

Réseaux de neurones profonds et fusion de données pour la segmentation sémantique d'images aériennes

Nicolas Audebert^{1,2}, Bertrand Le Saux¹, Sébastien Lefèvre²

¹ ONERA, *The French Aerospace Lab*

² Université de Bretagne-Sud/IRISA

Journées des jeunes chercheurs en vision par ordinateur (ORASIS)

La cartographie pour l'observation de la Terre

Apprentissage profond pour la cartographie sémantique

Extension à la cartographie sémantique multi-capteurs

Conclusion

La cartographie pour l'observation de la Terre





15

5

32

28

24

5

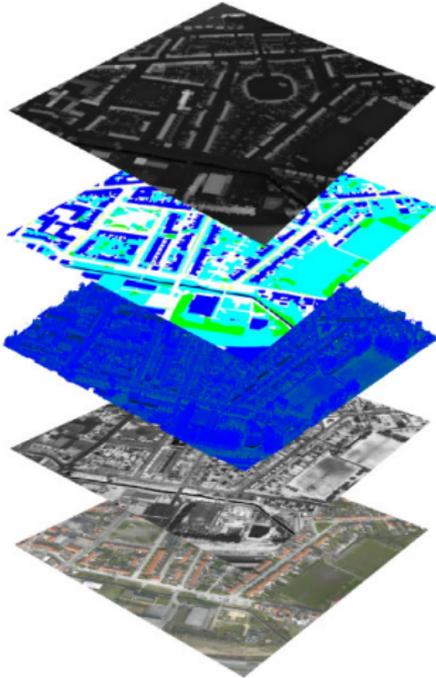
le Cavey

Parc
de loisirs

RTE DU CAP JOE DAWSON

50

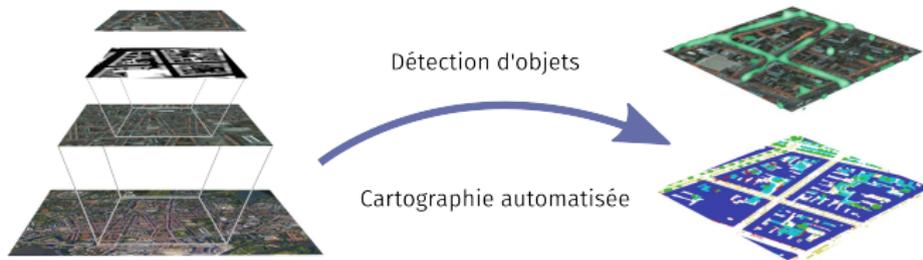
61



Images de télédétection

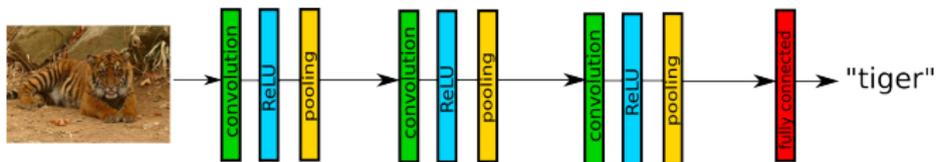
La communauté scientifique produit des images en abondance

- ▶ Capteurs aéroportés ou satellites (Sentinel, SPOT, AVIRIS)...
- ▶ Données hétérogènes et auxiliaires (modèles de surface, données cadastrales) recalées géographiquement
- ▶ Quelques bases de données d'images annotées par des experts!



Générer des cartes informatives à partir des images

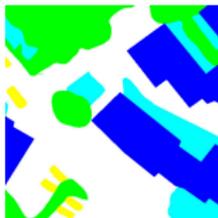
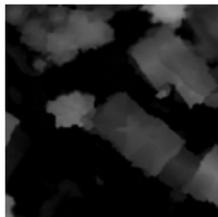
- ▶ **Classification** la plus **dense** possible
- ▶ **Cartes thématiques** : espèces végétales, routes & bâtiments...
- ▶ Première étape pour des études d'urbanisation, d'estimation de la biomasse, d'analyse du trafic...



Pourquoi le deep learning?

- ▶ Les **Convolutional Neural Networks** ont établi un nouvel état-de-l'art dans la plupart des tâches visuelles (classification, segmentation...),
- ▶ La quantité d'annotations a atteint la masse critique pour travailler en *deep learning* supervisé sur des images aéroportées et satellite,
- ▶ Pourquoi pas?

