Un outil de segmentation d'images au service de la restauration de fonds de livres anciens

Michel JORDAN[1] Camille SIMON CHANE[1] Merouane SENNOUN[1] Valérie LEE-GOUET[1,2]

[1]ETIS, UMR8051, CY Cergy Paris Université, ENSEA, CNRS, avenue du Ponceau, 95000 Cergy, France

[2]AGORA, CY Cergy Paris Université, boulevard du Port, 95000 Cergy, France

Résumé – Nous présentons dans cet article les apports de méthodes de segmentation basées *deep-learning* en vue de l'amélioration et de l'extension d'un ensemble de traitements pour la détection et l'identification d'altérations dangereuses sur des dos de reliures de livres patrimoniaux. Les améliorations apportées permettent d'étendre un pipeline pré-existant dans deux directions : d'une part la précision des détections d'altérations, d'autre part le traitement de fonds d'ouvrages avec des reliures plus variées.

Abstract – In this paper, we show how deep-learning based image segmentation methods can help to better detect alterations on old book bindings. This helps to extend previous versions of a specialized pipeline in two different directions: a better accuracy of alteration detection on one side, and on a second side, the possibility for processing various collections of old books.

1 Introduction

Dans tous les domaines, la conservation des objets patrimoniaux débute par une étape essentielle : le constat d'état. Dans le cas des livres anciens, cette opération représente un véritable défi pour les bibliothèques et les archives dont les fonds dépassent souvent plusieurs milliers de volumes. Par manque de temps ou de personnel qualifié, ces institutions peinent à disposer d'une vision d'ensemble de l'état de leurs livres patrimoniaux. Nous pensons que l'intelligence artificielle pourrait automatiser la tâche fastidieuse de détection des altérations des reliures et ainsi aider à hiérarchiser les interventions de conservation, allant d'une mise en boîte à une restauration complète, de manière rapide et économique.

Dans le cadre de la thèse de doctorat de Valérie Lee-Gouet, nous avons conçu un *pipeline* basé sur des photographies de livres patrimoniaux conservés sur leurs rayonnages et prises à l'aide d'un smartphone. Ce choix de prise de vue permet à n'importe quel membre du personnel de constituer rapidement un corpus d'images exploitables. À partir de ces images, le pipeline détecte automatiquement sept types d'altérations dangereuses observés sur les reliures et signalés dans les constats d'état écrits manuellement [3] : de haut en bas et de gauche à droite sur la figure 1, coiffe incomplète, sangle, mors fendu, cahiers visibles, lacune de cuir, tranchefile manquante, ainsi que livre dans une boîte (non représenté sur la figure).

Afin de détecter les altérations à partir des photographies des rayonnages, nous avons développé des algorithmes d'apprentissage profond permettant de réaliser deux tâches séparées :

1. **détection et identification des cotes :** l'objectif est de localiser et de lire les étiquettes sur chaque livre ; nous utilisons PSENET [5] pour la détection des textes, puis la reconnaissance des caractères par [4] ; le taux de succès pour la reconnaissance des cotes est ici supérieur à 90% pour toutes les bases d'image que nous avons traitées ;

2. **détection des altérations :** après la segmentation des images pour créer des « vignettes » (une par livre présent), détection de la ou des altérations présentes sur chaque ouvrage.

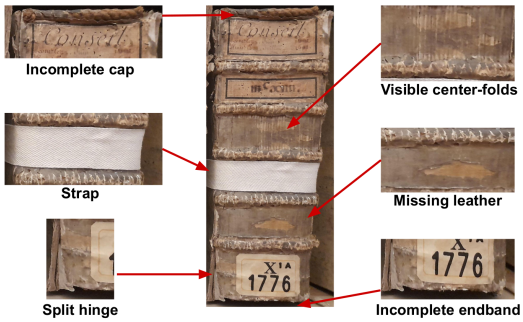

FIGURE 1 : Altérations les plus critiques sur des reliures de livres anciens.

