

Étude morphologique des parcelles cadastrales

Quel apport des réseaux de neurones ?

Stéphanie Himpens, Mathilde Poulhes et François Sémécurbe

7 février 2019

SSP Lab, Insee

Table of contents

1. Introduction
2. L'hypothèse de la forme des parcelles
3. Mise en œuvre des méthodes de Deep Learning
4. Conclusions

Introduction

La localisation des nouvelles constructions représente un enjeu majeur des politiques de développement durable (environnement, ségrégation, transport).

- En particulier, le «mitage» fait l'objet d'une lutte de la part des pouvoirs publics, qui cherchent à promouvoir le lotissement (Herrmann, 2018).
- Il est donc primordial de comprendre les phénomènes qui expliquent son inégale répartition dans l'espace ?

Explications potentielles : deux échelles interdépendantes d'analyse :

- Les agents sont différents : préférences, culture, etc.
- L'environnement est différent : réseaux plus ou moins denses d'électricité, prix des terres agricoles, forme des parcelles cadastrales (Watteaux 2005).

Motivation

Le mitage ne relève pas totalement de la géographie urbaine... (Albizzati et al 2017)

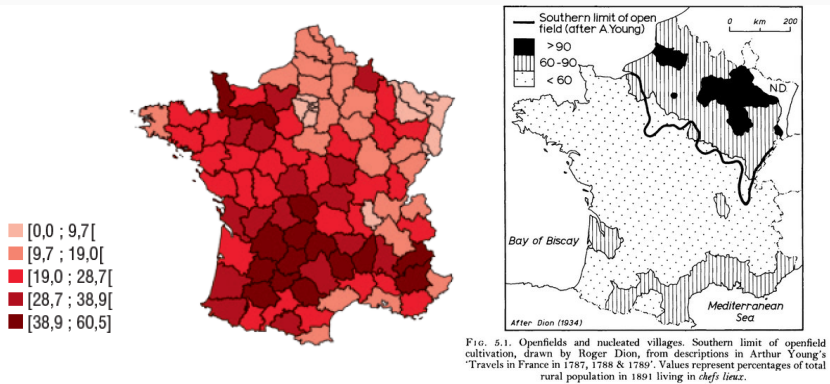
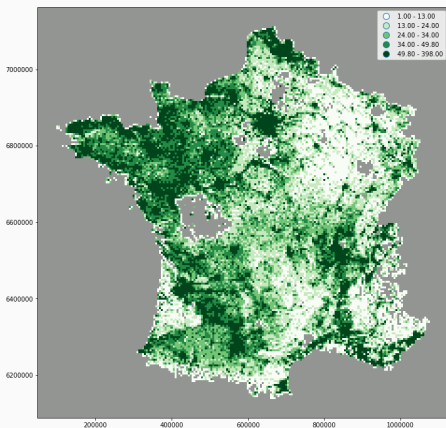


FIG. 5.1. Openfields and nucleated villages. Southern limit of openfield cultivation, drawn by Roger Dion, from descriptions in Arthur Young's 'Travels in France in 1787, 1788 & 1789'. Values represent percentages of total rural population in 1891 living in *chefs lieux*.

«Ces disparités peuvent être également expliquées par des modes d'habitat hérités de l'histoire, notamment agricole, de ces provinces (Demangeon, 1927 ; Le Bras et Todd, 2013).»

Nombre de postes de distribution publique électricité



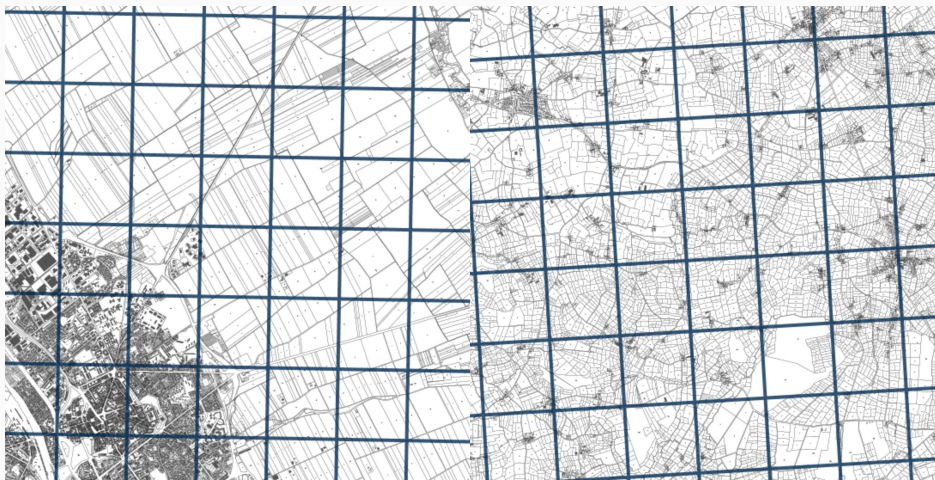
Notes : Un poste HTA/BT (ou poste de transformation HTA/BT) est un local, inaccessible au public, assurant la liaison entre le réseau haute tension HTA (HTA) et le réseau basse tension (BT). Les valeurs manquantes sont des zones couvertes par des réseaux locaux de distribution.

