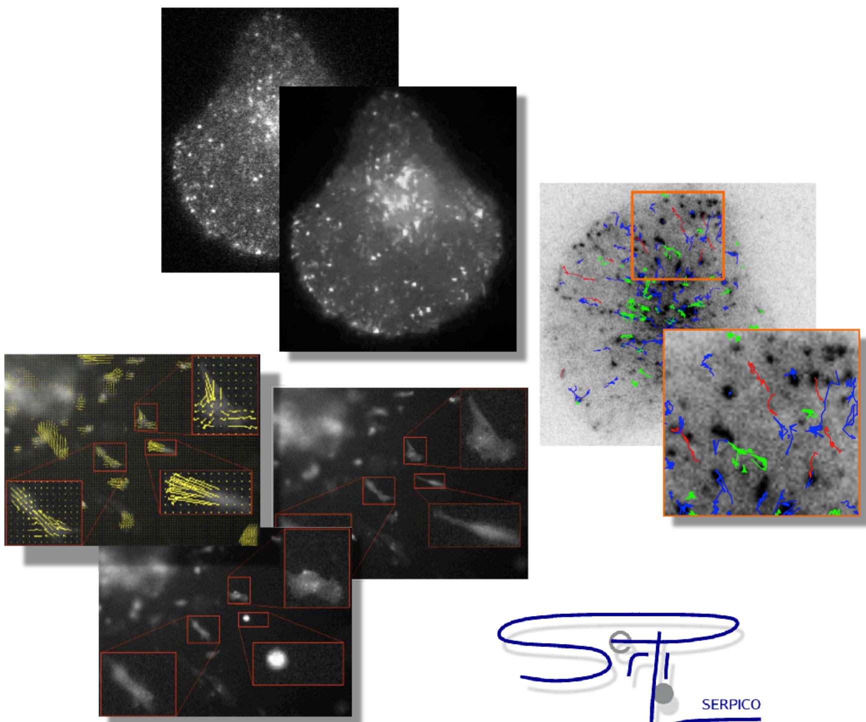
A fast Statistical Colocalization Method for 3D Live Cell Imaging and Super-Resolution microscopy



Charles Kervrann

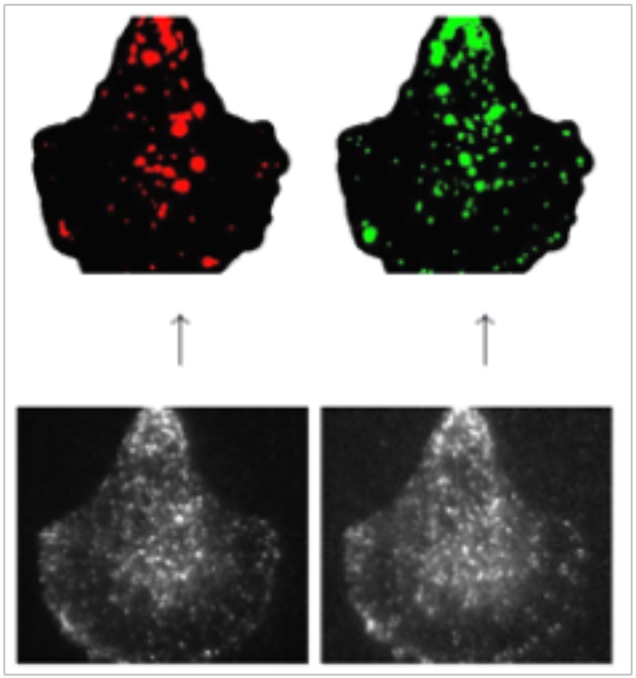
SERPICO Project-Team
http://www.serpico.rennes.inria.fr

Inria Rennes – Bretagne Atlantique
Campus Universitaire de Beaulieu, 35042 Rennes Cedex France

UMR 144 CNRS
26 rue d'Ulm, 75005 Paris

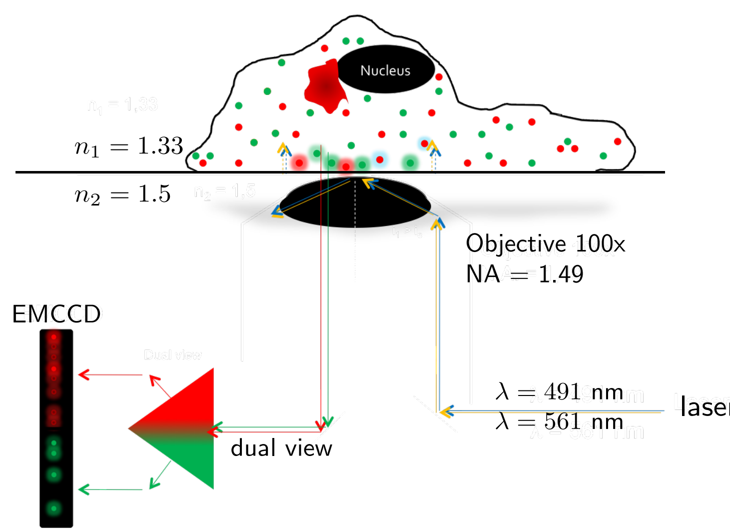
In collaboration with F. Lavancier, T. Pécot and L. Zengzhen

Traffic of two interacting fluorescently tagged proteins in a micro-patterned cell



Rab11 / Langerin proteins interactions in 2D/3D TIRM microscopy

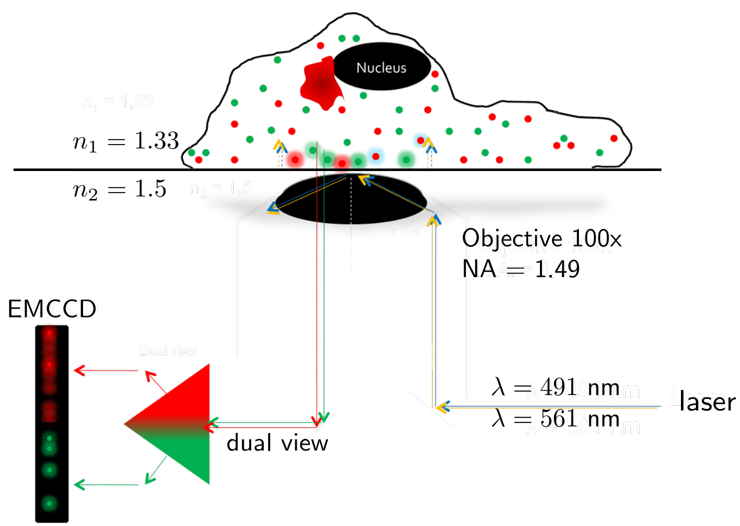
TIRF microscopy for exocytosis event analysis
spatial resolution : 200 nm
acquisition time: 50 ms/frame



Motivation: algorithms for image colocalization

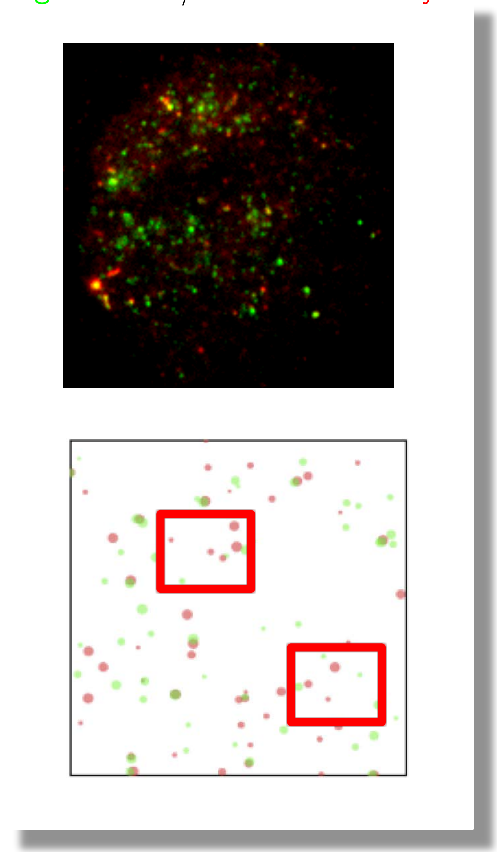
YES or NO ?

- **Colocalization is an open problem** for which no satisfying solution has been found up to now.
- It is still a struggling point of analysis and is usually **badly interpreted**.



2D+time TIRF microscopy (50 ms/frame)
3D TIRF (200 nm × 200 nm × 50 nm)

Traffic of fluorescently tagged proteins
Langerin-YFP / Rab11A-mCherry



Motivation: algorithms for image colocalization

- **Colocalization is an open problem** for which no satisfying solution has been found up to now.
- It is still a struggling point of analysis and is usually **badly interpreted**.