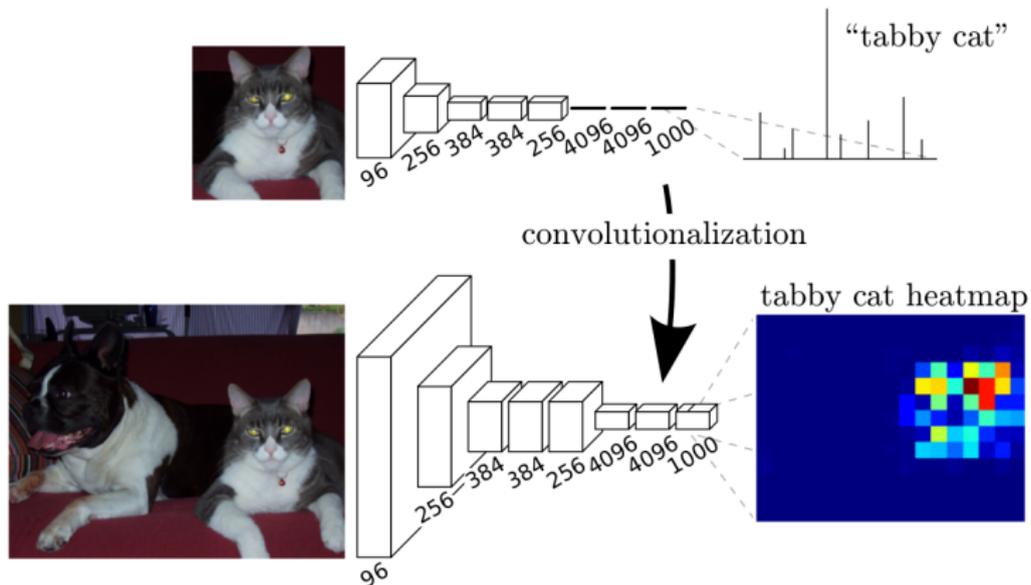
Apprentissage profond pour la cartographie sémantique



ISPRS Vaihingen

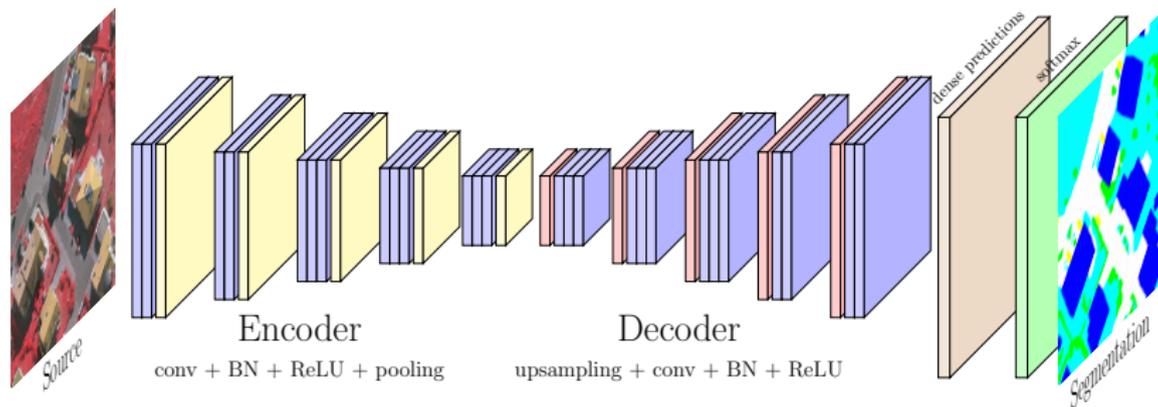
- ▶ Images ortho-rectifiées :
Infrarouge-Rouge-Vert
- ▶ Nuage de points Lidar
 - Modèle de surface (DSM)
 - Modèle de surface normalisé (NDSM)
- ▶ Résolution : 9 cm/pixel
- ▶ Annotations dense au niveau pixel sur 6 classes
 - ▶ routes, bâtiments, végétation basse, arbres, véhicules, “bazar”

Segmentation sémantique mono-capteur



Fully Convolutional Networks for Semantic Segmentation, Long et al., CVPR'15.

Segmentation sémantique mono-capteur



Architecture SegNet (Badrinarayanan et al, TPAMI, 2017) adaptée aux images IRRV de télédétection.

Pré-traitement

Les images haute résolution sont traitées *via* une fenêtre glissante 128×128 avec un recouvrement de 75%.