L'hypothèse de la forme des parcelles

Intuition : «Openfield» versus «Bocage»



Intuition : «Openfield» versus «Bocage»



Les notions de « bocage » et « d'openfield » ne font pas consensus parmi les historiens et les géographes. Ceci implique, que nous ne pouvons pas bâtir notre étude des parcelles à l'aide de définitions univoques. Pour surmonter ce problème, nous nous sommes appuyés sur les méthodes de Machine Learning.

Trois étapes :

1. Créer un échantillon d'apprentissage en Labellisant **manuellement** en « bocage » ou « openfield » des configurations parcellaires spécifiques.
2. Apprendre à un algorithme de ML à prédire le label d'une configuration parcellaire à partir de l'échantillon d'apprentissage
3. Appliquer cet algorithme à l'ensemble des espaces métropolitains.

Mise en œuvre des méthodes de Deep Learning

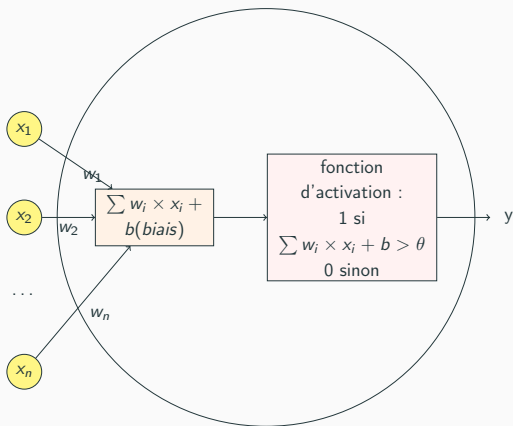
Qu'est-ce qu'un réseau de neurones ?

Un réseau de neurones artificiels, ou réseau neuronal :

1. est un système qui s'inspire des réseaux de neurones biologiques
2. est une méthode d'apprentissage statistique supervisée (ou non supervisée)
3. peut être utilisé afin de classer des images

Qu'est-ce qu'un réseau de neurones ?

L'élément de base d'un réseau de neurones : *Le perceptron*

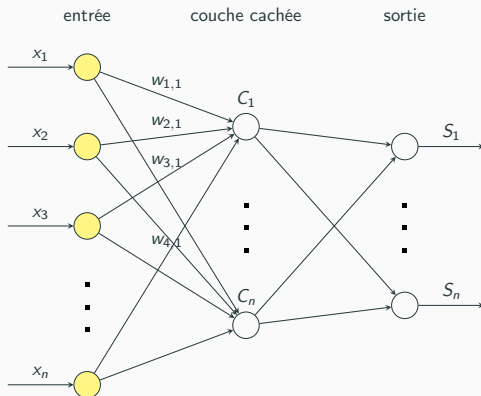


Ci-dessus : Le neurone formel de W. Mc Culloch et W. Pitts.

Entraîner le réseau consiste à estimer les poids w_i et le paramètre b .

Le perceptron multicouche

D. Rumelhart et simultanément Yann LeCun introduisent le perceptron multi-couche en 1986 afin de contourner certaines limitations du perceptron :

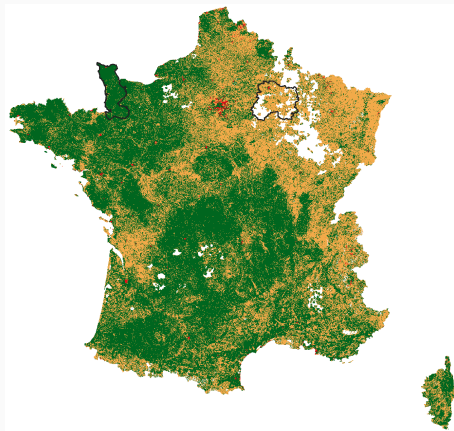


Un perceptron multicouche est un algorithme d'apprentissage supervisé :

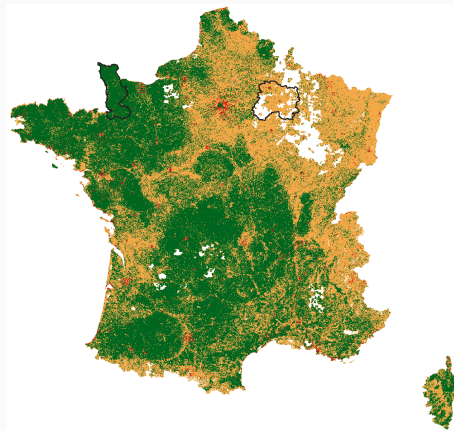
- L'objectif de la phase d'entraînement est d'estimer les paramètres de sorte que les prévisions du réseau (en sortie) correspondent aux labels prédéfinis.
- Une fois les paramètres estimés, le réseau peut-être appliqué à de nouvelles données (non labellisées) afin de prédire leurs catégories.

