

# Classification de données massives de télédétection

---

Nicolas Audebert

Encadrant : Bertrand Le Saux (ONERA/IVA)

Directeur de thèse : Sébastien Lefèvre (Univ. Bretagne-Sud, IRISA)

Soutenance de thèse – mercredi 17 octobre 2018

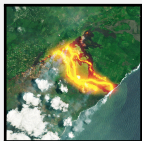


## Contexte

---

# Pourquoi observer la Terre ?

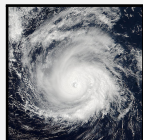
## Gestion des catastrophes



Intervention des secours après un cataclysme, aide humanitaire...

CNRS Le Journal (2018), « La cartographie au service des secours »

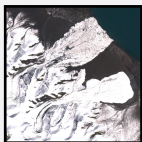
## Météorologie



Mesure des vents, des courants, de la température, du champ magnétique...

La Tribune (2018), « Avec le satellite Aeolus d'Airbus, Météo-France va mieux mesurer les vents à l'échelle mondiale »

## Changement climatique

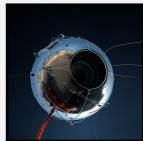


Fonte des glaces, mesures de pollution atmosphérique...

Le Monde (2017), « Un satellite franco-israélien pour lutter contre le réchauffement climatique »

La Croix (2018), « L'Observatoire spatial du climat prend son envol »

## Suivi des activités humaines



Contrebande, pêche ou déforestation illégales, expansion urbaine...

BBC (2018), « NovaSAR : UK radar satellite launches to track illegal shipping activity »

The Guardian (2016), « New satellite mapping a 'game changer' against illegal logging »

# Problématiques techniques

- ▶ Outil précis : équivalent à l'humain
- ▶ Outil adapté aux capteurs : infrarouge, multispectral
- ▶ Outil multimodal : plusieurs images d'une même zone
- ▶ Outil à large couverture : variété des environnements terrestres

L'interprétation doit permettre de renseigner sur :

- ▶ L'occupation des sols (trouver les routes, les bâtiments, les points d'eau...)
- ▶ L'usage des sols et les comportements associés (zones résidentielles et leur densité de population, industries lourdes, surfaces agricoles...)



# Volume de données

En 2017, 620 satellites opérationnels sont listés sous l'appellation "observation de la Terre".<sup>1, 2</sup>

Chaque jour, les satellites Sentinel-2 acquièrent 6 To d'images pour un total de 1 Po en 2020. La Terre entière est imagée tous les 5 jours.<sup>3</sup>

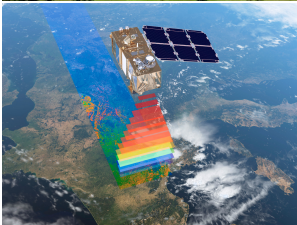
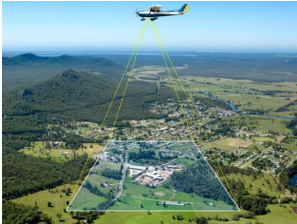
High-altitude aircraft and satellites will constitute a major primary data acquisition source in the future and will be generating vast amounts of imagery suitable for photomapping. In fact, photomapping would appear to be the only way to take reasonable advantage of these future data sources.

Cartography 1950-2000, Robinson et al, 1977, Transactions of the Institute of British Geographers

1. UCS Satellite Database
2. Pixalytics, « Earth Observation satellites in space in 2017 »
3. Sentinel Data Access Annual Report

# Le traitement automatisé des données

Trois problématiques scientifiques apparaissent :

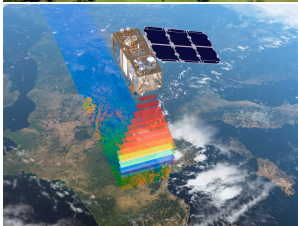
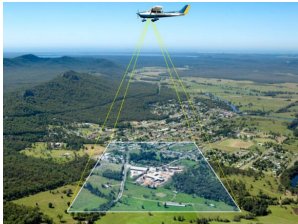


- ▶ Peut-on déléguer à la machine l'extraction de connaissances à partir des images d'observation de la Terre ?
- ▶ Est-il possible de combiner différentes sources d'images pour enrichir les connaissances ?
- ▶ Comment tirer profit de l'afflux constant de données ?

**Solution envisagée :** vision par ordinateur et apprentissage automatique.

# Le traitement automatisé des données

Trois problématiques scientifiques apparaissent :

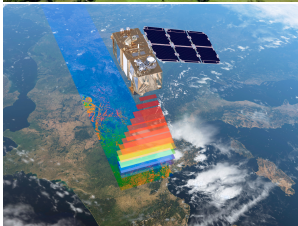
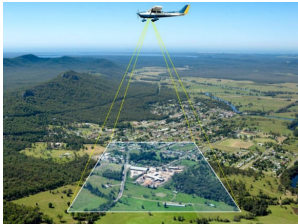


- ▶ Peut-on déléguer à la machine l'extraction de connaissances à partir des images d'observation de la Terre ?
- ▶ Est-il possible de combiner différentes sources d'images pour enrichir les connaissances ?
- ▶ Comment tirer profit de l'afflux constant de données ?

**Solution envisagée :** vision par ordinateur et apprentissage automatique.

# Le traitement automatisé des données

Trois problématiques scientifiques apparaissent :

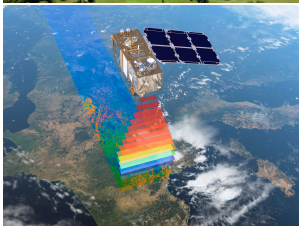
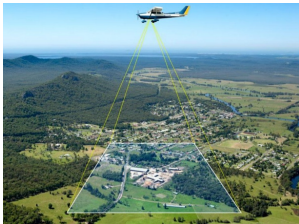


- ▶ Peut-on déléguer à la machine l'extraction de connaissances à partir des images d'observation de la Terre ?
- ▶ Est-il possible de combiner différentes sources d'images pour enrichir les connaissances ?
- ▶ Comment tirer profit de l'afflux constant de données ?

Solution envisagée : vision par ordinateur et apprentissage automatique.

# Le traitement automatisé des données

Trois problématiques scientifiques apparaissent :



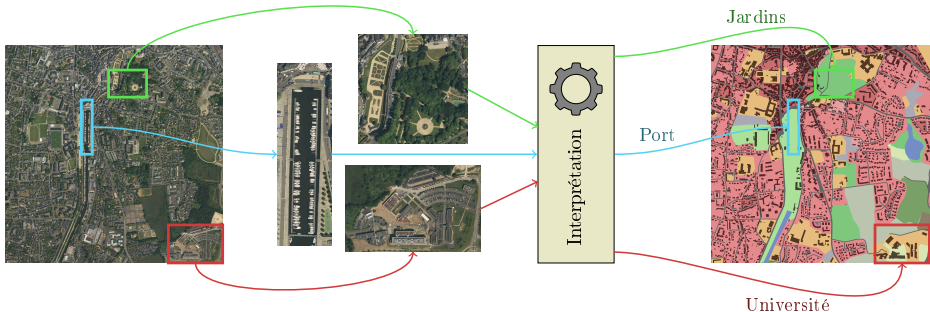
- ▶ Peut-on déléguer à la machine l'extraction de connaissances à partir des images d'observation de la Terre ?
- ▶ Est-il possible de combiner différentes sources d'images pour enrichir les connaissances ?
- ▶ Comment tirer profit de l'afflux constant de données ?

**Solution envisagée :** vision par ordinateur et apprentissage automatique.

# Cartographie d'images aériennes

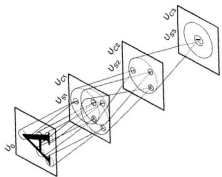
## Objectif

Réaliser automatiquement la cartographie d'une zone (urbaine ou rurale) à partir d'acquisitions aériennes ou satellites.



# Les réseaux de neurones profonds pour la vision par ordinateur

- ▶ 1980 : Neocognitron (reconnaissance de formes) (Fukushima et al.)
- ▶ 1989 : rétropropagation du gradient (Werbos 75, réintroduit par LeCun et al.)
- ▶ 2012 : implémentation sur GPU (Krizhevsky et al.)
- ▶ 2012 : challenge ImageNet (1000 classes, 1 000 000 d'images) (Deng et al.)

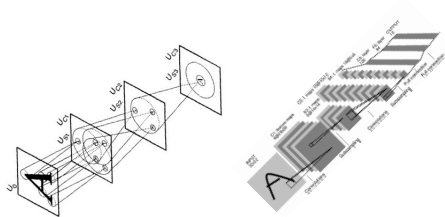


## Objectif

Explorer les architectures de réseaux de neurones profonds pour résoudre les problématiques de télédétection.

