



# Sourcerer

## A Robust, Multi-Scale Source Extraction Tool Suitable for Faint and Diffuse Objects

**Michael Wilkinson<sup>1</sup>**, Caroline Haigh<sup>1</sup>, Simon Gazagnes<sup>1</sup>,  
Paul Teeninga<sup>1</sup>, Nushkia Chamba<sup>2</sup>, Thanh Xuan Nguyen<sup>3</sup>,  
Laurent Najman<sup>3</sup>, Benjamin Perret<sup>3</sup>, Giovanni Chierchia<sup>3</sup>,  
Aku Venhola<sup>4</sup>, Reynier Peletier<sup>1</sup>

<sup>1</sup>University of Groningen, Groningen, The Netherlands

<sup>2</sup>Instituto de Astrofísica de Canarias, Spain

<sup>3</sup>ESIEE Paris, Noisy-Le-Grand, France

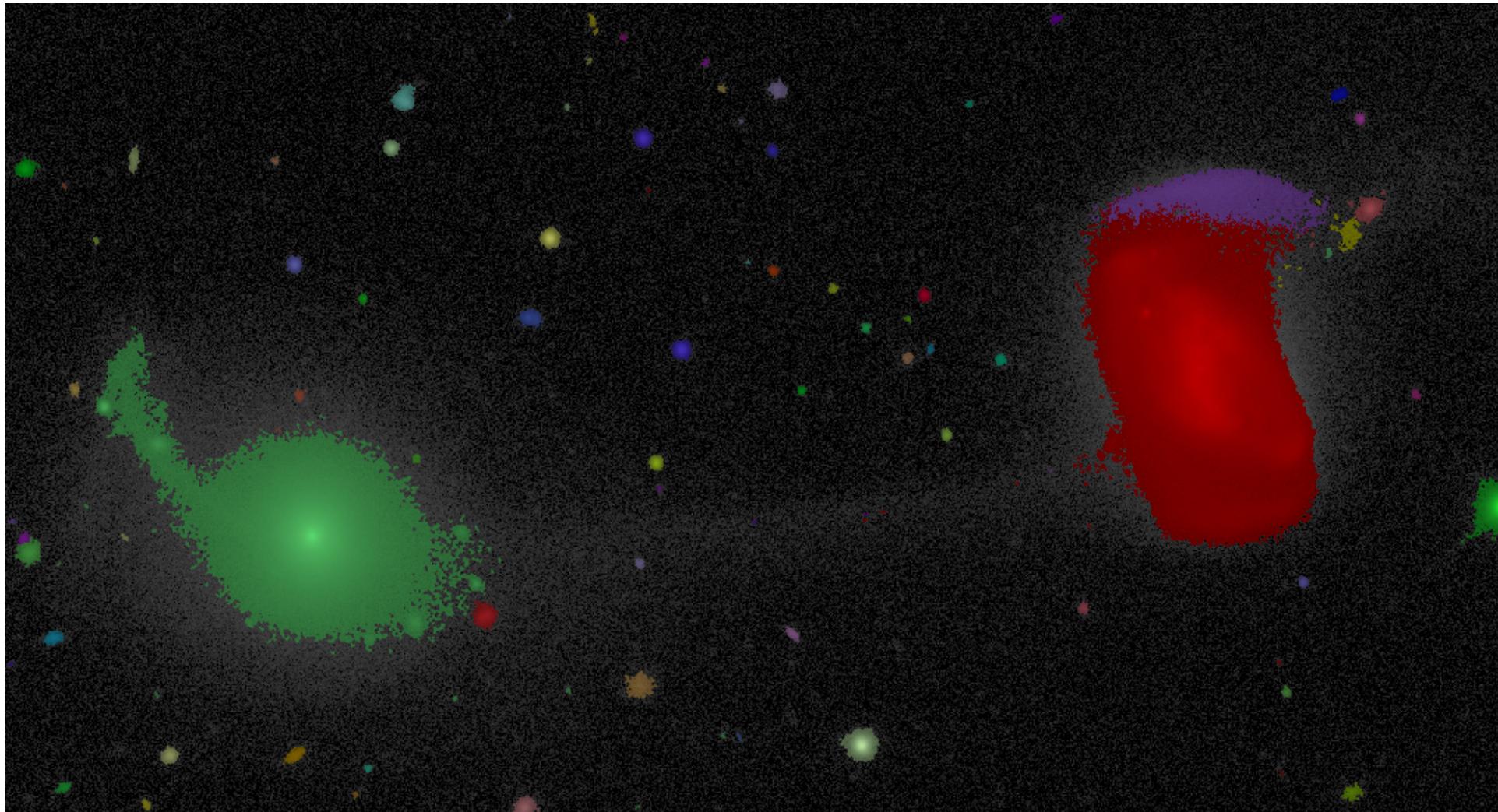
<sup>4</sup>University of Oulu, Oulu, Finland





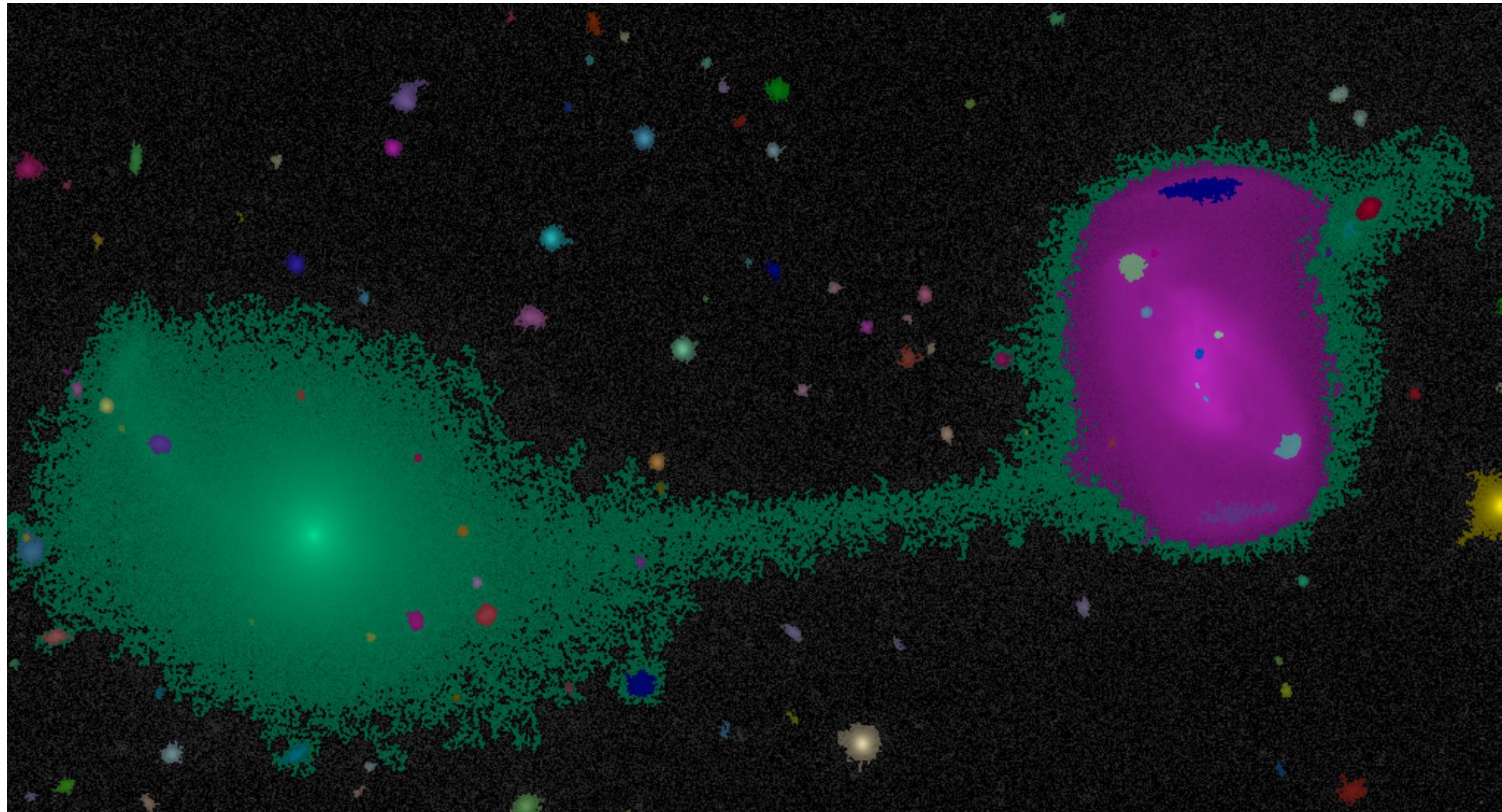
university of  
groningen

# Source Finding using Trees



SExtractor

[ Teeninga, Moschini, Trager, & Wilkinson. Mathematical Morphology, 2016 ]

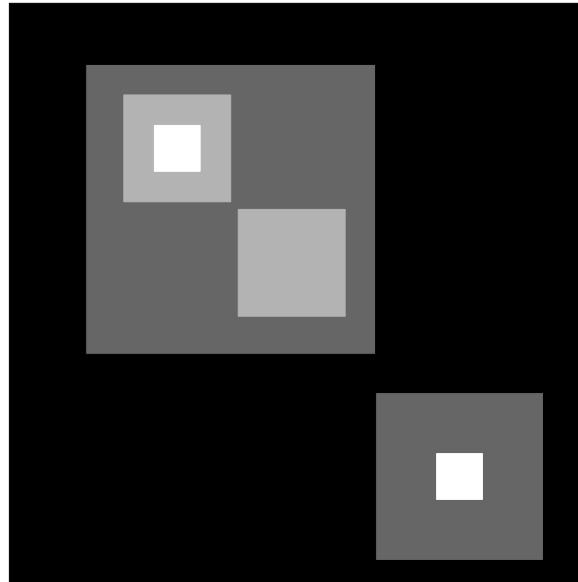


MTObjects

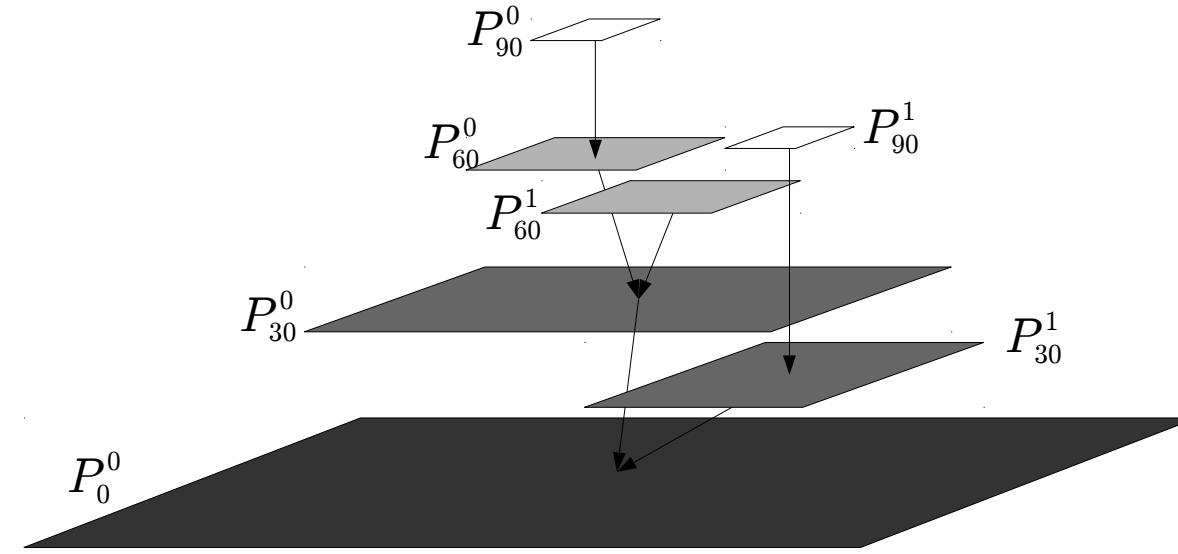
[ Teeninga, Moschini, Trager, & Wilkinson. Mathematical Morphology, 2016 ]



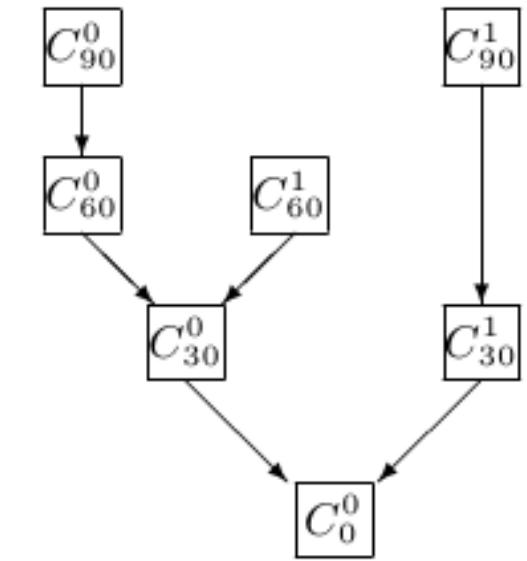
- Based on decomposition of image into connected components



