



# Localisation d'objets urbains à partir de sources multiples dont des images aériennes, et analyse de leur état, leur évolution temporelle ou leur relation avec leur voisinage

---

Soutenance de thèse de doctorat présentée publiquement le 30 novembre 2018 par :  
**Lionel Pibre**

---

Jenny Benois-Pineau, Pr, Univ. Bordeaux/CNRS/IPB-ENSEIRB - Rapportrice  
Sébastien Lefèvre, Pr, Univ. de Bretagne Sud/IRISA IUT Vannes - Rapporteur  
Josiane Zerubia, DR, INRIA Sophia-Antipolis Méditerranée - Examinatrice  
Georges Quénot, DR, CNRS/LIG Grenoble - Examinateur  
Jérôme Azé, Pr, Univ. Montpellier/LIRMM - Examinateur

Marc Chaumont, MCF, Univ. Nîmes/LIRMM - Directeur  
Gérard Subsol, CR, CNRS/LIRMM - Co-encadrant  
Dino Ienco, CR, IRSTEA/TETIS - Co-encadrant  
Laurent Deruelle, LRA, Berger-Levrault - Invité

## ⊙ Thèse CIFRE

- Équipe ICAR du LIRMM
  - Image et interaction
- Groupe international Berger-Levrault
  - Éditeur de logiciels
  - Collectivités locales, administrations

## ⊙ "Ville intelligente"

- Technologies de l'information et de la communication
  - Améliorer la qualité des services urbains
  - Réduire les coûts des services urbains
- Gestion des objets urbains
  - Produit ATAL II



C. Rochet, "Les villes intelligentes, enjeux et stratégies pour de nouveaux marchés".  
Étude réalisée pour ESCP Europe en 2014.



[http://fr.wikipedia.org/wiki/Ville\\_intelligente](http://fr.wikipedia.org/wiki/Ville_intelligente)



- Localisation d'objets urbains
  - ⇒ Recensement
  - ⇒ Suivi temporel

Données :

- Imagerie aérienne
  - Acquisition à l'échelle d'une ville
  - Très haute résolution
  - Rapide



# Acquisition aérienne

⊙ Ville de Potsdam (Allemagne)

⊙ Résolution au sol : 5 cm×5 cm

⊙ Taille : 6 000 × 6 000 pixels

⊙ Modalités : RVB



F. Mohammed and A. Idries and al., "UAVs for Smart Cities: Opportunities and Challenges".  
In *International Conference on Unmanned Aircraft Systems (ICUAS)*, 2014.

# Acquisition aérienne

⊙ Ville de Potsdam (Allemagne)

⊙ Taille : 6 000 × 6 000 pixels

⊙ Résolution au sol : 5 cm×5 cm

⊙ Modalités : RVB



F. Mohammed and A. Idries and al., "UAVs for Smart Cities: Opportunities and Challenges".  
In *International Conference on Unmanned Aircraft Systems (ICUAS)*, 2014.

*"Ensemble des objets ou dispositifs publics ou privés installés dans l'espace public et liés à une fonction ou à un service offert par la collectivité."*

Comment caractériser les objets urbains?

- ⊙ Objets relativement petits
- ⊙ Pas toujours référencés
- ⊙ Nombreux
- ⊙ Formes et apparences variables
- ⊙ Occultations partielles
- ⊙ Ombre
- ⊙ Évolution. . .

Notre cible:

- ⊙ **Les arbres d'alignement**

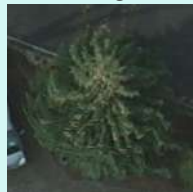


[https://fr.wikipedia.org/wiki/Mobilier\\_urbain](https://fr.wikipedia.org/wiki/Mobilier_urbain)



[https://www.plante-et-cite.fr/data/fichiers\\_ressources/patrimoine\\_arbore\\_lyon.pdf](https://www.plante-et-cite.fr/data/fichiers_ressources/patrimoine_arbore_lyon.pdf)

Arbre d'alignement



Poteau

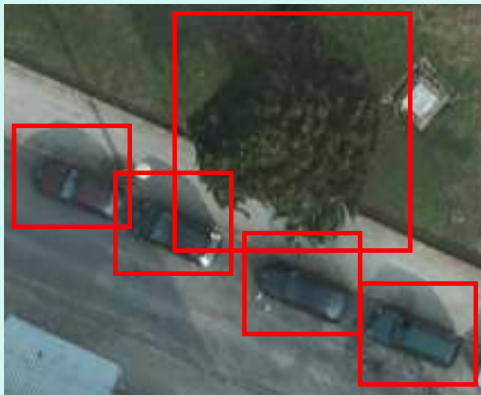


Plaque d'égout



Poubelles





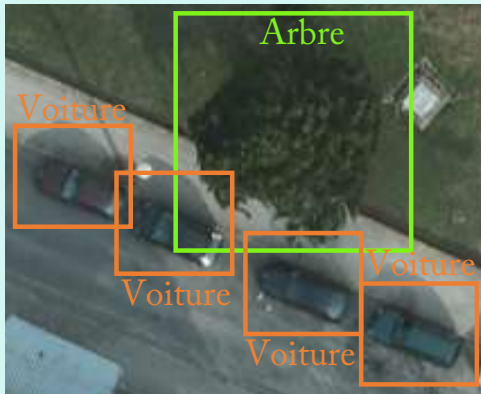
J. Ardila and W. Bijker and al., "Context-Sensitive Extraction of Tree Crown Objects in Urban Areas using VHR Satellite Images".  
In *International Journal of Applied Earth Observation and Geoinformation*, 2012.



G. Cheng and Z. Peicheng and al., "Learning Rotation-invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images".  
In *IEEE Transactions on Geoscience and Remote Sensing*, 2016.



G. Cheng and J. Han, "A Survey on Object Detection in Optical Remote Sensing Images".  
In *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016.



J. Ardila and W. Bijker and al., "Context-Sensitive Extraction of Tree Crown Objects in Urban Areas using VHR Satellite Images".  
In *International Journal of Applied Earth Observation and Geoinformation*, 2012.



G. Cheng and Z. Peicheng and al., "Learning Rotation-invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images".  
In *IEEE Transactions on Geoscience and Remote Sensing*, 2016.



G. Cheng and J. Han, "A Survey on Object Detection in Optical Remote Sensing Images".  
In *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016.

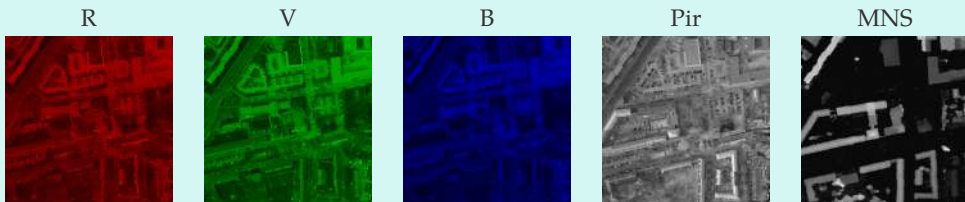




- ⊙ Image aérienne RVB
- ⊙ Histogramme de gradient orienté (HOG)
- ⊙ Réseau de Séparateur à Vaste Marge (SVM)
- ⊙ Intégration des CNN au réseau de SVM

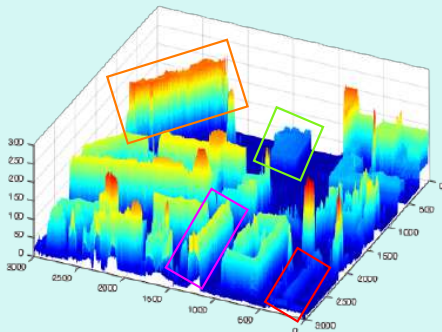
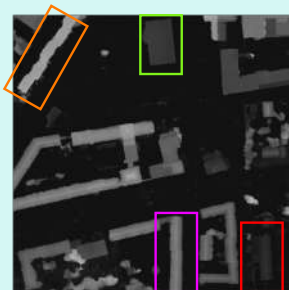


J. Pasquet, "Modélisation, détection et classification d'objets urbains à partir d'images photographiques aériennes".  
*Thèse de doctorat en informatique, Université de Montpellier, 2016.*



MNS

Projection en 3D



## Les données multimodales sont faciles d'accès Informations supplémentaires / complémentaires

- ⊙ Est-ce que les données multimodales améliorent-elles les performances?
- ⊙ Comment combiner les différentes modalités?