Ce *pipeline* est illustré en figure 2. La détection des altérations se fait par une classification multi-labels à l'aide d'un *Vision Transformer* pré-entraîné sur ImageNet 21k [1].

On voit que pour les deux tâches, l'étape de segmentation des images et de création des vignettes images est essentielle : chaque vignette ne doit contenir qu'un seul livre, et un livre ne doit être présent que sur une seule vignette. Dans les premières versions de notre pipeline, la tâche de segmentation est effectuée à l'aide d'un réseau Mask R-CNN avec backbone ResNet 101 et FPN, puis les vignettes sont calculées à partir des boîtes englobantes des régions de l'image identifiées comme « livre » (*cf.* fig. 3, les registres du Parlement de Paris ont été reconnus avec la probabilité de 100%). Ceci a donné des résultats satisfaisants dans notre premier cas d'étude, le fonds des archives du Parlement de Paris, composée de registres de grande taille et d'apparence très homogène, comme illustré dans la figure. Dans le cas de rayonnages plus disparates, les résultats de segmentation sont insuffisants pour garantir une bonne évaluation des altérations sur chaque livre.

Une description plus complète de notre *pipeline* et des résultats obtenus sur la base des images des archives du Parlement de Paris est disponible en [3].

FIGURE 2 : *Pipeline* de détection d'altérations sur des reliures de livres anciens.

FIGURE 3 : Segmentation des livres détectés sur un rayonnage des archives du Parlement de Paris (Archives Nationales, Paris).

2 Segmentation des images

A la suite de tests effectués avec le *pipeline* présenté au paragraphe précédent sur des images illustrant des fonds de livres patrimoniaux moins homogènes que celle des archives du Parlement de Paris (*cf.* fig 4 par exemple des images des bibliothèques du musée du quai Branly et de l'Ecole française de Rome), nous avons réfléchi aux améliorations possibles de l'étape de segmentation des images. Par ailleurs, une segmentation plus précise permet également, pour les responsables de ces fonds patrimoniaux, de simplifier le protocole de prise de vues.

2.1 Segmentation par apprentissage profond

Nous nous sommes donc intéressés à l'utilisation de méthodes de segmentation basées sur l'apprentissage profond telles que SAM (*Segment Anything Model* [2]). SAM est un modèle de fondation spécialisé pour la segmentation d'images, présenté

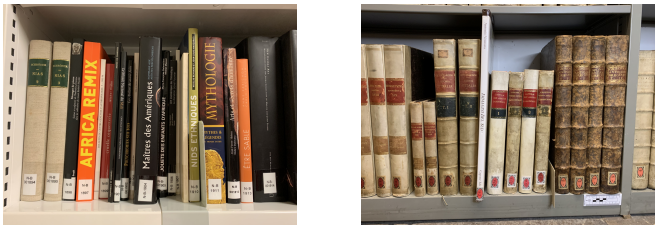

FIGURE 4 : Deux images de rayonnages des bibliothèques du musée du quai Branly (à gauche) et de l'Ecole française de Rome (à droite).

en 2023 par Meta AI, le modèle est « promptable » et peut s'appliquer sans apprentissage additionnel à de nouvelles catégories d'objets.

Pour la segmentation des images de rayonnages de bibliothèques, nous avons utilisé un prompt sous forme de grille régulière de points ; le paramétrage de la grille (nombre de points horizontalement et verticalement) a une influence sur le nombre de masques de région obtenus et sur le temps de calcul (même si nous n'avons pas pour cette application d'objectif de traitement temps-réel). Pour les images des registres du Parlement de Paris (Archives nationales), nous avons obtenu des résultats optimaux avec une grille régulière de 35 points dans chaque direction, deux exemples de masque de segmentation obtenus sont donnés en figure 5.

