

ANALYSE D'IMAGES DE LAMES ENTIERES AVEC LE LOGICIEL LIBRE QUPATH

PECOT Thierry

*Université de Rennes, plateforme FAIA
Biosit (UAR 3480 CNRS – US 18 Inserm)
2, Avenue du Professeur Léon Bernard – 35042 Rennes, France*

thierry.pecot@univ-rennes.fr

WHOLE-SLIDE IMAGE ANALYSIS WITH OPEN-SOURCE SOFTWARE QUPATH

ABSTRACT

Whole-slide imaging thanks to the use of scanners allows acquiring huge amounts of data. Analysis of these data potentially offers the quantification of biological processes at large scale. However, dedicated pieces of software are required to process such large data. Open-source software QuPath was precisely created with that purpose. It can load bright field images as well as fluorescence images, whatever the number of channels is. It allows the visualization of whole-slide images by using pyramidal representation of data. It offers a large variety of image annotations with an ergonomic interface. It enables stain deconvolution for brightfield whole-slide images. QuPath includes classical machine learning tools for pixel classification to segment tissue. It also enables the use of deep learning approaches for the segmentation of nuclei and cells that can be followed by an object classification in order to phenotype cells. Finally, QuPath offers spatial analysis tools for clustering and distance computation.

KEY WORDS

Image analysis, Machine learning methods, Object classification, Pixel classification, Spatial analysis, Whole-slide images.

RESUME

La numérisation de lames entières par l'utilisation de scanners permet d'accéder à de gros volumes de données. L'analyse de ces données offre la promesse de quantification de processus biologiques à grande échelle. Toutefois, étant donnée la taille de ces données, des logiciels dédiés sont

requis. Le logiciel libre QuPath a été créé précisément pour ce type d'analyse. Il peut charger des images acquises en lumière blanche comme des images de fluorescence, sans limite sur le nombre de canaux. Il permet une visualisation fluide des lames entières grâce à la prise en compte du format pyramidal. Il offre une grande variété d'annotations d'image avec une interface ergonomique. Il propose la déconvolution couleur des lames acquises en lumière blanche. QuPath intègre des outils d'apprentissage classique afin de réaliser une classification de pixels pour la segmentation de tissus. Il permet également l'utilisation d'outils d'apprentissage profond pour la segmentation de noyaux et de cellules, qui peut être suivi d'une classification d'objets afin d'obtenir un phénotypage des cellules. Enfin, QuPath propose des outils d'analyse spatiale par regroupement de données et calculs de distance.

MOTS CLES

Analyse d'images, Analyse spatiale, Classification d'objets, Classification de pixels, Lames entières, Méthodes d'apprentissage,

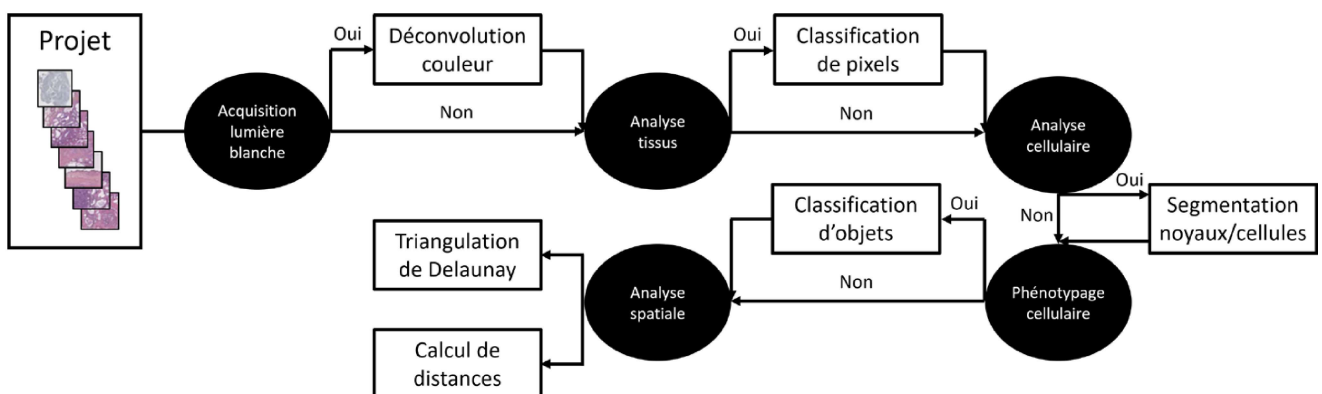


Figure 1 : Etapes successives d'analyse les plus courantes d'un projet d'images de lames entières avec QuPath.