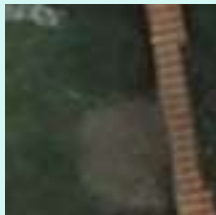
J. Wagner and V. Fischer and al., "Multispectral Pedestrian Detection using Deep Fusion Convolutional Neural Networks".  
In *European Symposium on Artificial Neural Networks (ESANN)*, 2016.

- ⊙ Comment faire lorsqu'elles ne sont pas disponibles toutes en même temps?



J. Hoffman and S. Gupta and al., "Learning with Side Information through Modality Hallucination".  
In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

RVB



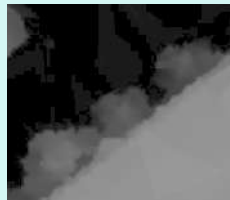
PirRV



PirRV



MNS



1. Introduction
2. Analyse des approches générales
3. Comment combiner les différentes modalités?
  - Contexte
  - Méthode et résultats expérimentaux
  - Conclusions
4. Modalités manquantes et réseaux de neurones convolutionnels
  - Introduction
  - Proposition et résultats expérimentaux
  - Conclusions
5. Conclusions & perspectives

## Analyse des approches générales

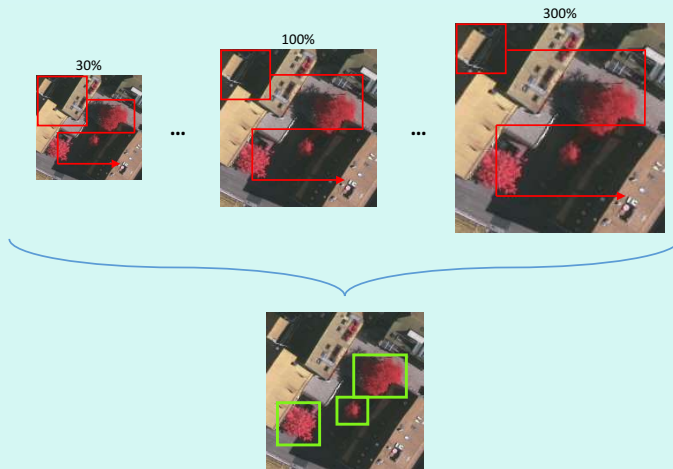
---

## Quelles sont les approches générales classiques?

### ⦿ Fenêtre glissante



V. Ferrari and L. Fevrier and al., "Groups of Adjacent Contour Segments for Object Detection".  
In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008.



## Quelles sont les approches générales classiques?

### ⦿ Fenêtre glissante



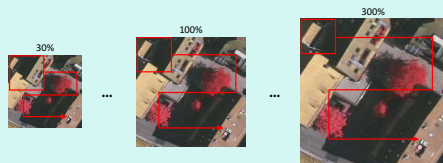
V. Ferrari and L. Fevrier and al., "Groups of Adjacent Contour Segments for Object Detection".  
In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008.

### ⦿ Apprentissage automatique

- Approche statistique
- Nécessite une base d'apprentissage annotée
- Obtient les meilleurs résultats



A. Maxwell and T. Warner and al., "Implementation of Machine-Learning Classification in Remote Sensing: An Applied Review".  
In *International Journal of Remote Sensing*, 2018.



## Quelles sont les approches générales classiques?

### ⦿ Fenêtre glissante



V. Ferrari and L. Fevrier and al., "Groups of Adjacent Contour Segments for Object Detection".  
In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008.

### ⦿ Apprentissage automatique

- Approche statistique
- Nécessite une base d'apprentissage
- Obtient les meilleurs résultats



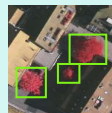
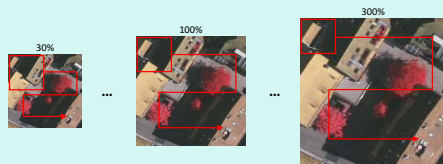
A. Maxwell and T. Warner and al., "Implementation of Machine-Learning Classification in Remote Sensing: An Applied Review".  
In *International Journal of Remote Sensing*, 2018.

## Quelle méthode d'apprentissage automatique utiliser?

- ⦿ Apprentissage automatique classique en deux étapes
- ⦿ Apprentissage automatique profond (*Deep Learning*)



Y. LeCun and Y. Bengio and al., "Deep Learning".  
*Nature*, 2015.





## Quelles sont les approches générales classiques?

### ⊙ Fenêtre glissante



V. Ferrari and L. Fevrier and al., "Groups of Adjacent Contour Segments for Object Detection".  
In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008.

### ⊙ Apprentissage automatique

- Approche statistique
- Nécessite une base d'apprentissage
- Obtient les meilleurs résultats



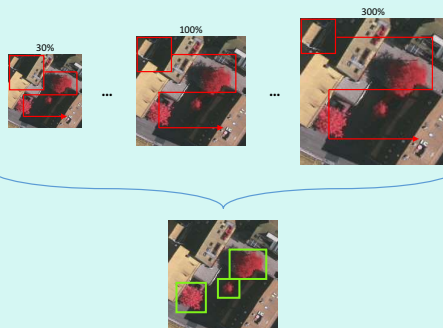
A. Maxwell and T. Warner and al., "Implementation of Machine-Learning Classification in Remote Sensing: An Applied Review".  
In *International Journal of Remote Sensing*, 2018.

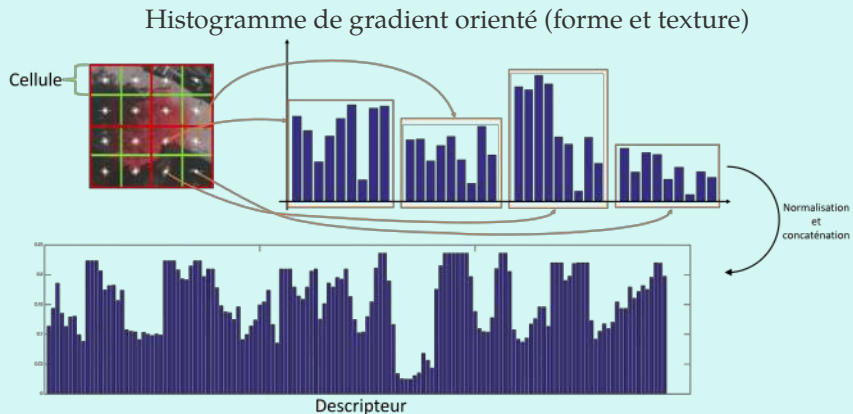
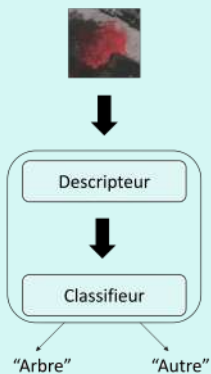
## Quelle méthode d'apprentissage automatique utiliser?

- ⊙ Apprentissage automatique classique en deux étapes
- ⊙ Apprentissage automatique profond (*Deep Learning*)



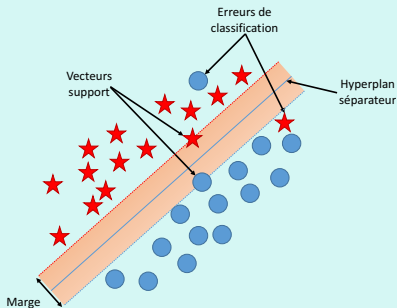
Y. LeCun and Y. Bengio and al., "Deep Learning".  
*Nature*, 2015.



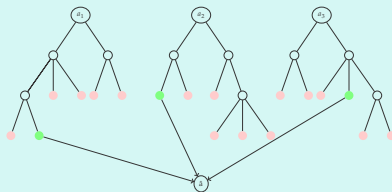


N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection".  
In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.


