

Annotation semi-automatique de base de données d'images complexes non standardisées

Guillaume PICAUD^{1,3} Marc CHAUMONT^{1,2} Gérard SUBSOL¹ Luc TEOT³

¹Equipe ICAR, LIRMM, Univ. Montpellier, CNRS, Montpellier, France

²Univ. Nîmes France

³Cicat-Occitanie, Montpellier, France

{ guillaume.picaud, marc.chaumont, gerard.subsol}@lirmm.fr
l-teot@chu-montpellier.fr

Résumé

Les plaies chroniques représentent un enjeu sanitaire mondial affectant lourdement le quotidien des patients et représentant un coût important pour les systèmes de santé. La prise en charge des plaies chroniques repose en partie sur une analyse visuelle, ce qui motive l'utilisation d'algorithmes d'analyse automatique. Des bases de données d'images spécialement dédiées à l'entraînement d'algorithmes Deep Learning ont été assemblées en respectant des conditions d'acquisitions rigoureuses. Cependant, une telle standardisation des images ne reflète pas le quotidien des soignants car la plupart des soins sont réalisés au domicile du patient et les infirmiers n'ont que leur smartphone pour prendre des photos. Le réseau Cicat-Occitanie est un réseau de conseil destiné à aider les équipes soignantes dans la prise en charge des patients porteurs de plaies complexes. Avec le temps, le réseau a constitué une base de données de plus 130 000 images de plaies chroniques. Bien que ces images soient décrites d'un point de vue médical, aucune annotation par boîte englobante n'est disponible. Or, l'annotation représente une charge de travail importante. Cet article propose une méthode d'annotation semi-automatique d'une base de données non labellisée via l'entraînement itératif d'un algorithme de détection. Elle vise à réduire la charge de travail de l'annotateur.

Mots clefs

Plaies chroniques, Deep Learning, détection d'objet, annotation semi-automatique

1 Introduction

Selon la Haute Autorité de Santé, une plaie est considérée comme chronique après 4 à 6 semaines d'évolution. De multiples facteurs peuvent favoriser leurs apparitions au sein de populations à risque comme les personnes âgées, les diabétiques ainsi que les personnes à mobilité réduite. Les plaies chroniques représentent un problème majeur de

santé publique. Elles présentent des complications comme l'amputation voire le décès. Elles ont aussi un coût pour la société. L'assurance maladie a estimé à plus d'un milliard d'euros la seule gestion des escarres et ulcères à domicile pour l'année 2011. Leur prévalence est en hausse [1] du fait de la croissance démographique mondiale ainsi que du vieillissement des populations occidentales.

La richesse de la littérature dédiée aux plaies chroniques témoigne de l'intérêt croissant des chercheurs, notamment dans le domaine du Deep Learning [2]. Quelques bases de données sont aujourd'hui accessibles comme le DFUC2020 [3] sur la détection, DFUC2021 [4] sur la classification, HealTech [5] également utilisable sur le thème de la classification ou encore FUseg [6] concernant la segmentation. Ici, les conditions d'acquisition sont standardisées, facilitant la convergence des algorithmes de Deep Learning. Pour répondre au besoin des équipes soignantes, il existe donc un réel besoin de concevoir des bases de données médicales annotées, moins standardisées et de proposer des IA plus robustes face aux variations des conditions d'acquisition.

Le « Réseau Cicat-Occitanie »¹ déploie sur l'ensemble de la région éponyme une aide destinée aux équipes de premier recours (médecins et infirmiers) dans le but d'améliorer la prise en charge des patients. Cette aide porte notamment sur l'analyse des plaies chroniques par des experts à distance au cours d'une téléconsultation [7]. Ainsi, près de 19 000 patients ont été suivis par le réseau, constituant une base de données structurée et significative contenant plus de 133 000 images rattachées aux dossiers médicaux archivés. La diversité des équipes et du matériel, le manque de temps ainsi que les conditions au domicile du patient rendent difficile l'acquisition d'images standardisées. La grande variabilité des images du point de l'éclairage, la distance, l'angle, le zoom, le flou, etc... complexifie leur analyse automatisée. Des exemples d'images issues de la base de données Cicat-Occitanie sont consultables à la figure 1.

