ENSG Géomatique

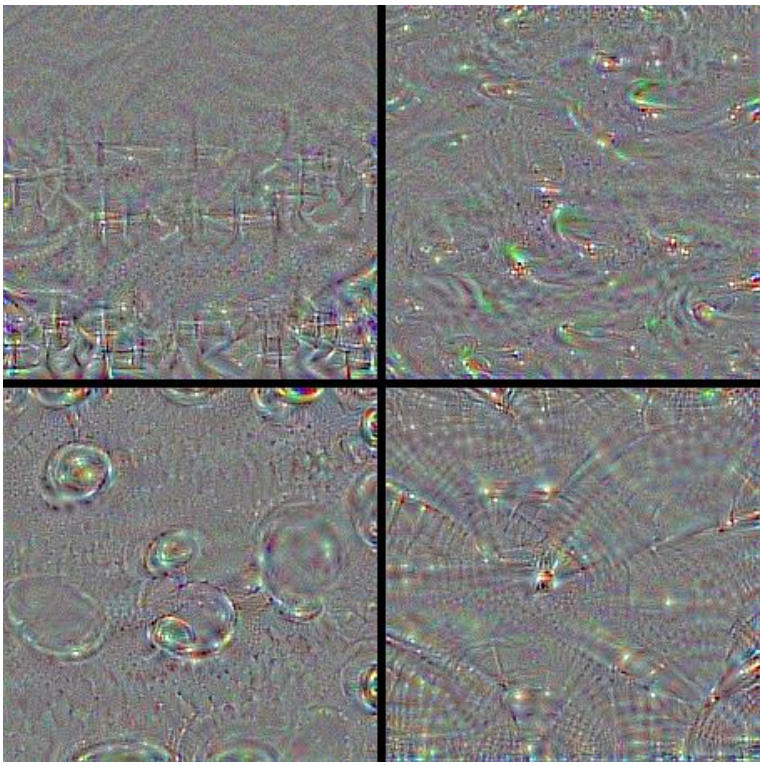
ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES

uOttawa

Rapport de stage

Cycle : ING2

SVM et réseaux neuronaux convolutifs pour la classification de scènes urbaines



Amaury ZARZELLI

le 13 Septembre 2017

☒Non confidentiel ☐ Confidentiel IGN ☐ Confidentiel Industrie jusqu'au ……

Commanditaire :

Michael SAWADA
Laboratory for Applied Geomatics and GIS Science (LAGGISS)
Department of Geography, Environment and Geomatics
60 University Pvt.
Ottawa, ON K1N 6N5
Canada

Encadrement du stage :

Michael SAWADA, LAGGISS, maître de stage
Tristan POSTADJIAN, DIAS, ENSG/IGN, rapporteur principal

Responsable pédagogique du cycle :

Serge BOTTON, DPTS, ENSG/IGN

Stage du 23/05/17 au 19/08/17
Diffusion Web : ☒ Internet ☒ Intranet ENSG
Nombre de pages : 42 dont 1 d’annexes

Remerciements

Je tiens à remercier les personnes qui ont contribué au bon déroulement de ce stage, ainsi qu'à la rédaction de ce rapport.

En premier lieu, je souhaite remercier Hugo BALTZ qui m'a permis de trouver mon stage qui correspondait totalement à mes attentes, et qui m'a apporté de nombreux conseils tout au long de ce stage.

Je remercie aussi vivement mon maître de stage, Dr Michael SAWADA, responsable du service dans lequel j'ai réalisé mon travail, pour son accueil chaleureux, la confiance qu'il m'a accordée ainsi que pour le suivi quasiment quotidien de mon travail, me permettant de tâcher au mieux de correspondre à ses attentes.

Je remercie également le personnel du Département de la Géographie, de l'Environnement et de la Géomatique de l'Université d'Ottawa pour son accueil et son aide concernant l'organisation technique de mon stage.

J'adresse mes remerciements à Tristan POSTADJIAN, mon professeur référent, dont les conseils et l'expertise technique m'ont servi tout au long du stage.

Enfin, je tiens à remercier ma famille et mes amis pour leur soutien infaillible.

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu

Résumé

L'accès de plus en plus répandu à des banques de données d'imagerie urbaine telles que StreetView de Google en corrélation avec le progrès des technologies en apprentissage machine facilite le développement de techniques permettant le traitement automatique des caractéristiques physiques du bâti sur de grandes zones urbaines.

L'une des applications de ce traitement peut être l'étude sociologique, et notamment la mesure de la gentrification, processus par lequel des classes aisées s'installent dans des quartiers historiquement moins favorisés. En effet, ce phénomène se caractérise souvent par une modification de l'aspect des habitations, qui peut être détectée par un modèle de classification.

Ce projet consiste à traiter toutes les étapes de cette classification, du téléchargement de l'imagerie urbaine jusqu'à la cartographie du phénomène étudié et peut être adapté à la qualification de n'importe quel phénomène urbain (accessibilité pour les piétons, structure des bâtiments…). Je me suis en particulier attaché à l'étape de la conception du modèle, en explorant notamment des techniques innovantes basées sur des réseaux de neurones convolutifs et dont les résultats sont prometteurs. Le travail a été effectué sur l'unité urbaine de la ville d'Ottawa au Canada.

L'ensemble des travaux réalisés au cours du projet sont accessibles sur le dépôt GitHub suivant : https://github.com/azarz/gentriNet.

Mots-clefs : Classification, Urbanisme, Gentrification, Ottawa, Réseaux neuronaux, SVM

Abstract

The wider access to databases of street imagery such as Google's StreetView in conjunction with the advances in machine learning allows the development of automated quantification and qualification of urban space over wide regions.

An application for that qualification could be neighbourhood studies and the measurement of gentrification, which is characterised by changes in the physical aspect of housing that can be detected by a classification model.

This project consists in studying all the steps of this classification, from the downloading of imagery to the mapping of the results, and can be adapted to any urban phenomenon (e.g. walkability or building structure). I mostly contributed in the design of the model. I explored innovative ways to build classifiers, in particular convolutional neural networks, with promising results. The area of interest was the city of Ottawa in Canada.

A GitHub repository is provided along with the project. It is accessible through the following URL address: https://github.com/azarz/gentriNet.

Keywords: Classification, Urban planning, Gentrification, Ottawa, ConvNets, SVM

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu

Sommaire

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu

Glossaire et sigles utiles

API	*Application Programming Interface,* interface de programmation applicative. Interface par laquelle un logiciel fournit des services à un autre logiciel.
SVM	*Support Vector Machine*, machine à vecteur de support ou encore Séparateur à Vaste Marge. Type de modèle de classification se basant sur une séparation linéaire entre des ensembles de vecteurs en essayant d'optimiser une marge d'erreur.
CNN	*Convolutional Neural Network,* réseau de neurones convolutif. Modèle de réseau de neurones pour le traitement d'images, basé sur un apprentissage de l'extraction d'attributs par filtres de convolution, puis de la classification.

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu

INTRODUCTION

L'organisme qui m'a accueilli pour le stage pluridisciplinaire de deuxième année est le Laboratoire pour la Géomatique Appliquée et la Science des SIG (LAGGISS), situé dans l'Université d'Ottawa au Canada. L'un des axes de recherche de ce laboratoire est l'étude des quartiers de la ville d'Ottawa. L'un des sujets de ces études est la gentrification, c'est-à-dire l'installation dans des quartiers peu favorisés de classes aisées, qui se traduit souvent par une modification des caractéristiques physiques des habitations.

La mise à disposition par Google de sa banque d'images urbaines StreetView à travers une API rend possible le développement d'outils permettant la classification de scènes urbaines à grande échelle. L'étude « StreetScore » (N. Naik *et al*, 2014)[1] qui présente des cartes de « sécurité perçue » pour plusieurs villes d'Amérique du Nord, en est l'exemple le plus connu.

En parallèle, le développement de nouvelles techniques d'apprentissage machine, appelées réseaux de neurones, et de bibliothèques permettant leur plus grande accessibilité, ainsi que l'accès de plus en plus aisé à de fortes puissances de calcul offrent de nouvelles perspectives quant au traitement de tels sujets.

L'objectif de ce stage a donc été d'apporter au LAGGISS la connaissance des réseaux de neurones convolutifs par l'intermédiaire d'une étude de la gentrification sur la ville d'Ottawa, de manière à ce que les résultats soient exploitables non seulement pour quantifier le phénomène étudié, mais aussi pour permettre aux membres du laboratoire d'aborder d'autres problématiques liées à l'imagerie urbaine en utilisant les CNN.