Input image



Thresholded sets(peaks)



Max-Tree

- Storage cost  $O(N)$  worst case
- Time complexity  $O(N \log N)$  typically



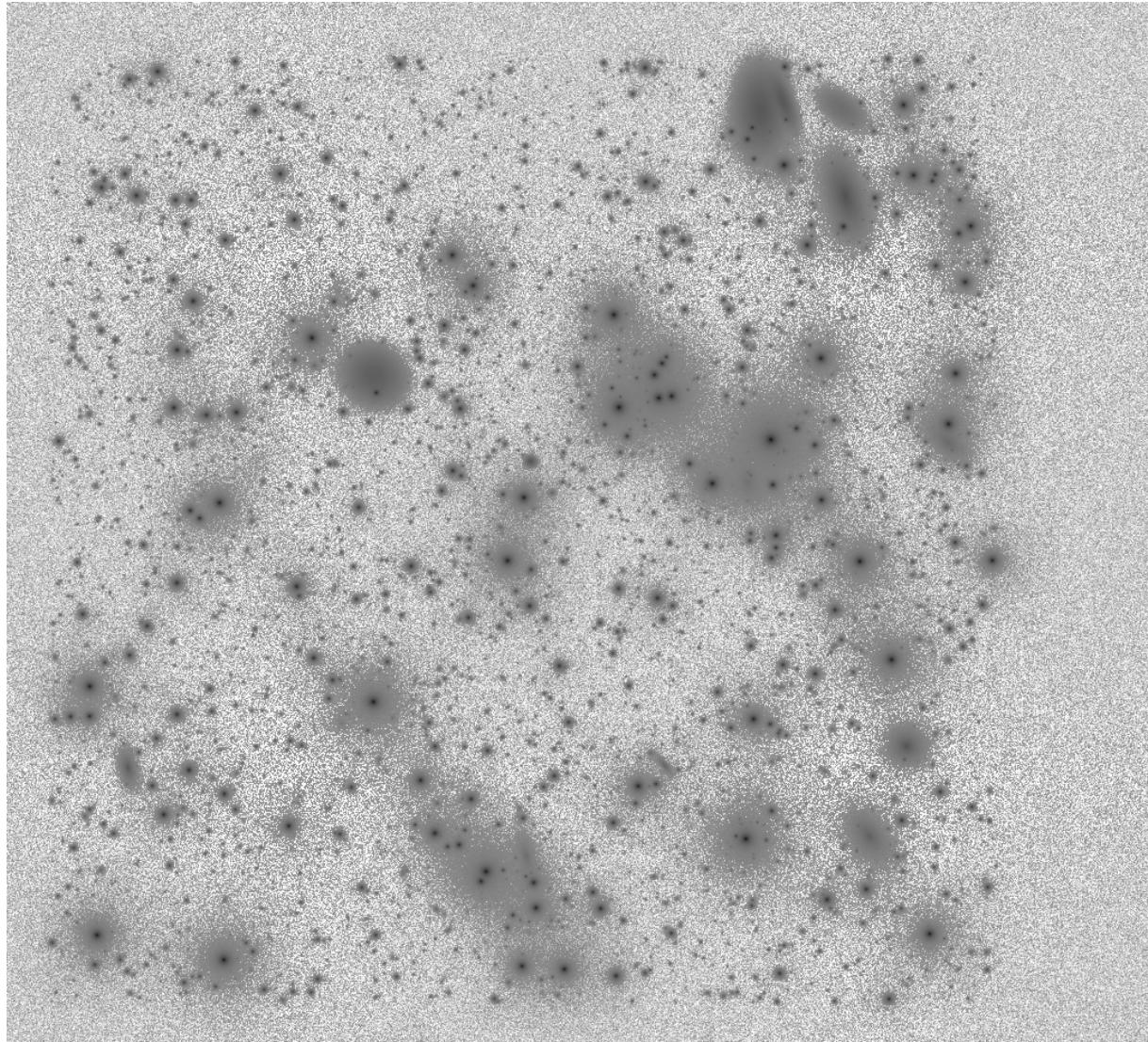
	SE	PF	NC	AD	MT	Sourcerer
core method	nested thresholds	watershed	watershed	"dendrogram"	max-tree	max-tree or other graphs
initial threshold	$\lambda\sigma$	$\lambda\sigma$	percentile	$\lambda\sigma$	0	0
nested objects	-	-	-	+	+	+
# thresholds	discrete	NA	NA	discrete/ $\infty$	$\infty$	$\infty$
detection by statistical test	-	-	-	-	+	+
parallel	+	?	+	-	-	+
# parameters	12	8	25	3	2	$\geq 2$



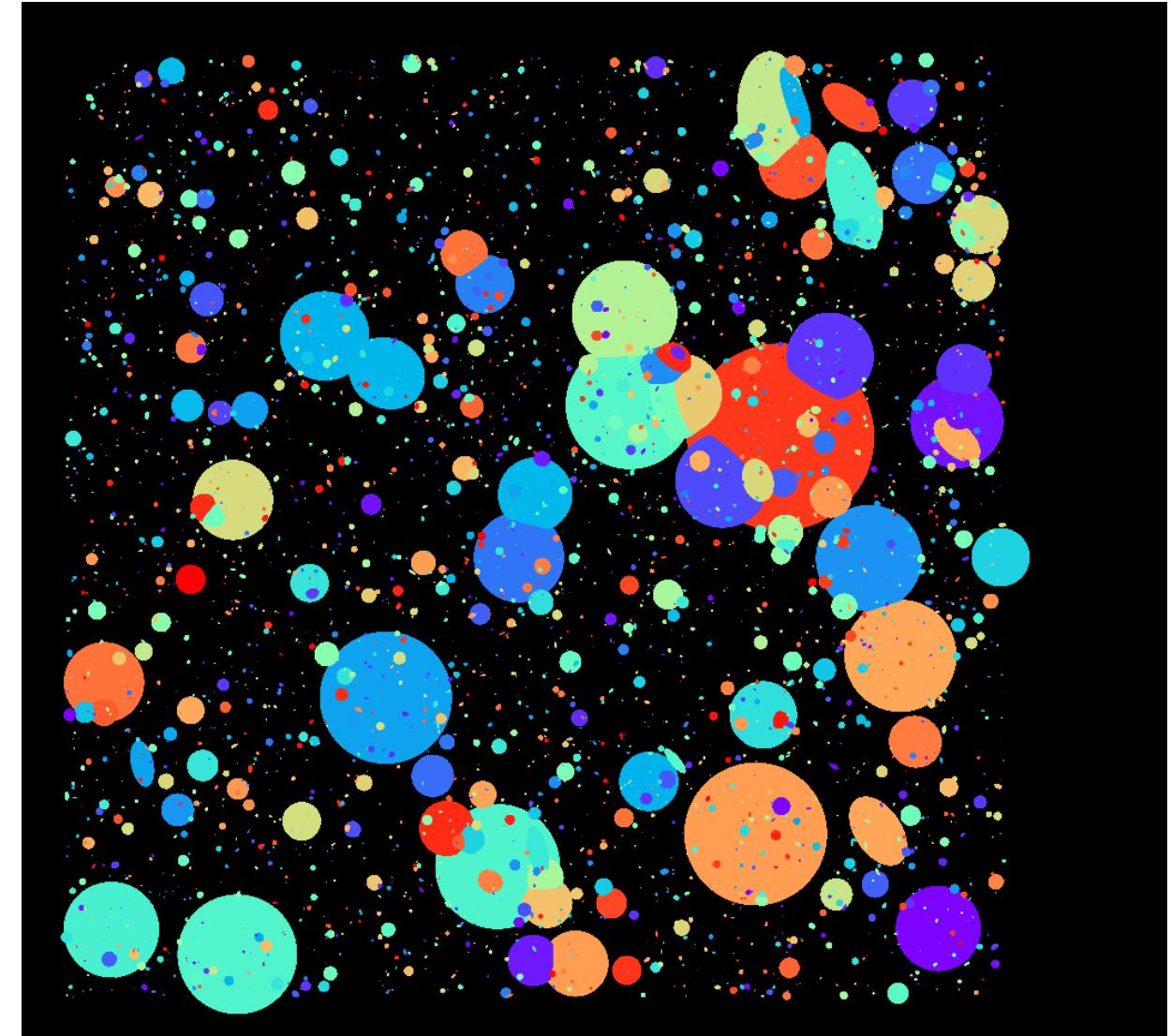
- Four source detection tools compared
  - SExtractor
  - Noisechisel + Segment
  - Profound
  - MTObjects
- Simulated FDS data
- Automatic parameter optimization
- Four different quality measures



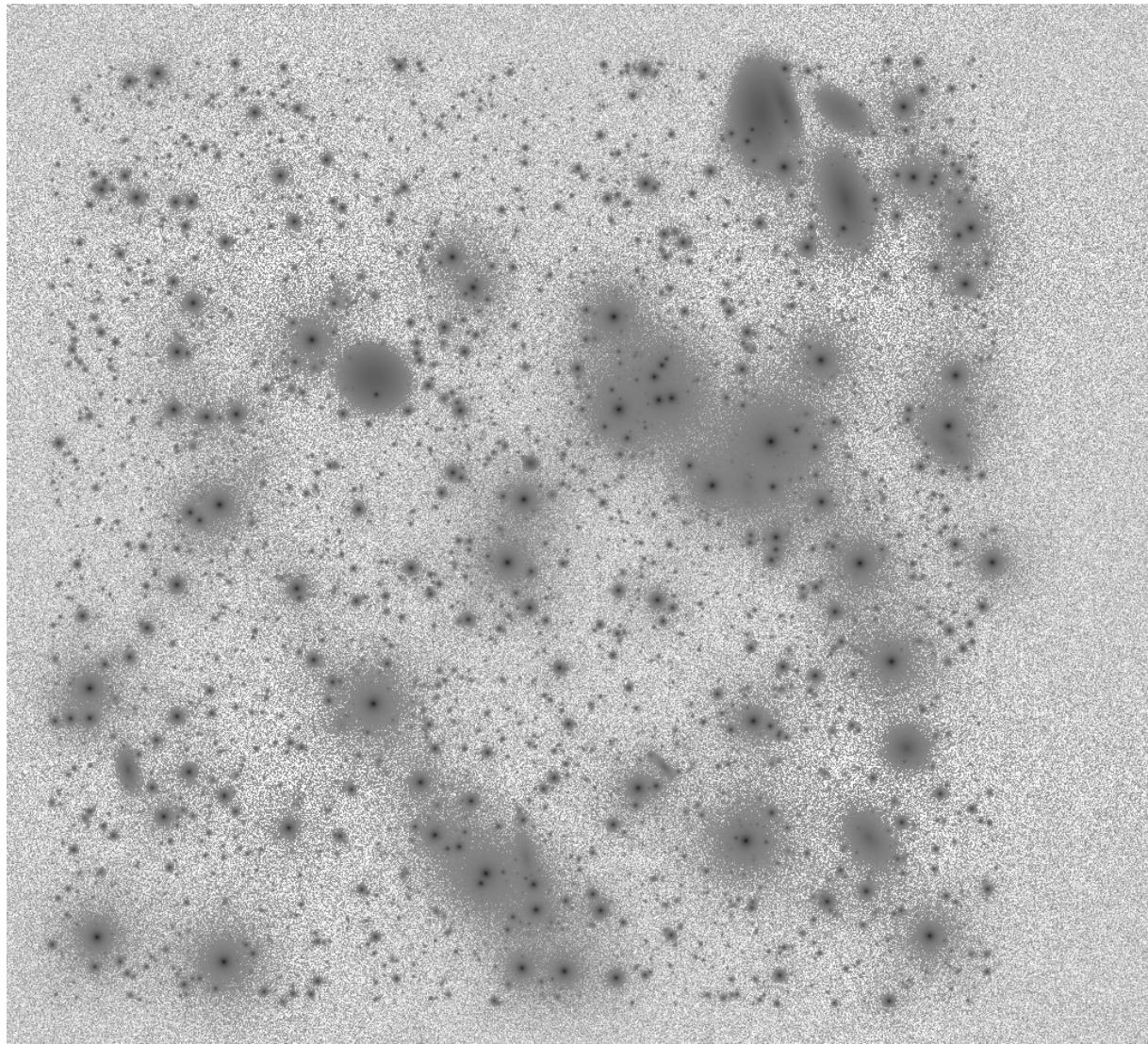
[Haigh et al. in preparation]