We propose an **automated, robust-to-noise** and **very fast** colocalization method which only needs the adjustment of **one parameter** guarantees more **reproducibility** and more **objective interpretation**.

Motivation: algorithms for image colocalization

- **Robustness** to any signal-to-noise ratio (e.g., low photon counts)
- **Fast** processing of 2D, 3D, 2D+time, 3D+time, and multi-spectral data
- **Flexible** to adapt to multiple image modalities (TIRF, PALM. . .)
- **User-friendly** algorithms with 1 or 2 **"normalized" parameters** (e.g., p-value)
- **Theoretical properties** and performance wrt state-of-the-art
- Adaptation **high-throughput imaging** and **high-content screening**

1

Previous work

Focus on two related colocalization approaches

Pearson's correlation coefficient (PCC)
Manders' co-localization coefficients (MCC)

- PCC and MCC are **fast** to compute and **very popular**.

- **Difficulties** of PCC and MCC:
  - Image background
  - Presence of noise and/or shift
  - Amount of photons/molecules
  - Practice: Impossible to define a clear and objective threshold for which there is actual colocalization in all situations.

$\rightsquigarrow$ **PCC and MCC do not provide any decision rule**.

Recent colocalization approaches

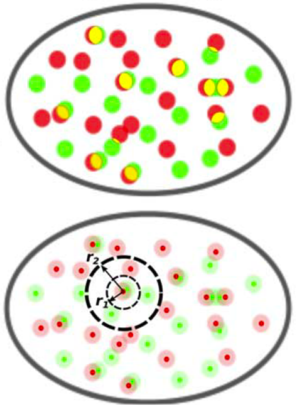
1. Sherman, E. et al. Immunity 35, 705–720 (2011)
2. Lehmann, M. et al. PloS Pathogens 7(12), e1002456 (2011)
3. Helmuth, J. et al., I. Sbalzarini BMC Bioinformatics 11, 372 (2010)
4. van Steensel, B.et al. J of Cell Science 109, 787–792 (1996)
5. Adler, J. & Parmryd, I. Cytometry part A 77, 733–742 (2010)
6. Szymborska et al. Science 341, 655 (2013)
7. Zinchuk, V. et al. Nature Protocols, 6(10), 1554–1566 (2011)
8. Rizk, A. et al. Nature Protocols 9(3), 586–596 (2014)
9. Endesfelder U. et al. Histochem Cell Biol (2014)
10. Malkusch S. et al. Histochem Cell Biol 137 (2012)
11. Malkusch S. & Heilemann M., Sci Reports 6:34486 (2016)

...

Focus on two related colocalization approaches

- **Intensity-based method**: the **Costes method**[1] is a permutation test applied to the PCC and requires $n = 1000$ simulated images !

- **Object-based method**: the **Lagache method**[2] consists in



  1. segmenting the two images to provide objects,
  2. reducing each object by its mass center,
  3. applying spatial statistics tools (Ripley's K function) to test dependance in the two point-patterns.

The Costes and Lagache methods return a **p-value** and provide a clear decision rule (i.e., co-localization if p-value < 0.05).

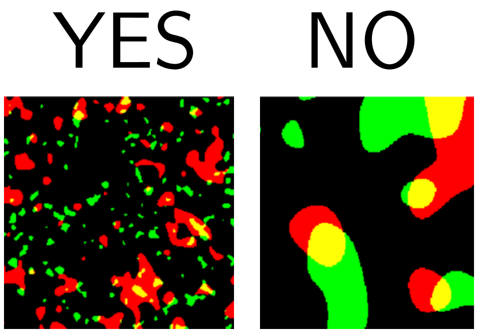
[1] Costes S.V., Daelemans D., Cho E.H., Dobbin Z., Pavlakis G., Lockett S., Biophys J. 86, 3993-4003 (2004).

[2] Lagache, T., Sauvonnet, N., Danglot, L., Olivo-Marin, J.-C., Cytometry Part A 87(6), 568-579 (2015).

Focus on two related colocalization approaches

- **Limitations of the patch-based Costes method:**
  - is computationally costly.
  - leads to too many false positives.

- **Limitation of the Lagache method dedicated to point-like objects:**
  - is not sufficiently sensitive to detect colocalization.
  - loses a lot of information by reducing each object to a point.



[1] Costes S.V., Daelemans D., Cho E.H., Dobbin Z., Pavlakis G., Lockett S., Biophys J. 86, 3993-4003 (2004).

[2] Lagache, T., Sauvonnet, N., Danglot, L., Olivo-Marin, J.-C., Cytometry Part A 87(6), 568-579 (2015).

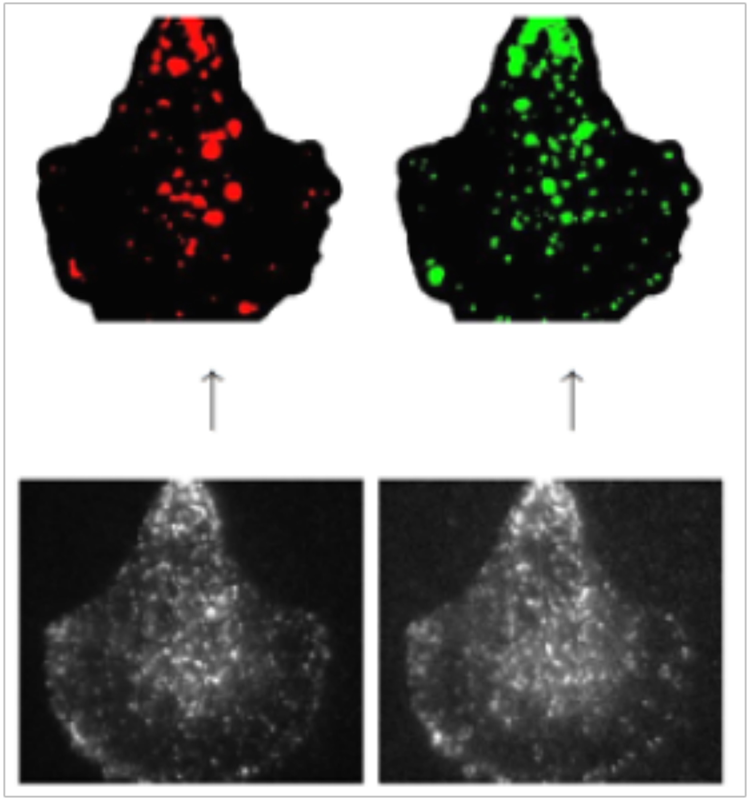
2

GcoPS: random set-based colocalization method

A new, versatile "swiss-knife" method for any types of objects

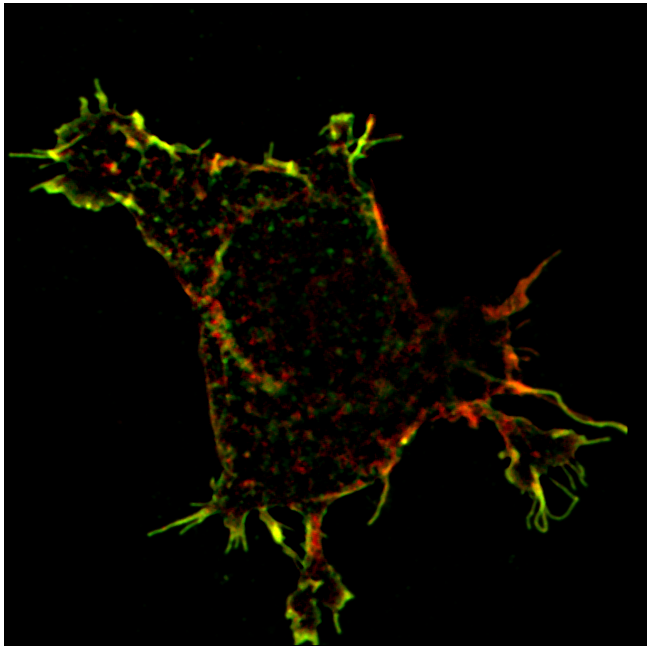
- **Each channel is segmented:**
  - not sensitive in our case
  - basic or sophisticated algorithm (e.g. Atlas[3])
- **Random sets framework:**
  - the two binary images are viewed as realization of two random sets