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu

GESTION DE PROJET

I. Organisation

Afin de m'organiser tout au long de mon stage, j'ai recouru à plusieurs méthodes. J'ai tout d'abord, après la définition précise du sujet, mis au point un planning prévisionnel et des fiches de tâches afin de planifier mon travail. Pour comparer mon avancement avec ce qui était prévu, j'ai tenu un journal de bord dans lequel j'ai consigné toutes les tâches réalisées et entamées durant mes journées. De plus, au moins une fois par semaine, je tenais une réunion avec mon maître de stage afin de rendre compte de mon avancement et de préciser les tâches à venir.

Afin de versionner mes travaux, j'ai utilisé le logiciel Git, en conjonction avec la plateforme GitHub, ce qui a permis et permettra l'accès à mes productions pour mon maître de stage et ses étudiants qui prendront la suite du projet.

II. Planification

Voici ci-après les calendriers prévisionnel et rétrospectif réalisés lors de mon stage. (Figures 1 et 2 page 13)

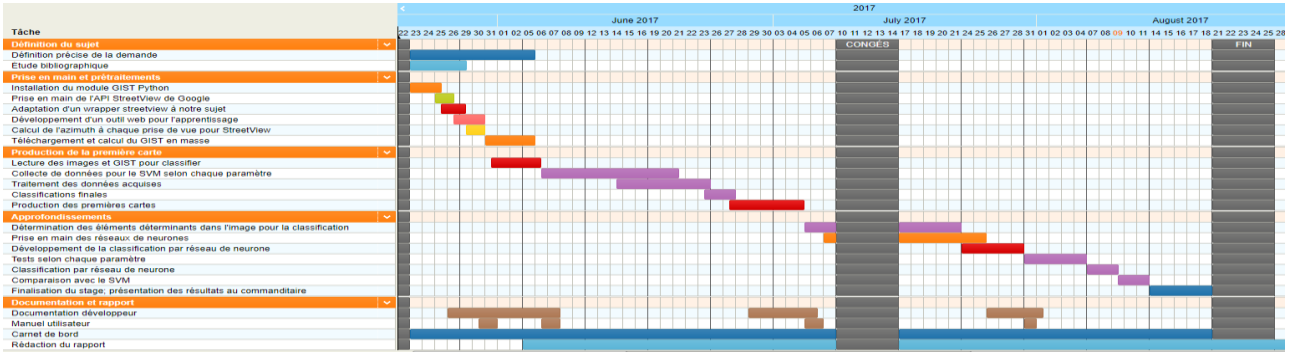


Figure 1 : Calendrier prévisionnel

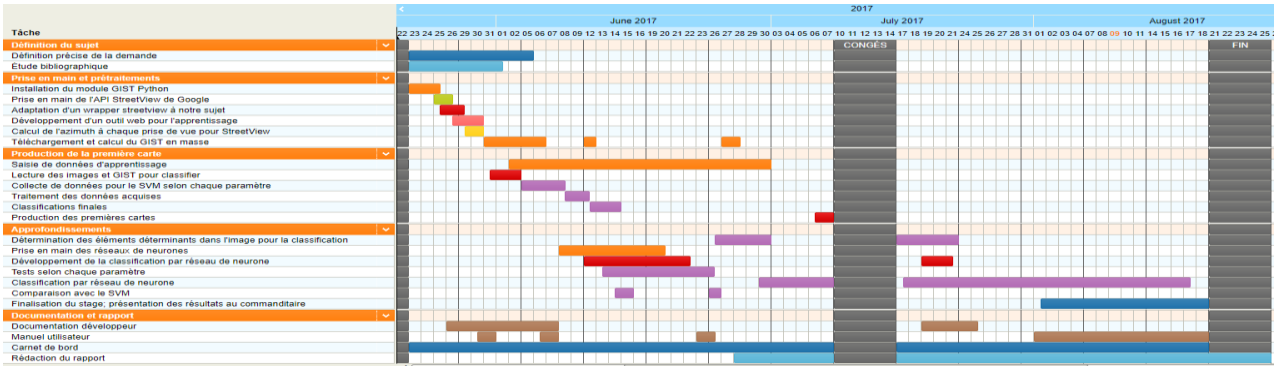


Figure 2 : Calendrier rétrospectif

En comparant ces deux plannings, on remarque que plus l'on s'éloigne du début du stage, plus le déroulement réel du stage est différent de ce qui était prévu.
En premier lieu, l'enchaînement des étapes a été fortement lié aux résultats obtenus lors des premières expériences. Étant donné que ce stage était axé sur la recherche, ces résultats étaient peu, voire non prévisibles *a priori*.
De plus, la plupart des tâches liées aux approfondissements ont été soit surestimées soit sous-estimées dans leur durée. Cela est dû à mon inexpérience initiale dans ce domaine. On note notamment la durée nécessaire à la classification par réseaux de neurones, qui est extrêmement longue, et que j'avais largement sous-estimée. Cela a été compensé par le temps passé sur la classification par SVM qui a été bien plus courte que prévu, et par la prise en main des réseaux de neurones qui a été rapide notamment grâce à l'existence de bibliothèques faciles d'usage pour les novices dont je ne connaissais pas l'existence.
Enfin, du fait des objectifs liés à l'apport de connaissances dans l'organisme d'accueil, la finalisation du stage est un point encore plus important que ce qui a été prévu à l'origine afin d'assurer la bonne transmission des connaissances.

En conclusion, le planning prévisionnel m'a permis, en conjonction avec le carnet de bord et les autres outils que j'ai utilisés, d'atteindre mes objectifs dans le temps imparti. Je note cependant la difficulté de prévoir le calendrier d'un projet axé sur la recherche dans un domaine encore peu exploré. Enfin, j'avais très largement sous-estimé l'importance et la durée de l'acquisition de données d'apprentissage, qui n'est pas spécifique aux réseaux de neurones, mais commun à tous les problèmes de classification.

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu

ÉTAPES DE LA CLASSIFICATION

Dans cette partie, je présente les étapes nécessaires à la classification lors de mon travail, en décrivant les outils mis en jeu et les choix que j'ai réalisés.

I. Téléchargement des images à l'aide de l'API Google StreetView

Afin de pouvoir réaliser un apprentissage puis une classification, il est nécessaire d'avoir des données à apprendre et à classifier. Dans le cadre de mon travail, ces données sont les scènes urbaines, dont la banque de données la plus accessible est celle offerte par le service StreetView de Google. Ces images peuvent être accessibles à l'aide de l'API proposée par Google [2] qui consiste en des requêtes http sous forme d'adresses URL contenant divers arguments, dont la localisation de l'image voulue que l'on peut remplacer par un identifiant unique. Or, si l'on passe par la localisation, seule la dernière image en date sera renvoyée. Pour contourner ce problème, l'utilisateur de GitHub robolyst a proposé un module Python [3], que j'ai adapté au sujet, qui permet par l'intermédiaire d'une autre API de Google d'obtenir les identifiants de toutes les images StreetView pour une position donnée, quelle que soit leur date. J'ai par conséquent utilisé cette méthode de manière automatisée pour télécharger l'imagerie StreetView historique à partir d'un ensemble de points géographiques sur la ville d'Ottawa, définis dans mon cas par un fichier au format ESRI Shapefile de type « point ».

II. Saisie des données d'apprentissage

Afin de réaliser une classification, il faut en premier lieu avoir des données d'apprentissage, c'est-à-dire un ensemble d'échantillons dont on connaît la classe, ou étiquette. Au cours de mon stage, mon superviseur et moi-même avons décidé de deux méthodes de saisie des données d'apprentissage différentes, amenant à deux types de données en entrée puis à deux méthodes de classification. D'un point de vue technique, afin de saisir ces données, j'ai conçu des interfaces web très basiques, hébergées sur mon poste de travail à l'université et donc accessibles sur le réseau local de mon département d'accueil. Ainsi, plusieurs personnes ont participé à la saisie de ces données.

1) Première méthode

La première méthode consiste à afficher, pour chaque localisation de l'ensemble d'apprentissage, tous les couples possibles d'images du même lieu au cours du temps, et à demander à chaque fois s'il y a gentrification ou non. On écrit à chaque validation de l'utilisateur une ligne dans un fichier csv comportant un identifiant pour le couple d'images, les coordonnées du lieu concerné, ainsi que l'étiquette donnée à ce couple correspondant à la présence ou non de gentrification.