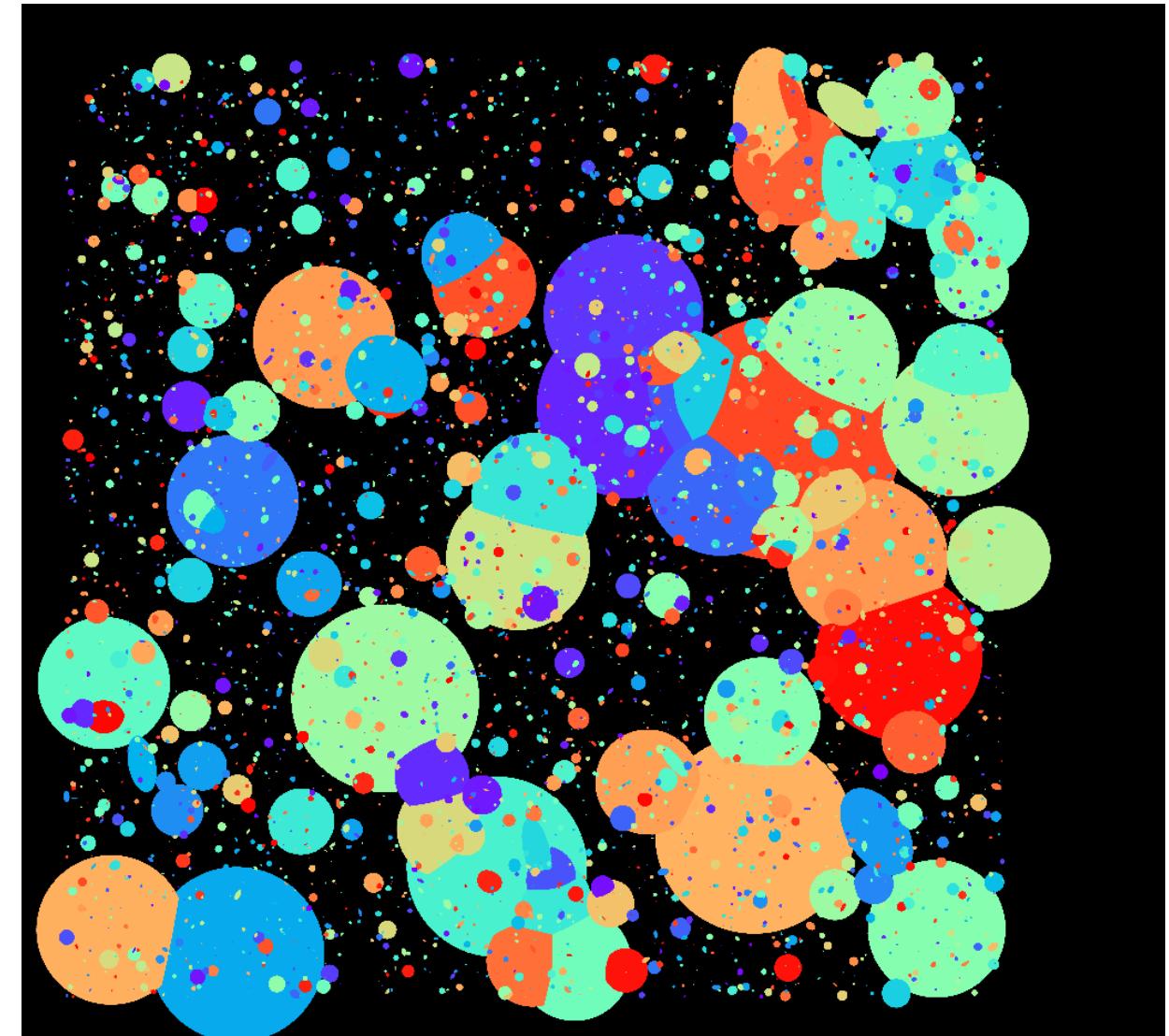
Simulated Image



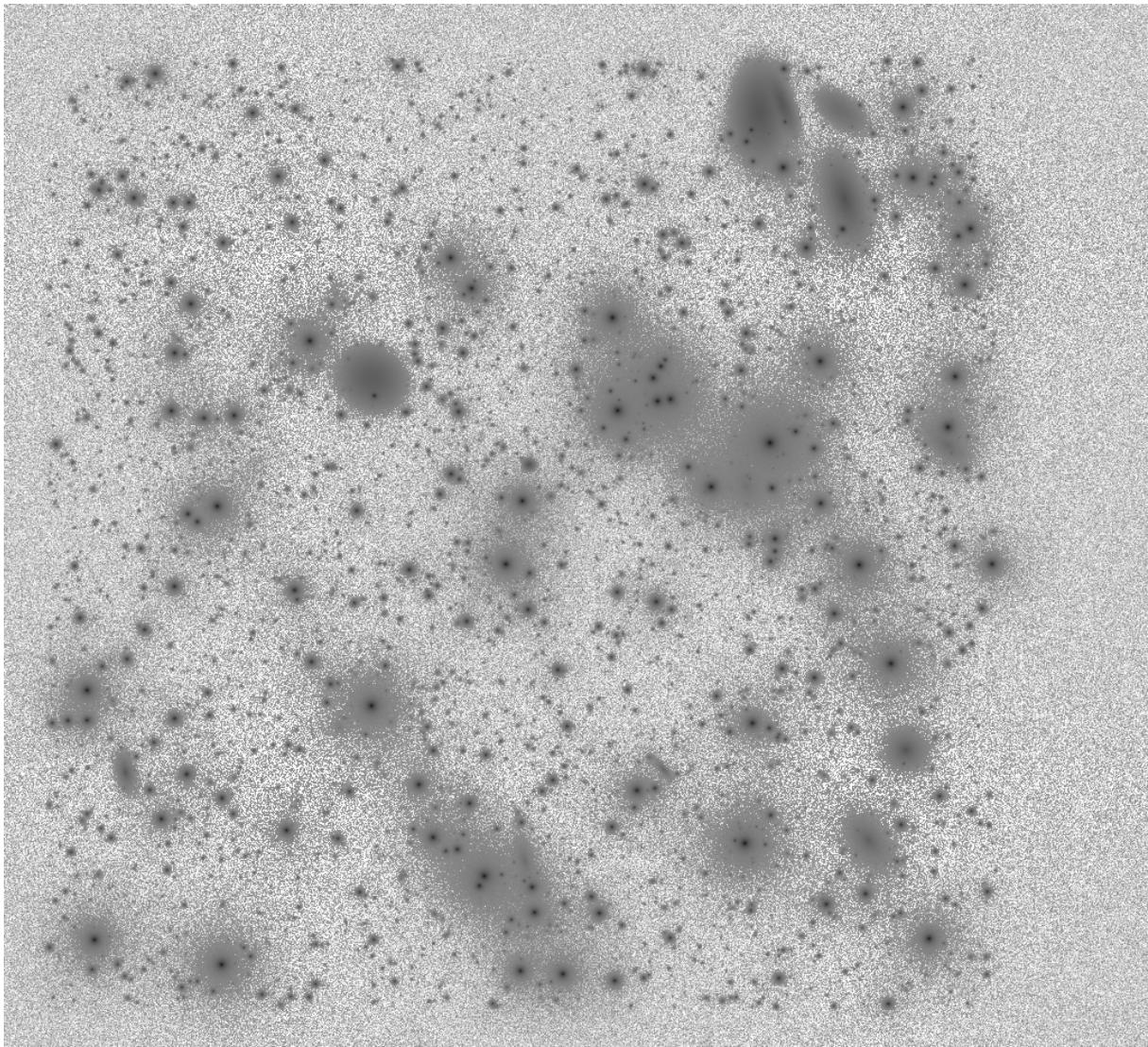
Ground truth at  $1.0\sigma$



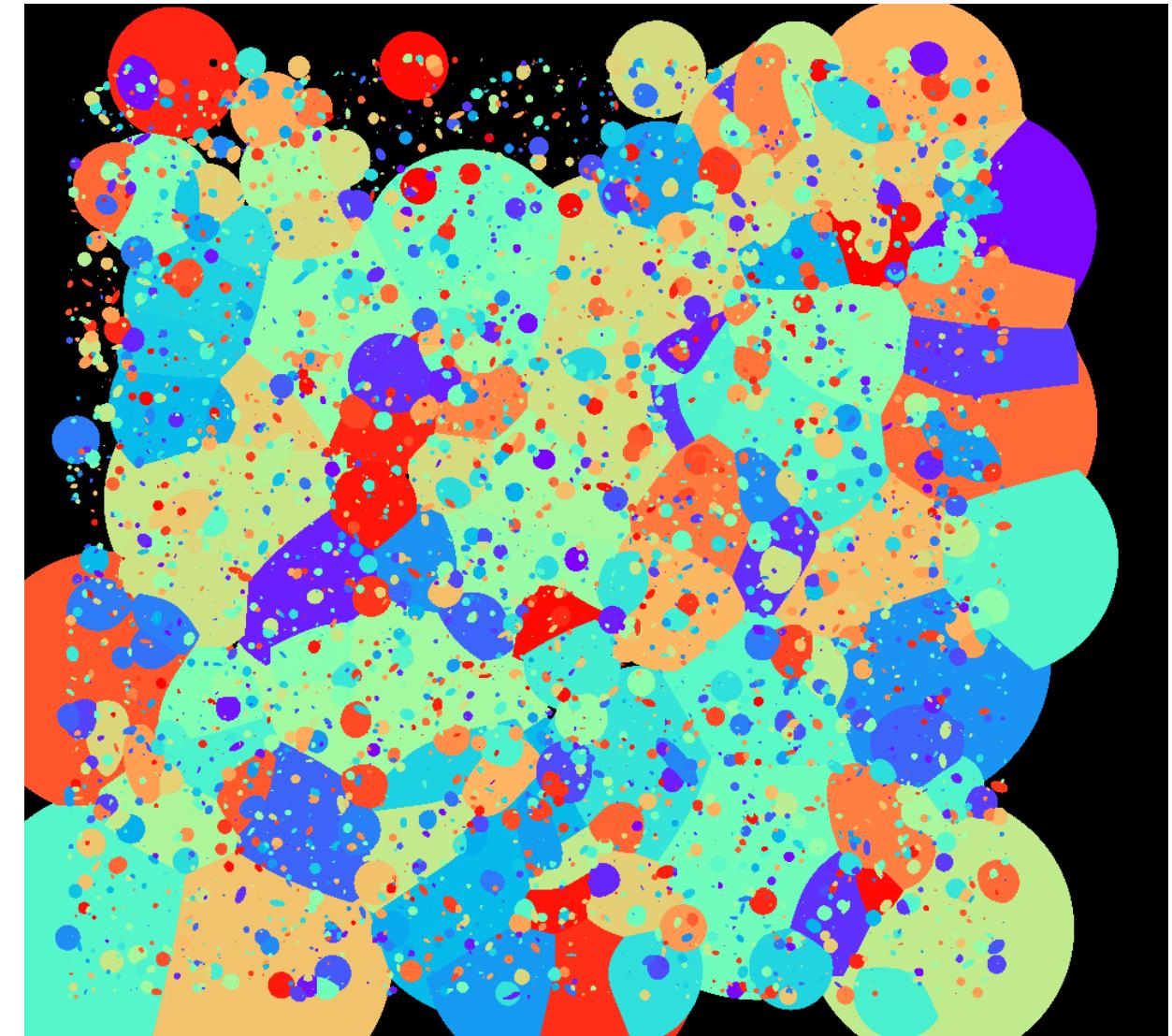
Simulated Image



Ground truth at  $0.5\sigma$



Simulated Image



Ground truth at  $0.1\sigma$



- $F_1$  score: Combines precision (purity) and recall (completeness) in pure detection task