# Les réseaux de neurones profonds pour la vision par ordinateur

- ▶ 1980 : Neocognitron (reconnaissance de formes) (Fukushima et al.)
- ▶ 1989 : rétropropagation du gradient (Werbos 75, réintroduit par LeCun et al.)
- ▶ 2012 : implémentation sur GPU (Krizhevsky et al.)
- ▶ 2012 : challenge ImageNet (1000 classes, 1 000 000 d'images) (Deng et al.)



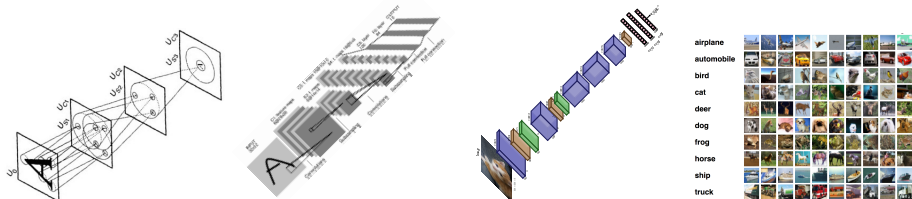
## Objectif

Explorer les architectures de réseaux de neurones profonds pour résoudre les problématiques de télédétection.



# Les réseaux de neurones profonds pour la vision par ordinateur

- ▶ 1980 : Neocognitron (reconnaissance de formes) (Fukushima et al.)
- ▶ 1989 : rétropropagation du gradient (Werbos 75, réintroduit par LeCun et al.)
- ▶ 2012 : implémentation sur GPU (Krizhevsky et al.)
- ▶ 2012 : challenge ImageNet (1000 classes, 1 000 000 d'images) (Deng et al.)

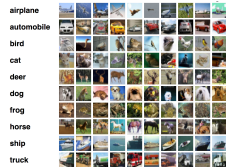
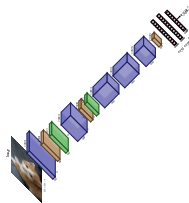
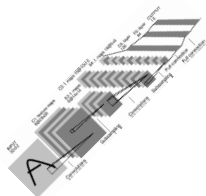
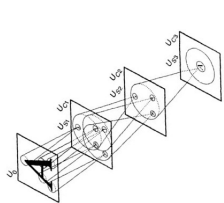


## Objectif

Explorer les architectures de réseaux de neurones profonds pour résoudre les problématiques de télédétection.

# Les réseaux de neurones profonds pour la vision par ordinateur

- ▶ 1980 : Neocognitron (reconnaissance de formes) (Fukushima et al.)
- ▶ 1989 : rétropropagation du gradient (Werbos 75, réintroduit par LeCun et al.)
- ▶ 2012 : implémentation sur GPU (Krizhevsky et al.)
- ▶ 2012 : challenge ImageNet (1000 classes, 1 000 000 d'images) (Deng et al.)



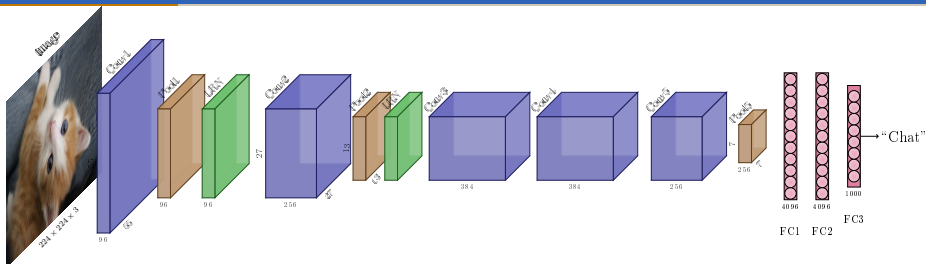
## Objectif

Explorer les architectures de réseaux de neurones profonds pour résoudre les problématiques de télédétection.

# Apprentissage profond pour l'imagerie aérienne

---

# État de l'art : approches par *patch*



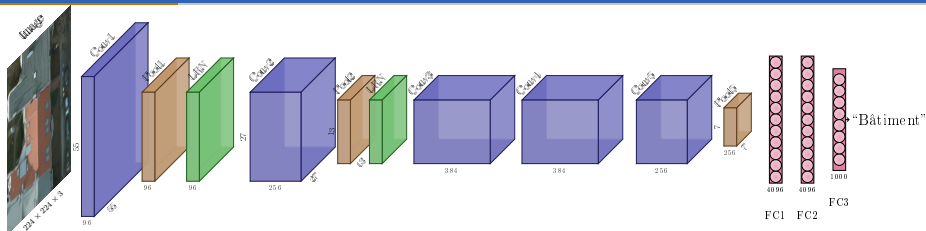
Krizhevsky et al., *ImageNet classification with deep Convolutional Neural Networks*, NIPS 2012

## État de l'art pré-thèse

Approche par *patch* : CNN pour la classification pixel à pixel

- ▶ Lent (incalculable pour des images 10 000 × 10 000)
- ▶ Susceptible au bruit poivre et sel  
→ nécessite un lissage a posteriori

# État de l'art : approches par *patch*



Penatti et al., *Do Deep Features Generalize from Everyday Objects to Remote Sensing and Aerial Scenes Domains?*, CVPRW 2015

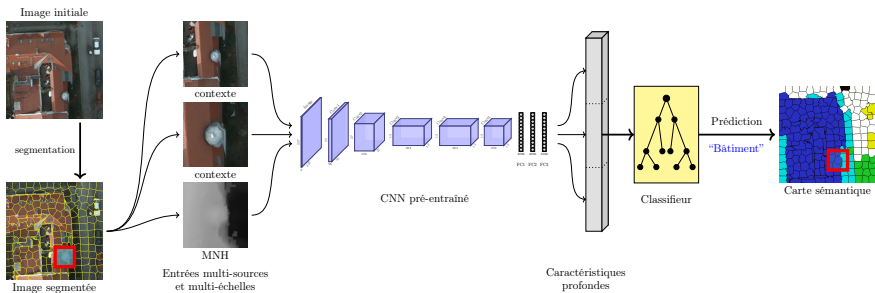
## État de l'art pré-thèse

Approche par *patch* : CNN pour la classification pixel à pixel

- ▶ Lent (incalculable pour des images  $10\,000 \times 10\,000$ )
- ▶ Susceptible au bruit poivre et sel  
→ nécessite un lissage a posteriori

Vakalopoulou et al., *Building detection in VHR multispectral data with deep learning features*, IGARSS 2015

# État de l'art : approches par région



## État de l'art pré-thèse

1. Pré-segmentation
2. Extraction d'attributs (experts ou profonds)
3. Classification par un modèle statistique

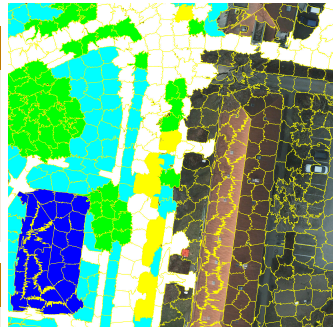
# Limites de l'état de l'art

## Choix des attributs

Les attributs “experts” sont insatisfaisants : spécifiques à une scène, complexes à choisir. Les attributs profonds les remplacent avantageusement.

## Impact de la segmentation

La pré-segmentation est limitante : non-supervisée, elle introduit des erreurs et des irrégularités impossibles à corriger.



Légende : blanc : routes, bleu : bâtiments, cyan : végétation basse, vert : arbres, jaune : véhicules, rouge : autre

## Publication associée

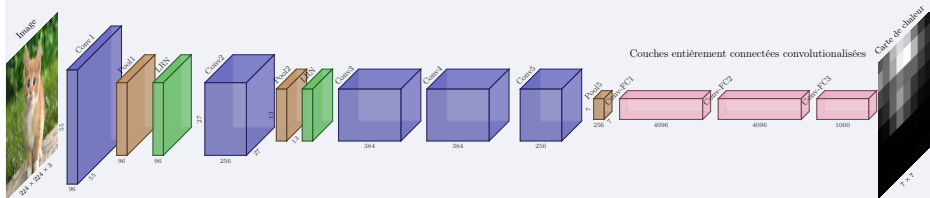
*How useful is region-based classification of remote sensing images in a deep learning framework?*, N. Audebert, B. Le Saux, S. Lefèvre, IGARSS 2016.

# Cartographie par classification dense

Le classifieur doit générer des prédictions **denses** de façon **efficace**.