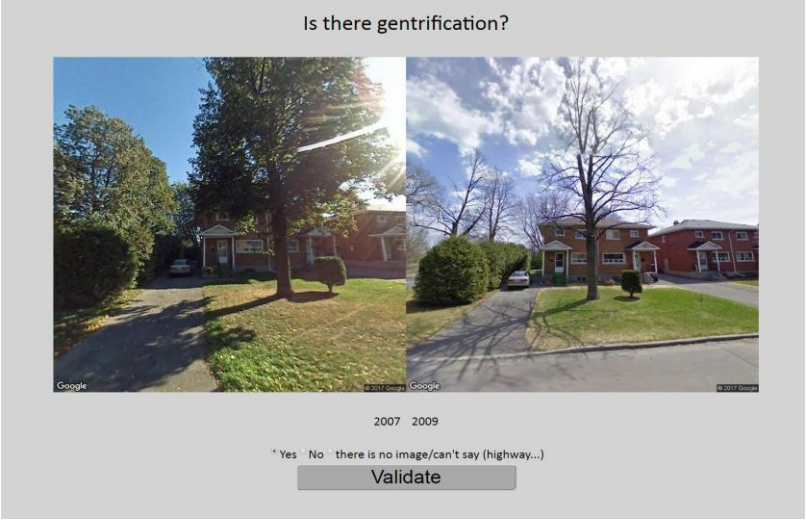


Figure 3 : Première méthode d'acquisition

2) Seconde méthode

Cette seconde méthode est directement inspirée de l'une des études qui a inspiré mon sujet de stage (A. Dubey *et al.* ; 2016)[4], et consiste à afficher deux images côte à côté choisies aléatoirement dans un sous-ensemble d'images. L'utilisateur doit alors indiquer laquelle des deux images semble provenir du voisinage le plus aisé, avec la possibilité de déclarer une « égalité ». À chaque validation, le résultat est stocké sous forme de ligne dans un fichier csv contenant un identifiant pour le couple d'images, l'identifiant Google StreetView de chacune des images, ainsi que le résultat de la comparaison entre ces images.

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu

Figure 4 : Seconde méthode d'acquisition

III. Extraction d'attributs

De nombreuses méthodes de classification se basent sur une représentation vectorielle des objets afin de pouvoir leur donner une étiquette. L'exemple le plus commun en télédétection est la classification de pixels, représentés par le vecteur tridimensionnel défini par les valeurs rouge-vert-bleu. Dans le cadre de mon travail, il s'agit de réaliser une classification d'images entières. Afin de représenter ces images de manière vectorielle, j'ai utilisé deux techniques issues de l'état de l'art, que j'ai choisies du fait de leur importance dans l'étude qui a inspiré mon sujet (N. Naik *et al*, 2014)[1]. Cette extraction d'attributs ne concerne dans mon projet que la classification par SVM.

1) Descripteur GIST

Le descripteur GIST (A. Oliva *et al*. 2001)[5] est un vecteur de dimensionnalité relativement faible (960 valeurs), pouvant décrire de manière assez robuste une scène (M. Douze *et al*, 2009)[6] en se basant sur une description spatiale, structurelle et des textures de l'image. L'équipe de recherche LEAR de l'INRIA a développé une implémentation en C du calcul de ce vecteur [7]. Cette implémentation a été adaptée pour Python sous Linux par l'utilisateur GitHub tuttieee [8], et j'ai moi-même adapté cette implémentation pour qu'elle fonctionne sur l'environnement Windows des ordinateurs de l'université. J'ai implémenté le calcul de ce vecteur de manière automatique au téléchargement des images sur lesquelles j'ai travaillé.

2) Approche par « Sacs de mots visuels »

Au moment où le vecteur GIST a été introduit, l'état de l'art de la description holistique des images se basait sur une approche par « Sac de mots visuels » (« *Bag of visual words* » en anglais) qui

est la suivante : on calcule des indicateurs locaux (par exemple l'indicateur SIFT) sur un ensemble d'images. On réalise ensuite sur ces indicateurs une classification non supervisée afin de les regrouper dans un certain nombre de catégories (ou « mots visuels »). On étudie enfin la distribution statistique des occurrences de chaque mot visuel dans les images afin de classifier lesdites images. Dans le cadre de mon travail, j'ai utilisé le calcul dense de l'indicateur SIFT pour réaliser cette approche.

IV. Entraînement du modèle

À partir des images auxquelles on a donné une étiquette et de leur description vectorielle, on peut entraîner un modèle à partir des caractéristiques communes des représentations d'images étiquetées dans la même catégorie, afin qu'il puisse prédire la classe d'une image dont la catégorie est inconnue.

1) Première méthode : classification

La première méthode de saisie implique une classification « classique » en catégories. En effet, on attribue à chaque couple d'images une étiquette binaire : présence ou non de gentrification. De fait, on cherche ici à trouver un moyen de discriminer les ensembles de couples d'images en deux catégories distinctes afin de pouvoir, à partir des caractéristiques d'un nouveau couple d'images, lui donner un étiquette qui corresponde à la réalité parmi ces deux classes.

La spécificité des données issues de la première méthode d'acquisition est qu'elles sont fortement biaisées : l'écrasante majorité des échantillons d'apprentissage sont négatifs, car les couples d'images sont pris sur l'ensemble de la ville, alors que la gentrification est très localisée. Ainsi, lors de l'entraînement des modèles, il faut réaliser une sélection des échantillons. Cette sélection a été faite de manière aléatoire.

a) Machines à vecteurs de support

Les machines à vecteurs de support (*Support Vector Machines* en anglais, SVM ou encore « séparateurs à vaste marge ») sont des outils mathématiques et informatiques qui permettent la classification dans des espaces de grande dimensionnalité. Ils font partie des classifieurs les plus populaires au sein de l'état de l'art, ce qui justifie leur utilisation au sein de mon travail. Le principe d'un classifieur SVM est de trouver une frontière linéaire (un hyperplan) optimale entre deux groupes de vecteurs ayant une classe définie *a priori* comme différente. À partir de cette frontière,

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu

tout nouveau vecteur se situera d'un côté ou de l'autre de l'hyperplan calculé et pourra donc ainsi être étiqueté.

Le vecteur GIST et les approches par sac de mots fournissent des descripteurs pour une seule image. Or, ce qu'il faut classifier ici, ce sont des couples d'images. Ainsi, afin de caractériser ces couples, j'ai utilisé une concaténation de ces vecteurs.

Afin de déterminer les meilleurs paramètres du modèle, j'ai procédé par validation croisée.

b) Réseaux neuronaux

La technologie des réseaux de neurones est une méthode de plus en plus utilisée pour l'intelligence artificielle et la classification. Le principe de leur fonctionnement est inspiré de celui du cerveau humain. Ce dernier est composé de très nombreuses cellules, les neurones, qui sont capables individuellement d'opérations très simples et qui sont connectés entre eux de manière très dense (c'est-à-dire qu'un neurone est connecté à de nombreux autres neurones). Ce qui permet à un cerveau humain d'apprendre est l'organisation de ces neurones entre eux, et notamment le renforcement de certaines liaisons neuronales au fil d'un apprentissage, afin que de mêmes données en entrées produisent un même résultat (F. Li, J. Johnson, S. Yeung, 2017, Lecture 4)[9].

En intelligence artificielle, un neurone est un opérateur mathématique simple, et un réseau de neurones est un empilement de « couches » constituées de neurones reliées entre elles par des liaisons de neurones. À ces liaisons sont associés des « poids », qui correspondent à des valeurs qui renforcent ou non les connexions entre les neurones.

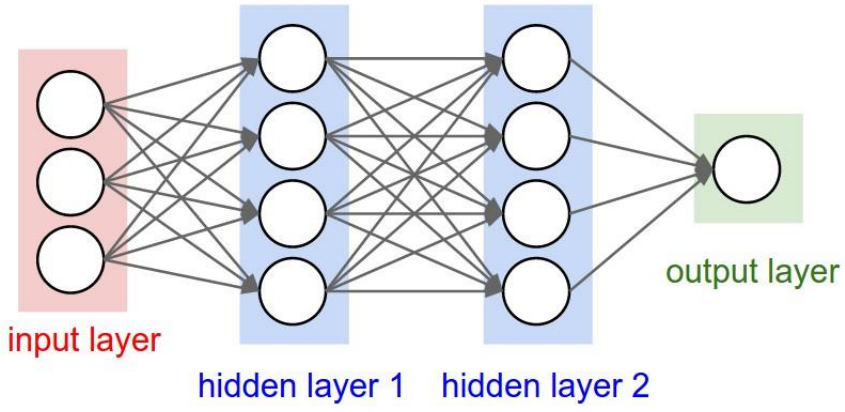


Figure 5 : Schéma d'un réseau de neurones. Chaque cercle est un neurone et chaque flèche une liaison Li, Johnson & Yeung, 2017

Les poids associés à ces liaisons sont généralement initialisés avec des valeurs suivant une distribution aléatoire (loi normale…). Lors de la phase d'apprentissage, on fournit au modèle des échantillons, dont il prédit la classe. Ensuite on lui fournit la réelle classe de l'échantillon et le modèle « ajuste » les poids à l'aide de la rétropropagation du gradient engendré par l'erreur éventuelle. Au fur et à mesure des itérations, et plus le modèle aura fait de mauvaises prédictions avant de se corriger, les poids s'ajustent et les erreurs sont moins fréquentes.

