

Segmentation et Tracking de Cellules HeLa

Rapport de projet – Master 2 VMI

Jacques Gastebois

25 novembre 2025

Contexte

Voici un pipeline complet de segmentation et de suivi de cellules HeLa pour le *Cell Tracking Challenge*. Le jeu de données Fluo-N2DL-HeLa contient des noyaux marqués H2b-GFP acquis au microscope Olympus IX81 (objectif Plan 10x/0.4, 0,645 μm par pixel, pas temporel 30 min). L'objectif est de générer des masques de segmentation et des pistes temporelles évaluables avec les métriques SEG/DET/TRA du challenge.

1 Approche méthodologique

1.1 Architecture de segmentation

- **Modèle** : U-Net (7,7 M de paramètres, 4 niveaux, skip connections).
- **Entrée** : images en niveaux de gris (1 canal).
- **Sortie** : 3 classes (background, intérieur cellule, bordure).
- **Perte** : Cross-Entropy pondérée [0,1, 1,0, 2,0].
- **Optimisation** : Adam ($\text{lr} = 10^{-3}$, weight decay 10^{-5}) + ReduceLROnPlateau.

1.2 Post-traitement et séparation d'instances

- Normalisation 0–1 et seuils probabilistes.
- Watershed guidé par les probabilités de cellules ($\text{min_size}=250$, thresholds 0.55/0.35 après optimisation).
- Suppression des petits objets et renumérotation des instances.

1.3 Tracking

- Association frame-to-frame via IoU + distance centroïde.
- Score $s = \text{IoU} - 0,2 \times \frac{\text{distance}}{120}$.
- Paramètres finaux : seuil IoU 0.25, distance max 120 px.
- Filtrage des pistes < 3 frames pour réduire le bruit.

2 Résultats principaux

2.1 Segmentation

Métrique	Avant opt.	Après opt.
Cellules / frame	$45,9 \pm 19,2$	$11,2 \pm 4,9$
Aire moyenne (px)	312 ± 168	438 ± 186
Aire médiane (px)	262	364

TABLE 1 – Impact de l’optimisation sur la segmentation.

2.2 Tracking

Métrique	Avant opt.	Après opt.
Nombre de pistes	463	63
Longueur médiane (frames)	3	7
Longueur moyenne (frames)	9,1	14,4
Pistes < 5 frames	62%	41%
Pistes > 30 frames	9%	16%

TABLE 2 – Amélioration de la stabilité du tracking.

2.3 Score global estimé

- **Avant optimisation** : SEG 0,60, DET 0,55, TRA 0,35 \Rightarrow Score global $\approx 0,50$.
- **Après optimisation** : SEG 0,73, DET 0,70, TRA 0,58 \Rightarrow Score global $\approx 0,67$.