$$F_1 = \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

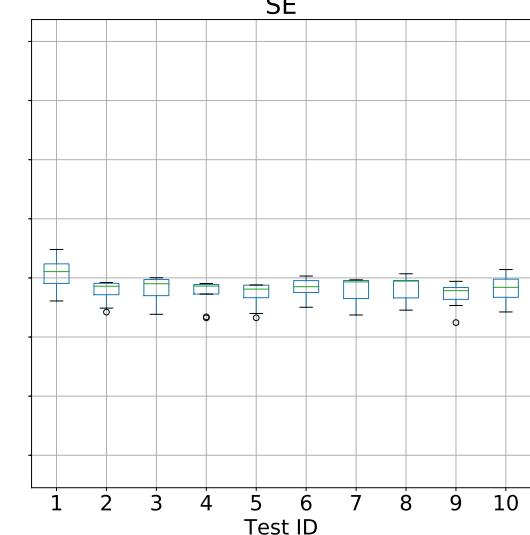
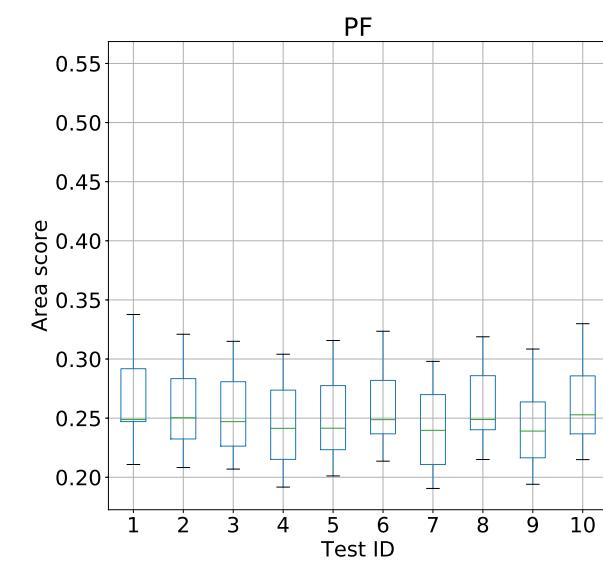
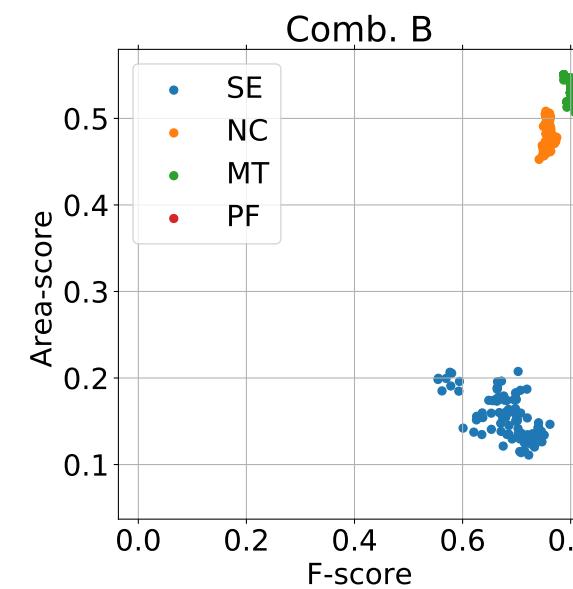
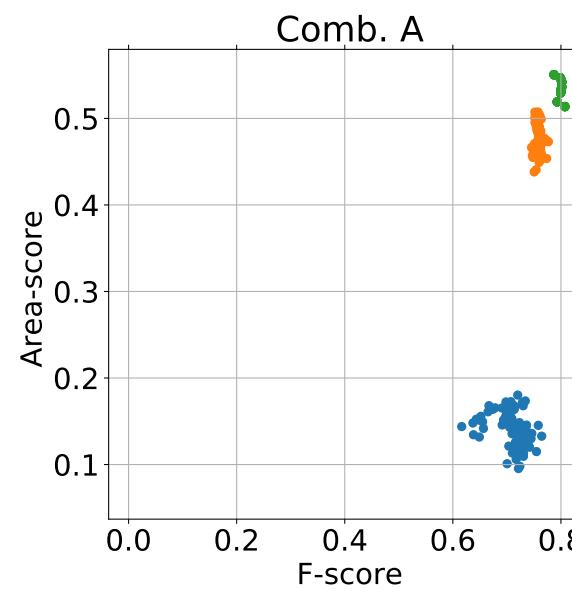
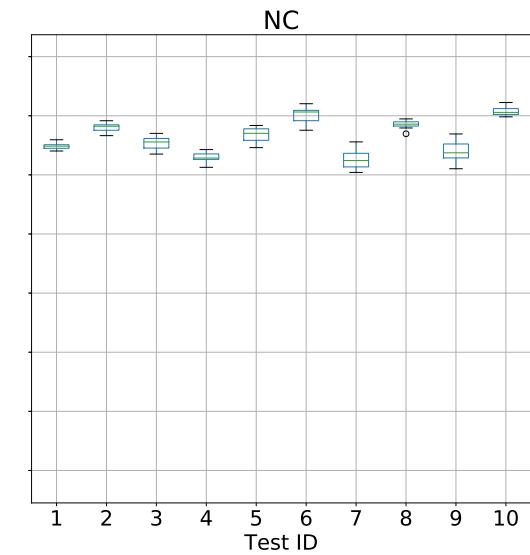
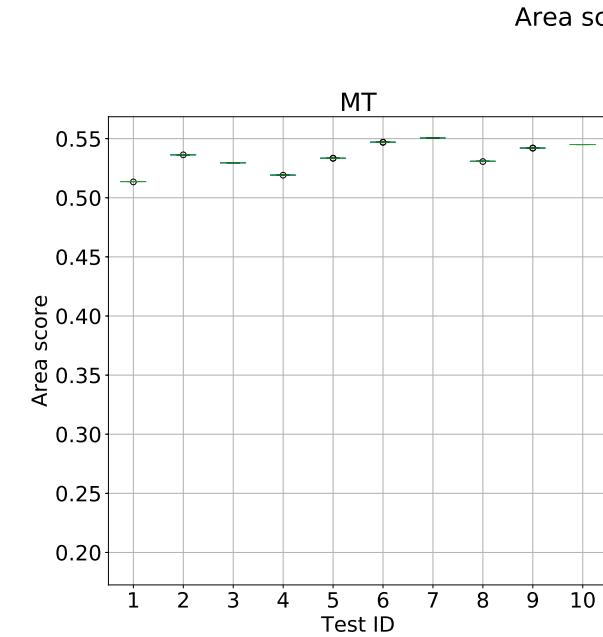
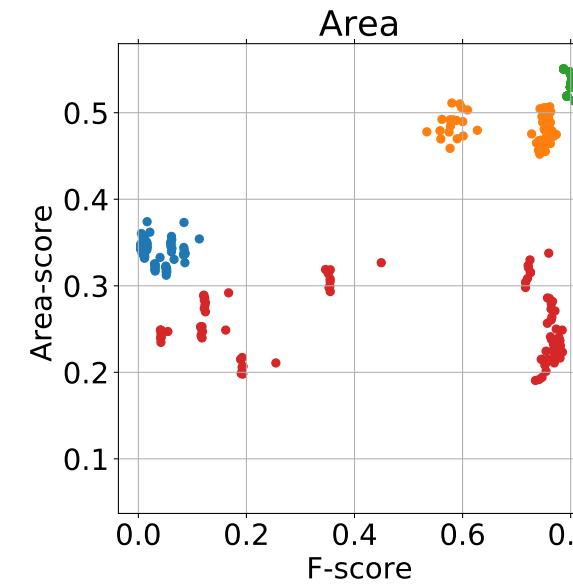
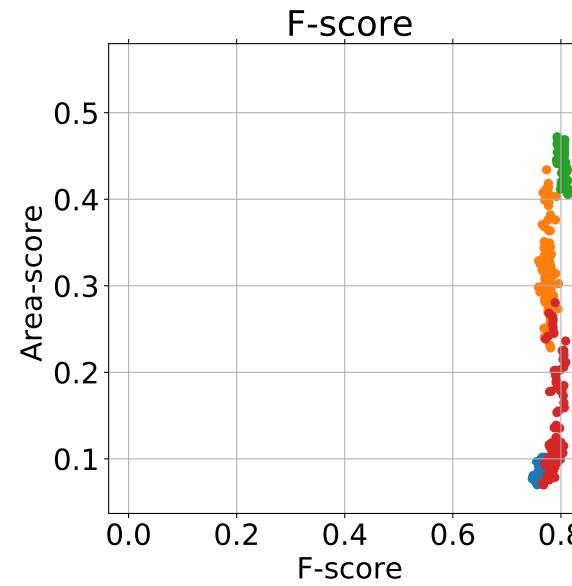
- Area score: Optimizes segmentation quality, combining undermerging error ( $UM$ ) and overmerging error ( $OM$ )

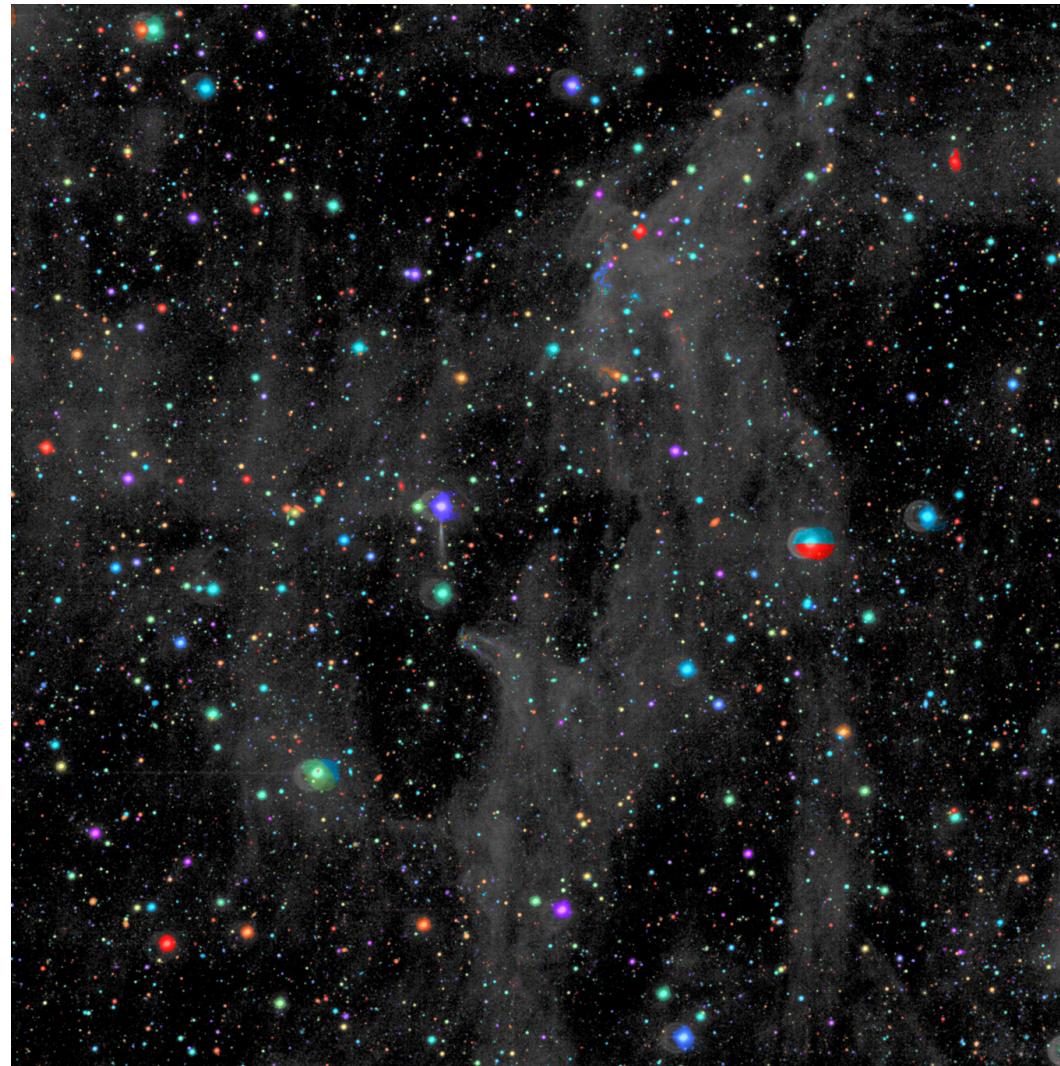
$$\text{Area-score} = 1 - \sqrt{OM^2 + UM^2}$$

- Combined score A:  $\sqrt{\text{Area-score}^2 + F_1^2}$
- Combined score B:  $\sqrt[3]{(1 - OM)(1 - UM)F_1}$