- ▶ les CNN réalisent déjà une extraction de caractéristiques spatialement denses
- ▶ il suffit de convolutionaliser les couches entièrement connectées pour conserver l'aspect spatial

## Fully Convolutional Networks



Long, Shelhamer et Dorell, *Fully Convolutional Networks for Semantic Segmentation*, CVPR 2015

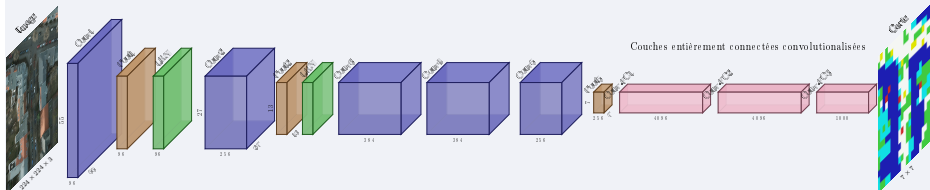


# Cartographie par classification dense

Le classifieur doit générer des prédictions **denses** de façon **efficace**.

- ▶ les CNN réalisent déjà une extraction de caractéristiques spatialement denses
- ▶ il suffit de convolutionnaliser les couches entièrement connectées pour conserver l'aspect spatial

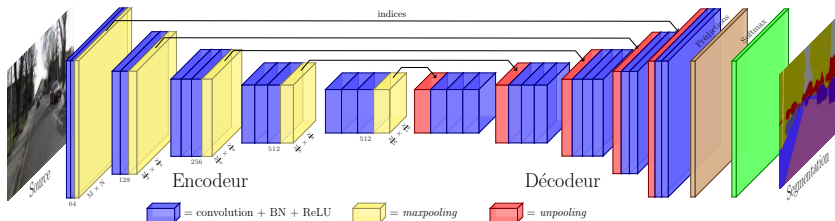
## Fully Convolutional Networks



Jamie Sherrah, *Fully Convolutional Networks for Dense Semantic Labeling of High-Resolution Aerial Imagery*, arXiv, 2016

Maggiore et al., *Fully Convolutional Neural Networks For Remote Sensing Image Classification*, IGARSS 2016

# Architectures symétriques : SegNet

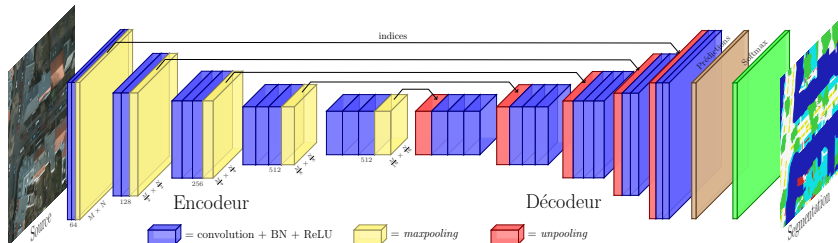


Badrinarayanan et al., *SegNet : A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation*, TPAMI, 2017.

- ▶ SegNet conserve la résolution de l'image d'entrée
- ▶ optimisation sur l'entropie croisée pixel à pixel :

$$\mathcal{L}(\text{softmax}(z), y) = -\frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N \sum_{p=1}^k y_p^{(i,j)} \log \left( \frac{\exp(z_p^{(i,j)})}{\sum_{q=1}^k \exp(z_q^{(i,j)})} \right)$$

# Architectures symétriques : SegNet



Audebert et al., *Semantic Segmentation of Earth Observation Data Using Multimodal and Multi-scale Deep Networks*, ACCV, 2016.

- ▶ SegNet conserve la résolution de l'image d'entrée
- ▶ optimisation sur l'entropie croisée pixel à pixel :

$$\mathcal{L}(\text{softmax}(z), y) = -\frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N \sum_{p=1}^k y_p^{(i,j)} \log \left( \frac{\exp(z_p^{(i,j)})}{\sum_{q=1}^k \exp(z_q^{(i,j)})} \right)$$

# Adaptation aux images aériennes et satellitaires



Image de télédétection (5000 × 5000)



Photo HD (1920 × 1080)



ImageNet (256 × 256)

- ▶ Apprentissage sur des imagerie aléatoires extraites des images haute résolution
- ▶ Prédiction sur une fenêtre glissante avec recouvrement et vote majoritaire pour lisser les discontinuités aux bords
- ▶ Augmentation de données par rotation (invariance à l'azimut)
- ▶ Encodeur initialisé à partir des poids de VGG-16 pré-entraîné

(Simonyan et al., ICLR 2014)

# Comparaison avec les méthodes précédentes

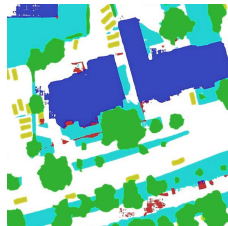
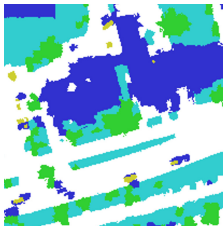
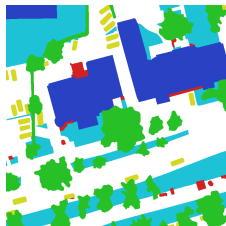
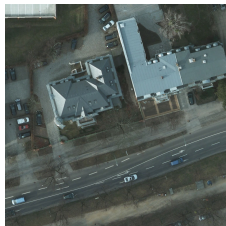


Image (RGB)

Vérité terrain

Segmentation + FA

SegNet

Résultats sur le jeu de données ISPRS Potsdam<sup>4</sup> (score  $F_1$  et exactitude).

Méthode	Routes	Bâtiments	Vég. basse	Arbres	Véhicules	Exactitude
Forêt aléatoire (FA)	77,0 %	79,7 %	73,1 %	59,4 %	58,8 %	74,2 %
FCN (Sherrah, 2016)	91,4 %	95,3 %	85,1 %	87,3 %	88,7 %	89,1 %
SegNet	93,0 %	92,9 %	85,0 %	85,1 %	95,1 %	89,7 %

4. Rottensteiner et al., *The ISPRS benchmark on urban object classification and 3D building reconstruction*, ISPRS Annals, 2012

# Nouvelles applications

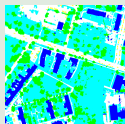
## Détection et reconnaissance de véhicules



Fenêtre glissante sur l'image RVB



réseau profond de segmentation



Carte sémantique

extraction des véhicules



Masque des véhicules



réseau profond classification



Véhicules classifiés

*Segment-before-Detect : Vehicle Detection and Classification through Semantic Segmentation of Aerial Images*, N. Audebert, B. Le Saux, S. Lefèvre, Remote Sensing, 2017

# Rappel des problématiques

- ✓ Peut-on déléguer à la machine l'extraction de connaissances à partir des images d'observation de la Terre?
- ▶ **Est-il possible de combiner différentes sources d'images pour enrichir les connaissances?**
- ▶ Comment tirer profit de l'afflux constant de données?

# Fusion de données et apprentissage multi-modal

---



# État de l'art en apprentissage multimodal

En vision par ordinateur, la fusion raster/raster est généralement mise en œuvre pour le traitement d'images RGB-D grâce aux :

- ▶ *Dual stream networks*

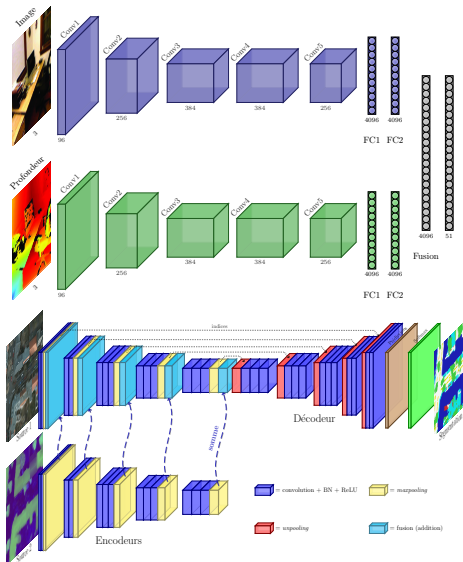
(Simonyan et al., NIPS 2014, Eitel et al., IROS 2015)

- ▶ Approches ensemblistes

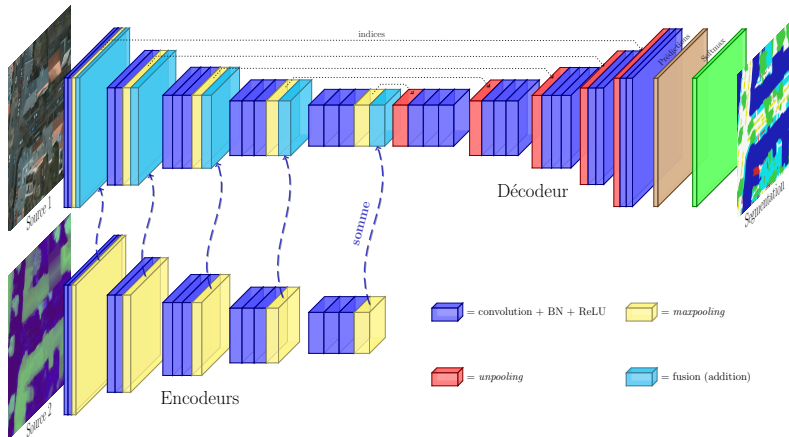
(Neverova et al., TPAMI 2015)

