



HAUTE ÉCOLE
D'INGÉNIERIE ET DE GESTION
DU CANTON DE VAUD
www.heig-vd.ch

HEIG-VD

RAPPORT INTERMÉDIAIRE

La Terre de nuit vue de l'espace

Antoine FRIANT
Haute École d'Ingénierie et de Gestion du
Canton de Vaud
Yverdon-les-Bains, VD, CH
antoine.friant@gmail.com

7 juin 2018

Table des matières

1	Cahier des charges	ii
1.1	Résumé du problème	ii
1.2	Objectifs	ii
1.3	Limitations	ii
1.4	Description fonctionnelle	1
1.5	Délais	1
2	Introduction	2
3	Exploration des données	3
3.1	Jeux de données	3
3.1.1	Images satellite	3
3.1.2	Grilles de population	5
3.1.3	Pays	5
3.2	Recherche de corrélation	7
3.3	Données à explorer	7
4	Modèle neuronal	8
4.1	Réseau de neurones	8
4.2	Environnement de développement	8
4.3	Métaparamètres	8
5	Conclusion	9
6	Authentification	11
7	Symboles et abréviations	12
8	Annexes	14

1 Cahier des charges

1.1 Résumé du problème

Les données géographiques sont nécessaires pour la prise de décisions importantes. Cependant la fiabilité et la disponibilité de ces données ne sont pas homogènes dans le temps et selon le lieu. Certaines de ces données ont une forte corrélation avec la lumière perçue par les satellites pendant la nuit.

Grâce à l'apprentissage automatique (*machine learning*), il est possible d'entraîner un réseau de neurones sur des données d'une date et d'un lieu connus pour reconstituer une carte de données géographiques à partir d'une image satellite nocturne.

Le travail à effectuer consiste à explorer différents types de données géographiques afin d'en choisir un, et faire de la prédiction sur ce type de données grâce à un réseau de neurones.

1.2 Objectifs

Le TB consiste dans un premier temps à explorer les données suivantes :

- Images satellites nocturnes de la Terre,
- Population humaine,
- Population animale,
- Densité végétale,
- PIB,

Et toutes autres données jugées pertinentes dans le but d'entraîner un réseau de neurones capable de prédire une estimation d'une donnée utile, à partir d'une image satellite de la terre de nuit.

La réalisation d'une application qui entraîne et exploite ce réseau de neurones est l'objectif de la seconde partie du TB.

Le but final est de pouvoir estimer, grâce au machine learning, des informations dont on ne possède pas de données à jour. Et cela à partir d'images satellites de nuit récentes, ou d'une combinaisons de ces images avec une autre donnée à jour.

1.3 Limitations

L'application sera compatible avec Windows 10 et Archlinux, et nécessitera l'installation de librairies tierces (telles que Keras et TensorFlow). Elle ne possèdera pas nécessairement d'interface utilisateur.

L'utilisateur sera responsable de fournir les données à l'application dans un format supporté.

1.4 Description fonctionnelle

L'application prend en argument au moins deux jeux de données géographiques de format imposé : une image satellite nocturne et un autre type de donnée à déterminer au cours du projet. Après un long temps d'entraînement (une semaine au maximum, dépend de la machine utilisée), un modèle est généré.

Une fois le modèle généré, il est sauvegardé et réutilisable sur une autre image satellite nocturne (d'une date et/ou d'une région différente). Lorsque le modèle est appliqué sur une image satellite, une carte est recrée, affichant le résultats des prédictions.

Par exemple, si au cours du travail de bachelor il s'avère que la population par kilomètre carré est une donnée utile et utilisable, l'application devra prendre en argument une image satellite nocturne ainsi qu'une carte des populations de même taille et de même résolution pour entraîner le réseau de neurones. Une fois le modèle généré, l'application devra être capable de régénérer une approximation de la carte de population par kilomètre carré à partir d'une image satellite.

