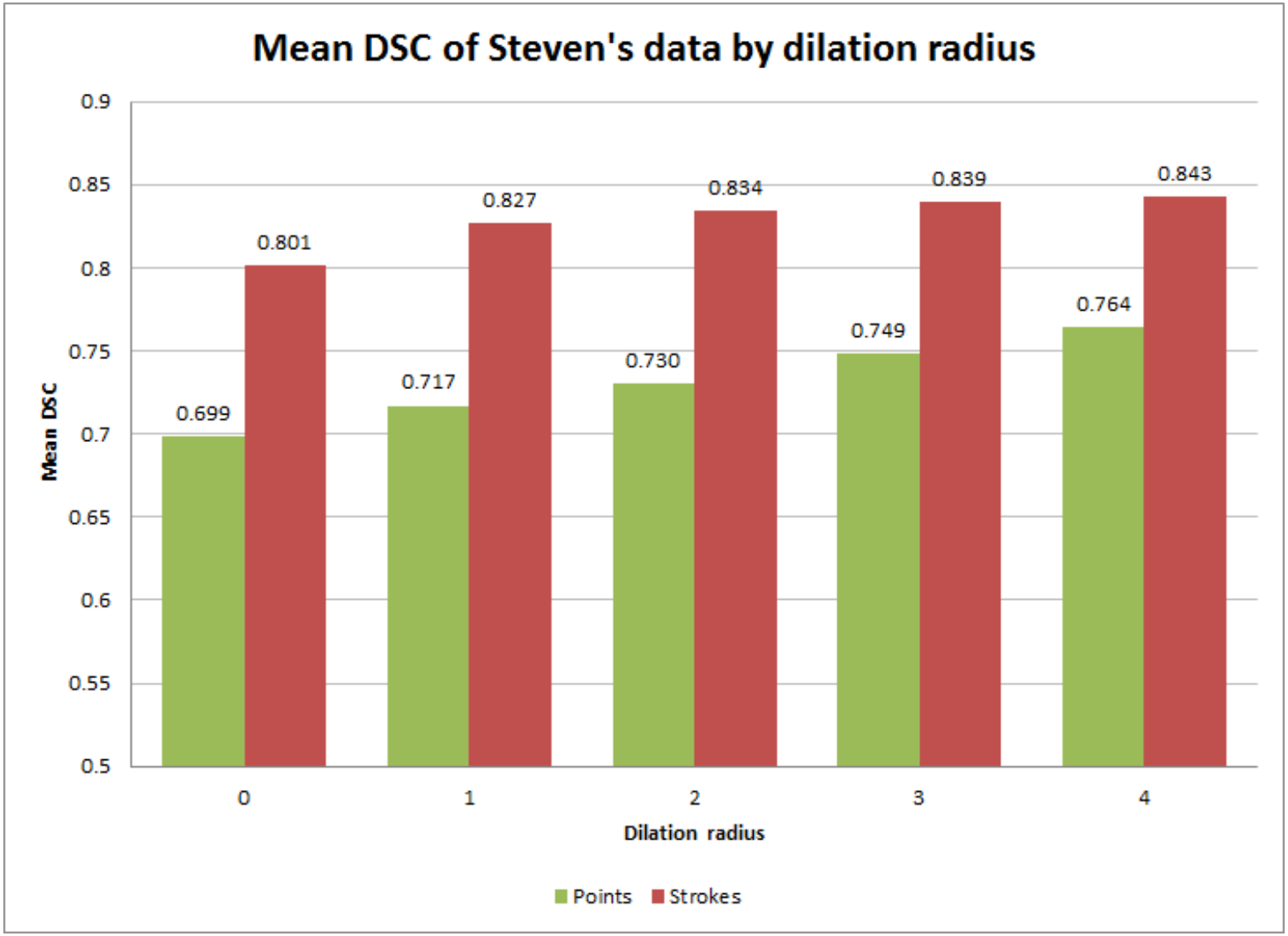
Segmentation programs



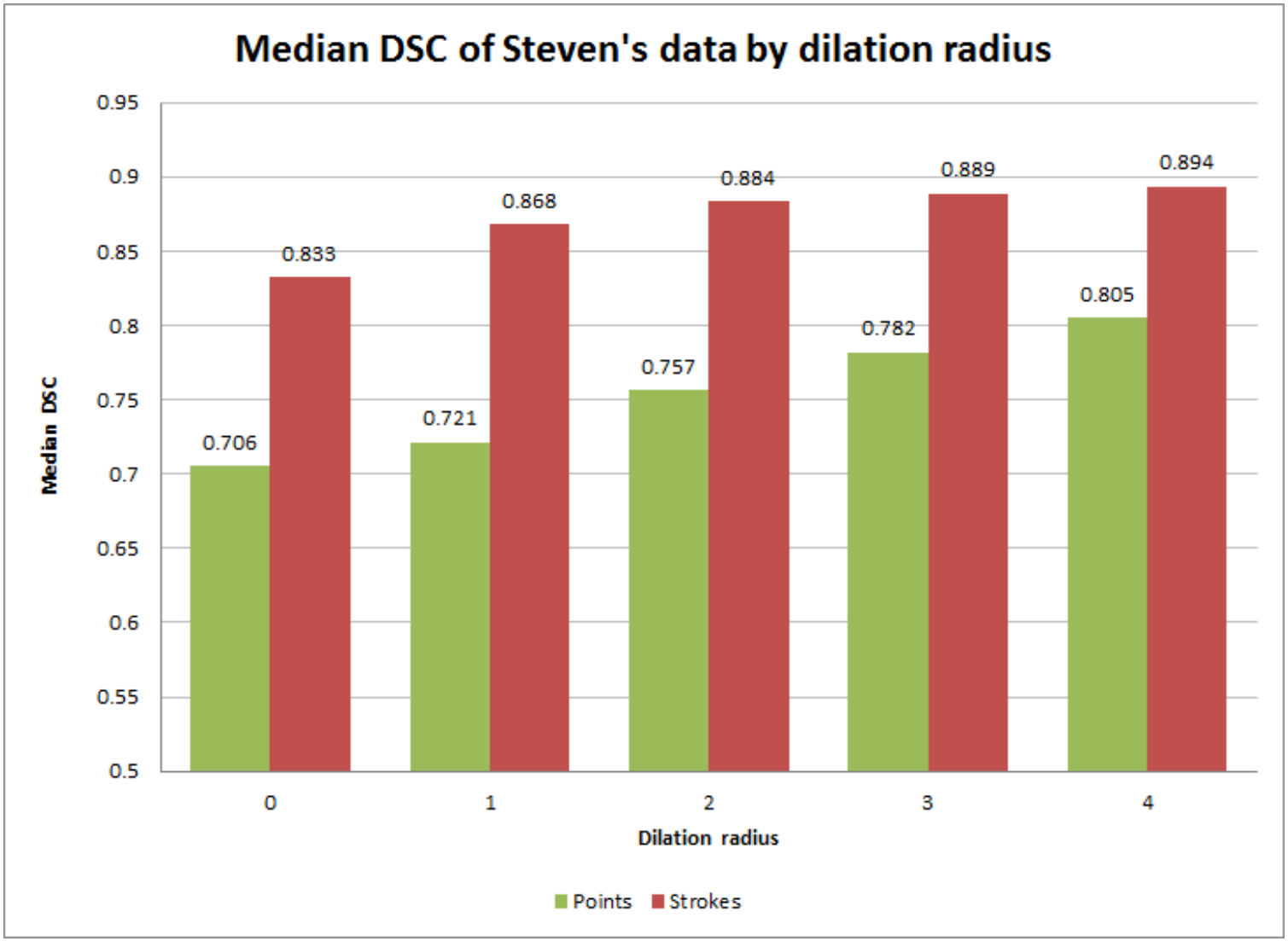
Analyzing accuracy

- Accuracy in image segmentation is most commonly measured with the Dice Similarity Coefficient.
- $\text{DSC} = \frac{2|X \cap Y|}{|X| + |Y|}$
- We use this to compare the segmented result to the ground truth.
- The value is between 0 and 1, where 0 = no match at all and 1 = perfect match.

