

# DÉTECTION DE PARTICULES ISOLÉES EN MICROSCOPIE PAR FLUORESCENCE 3D À PARTIR DE PEU D'EXEMPLES

Luc Vedrenne,<sup>1</sup> Denis Fortun,<sup>1</sup> Etienne Baudrier<sup>1</sup>  
<sup>1</sup>ICube, CNRS, Université de Strasbourg

## CONTEXTE ET OBJECTIFS

### Microscopie par fluorescence

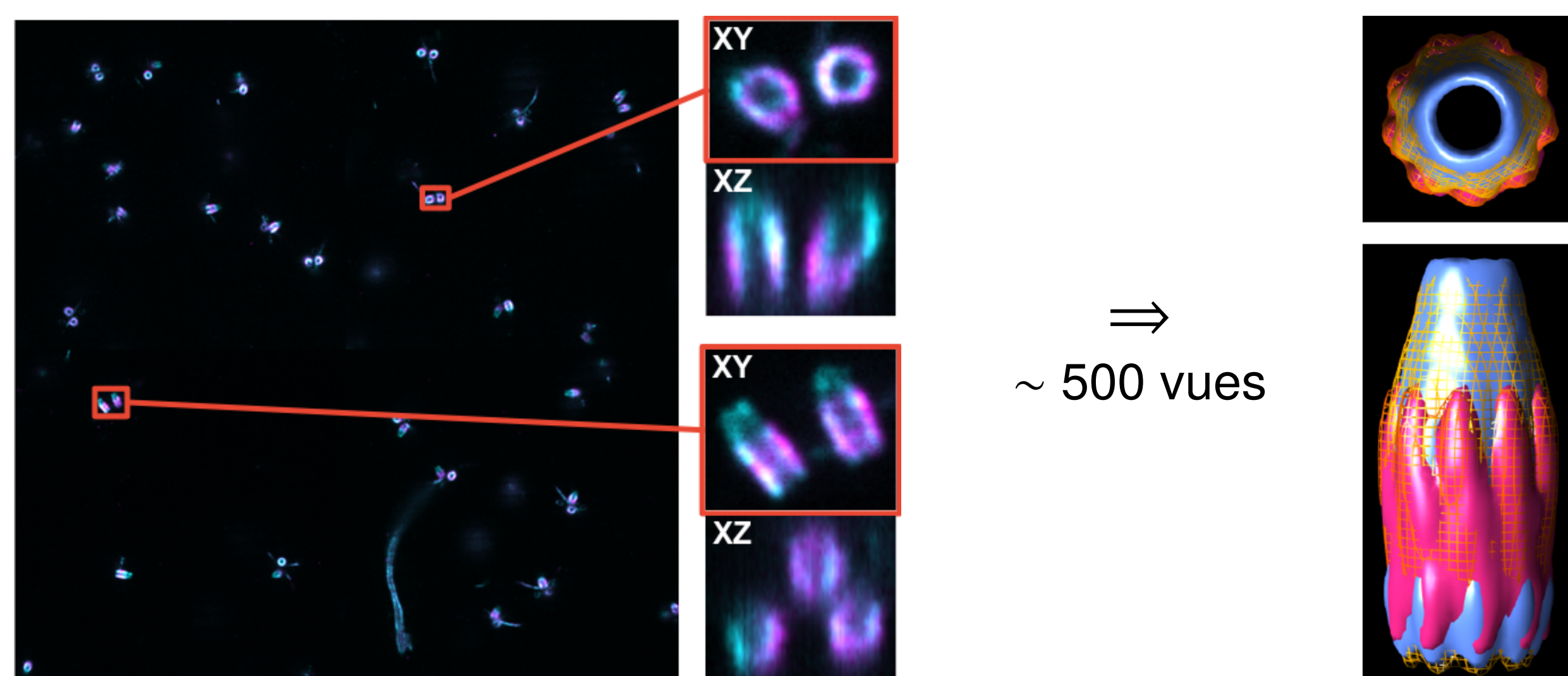
Permet de localiser des protéines spécifiques dans les assemblages macromoléculaires en biologie structurale

### Limites

► Anisotropie de résolution ► Hétérogénéité de l'intensité

### Reconstruction de particule isolée

Reconstruction d'une particule unique à partir de vues multiples aléatoirement orientées (Mahecic et al., 2020)



### Prérequis : détection de toutes les particules sur l'image 3D

Tâche fastidieuse, donc à **automatiser**

## FORMULATION DU PROBLÈME

- Détection par classification binaire dans une fenêtre glissante
- Pas d'*a priori* sur la forme de la particule
  - Entraînement entièrement supervisé impossible
- Approche semi-automatique : détection manuelle d'un petit nombre de particules dans les données
  - **Entraînement faiblement supervisé**

## POSITIVE-UNLABELLED LEARNING 2D (Bepler et al., 2019)

$g$  : estimateur (classifieur)

$\mathcal{P}$  : faible portion de l'image annotée positivement

$\mathcal{U}$  : reste de l'image non labellisée

$\rho$  : proportion d'éléments positifs dans  $\mathcal{U}$ , **supposée connue**