L'une des contraintes principales de ce type de modèle est le volume très important de données nécessaire pour leur phase d'apprentissage, qui dépasse de plusieurs ordres de grandeur les besoins d'un SVM.

Afin d'obtenir des résultats correspondant à l'état de l'art actuel, j'ai utilisé des réseaux de neurones à l'aide de Keras, une API d'apprentissage profond de haut niveau d'abstraction disponible en Python.

Pour les modèles d'apprentissage profond utilisés, les données en entrée ont subi un prétraitement en étant centrées sur zéro : on soustrait à chaque vecteur de l'ensemble d'apprentissage la valeur du vecteur moyen.

Réseaux denses

Un réseau de neurones dense est un réseau dans lequel chaque neurone d'une couche est connecté à tous les neurones de la couche suivante. Il prend en entrée un vecteur, à l'instar du SVM, et fournit en sortie un vecteur de probabilité de classe. Dans le cadre de mon sujet, le problème de classification étant binaire, la dernière couche est composée d'un seul neurone, et la sortie est donc unidimensionnelle, et comprise entre 0 et 1. On interprète cette valeur comme suit : plus la valeur est proche de 0, plus la probabilité, d'après le modèle, d'appartenir à la classe « 0 » (pas de gentrification) est forte, et, inversement, plus la valeur est proche de 1, plus la probabilité d'appartenir à la classe « 1 » (gentrification présente) est forte. Afin d'obtenir une décision pour chaque couple d'images rencontré, on place la valeur seuil à 0,5. Autrement dit, si la valeur de sortie est inférieure à 0,5, on considèrera que le modèle a déterminé la classe du couple comme étant « 0 », et si la sortie est supérieure à 0,5, la classe prédite est « 1 ».

Réseaux convolutifs

Afin de réaliser la classification d'images, la technique de l'état de l'art en 2017 est celle des réseaux de neurones convolutifs (*Convolutional Neural Networks*, ou CNN), qui plutôt que de prendre en entrée un vecteur, prennent directement les images. Cela permet de prendre en

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu

compte moins de paramètres par rapport aux réseaux denses, tout en étant plus adaptés au traitement d'image de par leur nature convolutive. Les premières couches de neurones ne sont donc plus cette fois constituées d'opérateurs mathématiques simples, mais de filtres de convolution (d'où le nom de ce type de réseau). Ainsi, on applique une série de n convolutions à l'image d'entrée, obtenant une image à n canaux, chaque canal correspondant à un filtre. On pourra ensuite réaliser des convolutions sur cette nouvelle image, et ce de manière itérative, afin d'extraire des attributs de hauts niveaux d'abstraction (F. Li, J. Johnson, S. Yeung, 2017, Lecture 5)[9]. L'apprentissage des poids détermine dans ce cas l'importance des différents filtres.

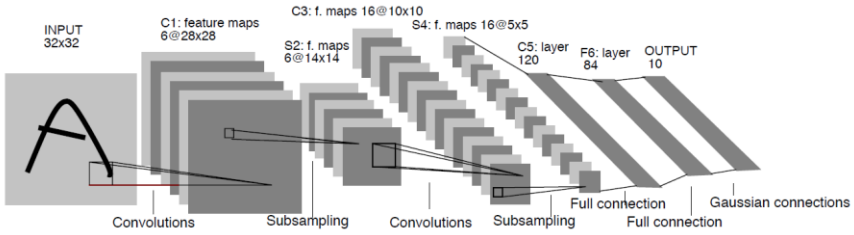


Figure 6 : Schéma d'un CNN

LeCun et al, 1998

Dans le cadre de mon projet, j'ai adopté l'architecture VGGNet-19 (K. Simonyan et A. Zisserman, 2015)[10], que j'ai adapté afin d'obtenir une architecture dite « siamoise » (S. Chopra *et al*, 2005)[11], car c'est celle qui obtient les meilleurs résultats dans l'étude qui a inspiré le sujet (A Dubey *et al*, 2016, p.10)[4]. Les poids ont été initialisés avec les poids d'un modèle pré-entraîné sur la base d'images ImageNet, et ceux concernant le premier bloc de filtres (bas niveaux d'abstraction) ont été figés afin d'éviter un temps d'apprentissage trop long, et du fait des attributs très communs auxquels ils correspondent.

Les dernières couches de neurones de cette architecture constituent un réseau dense, qui à partir des résultats des convolutions, fournit un scalaire compris entre 0 et 1 en sortie. On détermine la classe prédite à partir de cette sorte de la même manière que précédemment, avec un seuil à 0,5.

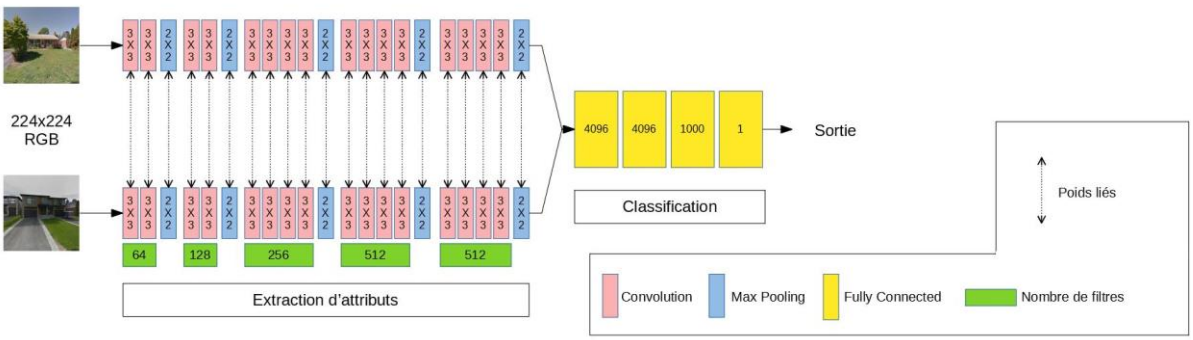


Figure 7 : Architecture utilisée

L'implémentation de tels réseaux neuronaux nécessite une forte capacité de mémoire et de calcul, qui est disponible sur des cartes graphiques (*Graphical Processing Unit* ou encore GPU) puissantes. Au sein de l'Université d'Ottawa, j'ai pu accéder à des cartes Nvidia Tesla K80 par l'intermédiaire d'un serveur de calculs, ce qui m'a permis de réaliser les expérimentations sur les CNN.

2) Seconde méthode : régression

La seconde méthode d'acquisition de données d'apprentissage, inspirée par des travaux antérieurs, nous donne comme résultat un ensemble de duels entre des images et leur résultat. Ces données permettent de réaliser un classement entre les images. À l'instar des rédacteurs du projet StreetScore (N. Naik *et al,* 2014 p.1-2)[1], je me suis servi de l'algorithme Microsoft Trueskill, dont une implémentation existe en Python, pour réaliser le classement. Ledit classement consiste en un score de probabilité de gagner un duel affecté à chaque image, et mis à jour au fur et à mesure des « confrontations ». On peut ensuite étaler ces scores entre 0 et 10 et ainsi donner un score à chaque image. Ce score correspondra à l'étiquette de l'image.

Étant donné que l'étiquette des images est ici déterminée selon une échelle continue et non un ensemble discret de classes, il s'agit d'un problème de régression : le but est ici d'estimer le score des nouvelles images, et non une classe. Les outils employés sont donc différents de ceux utilisés lors d'une classification par catégories.

a) SVR

Il existe une version du SVM adaptée à la régression (*Support Vector Regression*, ou SVR), qui plutôt qu'optimiser un hyperplan formant une frontière entre des vecteurs, optimise un hyperplan correspondant à une fonction affine des vecteurs donnant leur valeur. À l'instar des auteurs du travail qui a inspiré mon stage (N. Naik *et al,* 2014 p.2-3)[[1], j'ai utilisé l'implémentation *nu-SVR* issue de *libsvm* implémentée en Python dans la librairie *scikit-learn*.

b) Réseaux de neurones

On l'a vu plus haut, la sortie des réseaux de neurones est en réalité un scalaire. De fait, il n'y a pas de changement structurel à réaliser sur un modèle entraîné à prédire des classes pour qu'il puisse prédire un score.

J'ai donc utilisé un CNN inspiré de l'architecture VGGNet-19 avec les poids initialisés avec un modèle pré-entraîné sur la base ImageNet pour réaliser la prédiction des scores.