- ▶ Approches conjointes

(Hazirbas et al., ACCV 2016)



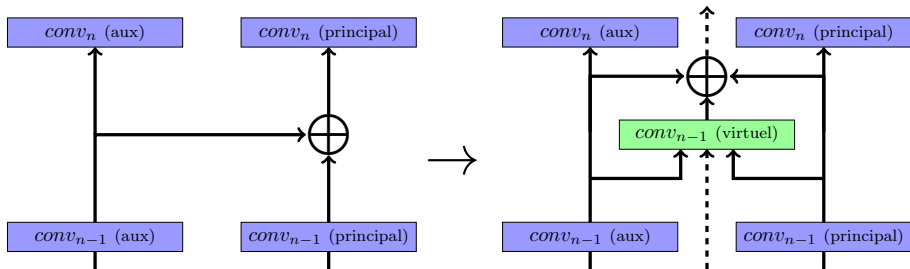
# Apprentissage multimodal : FuseNet



Approche par fusion : apprentissage conjoint sur les deux capteurs

Hazirbas et al., *FuseNet : Incorporating Depth into Semantic Segmentation via Fusion-based CNN Architecture*, ACCV 2016

# V-FuseNet : fusion résiduelle et symétrisation



Fusion par somme (Hazirbas et al.)

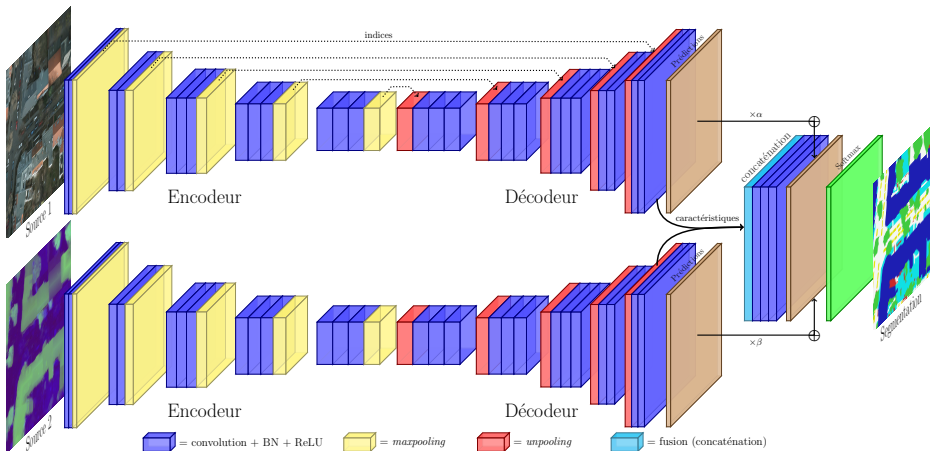
Fusion résiduelle (Audebert et al.)

- ▶ symétrisation de FuseNet par l'introduction d'un encodeur virtuel
- ▶ apprentissage de la fusion sur un mode résiduel (He et al., CVPR 2016)

## Publication associée

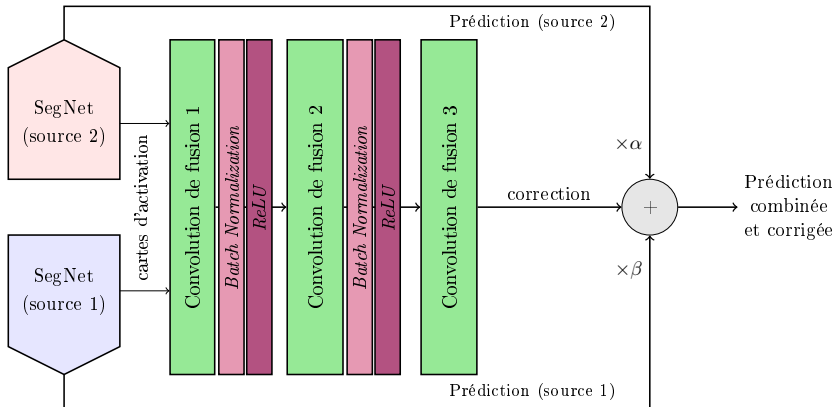
**N. Audebert**, B. Le Saux, S. Lefèvre, *Beyond RGB : Very high resolution urban remote sensing with multimodal deep networks*, ISPRS Journal, 2018.

# Fusion tardive par correction résiduelle



Approche par correction résiduelle : un réseau prédicteur par capteur + fusionneur

# Module de correction résiduelle



## Publication associée

N. Audebert, B. Le Saux, S. Lefèvre, *Semantic Segmentation of Earth Observation Data Using Multimodal and Multi-scale Deep Networks*, ACCV 2016

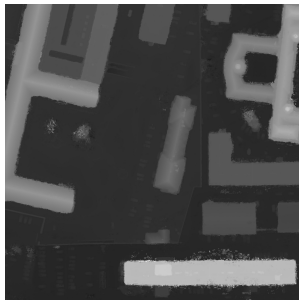
# Combinaison optique et Lidar

## Problématique

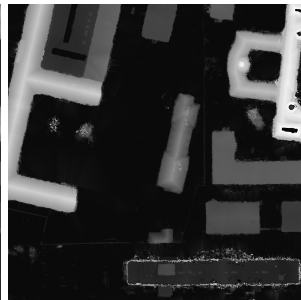
Du Lidar se dérive le modèle numérique de surface, c'est-à-dire une information de hauteur absente de l'image optique. Comment exploiter conjointement ces deux capteurs ?



Image RVB



MNS



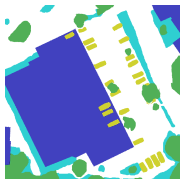
MNS normalisé

# Résultats quantitatifs : ISPRS Potsdam

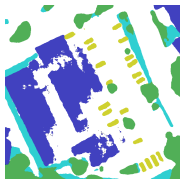
Modèle	Routes	Bâtiments	Vég. basse	Arbres	Véhicules	Exactitude
FCN + CRF expert (Liu et al., 2017)	91,2	94,6	85,1	85,1	92,8	88,4
FCN (Sherrah, 2016)	92,5	96,4	86,7	88,0	94,7	90,3
SegNet (IRRV)	92,4	95,8	86,7	87,4	95,1	90,0
SegNet-CR <sup>5</sup>	<b>93,3</b>	<b>97,3</b>	87,6	<b>88,3</b>	<b>95,8</b>	<b>91,0</b>
FuseNet	93,0	97,0	87,3	87,7	95,2	90,6
V-FuseNet	93,2	97,2	<b>87,9</b>	88,2	95,0	<b>91,0</b>



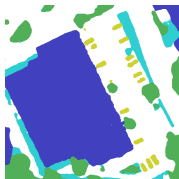
Image (IRRV)



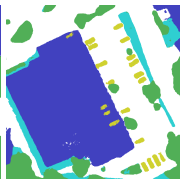
Vérité terrain



SegNet (IRRV)



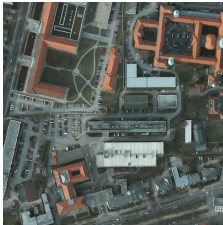
FuseNet



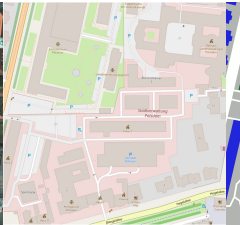
SegNet-CR

5. CR : correction résiduelle

# Combinaison OpenStreetMap et optique



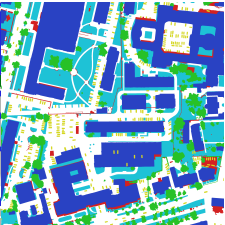
Image



OSM



Raster OSM



Vérité terrain

## Prise en compte des données géographiques existantes

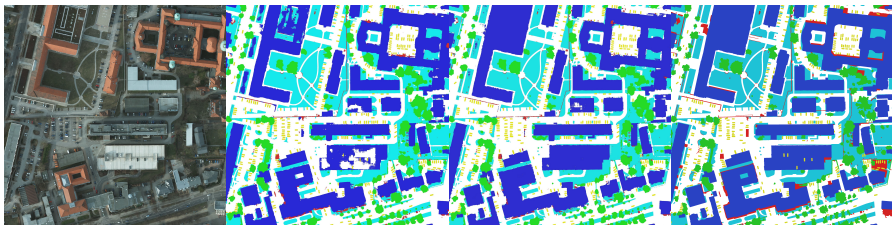
Les informations géographiques d'OpenStreetMap permettent d'établir un a priori fort sur la nature des objets observés. Mais :

- ▶ Les classes ne correspondent pas à celles recherchées,
- ▶ Il existe un décalage temporel entre les images et OSM.