Dans notre application, nous avons utilisé 9 indicateurs géométriques (croisement entre les terciles de surface et de forme, indice de Gravelius) en entrée pour prévoir l'appartenance d'une image à l'une des trois classes de sortie ('50', '51' et 'villes').

Régression logistique :



Réseau de neurones :



Un réseau spécialisé dans le traitement d'image

Les perceptrons multicouches ne sont pas adaptés aux images :

1. Non prise en compte la **proximité spatiale** des valeurs en entrée (des noeuds x_i)
2. **Le nombre de poids à estimer w_i devient vite considérable** avec des images.



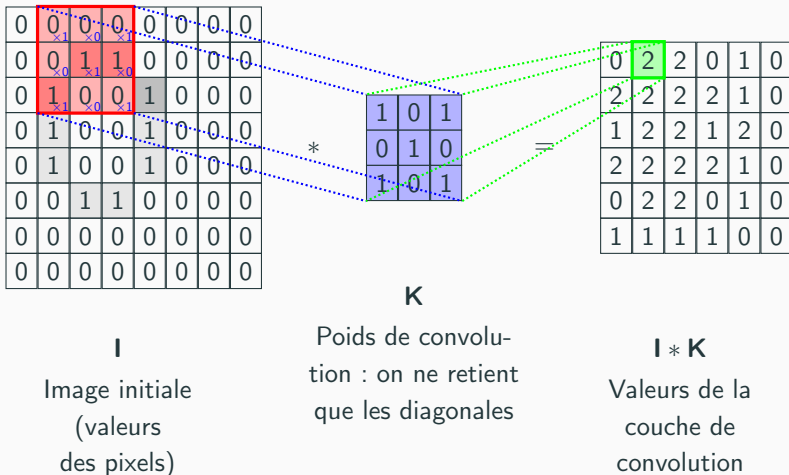
Introduction des réseaux convolutifs.

Les réseaux convolutifs (ou convnet) :

1. sont bâtis sur le modèle des perceptrons multicouches.
2. possèdent des couches spécifiques (couches de *convolutions* et de pooling notamment)

Un réseau spécialisé dans le traitement d'image

La convolution permet de prendre en compte la proximité spatiale :



Un réseau spécialisé dans le traitement d'image

Les couches de pooling permettent d'échantillonner l'image :

0	2	2	0	1	0
2	2	2	2	1	0
1	2	2	1	2	0
2	2	2	2	1	0
0	2	2	0	1	0
1	1	1	1	0	0

Valeurs en sortie
de la convolution

1	0
1	0

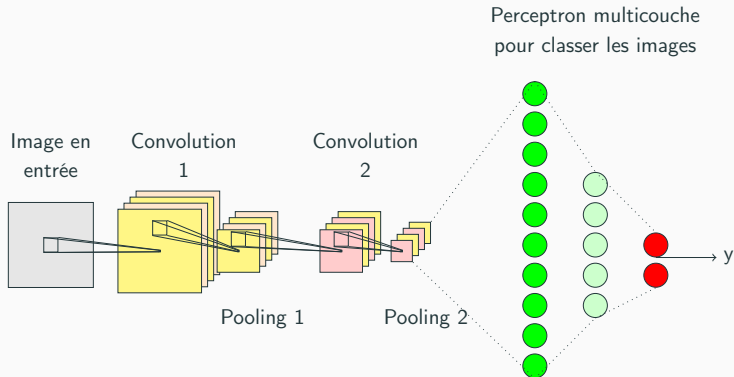
Pooling : on
prend le maximum
(ou la moyenne
ou la somme)

2	2	1
2	2	2
2	2	1

Valeurs en sor-
tie du pooling

Un réseau spécialisé dans le traitement d'image

Les différents types de couches sont combinées (à la manière d'un perceptron multicouche) afin de constituer le réseau :



L'empilement de couches de convolutions est suivie d'un perceptron multicouche qui permet de classer les différentes images (ici en deux classes). L'ensemble du réseau est estimé simultanément.