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu

V. Classification

Une fois le modèle permettant de prédire la classe ou le score d'un échantillon entraîné sur un ensemble dit d'apprentissage, sa finalité est de prédire la classe d'un ensemble d'échantillons dont on veut déterminer la classe.

1) Première méthode

La première méthode est une classification de couples d'images du même lieu à des dates différentes. Afin de constituer les entrées pour le modèle, on parcourt l'ensemble des lieux que l'on veut classifier sur la région d'Ottawa, puis on extrait tous les couples possibles d'images pour chaque lieu. Pour les modèles SVM et réseau dense, on réalise une concaténation des vecteurs GIST des images constituant ce couple afin de prendre une décision. Le modèle convolutif siamois est quant à lui à double entrée : un couple d'images.

Le fichier en sortie de la classification contient un ensemble de lignes correspondant à l'ensemble des couples. Les valeurs de chaque ligne sont les coordonnées WGS84, l'année de la première prise de vue, l'année de la seconde prise de vue et la classe (0 ou 1) prédite par le modèle.

2) Seconde méthode

Afin de déterminer la présence ou non de gentrification à l'aide de la seconde méthode, on réalise la comparaison entre les images *a posteriori*, contrairement à la première méthode. Une fois le score de chaque image prédit, on regarde pour chaque lieu s'il y a une amélioration du score au fil du temps. Si c'est le cas, alors il y a gentrification, et sinon, on considère que le modèle n'a pas détecté de gentrification.

RÉSULTATS OBTENUS

Dans cette partie, je présente et interprète les résultats obtenus suite à l'application des étapes décrites dans la partie précédente.

I. Première méthode de saisie

Cette méthode correspond à la classification de couples d'images d'un même lieu à des dates différentes.

1) SVM

Les paramètres testés lors de la validation croisée pour la classification par SVM ont été :

- Le type de noyau, linéaire ou gaussien ;
- Le paramètre C (pour « coût »), qui correspond à l'inverse de la tolérance aux erreurs dans l'ensemble d'apprentissage : si C est faible, beaucoup d'erreurs sont tolérées et la frontière sera « lisse », et si C est fort, aucune erreur de classification ne sera tolérée, avec un risque de sur-apprentissage.
- Pour le noyau gaussien, le paramètre gamma, qui correspond à l'inverse du rayon d'influence des échantillons d'apprentissage lors de la détermination de frontières.

Les meilleurs résultats sont obtenus avec un noyau linéaire et une valeur de C de l'ordre de grandeur de l'unité (dont la valeur optimale peut être déterminée à la volée à chaque classification).

La classification par SVM a assez vite montré ses limites dans le cadre de mon sujet. Je rappelle en effet que dans l'ensemble d'apprentissage, les cas négatifs sont en très large majorité. Or, si l'on essaye de tenir compte de cette dissymétrie lors de l'entraînement du modèle SVM, il aura tendance à classifier tous les échantillons en tant que négatifs pour obtenir une bonne précision : si 80% des échantillons sont négatifs, alors classifier tous les échantillons négativement amènera à une précision de 80%. Or le rappel aura une valeur bien plus faible. Pour pallier ce problème on pourra choisir de réaliser un apprentissage sur un ensemble contenant autant de cas positifs que de cas négatifs, mais cela provoque la mauvaise représentativité de la réalité pour la partie « négative » : en effet, les cas ne présentant pas de gentrification présentent une grande variété (aucun changement en ville, aucun changement en campagne, changements dus à la différence

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu

d'angle entre les prises de vues…), variété qui ne peut pas être totalement prise en compte s'il y a aussi peu de négatifs que de positifs.

C'est pour ces raisons que les scores que j'ai choisi pour qualifier les classifieurs SVM sont le F1-score, qui est un rapport tenant compte de la précision et du rappel pour chacune des classes, ainsi que le score kappa de Cohen qui mesure l'accord entre deux observateurs pour un problème de classification, avec une préférence pour ce dernier du fait de sa facilité de calcul pour tous les autres types de modèles.

Du fait de ce problème, j'ai réalisé l'apprentissage sur un ensemble équitablement réparti entre positifs et négatifs, les négatifs étant choisis à chaque fois aléatoirement parmi l'ensemble des données d'apprentissage. Afin de pouvoir qualifier l'efficacité de la classification, j'ai donc réalisé mes calculs sur un grand nombre d'« itérations », avec à chaque itération un choix différent d'échantillons négatifs.

Utilité des différents attributs

Je rappelle qu'afin de réaliser une classification des couples d'images, j'ai réalisé une extraction d'attributs pour vectoriser chacune des images puis réaliser une concaténation des vecteurs. Les attributs extraits ont été le vecteur GIST et le SIFT dense.

De manière générale, l'apport du SIFT dense dans la précision de classification est très faible, alors que sa prise en compte dans les calculs multiplie par 100 la durée des traitements, du fait de la forte dimensionnalité du vecteur et de la nécessité de réaliser une classification non supervisée à chaque itération. De fait, les résultats suivant sont ceux obtenus à l'aide d'une classification utilisant uniquement le vecteur GIST.

Résultats quantitatifs obtenus

Les résultats suivants sont les moyennes des scores obtenus pour une suite de 1000 itérations, en utilisant uniquement la concaténation des vecteurs GIST.

- F1-score : 0,59 (écart-type 0,09)
- Kappa : 0,2927 (écart-type 0,11)

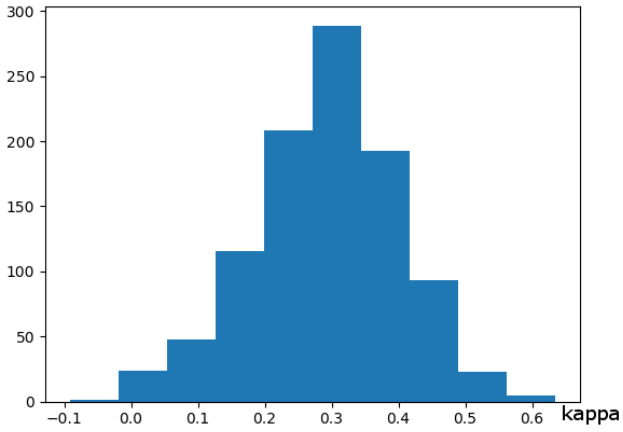


Figure 8 : Histogramme du nombre de classifieurs parmi les 1000 itérations en fonction du kappa obtenu

On constate sur la figure 8 une distribution normale du score kappa autour de la moyenne située à 0,3 environ, ce qui correspond à un accord faible. Le F1-score de 0,59 en moyenne est très peu supérieur à une prévision aléatoire, qui donnerait 0,5, notamment lorsqu'on tient compte des marges d'erreurs. De plus, encore une fois, étant donné que le modèle SVM n'est entrainé que sur une faible portion de l'ensemble d'apprentissage, sa capacité à prédire tous les cas de manière exacte est très faible.

N'étant pas satisfait des résultats apportés par cette méthode, je me suis donc tourné vers des techniques de classification plus proches de l'état de l'art actuel, les réseaux neuronaux, en commençant par les réseaux denses.

2) Réseaux neuronaux denses

Suite à divers tests préliminaires, la structure que j'ai adoptée est une structure comprenant une couche en entrée, une couche dite « cachée », et un neurone de sortie. Les paramètres que j'ai cherché à optimiser à l'aide de la validation croisée sont les suivants :

- Le nombre de neurones dans chacune des deux premières couches ;
- Le taux d'apprentissage, qui est une valeur agissant sur la mémoire à long terme du modèle ;
- Le type d'algorithme d'optimisation pour la descente du gradient ;
- Le nombre d'époques d'apprentissage pour chaque itération, une époque correspondant à une présentation des échantillons au modèle ;

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu

- Le mode d'initialisation aléatoire ;
- La contrainte sur la norme maximale des poids ;
- Le « taux d'abandon » (*Dropout rate*), d'une couche spéciale dite de « Dropout » qui vise à limiter le sur-apprentissage.

Je ne détaillerai pas ici l'ensemble des résultats obtenus suite à cette validation croisée, qui dépendent fortement du problème rencontré.

L'avantage du réseau de neurones comparativement au SVM est le fait que la prise en compte de plusieurs ensembles est possible. Aussi, à chaque itération sélectionnant un nouvel ensemble d'apprentissage, les poids du réseau s'ajustent un peu plus pour tendre vers le résultat voulu.

Cependant, les résultats obtenus à l'aide de ce modèle ne sont pas meilleurs que ceux obtenus précédemment.