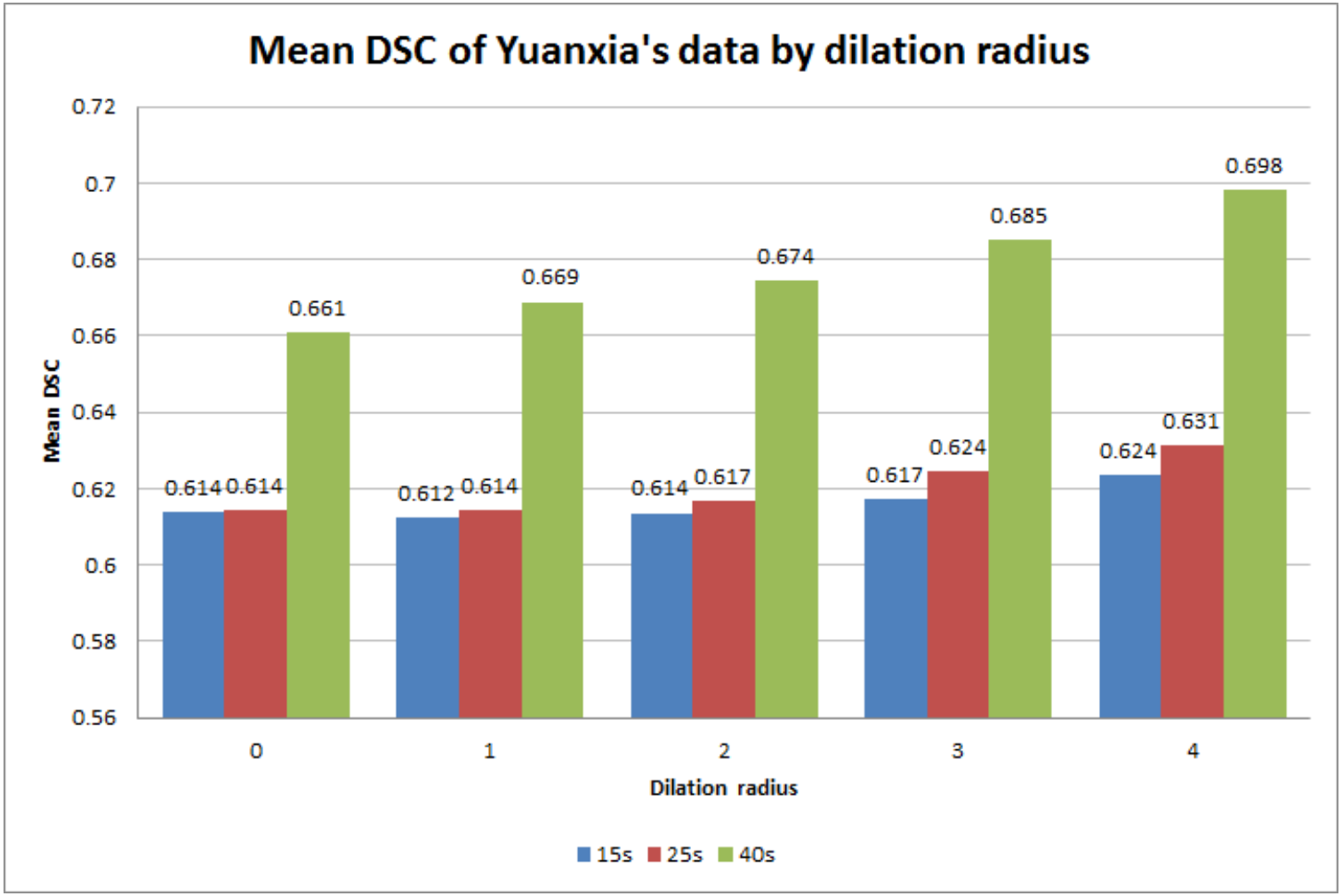
Accuracy results



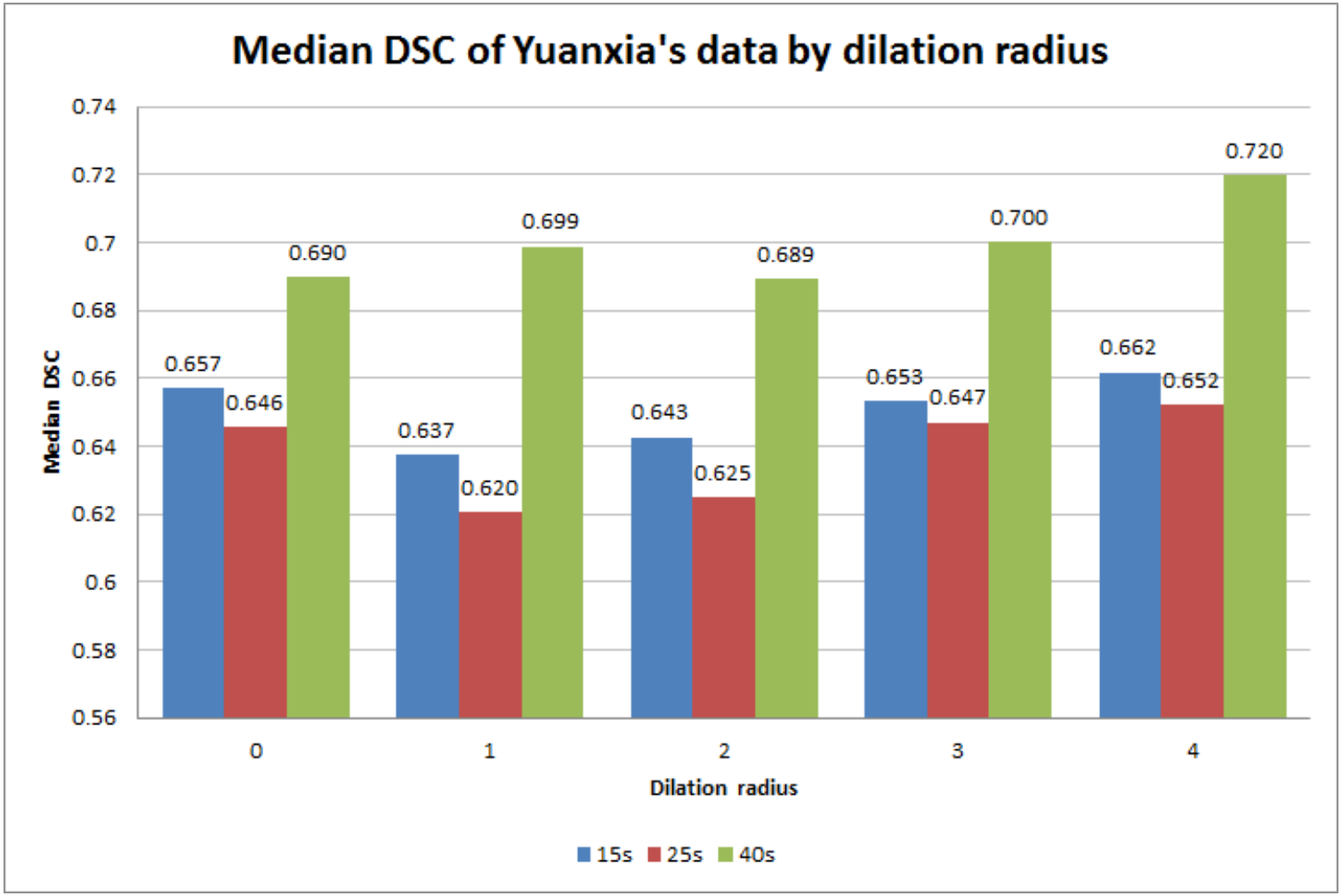
Accuracy results



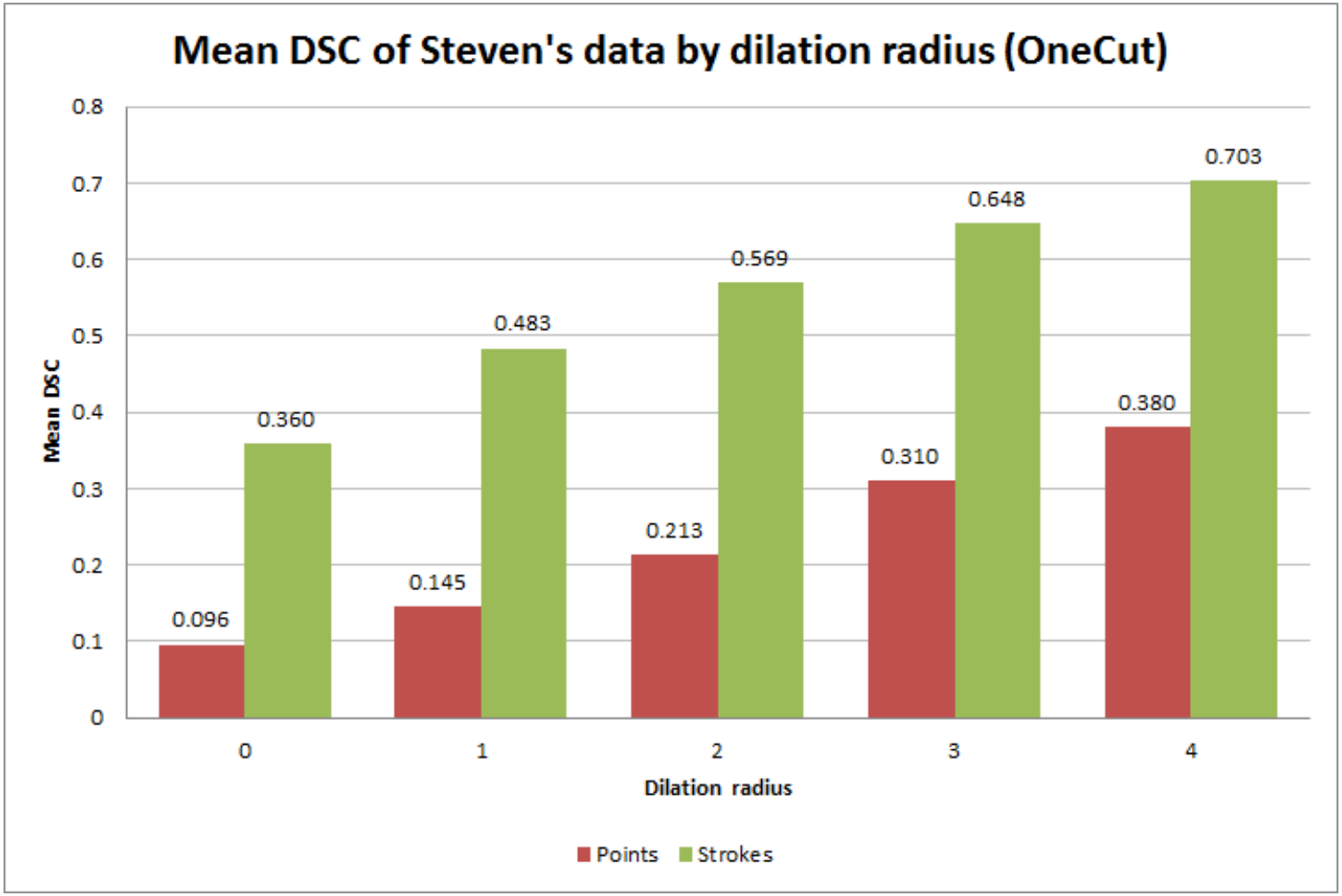
Accuracy results