Fonction objectif intégrant cet *a priori* :

$$\mathbb{E}_{\mathbf{x} \sim \mathcal{P}} [L(g(\mathbf{x}), 1)] + \lambda D_{\text{KL}}(\mathbb{E}_{\mathbf{x} \sim \mathcal{U}} [g(\mathbf{x})] \parallel \rho)$$

Descente de gradient stochastique  $\Rightarrow$  estimateur biaisé

Mais : un *batch* de  $\mathcal{U}$  de taille  $n$  avec  $k$  éléments positifs suit une loi binomiale  $p(k)$  de paramètre  $\rho$ . D'où

$$\mathbb{E}_{\mathbf{x} \sim \mathcal{P}} [L(g(x), 1)] + \lambda \sum_{k=1}^n q(k) \log(p(k))$$

où  $q(k)$  est la distribution des prédictions positives de  $g$ .

## LIMITES DU PU-LEARNING

### Passage 2D $\rightarrow$ 3D

$r$  : rayon d'une particule

$k$  : nombre de particules total dans l'image

$D = d \times h \times w$  : nombre de voxels total de l'image

Alors  $\rho = \frac{k}{D} \pi r^3$  et  $\text{batch size}_{\min} = \lceil \rho \rceil$

**Le passage à la 3D impose une taille de batch  $\sim 10^6$**

**Régime très faiblement supervisé** ( $\sim 10$  annotations)

**Les performances dépendent trop du choix des annotations**

## APPORTS

### Extension à la 3D

$\rho = \frac{|\mathcal{P}|}{|\mathcal{U}|} \nearrow \Rightarrow |\mathcal{U}| \searrow$  car  $|\mathcal{P}|$  fixe.

→ Création d'un masque : seuillage par mélange de deux gaussiennes, correction d'histogramme et opérateurs morphologiques