Rab11 / Langerin proteins interactions in 2D/3D TIRM microscopy

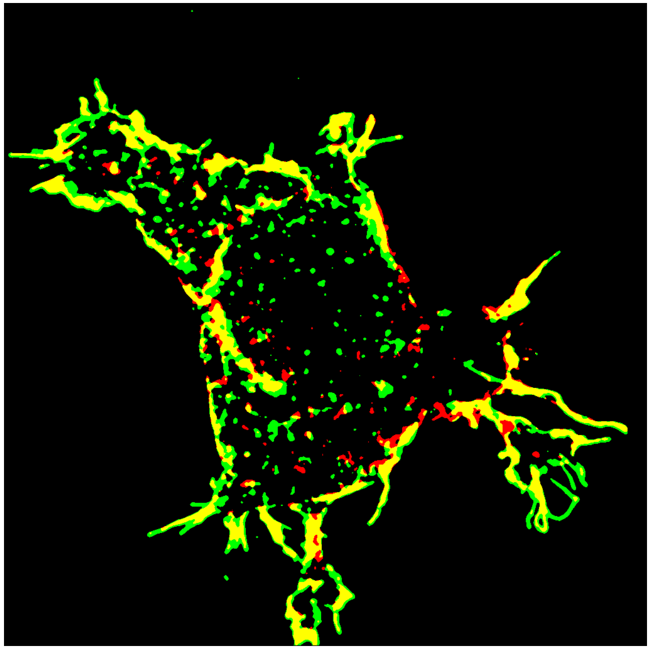


[3] Basset, A., Boulanger, J., Salamero, J., Bouthemy, P., Kervrann, C. Adaptive spot detection with optimal scale selection in fluorescence microscopy images, IEEE Transactions on Image Processing, 24(11):4512-4527 (2015)

A new, versatile "swiss-knife" method for any types of objects



c-Src tyrosine kinase
serotonin receptor



Segmentation of the 2 channels

Idea and probabilistic theory

- **Notation:** Define Γ_1, Γ_2 two *random sets* $\mathbb{R}^d$ (segmented objects in channels 1 and 2). For any point $x \in \Omega \subset \mathbb{R}^d$

$$p_1 = P(x \in \Gamma_1), \quad p_2 = P(x \in \Gamma_2), \quad p_{12} = P(x \in \Gamma_1 \cap \Gamma_2).$$

- **Independence hypothesis:**

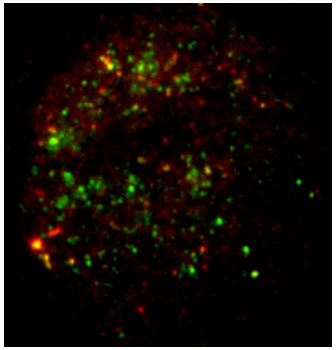
> **If Γ_1 and Γ_2 are independent $p_{12} = p_1 p_2$.**

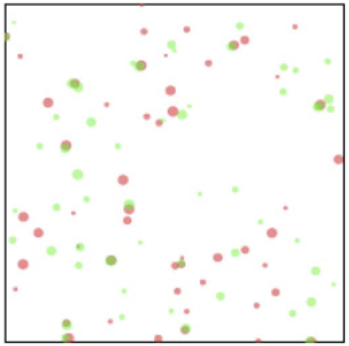
- **Empirical approximation:**

$$\hat{p}_1 = \frac{1}{|\Omega|} \sum_{x \in \Omega} \mathbf{1}_{\Gamma_1}(x), \qquad \hat{p}_{12} = \frac{1}{|\Omega|} \sum_{x \in \Omega} \mathbf{1}_{\Gamma_1}(x) \mathbf{1}_{\Gamma_2}(x)$$

$$\hat{p}_2 = \frac{1}{|\Omega|} \sum_{x \in \Omega} \mathbf{1}_{\Gamma_2}(x),$$

> **What are the fluctuations of $D = \hat{p}_{12} - \hat{p}_1 \hat{p}_2$?**





GcoPS (Geo-coPositioning System) testing procedure

1. Each image is **segmented** into a set of objects.
2. The **colocalization score** is computed as

$$T = \frac{D}{\sqrt{\hat{V}_D}}$$

   where $\hat{V}_D$ is the asymptotic variance of D.

3. The null hypothesis of **no-positive colocalization** is rejected if $T > q(\alpha)$ corresponding to

$$\boxed{p\text{-value} \quad = \quad 1 - \Phi(T)}$$

   where Φ denotes the cdf of $\mathcal{N}(0, 1)$.

If T is large, p-value is small and colocalization is positive if p-value $< \alpha$ (e.g. $\alpha = 0.05$).

Some theory and consistency

- If Γ_1 and Γ_2 are independent (and stationary), $\mathbb{E}[D] = 0$,

$$\text{Var}[D] \approx |\Omega|^{-2} \sum_{x \in \Omega} \sum_{y \in \Omega} C_1(x-y) C_2(x-y) \quad \text{as} \quad |\Omega| \to \infty.$$

  where C_1 and C_2 are the **auto-covariance** functions of Γ_1 and Γ_2.

- If Γ_1 and Γ_2 are **stationary independent random sets**, and if $C_1(x-y)$ and $C_2(x-y)$ tend fast enough to 0 as $|x-y| \to \infty$, then

$$\boxed{\frac{D}{\sqrt{\hat{V}_D}} \to \mathcal{N}(0,1) \quad \text{as} \quad |\Omega| \to \infty}$$

  where $\hat{V}_D = |\Omega|^{-2} \sum_{x \in \Omega} \sum_{y \in \Omega} \hat{C}_1(x-y) \hat{C}_2(x-y)$

  and $\hat{C}_1$ and $\hat{C}_2$ are the **estimates of C_1 and C_2 by FFT**.

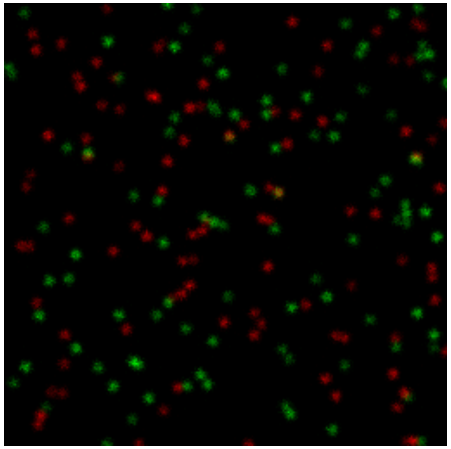
3

Experimental results and algorithm comparisons

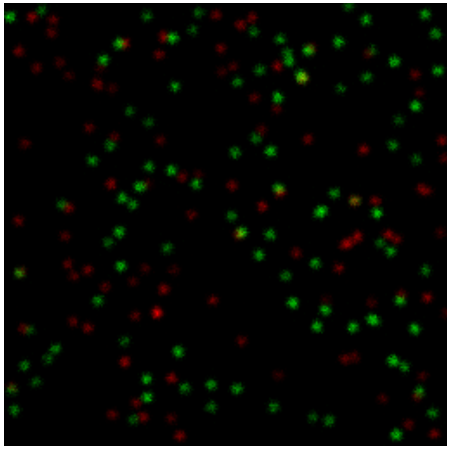
Two simulators to generate artificial colocalized images

1. Simulator from Lagache[2]:

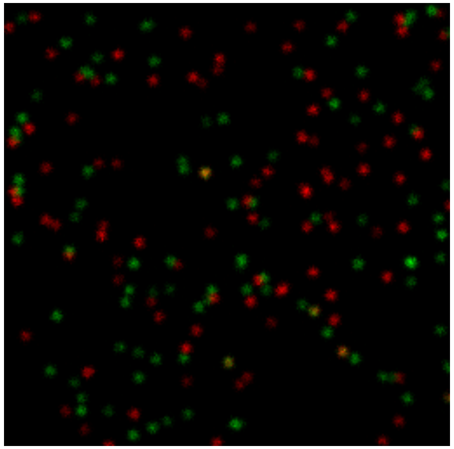
- Generate randomly distributed "red" particles.
- Simulate a proportion of "green" particles nearby "red" particles. The proportion is referred to as the *level of colocalization*.
- The other "green" particles are drawn randomly and independently.
- Add a diffraction effect and possibly a noise and/or a shift.