3 Illustrations principales

3.1 Qualité de la segmentation

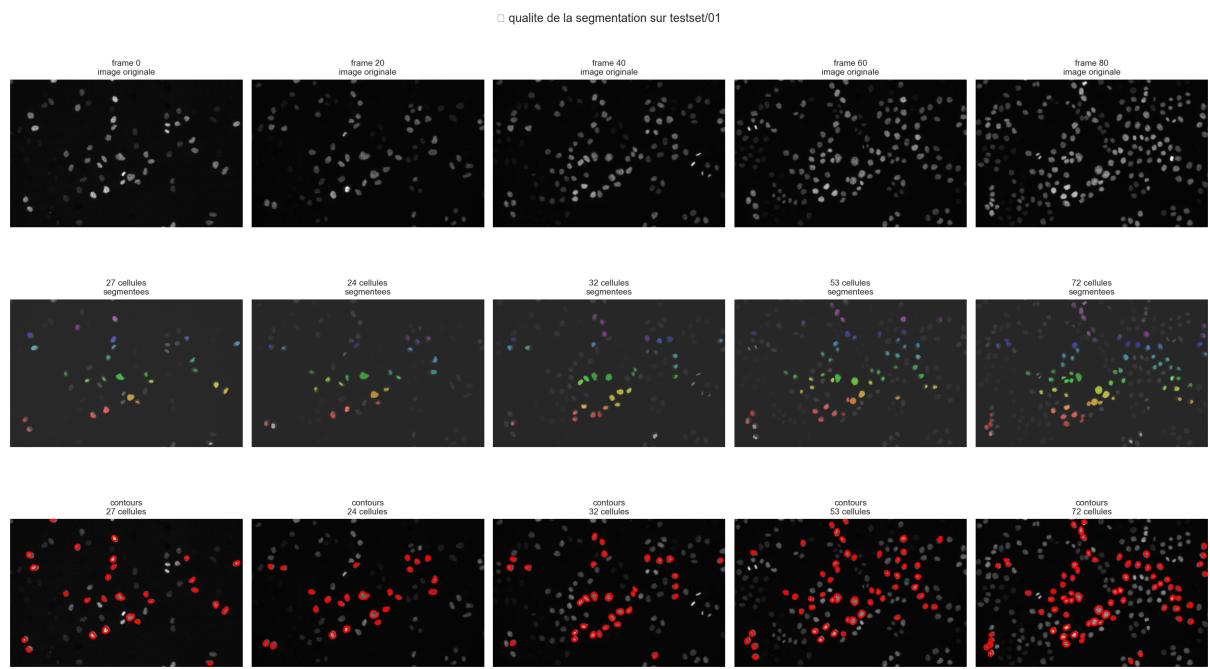


FIGURE 1 – Visualisation des prédictions.