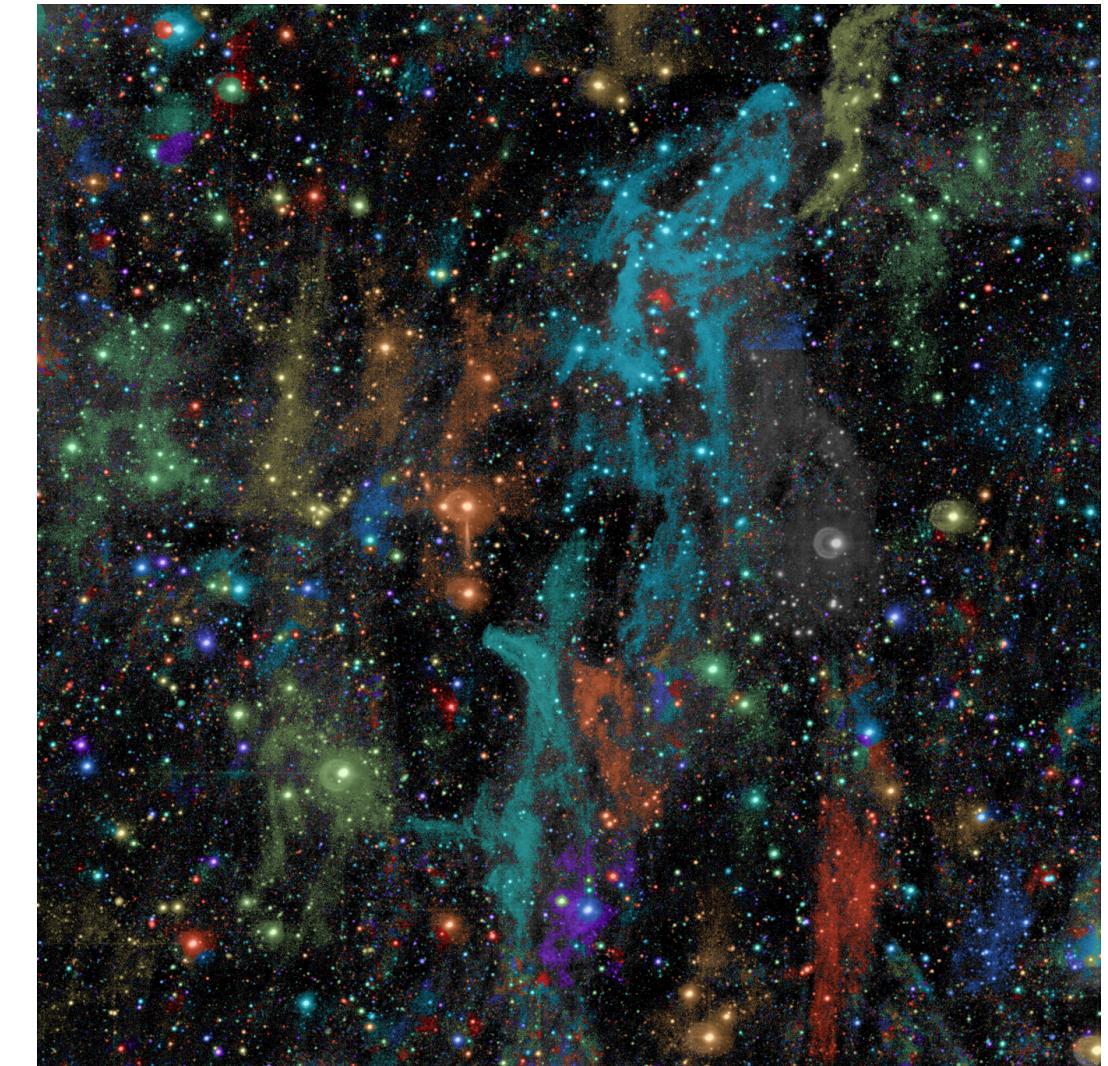
- Ten simulated images are used
- Bayesian optimisation is performed on each image for each quality measure
- Each of the settings is tested on the remaining 9 images
- Method implemented using GPyOpt optimisation library