## SVM



## Forêt Aléatoire

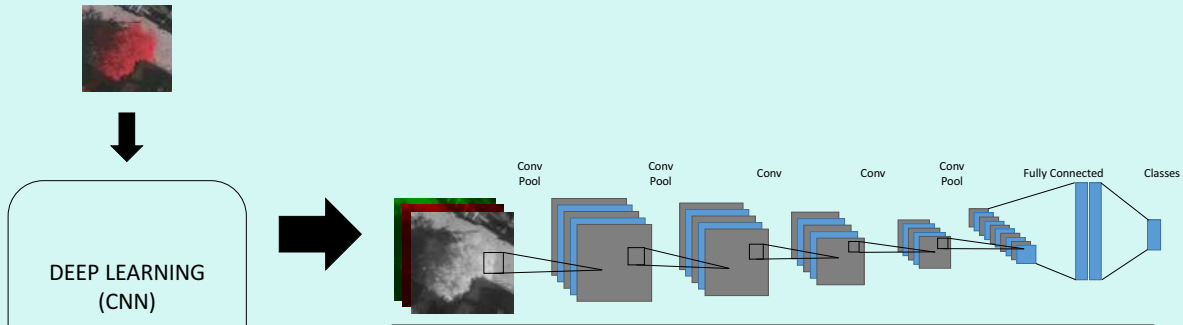


- ⊙ Robustes
  - Peu de données
  - Points aberrants
- ⊙ Peu de paramètres
- ⊙ Utilisées en télédétection

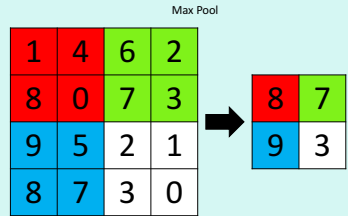
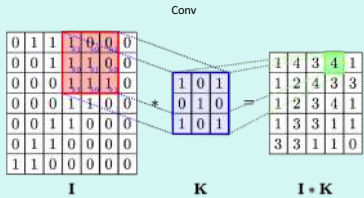
 G. Mountrakis and J. Im and al., "Support Vector Machines in Remote Sensing: A Review".  
In *ISPRS Journal of Photogrammetry and Remote Sensing*, 2011.

 M. Belgiu and L. Drăguț, "Random Forest in Remote Sensing: A Review of Applications and Future Directions".  
In *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016.

# Apprentissage automatique profond (*Deep learning*)



“Arbre”      “Autre”





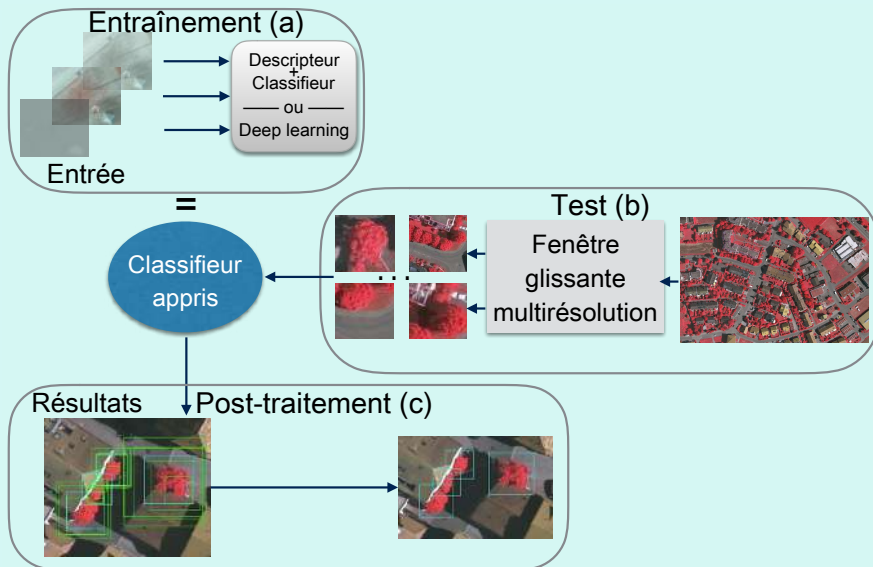
Y. LeCun and Y. Bengio and al., "Deep Learning".  
Nature, 2015.

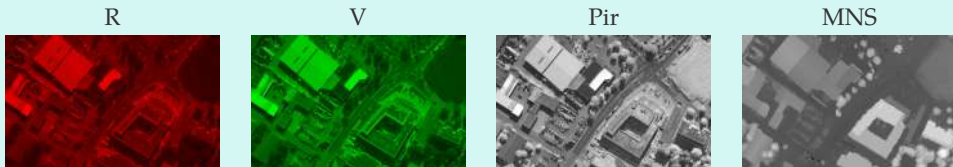
# Étude comparative

- ⊙ **Tâche:** Localisation d'arbres urbains
- ⊙ **Modalités:** PirRV et MNS

## Les algorithmes en compétition :

- ⊙ Apprentissage automatique classique en deux étapes
  - HOG + SVM
  - HOG + RF
- ⊙ *Deep learning*
  - AlexNet
    -  A. Krizhevský and I. Sutskever and al., "ImageNet Classification with Deep Convolutional Neural Networks".  
*Conference on Neural Information Processing Systems (NIPS), 2012.*
  - GoogLeNet
    -  C. Szegedy and W. Liu and Y. Jia and al., "Going Deeper with Convolutions".  
*In IEEE conference on Computer Vision and Pattern Recognition (CVPR), 2015.*





- ⊙ **Base de données publique** : Vaihingen (Allemagne)
- ⊙ **Modalités** : PirRV et MNS
- ⊙ **Entraînement** : 4 500 imagerie "arbre" et 36 000 imagerie "autre" redimensionnées en  $64 \times 64$  pixels
- ⊙ **Test** : 20 images de taille variable (de  $125 \times 150$  pixels à  $550 \times 725$  pixels) qui contiennent une centaine d'arbres

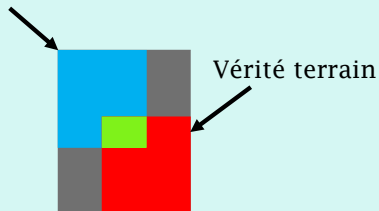


The Vaihingen data set was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) [3]: <http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html>.

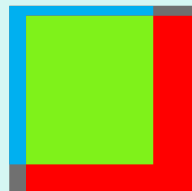
*Localisation* =  $\begin{cases} \text{Bonne} \\ \text{Mauvaise} \end{cases}$

Si  $\frac{\text{Intersection}}{\text{Union}} > 0.5$   
Sinon

Localisation



Mauvaise localisation



Bonne localisation



M. Everingham and L. Van Gool and al., "The Pascal Visual Object Classes (VOC) Challenge".  
*International Journal of Computer Vision*, 2010.



- ⊙ En vert : **Vrais Positifs**
- ⊙ En jaune : **Faux Positifs**
- ⊙ En bleu : **Faux Négatifs**



⊙  $Rappel = \frac{VP}{VP+FN}$

⊙  $Précision = \frac{VP}{VP+FP}$

⊙  $F - Mesure = \frac{2 \times Rappel \times Précision}{Rappel + Précision}$

# Résultats expérimentaux et conclusions

	HOG +SVM	HOG +RF	AlexNet	GoogLeNet
F-Mesure	2.88%	13.78%	30.63%	<b>33.32%</b>
Rappel	21%	33.47%	<b>49.28%</b>	48.96%
Précision	1.54%	10.47%	22.57%	<b>25.71%</b>

- ⊙ Le *Deep learning* a de meilleures performances
  - Gain d'environ +20% sur la F-Mesure par rapport à l'approche classique
- ⊙ GoogLeNet plus performant que AlexNet

⇒ Par la suite nous utiliserons des approches de *Deep Learning*

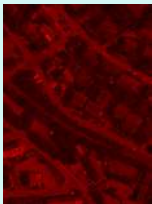


L. Zhang and L. Zhang and al., "Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art".  
In *IEEE Geoscience and Remote Sensing Magazine*, 2016.

Comment combiner les différentes  
modalités?