INTRODUCTION

L'acquisition en routine sur les plateformes d'histologie de lames entières permet l'accès à des données image de grande dimension. Ces données sont riches et potentiellement très informatives mais nécessitent des outils

appropriés afin d'être analysées pour extraire des données biomédicales de manière quantitative. QuPath [1] a été créé spécifiquement dans ce but. Ce logiciel libre permet de visualiser de manière fluide des images de lames entières et d'en extraire des mesures quantitatives, en allant de la segmentation à l'analyse spatiale. QuPath est particulièrement adapté à une analyse centrée sur la cellule. Les principaux types d'analyse proposés par QuPath sont résumés dans le diagramme présenté **Figure 1**.

Représentation pyramidale des lames entières

Une fois scannées, les images de lames entières représentent des données de grande dimension, typiquement plusieurs dizaines de milliers de pixels de côté. Quand une image digitale est ouverte par un appareil électronique, la totalité de cette image est transférée dans la mémoire RAM de cet appareil. Malheureusement, les images de lames entières sont généralement plus grandes (plusieurs dizaines de gigaoctets une fois décompressées) que la quantité de mémoire RAM disponible, ce qui empêche de les visualiser avec des logiciels classiques tels que ImageJ/Fiji [2]. Pour contourner ce problème, les images de lames entières sont définies dans un format pyramidal, c'est-à-dire que la même image est disponible à plusieurs résolutions différentes. Cela permet de ne mettre dans la mémoire RAM qu'une image à faible résolution pour visualiser la lame dans son entièreté puis ne mettre que l'information utile dans la mémoire RAM à mesure que l'utilisateur zoome dans la lame. Cette représentation des données image permet une navigation fluide, même avec un ordinateur ne possédant qu'une quantité limitée de mémoire RAM.

Annotations

Comme les images de lames entières représentent des données de grande dimension, QuPath encapsule les différentes analyses sous forme d'annotations, une représentation compacte et efficace. QuPath propose un certain nombre d'outils d'annotation manuelle, allant d'annotations unidimensionnelles (lignes, poly-lignes) à des annotations bidimensionnelles

régulières (rectangles, ovales, polygones, brosse) avec la possibilité de définir des lignes ou polygones à main levée. QuPath propose également une baguette magique qui permet d'annoter des objets suivant la régularité des structures présentes dans l'image en prenant en compte la résolution utilisée. Il est possible de modifier des annotations pour les agrandir, les rétrécir, soustraire des régions ou combiner des annotations entre elles.

Déconvolution couleur

Pour des images acquises en lumière blanche, il est potentiellement utile de pouvoir passer d'un espace Rouge-Vert-Bleu (canaux utilisés pour l'acquisition) à un espace Hématoxyline-DAB-Résidu pour des images de tissus marqués par hématoxyline et DAB par exemple. En effet, si le but est de segmenter les noyaux, cette tâche sera beaucoup plus aisée à effectuer à partir d'une estimation du marquage hématoxyline plutôt qu'à partir des canaux Rouge-Vert-Bleu. QuPath offre la possibilité d'utiliser la méthode automatique de déconvolution couleur décrite par Ruifrok *et al.* [3] (voir **Figure 2**). Il est aussi possible de définir manuellement des régions de marquage « pur » (Hématoxyline et DAB par exemple) afin de réaliser la déconvolution couleur.

Classification de pixels

QuPath intègre des outils d'apprentissage machine classique (perceptron multicouches [4], forêts aléatoires [5], régression logistique [6], K voisins les plus proches [7]) afin de réaliser une classification de pixels, c'est-à-dire l'estimation en chaque pixel d'une étiquette, appelée classe en apprentissage. Ces outils sont très utiles afin de segmenter des tissus. Par exemple, en utilisant les classes « épithélium », « stroma » et « fond », il est potentiellement possible d'estimer pour chaque pixel s'il fait partie de l'une de ces classes. Ces outils font partie de la famille des méthodes d'apprentissage, cela implique que l'utilisateur entraîne un algorithme à estimer une classe en chaque pixel en lui présentant des exemples. Si l'on considère l'exemple « épithélium/stroma/fond », l'utilisateur utilise les outils d'annotation de QuPath pour donner des