1. <https://www.cicat-occitanie.org/a-propos-1>

En l'état, la base d'image du Cicat-Occitanie ne possède aucune annotation par boîte englobante. L'annotation de base de données par un opérateur coûte cher en temps et en ressources. Une automatisation de ce travail par des algorithmes de Deep Learning serait utile mais ne garantit pas l'absence d'erreurs d'annotation. Nous trouvons dans la littérature des travaux dédiés à la réduction du temps de travail de l'annotateur. Une solution étudiée dans [8] et [9] est fondée sur l'Active Learning afin que l'opérateur humain n'ait qu'une fraction de la base de données à annoter. Ces deux travaux ainsi que [10] ont aussi démontré l'efficacité d'une interface ergonomique dédiée à l'annotation afin de réduire le temps de travail de l'opérateur. Dans cet article, nous proposons une méthode d'annotation semi-automatique, incrémentale, dont le but est de réduire le temps de travail de l'annotateur tout en garantissant la qualité des annotations. Cette méthode permet de vérifier manuellement l'ensemble de la base de données en s'appuyant sur des outils open source faciles à utiliser.



FIGURE 1 – Exemples d'images issues de la base de données Cicat-Occitanie. Certaines images sont de bonne qualité tandis que certaines sont plus difficiles à analyser par le flou ou l'occultation de la plaie.

2 Annotation supervisée incrémentale

Dans cette section, nous présentons notre méthode incrémentale pour annoter la base de données Cicat-Occitanie de manière semi-automatique à l'aide de boîtes englobantes. Nous proposons tout d'abord une vue d'ensemble du protocole d'annotation avant de présenter en détail les différentes étapes.

2.1 Vue d'ensemble

La figure 2 présente le processus itératif global d'annotation semi-automatique d'une base de données B par boîtes englobantes. Il consiste à entraîner incrémentalement un algorithme de détection R_i à partir d'une sous-partie de B

correctement annotée B_i^* . On obtient une version R_{i+1} . On peut alors réaliser une inférence sur le lot L_i d'images non annotées issu de B . Les prédictions L_i' sont alors vérifiées, validées ou corrigées par l'annotateur. Le lot corrigé L_i^* est intégré dans une nouvelle version de la base de données d'entraînement B_{i+1}^* .

L'initialisation du processus nécessite de choisir une base de données de référence B_0^* différente de B et qui ne sera pas utilisée par la suite pour la constitution des B_i pour tout $i > 0$. Il faut également choisir un algorithme de détection pré-entraîné R_0 puis l'entraîner sur B_0^* donnant ainsi la version R_1 de l'algorithme. Nous l'utilisons alors pour prédire les boîtes englobantes d'un lot d'images L_0 , issu de la base de données à annoter. Ce lot d'images désormais doté de boîtes englobantes à vérifier se nomme L_0' . Les prédictions sont acceptées, corrigées ou supprimées par l'annotateur et le lot d'images corrigé se nomme alors L_0^* .

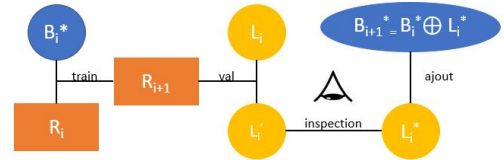


FIGURE 2 – Schéma du processus itératif d'annotation semi-automatique. Nous entraînons le réseau R_i sur la base de données d'entraînement annotée B_i^* . Nous utilisons R_{i+1} , le résultat de l'entraînement, sur le lot d'images L_i afin d'obtenir les prédictions L_i' . L'opérateur vérifie, corrige ou supprime produisant ainsi L_i^* qui sera utilisé pour construire B_{i+1}^* .