0% colocalization



2.5% colocalization

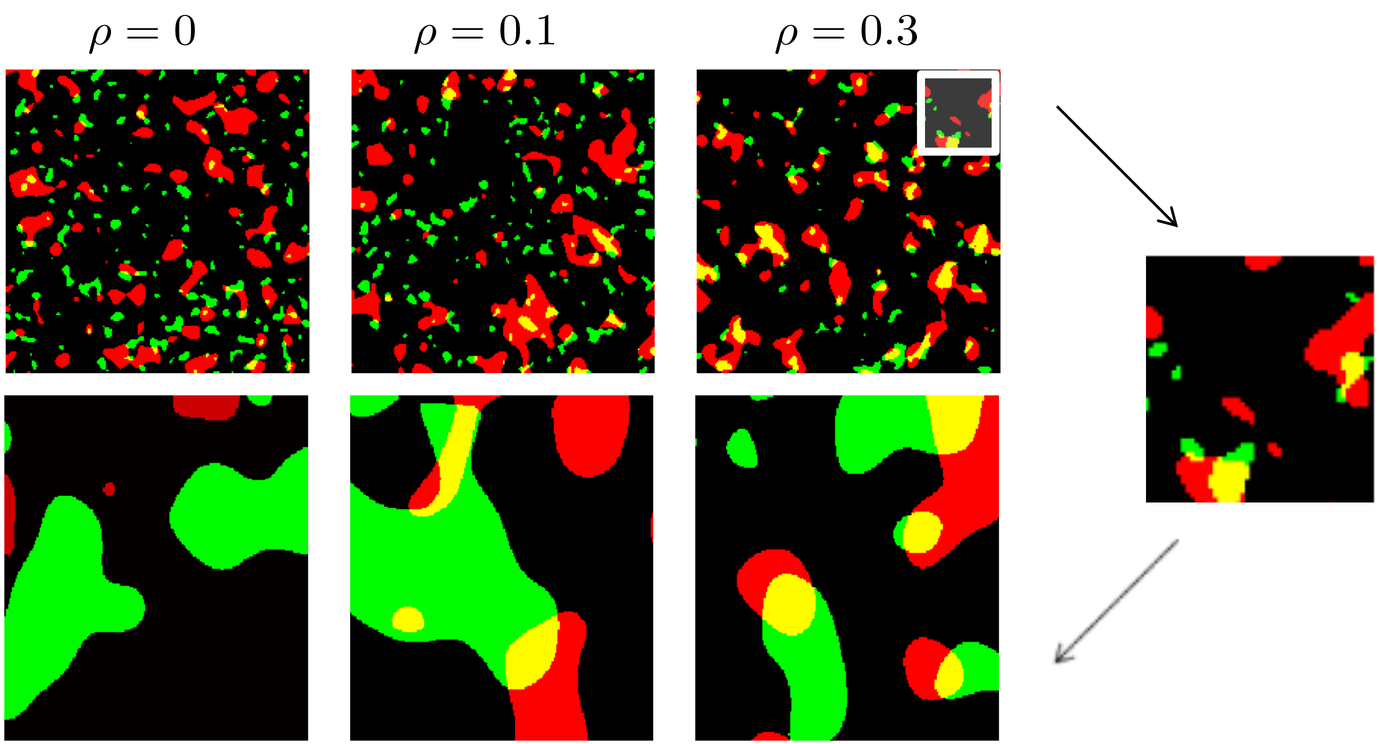


5% colocalization

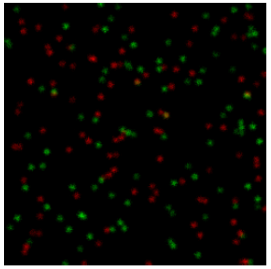
[2] Lagache, T., Sauvonnet, N., Danglot, L., Olivo-Marin, J.-C., Cytometry Part A 87(6), 568-579 (2015).

Two simulators to generate artificial colocalized images

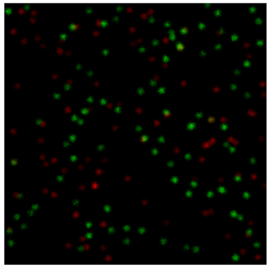
2. **Simulator from Gaussian level sets:**
   - simulation of two different image resolutions.
   - simulation of two weakly/well deconvolved images.
   - simulation of large/small objects and arbitrary shapes.