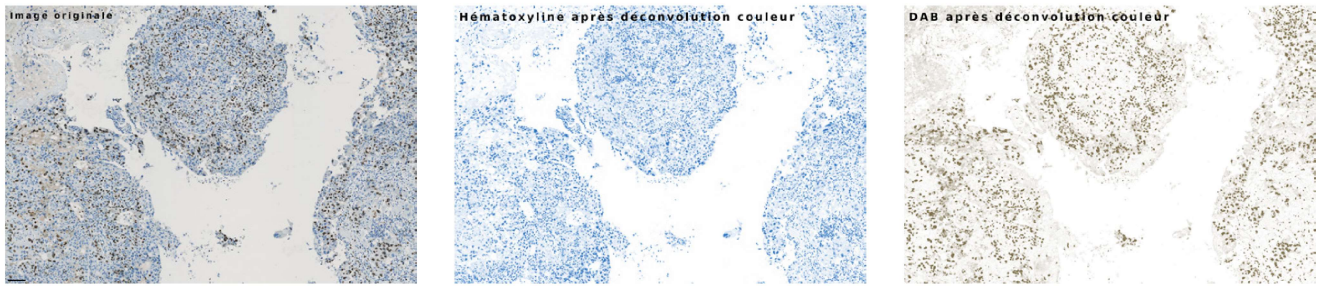


Figure 2. Déconvolution couleur d'une lame Hématoxyline-DAB. Gauche : image Hématoxyline-DAB d'une biopsie humaine de poumon où le marquage DAB permet d'identifier le marqueur de prolifération Ki67. La barre d'échelle correspond à 100 μ m. Milieu : estimation de la composante hématoxyline de l'image de gauche. Droite : estimation de la composante DAB de l'image de gauche.

exemples de pixels appartenant à l'épithélium, au stroma et au fond. Ces approches extraient de l'image des caractéristiques définies par l'utilisateur (gaussienne, laplacien de gaussienne, déterminant de la hessienne, ...) à différentes résolutions afin d'apprendre comment identifier les différentes classes. Ces méthodes étant très rapides, il est possible de visualiser le résultat de la classification, d'identifier des régions mal classées et d'ajouter de nouveaux exemples afin d'affiner de manière interactive et en temps réel la classification. Un exemple de classification de pixels est présenté **Figure 3**.

Segmentation de cellules

Si la classification de pixels est performante pour l'identification de tissus, elle est généralement sous-optimale pour la segmentation de cellules ou noyaux. En effet, si des cellules se touchent, ce qui arrive fréquemment quand on observe des tissus, l'obtention d'un label en chaque pixel n'est pas suffisante pour identifier des cellules individuelles. Ces dernières années, des approches d'apprentissage profond pour la segmentation de noyaux et de cellules sont apparues et permettent d'obtenir des performances nettement supérieures à ce qu'il était possible d'obtenir auparavant [8]. QuPath permet d'intégrer les deux approches les plus populaires en microscopie, à savoir Stardist [9] pour la segmentation de noyaux et Cellpose [10] pour la segmentation de cellules. De plus, une liste de mesures morphologiques et d'intensité sont