Taille des échantillons :

1. 42 096 individus pour l'échantillon d'apprentissage
2. 42 133 individus pour l'échantillon tests

Construction des labels (apprentissage supervisé) :

1. Utilisation d'un échantillon d'images de 256×256 pixels (résolution de 4 mètres) des départements de la Manche (50) et de la Marne (51).
2. Labellisation : les données du département 50 sont assimilées à du « bocage », celles du département 51 à de l'« openfield ».
3. Une troisième catégorie est introduite, celle des villes : il s'agit des images dont la surface dans une tâche urbaine calculée à partir de la BD TOPO est de plus de 50 %.

Un réseau ConvNet sur les parcelles cadastrales

L'architecture du réseau :

Type de couche	Dimensions (<i>hauteur</i> × <i>largeur</i> × <i>profondeur</i>)
Image en entrée	256 × 256 × 1
Convolution 1 (3 × 3) + ReLu	256 × 256 × 16
Max Pooling 1 (2 × 2)	254 × 254 × 16
Convolution 2 (3 × 3) + ReLu	254 × 254 × 32
Max Pooling 2 (2 × 2)	63 × 63 × 32
Convolution 3 (3 × 3) + ReLu	63 × 63 × 32
Max Pooling 3 (2 × 2)	31 × 31 × 32
Convolution 4 (3 × 3) + ReLu	31 × 31 × 64
Max Pooling 4 (2 × 2)	15 × 15 × 64
Convolution 5 (3 × 3) + ReLu	15 × 15 × 64
Max Pooling 5 (2 × 2)	7 × 7 × 64
Couche de perceptrons 1 + ReLu	128 noeuds
Couche de perceptrons 2 + ReLu	64 noeuds
Couche de perceptrons 3	3 noeuds

Un réseau ConvNet sur les parcelles cadastrales

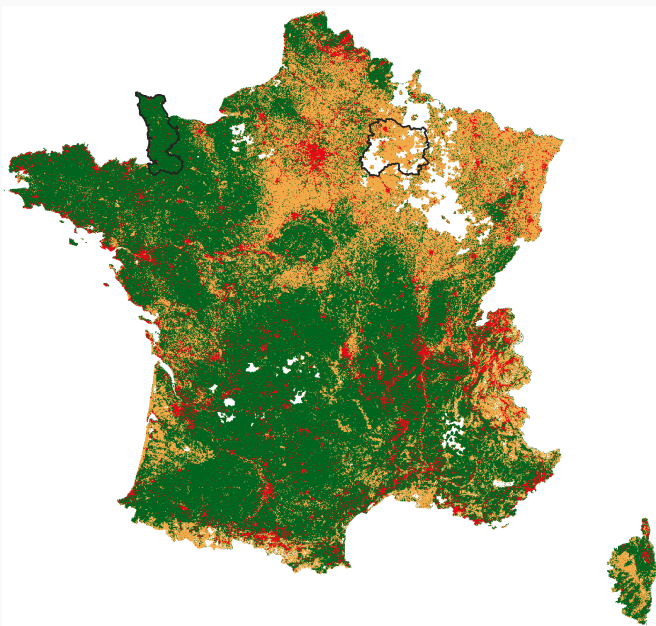
Sur l'échantillon de tests :

- *Pourcentage de biens classés* : 96,5 %

Répartition des images au sein des classes définies par le réseau :

		Labels du réseau		
		0	1	2
Labels des images en entrée	Manche	98,2%	0,7%	1,1%
	Champagne	3,2%	94,7%	2,1%
	Villes	3,4%	0,5%	96,0%

Un réseau ConvNet sur les parcelles cadastrales



Conclusions

Le réseau convolutif semble donner de meilleurs résultats :

- Les zones de « bocage » et d'« openfield » sont plus homogènes.
- Plus de 95% des images de l'échantillon test sont bien classées contre près de 91% pour la première méthode (avec les indicateurs).

Limites :

- La labellisation de l'échantillon d'apprentissage est très sommaire : pas de zones de montagnes. Les parcelles sont assignées à l'un ou l'autre type suivant leur appartenance à un département.
- Le « bocage » et l'« openfield » ne sont pas homogènes.
- Il n'y a pas de déterminisme de l'occupation des sols en fonction de la forme des parcelles.

Améliorer le modèle :

- inclure d'autres catégories : parcelles de montagne, de vignes, de type de bocages ...
- utiliser plus de départements dans l'échantillon d'apprentissage

Utiliser les zones définies pour étudier le mitage :

- Existe-t-il un lien entre les différentes formes d'extension de la ville et la forme des parcelles ?
- Délimiter automatiquement les différentes zones afin d'identifier des frontières
- Analyser les caractéristiques du mitage dans les zones frontières : Y a-t-il plus de mitage dans les zones de « bocage » ? Y a-t-il une influence de la politique locale ?