2.2 Présentation et préparation des bases de données

Le Diabetic Foot Ulcer Challenge est une compétition annuelle² pour la promotion et le développement d'algorithmes capables d'analyser des photos de plaies du pied diabétique. En 2020, le thème de la compétition était la détection par boîtes englobantes [3]. La base de données mise à disposition jouit de conditions d'acquisition rigoureuses : un nombre restreint d'experts a utilisé un nombre restreint d'appareils photos dans un nombre restreint de salles de consultations garantissant ainsi l'homogénéité de la base de données. 2000 images annotées composent la base d'entraînement et seront utilisées dans l'expérience en tant que B_0^* .

La base de données du Cicat-Occitanie rassemble elle plus de 133 000 images de plaies chroniques de tout type (escarre, ulcère, plaie du pied diabétique, etc...). Un premier filtrage de la base de données est réalisé afin de standardiser les images aux dimensions 450x600 en excluant celles dont le ratio d'origine est trop éloigné de la valeur 1,33. B est composé de 9000 images extraites de la base de données Cicat-Occitanie par tirage aléatoire sans remise. Nous

2. <https://dfu-challenge.github.io/>

l'avons divisé en 3 lots de tailles identiques, L_0 , L_1 et L_2 . En plus, un dernier lot de 538 images, L_{val} , a été conçu. Il est annoté manuellement à l'aide du logiciel opensource LabelImg³.

3 Expériences

3.1 Implémentation

Comme algorithme de détection, nous avons choisi Yolov5 développé par Ultralytics [11]. Ce choix est motivé par sa simplicité de prise en main et par sa capacité à détecter des objets de petites comme de grandes tailles. Nous disposons pour nos entraînements d'une carte graphique GeForce GTX TITAN X 12Go ce qui nous permet d'utiliser la version Yolov5x à 86 millions de paramètres. R_0 correspond à Yolov5x pré-entraîné sur la base de données ImageNet avec le jeu d'hyperparamètres proposé par Ultralytics pour l'entraînement sur la base de données VOC. Tous les entraînements sont réalisés avec des mini-batches de 8 images durant 60 epochs. Nous initialisons la base de données B_0^* avec celle du DFUC2020.

3.2 Métriques d'évaluation

L'algorithme de détection propose quatre coordonnées servant à localiser la boîte englobante ainsi qu'une valeur appelée score de confiance qui quantifie la confiance de l'algorithme dans la prédiction de chaque boîte.

Les vrais positifs T_P , les faux positifs F_P ainsi que les faux négatifs F_N sont déterminés par rapport à une valeur seuil du calcul de l'Intersection over Union, dont l'équation est établie ci-dessous, entre les boîtes englobantes issues de la prédiction de l'algorithme et celle correspondant à la vérité terrain. Si l'IoU entre la prédiction et la vérité terrain dépasse le seuil de 0.5 alors la prédiction sera comptée comme T_P , sinon elle sera comptée comme F_P . Si la boîte englobante vérité terrain d'une plaie n'est associée à aucune prédiction T_P alors nous compterons un faux négatif F_N . Nous sommes ainsi en mesure de calculer les métriques Précision et Rappel, dont les abréviations sont respectivement P et R, à partir de toutes les boîtes englobantes prédites.

$$IoU = \frac{Vt \cap Pred}{Vt \cup Pred} \quad P = \frac{T_P}{T_P + F_P} \quad R = \frac{T_P}{T_P + F_N}$$

La Précision moyenne à 50 ou AP_{50} est mesurée en fixant le seuil d'IoU à 50% puis en faisant varier le seuil du score de confiance. Ainsi, nous obtenons pour chaque seuil un couple (Précision, Rappel) nous permettant de construire la courbe précision sur rappel. AP_{50} correspond à la mesure de l'aire sous cette courbe.

La Précision moyenne entre 0.5 et 0.95 ou $AP_{[0.5:0.95]}$ est mesurée en fixant cette fois-ci le seuil du score de confiance à 0.01 puis en faisant varier le seuil d'IoU entre 0.5 et 0.95 par pas de 0.05. Nous pouvons alors construire la courbe précision sur rappel et mesurer son aire.