---

R



V



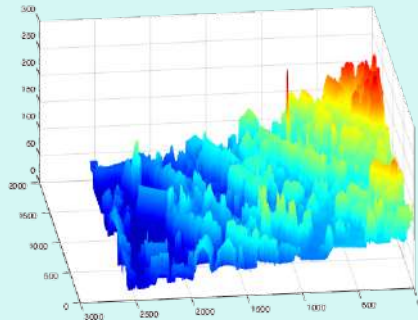
Pir



MNS



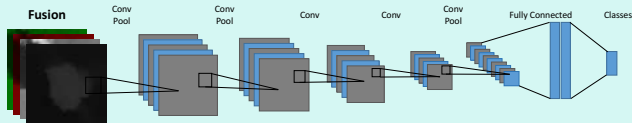
Projection en 3D



## Quelle est la difficulté?

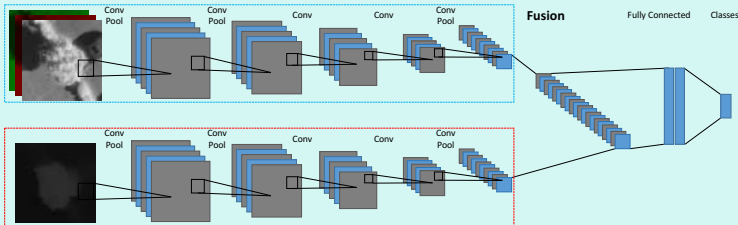
- ⊙ La fusion de différentes modalités est complexe
  - Informations de nature différentes
  - Échelles différentes
  - Problème de normalisation

## Early Fusion



- ⊙ 1 modalité = 1 canal
- ⊙ Normalisation par le CNN

## Intermediate / Late Fusion



- ⊙ 1 modalité = 1 branche
- ⊙ Combinaison avant la classification



J. Wagner and V. Fischer and al., "Multispectral Pedestrian Detection using Deep Fusion Convolutional Neural Networks".  
*European Symposium on Artificial Neural Networks (ESANN), 2016.*



K. Aderghal and J. Benois-Pineau and al., "FuseMe: Classification of sMRI images by fusion of Deep CNNs in 2D+  $\epsilon$  projections".  
*International Workshop on Content-Based Multimedia Indexing (CBMI), 2017.*

- ⊙ **Base de données** : Vaihingen
- ⊙ **Modalités** : R, V, Pir, MNS + *Normalized Difference Vegetation Index* (NDVI)

$$NDVI = \frac{Pir - R}{Pir + R}$$

- ⊙ **Entraînement** : 6 000 imagerie "arbre" and 40 000 imagerie "autre" redimensionnées en  $64 \times 64$  pixels
- ⊙ **Test** : 20 images de taille variable (de  $125 \times 150$  pixels à  $550 \times 725$  pixels) qui contiennent une centaine d'arbres



F. Krieglér and W. Malila and al., "Preprocessing Transformations and Their Effects on Multispectral Recognition".  
In *Remote Sensing of Environment*, 1969.

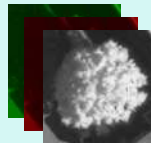
# Résultats en utilisant une seule modalité

Résultats en utilisant une seule modalité

Modalité	PirRV	MNS	NDVI
F-Mesure	60.45%	62.47%	<b>63.97%</b>
Rappel	57.89%	57.62%	<b>62.34%</b>
Précision	63.44%	<b>68.56%</b>	67.04%

- ⊙ Meilleure précision avec le MNS
- ⊙ NDVI > PirRV
- ⊙ Meilleure F-Mesure avec le NDVI

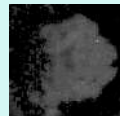
PirRV

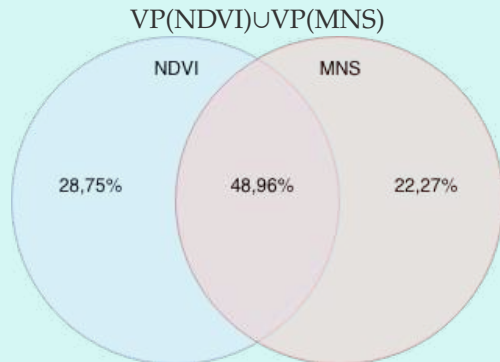
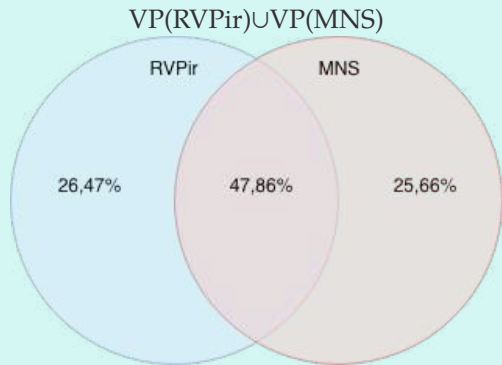


MNS



NDVI





- ⊙ 50% des arbres sont trouvés dans les deux modalités
- ⊙ Les autres 50% sont distribués dans les deux modalités  
⇒ Cela montre l'intérêt de combiner différentes modalités



## Early Fusion

	PirRV+MNS	NDVI+MNS
F-Mesure	67.12%	<b>75.30%</b>
Rappel	65.40%	<b>68.37%</b>
Précision	69.54%	<b>84.11%</b>

## Intermediate Fusion

	PirRV+MNS	NDVI+MNS
F-Mesure	62.14%	<b>72.57%</b>
Rappel	62.54%	<b>70.99%</b>
Précision	62.65%	<b>74.83%</b>

## 1 modalité

Modalité	NDVI
F-Mesure	63.97%
Rappel	62.34%
Précision	67.04%

- ⊙ L'utilisation de données multimodales apporte un gain de +11% sur la F-Mesure
- ⊙ **Early Fusion** : meilleure F-Mesure, meilleure Précision
  - **On réduit le nombre de Faux Positifs**
- ⊙ **Intermediate Fusion** : meilleur Rappel
  - **Retrouve plus d'arbres que l'architecture Early Fusion**



N. Audebert and B. Le Saux and al., "Beyond RGB: Very High Resolution Urban Remote Sensing with Multimodal Deep Networks".  
*ISPRS Journal of Photogrammetry and Remote Sensing*, 2018.

- ⊙ Le NDVI+MNS donne les meilleurs résultats

⇒ Intégrer la détection dans le CNN (Faster R-CNN, YOLO, Mask R-CNN. . .)

⇒ Expérimenter sur d'autres bases de données



S. Ren and K. He and al., "Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks".  
In *Advances in Neural Information Processing Systems (NIPS)*, 2015.



J. Redmon and S. Divvala and al., "You Only Look Once: Unified, Real-time Object Detection".  
In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.



K. He and G. Gkioxari and al., "Mask R-CNN".  
In *IEEE International Conference on Computer Vision (ICCV)*, 2017.

## Modalités manquantes et réseaux de neurones convolutionnels

---

Problème avec des données multimodales = supposées complètes

## Objectif?

- ⊙ Traiter des données complètes et incomplètes

## Pourquoi?

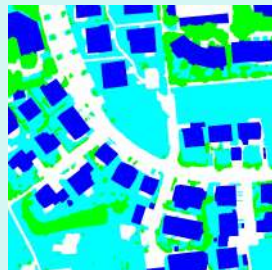
- ⊙ Cas réels
  - Panne du capteur
  - Fréquences d'acquisition différentes

## Comment?

- ⊙ Une architecture adaptée

## Tâche?

- ⊙ Segmentation sémantique

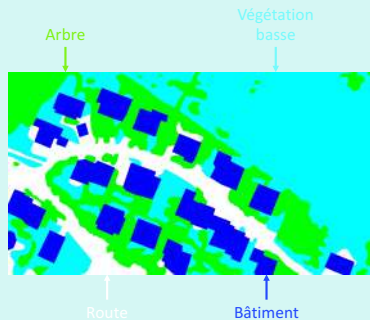


Qu'est-ce que la segmentation sémantique?

- ⊙ Étiquetage supervisé de chaque pixel
- ⊙ Le nombre de classes est connu

Pourquoi faire de la segmentation sémantique?

- ⊙ Base de données multimodales annotée accessible
- ⊙ Étudier un autre problème permettant de localiser les arbres



## Quelles sont les méthodes classiques?

- ⊙ Supprimer les données/modalités incomplètes
- ⊙ Générer les modalités incomplètes

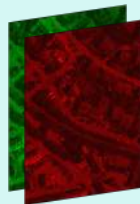


I. Goodfellow and J. Pouget-Abadie and al., "Generative Adversarial Nets".  
In *Advances in Neural Information Processing Systems (NIPS)*, 2014.