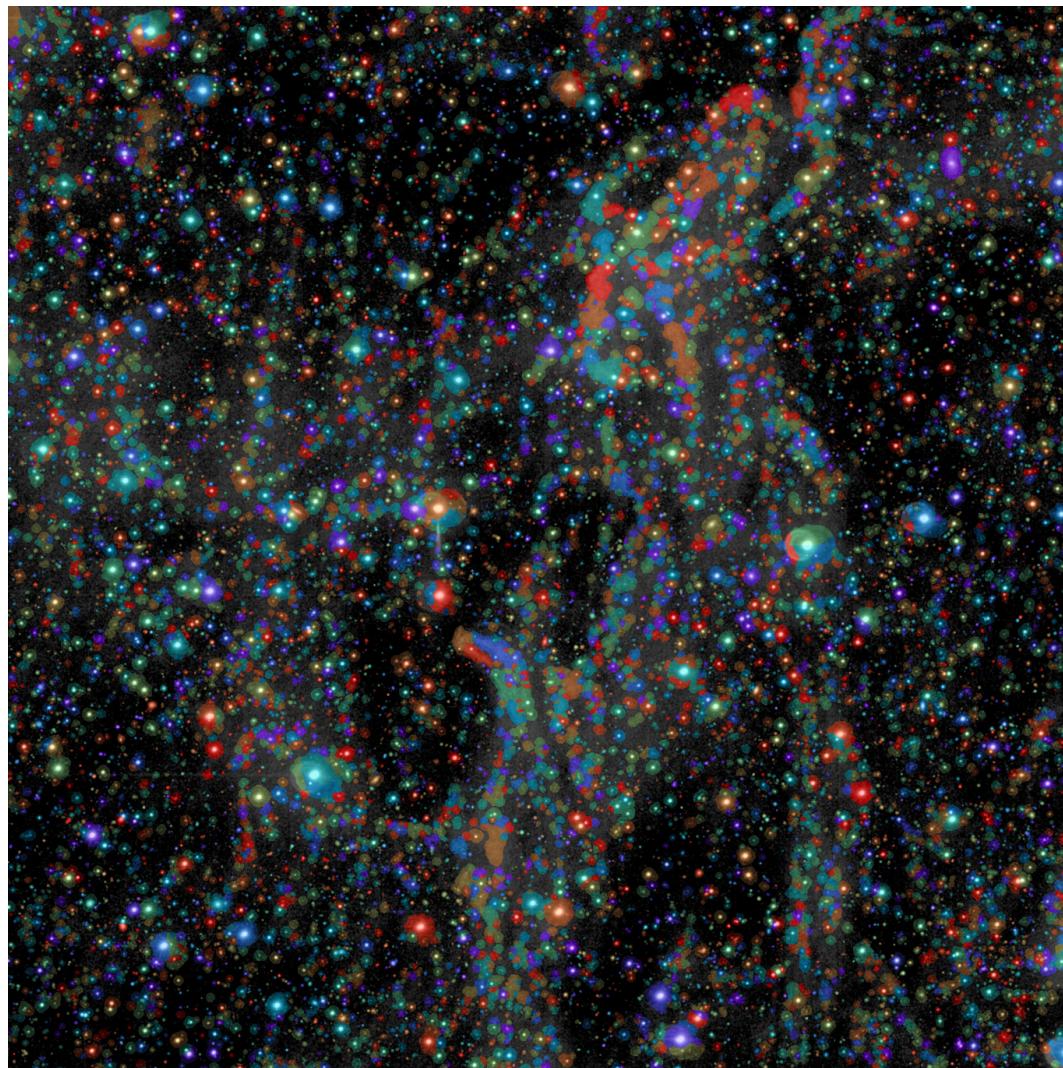


Optimized for F1

SExtractor

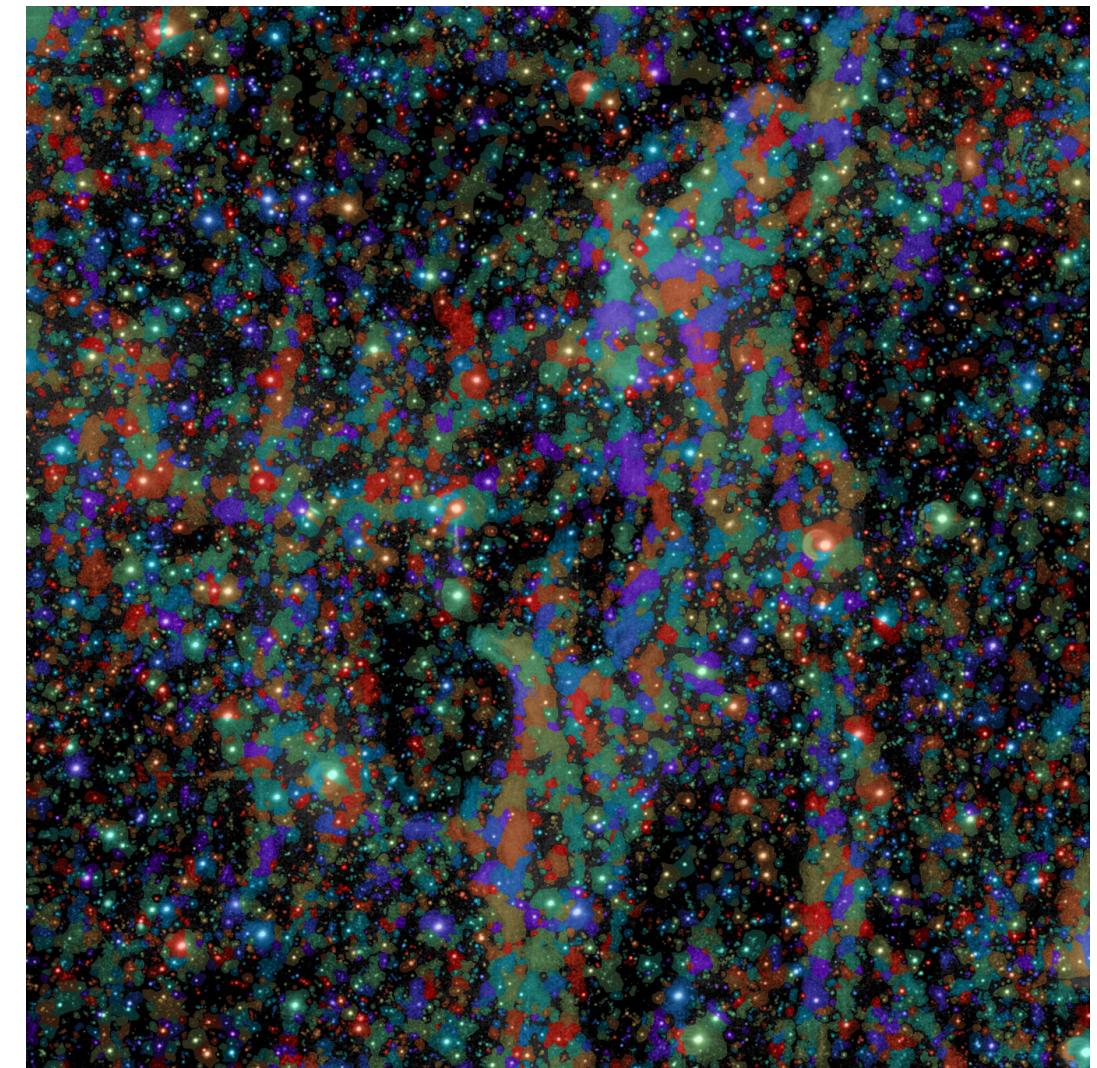


Optimized for Area score

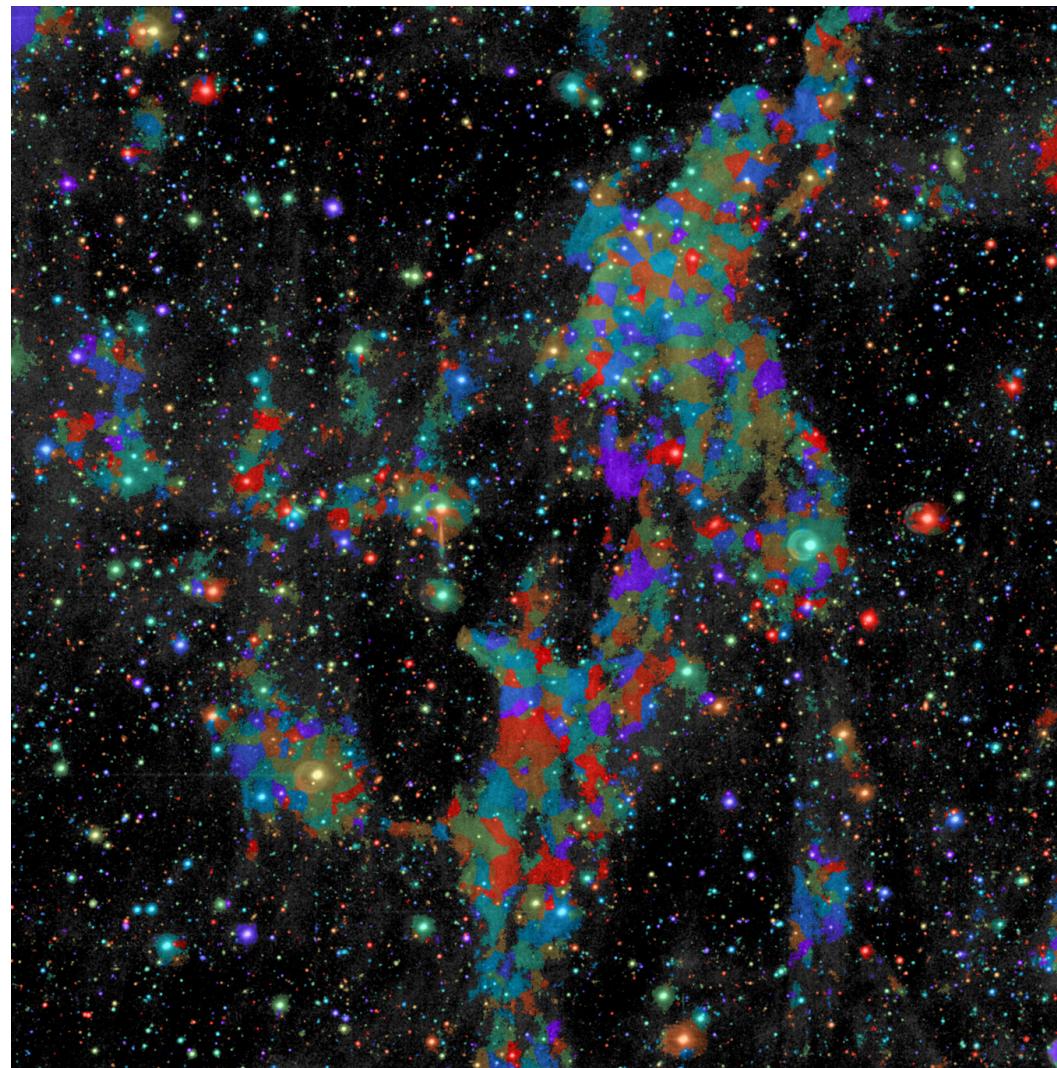


Optimized for F1

ProFound

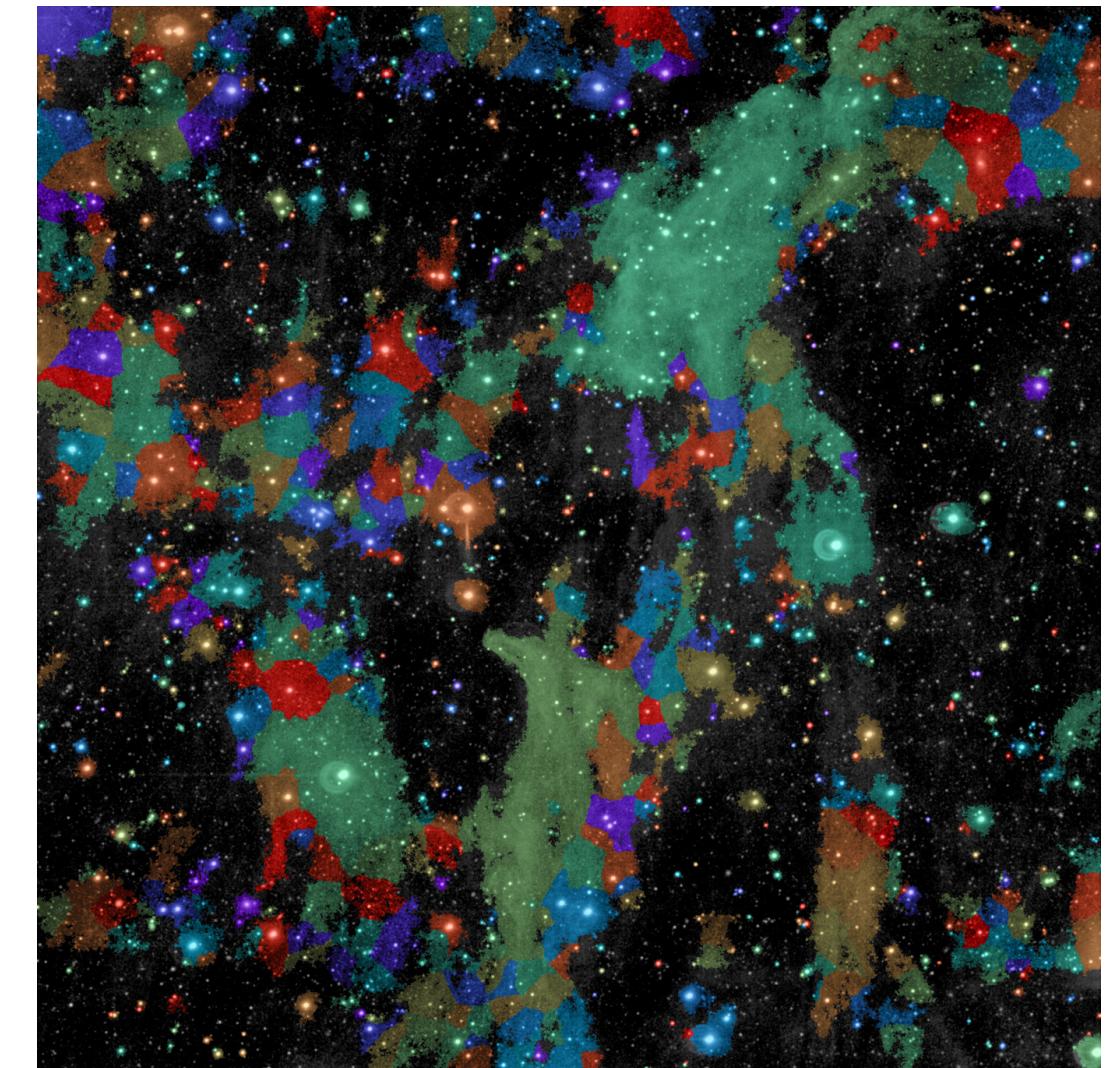


Optimized for Area score

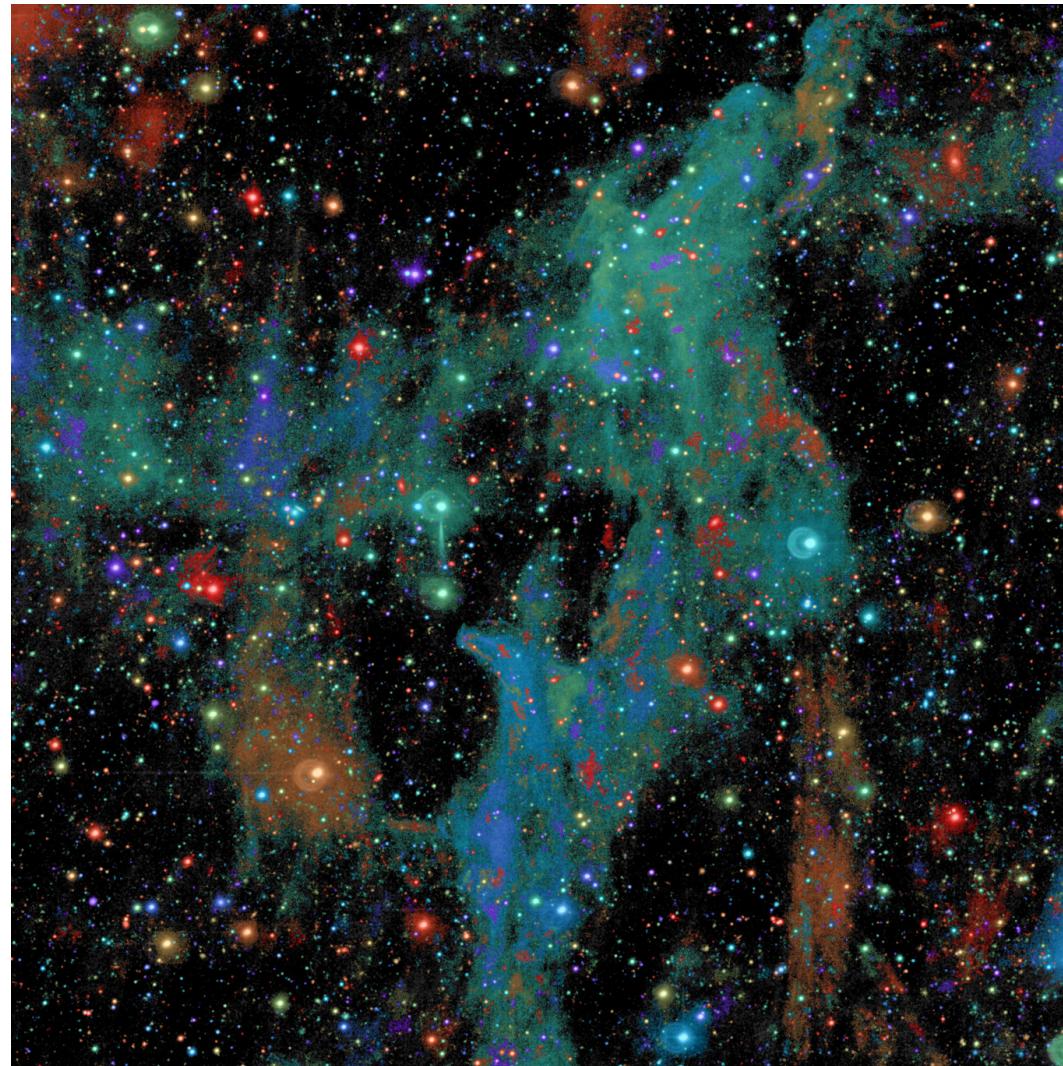


Optimized for F1

NoiseChisel + Segment

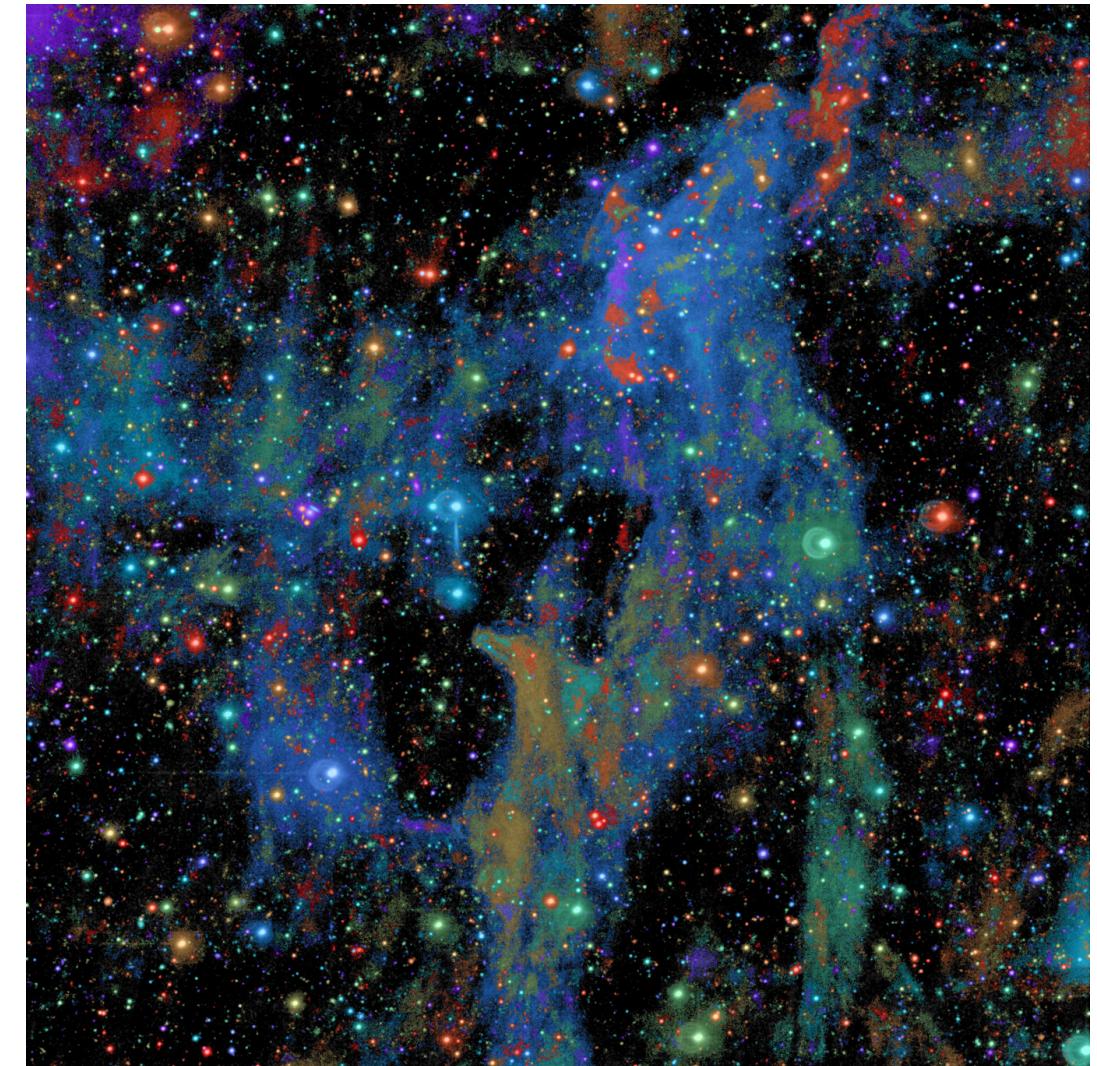


Optimized for Area score

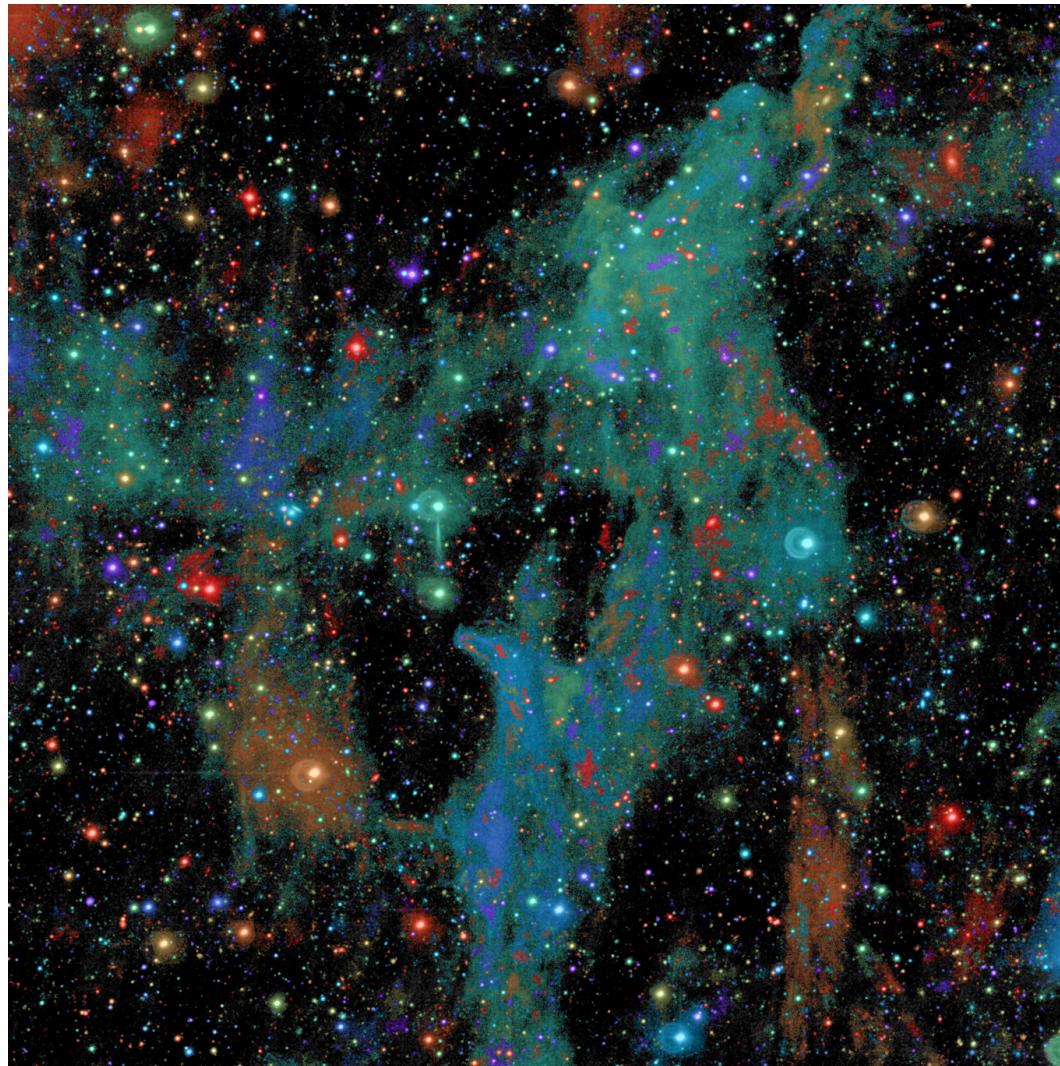


Optimized for F1

MTObjects

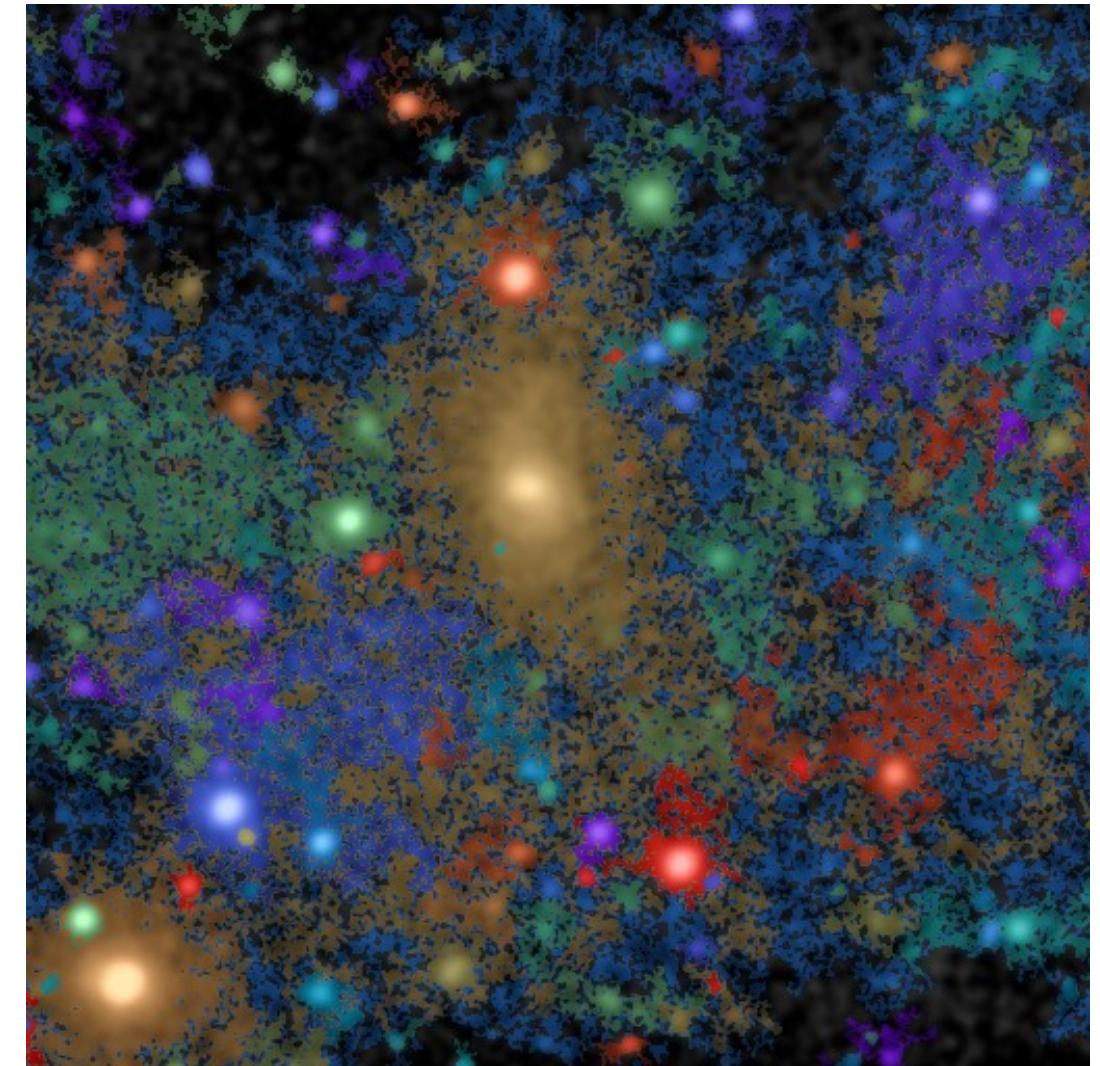