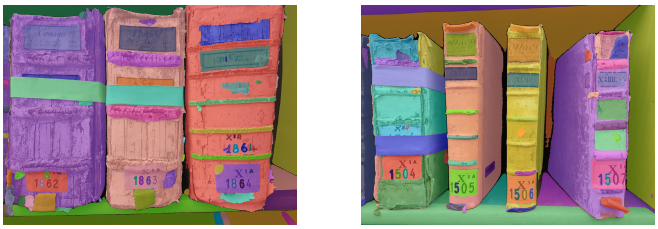

FIGURE 5 : Deux exemples de résultat du modèle SAM sur des images des archives du Parlement de Paris.

Sur ces résultats, on constate que la segmentation obtenue est précise : il est possible d'avoir une délimitation fine des régions « livre », mais un grand nombre de régions autres (étagères, fond des rayonnages) et de sous-régions (étiquettes, taches, sangles, éléments de reliures, *etc.*) sont également isolées. Des post-traitements sont donc nécessaires pour obtenir les masques des régions « livre ».

2.2 Post-traitements

L'objectif des post-traitements appliqués aux résultats de segmentation obtenus à l'aide de SAM est d'éliminer de ces résultats toutes les régions qui ne représentent pas un livre et un seul ; la figure 6 montre l'ensemble des masques de segmentation obtenus sur une image exemple des archives du Parlement de Paris ; sur cette figure, les masques n°1, 2 et 4 en haut à gauche de la figure correspondent aux livres à détecter et donc aux masques à conserver. Pour ce faire, nous avons donc appliqué les opérations suivantes :

Suppression des « petites » régions : toutes les régions dont la surface est inférieure à un seuil donné (en pixels) sont supprimés ; le seuil peut être paramétré en fonction de

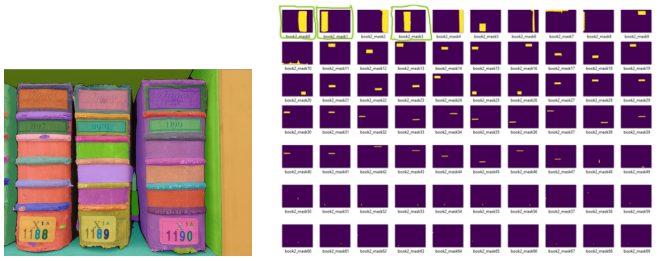

FIGURE 6 : Masques de segmentation (droite) obtenus sur l'image des archives du Parlement de Paris (gauche).

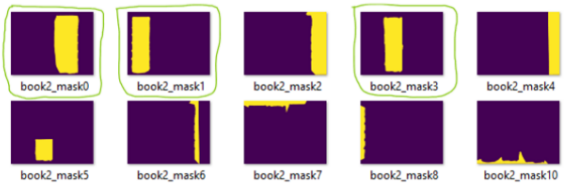

FIGURE 7 : Masques de segmentation obtenus sur l'image de droite de la figure 6 après filtrage sur la dimension des images.

la dimension des images originales, du nombre et de la taille des ouvrages présents dans l'image (pour les images de la base **AN3**, comportant 3 à 5 livres par image, le seuil est fixé à 30.000 pixels) ; les masques restant après cette étape sont présentés en figure 7, les masques corrects sont bien conservés (n° 1, 2 et 4 en haut à partir de la gauche).

Suppression des régions en bord d'image : les régions en bord d'image ne sont pas intéressantes pour notre application : en effet, elles représentent le fond des rayonnages, les étagères elle-mêmes, ou des ouvrages vus partiellement sur cette image, et qui seront vus en entier sur une photographie voisine. Toutes les régions en bord d'image sont donc supprimées (*cf.* figure 8).

Suppression des régions entièrement incluses dans une région plus grande : certaines régions détectées par SAM correspondent à des étiquettes, des taches ou des éléments de reliure des livres, ces régions doivent donc être supprimées (*cf.* figure 9).

3 Résultats

Nous présentons ici les résultats obtenus sur deux bases d'image différentes, avec des niveaux de complexité croissants : *(i)* la collection des archives du Parlement de Paris, très homogène (**AN3**) et *(ii)* des images de la bibliothèque de l'École française de Rome (**Rome**).