# Résultats sur le jeu de données ISPRS Potsdam

Méthode	Routes	Bâtiments	Vég. basse	Arbres	Véhicules	Global
SegNet (RVB)	93,0%	92,9%	85,0%	85,1%	95,1%	89,7%
SegNet-CR (RVB + OSM)	93,9%	92,8%	85,1%	<b>85,2%</b>	95,8%	90,6%
FuseNet (RVB + OSM)	<b>95,3%</b>	<b>95,9%</b>	<b>86,3%</b>	85,1%	<b>96,8%</b>	<b>92,3%</b>



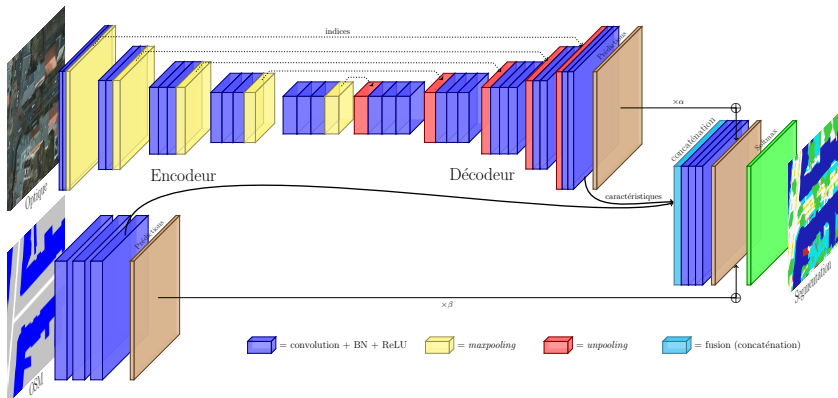
Image

SegNet

FuseNet (+ OSM)

Vérité terrain

# Mise à jour de données géographiques



La vérité terrain étant réalisée sur la donnée optique (qui sert de référence), les désaccords entre OSM et les prédictions du réseau permettent de mettre à jour OSM dans le temps par raffinement.

# Rappel des problématiques

- ✓ Peut-on déléguer à la machine l'extraction de connaissances à partir des images d'observation de la Terre ?
- ✓ Est-il possible de combiner différentes sources d'images pour enrichir les connaissances ?
- ▶ **Comment tirer profit de l'afflux constant de données ?**

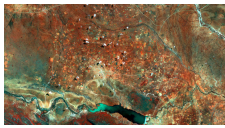
## Passage à l'échelle

---

# Constitution d'un jeu de données à grande échelle



Mont Saint-Michel



Serowe (Botswana)



Désert d'Atacama



Egg Island (Bahamas)

Crédits : Copernicus Sentinel-2 (ESA), CC BY-SA 3.0 IGO

Comment s'adapter à la variété des environnements terrestres ?

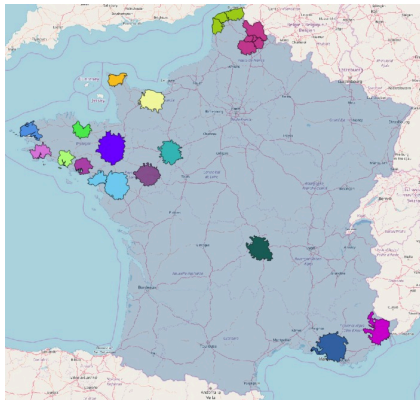
- ▶ Impossible de généraliser à partir d'une seule région
- ▶ Succès en vision : pré-entraîner sur une grande variété d'images  
→ même besoin en télédétection

Deux solutions :

- ▶ Adaptation de domaine (inadapté sur une aussi grande variété)
- ▶ Entraînement sur une vaste quantité de données diversifiées

# MiniFrance : jeu de données aérien multiscène multiclasse

## Cahier des charges



▶ Aérien ou satellitaire sur de multiples scènes

→ *BD ORTHO de l'IGN (RVB 50cm/px)*

▶ Redistribuable

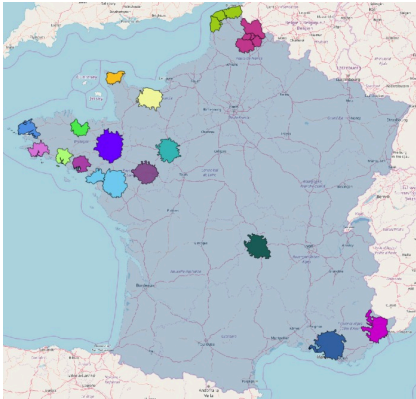
→ *Données sous licence ouverte*

▶ Annotations disponibles pour l'apprentissage supervisé

→ *Projet Copernicus Urban Atlas*

→ images couleur et annotations semi-automatiques d'usage des sols sur les agglomérations >30 000 habitants

# MiniFrance : jeu de données aérien multiscène multiclasse



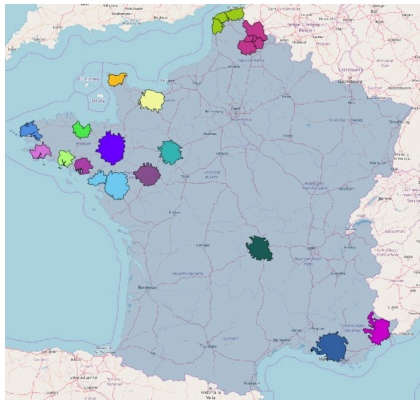
## Cahier des charges

- ▶ Aérien ou satellitaire sur de multiples scènes
- **BD ORTHO de l'IGN (RVB 50cm/px)**
- ▶ Redistribuable
- *Données sous licence ouverte*
- ▶ Annotations disponibles pour l'apprentissage supervisé
- *Projet Copernicus Urban Atlas*

→ images couleur et annotations semi-automatiques d'usage des sols sur les agglomérations >30 000 habitants

# MiniFrance : jeu de données aérien multiscène multiclasse

## Cahier des charges

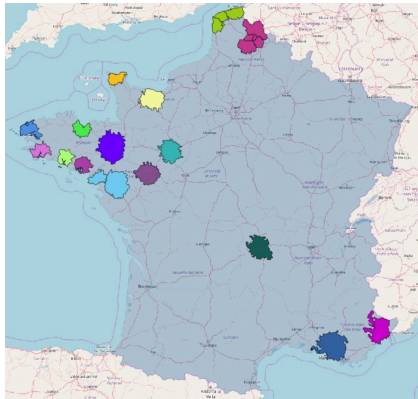


- ▶ Aérien ou satellitaire sur de multiples scènes
- *BD ORTHO de l'IGN (RVB 50cm/px)*
- ▶ Redistribuable
- *Données sous licence ouverte*
- ▶ Annotations disponibles pour l'apprentissage supervisé
- *Projet Copernicus Urban Atlas*

→ images couleur et annotations semi-automatiques d'usage des sols sur les agglomérations >30 000 habitants



# MiniFrance : jeu de données aérien multiscène multiclasse



## Cahier des charges

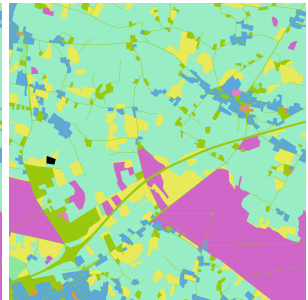
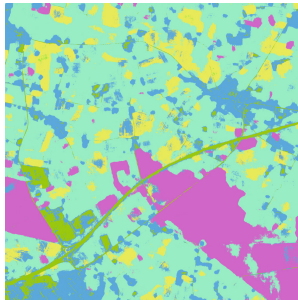
- ▶ Aérien ou satellitaire sur de multiples scènes
- ***BD ORTHO de l'IGN (RVB 50cm/px)***
- ▶ Redistribuable
- ***Données sous licence ouverte***
- ▶ Annotations disponibles pour l'apprentissage supervisé
- ***Projet Copernicus Urban Atlas***

→ images couleur et annotations semi-automatiques d'usage des sols sur les agglomérations >30 000 habitants

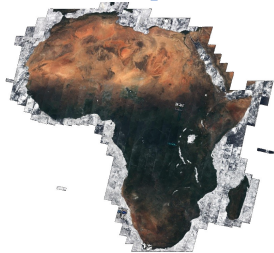
# Premiers résultats et défis

Entraînement sur	OA	<i>mIoU</i>
100% des images	52,40	15,79
10% des images	50,14	15,25
Caen uniquement	42,09	10,05

- ▶ Apprendre avec Urban Atlas est plus difficile qu'avec les annotations ISPRS (taxonomie complexe, zones imprécises, décalage temporel)



## Extension de l'African CCI Land Cover (Sentinel-2)



- ▶ En collaboration avec le  $\Phi$ -lab de l'ESA
- ▶ Apprendre l'occupation des sols à partir d'images S2 et de la classification CCI African Land Cover 2016
- ▶ CCI African Land Cover : classification hors-ligne sur une série temporelle 2016
- ▶ Objectif : mise à jour à chaque nouvelle image Sentinel-2 par réseau profond
- ▶  $\simeq$  50% d'accord sur les zones tests en Afrique de l'est

## Conclusion

---

# Conclusion

Les réseaux de neurones profonds sont d'excellents outils pour le traitement automatique des images aériennes et satellites.