- ⊙ Guider l'espace de représentation des données



J. Hoffman and S. Gupta and al., "Learning with Side Information through Modality Hallucination".  
In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.



## Quelles sont les méthodes classiques?

- ⊙ Supprimer les données/modalités incomplètes

- ⊙ Générer les modalités incomplètes

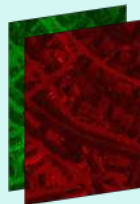


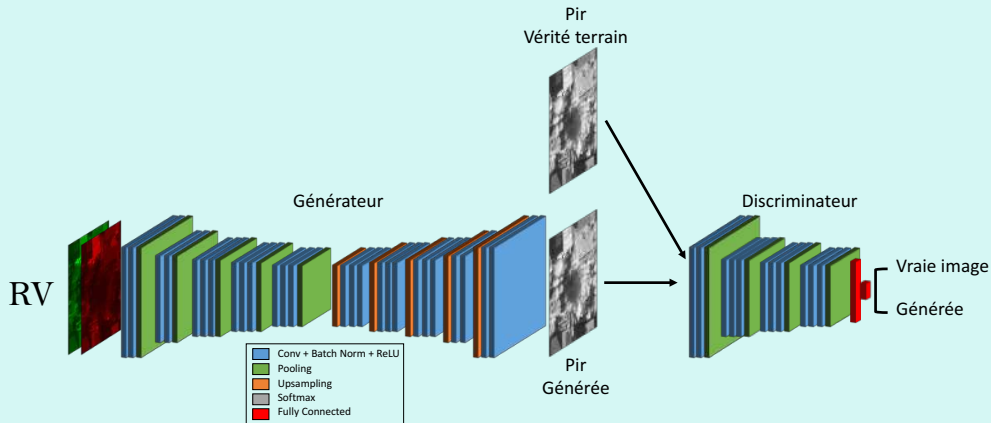
I. Goodfellow and J. Pouget-Abadie and al., "Generative Adversarial Nets".  
In *Advances in Neural Information Processing Systems (NIPS)*, 2014.

- ⊙ Guider l'espace de représentation des données



J. Hoffman and S. Gupta and al., "Learning with Side Information through Modality Hallucination".  
In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.



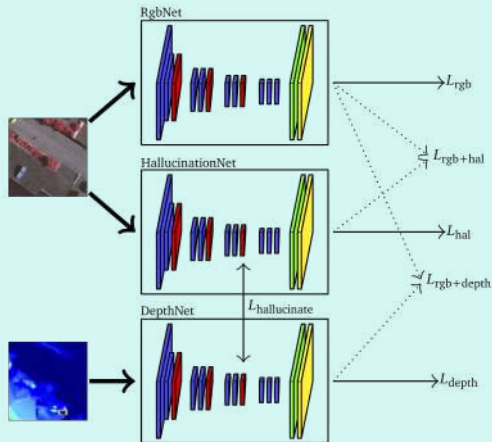


- ⊙ Réseaux antagonistes génératifs (GAN)
- ⊙ Générer la modalité incomplète à partir des modalités complètes



X. Liu and Y. Wang and al., "Psgan: A Generative Adversarial Network for Remote Sensing Image Pan-Sharpning".  
In *IEEE International Conference on Image Processing (ICIP)*, 2018.





- ⊙ Apprentissage sur des données complètes
- ⊙ Test sur des données incomplètes
- ⊙ 1 réseau par modalité
- ⊙ 1 réseau HallucinationNet
  - Corréler les modalités



J. Hoffman and S. Gupta and al., "Learning with Side Information through Modality Hallucination".  
In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.



M. Kampffmeyer and A.-B. Salberg and al., "Urban Land Cover Classification with Missing Data Using Deep Convolutional Neural Networks".  
In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017.

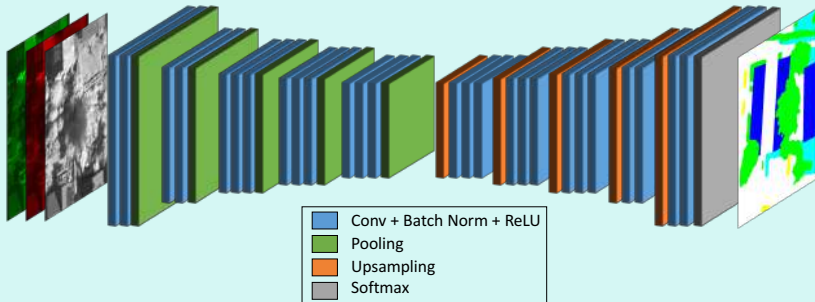
# Limites de ces méthodes

- ⊙ Supprimer les données/modalités incomplètes
  - Supprime de l'information utile
- ⊙ Générer les modalités incomplètes
  - Peut introduire un biais si peu de données complètes
- ⊙ Guider l'espace de représentation des données
  - Nécessite des données complètes pour l'apprentissage

## ⇒ Notre approche

- ⊙ **Ne supprime aucune information**
- ⊙ **N'introduit pas de biais**
- ⊙ **Gère l'apprentissage et le test**

## PirRV



⊙ 1 encodeur

⊙ 1 décodeur

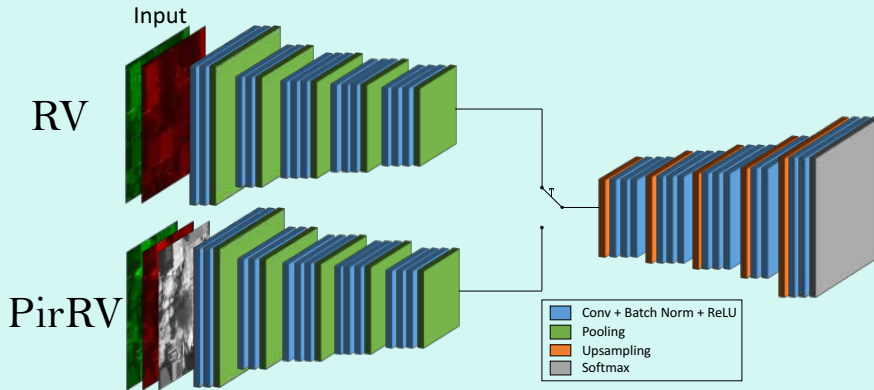


H. Noh and S. Hong and al., "Learning Deconvolution Network for Semantic Segmentation".  
In *IEEE International Conference on Computer Vision (ICCV)*, 2015.



I. Kokkinos, "UberNet: Training a Universal Convolutional Neural Network for Low-, Mid-, and High-Level Vision Using Diverse Datasets and Limited Memory".

