

Contacts :  
D. Andrieu : dominique.andrieu@univ-tours.fr  
R. Raveaux : romain.raveaux@univ-tours.fr

Dominique ANDRIEU<sup>1</sup>, Romain RAVEAUX<sup>2</sup>

<sup>1</sup>USR 3501 MSH Val de Loire, <sup>2</sup>EA 6300 LIFAT



### Enjeu scientifique

L'objectif du projet est d'établir une **preuve du concept**. Elle vise à valider la faisabilité de la vectorisation de cartes numérisées anciennes par l'intelligence artificielle (AI), afin d'en **extraire les informations géographiques** historiques qu'elles contiennent.

Les sources cartographiques imprimées sont très variées (Cartomundi, Gallica, bibliothèques universitaires, archives départementales, etc.). Une typologie des problèmes réalisée par la MMSH permet d'appréhender la complexité du traitement à développer.

### Exemples de contraintes à l'extraction d'informations géographiques



Cadastre napoléonien, tableau d'assemblage de la commune de Nice, 1871.  
*Document abîmé : déchiré, sali, sommairement réparé, etc.*



Carte topographique de Dufour, feuille n°2 Belfort-Basel, Swisstopo, 1932.  
*Information difficile à différencier.*

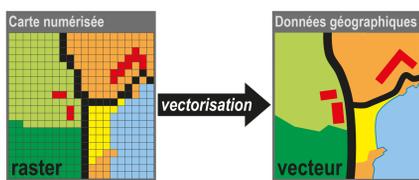


Cadastre napoléonien de Puget-sur-Argens ; © Archives départementales du Var, droits réservés.  
*Signes ayant plusieurs significations : les lignes représentent des limites de parcelles ou de bâtiments.*



Carte topographique 1:100 000, feuille 023, IGN 1958.  
*Signes différents pour représenter la même information : les zones bâties sont roses et noires.*

### Problématique géomatique



**Vectorisation** : opération consistant à transformer une image pixellisée en objets géométriques exploitables par les outils de la géomatique, elle interprète l'espace numérisé.

La **vectorisation manuelle** est connue des géomaticiens. Les outils existent mais l'opération est longue et chronophage.

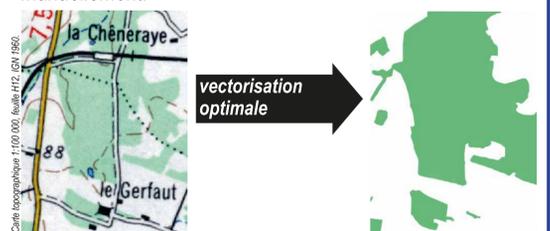
La **vectorisation automatique** est répandue pour les images satellitaires. Elle emploie des développements spécialisés mais non adaptés aux documents cartographiques.

La vectorisation automatique extrait des informations en fonction du code numérique porté par les pixels qui représentent des couleurs. Elle délimite des plages de couleurs comme l'illustre la figure 1. Cette figure montre que la carte n'est pas une juxtaposition de couleurs comme l'image satellitaire. La carte est une image abstraite de l'espace superposant de nombreuses informations. L'appréhension de l'espace de la carte fait appel un **processus cognitif** qui rend complexe l'automatisation de sa vectorisation. Une **discontinuité graphique** ne signifie pas systématiquement une **discontinuité spatiale** comme le montre la figure 1. Les massifs boisés à extraire de cette carte numérisée doivent être représentés par les polygones de la figure 2.

Figure 1 : vectorisation automatique imparfaite due à la surimpression d'informations sur les surfaces à extraire.



Figure 2 : vectorisation optimale de l'image pratiquée manuellement.



### Intelligence artificielle

L'**intelligence artificielle** (IA) est un système informatique prenant des décisions en fonction d'un apprentissage automatique réalisé au préalable. Dans le projet **Veccar**, les pixels de la carte numérisée composent l'information à traiter. L'IA produit une **image numérique segmentée** de la carte à traiter, où chaque pixel est catégorisé suivant qu'il répond ou non à l'objet du traitement.

L'IA est composée d'un modèle (**réseau de neurones**) dont les paramètres sont trouvés par un algorithme d'apprentissage (**deep learning**). L'algorithme d'apprentissage opère sur des données représentant la connaissance du problème.

Si l'IA est répandu pour traiter les images satellitaires, elle l'est moins pour les cartes pour les raisons décrites précédemment. Les développements existants ne sont pas accessibles aujourd'hui : par exemple Historical map, plug-in obsolètes de l'application QGIS.

### HYPOTHÈSE

**Le développement de l'IA et plus spécifiquement celui du deep-learning peut-il répondre à l'acquisition d'informations géographiques à partir des cartes numérisées malgré les contraintes particulières de ces images ?**

L'**apprentissage** est une étape importante puisque sur lui repose la réussite du traitement de l'information par l'IA. Il se compose de **vérités-terrain** (VT). Une VT est l'information montrant le résultat à produire. Si l'objectif est d'extraire les surfaces boisées d'une carte numérisée, la VT sera une image masquant toutes les surfaces ne représentant pas les bois et forêt (cf. figure 2). La connaissance de l'expert est transmise par la VT à l'IA. Elle est exploitée par l'algorithme d'apprentissage afin d'optimiser le réseau de neurones qui produira une image segmentée (raster) de la carte en vue de sa vectorisation.

Extrait du Projet Recherche & Développement de Yoann Dupas, élève ingénieur en 5e année du Département Informatique de l'Ecole Polytechnique de l'Université de Tours



Carte source numérisée : carte topographique 1:100 000, feuille H11, IGN 1960.



Vérité-terrain : trois informations géographiques sont renseignées : les routes nationales (pixels rouges), les routes départementales (pixels jaunes) et les espaces boisés (pixels verts).



Image segmentée : après apprentissage, l'IA propose cette segmentation de la carte. **Points positifs** : la continuité des massifs boisés et des routes est respectée. **Points à améliorer** : confusion des pixels verts entre bois et vignes, réseau routier à affiner.

### Perspectives

Une **étude expérimentale** en cours est portée par Iliad3, centre d'expertise et de transfert de l'Université de Tours. Elle vise à comparer et tester différents modèles de réseau de neurones afin de mettre à l'épreuve la qualité des images segmentées produites par l'IA répondant aux attentes des experts. Cette étude permettra de déposer un projet de recherche à l'ANR.

