

Sketchpose 2.0 – Entraîner des Modèles de Segmentation d’Instances en Biologie à partir d’Annotations Partielles

Florian SARRON^{1,2} Pierre WEISS^{1,2}

¹IRIT (Institut de Recherche en Informatique de Toulouse)

²CBI (Centre de Biologie Intégrative, Toulouse)

Résumé – La segmentation d’images biologiques présente de nombreux défis. Les microscopes utilisés sont variés et souvent optimisés en fonction des spécificités de l’objet à observer, ce qui entraîne une grande diversité de types d’images et d’objets à analyser. Ainsi, il est souvent nécessaire d’entraîner ou d’affiner les poids des réseaux de neurones pour chaque problème spécifique. Dans ce travail, nous proposons une méthode d’apprentissage permettant d’optimiser les poids d’un modèle à partir d’annotations très partielles, telles que des points, des segments de courbes ou des coups de pinceau appliqués sur l’intérieur des objets à segmenter. Notre approche repose sur l’équation eikonale, qui définit la distance aux frontières des objets. Les résultats préliminaires sont prometteurs, et une implémentation sous forme de plugin Napari permet de tester la solution proposée.

Abstract – Biological image segmentation presents several challenges. Microscopes vary widely and are often optimized for specific objects, leading to a great diversity of image types and structures to analyze. As a result, it is often necessary to train or fine-tune neural network weights for each specific problem. In this work, we propose a training method that allows model weights to be optimized using very partial annotations, such as points, curve segments, or brush strokes applied to the interior of the objects to be segmented. Our approach is strongly based on the eikonal equation, which defines the distance to object boundaries. Preliminary results are promising, and a Napari plugin has been developed to test the proposed solution.

1 Introduction

Le domaine de la segmentation d’images n’échappe pas aux transformations majeures engendrées par l’apprentissage machine. Aujourd’hui, les outils les plus performants pour des tâches complexes d’analyse d’image reposent sur l’utilisation de réseaux de neurones. En biologie – le domaine ciblé par ce papier – les outils CellPose [15, 8, 14], Omnipose [4], StarDist [12] ou d’autres [6] ont essentiellement remplacé les outils traditionnels de segmentation dans les laboratoires de biologie.

Un problème essentiel cependant est le besoin de générer des données labellisées. Par exemple, entraîner un modèle requiert habituellement de générer des centaines de masques de segmentation de cellules ou de noyaux qui permettent de construire des fonctions de régression. Dans un travail récent [1], nous avons proposé une technique d’entraînement de réseaux de neurones 2D, fonctionnant en délimitant seulement quelques bords d’objets ainsi que leur extérieur et intérieur. Une des conséquences est que seuls 25% environ des annotations sont suffisants, pour obtenir des résultats de performance similaire à des modèles entraînés sur données complètes. Une personne mal intentionnée pourrait dire que cet avantage est plutôt léger, car on dispose aujourd’hui de nombreuses bases de données assez riches. Si ce point de vue se défend pour des images 2D, on peut cependant affirmer que la situation est très différente pour des jeux de données 3D : aujourd’hui il n’y a essentiellement pas d’outil clé en main pour ce domaine. La solution que nous avons proposée dans [1] (Sketchpose 1.0) ne résout pas ce dilemme, car elle nécessiterait de tracer des bouts de surfaces des objets à retrouver, ce qui est peu pratique.

Dans ce papier, nous proposons une solution alternative permettant de prendre en compte des annotations très partielles. Un réseau est entraîné à prédire la fonction de distance au

bord des objets, à partir de quelques points ou courbes tracés sur les frontières des objets à retrouver. Nous introduisons des fonctions pertes reposant sur des équations aux dérivées partielles et en particulier l’eikonale, qui nous permettent de prédire la fonction distance au bord des objets, grâce à des réseaux de neurones convolutionnels.

1.1 Idée générale

Nous considérons un problème de segmentation d’instances supervisée. Soit \mathcal{X} le domaine de l’image. Nous voulons le diviser en deux composantes : \mathcal{X}_0 et \mathcal{X}_1 . Pour fixer les idées, \mathcal{X}_0 représente l’arrière-plan et \mathcal{X}_1 représente les objets à segmenter (par exemple, les cellules, les noyaux). Le domaine \mathcal{X}_1 peut être partitionné comme suit

$$\mathcal{X}_1 = \bigsqcup_{n=1}^N \omega_n, \quad (1)$$

où chaque sous-ensemble ω_n désigne un objet à segmenter. Nous supposons que les objets sont simplement connectés et ne peuvent s’intersecter qu’au niveau de leur frontière. Nous définissons l’ensemble des bords des objets comme l’union des frontières des domaines ω_n : $\mathcal{B} = \cup_{n=1}^N \partial\omega_n$.