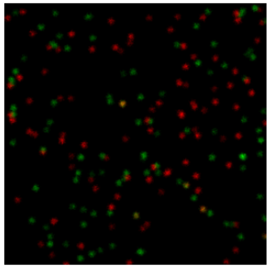
Evaluation on synthetic images: results



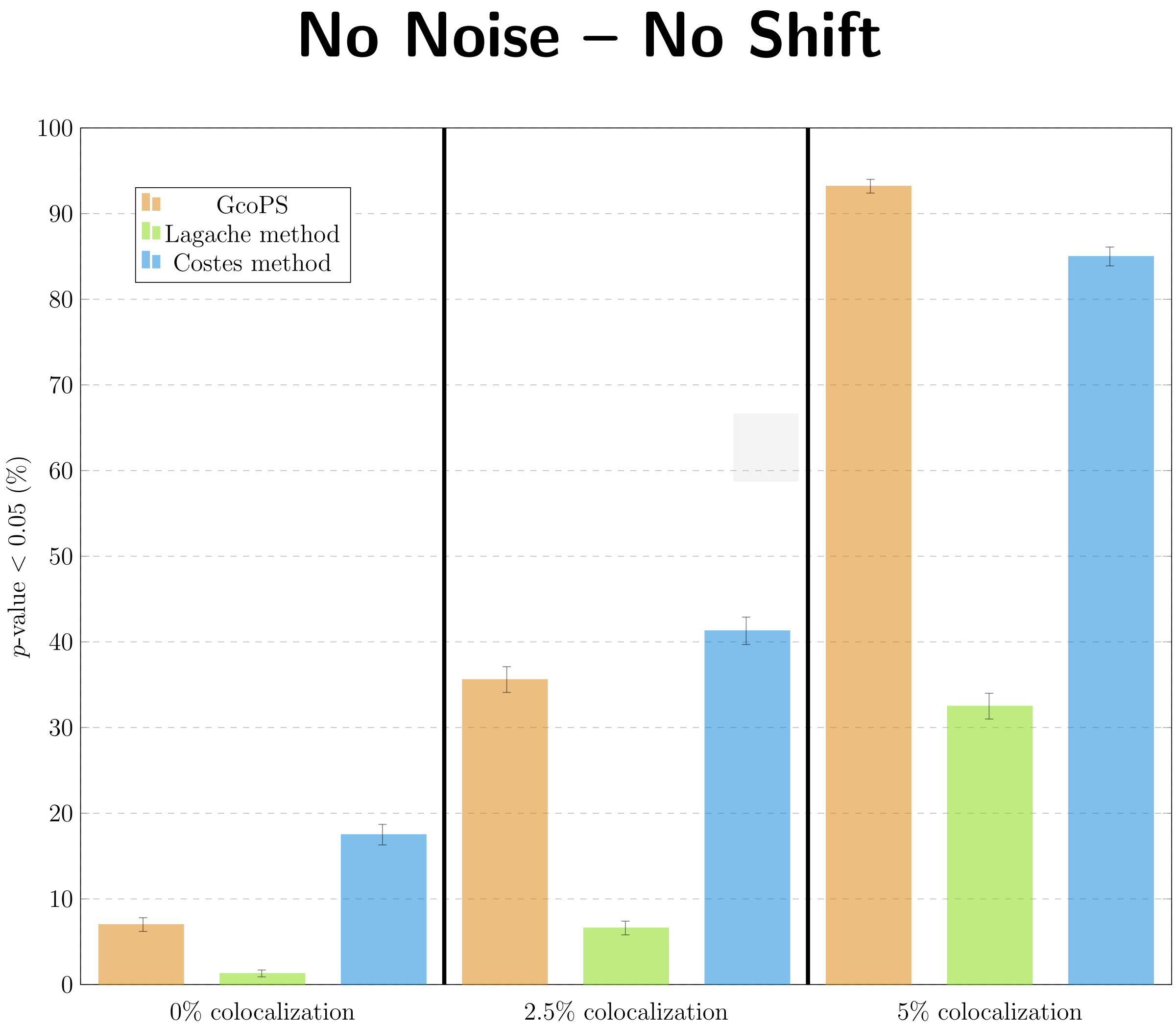
0% colocalization



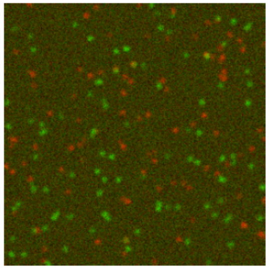
2.5% colocalization



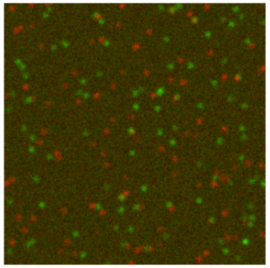
5% colocalization



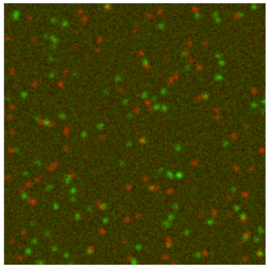
Evaluation on synthetic images: results



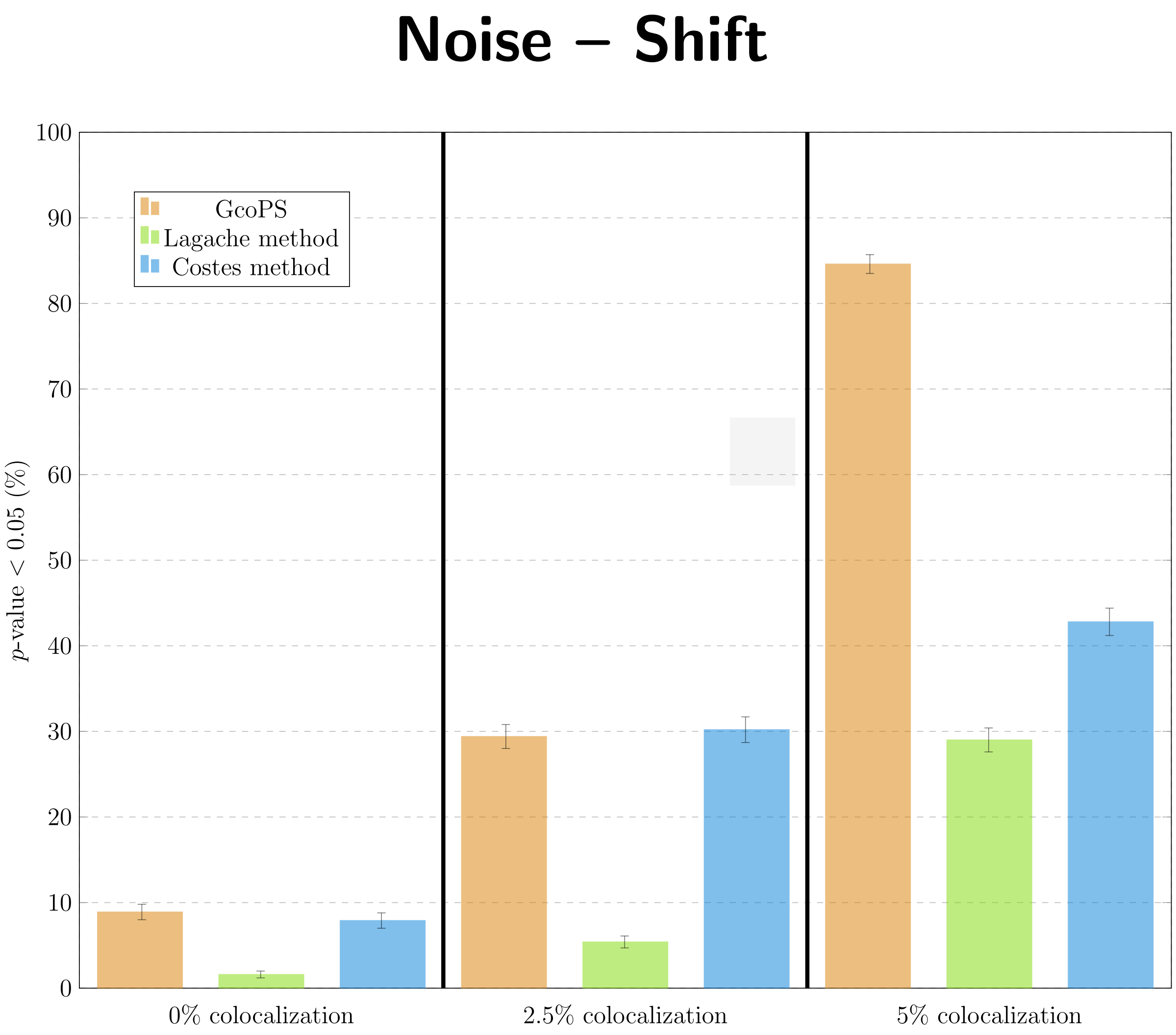
0% colocalization



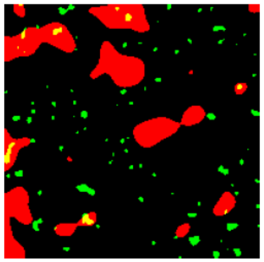
2.5% colocalization



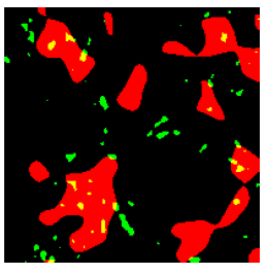
5% colocalization



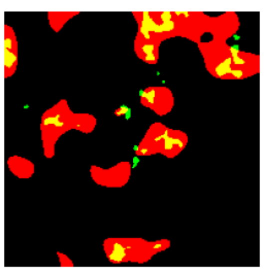
Evaluation on synthetic images: results



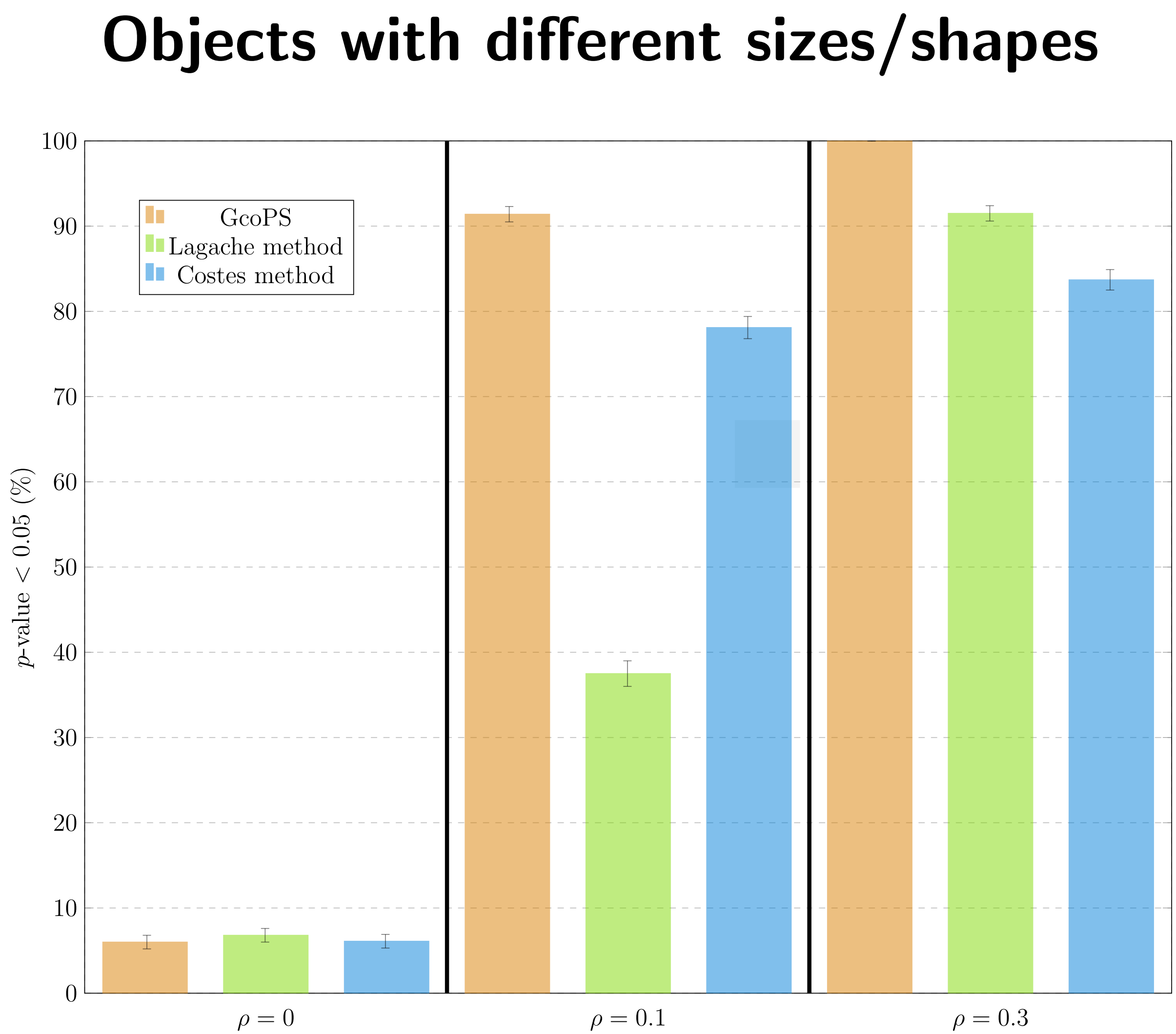
$\rho = 0$



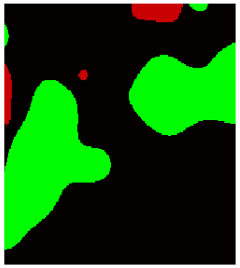