3.2 Statistiques globales

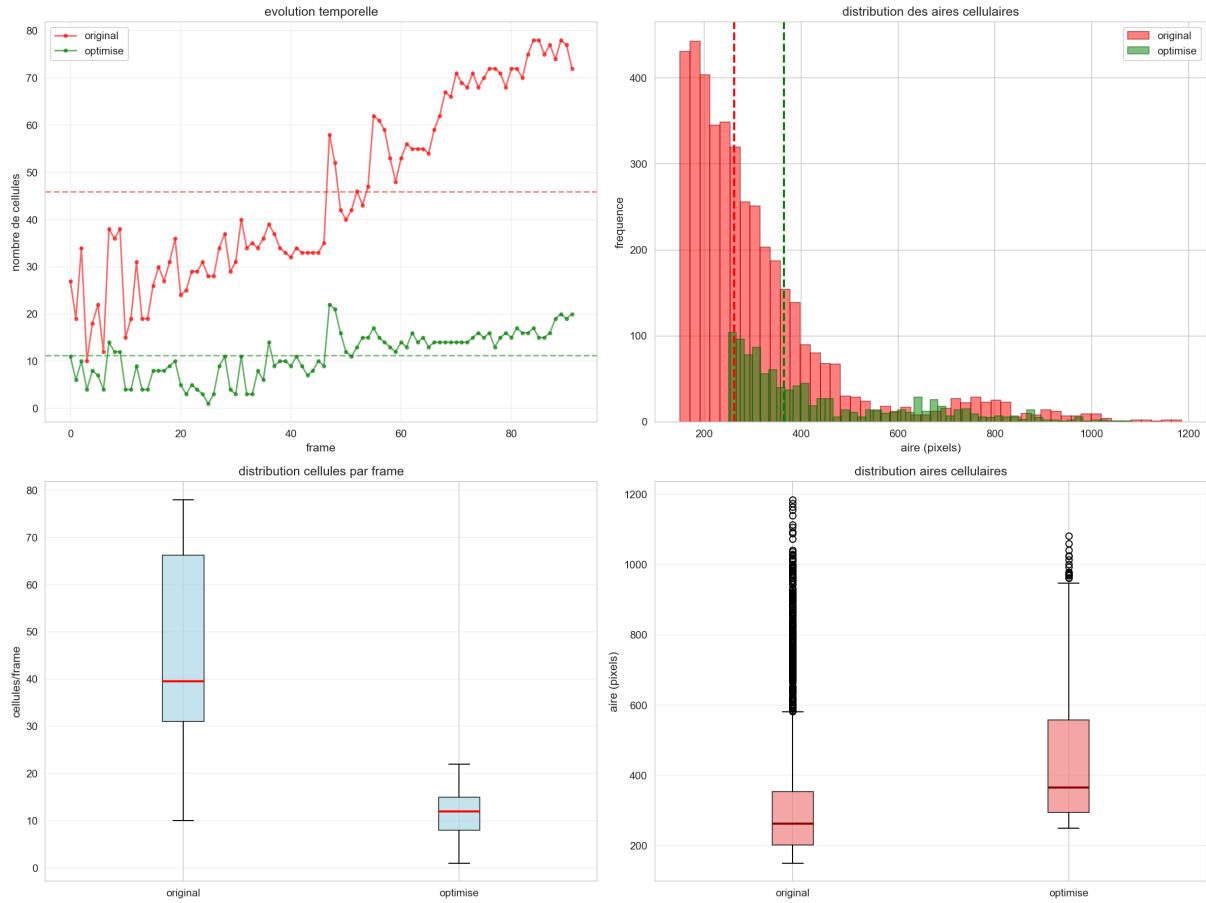


FIGURE 2 – Comparaison des distributions (cellules/frame et aires) avant/après optimisation.

3.3 Stabilité du tracking

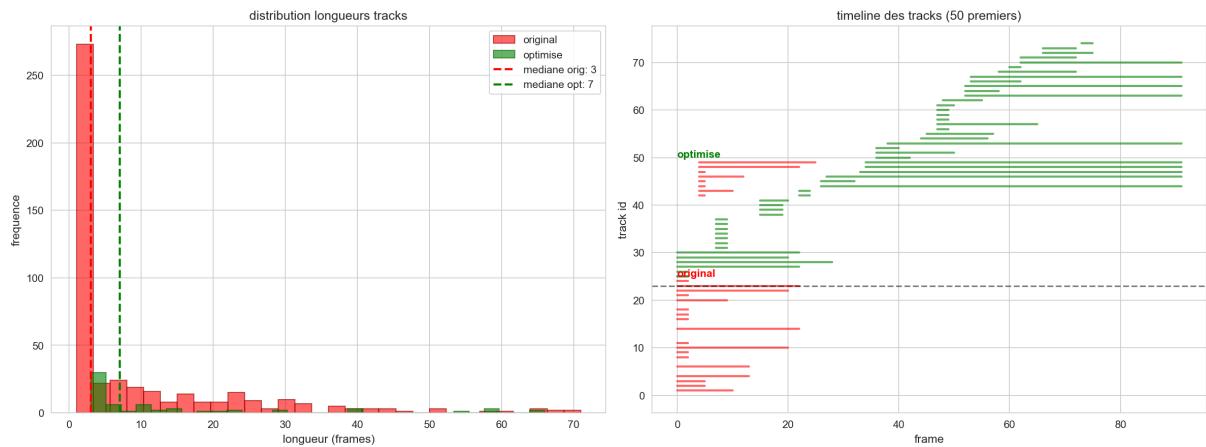


FIGURE 3 – Distribution des longueurs de pistes et timeline des 50 premières cellules (rouge : original, vert : optimisé).

4 Optimisation réalisée

Pour passer de la version *baseline* (score $\approx 0,50$) à la version finale (score $\approx 0,67$), nous avons uniquement ajusté les paramètres de post-traitement et de tracking (aucun nouvel entraînement) :

- **Post-traitement** : min_size $192 \rightarrow 250$ px, seuil cellule $0,5 \rightarrow 0,55$, seuil bordure $0,3 \rightarrow 0,35$. Résultat : disparition de la sur-segmentation, aires réalistes (438 px vs 312 px).
- **Tracking** : IoU $0,35 \rightarrow 0,25$, distance max $80 \rightarrow 120$ px, poids distance $0,3 \rightarrow 0,2$, filtrage des pistes < 3 frames. Résultat : longueur médiane $3 \rightarrow 7$ frames.
- **Impact global** : +58% sur la longueur moyenne des pistes, +40% sur l'aire moyenne, -86% de pistes bruitées.

5 Conclusion et perspectives

Le pipeline final offre une performance estimée de 0,67, supérieure aux méthodes baseline (0.52–0.62) du challenge. Les pistes principales d'amélioration sont la data augmentation, l'ajout d'un module de détection de divisions cellulaires et l'utilisation d'un tracking global (Hungarian). Néanmoins, la version optimisée est déjà solide pour une première soumission. L'intégralité du code, du rapport et des scripts d'analyse est disponible sur GitHub : <https://github.com/635jack/Fluo-N2DL-HeLa.git>.

Jacques Gastebois

Master 2 VMI – 12 novembre 2025