1.5 Délais

15 juin 2018 : Rapport intermédiaire

27 juillet 2018 : Rapport final et application fonctionnelle

Entre le 3 et le 14 septembre 2018 : Soutenance du travail de bachelor

2 Introduction

Les produits d'imagerie satellite sont devenus abondants et largement accessible au cours des vingt dernières années. De nombreux satellites prennent des photographies de la Terre à chaque heure du jour *et de la nuit*. Ces observations nocturnes révèlent des caractéristiques peu évidentes de jour, parfois même cachées. Les routes apparaissent, les villes montrent leurs lumières, même les bateaux de pêche aveuglent les océans avec des projecteurs pour attirer les poissons.

La disponibilité, la résolution et l'uniformité de la qualité de ces données contraste fortement avec le manque de fiabilité d'autres informations géographiques utiles lors de prises de décisions importantes. Par exemple, la densité de la population est une estimation précise en Suisse mais très approximative au Kenya. D'autres mesures intéressantes incluent : la consommation en électricité, les émissions de CO_2 , la couverture végétale et la présence de faune. Les lumières nocturnes observées depuis l'espace donnent des indications sur chacune de ces mesures alors qu'elles peuvent manquer dans une région à une date donnée.

Le but de ce projet est d'extraire autant d'information que possible de l'imagerie satellite nocturne en utilisant l'apprentissage automatique (*machine learning*) sous la forme de réseau de neurones.

3 Exploration des données

3.1 Jeux de données

3.1.1 Images satellite

NASA Worldview

La première source de données explorée est l'application "Worldview" de la NASA [2]. Elle permet de visionner un grand nombre d'images satellite composites sur un globe en trois dimensions (voir figure 3.1). Parmi les jeux de données disponibles sont trois jeux d'images nocturnes.

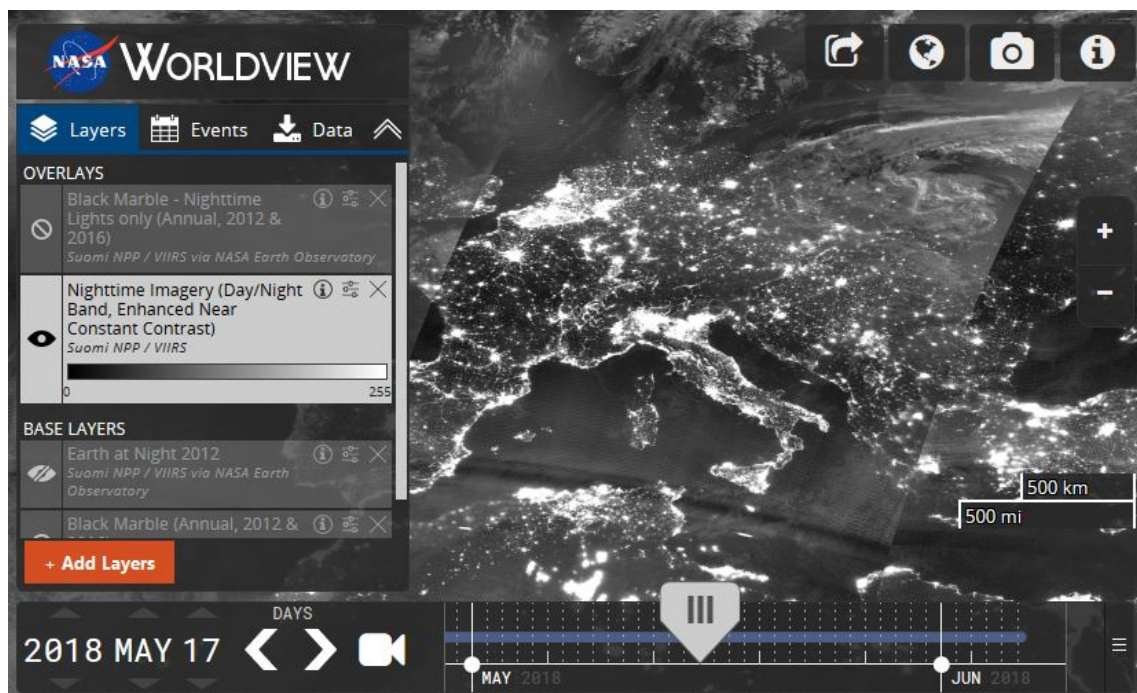


FIGURE 3.1 – Outil de visualisation NASA Worldview [2].