Soit $\phi : \mathcal{X} \rightarrow \mathbb{R}$ la distance signée à la frontière \mathcal{B} des objets et $v : \mathcal{X} \rightarrow \mathbb{R}^D$ son gradient $v = \nabla\phi$. La fonction v est bien définie, presque partout sous des conditions de régularité appropriées sur la frontière \mathcal{B} .

L’idée d’Omnipose [4] – et de presque tous les meilleurs modèles de segmentation d’instance actuels en imagerie biologique – est de prédire la fonction de distance au bord des objets. Avec nos notations, ceci revient à régresser le champ de vecteurs v et/ou la distance ϕ .

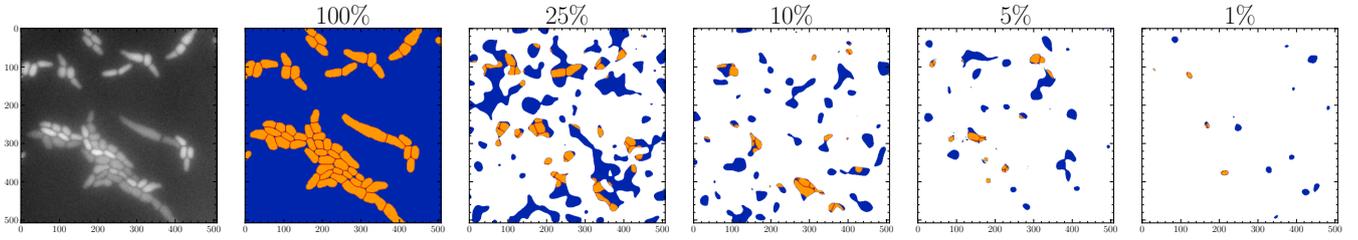


FIGURE 1 : Exemples d’annotations partielles sur une image 512×512 . L’arrière plan \mathcal{S}_0 est représenté en bleu, les objets \mathcal{S}_1 en jaune et les bords $\mathcal{B}_{\text{part}}$ en rouge. Le blanc correspond à une absence d’annotations.

Soit u une image et w les poids d’un réseau. Omnipose entraîne un réseau de neurones convolutionnel de type U-Net [5] $N(u, w)$ qui produit deux composantes $N_\phi(u, w)$ et $N_v(u, w)$. Elles sont construites de sorte à coïncider respectivement avec la fonction de distance ϕ et son gradient v . Dans sa version la plus basique, l’entraînement revient à minimiser le risque empirique suivant :

$$\inf_w \frac{1}{I} \sum_{k=1}^K \|N_v(u_k, w) - v_k\|_2^2 + \frac{1}{I} \sum_{k=1}^K \|N_\phi(u_k, w) - \phi_k\|_2^2 \quad (2)$$

où ϕ_k et v_k sont, respectivement, les vérités terrains de la fonction distance et du champ de vecteur pour $K \in \mathbb{N}$ images totalement annotées. Sketchpose 1.0 [1] étend cette logique aux annotations partielles en restreignant le risque empirique (2) aux pixels où les vérités terrain peuvent être calculées malgré l’annotation partielle.

Le masque de segmentation prédit $\hat{\mathcal{X}}_1$ est simplement obtenu par seuillage $\hat{\mathcal{X}}_1 = \{x \in \mathcal{X}, N_\phi(u_k, w)(x) \geq \epsilon\}$, où $\epsilon \geq 0$ est un seuil sur la fonction distance. Pour obtenir chaque instance dans l’ensemble résultant, le flot $N_v(u_k, w)$ est utilisé pour générer des trajectoires avec une technique d’intégration d’Euler. Les points appartenant à des trajectoires finissant dans une même composante connexe sont alors déclarés comme appartenant à la même instance.