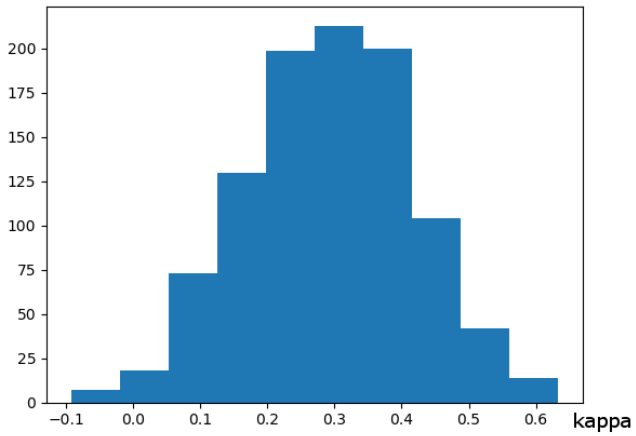


Figure 9 : Histogramme du nombre de classifieurs parmi les 1000 itérations en fonction du kappa obtenu

On remarque encore une fois sur la figure 9 une distribution normale des scores kappa, autour d'une moyenne de 0,2964 avec un écart-type de 0,12.

Afin d'obtenir de meilleurs résultats, et étant donné que mon sujet est un problème de classification d'images, je me suis tourné vers la technique actuelle utilisée dans ce genre de problème : les réseaux convolutifs, qui, plutôt que d'apprendre une classification à partir d'attributs pré-extraits, apprend en plus l'extraction d'attributs optimaux pour un problème donné.

3) Réseaux neuronaux convolutifs

J'ai réalisé l'entraînement des réseaux convolutifs avec les proportions entraînement-validation-test respectives suivantes : 55,6% - 11,1% - 33,3%. Le score kappa est calculé en mesurant l'accord entre les étiquettes prédites par le modèle sur l'ensemble de test et ses étiquettes réelles.

Les réseaux convolutifs que j'ai entraînés ont fait montre d'un problème récurrent : le sur-apprentissage et la présence de faux positifs trop nombreux. En effet, lors des premiers tests, le score kappa obtenu sur l'ensemble de test était situé autour de 0,95, ce qui est trop élevé et est un indice de sur-apprentissage.

Afin de pallier ce problème, une fois la classification terminée, j'ai parcouru manuellement les résultats du modèle en ajoutant les faux positifs en tant que cas négatifs et les vrai positifs en tant que positifs au fichier d'apprentissage, afin d'entraîner un nouveau modèle. J'ai procédé de la sorte de manière itérative. Ainsi, l'ensemble d'apprentissage, qui contenait à l'origine très peu de cas positifs, a été augmenté, tout en prenant en compte les précédentes erreurs de classification afin d'éviter de les réitérer par la suite.

Pour encore réduire les problèmes de sur-apprentissage, j'ai fortement réduit le taux d'apprentissage du modèle et augmenté le nombre d'itérations, et donc d'époques d'apprentissage. Avec cette méthode, j'ai obtenu au bout de 12 augmentations du jeu de données un score kappa de 0,82, ce qui est très bon sans pour autant sembler être du sur-apprentissage. Il reste cependant après cela de nombreux cas de faux positifs. En effet, le modèle est trop sensible aux changements, ce qui limite par ailleurs le nombre de faux négatifs. Dans le cadre de mon projet, qui est une recherche de toute la gentrification dans la ville d'Ottawa, le commanditaire favorise le rappel (la détection de tous les cas positifs) à la précision (l'absence de faux positifs). De plus, du fait de la plus petite proportion de positifs, il est plus facile de calculer un taux de faux positifs *a posteriori* qu'un taux de faux négatifs.

Le dernier modèle a été entraîné sur 500 itérations (nouveau choix aléatoire de cas positifs et négatifs) de 10 époques chacune, ce qui a duré sur le GPU Nvidia Tesla K80 de l'université environ 110 heures.

Visualisation du modèle

La particularité des modèles de classification par réseaux de neurones convolutifs est que les attributs sont extraits par les filtres de convolution du modèle, qui sont donc optimisés lors de la phase d'apprentissage, et ce de manière spécifique à chaque problème. De plus, du fait de leur

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu

nature convolutive, on peut aisément visualiser les filtres impliqués dans le modèles et ce de différentes manières.

Les visualisations que j'ai réalisées ont été faites sur la partie convolutive du dernier modèle siamois entraîné au cours de mon stage.

La première est la visualisation des filtres (F. Chollet, 2016)[12] : lorsqu'un modèle est entraîné et les poids liés aux filtres optimisés, on peut, à partir d'une image générée en bruit gaussien, maximiser l'activation des filtres afin d'obtenir une représentation picturale de ces filtres et des motifs qu'ils détectent. J'ai réalisé cette opération sur la dernière couche de filtres, car elle correspond au plus haut niveau d'abstraction des motifs recherchés par les filtres (par exemple, un filtre de bas niveau détectera des contours et un filtre de haut niveau des plumes). Cela permet de constater les motifs recherchés discriminants dans une image.

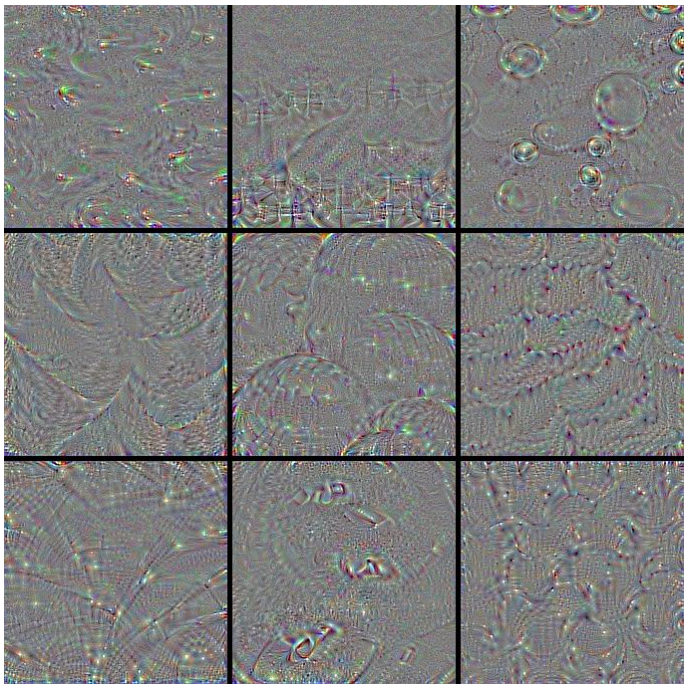


Figure 10 : Visualisation des 9 filtres les plus discriminants de la dernière couche de convolution, triés par ordre d'importance de gauche à droite puis de haut en bas

Parmi les 9 filtres les plus discriminants que j'ai pu calculer, on peut constater les motifs qui permettent de distinguer les parties des images qui correspondent à la gentrification. Par exemple, sur le deuxième filtre, la présence de contours horizontaux, verticaux et diagonaux sur le bas de l'image (sol) est importante, de même que la présence de cercles dans l'image (troisième filtre). Une analyse poussée de ces filtres et leur comparaison avec les images en jeu dans la classification pourra amener à une meilleure compréhension des motifs caractérisant la

gentrification. Le cinquième filtre pourra par exemple correspondre à la cime de buissons. Il faut toutefois noter que la discrimination ne se fait pas ici entre des images, mais bien entre des couples d'images. Cela rend l'analyse des filtres plus complexe.

Une autre manière de visualiser le modèle, et notamment son comportement par rapport aux images en entrée, est de visualiser les activations du modèle. À partir d'une image en entrée, pour chaque filtre de la dernière couche de convolutions du modèle, on détecte les zones de l'image qui correspondent aux motifs recherchés par le filtre.

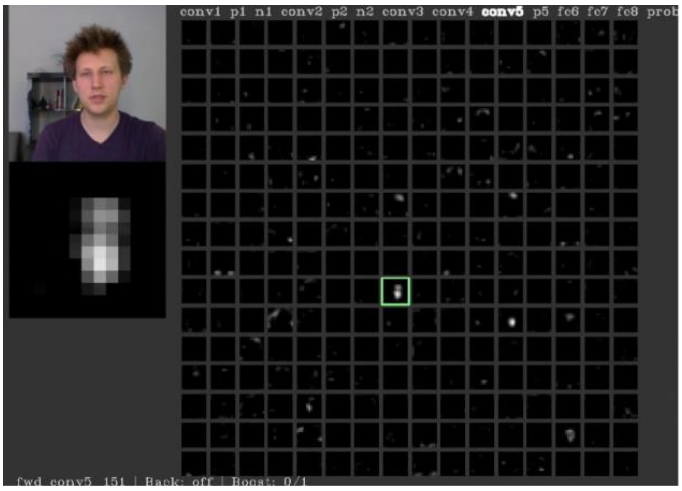


Figure 11 : Exemple d'activation pour un filtre pour la détection de visages
Yosinski et al, 2014

En faisant une somme de toutes les activations de la dernière couche de filtres, on peut réaliser une carte des activations du modèle sur une image donnée. On peut ensuite comparer les différences de ces cartes de chaleur entre un cas positif est un cas négatif, permettant de constater les éléments discriminants pour le modèle dans une image. Sur les figures suivantes, le bleu correspond à la valeur minimale et le rouge à la valeur maximale de la somme de ces activations.