Optimized for Area score



Optimized for F1

MTObjects



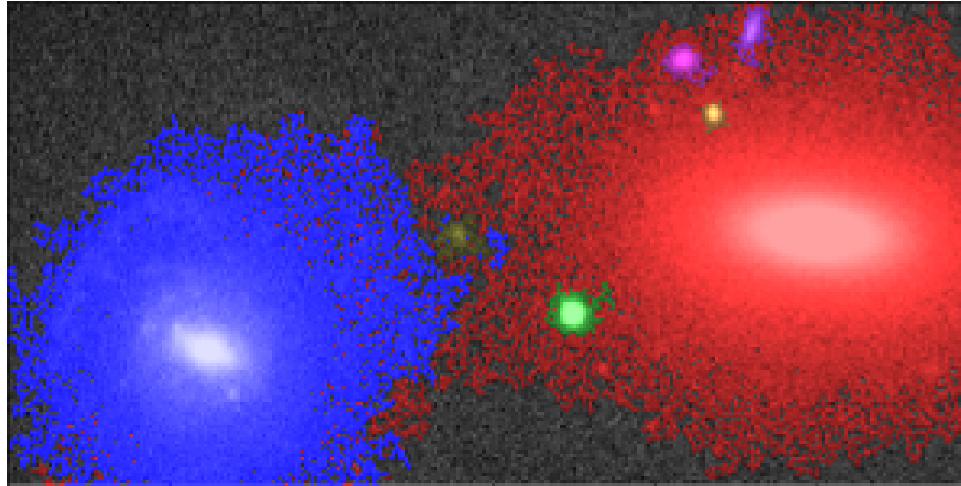
Cut-out



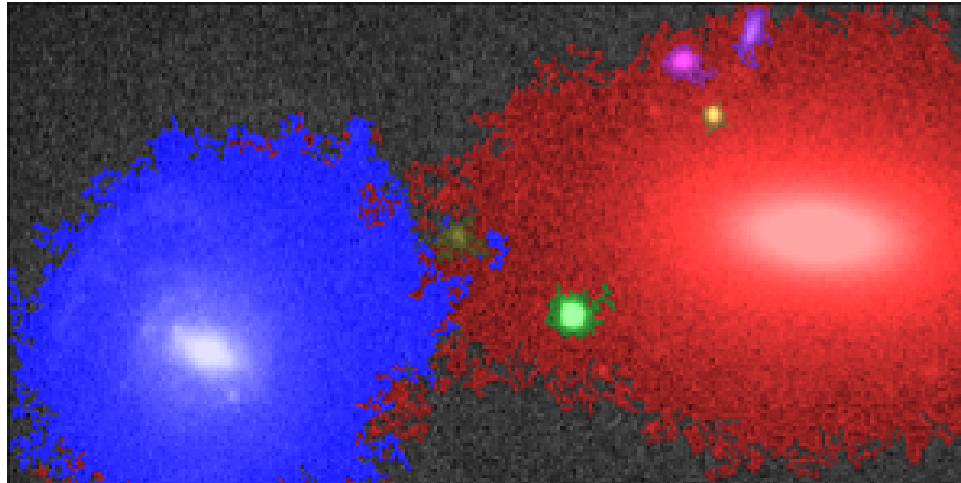
- + Other graph structures
- + Adaptive smoothing methods
- + Other statistical tests
- + Extraction of new features
- + Handling of multi-band images
- + PSF removal
- + Parallelism
- + Handling of massive images



+ Other graph structures



Max-Tree

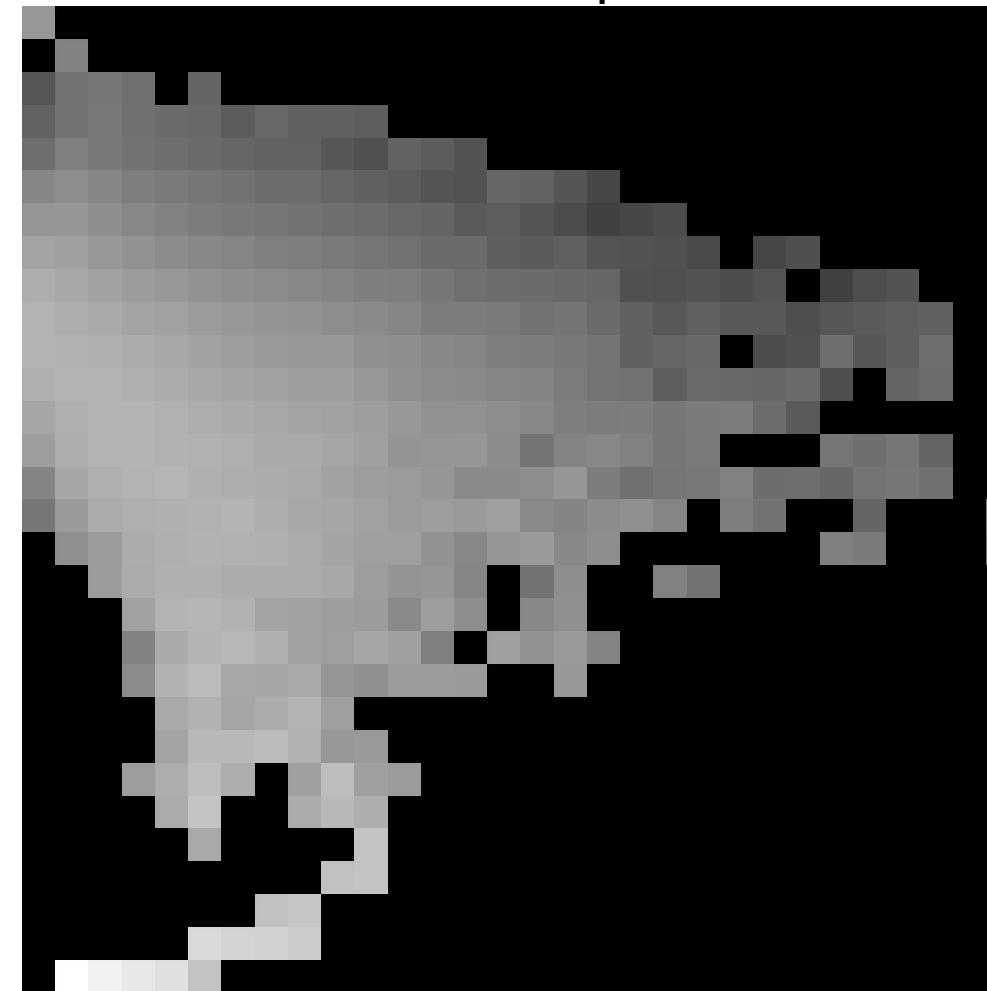
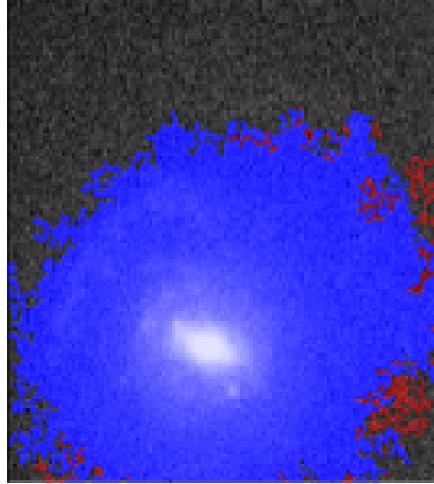
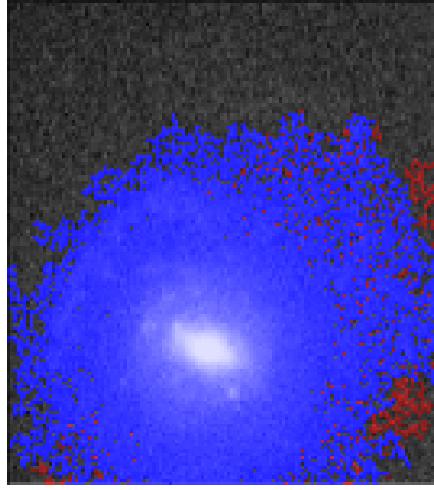