In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

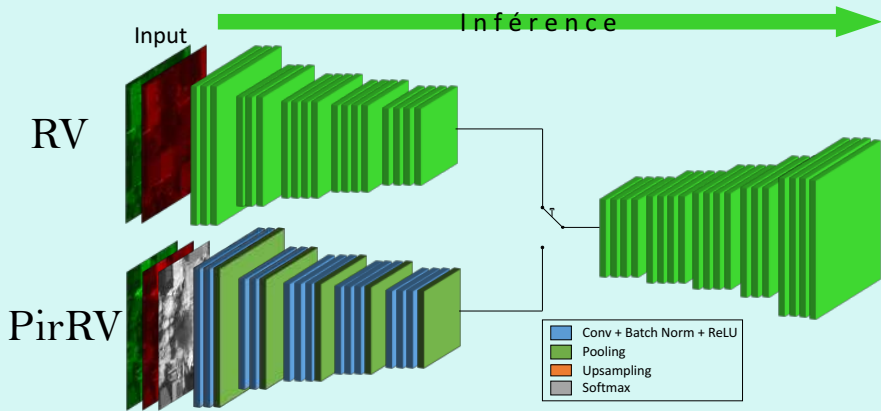


⊙ 2 sous-réseaux

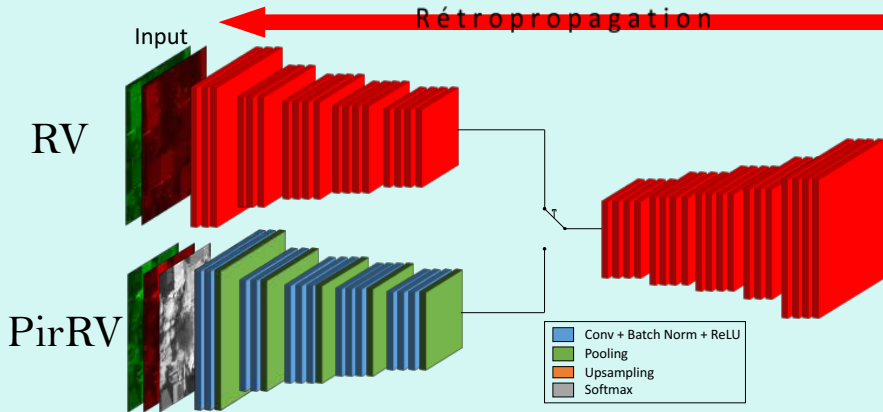
⊙ 1 sortie commune

⊙ Espace latent commun

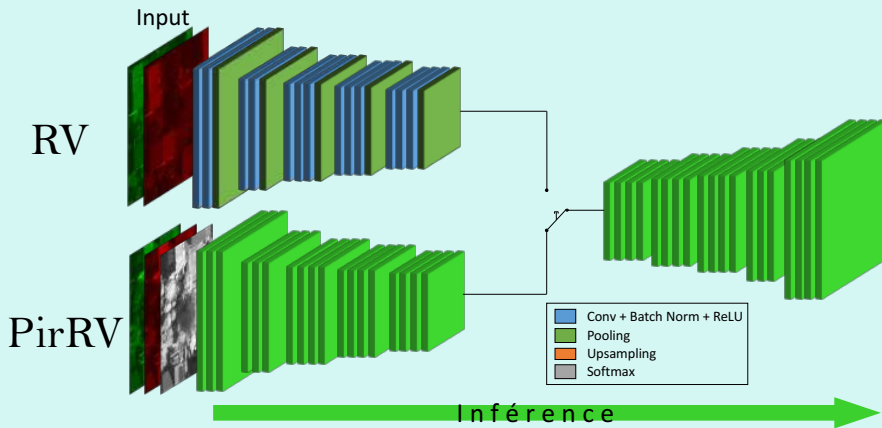
⊙ Apprentissage alterné



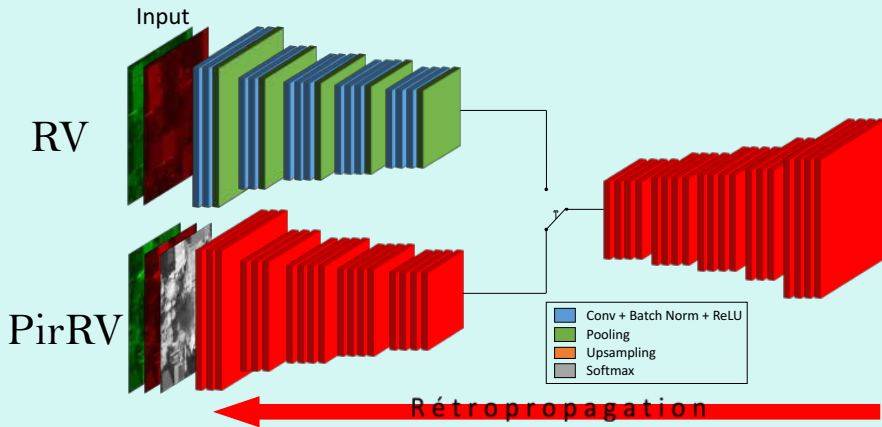
⊙ Inférence avec l'encodeur RV et le décodeur



⊙ Rétropropagation sur le décodeur et l'encodeur RV



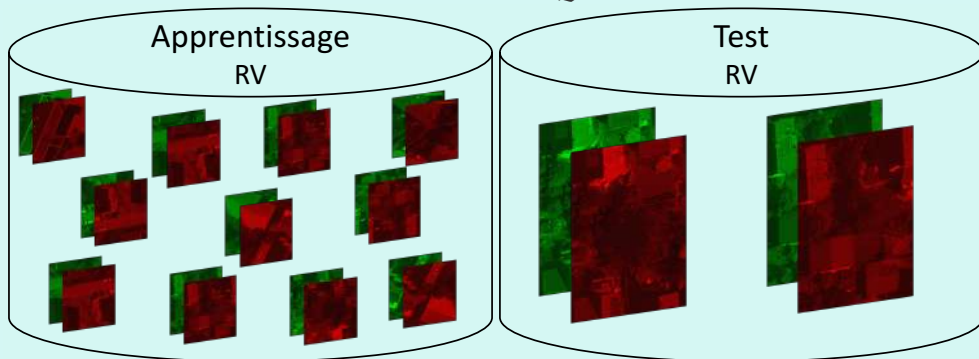
⊙ Inférence avec l'encodeur PirRV et le décodeur



⊙ Rétropropagation sur le décodeur et l'encodeur PirRV

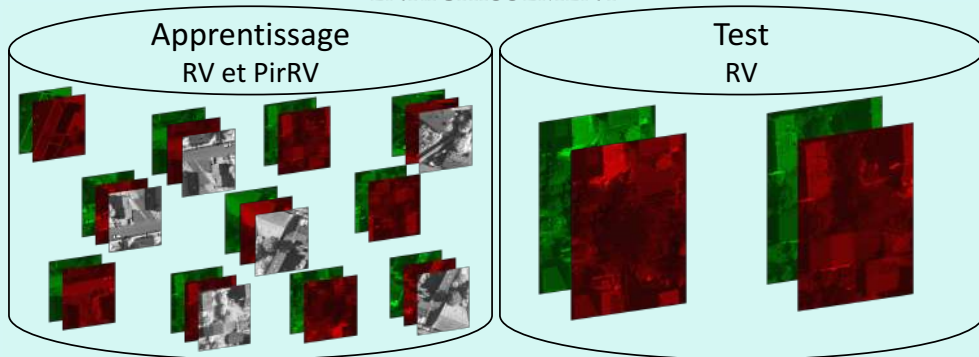


## BASE CLASSIQUE



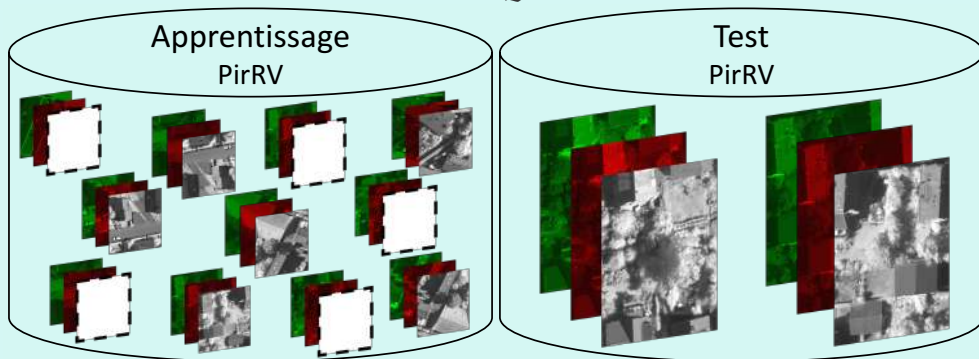
- ⊙ Apprentissage et test sur les modalités RV
- ⊙ Accès à une autre modalité?