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu

Figure 12 : Carte des activations pour un cas négatif

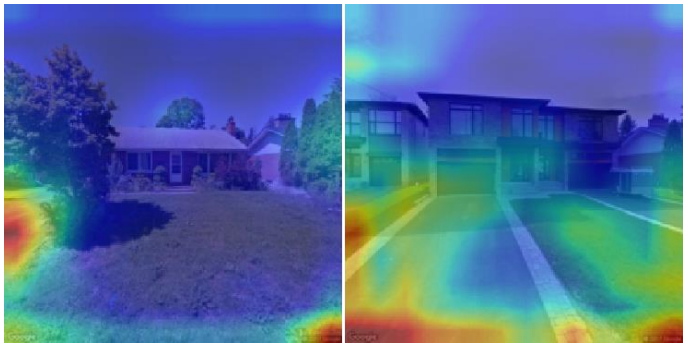


Figure 13 : Carte des activations pour un cas positif

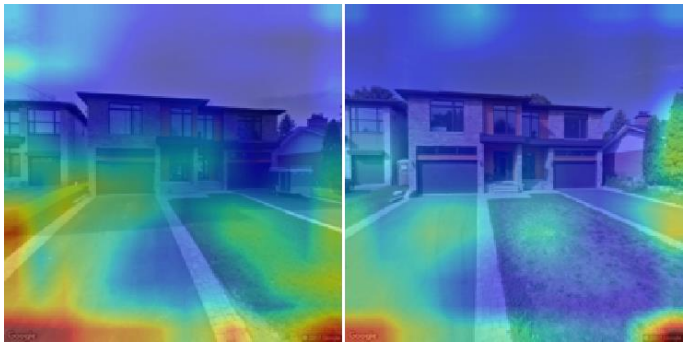


Figure 14 : Carte des activations pour un cas négatif

Le modèle semble donner une réponse plus forte pour les images qui présentent des bâtiments après gentrification. Or, étant donné que la classification se fait sur des couples, les cas positifs ne sont que ceux dont la seconde image présente une réponse forte, et les autres cas (réponse forte ou réponse faible pour les deux images du couple) sont négatifs. De plus, on constate que les zones qui activent le plus les filtres de convolution sont situés sur les zones de sol et de bâti.

De fait, les cas de faux positifs que l'on peut trouver peuvent être liés à la suppression de végétation verticale, comme sur le cas suivant.

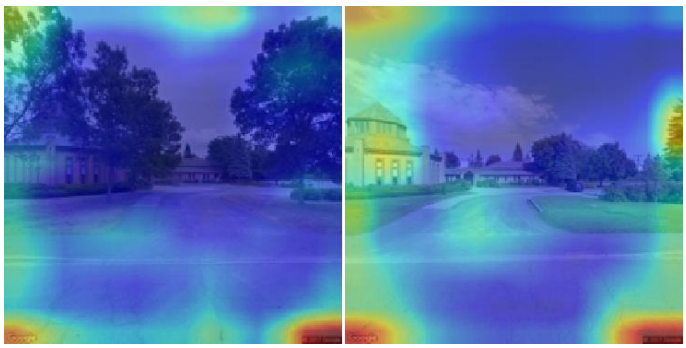


Figure 15 : Carte des activations pour un faux positif

On remarque de plus un effet de bord : les extrémités des images sont souvent plus activent souvent plus les neurones que leur centre.

Améliorations éventuelles

Afin d'obtenir de meilleurs résultats, il y a plusieurs pistes à explorer :

- Les cas de nouvelles constructions de pavillons en banlieue étaient ici considérés comme des cas positifs. On pourra essayer de les considérer comme des cas négatifs.

- Une autre approche au problème serait non pas d'utiliser une architecture siamoise, mais une architecture simple qui ferait une classification d'une seule image, résultant d'une fusion des deux images du couple. On peut par exemple penser à une soustraction des valeurs RGB pour chaque pixel.

- Enfin, on pourra essayer de tenir compte du contexte géographique de chaque prise de vue, en incluant par exemple un poids lié à la position des photos. Pour réaliser cela au sein du réseau neuronal, on pourra au moment de la fusion des deux « branches » de l'architecture siamoise ajoutée des métadonnées à chaque vecteur de sortie.

II. Seconde méthode de saisie

L'ensemble d'apprentissage de la seconde méthode de saisie est obtenu en comparant des couples de scènes urbaines. Dans l'étude StreetScore, le score de chaque image estimé *via* Trueskill est stable après environ 16 duels par image (N. Naik *et al,* 2014 p.2)[1]. Dans le cadre de mon stage, étant donné le peu de personnes pouvant participer à l'apprentissage des données, et un nombre assez important de types d'images à comparer, j'ai totalisé 21 920 « clics », amenant à une moyenne de 11 clics par image, ce qui est insuffisant pour obtenir des scores stables.

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu

De fait, lorsqu'on essaye de prédire les notes d'un ensemble d'images avec le modèle SVR, on ne constate aucune corrélation entre note prédite et note réelle.

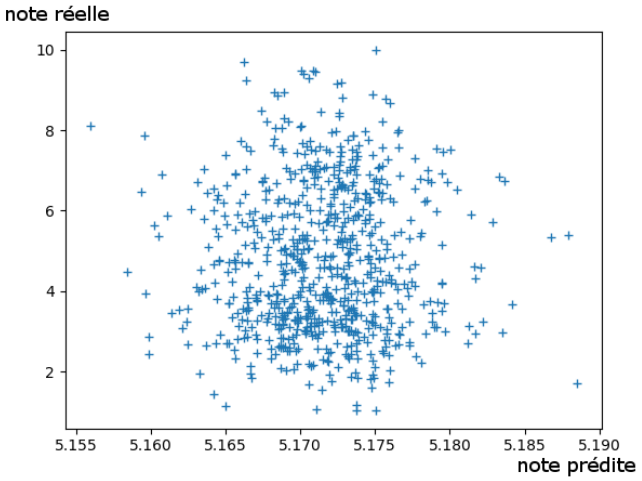


Figure 16 : Note réelle en fonction de la note prédite. Le cas idéal serait la fonction identité y=x

Le modèle par réseau de neurones convolutif ne donne pas de meilleurs résultats. En effet, le problème ne vient pas du modèle, mais bien des données d'apprentissage qui sont trop peu nombreuses.

III. Cartographie des résultats

L'un des objectifs initiaux de mon sujet de stage était de produire une carte de la gentrification sur la ville d'Ottawa à partir des résultats obtenus par classification. Je présente ici comment j'ai procédé pour réaliser cette carte, qui a été obtenue avec le dernier modèle CNN entraîné au cours de mon stage.

Le responsable du LAGGISS m'a fourni pour télécharger l'imagerie urbaine sur toute la ville d'Ottawa un fichier de formes de type ponctuel dans lequel chaque entité représente une habitation. De fait, le jeu d'images qui est à disposition pour la classification recouvre densément toute l'unité urbaine d'Ottawa. Il suffit donc de classifier tous ces points pour obtenir une idée de la répartition de la gentrification sur le territoire étudié.

À chaque couple d'image dont le modèle prédit la classe est associé ses coordonnées WGS84, soit la latitude et la longitude. Du fichier de sortie qui contient tous les résultats de classification, on extrait tous les couples prédits comme positifs. L'ensemble ainsi obtenu est un nuage de points que l'on peut charger par exemple dans QGIS.

Afin d'obtenir un rendu qui rend compte des modifications des scènes urbaines potentiellement multiples au cours du temps, et d'avoir une idée des zones dans lesquelles cette gentrification a lieu, j'ai choisi de représenter ce nuage de points à l'aide d'une carte de chaleur. Cette représentation permet également une représentation efficace à plusieurs échelles.