1.2 Gérer des annotations partielles

L’idée principale que nous utiliserons dans ce travail est d’exploiter une caractérisation fonctionnelle de la fonction de distance ϕ à travers l’équation eikonale :

$$\begin{aligned} |\nabla\phi| &= 1 \quad \text{presque partout sur } \mathcal{X} & (\mathcal{P}_{\text{continuous}}) \\ \phi(x) &= 0 \quad \forall x \in \mathcal{B}. \end{aligned}$$

Ce système d’équations est connu pour admettre une infinité de solutions, et celle qui nous intéresse est la *solution de viscosité* [2], qui peut être interprétée comme la solution la moins oscillante. En particulier, les fonctions de distance signées et non signées sont toutes deux des solutions. Les solutions de cette équation sont généralement obtenues grâce à la littérature sur les équations de Hamilton-Jacobi, mais nous adopterons ici une approche différente. Nous présentons les idées principales en 2D, car les notations sont plus simples. L’extension en 3D est directe et nos codes sont pour le moment 2D.

Discrétisation Nous travaillons ainsi sur une grille de pixels. Un pixel est indexé par ses coordonnées entières $(i, j) \in \mathbb{Z}^2$. La valeur d’une image ou d’un tableau ϕ est notée $\phi_{i,j}$ ou $\phi[i, j]$.

Étant donné un ensemble discret de points de bord \mathcal{B} de \mathcal{X} , nous cherchons à calculer la distance signée à cet ensemble en tout point de la grille. En suivant [10, 16, 11], nous proposons de discrétiser l’équation eikonale ($\mathcal{P}_{\text{continuous}}$), en utilisant un schéma de différences amont, qui permet d’obtenir la solution de viscosité. Définissons ϕ_0^{\min} et ϕ_1^{\min} comme suit :

$$\begin{aligned} \phi_0^{\min}[i, j] &\stackrel{\text{def}}{=} \min(\phi[i-1, j], \phi[i+1, j]) \\ \phi_1^{\min}[i, j] &\stackrel{\text{def}}{=} \min(\phi[i, j-1], \phi[i, j+1]). \end{aligned}$$

Nous définissons ensuite le gradient non linéaire amont ∇_g (l’indice $_g$ faisant référence au nom Godunov) comme :

$$\nabla_g \phi = \begin{pmatrix} (\phi - \phi_0^{\min})_+ \\ (\phi - \phi_1^{\min})_+ \end{pmatrix} \quad (3)$$

et

$$|\nabla_g \phi|[i, j] \stackrel{\text{def}}{=} \left(\sqrt{(\phi - \phi_0^{\min})_+^2 + (\phi - \phi_1^{\min})_+^2} \right) [i, j].$$

La fonction de distance discrète ϕ est alors obtenue en utilisant l’équation d’eikonale discrétisée :

$$\begin{aligned} \phi[i, j] &= 0 \quad \text{si } (i, j) \in \mathcal{B} & (\mathcal{P}_{\text{discret}}) \\ |\nabla_g \phi| &= 1 \quad \text{sur } \mathcal{X} \setminus \mathcal{B} \end{aligned}$$

Remarquons que le gradient de Godunov $|\nabla_g \phi|$ s’annule en tout minimum local de ϕ . En particulier, il s’annule sur la frontière, ce qui explique pourquoi la contrainte de norme unitaire n’est valide que sur $\mathcal{X} \setminus \mathcal{B}$.

Formulation variationnelle et annotations partielles Supposons qu’un utilisateur délimite certaines parties de frontière $\mathcal{B}_{\text{part}} \subset \mathcal{B}$ et esquisse l’extérieur \mathcal{X}_0 et l’intérieur \mathcal{X}_1 à l’aide de coups de pinceaux \mathcal{S}_0 et \mathcal{S}_1 . Nous supposons que les ensembles $\mathcal{B}_{\text{part}}$, \mathcal{S}_0 et \mathcal{S}_1 sont disjoints, ce qui est facile à imposer dans une interface graphique de dessin.

Afin d’entraîner un réseau de neurones, nous souhaitons transformer les équations ($\mathcal{P}_{\text{discret}}$) sous forme d’une fonction de perte. Pour cela, soit $\phi = N_\phi(u, w)$ la distance prédite par le réseau de neurones. Une formulation variationnelle naturelle est alors la suivante :

$$\begin{aligned} \inf_w \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} (|\nabla_g \phi|(x) - 1)^2 &+ \frac{1}{|\mathcal{B}_{\text{part}}|} \sum_{x \in \mathcal{B}_{\text{part}}} \phi^2(x) \\ &+ \frac{1}{|\mathcal{S}_0|} \sum_{x \in \mathcal{S}_0} \phi_+^2(x) + \frac{1}{|\mathcal{S}_1|} \sum_{x \in \mathcal{S}_1} \phi_-^2(x), \end{aligned}$$

où ϕ_+ et ϕ_- désignent respectivement les parties positive et négative de ϕ . Le premier terme favorise un gradient de norme

un, le second impose que ϕ s'annule sur \mathcal{B} , et les deux derniers garantissent les conditions $\phi|_{\mathcal{S}_0} \leq 0$ et $\phi|_{\mathcal{S}_1} \geq 0$, conformément aux attentes pour une fonction de distance signée.