## ENRICHISSEMENT



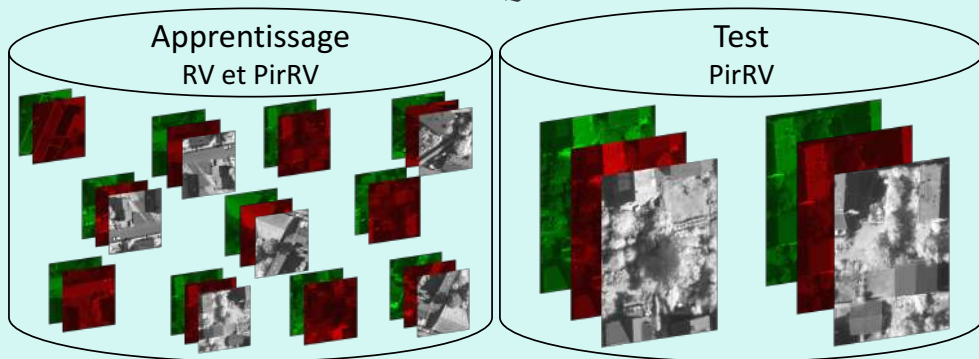
- ⊙ Accès à la modalité Pir sur certaines données
- ⊙ Enrichir le modèle quand la modalité Pir est disponible

## MANQUE



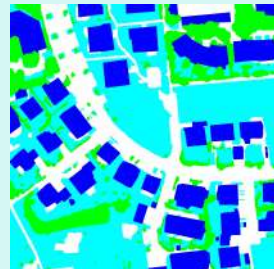
- ⊙ Apprentissage et test sur les modalités PirRV
- ⊙ Problème avec la modalité Pir sur certaines données

## MANQUE

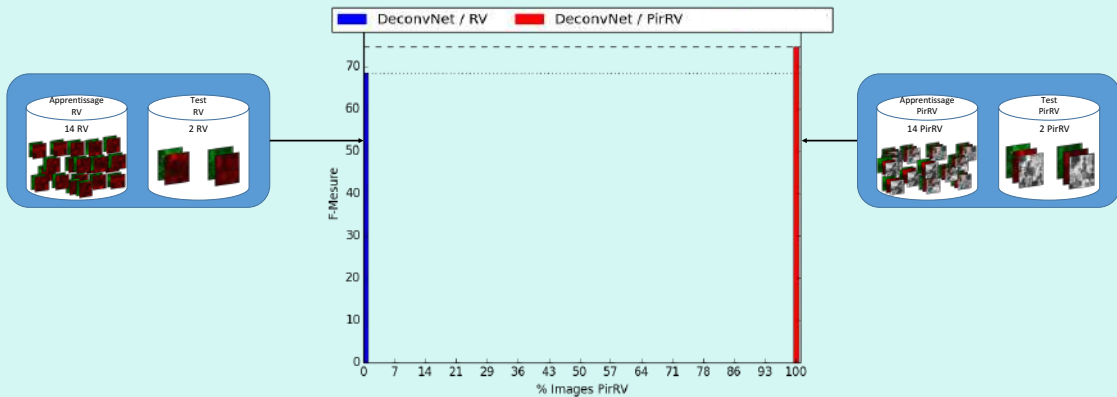


- ⊙ Utilisation des modalités disponibles
- ⊙ Compenser l'absence de la modalité Pir avec les autres modalités

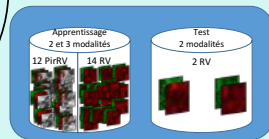
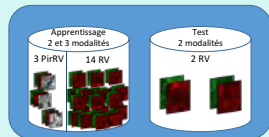
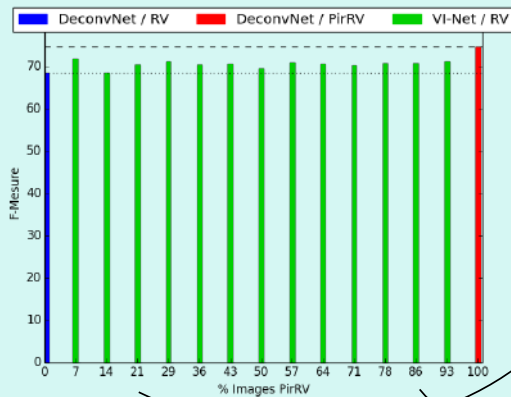
- ⊙ **Base de données** : Vaihingen
- ⊙ **Modalités** : Pir,R,V
- ⊙ **4 classes** : arbre, route, bâtiment, végétation basse
- ⊙ **Base d'apprentissage** : imagerie de  $320 \times 320$  pixels extraites aléatoirement à chaque itération sur 14 images d'environ  $2500 \times 2000$  pixels
- ⊙ **Base de test** : 2 images de  $1388 \times 1388$  pixels



# Enrichissement



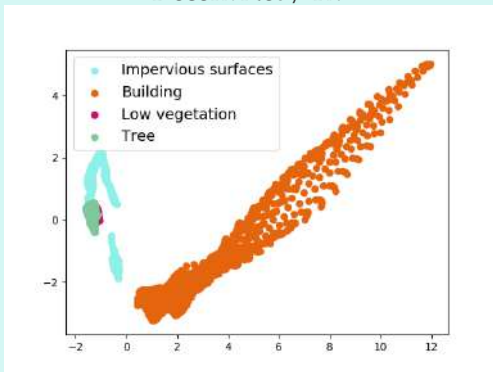
# Enrichissement



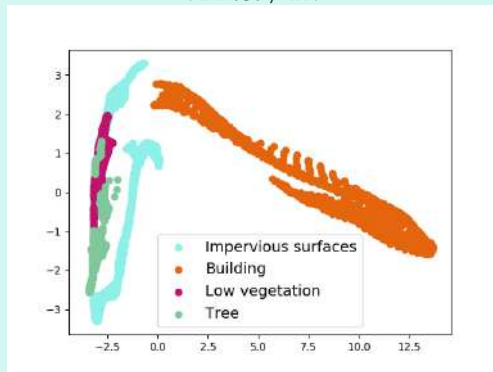
⊙ Notre stratégie améliore les performances

# Enrichissement

DeconvNet / RV



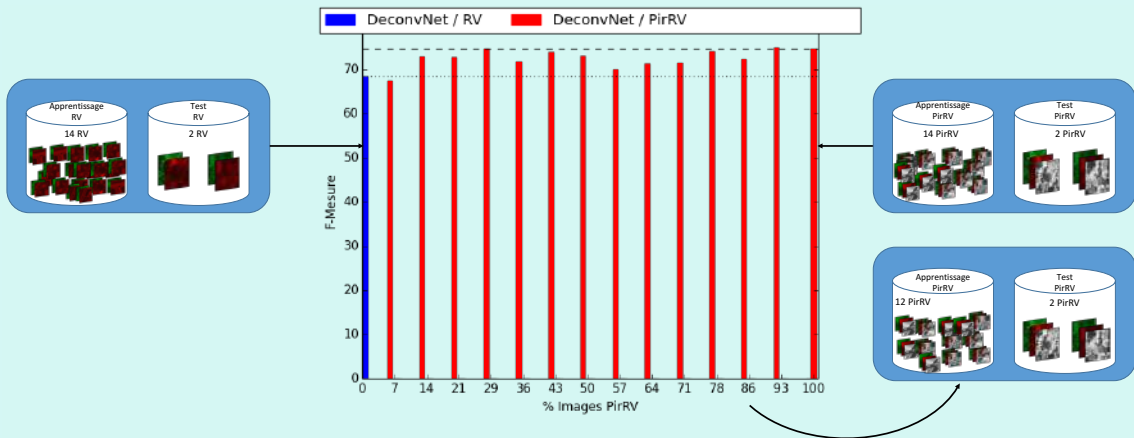
VI-Net / RV



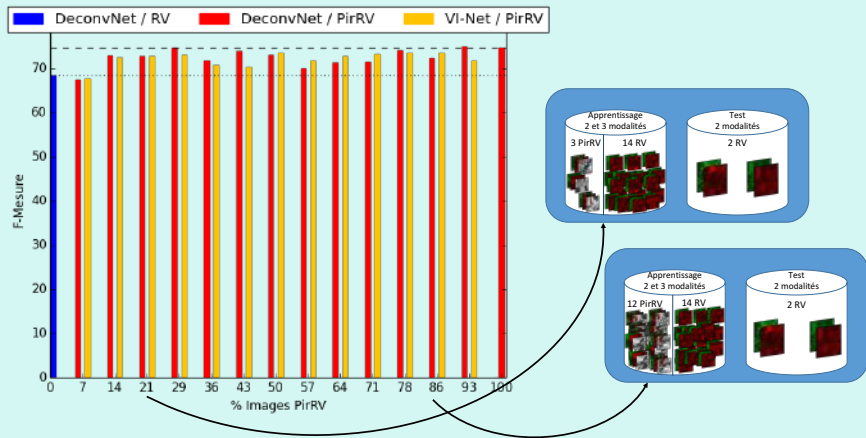
- ⊙ Extraction sur la dernière couche du décodeur
- ⊙ Analyse en Composantes Principales (ACP)
- ⊙ Représentation en 2 dimensions de nos caractéristiques

