

Titre du projet : CANAPAI – Extraction d’information spatiale des CADastres NAPoléoniens par l’Intelligence Artificielle

Coordonnées du porteur n°1 :

Nom : RAVEAUX

Prénom : Romain

Nom du Laboratoire (indiquer notamment le numéro de labellisation EA, UMR, IFR...) : UR 6300 LIFAT

Ecole Doctorale : MIPTIS

N° téléphone : 02 47 36 14 27

Courriel : romain.raveaux@univ-tours.fr

Coordonnées du porteur n°2 :

Nom : RIGOT

Prénom : Jean-Baptiste

Nom du Laboratoire (indiquer notamment le numéro de labellisation EA, UMR, IFR...) : UMR 7324 CITERES-LAT

Ecole Doctorale : HL

N° téléphone : 0678369649

Courriel : jb.rigot@univ-tours.fr

Bilan scientifique

L’enjeu du projet CANAPAI est d’extraire automatiquement des données géographiques à partir de cartes anciennes numérisées et plus spécifiquement de cadastres napoléoniens. Cette action est appelée la vectorisation en sciences de l’information géographique. L’apport de l’intelligence artificielle et *du deep learning* lors de ce travail réalisé manuellement dans des applications professionnelles, permettrait de réduire le temps et le coût d’acquisition de données anciennes.

A partir de feuilles de cadastres napoléonien, l’objectif est de produire une donnée de référence composée des limites de parcelles. A l’issue du projet VECCAR financé par le RnMSH en 2021-22, l’apprentissage d’un réseau de neurones (RN) s’appuyait sur un trop faible volume de données. Il a été nécessaire de trouver d’autres feuilles cadastrales pour lesquelles la vectorisation manuelle est réalisée. La donnée servant à l’apprentissage associe ici une feuille cadastrale numérisée et la donnée vectorisée manuellement ; cette dernière se nomme la vérité-terrain ou donnée annotée. L’hypothèse faite dans le cadre du projet CANAPAI porte sur l’amélioration des résultats d’une vectorisation automatique en augmentant le volume de données d’apprentissage du RN. Pour ce faire le projet a utilisé les parcelles cadastrales anciennes vectorisées manuellement sur la commune de Noizay (Indre-et-Loire) par Pauline Thonnet qui a analysé l’évolution du paysage sur le temps long dans le val de Cisse (thèse soutenue en 2021).

Ainsi par rapport aux données d’apprentissage du RN faites sur Fréjus à l’issue du projet VECCAR, le cadastre de Noizay (Figure 1) offre deux caractéristiques testées dans les résultats de la vectorisation automatique :

- La configuration géographique : en étoffant la variété des formes des parcelles, de leur agencement en fonction du relief ou de la mise en valeur humaine des terres, l’extraction des données devrait être améliorée. Noizay présente des formes de parcelles extrêmement fines et longue.
- La qualité de la facture graphique des feuilles cadastrales : au XIXe siècle le cadastre est dessiné à la main et par différentes personnes d’un bout à l’autre de l’hexagone. Son graphisme n’est donc pas uniforme sur tout le territoire national et il apparaît nécessaire d’étoffer là aussi le volume de données sur cette caractéristique du cadastre.

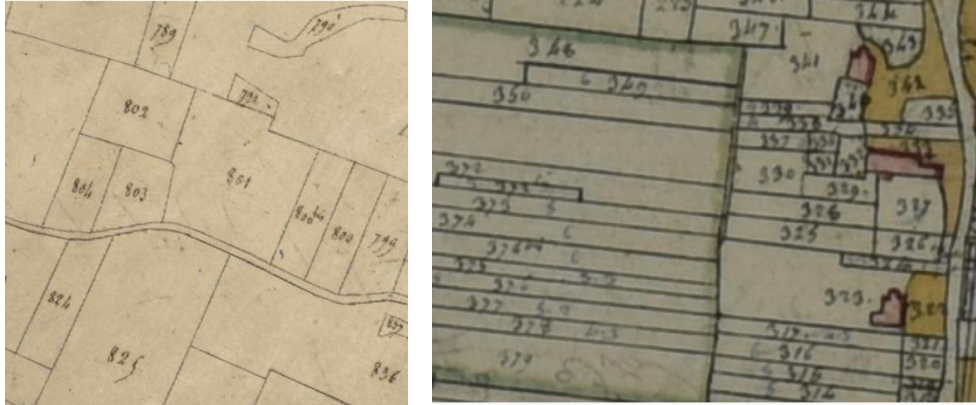


Figure 1 : Extrait du cadastre de Fréjus (Var) à gauche et Noizay (Indre-et-Loire) à droite