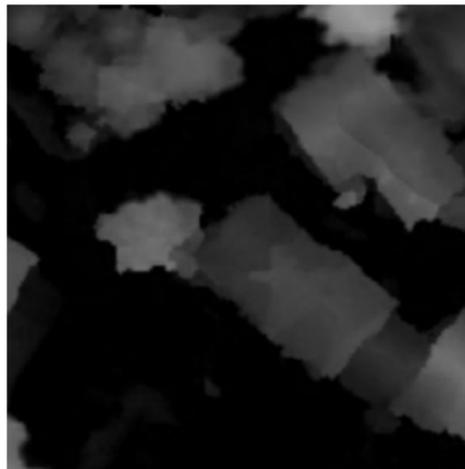
- + Diminue la quantité mémoire GPU requise
- + Apprentissage : augmentation de données
- + Inférence : lissage par moyenne de prédictions
- Temps de calcul linéairement plus long

Optimisation

- ▶ **SGD** avec moment et rétro-propagation classique
- ▶ **Log-loss moyennée** sur les pixels du patch observé (aucune régularisation spatiale explicite)

Problème : approche multi-capteurs

Les données disponibles contiennent aussi bien de l'information optique (IRRV) que Lidar (DSM/NDSM). Est-ce que l'on peut exploiter ces deux capteurs conjointement ?



Extension à la cartographie sémantique multi-capteurs

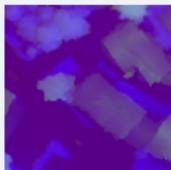
Fusion orientée données ou orientée prédictions ?

- ▶ La fusion orientée données travaille sur les données pour les transformer en une représentation combinée.
- ▶ La fusion orientée prédictions combine les sorties de plusieurs classifieurs.

Ici, on s'intéresse* à la fusion orientée prédictions, dite **fusion tardive**.

* en fait, la fusion orientée données ne donne pas de très bons résultats expérimentaux...