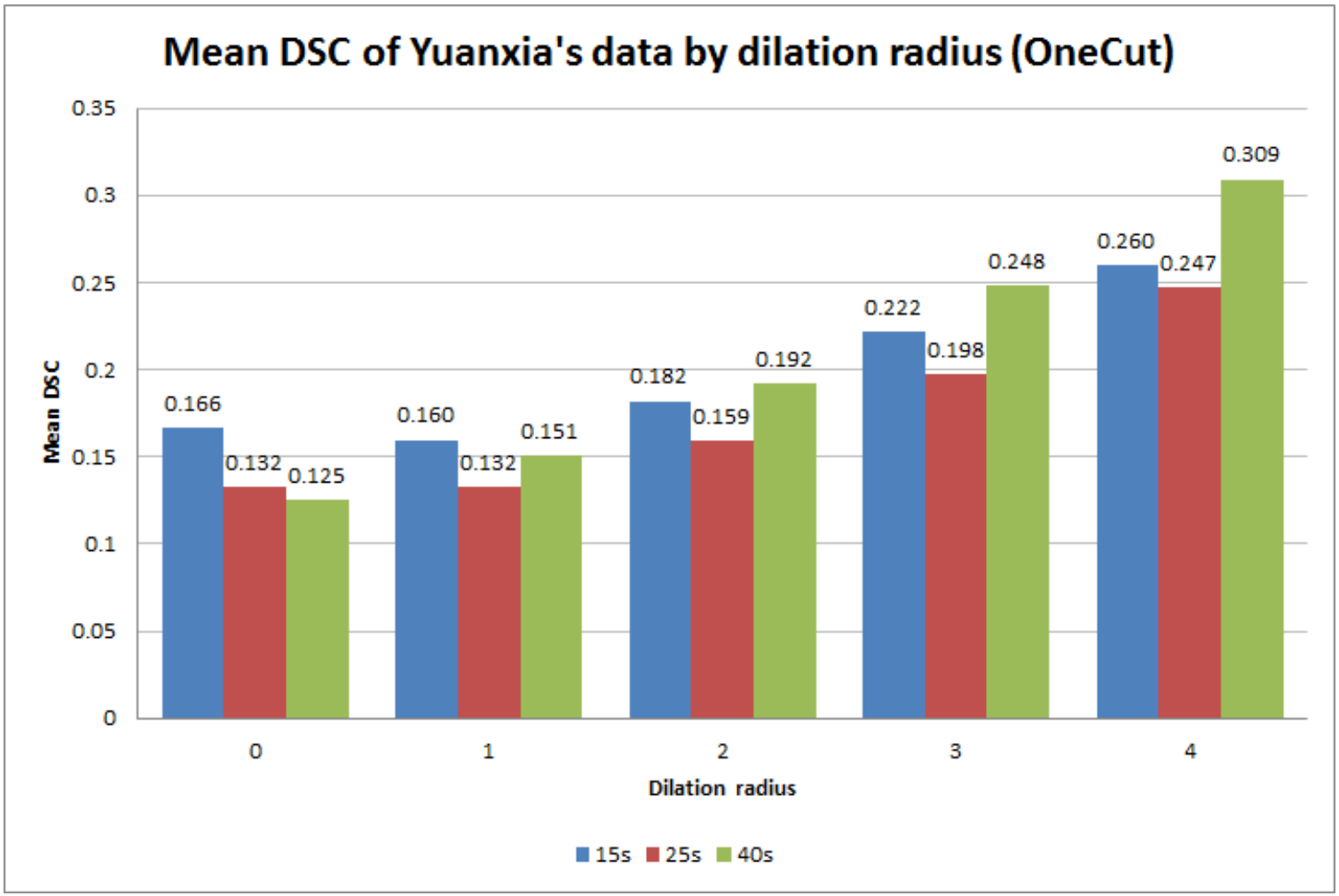
Accuracy results



Accuracy results



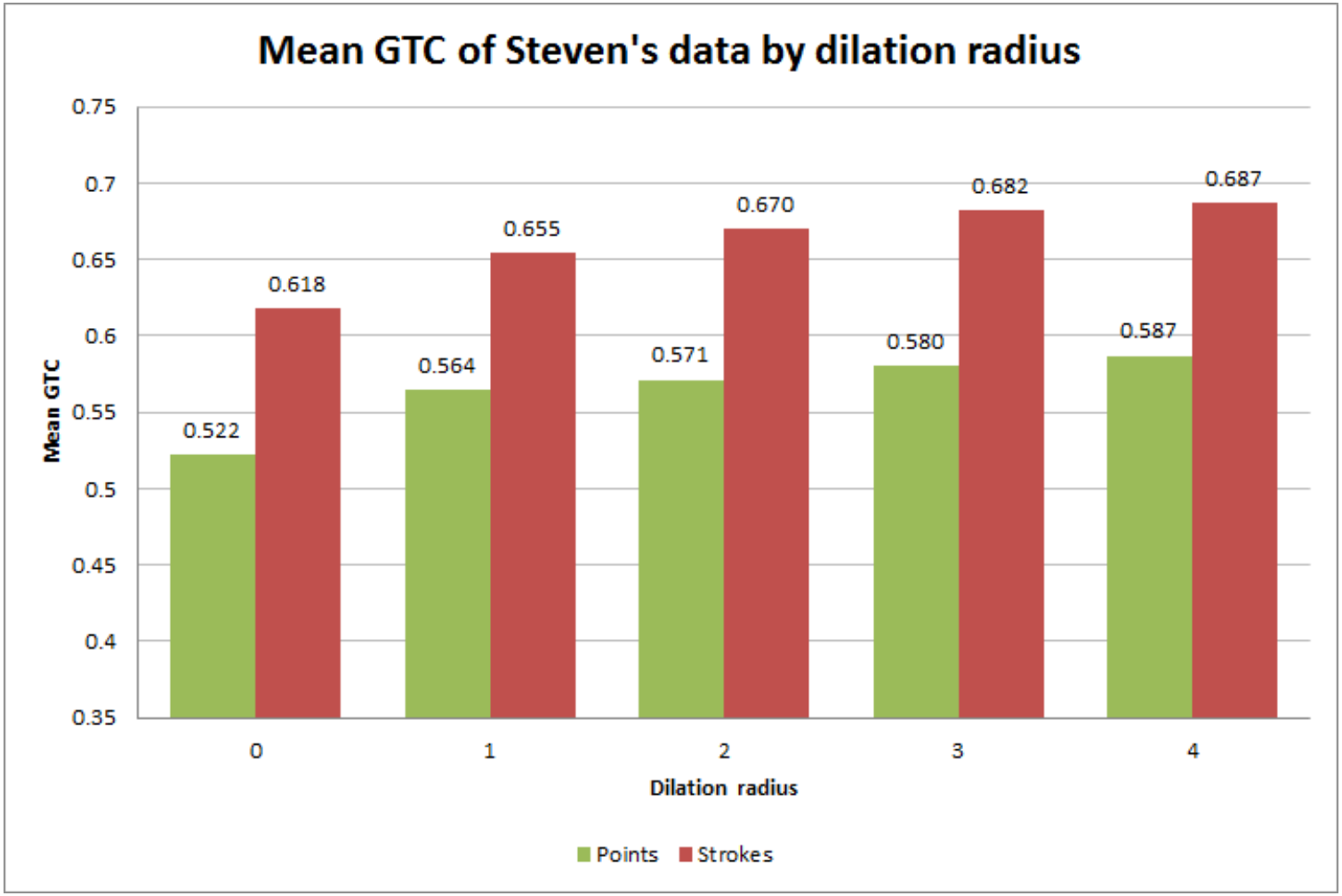
Accuracy results



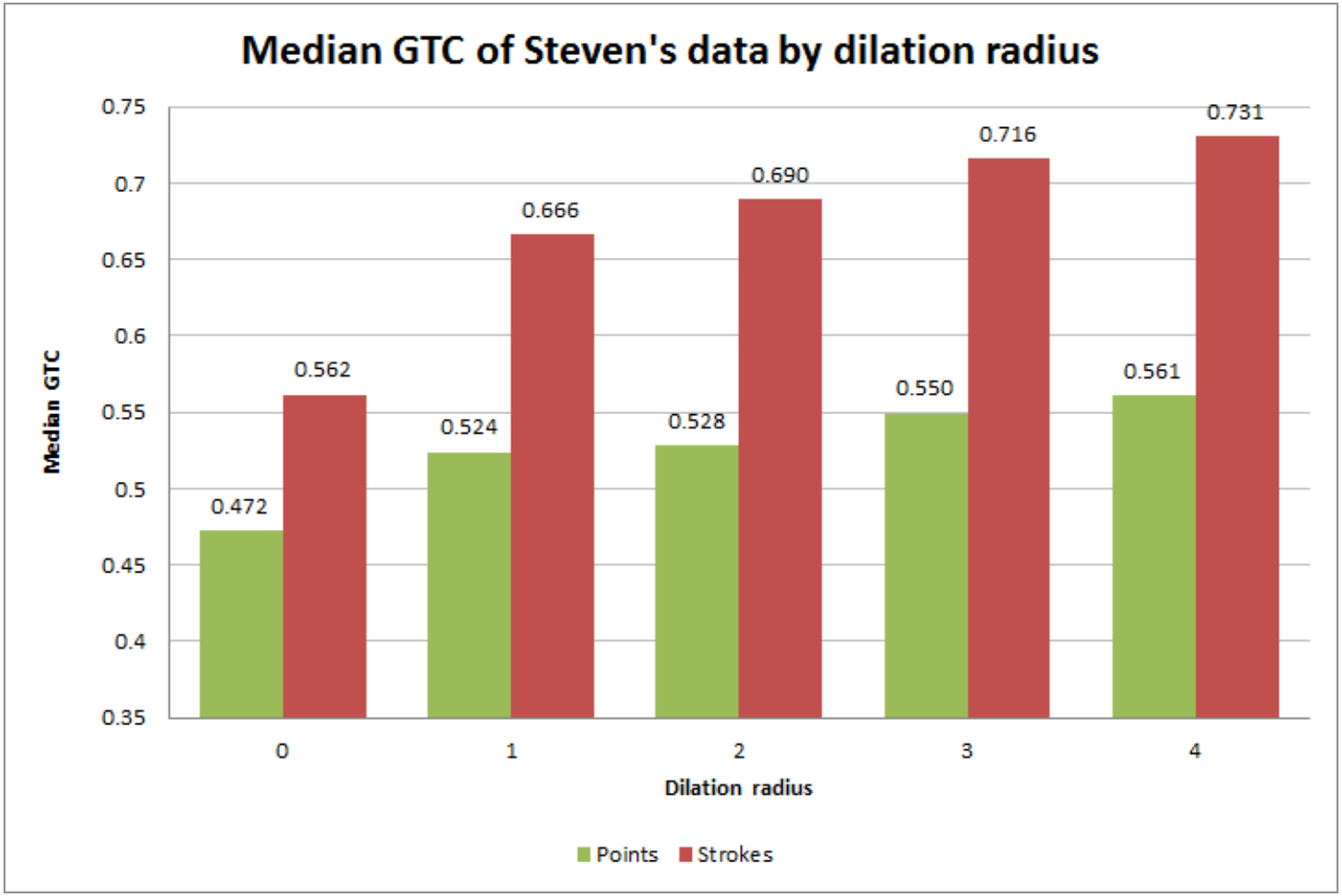
Analyzing reproducibility

- Reproducibility is a measure of consistency in the segmentation.
- We want to know how consistently the algorithm handles different users segmenting the same image.
- Measured with the Generalized Tanimoto Coefficient.
- $\text{GTC} = \frac{\Sigma(X_i \wedge Y_i)}{\Sigma(X_i \vee Y_i)}$
- The value is between 0 and 1, where 0 = no consistency between samples and 1 = all samples the same.