Le premier jeu de données est série d'images composites capturées par le satellite Suomi NPP opéré par la NASA, la NOAA et le Département de la Défense des États-Unis. Il est mis à jour toutes les quelques heures, et présente une image composite chaque jour depuis le 30 novembre 2016. Elle possède deux défauts éliminatoires : la période d'observation actuellement disponible (à peine plus d'une année) n'est pas suffisamment longue pour observer une évolution significative des villes depuis l'espace, et les images ne sont pas traitées. Cela signifie que celles-ci sont très fortement bruitées par les nuages et la lumière ambiante due aux différentes phases de la Lune.

Les deux autres jeux de données nocturnes sont des images composites : des clichés pris tout

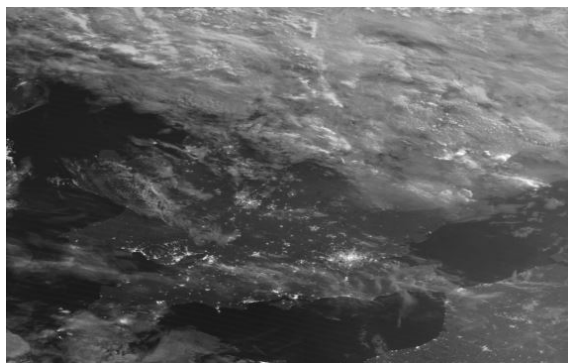


FIGURE 3.2 – Image satellite quotidienne servie par NASA Worldview [2], représentant la Grande Bretagne et son climat nuageux.

au long de l'année ont permis de fabriquer une seule image du globe dont la luminosité ambiante est constante (moyennée) et sur laquelle les nuages n'apparaissent pas. Malheureusement, l'outil Worldview ne permet pas un téléchargement direct de ces images *dans leur pleine résolution*. Heureusement, la NASA a mis à disposition une API REST pour télécharger des "tuiles" de n'importe laquelle de leur image. Seulement le format PNG est disponible. Ce format ne contient pas d'information géographique, ce qui complique leur utilisation pour la suite de ce travail. Un script Python suffit pour télécharger et assembler les tuiles (figure 3.3) pour reconstituer une image complète du globe de plus de 800 millions de pixels (figure 3.4).

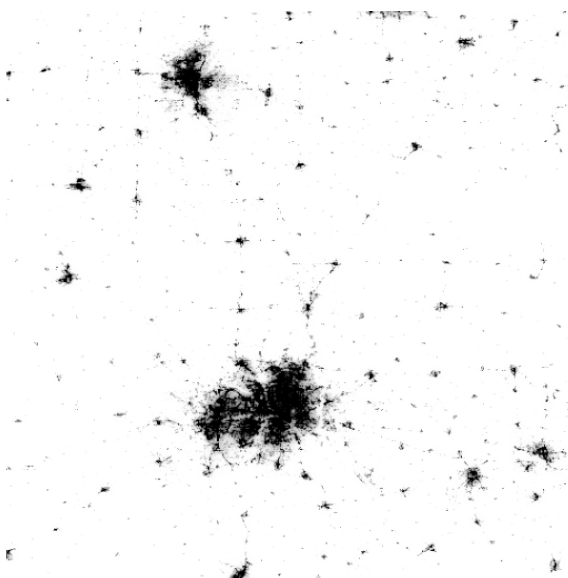


FIGURE 3.3 – Une tuile de l'image de 2016 [nasa-api] montrant la ville de Dallas (USA) après avoir été mise en couleurs négatives.

Le script python utilisé se trouve dans Code/scraper/generate_all.py et nécessite l'installation de la librairie Pillow pour le traitement des images, ainsi qu'urllib pour le téléchargement en soi. Son exécution peut demander plus d'une heure pour le téléchargement (vitesse limitée par le serveur), et plus de 4 Go de RAM pour l'assemblage des tuiles.