- ▶ Les réseaux de neurones permettent d'établir de nouveaux états de l'art en traitement d'images d'observation de la Terre (cartographie et détection d'objets).
- ▶ Ces modèles peuvent être utilisés dans un contexte multicapteur pour enrichir les inférences.
- ▶ Les données pour des applications à large échelle (pays voire continent) existent et les premiers résultats sont prometteurs.

# Conclusion

Les réseaux de neurones profonds sont d'excellents outils pour le traitement automatique des images aériennes et satellites.

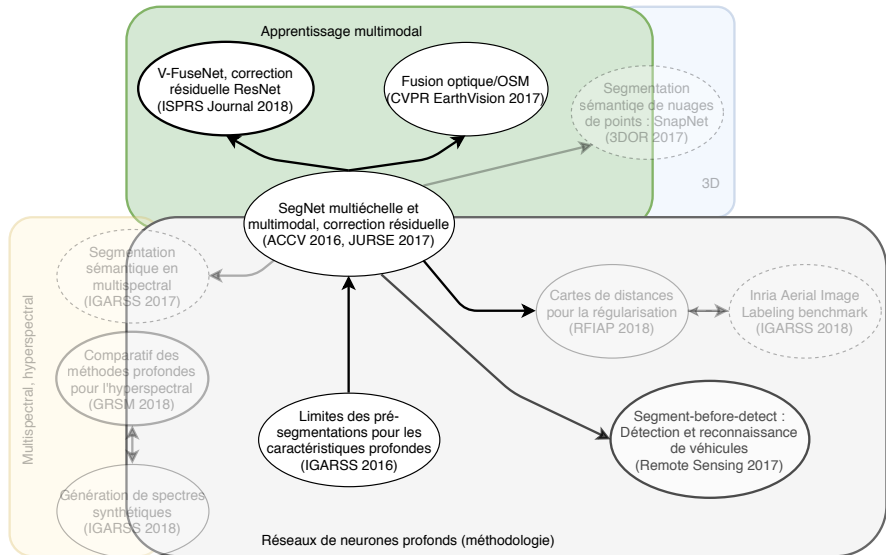
- ▶ Les réseaux de neurones permettent d'établir de nouveaux états de l'art en traitement d'images d'observation de la Terre (cartographie et détection d'objets).
- ▶ Ces modèles peuvent être utilisés dans un contexte multicapteur pour enrichir les inférences.
- ▶ Les données pour des applications à large échelle (pays voire continent) existent et les premiers résultats sont prometteurs.

# Conclusion

Les réseaux de neurones profonds sont d'excellents outils pour le traitement automatique des images aériennes et satellites.

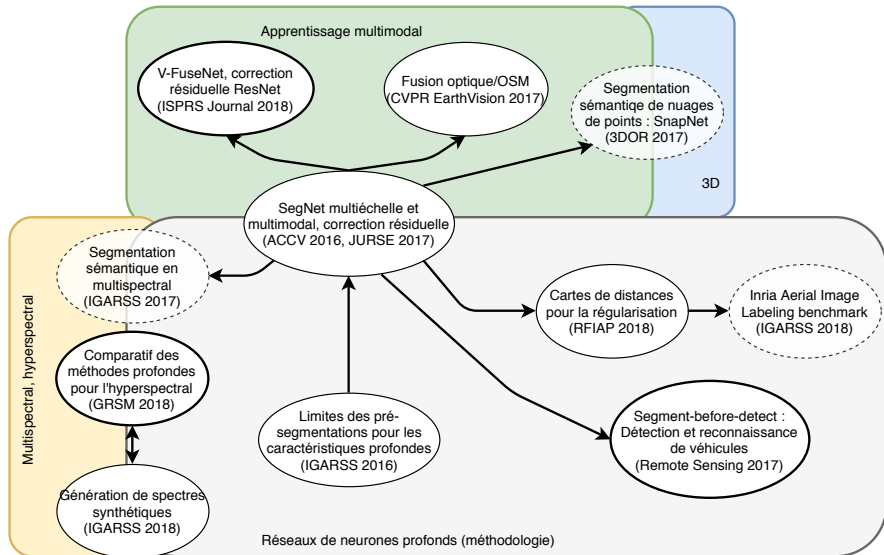
- ▶ Les réseaux de neurones permettent d'établir de nouveaux états de l'art en traitement d'images d'observation de la Terre (cartographie et détection d'objets).
- ▶ Ces modèles peuvent être utilisés dans un contexte multicapteur pour enrichir les inférences.
- ▶ Les données pour des applications à large échelle (pays voire continent) existent et les premiers résultats sont prometteurs.

# Graphesynoptique de la thèse





# Graphes synoptique de la thèse



# Perspectives

- ▶ Suivi temporel pour la détection de changements et l'étude des évolutions du territoire
- ▶ Apprentissage semi-supervisé pour l'exploitation des vastes quantités de données non-annotées
- ▶ Approches multimodales incluant des capteurs non-optiques (radar) et images au sol

# Publications

## Journaux

**N. Audebert**, B. Le Saux, S. Lefèvre (2018). Deep Learning for Classification of Hyperspectral Data : A Comparative Review, Geosciences and Remote Sensing Magazine, 2018 (soumis).

**N. Audebert**, B. Le Saux, S. Lefèvre (2017). Beyond RGB : Very High Resolution Urban Remote Sensing With Multimodal Deep Networks, ISPRS Journal of Photogrammetry and Remote Sensing, Elsevier.

**N. Audebert**, B. Le Saux, S. Lefèvre (2017). Segment-before-Detect : Vehicle Detection and Classification through Semantic Segmentation of Aerial Images, Remote Sensing, MDPI.

A. Boulch, J. Guerry, B. Le Saux, **N. Audebert** (2017). SnapNet : 3D Point Cloud Semantic Labeling with 2D Deep Segmentation Networks, Computer & Graphics, Elsevier

# Publications

## Conférences internationales

**N. Audebert**, B. Le Saux, S. Lefèvre (2018). Deep generative models for realistic synthesis of hyperspectral samples, IGARSS.

**N. Audebert**, B. Le Saux, S. Lefèvre (2017). Joint Learning from Earth Observation and OpenStreetMap Data to Get Faster Better Semantic Maps. CVPR Workshop - EarthVision.

A. Ben Hamida, A. Benoit, P. Lambert, L. Klein, C. Ben Amar, **N. Audebert**, S. Lefèvre (2017). Deep learning for semantic segmentation of remote sensing images with rich spectral content, IGARSS.

**N. Audebert**, B. Le Saux, S. Lefèvre (2016). Semantic segmentation of Earth Observation data using multi-scale and multimodal deep networks, ACCV.

**N. Audebert**, B. Le Saux, S. Lefèvre (2017). Fusion of Heterogeneous Data in Convolutional Networks for Urban Semantic Labeling, JURSE, *2nd best student paper award*.

**N. Audebert**, B. Le Saux, S. Lefèvre (2016). On the usability of deep networks for object-based image analysis. GEOBIA, *Best contribution to the benchmark award*.

**N. Audebert**, B. Le Saux, S. Lefèvre (2016). How useful is region-based segmentation in a deep learning framework?, IGARSS.

# Publications

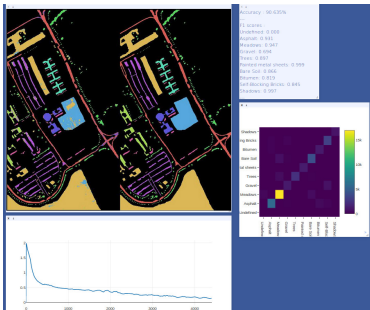
## Codes sources

DeepHyperX, boîte à outils de deep learning pour l'imagerie hyperspectrale :

<https://gitlab.inria.fr/naudeber/DeepHyperX>

Deep Nets for Earth Observation, réseaux de neurones pré-entraînés pour la cartographie automatisée d'images optiques aériennes et satellites :

<https://github.com/nshaud/DeepNetsForEO>



### Semantic segmentation of aerial images with deep networks

This notebook presents a straightforward PyTorch implementation of a Fully Convolutional Network for semantic segmentation of aerial images. More specifically, we aim to automatically perform scene interpretation of images taken from a plane or a satellite by classifying every pixel into several land cover classes.

As a demonstration, we are going to use the [SedNet architecture](#) to segment aerial images over the cities of Valhøgen and Potsdam. The images are from the [SPV5-30 Semantic Labeling dataset](#). We will train a network to segment roads, buildings, vegetation and cars.