Si l'annotation est complète, c'est-à-dire si $\mathcal{X} = \mathcal{B}_{\text{part}} \cup \mathcal{S}_0 \cup \mathcal{S}_1$, on remarque que la fonction de distance signée à \mathcal{B} satisfaisant $(\mathcal{P}_{\text{discret}})$ est un minimiseur global de cette fonctionnelle.

En pratique, nous avons observé que minimiser uniquement la première partie N_ϕ du réseau de neurones converge difficilement. Pour contourner ce problème, nous proposons d'imiter le mécanisme implémenté dans Omnipose et d'entraîner N_ϕ et N_v conjointement. L'intuition sous-jacente est qu'en relâchant la contrainte selon laquelle le champ de vecteurs v est le gradient d'un potentiel, le réseau de neurones dispose de plus de degrés de liberté pour optimiser ses poids et trouver ainsi de meilleurs minimiseurs.

Soit $\phi = N_\phi(u, w)$ et $v = N_v(u, w)$. La fonction de perte complète que nous considérons est la suivante :

$$\begin{aligned}
 E(v, \phi) \stackrel{\text{def}}{=} & \underbrace{\frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} (|v(x)| - 1) + \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} (|\nabla_g \phi(x)| - 1)}_{\text{Champ de vecteurs unitaire}} \\
 & + \underbrace{\frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} (\nabla_g \phi - v)^2}_{\text{Couplage}} + \underbrace{\frac{1}{|\mathcal{B}_{\text{part}}|} \sum_{x \in \mathcal{B}_{\text{part}}} \phi^2(x)}_{\text{0 sur la frontière}} \\
 & + \underbrace{\frac{1}{|\mathcal{S}_0|} \sum_{x \in \mathcal{S}_0} \phi_+(x)}_{\text{Négatif sur } \mathcal{S}_0} + \underbrace{\frac{1}{|\mathcal{S}_1|} \sum_{x \in \mathcal{S}_1} \phi_-(x)}_{\text{Positif sur } \mathcal{S}_1}
 \end{aligned}$$

Il est important de noter ici que la fonctionnelle introduite ne dépend que des annotations partielles $\mathcal{B}_{\text{part}}$, \mathcal{S}_1 , \mathcal{S}_0 . On exploite donc les biais implicites du réseau de neurones pour favoriser l'émergence de solutions régulières coïncidant avec les données sous-jacentes. En toute généralité, les différents termes de la fonction de perte pourraient être pondérés, et les poids optimisés comme des hyperparamètres. Ce point sera abordé dans des travaux ultérieurs.

2 Résultats

2.1 Génération d'annotations partielles

Pour tester les performances de notre méthode, nous avons entraîné le réseau de neurones sur le jeu de données public `bact_fluor` d'Omnipose, qui contient 143 images et plus de 19 000 objets. Ce jeu de données comprend des images de colonies bactériennes obtenues par microscopie à fluorescence en champ large. Les objets à segmenter présentent des morphologies variées, incluant des bactéries très allongées ainsi que d'autres en forme de cacahuète ou de ballon de rugby. Les tailles dans la direction transversale sont relativement faibles, variant entre 5 et 30 pixels, ce qui rend la tâche particulièrement complexe. Ce jeu de données a été conçu pour illustrer les avantages d'Omnipose par rapport à Cellpose 1.0, qui est incapable de segmenter correctement ces formes.

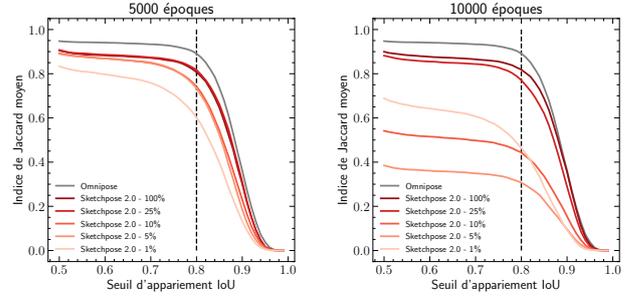


FIGURE 2 : Résultats quantitatifs sur le jeu de données `bact_fluor`. Index de Jaccard moyen en fonction du seuil IoU utilisé pour décider si deux masques coïncident. A gauche : entraînement à 5000 époques. A droite : entraînement à 10000 époque. On peut observer l'effet de surentraînement.