### Régime faiblement supervisé robuste

Manque d'annotations  $\Rightarrow$  moins de généralisation

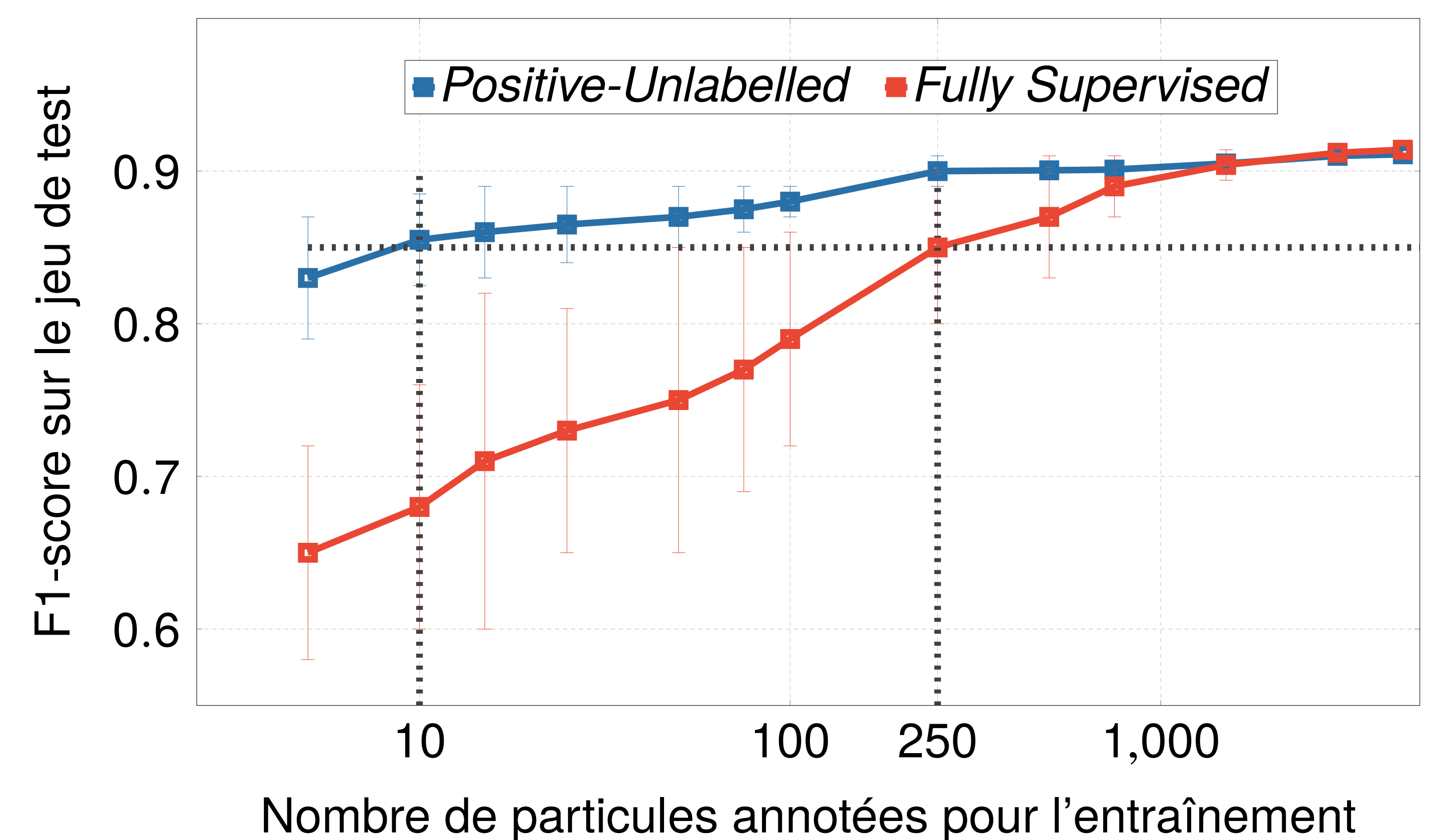
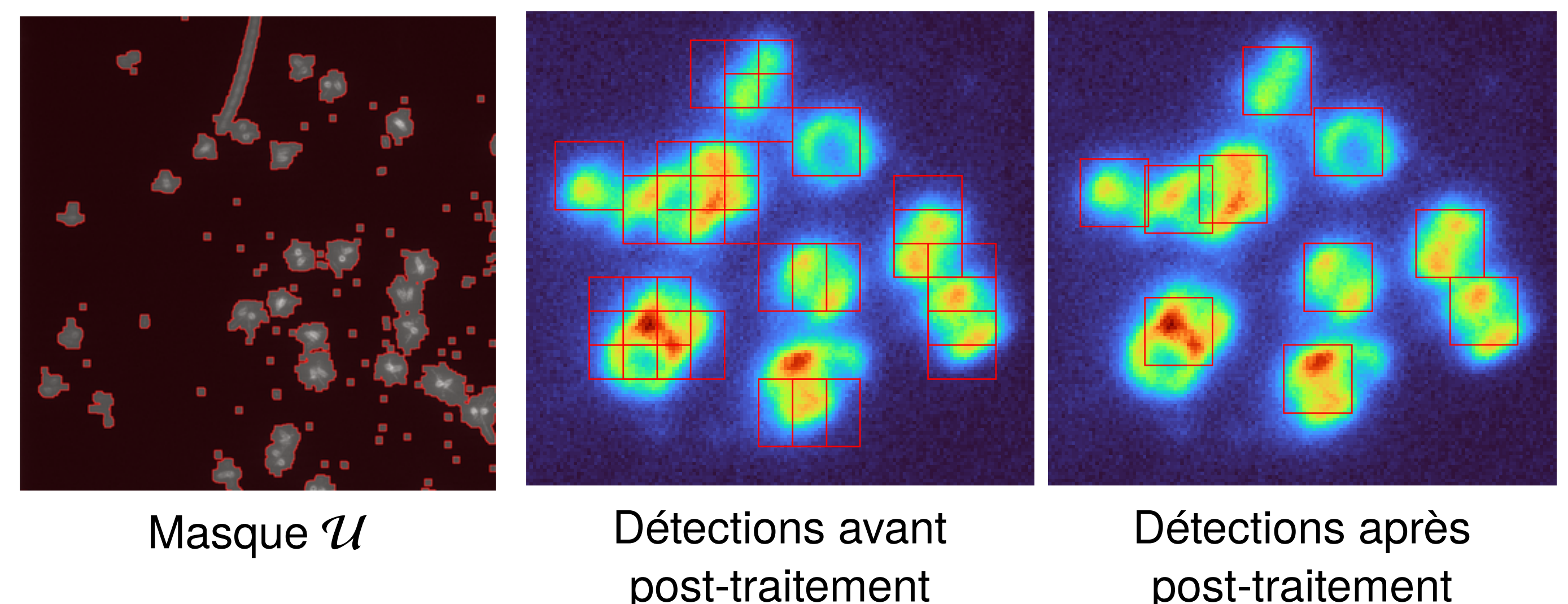
→ *Stochastic Weight Averaging* (Izmailov et al., 2018) : améliore la généralisation sans surcoût computationnel

### Amélioration fine des détections

Classification binaire de *patches*  $\Rightarrow$  trop de faux positifs, prédictions mal centrées

→ Post-traitement itératif basé sur une NMS, une détection de composantes connexes *soft* suivie d'une correction du centre de masse

## RÉSULTATS



Par rapport au *fully-supervised learning*, la pipeline proposée est

- **Plus performante avec peu d'annotations**  
250  $\rightarrow$  10 particules pour atteindre 0.85 de F1-score
- **Plus indépendante du choix des annotations**  
 $\sigma$  0.16  $\rightarrow$  0.06 avec 10 particules annotées

## REFERENCES

- Bepler, Tristan et al. (2019). "Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs". In: *Nature Methods*.
- Izmailov, Pavel et al. (2018). "Averaging Weights Leads to Wider Optima and Better Generalization". In: *Conference on Uncertainty in Artificial Intelligence 2018*.
- Mahecic, Dora et al. (2020). "Homogeneous multifocal excitation for high-throughput super-resolution imaging". In: *Nature Methods*.