This work is a PyTorch implementation of the baseline presented in "[3Dconv-FC8: Very High Resolution Urban Remote Sensing](#)" [Yifu Wang, Dong Sun, Yueshan Chen, Nicolas Audebert, Bertrand Le Saux and Sébastien Lefèvre, ISPRS Journal, 2018](#).

#### Requirements

This notebook requires a few useful libraries, e.g. torch, scikit-image, numpy and matplotlib. You can install everything using `pip install -r requirements.txt`.

This is expected to run on a GPU, and therefore you should use torch in combination with CUDA/cuDNN. This can probably be made to run on CPU but we warned that:

- you have to rename all calls to torch\_Tensor (cudnn) throughout this notebook,
- this will be very slow.

A "small" GPU should be enough, e.g. this runs fine on a 4-7GB Tesla K80. It uses quite a lot of RAM as the dataset is stored in memory (about 50GB for Valhøgen). You can spare some memory by disabling the caching below. 4GB should be more than enough without caching.

```
In [1]: % imports and stuff
import numpy as np
from skimage import io
from glob import glob
from tqdm import tqdm
from sklearn.metrics import confusion_matrix
import random
import itertools
# matplotlib
import matplotlib.pyplot as plt
matplotlib.rcParams['font.family'] = 'serif'
# Torch imports
import torch
import torch.nn as nn
import torch.nn.functional as F
import torch.optim as optim
import torch.optim.lr_scheduler
import torch.nn.init
from torch.autograd import Variable
```

# Remerciements, questions, contact



## Remerciements

Ces travaux sont financés par le projet de recherche commun NAOMI entre l'ONERA et Total.



Pour plus d'informations, questions et remarques

<https://nicolas.audebert.at>

[nicolas.audebert@onera.fr](mailto:nicolas.audebert@onera.fr)

## Expériences préliminaires

---

## Transfert ImageNet → télédétection

- ▶ ImageNet : vie quotidienne (animaux, objets, personnes...)
  - symétrie gauche/droite, perspective et changements d'échelle
- ▶ Télédétection : structures au nadir (bâtiments, forêts, véhicules...)
  - équivariance à l'azimut, aucune perspective, échelle fixe

Les filtres appris sur ImageNet se transfèrent au moins partiellement sur des images de télédétection.<sup>6,7</sup>

Comparaison de différentes initialisations (ISPRS Vaihingen)

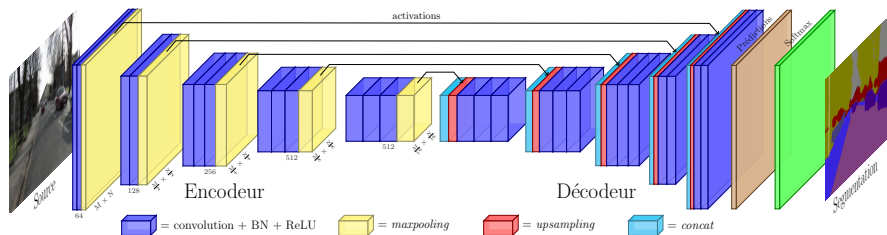
Initialisation	Aléatoire	VGG-16 (ImageNet)			
Variabilité de l'encodeur $\frac{\alpha_e}{\alpha_d}$	1	1	0,5	0,1	0
Exactitude	87,0%	87,2%	<b>87,8%</b>	86,9%	86,5%

6. Penatti et al., *Do Deep Features Generalize from Everyday Objects to Remote Sensing and Aerial Scenes Domains?*, CVPRW 2015

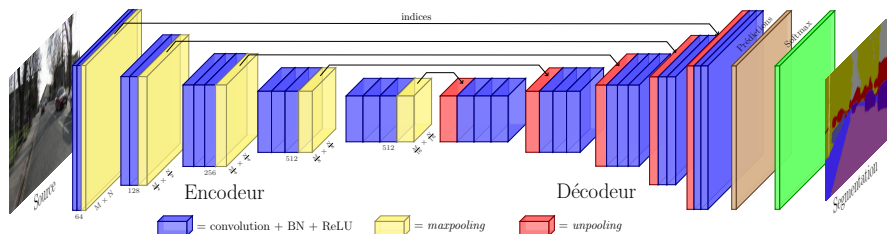
7. Razavian et al., *CNN Features off-the-shelf: an Astounding Baseline for Recognition*, CVPRW 2014



## SegNet/UNet



Ronneberger et al., *U-Net : Convolutional Networks for Biomedical Image Segmentation*, MICCAI 2015



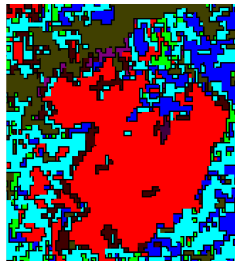
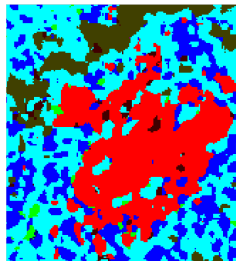
Badrinarayanan et al., *SegNet : A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation*, TPAMI 2017, arXiv 2015

## Extension au cas multispectral

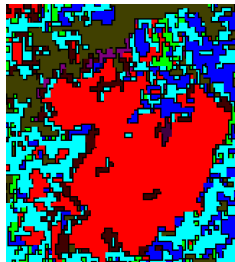
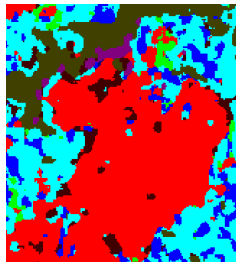
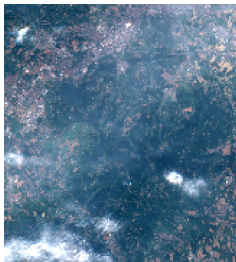
---

# Résultats

D1 (sans nuage)



D2 (avec nuages)

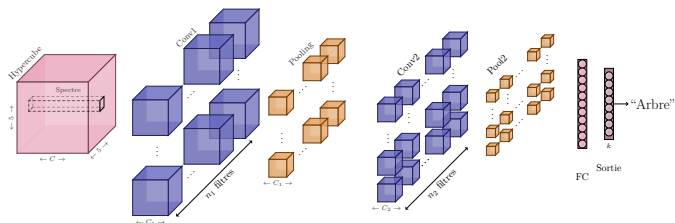
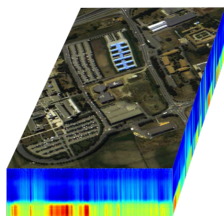


Couleur naturelle

Prédiction

Vérité terrain

# L'imagerie hyperspectrale



## Motivation

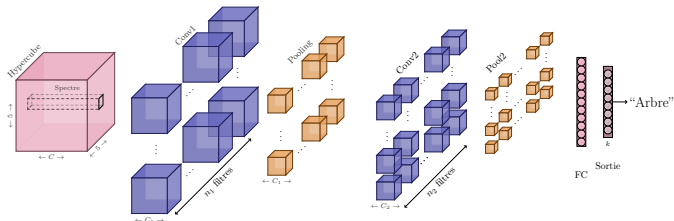
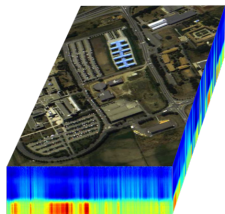
Capteur hyperspectral = centaines de longueurs d'onde

Des réseaux 3D permettent de traiter le "cube" hyperspectral.

## Problématique

Les réseaux profonds sont très gourmand en données, mais peu d'exemples annotés en hyperspectral...

# L'imagerie hyperspectrale



## Motivation

Capteur hyperspectral = centaines de longueurs d'onde

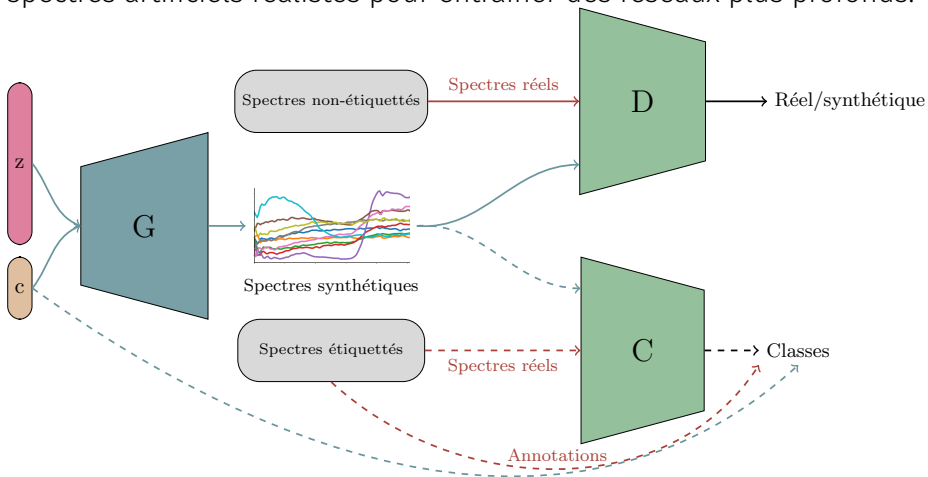
Des réseaux 3D permettent de traiter le "cube" hyperspectral.