3.1 Analyse quantitative

Pour les données **AN3** et **Rome**, nous disposons d'une vérité-terrain : l'annotation manuelle (détourage de chaque livre présent sur l'image) d'images de chacune des bases de données. Nous pouvons donc calculer sur ces vérités-terrain la proportion de livres correctement retrouvés, ainsi que de faux positifs et de faux négatifs. Pour **AN3**, 100 images ont été ainsi manuellement annotées, représentant 342 livres. Pour **Rome**,

FIGURE 8 : Masques de segmentation obtenus sur l'image de la figure 7 après suppression des régions en bord d'image.

FIGURE 9 : Masques de segmentation obtenus sur l'image de la figure 8 après suppression des régions incluses dans une région plus grande.

ce sont 104 images qui ont été annotées pour un total de 1277 livres. Pour chacun de ces deux datasets, nous présentons en table 1 le nombre total de masques de segmentation calculés par SAM, ainsi que le nombre de masques restant après chaque étape de post-traitement. Pour les deux bases, nous constatons bien une convergence du nombre de livres détectés vers la valeur de la vérité-terrain.

Pour les résultats finaux, nous montrons en table 2 la proportion de bonnes et de mauvaises détections : les « vrais positifs » (TP) sont les masques détectés ayant un taux de recouvrement de plus de 80% avec la vérité-terrain pour la base **AN3**, et de plus de 60% pour la base **Rome** ; dans la base **Rome**, les livres sont généralement plus fins, les reliures portent souvent des étiquettes qui couvrent complètement la largeur du livre et qui causent la sur-segmentation, et donc un taux de recouvrement plus faible pour des livres que nous considérons cependant correctement détectés (l'objectif final étant de détecter des altérations sur les reliures, des altérations éventuelles sous les étiquettes ne sont évidemment pas visibles). FP désigne le nombre de « faux positifs » et FN le nombre « de faux négatifs » (livres présents dans la vérité-terrain et non détectés par l'algorithme). Le nombre de livres non détectés (FN) est faible pour les deux bases de données.

Les histogrammes de la figure 10 montrent la répartition des taux de recouvrement (IoU) entre livres détectés et vérités-terrain pour les deux bases d'images. On observe que la grande majorité des livres détectés correspondent à plus de 80% aux vérités terrains détourées par l'experte.

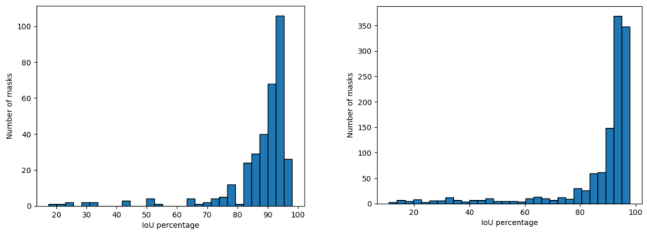

FIGURE 10 : Histogrammes des taux de recouvrement entre livres détectés et vérité-terrain pour les bases **AN3** (à gauche) et **Rome** (à droite).

Base	Nb de livres	Nb de masques après suppression des régions			
		SAM	petites	au bord	incluses
AN3	342	6687	1260	809	353
Rome	1277	14380	3266	1975	1489

TABLE 1 : Nombre de masques de segmentation (régions) détectés après chaque étape de post-traitements pour les deux datasets **AN3** et **Rome**.

Base	TP (en %)	FP	FN
AN3	337 (98,5%)	16	5
Rome	1105 (86,5%)	384	172

TABLE 2 : Taux de bonnes et mauvaises détections de livres dans les bases **AN3** et **Rome**.