$\rho = 0.1$



$\rho = 0.3$



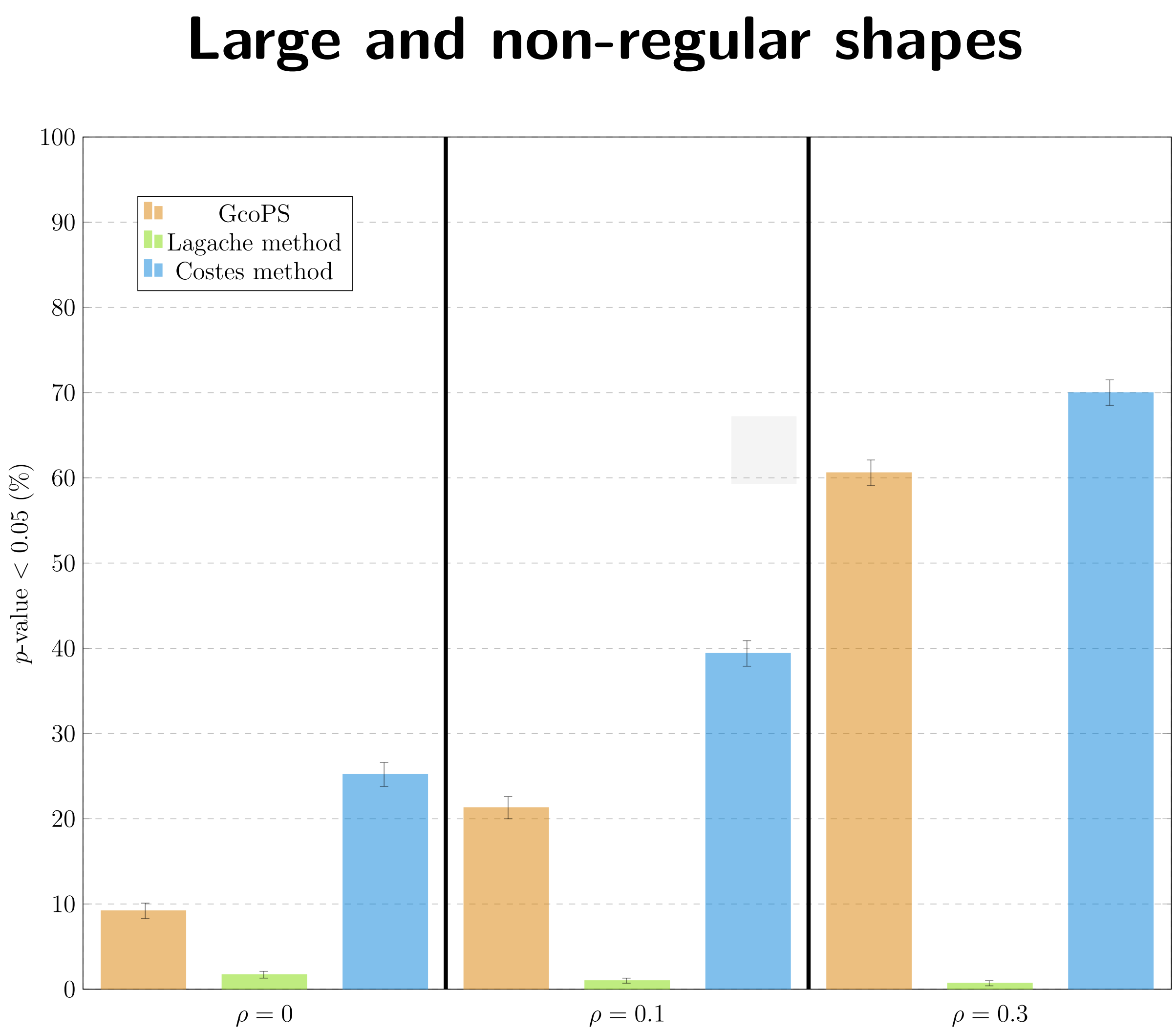
Evaluation on synthetic images: results



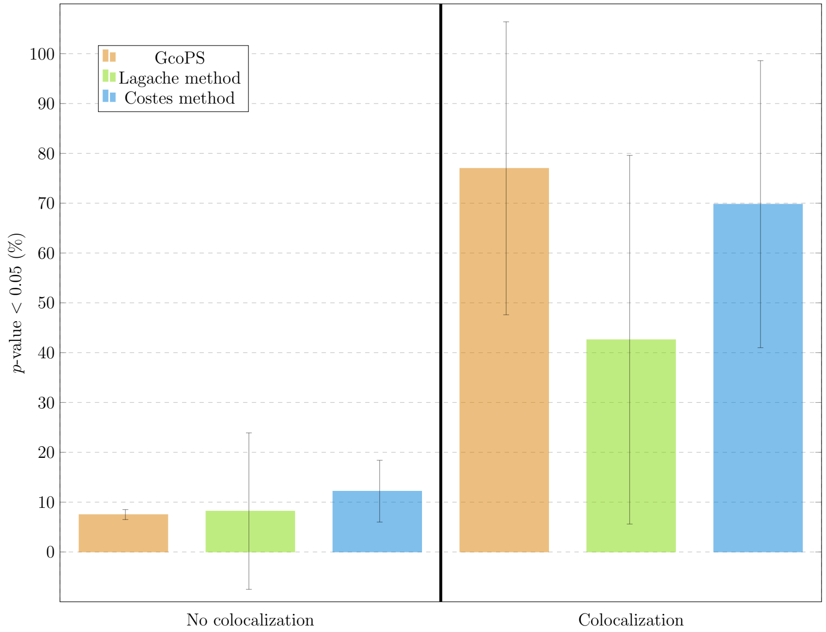
$\rho = 0$

$\rho = 0.1$

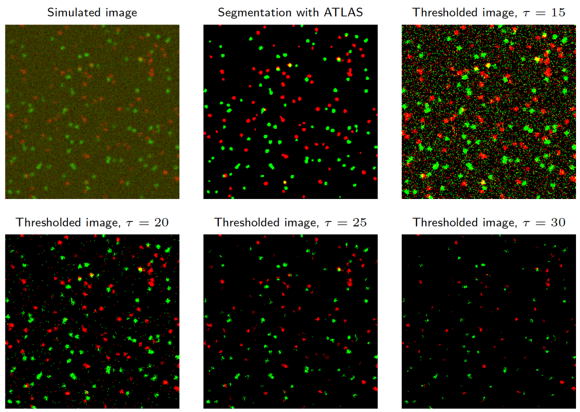
$\rho = 0.3$



Evaluation on synthetic images: results



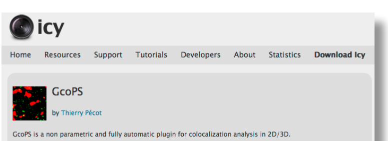
Overall results



	Costes (2004)	Lagache (2015)	GcoPS (2017)
CPU Time	★	★★★	★★★★
Sensitivity to method parameters	★	★★	★★★
Numbers of false positives	★	★★★★	★★★
Sensitivity to colocalization (true positives)	★★★★	★	★★★★
Robustness to segmentation outputs	★★★★	★★	★★★★
Robustness to non-regular shaped objects	★★	★	★★★★
Robustness to a different optical resolution	★★★	★★	★★★★