## Problématique

Les réseaux profonds sont très gourmand en données, mais peu d'exemples annotés en hyperspectral...

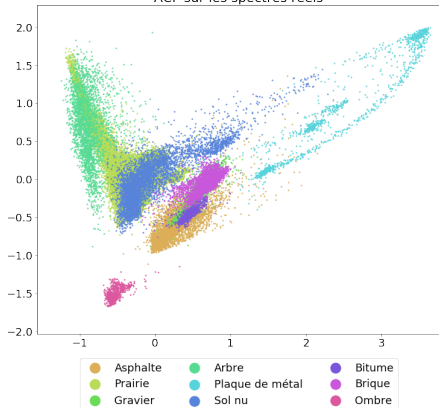
# Génération de données

Idee : Générer des exemples d'apprentissage synthétiques, *i.e.* des spectres artificiels réalistes pour entraîner des réseaux plus profonds.

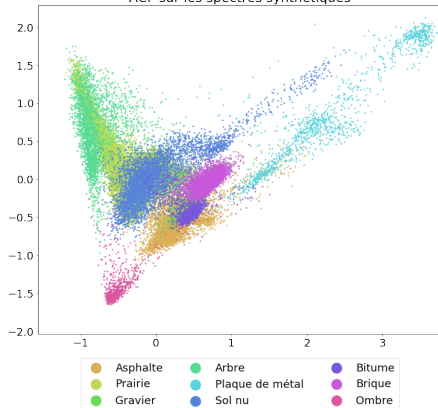


# Comparaison avec la distribution réelle

ACP sur les spectres réels

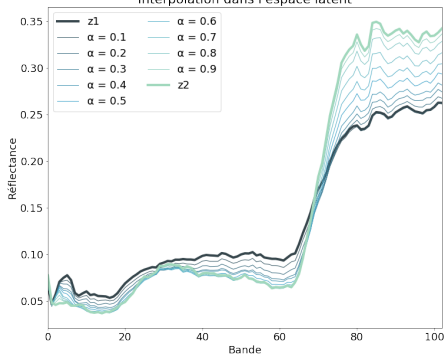


ACP sur les spectres synthétiques

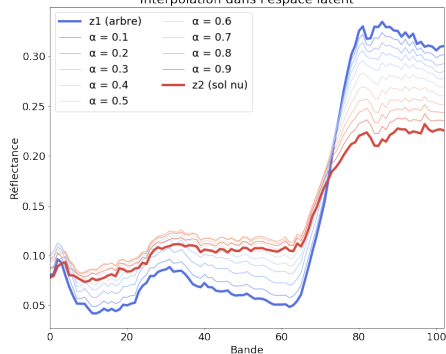


# Interpolations dans l'espace latent

Interpolation dans l'espace latent



Interpolation dans l'espace latent





## Régularisation par carte de distances

---

# Transformée de distance euclidienne

## Idée

Estimer la transformée de distance euclidienne tronquée (Ye et al, ICPR 1998) pour chaque classe afin d'incorporer le voisinage spatial dans la fonction de coût.

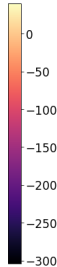
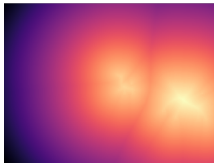


$$y[i, j] = (0, \dots, 0, 1, 0, \dots, 0) \quad \text{vs} \quad d[i, j] = (-1, -0.3, \dots, 0.8, -1, \dots, -0.3)$$

# Transformée de distance euclidienne

## Idée

Estimer la transformée de distance euclidienne tronquée (Ye et al, ICPR 1998) pour chaque classe afin d'incorporer le voisinage spatial dans la fonction de coût.

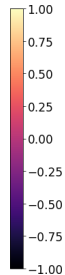
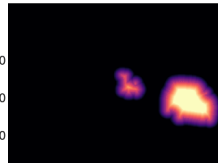
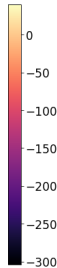
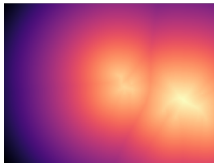


$$y[i, j] = (0, \dots, 0, 1, 0, \dots, 0) \quad \text{vs} \quad d[i, j] = (-1, -0.3, \dots, 0.8, -1, \dots, -0.3)$$

# Transformée de distance euclidienne

## Idée

Estimer la transformée de distance euclidienne tronquée (Ye et al, ICPR 1998) pour chaque classe afin d'incorporer le voisinage spatial dans la fonction de coût.

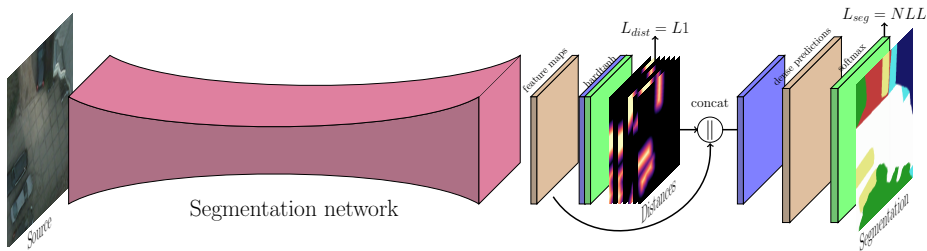


$$y[i, j] = (0, \dots, 0, 1, 0, \dots, 0) \quad \text{vs} \quad d[i, j] = (-1, -0.3, \dots, 0.8, -1, \dots, -0.3)$$

# Apprentissage multitâche

## Architecture multitâche

La régression des CDS est utilisée comme proxy pour la classification.

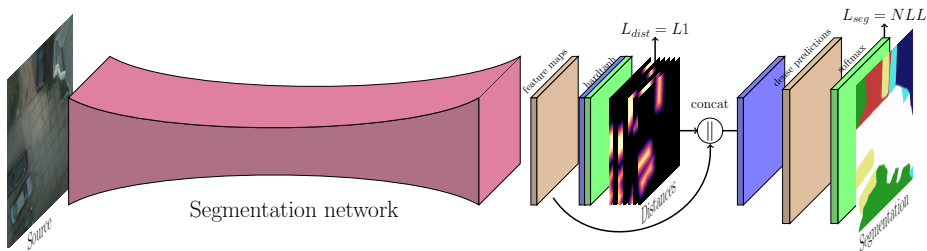


$$\mathcal{L}_{total} = \underbrace{\text{NLLLoss}(\text{softmax}(\hat{y}), \text{softmax}(y))}_{\text{entropie croisée}} + \underbrace{\lambda \cdot |\hat{D}_y - D_y|}_{\substack{\text{L1 sur les distances} \\ \text{force de la régularisation}}}$$

# Apprentissage multitâche

## Architecture multitâche

La régression des CDS est utilisée comme proxy pour la classification.



$$\mathcal{L}_{total} = \underbrace{\text{NLLLoss}(\text{softmax}(\hat{y}), \text{softmax}(y))}_{\text{entropie croisée}} + \underbrace{\lambda \cdot |\hat{\mathcal{D}}_y - \mathcal{D}_y|}_{\substack{\text{L}_1 \text{ sur les distances} \\ \text{force de la régularisation}}}$$

# Inria Aerial Image Labeling

▶ SegNet : 71.02% IoU (+6.98%), 95.63% OA (+0.89%)

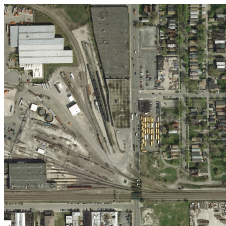
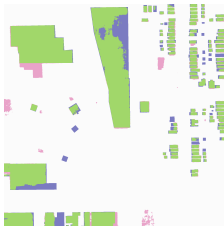
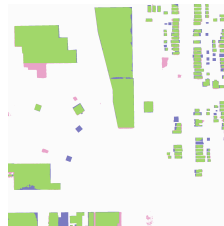


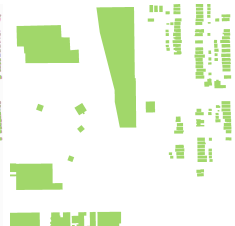
Image RVB



SegNet  
(classification)



SegNet  
(multitâche)



Vérité terrain

Vert : vrais positifs, rose : faux positifs, bleu : faux négatifs, blanc : vrais négatifs.