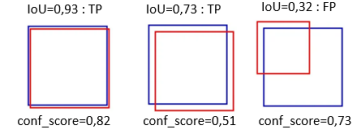


FIGURE 3 – Illustration du calcul de l'IoU et la prédiction ainsi que du score de confiance prédit par l'algorithme.

3.3 Performances du processus d'annotation semi-automatique

La version R_1 a été évaluée sur les 2000 images composant la base de données test du DFUC2020 donc distincte de B_0^* . On obtient des performances de 0.64 en mAP et de 0.69 en $F1_{score}$. Le tableau 1 présente les performances des versions de l'algorithme, au fil des itérations, évaluées sur L_{val} .

B_i^*	P	R	AP_{50}	$AP_{0.5:0.95}$
$B_0^*=DFUC2020$	0.93	0.84	0.88	0.62
$B_1^*=L_1^* (2377)$	0.92	0.88	0.94	0.66
$B_2^*=B_1^* \oplus L_2^* (4608)$	0.92	0.89	0.94	0.66
$B_3^*=B_2^* \oplus L_3^* (6894)$	0.91	0.88	0.94	0.68

TABEAU 1 – Evolution des performances des versions de l'algorithme de détection sur le lot de validation L_{val} en fonction de la base de données d'entraînement B_i^* .

Le tableau 2 présente l'évaluation des performances de l'inférence L_i' du détecteur R_{i+1} par rapport à l'ensemble d'images acceptées ou corrigées par l'annotateur L_i^* .

L_i^*	P	R	AP_{50}	$AP_{0.5:0.95}$
DFUC2020	0.94	0.86	0.93	0.86
$L_1^* (2377)$	0.94	0.93	0.97	0.92
$L_2^* (2231)$	0.94	0.94	0.97	0.93
$L_3^* (2286)$	0.94	0.93	0.98	0.93

TABEAU 2 – Performances des prédictions proposées par l'algorithme de détection R_{i+1} sur les lots L_i au cours des itérations.

Le tableau 3 illustre l'impact de l'algorithme de détection semi-automatique sur le temps de travail de l'annotateur afin d'inspecter un lot.

R_i	Lot	t_{total}	$t_{moyen}/image$
aucun	$L_{val} (538)$	49m	5.5s
R_1	$L_1' (3000)$	2h05m	2.5s
R_2	$L_2' (3000)$	1h44m	2.1s
R_2	$L_3' (3000)$	1h33m	1.9s

TABEAU 3 – Evaluation du temps de travail nécessaire à l'annotateur pour vérifier un lot en fonction de la version de l'algorithme réalisant la pré-annotation.

3. <https://github.com/tzutalin/labelImg>

3.4 Discussion

A l'aide du tableau 1, nous remarquons que chaque itération semble conserver voire améliorer légèrement les prédictions de l'algorithme sur L_{val} .

Le tableau 2 retranscrit le comportement de l'annotateur face aux propositions de l'algorithme de détection. La précision semble être maintenue au cours des itérations. Le nombre de F_P ne semble pas se réduire malgré l'augmentation itérative de la base de données d'entraînement B_i^* . Cependant, le rappel augmente ce qui signifie que l'apprentissage itératif permet de réduire le nombre de F_N . On remarque dans la figure 4 que l'algorithme réalise des erreurs comme la détection de certains motifs dans l'arrière-plan ou la confusion de zones anatomiques.

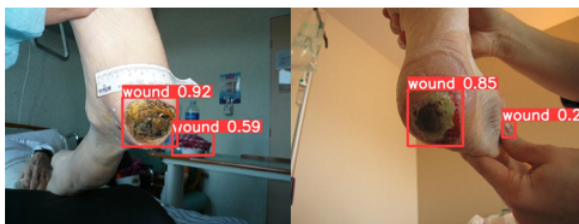


FIGURE 4 – Confusion de détection de plaies avec un motif en arrière-plan ou une autre zone anatomique.