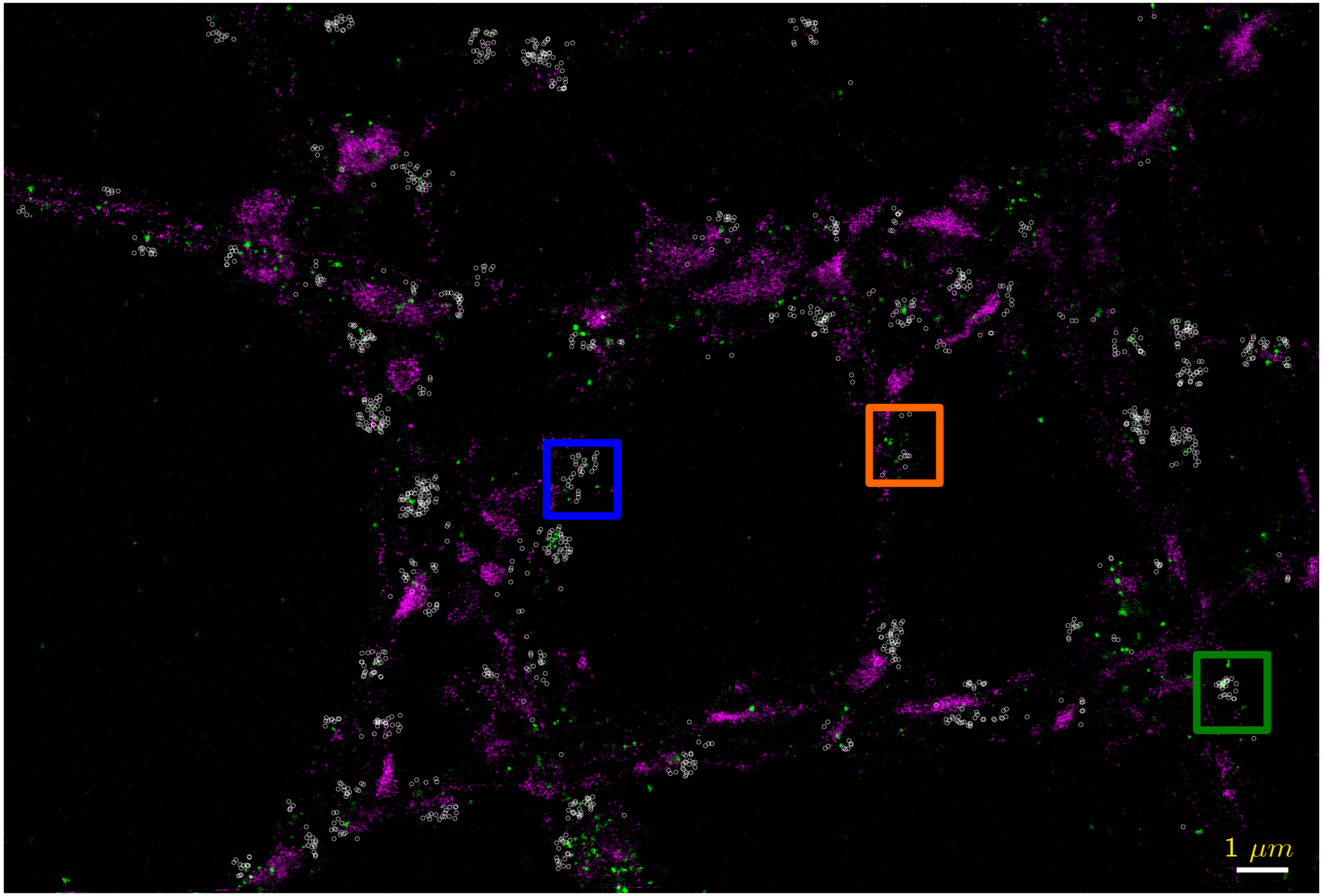
Evaluation on synthetic images: results

	2D image 256×256 50 objects	**2D image** 256×256 200 objects	**2D image** 256×256 3500 objects	**2D+time image** $256 \times 256 \times 1000$ 100 objects	**3D image** $256 \times 256 \times 60$ 1000 objects	**3D image** $256 \times 256 \times 60$ 2000 objects
Costes ImageJ	6.1 sec	6.2 sec	6.1 sec	38 min 20 sec	3 min 3 sec	3 min 10 sec
Lagache Icy	1 sec	1.96 sec	12.38 sec	12 min 39 sec	25 sec	60 sec
GcoPS C++	**0.18 sec**	**0.2 sec**	**0.19 sec**	**29.5 sec**	**10 sec**	**9.8 sec**
GcoPS Icy	**0.77 s**	**0.86 sec**	**0.82 sec**	**2 min 50 sec**	**22 sec**	**21 sec**



Evaluation on real super-resolution images:

dSTORM (direct Stochastic Optical Reconstruction Microscopy)

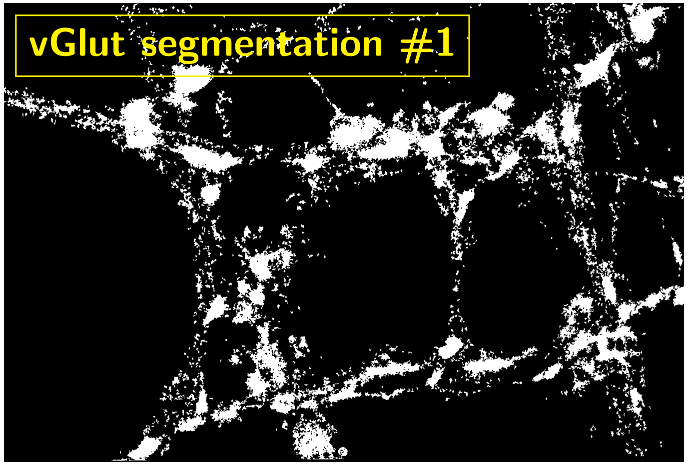


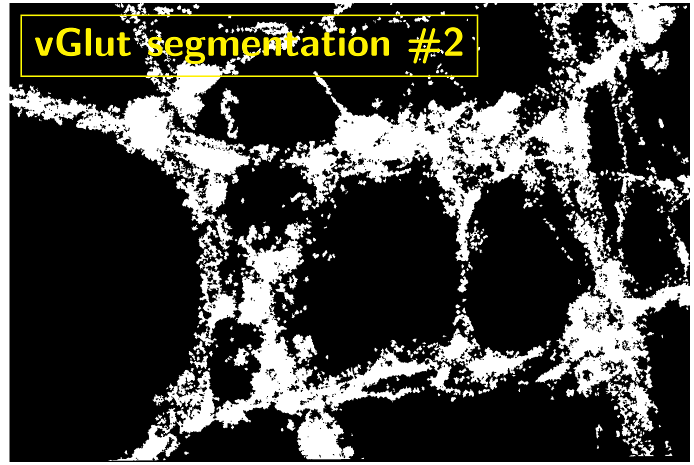
dSTORM acquisition of cells from hippocampi of mice expressing BDNF proteins (green) and vGlut (purple) (courtesy of M. Sauer, University Würzburg, Germany).

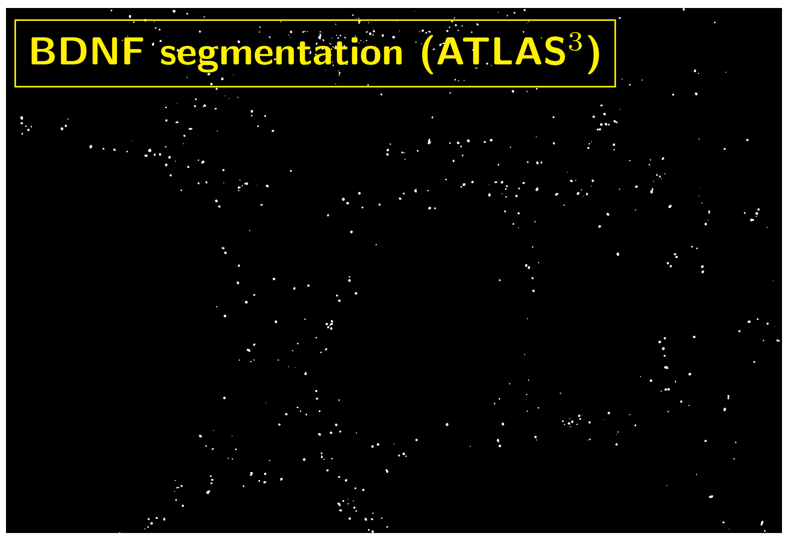
(2547×1724 image, scale bar: $1\mu m$, pixel: $30nm$)

Andreska, T. et al., R. High abundance of BDNF within glutamatergic presynapses of cultured hippocampal neurons. Front Cell Neurosci 8 , 1–15 (2014).