Les annotations complètes disponibles pour ce jeu de données sont des cartes d'instances, où chaque objet est identifié par un numéro unique. Pour nos tests, nous avons utilisé ces annotations complètes afin de générer divers jeux d'annotations partielles. Nous avons produit aléatoirement des masques binaires en seuillant un processus gaussien obtenu (convolution d'un bruit blanc avec un filtre gaussien de variance $\sigma^2 = 100$ pixels), de sorte que seules les zones au-dessus du seuil soient annotées. Nous considérons ici différents pourcentages de pixels annotés : 1%, 5%, 10%, 25% et 100%. Une illustration des annotations partielles utilisées pour une image du jeu de données est présentée en Fig. 1.

2.2 Performances

Pour évaluer quantitativement les performances de la méthode proposée en fonction du pourcentage d'annotations, nous avons calculé l'indice de Jaccard moyen sur l'ensemble du jeu de données d'évaluation (75 images, $\sim 15\,000$ objets), en fonction du seuil de recouvrement utilisé pour considérer que deux masques coïncident. Le seuil de référence est fixé à $\text{IoU} > 80\%$, pour lequel les masques et la vérité terrain deviennent indiscernables pour l'œil expert humain [4].

Après 5000 epochs d'entraînement, l'indice de Jaccard moyen passe de 0.82 pour les annotations complètes à 0.61 pour 1% d'annotations partielles (0.75 pour 5%). Ces performances restent inférieures à celles du modèle public d'Omnipose (0.88) même en annotations complètes. Cependant, contrairement à Omnipose, notre méthode ne régresse pas la distance signée à \mathcal{B} , mais encourage ϕ, v à respecter $(\mathcal{P}_{\text{discret}})$ en minimisant la fonctionnelle $E(v, \phi)$. Or, bien qu'elle minimise globalement $E(v, \phi)$, la distance signée à \mathcal{B} n'en est pas l'unique minimiseur. On peut donc supposer qu'avec des annotations complètes, notre stratégie d'entraînement est moins efficace pour prédire la distance signée à \mathcal{B} que la régression.

Par ailleurs, nous observons que notre méthode sur-apprend lorsqu'elle est entraînée avec des annotations très partielles : les résultats pour 5000 époques sont meilleurs que pour 10000 époques. Ces résultats sont illustrés en Fig. 2.2. Une comparaison qualitative entre la segmentation prédite, la vérité terrain et le modèle Omnipose, pour différents pourcentages d'annotations sur une image du jeu d'évaluation, est présentée en Fig. 2.2.

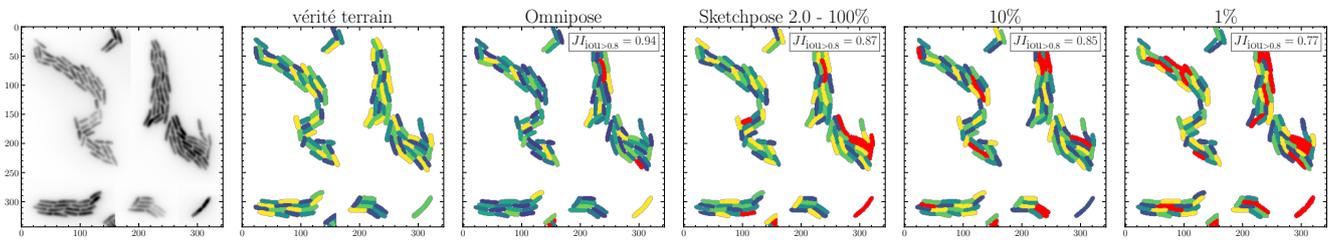


FIGURE 3 : Résultat de segmentation sur une image test, après l’entraînement proposé sur des données avec différents pourcentages d’annotations. Les résultats sont comparés à la vérité terrain et aux prédictions d’Omnipose. Les cellules retrouvées avec un IoU < 80% sont colorées en rouges. L’indice de Jaccard calculé à IoU > 80% sur cette image test est indiqué dans les légendes.