Sources d'entrée

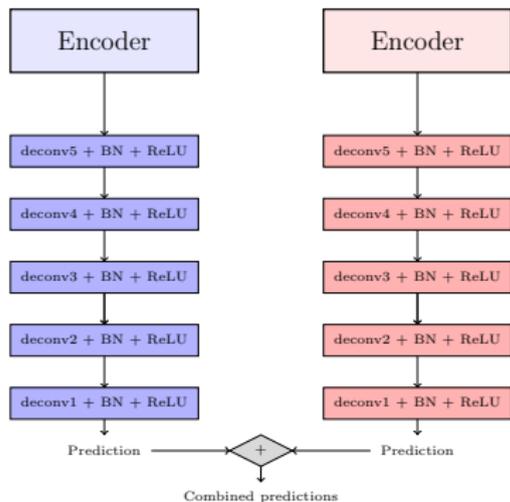


- ▶ Image optique IRRV
- ▶ Image composite DSM/NDSM/NDVI

Segmentation sémantique multi-capteurs

SegNet IRRG

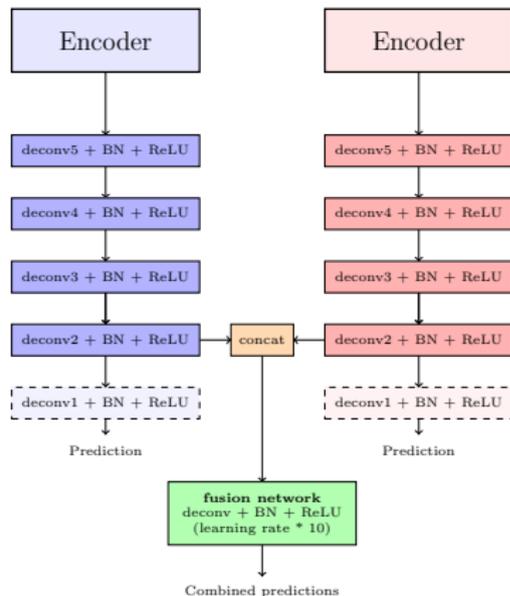
SegNet DSM/NDSM/NDVI



Fusion naïve par moyenne

SegNet IRRG

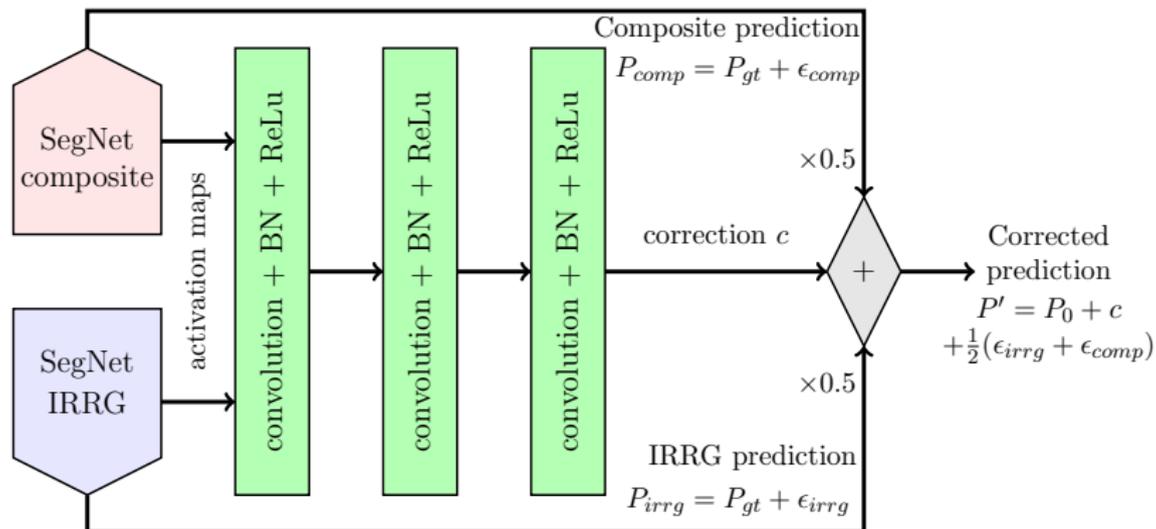
SegNet DSM/NDSM/NDVI



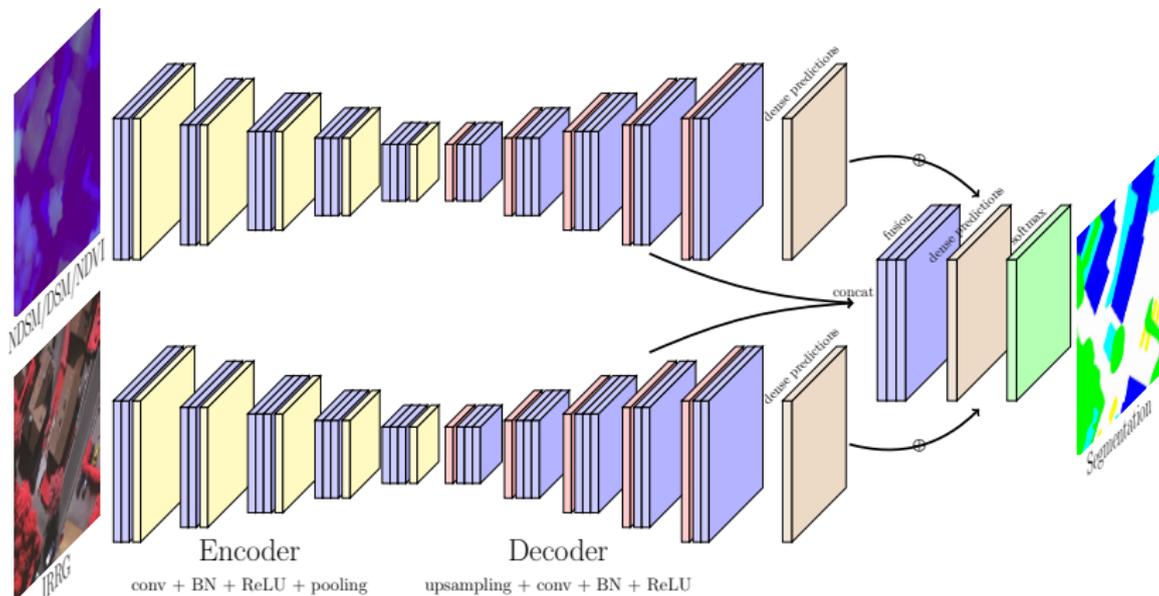
Fusion apprise automatiquement

Correction résiduelle

En s'inspirant du principe d'**apprentissage par résidu** (He et al., CVPR 2016), on peut imaginer un module type "correction d'erreur" combinant les prédictions multi-sources.



Modèle multi-capteurs



SegNet à double flux pour la cartographie multi-capteurs

Jeu de données ISPRS 2D Semantic Labeling Challenge Vaihingen

Method	imp surf	building	low veg	tree	car	Accuracy
RF + CRF ("HUST")	86.9%	92.0%	78.3%	86.9%	29.0%	85.9%
CNN ensemble ("ONE_5")	87.8%	92.0%	77.8%	86.2%	50.7%	85.9%
FCN ("DLR_2")	90.3%	92.3%	82.5%	89.5%	76.3%	88.5%
FCN + RF + CRF ("DST_2")	90.5%	93.7%	83.4%	89.2%	72.6%	89.1%
SegNet++	91.5%	94.3%	82.7%	89.3%	85.7%	89.4%
SegNet++ + fusion	91.0%	94.5%	84.4%	89.9%	77.8%	89.8%



Image IRRV

Vérité terrain

Prédiction
IRRV

Prédiction
comp.