Résultats

Qualité des jeux de

données

- De nouvelles expérimentations mettent en jeu les techniques et algorithmes retenus sur un nouveau corpus. Ces expérimentations ont pu mettre en évidence à la fois les limites du pouvoir de généralisation des algorithmes d'apprentissage profond sur ces types de données.
- Les formes très étirées des parcelles ont imposé une augmentation de la qualité de la numérisation : une résolution standard à 300 DPI dans un format compressé (JPG) empêche une bonne extraction des limites sur ces configurations particulières. Des images en 600 DPI dans un format compressé qui ne recode pas les pixel (TIFF) produisent de meilleurs résultats. Dans ce contexte, des nouvelles numérisations ont été recueillies pour améliorer la qualité des jeux de données disponibles.

Méthodes et traitements

- Conséquence de la facture graphique de la feuille de Noizay : des lignes ont été extraites alors qu'elles ne constituent aucunement une information liée au cadastre. Ces lignes forment une quadrillage technique utilisée pour la construction du plan.
- Bonne qualité des résultats en termes de quantité d'information extraite mais la qualité de celle-ci n'est pas exploitable sans des traitements supplémentaires propre à la géomatique qui n'ont pas été intégrés dans la chaîne de traitement.

Discussion

Si cette qualité est peu satisfaisante dans l'optique de produire une donnée de référence au sens géomatique, elle peut selon certaines hypothèses scientifiques d'analyse spatiale (évolution globale des parcelles dans le temps, nombre, surface...), être suffisante pour l'obtention de données statistiques massives sur les parcelles. Ces types d'analyses spatiales sont pratiquées en archéologie. Cette interrogation ouvre de nouvelles pistes de recherche en lien avec l'archéologie permettant de croiser d'autres informations telles que des données de microreliefs acquises par LIDAR.

Perspectives

Les développements de jeux de données et les résultats d'expérimentation des méthodes implémentées sont un acquis à la fois scientifique et technique. Ils nous permettent d'ancrer des recherches dans une collaboration entre les domaines informatique (apprentissage automatique), géomatique et archéologique.

Ainsi, les relations scientifiques et l'expérience acquise dans un domaine peu investi par l'intelligence artificielle, c'est-à-dire la vision par ordinateur sur des cartes anciennes numérisées, a permis d'intégrer la demande de labellisation du **consortium national PicorIA** auprès de l'IR* Huma-Num. Ce nouveau consortium labellisé en janvier 2024 est dédié à l'analyse de corpus visuels numériques en sciences humaines et sociales par le biais d'outils d'intelligence artificielle. La MSH Val de Loire est partenaire pour l'Université de Tours. La mission de PictorIA répond aux constats

du projet CANAPAI en offrant cette opportunité de créer une base de données d'apprentissage sur la cartographie. L'objectif du consortium est d'améliorer l'apprentissage de RN en constituant des jeux de données solides et pertinents, permettant à la fois l'apprentissage massif et une reproductibilité des expérimentations à venir pour les communautés. Ils augmenteront la visibilité à l'échelle nationale et internationale des travaux des unités (UR LIFAT, UMR Citeres, UAR MSH) associées au projet.

<https://pictoria.hypotheses.org/>

Actions scientifiques

- La page Internet :

<https://www.msh-vdl.fr/projet/canapai-extraction-dinformation-spatiale-des-cadastres-napoleoniens-par-lintelligence-artificielle/>

- Gitlab : <https://scm.univ-tours.fr/projetspublics/lifat/veccar-bdcn>

- Séminaire diffusé nationalement, une collaboration MSH / CRIHAM (U. Poitiers) :

<https://www.msh-vdl.fr/actualite/seminaire-lintelligence-artificielle-dans-le-traitement-dimages-numeriques-en-shs/>

- Séminaire en ligne sur Canal U :

<https://www.canal-u.tv/chaines/mshvalde Loire/l-intelligence-artificielle-dans-le-traitement-d-images-numeriques-en-shs>