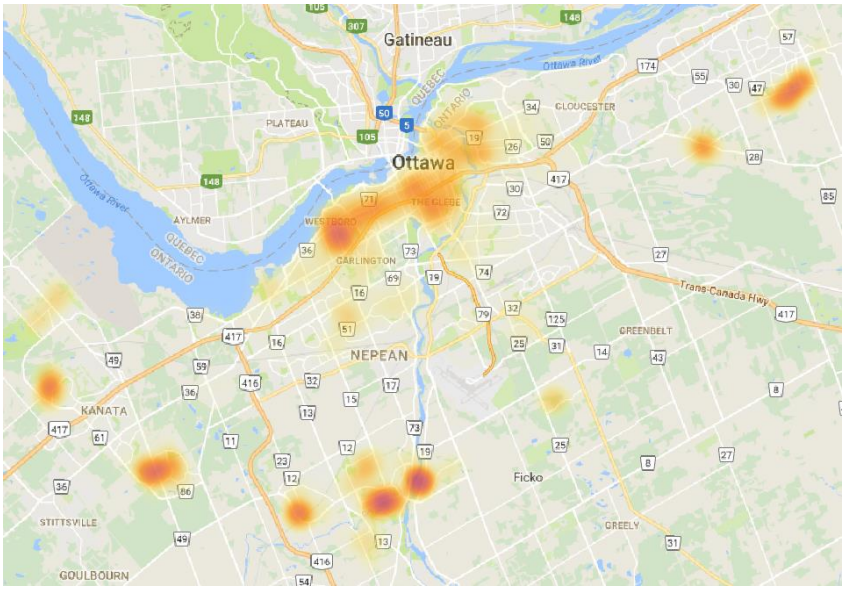


Figure 17 : Carte de chaleur de la gentrification sur l'ensemble de l'aire urbaine d'Ottawa

Fond de carte par Google Maps, 2017

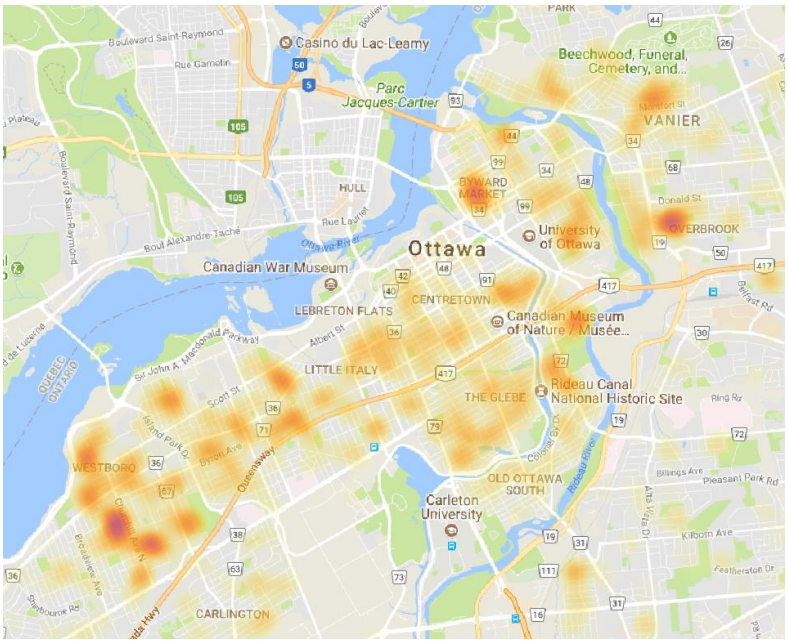


Figure 18 : Carte de chaleur de la gentrification sur la partie centrale de la ville

Fond de carte par Google Maps, 2017

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu

La première carte (figure 17) comprend de nombreux cas positifs en périphérie de la ville. Cela est dû aux constructions de nouveaux arrondissements, qui sont détectés comme de la gentrification. En effet, un terrain vague se transforme en habitation en l'espace de deux photos.

Un changement d'échelle centré sur la ville historique d'Ottawa permet de déterminer plus précisément les quartiers détectés comme en gentrification. On constate une corrélation plutôt forte avec la réalité, notamment dans les quartiers au sud-ouest et au nord-est de la ville qui sont effectivement considérés en pleine gentrification par les sociologues et les agents immobiliers.

CONCLUSION

Concernant son objectif principal, qui est la prise en main pour le LAGGISS de la technologie des réseaux de neurones convolutifs, le stage est une réussite. En effet, par l'intermédiaire de la problématique concernant la gentrification, j'ai pu comprendre et expliquer le fonctionnement de tels modèles et fournir des exemples de programmation. L'apport de cette connaissance assure la continuité du projet et l'amélioration de ses résultats, ainsi que son adaptation à d'autres problématiques.

Cela étant dit, la carte qui a été produite à la fin de la durée de mon stage, bien que prometteuse, est encore insuffisante pour qualifier la gentrification de manière précise sur la ville d'Ottawa. Cependant, les modèles par réseaux de neurones convolutifs ont montré de très bons résultats sur d'autres problématiques abordées par le laboratoire, notamment la classification des bâtiments selon leur structure, qui est une information utile entre autres pour la prévention sismique.

Le fait d'avoir adopté différentes méthodes pour la classification m'a permis, après avoir fait le constat que l'une d'entre elle prenait trop de temps, de me concentrer sur l'une des deux. Cela offre une autre perspective de recherche pour les personnes qui prendront ma suite dans le projet, qui est de terminer le travail entamé sur la seconde méthode d'acquisition.

D'un point de vue personnel, ce stage m'a permis de travailler sur des techniques à la pointe de la recherche, et ce dans un domaine qui m'intéresse particulièrement. La connaissance des réseaux de neurones est une nouvelle compétence acquise durant ce stage, et me permet de mieux comprendre le fonctionnement de certains logiciels récents. J'ai de plus acquis une connaissance du monde de la recherche et de son fonctionnement, et appris à travailler sur un projet pendant une durée longue et continue.

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu

BIBLIOGRAPHIE

[1]

Nikhil Naik, Jade Philipoom, Ramesh Raskar, César Hidalgo. “Streetscore - Predicting the Perceived Safety of One Million Streetscapes”, *CVPR Workshop on Web-scale Vision and Social* Media, 2014

http://streetscore.media.mit.edu/static/files/streetscore_paper.pdf

[2]

Référence de l’API Google StreetView. [30/05/2017],

https://developers.google.com/maps/documentation/streetview/

[3]

Module Python permettant le téléchargement d’images StreetView historiques. [07/06/2017],

https://github.com/robolyst/streetview/

[4]

Abhimanyu Dubey, Nikhil Naik; Devi Parikh, Ramesh Raskar, César Hidalgo. “Deep Learning the City: Quantifying Urban Perception at a Global Scale”, *European Conference on Computer Vision*, 2016

https://arxiv.org/pdf/1608.01769v2.pdf

[5]

Aude Oliva, Antonio Torralba. “Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope”, *International Journal of Computer Vision 42(3) 145–175*, 2001 http://cvcl.mit.edu/Papers/IJCV01-Oliva-Torralba.pdf

[6]

Matthijs Douze, Hervé Jégou, Sandhawalia Harsimat, Laurent Amsaleg, Cordelia Schmid. “Evaluation of GIST descriptors for web-scale image search”, *International Conference on Image and Video Retrieval,* 2009

http://www.quaero.org/media/files/bibliographie/inria_cpr6_douze_gist_evaluation.pdf

[7]

Liste des logiciels programmés par l’équipe LEAR [16/08/2017],

https://thoth.inrialpes.fr/software

[8]
Implémentation en Python du calcul du vecteur GIST. [25/05/2017],
https://github.com/tuttieee/lear-gist-python

[9]
Fei-Fei Li, Justin Johnson, Serena Yeung, “CS231n: Convolutional Neural Networks for Visual Recognition.” [02/08/2017],
Cours de l’Université de Stanford décrivant les réseaux de neurones et leur utilisation à l’aide de Python
http://cs231n.stanford.edu/

[10]
Karen Simonyan, Andrew Zisserman. “Very Deep Convolutional Networks for large-scale image recognition”, *International Conference on Learning Representations,* 2015
http://www.quaero.org/media/files/bibliographie/inria_gpr6_douze_gist_evaluation.pdf

[11]
Sumit Chopra, Raia Hadsell, Yann LeCun “Learning a Similarity Metric Discriminatively, with Application to Face Verification” *Computer Vision and Pattern Recognition,* 2005
http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1467314

[12]
François Chollet, “How convolutional neural networks see the world” [03/07/2017]
https://blog.keras.io/how-convolutional-neural-networks-see-the-world.html

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu

ANNEXES

L’ensemble des sources et des documentations produites durant le projet sont disponibles sur le dépôt GitHub suivant : https://github.com/azarz/gentriNet.

Table des matières

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu

Liste des figures

ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES
6 et 8 avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 MARNE-LA-VALLÉE CEDEX 2
01 64 15 31 00 – stages@ensg.eu