Evaluation on real images: dSTORM



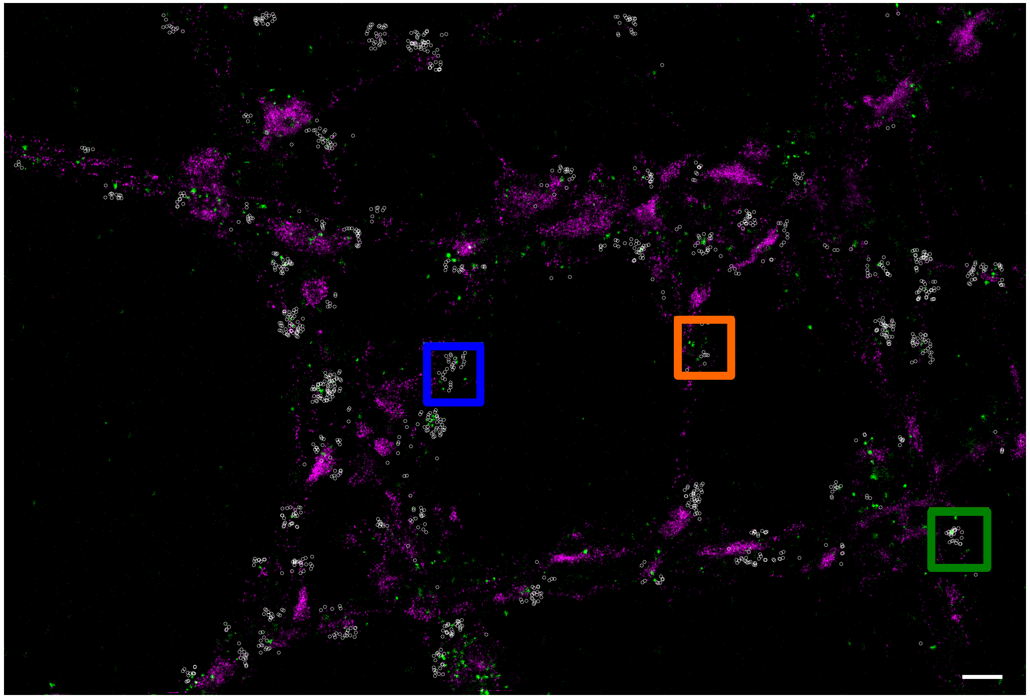




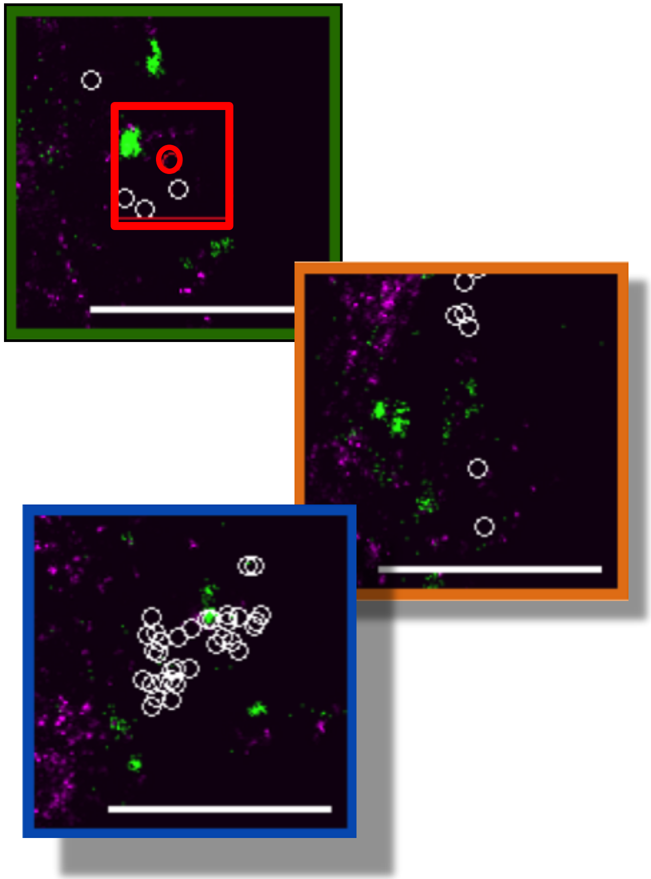
[3] Basset, A., Boulanger, J., Salamero, J., Bouthemy, P., Kervrann, C. Adaptive spot detection with optimal scale selection in fluorescence microscopy images, IEEE Transactions on Image Processing, 24(11):4512-4527 (2015)

Global colocalization score: $T = 4.34$
$p - value = 0.000000703$

Evaluation on real images: dSTORM

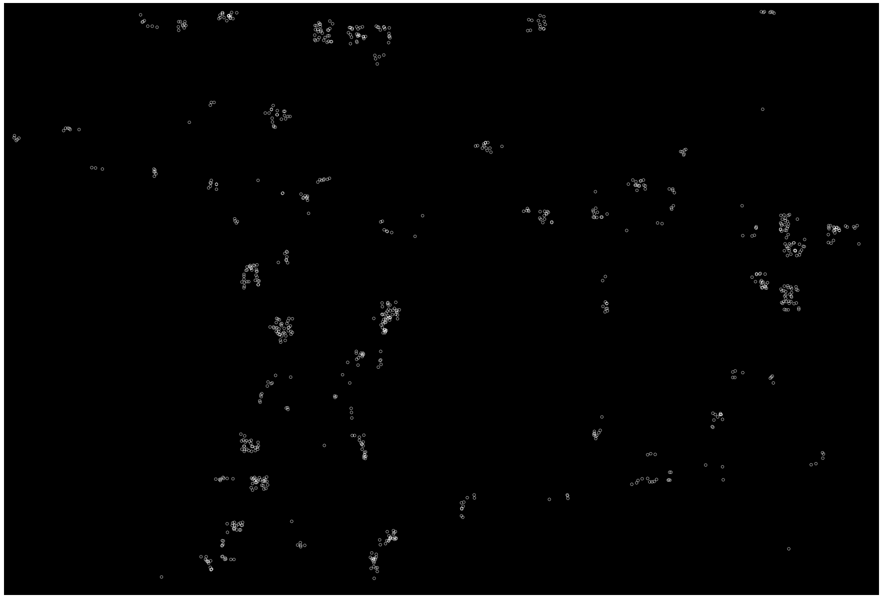


dSTORM acquisition of cells from hippocampi of mice expressing BDNF proteins (green) and vGlut (purple) (courtesy of M. Sauer, Julius-Maximilians-University Würzburg, Germany)

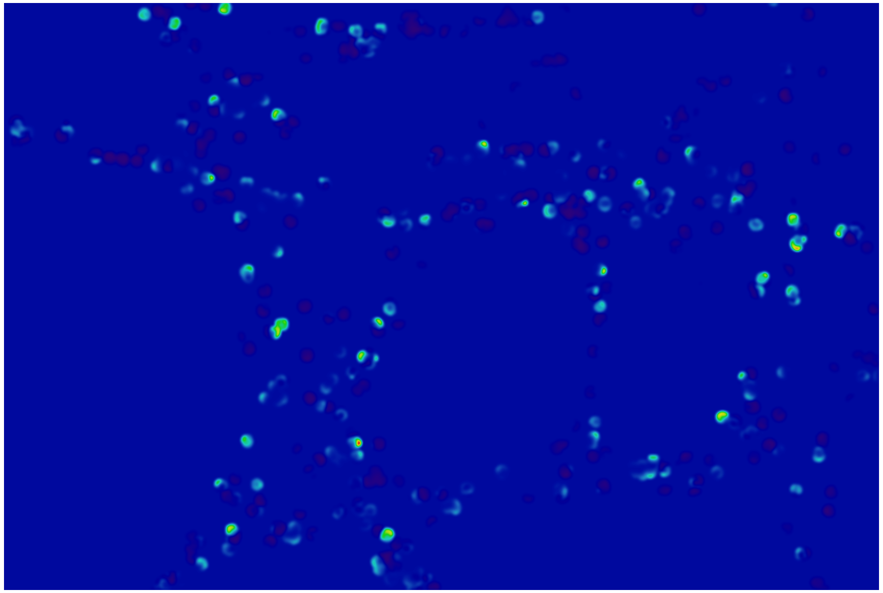


The red rectangle represents the window used to find the hit shown as a red circle

Evaluation on real images: dSTORM



Sparse colocalization (hits)



Dense map of colocalization scores

dSTORM acquisition of cells from hippocampi of mice expressing BDNF proteins (green) and vGlut (purple) (courtesy of M. Sauer, Julius-Maximilians-University Würzburg, Germany)

Global colocalization score: $T = 4.34$
$p - value = 0.000000703$

Messages to take away

F. Lavancier, T. Pécot, L. Zhengzhen, C. Kervrann, A fast automatic co-localization method for 3D live cell and super-resolution microscopy, 2017 (under review)

GcoPS, controlled by a p-value, tests whether the normalized Pearson correlation between two binary images is significantly positive.

- **Objective** procedure to test co-localization
- **Fast** and **reliable** approach
- Adapted to **any size** and **any shape** of 2D and 3D objects, and to a **different resolution** in the channels
- **Icy plugin** available

- **And even more...**
  - enable to localize co-localization (geo-colocalisation)
  - provides temporal profiles in an image sequence
  - equivalent procedure to test anti-colocalization

We thank J. Salamero (UMR 144 CNRS Institut Curie, Paris, France), and M. Sauer, S. Doose, S. Aufmolk (Julius-Maximilians-University Würzburg, Germany), for assistance with experiments.

THANK YOU !