2.3 Conclusion

Nous avons introduit un principe original reposant sur des discrétisations d’équations aux dérivées partielles, permettant d’utiliser des annotations partielles pour entraîner un réseau de neurones convolutionnel de segmentation. Cette technique permet d’obtenir des réseaux de segmentation de bonne qualité avec seulement 5-10% de pixels annotés (contre 25% pour Sketchpose 1.0), même si nous n’atteignons pas encore les performances des meilleures méthodes existantes telles que Omnipose (Index de Jaccard $\approx 85\%$ contre $\approx 90\%$). Nous travaillons actuellement à l’amélioration des performances avec d’autres architectures de réseau, à l’extension au cas multi-classe, ainsi qu’au portage de cette technique en 3D, en utilisant des annotations partielles en 2D dans différents plans.

3 Bibliographie

Références

- [1] Clément CAZORLA, Nathanaël MUNIER, Renaud MORIN et Pierre WEISS : Sketchpose : Learning to segment cells with partial annotations. 2023.
- [2] Michael G CRANDALL et Pierre-Louis LIONS : Viscosity solutions of hamilton-jacobi equations. *Transactions of the American mathematical society*, 277(1):1–42, 1983.
- [3] Keenan CRANE, Clarisse WEISCHEDEL et Max WARDZKY : Geodesics in heat : A new approach to computing distance based on heat flow. *ACM Transactions on Graphics (TOG)*, 32(5):1–11, 2013.
- [4] Kevin J CUTLER, Carsen STRINGER, Teresa W LO, Luca RAPPEZ, Nicholas STROUSTRUP, S BROOK PETERSON, Paul A WIGGINS et Joseph D MOUGOUS : Omnipose : a high-precision morphology-independent solution for bacterial cell segmentation. *Nature methods*, 19(11):1438–1448, 2022.
- [5] Thorsten FALK, Dominic MAI, Robert BENSCH, Özgün ÇIÇEK, Ahmed ABDULKADIR, Yassine MARRAKCHI, Anton BÖHM, Jan DEUBNER, Zoe JÄCKEL, Katharina SEIWALD *et al.* : U-net : deep learning for cell counting, detection, and morphometry. *Nature methods*, 16(1):67–70, 2019.
- [6] Peter NAYLOR, Marick LAÉ, Fabien REYAL et Thomas WALTER : Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE transactions on medical imaging*, 38(2):448–459, 2018.
- [7] Stanley OSHER et James A SETHIAN : Fronts propagating with curvature-dependent speed : Algorithms based on hamilton-jacobi formulations. *Journal of computational physics*, 79(1):12–49, 1988.
- [8] Marius PACHITARIU et Carsen STRINGER : Cellpose 2.0 : how to train your own model. *Nature methods*, 19(12):1634–1641, 2022.
- [9] Ramya RANGAN, Ryan FEATHERS, Sagar KHAVNEKAR, Adam LERER, Jake D. JOHNSTON, Ron KELLEY, Martin OBR, Abhay KOTECHA et Ellen D. ZHONG : Cryodrgnet : deep reconstructing generative networks for visualizing dynamic biomolecules inside cells. *Nature Methods*, 21(8):1537–1545, juillet 2024.
- [10] Elisabeth ROUY et Agnes TOURIN : A viscosity solutions approach to shape-from-shading. *SIAM Journal on Numerical Analysis*, 29(3):867–884, 1992.
- [11] Giovanni RUSSO et Peter SMERKA : A remark on computing distance functions. *Journal of computational physics*, 163(1):51–67, 2000.
- [12] Uwe SCHMIDT, Martin WEIGERT, Coleman BROADDUS et Gene MYERS : Cell detection with star-convex polygons. In *Medical image computing and computer assisted intervention—MICCAI 2018 : 21st international conference, Granada, Spain, September 16-20, 2018, proceedings, part II 11*, pages 265–273. Springer, 2018.
- [13] James A SETHIAN *et al.* : *Level set methods and fast marching methods*, volume 98. Cambridge Cambridge UP, 1999.
- [14] Carsen STRINGER et Marius PACHITARIU : Cellpose3 : one-click image restoration for improved cellular segmentation. *Nature Methods*, 22(3):592–599, février 2025.
- [15] Carsen STRINGER, Tim WANG, Michalis MICHAELOS et Marius PACHITARIU : Cellpose : a generalist algorithm for cellular segmentation. *Nature methods*, 18(1):100–106, 2021.
- [16] Hongkai ZHAO : A fast sweeping method for eikonal equations. *Mathematics of computation*, 74(250):603–627, 2005.