Agence américaine d'observation océanique et atmosphérique

La source d'images satellite retenue pour la suite du travail est celle de l'Agence américaine d'observation océanique et atmosphérique (abrégée NOAA)[noaa]. Des images satellites nocturnes composites ("Average Lights X Pct") sont disponibles pour les années 1992 à 2013, une période suffisante pour observer des changements depuis l'espace. De plus, ces données sont disponibles



FIGURE 3.4 – Image globale annuelle (2016) reconstituée à partir de tuiles téléchargées [nasa-api], puis mise en couleurs négatives.

en format GeoTIFF, qui contient les informations géographiques nécessaires pour superposer cette carte à sur autre. Il est donc possible d'explorer et manipuler ces cartes à l'aide de logiciels libres tels que QGIS. Il s'agit en réalité des mêmes clichés fournis par Worldview (et Google Earth), mais plus nombreux et dans un format plus exploitable.

Ces images ont été créées en moyennant la valeur de luminosité de chaque pixel sur une année, en ignorant les pixels couverts par des nuages, et en multipliant cette moyenne par la fréquence de détection de lumière sur le pixel au cours de l'année.

3.1.2 Grilles de population

Sedac [1] met à disposition des grilles de populations pour le monde entier, sous forme de fichier GeoTIFF. Chaque "case" de 1 km^2 est représentée comme un pixel et contient une estimation du nombre de personnes vivant dans cette case. QGIS s'est à nouveau montré d'une grande aide pour visualiser et manipuler ces données volumineuses (exemple en figure 3.5).

On remarque que la valeur de densité de population ne varie pas à l'intérieur d'un sous-région. Tous les pixels d'une grande ville ou d'un département possèdent la même valeur. Ces données ont été créées à partir de plusieurs mesures lorsque le décompte de la population pour une sous-région n'a pas été considéré fiable. Les images satellites font d'ailleurs partie de ces mesures supplémentaires.

Ces cartes sont disponibles pour les années 2000, 2005, 2010, 2015 et 2020.

3.1.3 Pays

Les grilles de données globales sont difficiles à créer, il n'en existe donc pas pour tous les types et toutes les dates. Afin de contourner ce problème, il existe des outils pour combiner des grilles (ici : l'image satellite) avec des données vectorielles telles que les frontières des pays.

Le jeu de données TODO [naturalearthdata] contient les frontières des pays à TODODATE telles que (TODO : politique de choix des frontières), ainsi que des méta-données sur chacun de ces pays (population estimée, indice de développement, différentes appellations et abréviations, etc.). Voici à quoi ressemble ces méta-données pour la Tunisie :



FIGURE 3.5 – Extrait de la grille de population [1] rendue par QGIS. Le blanc indique une absence d'habitants, le noir indique au moins 1000 habitants par mètre carré.

```
( 'scalerank', 0), ( 'featurecla', 'Admin-0 country'), ( 'LABELRANK', 3.0), ( 'SOVEREIGNT', 'Tunisia'), ( 'SOV_A3', 'TUN'), ( 'ADM0_DIF', 0.0), ( 'LEVEL', 2.0), ( 'TYPE', 'Sovereign country'), ( 'ADMIN', 'Tunisia'), ( 'ADM0_A3', 'TUN'), ( 'GEOU_DIF', 0.0), ( 'GEOUNIT', 'Tunisia'), ( 'GU_A3', 'TUN'), ( 'SU_DIF', 0.0), ( 'SUBUNIT', 'Tunisia'), ( 'SU_A3', 'TUN'), ( 'BRK_DIF', 0.0), ( 'NAME', 'Tunisia'), ( 'NAME_LONG', 'Tunisia'), ( 'BRK_A3', 'TUN'), ( 'BRK_NAME', 'Tunisia'), ( 'BRK_GROUP', None), ( 'ABBREV', 'Tun.'), ( 'POSTAL', 'TN'), ( 'FORMAL_EN', 'Republic of Tunisia'), ( 'FORMAL_FR', None), ( 'NAME_CIAWF', 'Tunisia'), ( 'NOTE_ADM0', None), ( 'NOTE_BRK', None), ( 'NAME_SORT', 'Tunisia'), ( 'NAME_ALT', None), ( 'MAPCOLOR7', 4.0), ( 'MAPCOLOR8', 3.0), ( 'MAPCOLOR9', 3.0), ( 'MAPCOLOR13', 2.0), ( 'POP_EST', 11403800.0), ( 'POP_RANK', 14.0), ( 'GDP_MD_EST', 130800.0), ( 'POP_YEAR', 2017.0), ( 'LASTCENSUS', 2004.0), ( 'GDP_YEAR', 2016.0), ( 'ECONOMY', '6. Developing region'), ( 'INCOME_GRP', '3. Upper middle income'), ( 'WIKIPEDIA', -99.0), ( 'FIPS_10_', 'TS'), ( 'ISO_A2', 'TN'), ( 'ISO_A3', 'TUN'), ( 'ISO_A3_EH', 'TUN'), ( 'ISO_N3', '788'), ( 'UN_A3', '788'), ( 'WB_A2', 'TN'), ( 'WB_A3', 'TUN'), ( 'WOE_ID', 23424967.0), ( 'WOE_ID_EH', 23424967.0), ( 'WOE_NOTE', 'Exact WOE match as country'), ( 'ADM0_A3_IS', 'TUN'), ( 'ADM0_A3_US', 'TUN'), ( 'ADM0_A3_UN', -99.0), ( 'ADM0_A3_WB', -99.0), ( 'CONTINENT', 'Africa'), ( 'REGION_UN', 'Africa'), ( 'SUBREGION', 'Northern Africa'), ( 'REGION_WB', 'Middle East & North Africa'), ( 'NAME_LEN', 7.0), ( 'LONG_LEN', 7.0), ( 'ABBREV_LEN', 4.0), ( 'TINY', -99.0), ( 'HOMEPART', 1.0), ( 'MIN_ZOOM', 0.0), ( 'MIN_LABEL', 3.0), ( 'MAX_LABEL', 8.0)
```