Tree-of-Shapes



+ Other graph structures

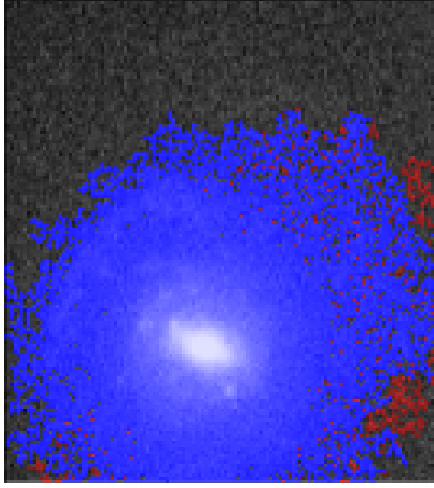
+ Pattern Spectra



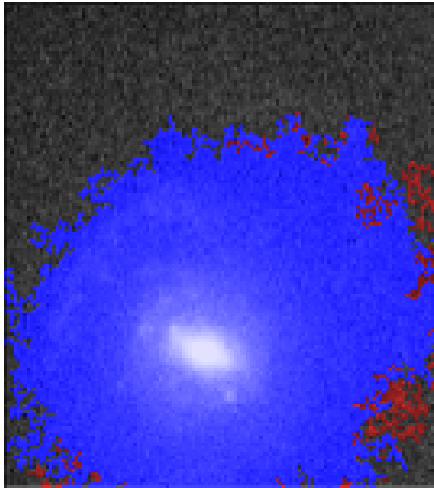
[Urbach, Roerdink, and Wilkinson, TPAMI 2007 ]



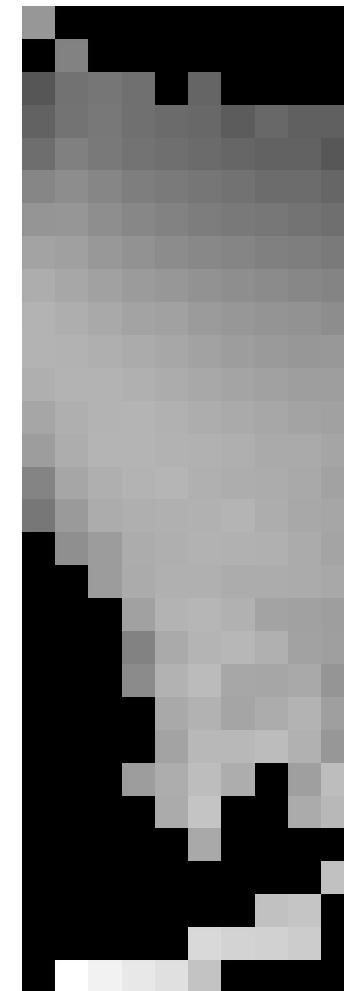
+ Other graph structures



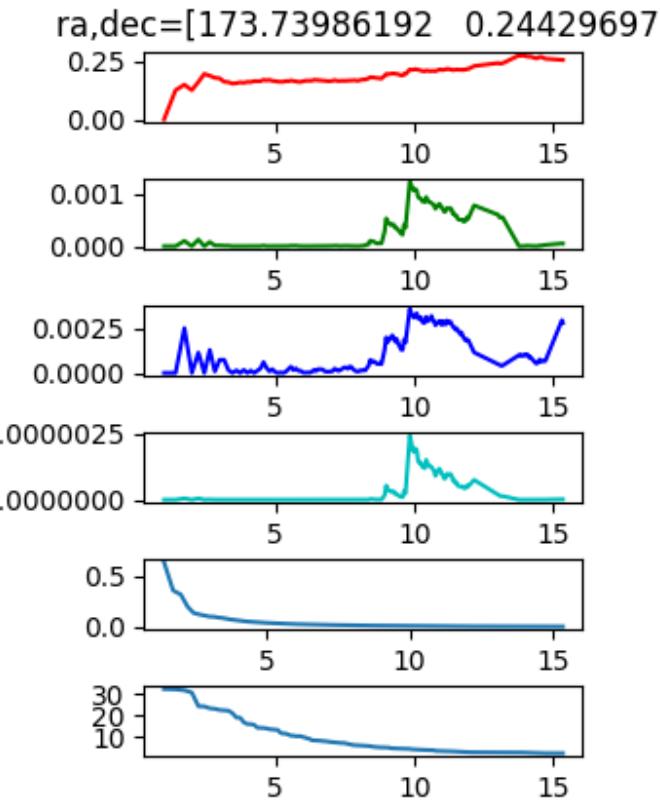
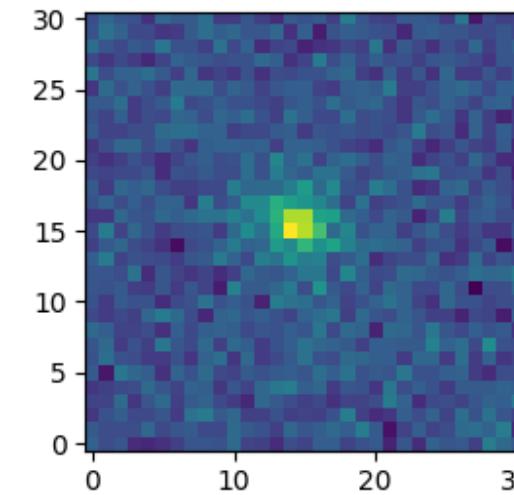
+ Pattern Spectra



+ Morphological Profiles

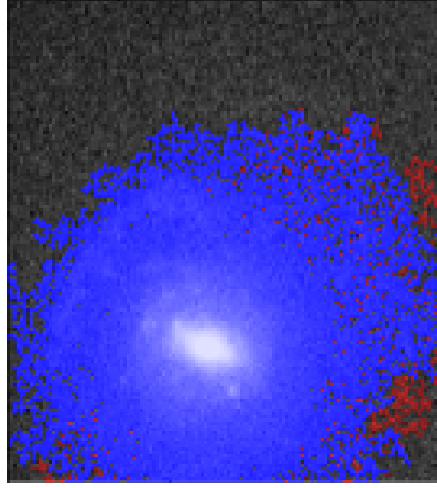


[Urbach, Roerd]

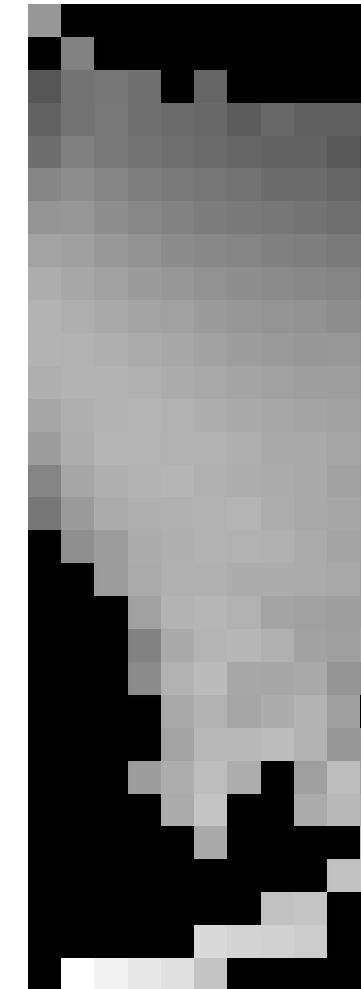




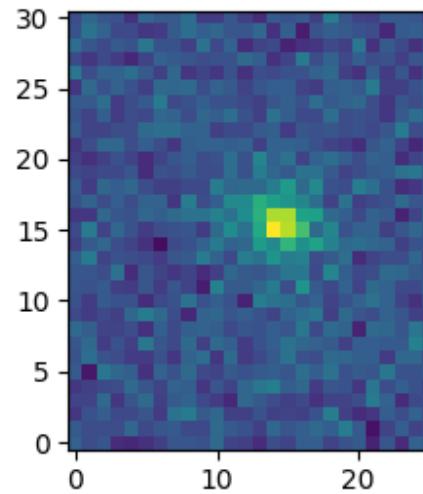
+ Other graph structures



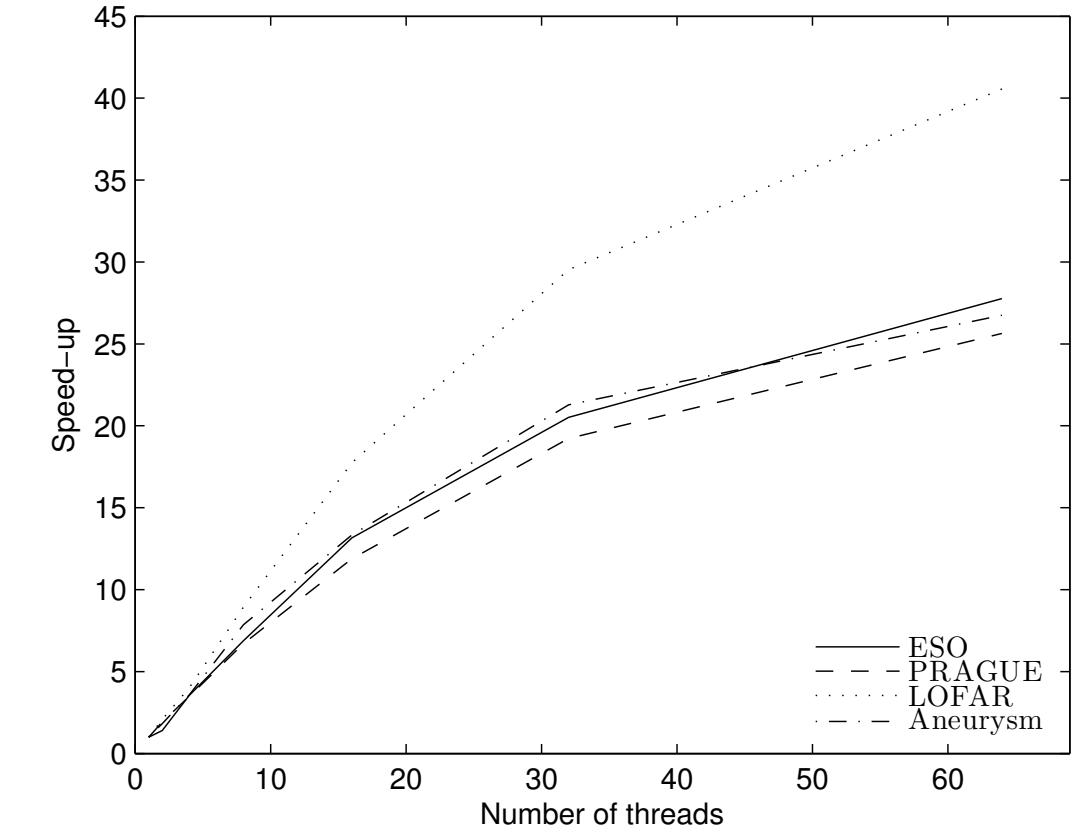
+ Pattern Spectra



+ Morphological Profiles



+ Parallel Execution



[Urbach, Roerd]

[ Moschini, Meijster and Wilkinson, TPAMI 2018 ]



- Bernoulli Institute:  
Caroline Haigh, Simon Gazagnes, Pieter-Jan Eilers,  
Zamir Amiri, Sten Sipma, Loran Oosterhaven,  
Paul Teeninga, Yuanqing Zhang, Michael Wilkinson
- Kapteyn Astronomical Institute:  
Reynier Peletier, Scott Trager
- University of Oulu:  
Aku Venhola
- ESIEE Paris:  
Thanh Xuan Nguyen, Laurent Najman, Giovanni Chierchia,  
Benjamin Perret, Hugues Talbot



We acknowledge financial support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 721463 to the SUNDIAL ITN network.