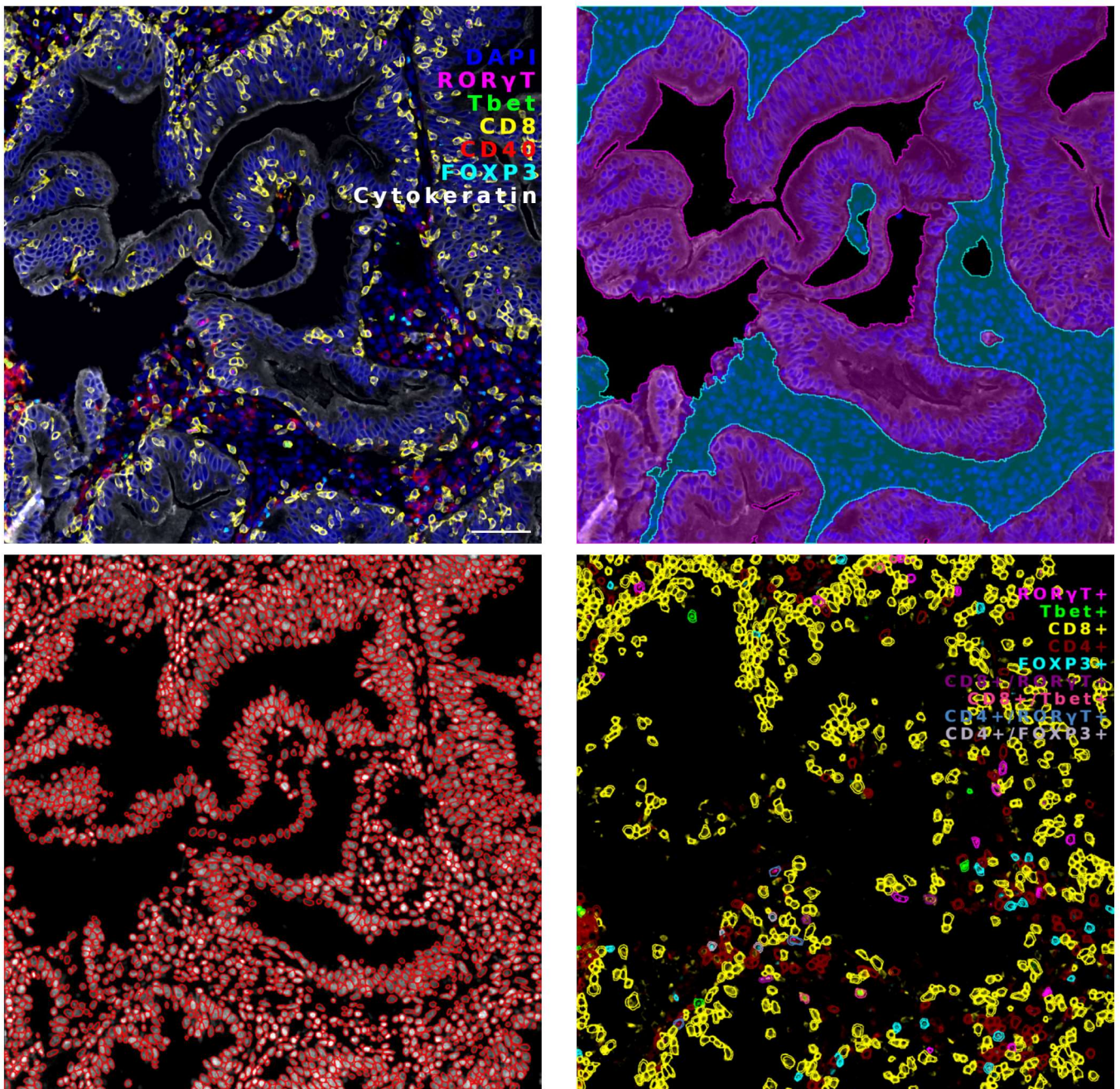


Figure 3. Classification de pixels, segmentation de noyaux et classification d'objets avec QuPath.

Haut gauche : image multiplexée d'une biopsie humaine de polype. La barre d'échelle correspond à 100 µm. Haut droite : segmentation de l'épithélium (magenta) et du stroma (cyan) par classification de pixels. Bas gauche : segmentation des noyaux avec la méthode d'apprentissage profond Stardist. Bas droite : identification des différents types de cellule par classification d'objet.

automatiquement extraites pour chaque objet segmenté. La **Figure 3** illustre un exemple de segmentation de noyaux réalisé avec Stardist entraîné sur les données disponibles dans [11-13].

Classification d'objets

Une fois tous les noyaux segmentés, QuPath permet l'identification de différents types de noyaux ou cellules par méthodes d'apprentissage machine classique (perceptron multicouches, forêts aléatoires, K voisins les plus proches). Cette étape est similaire à la classification de pixels mais cette fois-ci, le but est de réaliser une classification d'objets en utilisant des caractéristiques morphologiques (taille, périmètre, ...) ou d'intensité (intensité moyenne, maximum, écart-type, ...) associées aux noyaux, cellules, régions cytoplasmiques et/ou régions membranaires. Une fois encore, la rapidité d'exécution de cette classification permet une utilisation interactive afin d'ajouter de nouveaux exemples et d'observer en temps réel l'impact sur la classification des cellules. **La Figure 3** illustre un exemple de classification d'objets.