Le script `Code/regionvertime/dumpstats.py` superpose ces données vectorielles à l'image satellite pour ajouter à chaque pays sa luminosité moyenne sur une échelle de 0 à 63 ainsi que l'écart-type de luminosité par pixel. Puis ces données sont enregistrées dans le fichier `stats.pickle` pour être utilisées plus tard.

Grâce à cette information de luminosité perçue par pays, on peut faire un parallèle avec une grande variété de données, dont le produit intérieur brut (GDP), la consommation en énergie, le niveau de développement, l'indice économique, les émissions de CO₂, etc.

3.2 Recherche de corrélation

Recherche de corrélation, graphes

Problème de format pour la superposition des rasters

GPD, groupement par pays, filtré par développement économique, graphes

deltas de luminosité par pays, incomplet

plot syrie, image corée du nord

3.3 Données à explorer

Données encore pas explorées : human footprint, ...

4 Modèle neuronal

4.1 Réseau de neurones

réseau de neurones à convolution pour faire de la régression => chercher des cas de problèmes similaires

4.2 Environnement de développement

installation : keras, tensorflow, setup nvidia

4.3 Méta-paramètres

méta-paramètres du réseau de neurones

5 Conclusion

Travail effectué, travail restant

Bibliographie

- [1] *Gridded Population of the World, Version 4 (GPWv4). Population Count Adjusted to Match 2015 Revision of UN WPP Country Totals, Revision 10.* CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK-CIESIN-COLUMBIA UNIVERSITY. URL : <https://doi.org/10.7927/h4jq0xzw> (visité le 06/06/2018).
- [2] *Worldview.* NASA EOSDIS. URL : <https://worldview.earthdata.nasa.gov/> (visité le 06/06/2018).

6 Authentication

7 Symboles et abréviations

Table des figures

3.1	Outil de visualisation NASA Worldview [2].	3
3.2	Image satellite quotidienne servie par NASA Worldview [2], représentant la Grande Bretagne et son climat nuageux.	4
3.3	Une tuile de l'image de 2016 [nasa-api] montrant la ville de Dallas (USA) après avoir été mise en couleurs négatives.	4
3.4	Image globale annuelle (2016) reconstituée à partir de tuiles téléchargées [nasa-api], puis mise en couleurs négatives.	5
3.5	Extrait de la grille de population [1] rendue par QGIS. Le blanc indique une absence d'habitants, le noir indique au moins 1000 habitants par mètre carré.	6

8 Annexes