3.2 Discussion

Pour notre application, la segmentation des images de rayonnage par SAM doit être complétée par plusieurs post-traitements dont nous avons réglé les paramètres de manière *ad hoc*, mais identiques pour toutes les images d'une même base de données. L'initialisation de SAM avec une grille de 35×35 points est adaptée au format des images acquises par smartphone et permet un bon compromis entre rapidité, ressources matérielles et précision des résultats. Le paramétrage des post-traitements est directement lié aux caractéristiques des images : nombre moyen de livres par image et dimensions (en pixels) des plus petits livres sur lesquels détecter des altérations, ce paramétrage est identique pour toutes les images d'une même base et pourra faire l'objet de recommandations globales pour les usagers. Il reste des cas de détection incorrecte non réglés par les post-traitements : par exemple en figure 11 dans une image de la base **Rome** ; s'ils s'avéraient trop nombreux, une étape supplémentaire de suppression des régions non orientées verticalement permettra de les régler.

FIGURE 11 : Exemple de fausse détection de livre sur une image de la base **Rome**.

L'utilisation de SAM et des post-traitements permet également de simplifier le protocole de prise de vue pour les bibliothèques : il n'est plus nécessaire de régler la distance de prise de vue de manière à ce que seuls les livres d'une étagère soit visible, les post-traitements permettent d'éliminer sans équivoque les livres vus partiellement sur les bords des images.

4 Conclusion et perspectives

Dans cet article, nous avons présenté un outil de segmentation efficace pour la détection de livres patrimoniaux sur des photographies de rayonnages en bibliothèques et archives. Dans cet outil, le modèle de fondation SAM est complété par trois étapes de post-traitements basés sur des heuristiques adaptées au contexte général de la conservation des livres patrimoniaux.

Comparé aux versions précédentes de segmentation d'images utilisées dans notre *pipeline* de détection des altérations, cette contribution nous donne la possibilité d'appliquer ce *pipeline* pour des fonds patrimoniaux aux reliures variées. Il reste toutefois nécessaire de tester le système sur des fonds encore plus diversifiés pour évaluer si le niveau d'automatisation atteint est réellement pertinent. Pour être utilisé sur le terrain, l'outil devra être convivial et présenter un taux d'erreurs inférieur à celui d'un humain.

Remerciements

Nous remercions les conservateurs et restaurateurs des Archives nationales (Paris), de la médiathèque du musée du quai Branly (Paris), de la Bibliothèque et des Archives nationales de Québec (BAnQ) et de la bibliothèque de l'Ecole française de Rome, pour nous avoir permis de photographier leurs fonds patrimoniaux ou transmis des images de ceux-ci.

Références

[1] Alexey DOSOVITSKIY, Lucas BEYER, Alexander KOLESNIKOV, Dirk WEISSENBORN, Xiaohua ZHAI, Thomas UNTERTHINER, Mostafa DEHGHANI, Matthias MINDERER, Georg HEIGOLD, Sylvain GELLY, Jakob USZKOREIT et Neil HOULSBY : An Image is Worth 16x16 Words : Transformers for Image Recognition at Scale. *In 9th International Conference on Learning Representations, ICLR 2021, May 3-7, 2021*, Virtual Event, Austria, 2021. OpenReview.net.

[2] Alexander KIRILLOV, Eric MINTUN, Nikhila RAVI, Hanzi MAO, Chloe ROLLAND, Laura GUSTAFSON, Tete XIAO, Spencer WHITEHEAD, Alexander C. BERG, Wan-Yen LO, Piotr DOLLÁR et Ross GIRSHICK : Segment anything, 2023.

[3] Valérie LEE-GOUET, Lahcen YAMOUN, Zacharie RODIÈRE, Camille SIMON CHANE, Michel JORDAN, Julien LONGHI et David PICARD : A deep learning-based pipeline for the conservation assessment of bindings in archives and libraries. *Multimedia Tools and Applications*, janvier 2025.

[4] Hui LI, Peng WANG, Chunhua SHEN et Guyu ZHANG : Show, attend and read : A simple and strong baseline for irregular text recognition. *In Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 8610–8617, 2019.

[5] Wenhai WANG, Enze XIE, Xiang LI, Wenbo HOU, Tong LU, Gang YU et Shuai SHAO : Shape robust text detection with progressive scale expansion network. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9336–9345, 2019.