Analyse spatiale

QuPath propose principalement deux approches pour l'analyse spatiale. Tout d'abord, il est possible de réaliser une triangulation de Delaunay sur l'ensemble des cellules, c'est-à-dire construire un graphe où chaque cellule est associée à un nœud et où les cellules voisines sont reliées par des arêtes. Cela permet de créer une carte de densité des cellules afin d'identifier des groupes de cellules (de même phénotype ou non) proches les unes des autres. Ensuite, QuPath propose également le calcul pour chaque cellule des distances aux cellules les plus proches pour chaque phénotype identifié.

Tutoriels

Des tutoriels vidéo montrant comment installer QuPath, créer un projet, définir des annotations manuelles, réaliser une déconvolution couleur, une classification de pixels, une segmentation de noyaux et une classification d'objets sont disponibles sur la page github <https://github.com/tpecot/WholeSlideImageAnalysisWithQuPath>.

CONCLUSION

QuPath est un logiciel libre qui permet de visualiser et d'analyser des images de lames entières. Il met à la disposition de non spécialistes des outils très performants permettant une analyse quantitative tissulaire et cellulaire très complète. Ce logiciel est toujours en développement et son avenir est assuré sur un horizon temporel long, ce qui en fait un outil de choix pour les utilisateurs de lames entières.

REFERENCES BIBLIOGRAPHIQUES

1. BANKHEAD P., LOUGHREY M.B., FERNANDEZ J.A., DOMBROWSKI Y., MCART D.G., DUNNE P.D., MCQUAID S., GRAY R.T., MURRAY L.J., COLEMAN H.G. & JAMES J.A. : QuPath: Open source software for digital pathology image analysis. *Scientific reports*, 2017, **7**(1), 1-7.
2. SCHINDELIN J., ARGANDA-CARRERAS I., FISE E., KAYNIG V., LONGAIR M., PIETZSCH T., PREIBISCH S., RUEDEN C., SAALFELD S., SCHMID B. & TINEVEZ J.Y. : Fiji: an open-source platform for biological-image analysis. *Nature methods*, 2012, **9**(7), 676-682.
3. RUIFROK A.C., KATZ R.L. & JOHNSTON D.A. : Comparison of quantification of histochemical staining by hue-saturation-intensity (HSI) transformation and color-deconvolution. *Applied Immunohistochemistry & Molecular Morphology*, 2003, **11**(1), 85-91.
4. ROSENBLATT F. : Principles of neurodynamics: Perceptrons and the theory of brain mechanisms. Spartan books, 1962.
5. BREIMAN L. : Random forests. *Machine learning*, 2001, **45**:5-32.
6. COX D.R. : The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1958, **20**(2), pp.215-232.
7. FIX E. & HODGES J.L. : Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. *Rapport pour USAF School of Aviation Medicine*, 1951.

8. CAICEDO J.C., ROTH J., GOODMAN A., BECKER T., KARHOHS K.W., BROISIN M., MOLNAR C., MCQUIN C., SINGH S., THEIS F.J. & CARPENTER A.E. : Evaluation of deep learning strategies for nucleus segmentation in fluorescence images. *Cytometry Part A*, 2019, **95**(9), 952–965.
9. SCHMIDT U., WEIGERT M., BROADDUS C. & MYERS G. : Cell detection with star-convex polygons. *21e Symposium Medical Image Computing and Computer Assisted Intervention*, Granada, Espagne, 2018, **11** 265–273.
10. STRINGER C., WANG T., MICHAELOS M. & Pachitariu M. : Cellpose: a generalist algorithm for cellular segmentation. *Nature methods*, 2021, **18**(1), 100–106.
11. GREENWALD N.F., MILLER G., MOENE., KONG A., KAGEL A., DOUGHERTY T., FULLAWAY C.C., MCINTOSH B.J., LEOW K.X., SCHWARTZ M.S., PAVELCHEK C. *et al.* : Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nature biotechnology*, 2022, **40**(4), 555–565.
12. PECOT T., CUITINO M.C., JOHNSON R.H., TIMMERS C. & LEONE G. : Deep learning tools and modeling to estimate the temporal expression of cell cycle proteins from 2D still images. *PLOS Computational Biology*, 2022, **18**(3), p.e1009949.
13. PECOT T., ALEKSEYENKO A. & WALLACE K. : A deep learning segmentation strategy that minimizes the amount of manually annotated images. *F1000Research*, 2022, 10.