Le tableau 3 permet de mesurer l'impact sur le temps de travail de l'annotateur. Dès la première itération, ce temps est divisé par 2 par rapport à une annotation purement manuelle. Cela montre que la plupart des prédictions de l'algorithme sont acceptées par l'annotateur. De plus, la valeur $t_{moyen}/image$ continue de diminuer légèrement au cours des itérations. Pour autant, le temps de lecture de l'image par l'annotateur reste incompressible.

4 Conclusion et perspectives

Dans cet article, nous avons proposé une méthode incrémentale d'annotation semi-automatique par boîte englobante permettant de diviser par 2 le temps de travail d'un annotateur. Cette méthode a été mise en application sur une base de données d'images complexes n'ayant pas d'acquisition standardisée. L'initialisation de la base de données et de l'algorithme peuvent être réalisées à partir de ressources publiques. Les itérations permettent ensuite d'adapter plus finement l'algorithme à la base de données cible.

La compétition DFUC2020⁴ a montré que certaines modifications architecturales permettent de dépasser les performances de YOLOv5 comme notamment la convolution déformable [12, 13] dont l'application dans l'architecture Faster R-CNN a atteint la première place de la compétition. Ainsi, il serait intéressant d'améliorer le processus d'annotation du Cicat-Occitanie en appliquant ce processus itératif sur de nouvelles architectures de détection.

4. <https://github.com/ryohachiuma/DFU-challenge>

Remerciements

Nous souhaitons remercier l'Association Nationale de la Recherche et de la Technologie ainsi que le réseau Cicat-Occitanie pour financer et soutenir la thèse au travers d'une Convention industrielle de formation par la recherche.

Références

- [1] L. Martinengo et al. Prevalence of Chronic Wounds in The General Population : Systematic Review and Meta-Analysis of Observational Studies. *Annals of epidemiology*, (29) :8–15, Janvier 2019.
- [2] R. Zhang et al. A Survey of Wound Image Analysis Using Deep Learning : Classification, Detection, and Segmentation. *IEEE Access*, 10 : 79502–79515, 2022.
- [3] B. Cassidy et al. The DFUC 2020 Dataset : Analysis Towards Diabetic Foot Ulcer Detection. *touchREV Endocrinol*, (17) :5–11, 2021 .
- [4] M. Hoon Yap et al. Analysis Towards Classification of Infection and Ischaemia of Diabetic Foot Ulcers. Dans *IEEE EMBS International Conference on Biomedical and Health Informatics BHI*, Juillet 2021.
- [5] O. Subba Reddy et al. HealTech-A System for Predicting Patient Hospitalization Risk and Wound Progression in Old Patients. Dans *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2463–2472, 2021.
- [6] C. Wang et al. FUSeg : The Foot Ulcer Segmentation Challenge. Dans *arXiv* , 2022.
- [7] L. Téot et al. Complex Wound Healing Outcomes for Outpatients Receiving Care via Telemedicine, Home Health, or Wound Clinic : A Randomized Controlled Trial. *The International Journal of Lower Extremity Wounds*, (19) :197–204, Juin 2020.
- [8] K. Gokalp Ince et al. Semi-Automatic Annotation For Visual Object Tracking. Dans *CVF International Conference on Computer Vision Workshops (ICCVW)*. IEEE, oct 2021.
- [9] D. P. Papadopoulos et al. We don't need no bounding-boxes : Training object class detectors using only human verification, 2017.
- [10] A. M. Obeso. Image annotation for Mexican buildings database. Dans *Optics and Photonics for Information Processing X*, volume 9970, pages 201–208. SPIE, 2016.
- [11] G. Jocher et al. YOLOv5 by Ultralytics, 5 2020, <https://github.com/ultralytics/yolov5>.
- [12] D. Jifeng et al. Deformable convolutional networks. Dans *Proceedings of the IEEE International Conference On Computer Vision*, pages 764–773, 2017.
- [13] X. Zhu et al. Deformable ConvNets V2 : More Deformable, Better Results. Dans *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9300–9308, 2019.