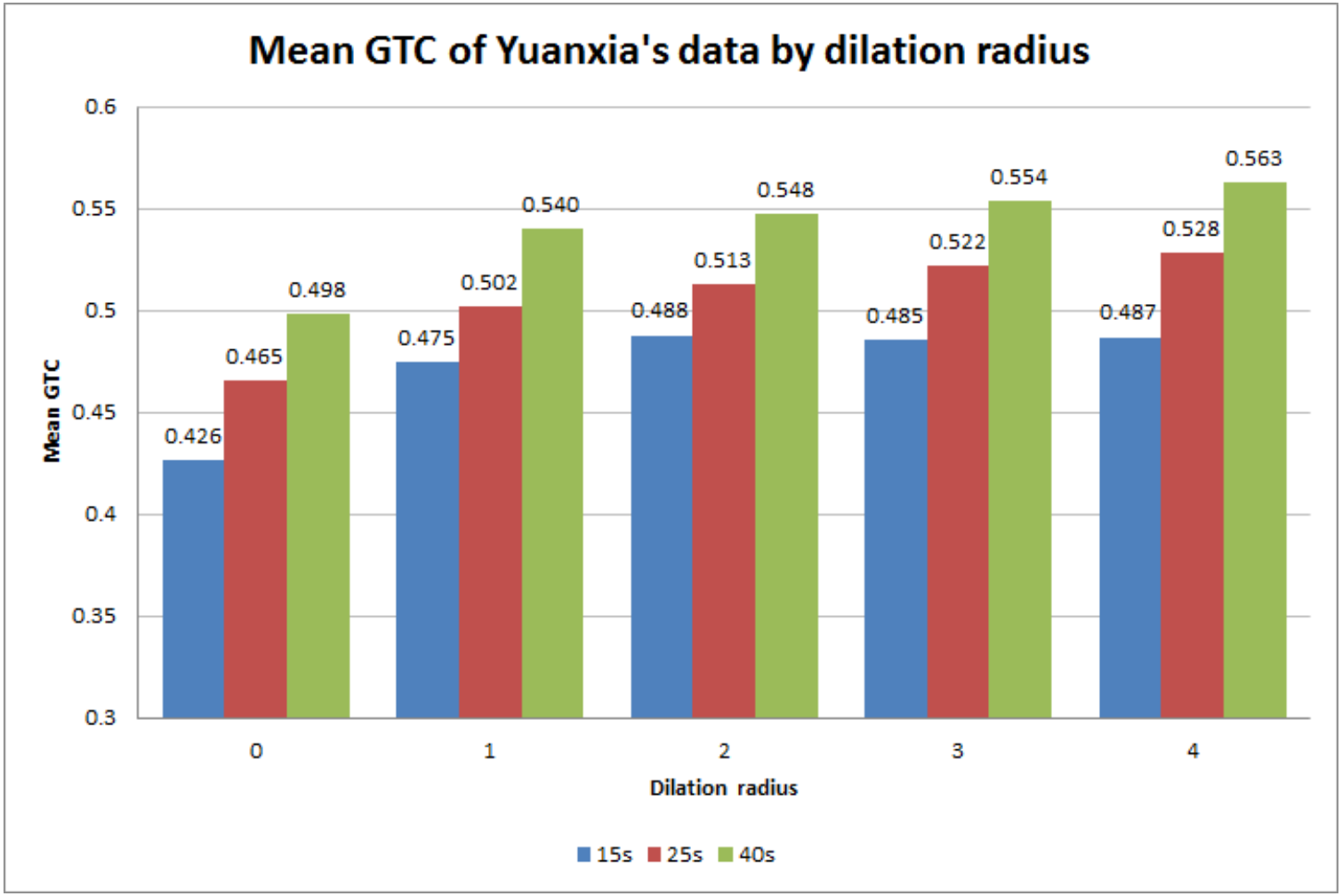
Reproducibility results



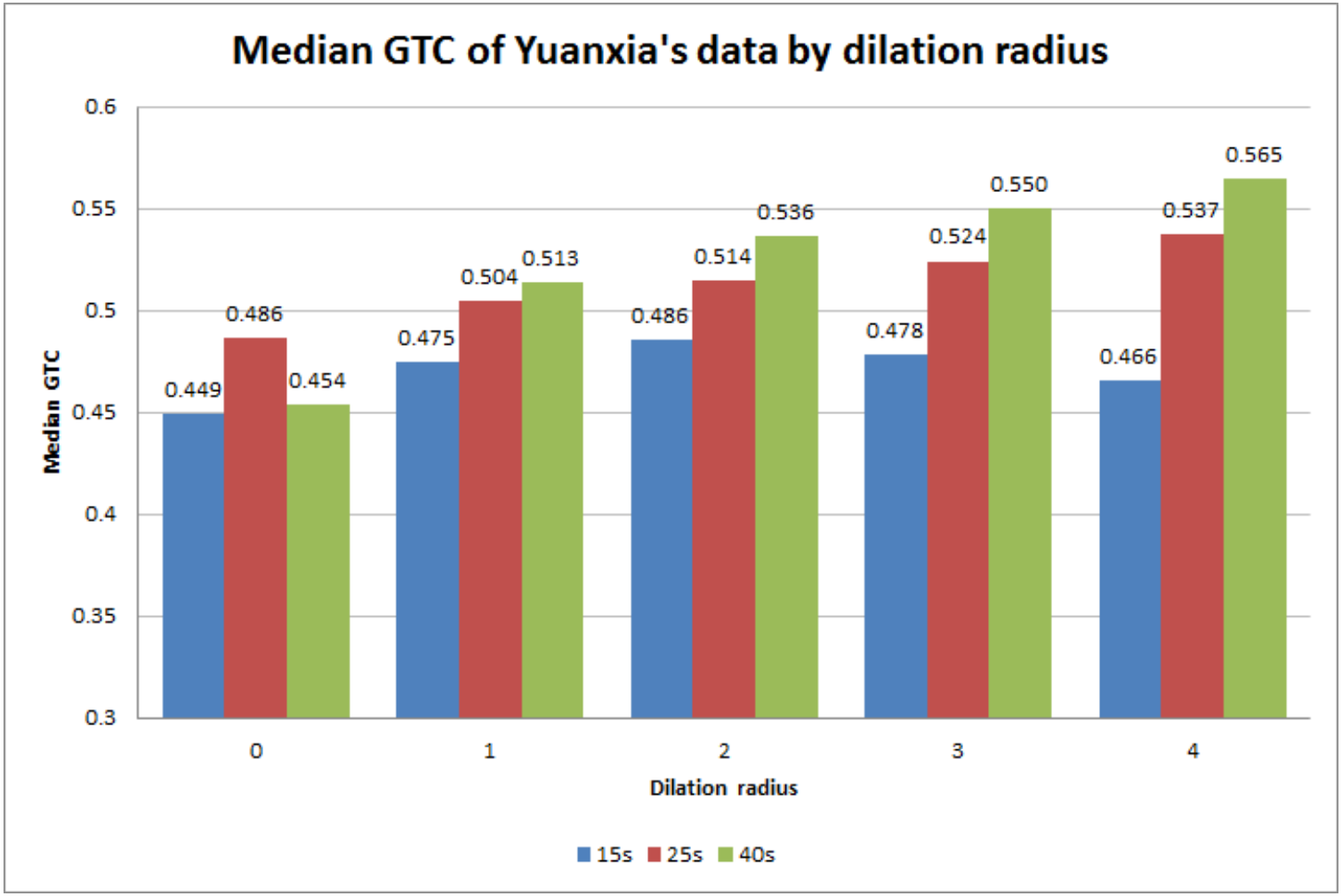
Reproducibility results



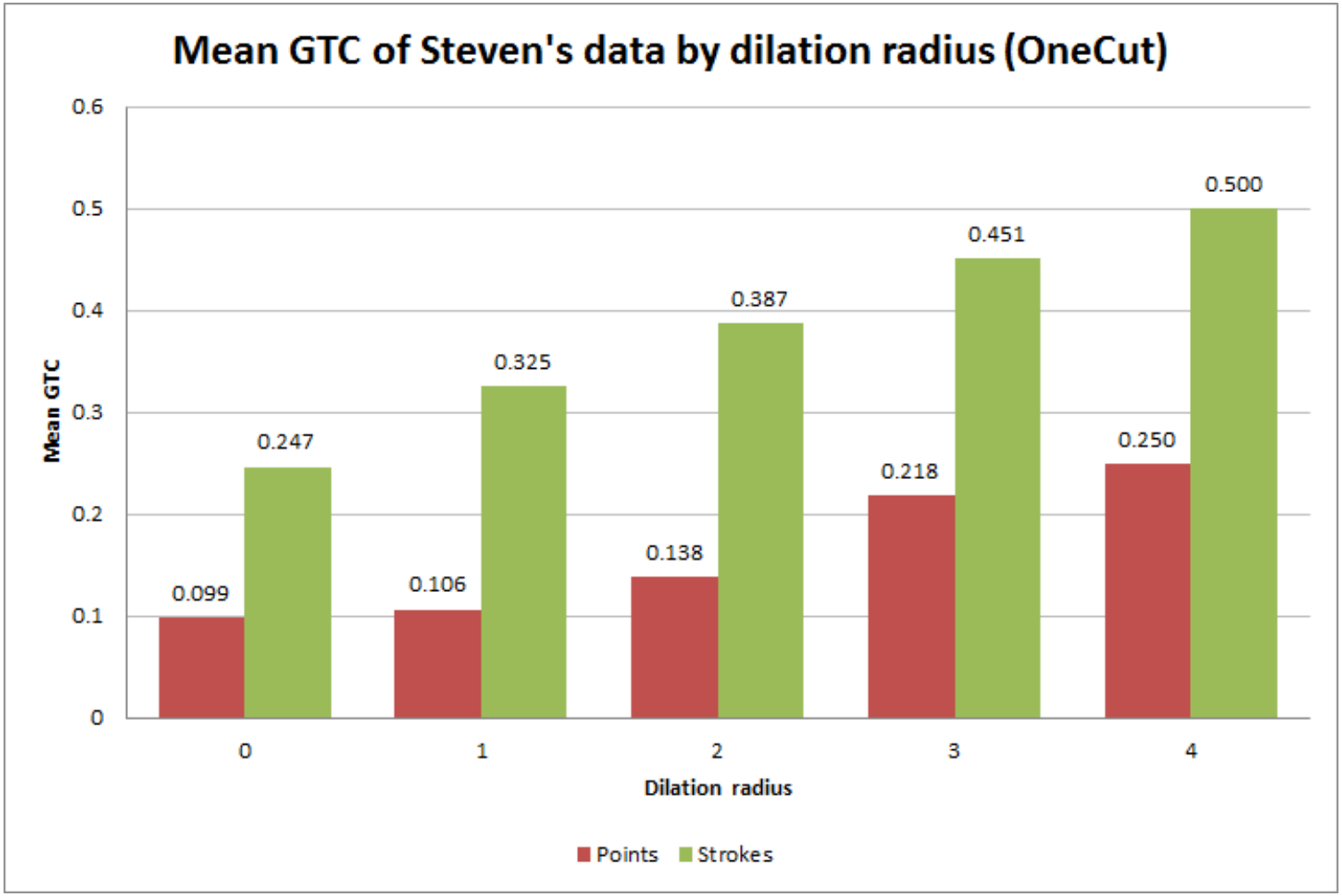
Reproducibility results



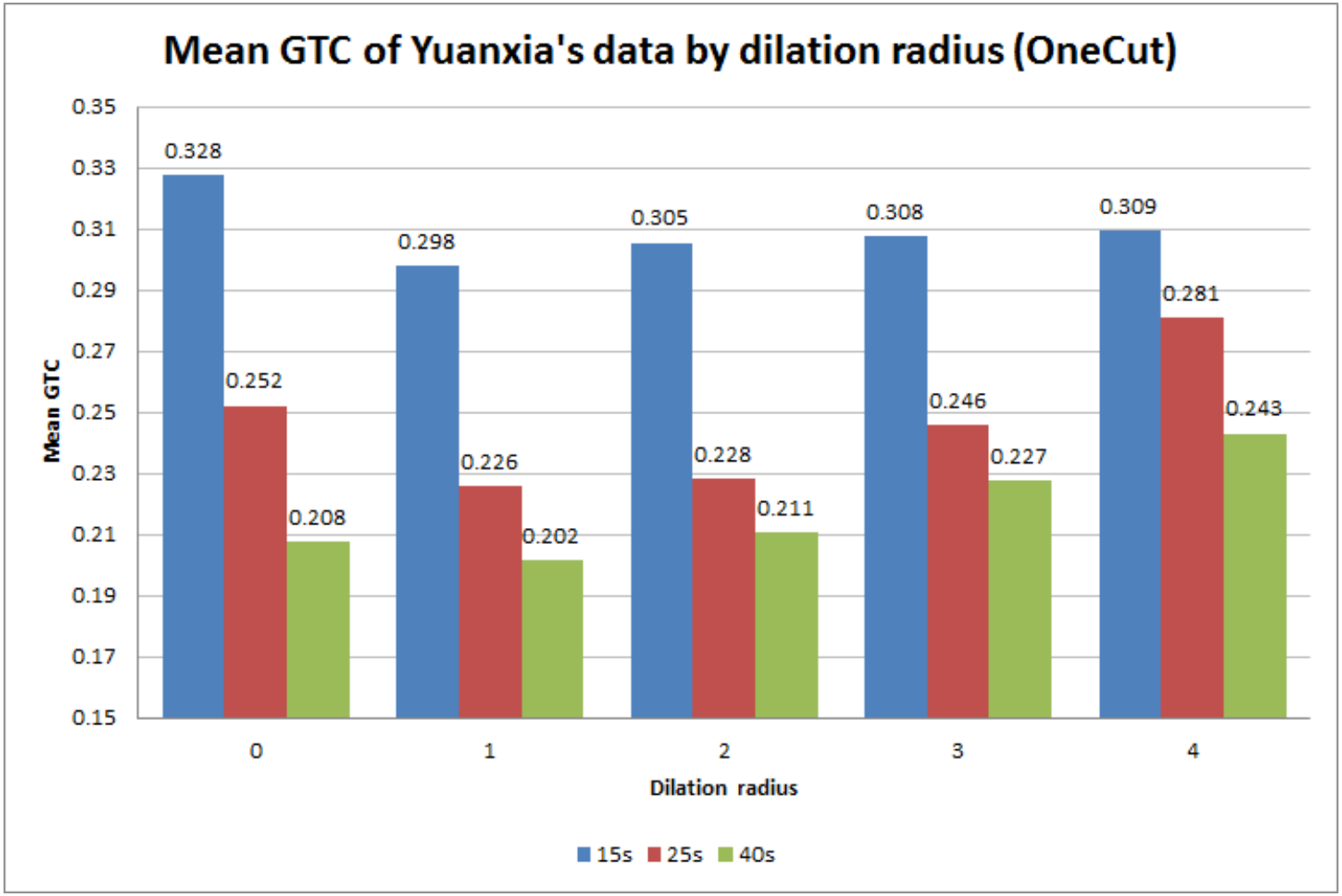
Reproducibility results



Reproducibility results



Reproducibility results



Testing for statistical significance

- In order to know that our results were statistically significant, we used the Friedman square chi test.
- This could only be used within groups of independent variables (eg, no comparing one study's data to another's).
- Was chosen because our data was not normally distributed (determined with D'Agostino's K^2 test)
- Differences statistically significant ($p < 0.01$) for all except Yuanxia's 40s time pressure group for the GTC under OneCut.

Conclusion

- Hence, we conclude **stroke width has a significant impact on our semi-automatic image segmentation programs**.
- This can't be generalized for all algorithms. But it seems likely that it may hold for many, or at least not negatively affect accuracy and reproducibility.
- The OneCut algorithm's results should be taken with a grain of salt as it often failed to segment in any useful way. However, there was certainly *some* cases where dilation allowed it to achieve a useful segmentation.

The end

Project code and data available at:
`https://github.com/KatrinaHoffert/stroke-radius-segmentation`

Questions?