Prédiction
fusion

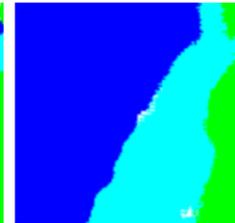
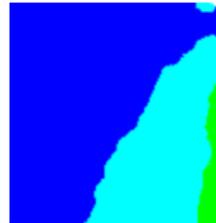
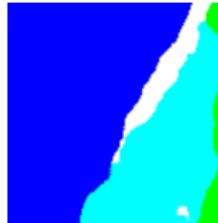


Image IRRV

Vérité terrain

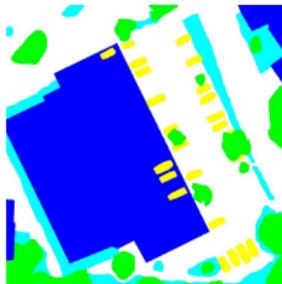
Prédiction
IRRV

Prédiction
comp.

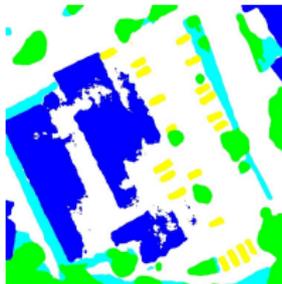
Prédiction
fusion



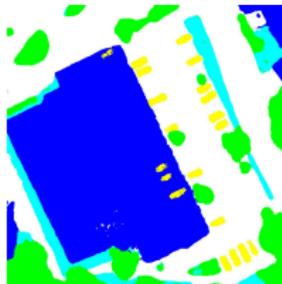
Image IRRV



Vérité terrain



Prédiction IRRV

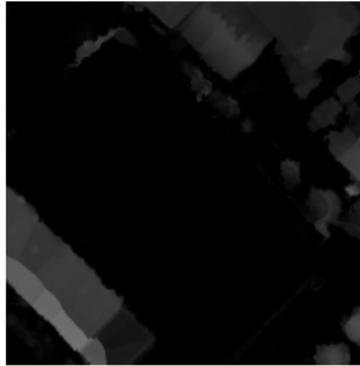


Prédiction
fusion

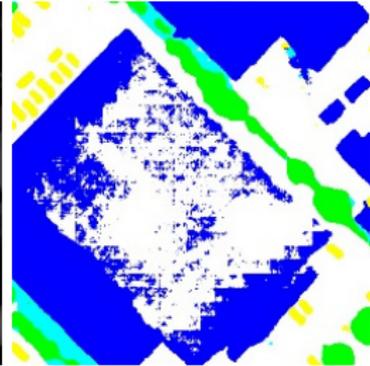
La texture du parking aérien met le modèle purement optique en défaut, mais pas le modèle multi-capteurs qui utilise l'information Lidar pour lever l'ambiguïté.



IRRV



NDSM



Fusion

L'absence d'un bâtiment dans le modèle de surface induit le réseau multi-sources en erreur, qui n'est pas récupérée automatiquement.

- ▶ *“Unstructured point cloud semantic labeling using deep segmentation networks”*, A. Boulch, B. Le Saux, N. Audebert, Eurographics 3DOR, 2017
 - Segmentation sémantique de nuages de points 3D : fusion RGB + données de profondeur
- ▶ *“Joint Learning from Earth Observation and OpenStreetMap Data to Get Faster Better Semantic Maps”*, N. Audebert, B. Le Saux, S. Lefèvre, CVPR EarthVision, 2017
 - ▶ Intégration de connaissances *a priori* : images RGB + données géographiques OpenStreetMap

Conclusion

Contributions principales :

- ▶ Baseline de cartographie sémantique par réseaux profonds
- ▶ Stratégie de fusion tardive multi-capteurs dite de “correction résiduelle”

Perspectives :

- ▶ Intégration de sources multiples (3 et +)
- ▶ Comparaison avec des méthodes de “*early fusion*”



Code et modèles pré-entraînés

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

Questions?

Merci de votre attention!

Pour toute question, remarque ou commentaire :
`nicolas.audebert@onera.fr`.

Remerciements

Ces travaux ont été financés dans le cadre du projet de recherche NAOMI, une collaboration entre l'ONERA et Total.



Code et modèles pré-entraînés

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