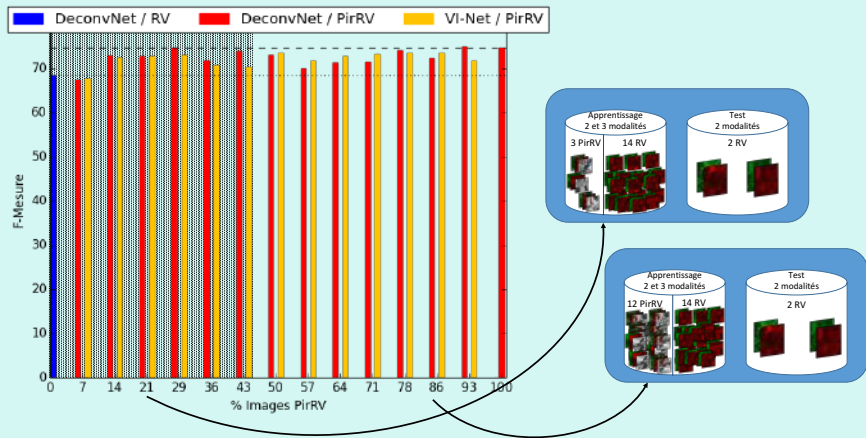
⇒ VI-Net améliore la séparation des classes "arbre" et "végétation basse"





⊙ Résultats de référence classique ⇒ nombre fixe de modalités (stratégie de suppression)





⊙ VI-Net améliore les performances

- ⊙ VI-Net améliore la F-mesure sur le scénario de l'**enrichissement** (test sur les images incomplètes)
- ⊙ Plus contrasté pour le scénario du **manque** (test sur les images complètes)
  - Amélioration de la F-mesure lorsque 50% des images sont complètes

- ⇒ Stratégie mixte sur le scénario du **manque**
- ⇒ Expérimenter sur d'autres bases de données
- ⇒ Expérimenter avec d'autres modalités
- ⇒ Expérimenter sur la tâche de localisation d'objets

## Conclusions & perspectives

---

## ① Étude comparative des méthodes générales



L. Pibre, M. Chaumont, G. Subsol, D. Ienco, and M. Derras, "Detection of Urban Trees in Multiple-Source Aerial Data (Optical, Infrared, DSM)". In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) M.Sc./Ph.D. Forum*, 2017.



L. Pibre, M. Chaumont, G. Subsol, D. Ienco, and M. Derras, "Détection d'arbres urbains à partir de données aériennes multi-sources (optique, infrarouge, MNS)". In *Atelier Télédétection pour l'Etude des Milieux Urbains (TEMU)*, 2017.

## ② Expérimentation de 2 approches pour combiner des données multimodales

- L'approche *Early Fusion* permet d'obtenir les caractéristiques les plus discriminantes
- L'approche *Intermediate Fusion* permet de trouver plus d'arbres
- Le choix de l'approche reste un problème ouvert



L. Pibre, M. Chaumont, G. Subsol, D. Ienco, and M. Derras, "How to Deal with Multi-source Data for Three Detection Based on Deep Learning". In *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2017.

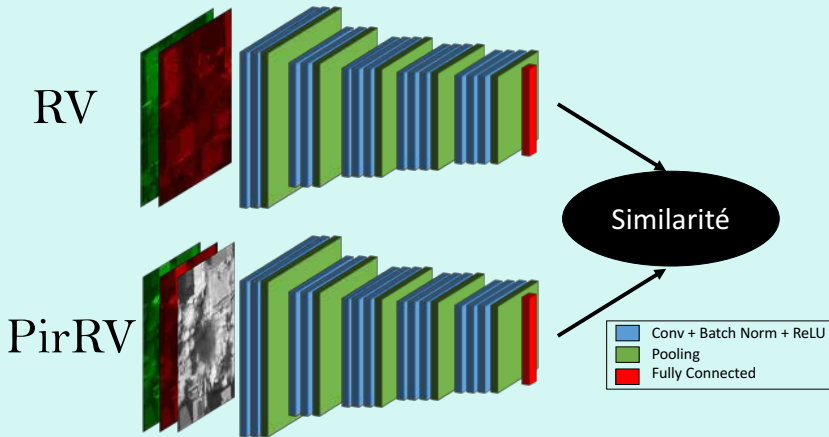
## ③ Méthode pour gérer des modalités manquantes durant l'entraînement et le test

- Améliore les performances sans supprimer ni générer les modalités incomplètes



Soumis : L. Pibre, M. Chaumont, G. Subsol, D. Ienco, and M. Derras, "Towards Dealing with Missing Modalities with Convolutional Neural Networks". In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019.

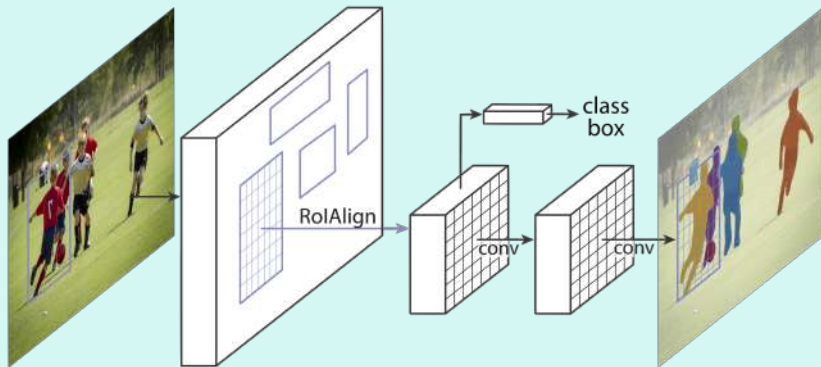
## ④ Transfert industriel en cours



- ⊙ Pré-entraîner les encodeurs sur les données complètes
- ⊙ Contraindre l'espace latent avec la distance



J. Bromley and I. Guyon and al., "Signature Verification using a "Siamese" Time Delay Neural Network".  
In *Advances in Neural Information Processing Systems (NIPS)*, 1994.



- ⊙ Détection et segmentation de chaque objet distinct
- ⊙ Tâche réalisée sur des données monomodales
- ⊙ Nécessite une annotation objet et pixel



K. He and G. Gkioxari and al., "Mask R-CNN".  
In *IEEE International Conference on Computer Vision (ICCV)*, 2017.



A. Böhm and A. Ücker and al., "ISOODL: Instance Segmentation of Overlapping Biological Objects using Deep Learning".  
In *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2018.



- ⊙ Détection des plaques d'égout
  - Collaboration avec le laboratoire HydroSciences



B. Commandre, D. En-Nejjary, **L. Pibre**, M. Chaumont, C. Delenne, N. Chahinian, "Manhole Cover Localization in Aerial Images with a Deep Learning Approach".

In *International Society for Photogrammetry and Remote Sensing Workshop (ISPRS)*, 2017.

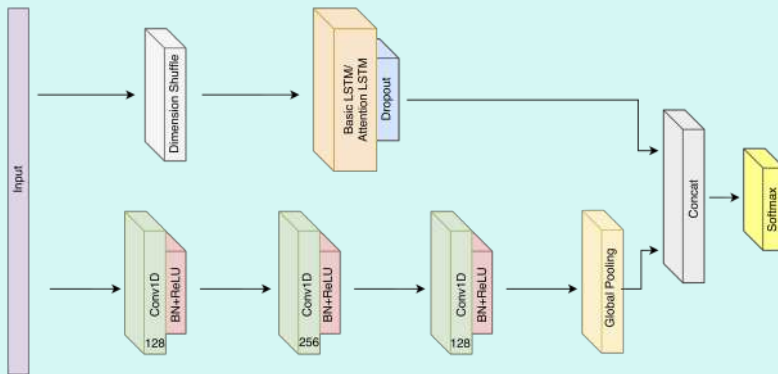


B. Commandre, D. En-Nejjary, **L. Pibre**, M. Chaumont, G. Subsol, L. Deruelle, M. Derras, C. Delenne, N. Chahinian, "Détection de regards de visite sur des images à très haute résolution spatiale par une méthode d'apprentissage".

In *Atelier Télédétection pour l'Etude des Milieux Urbains (TEMU)*, 2018.

Localisation d'objets urbains à partir de sources multiples dont des images aériennes, et analyse de leur état, leur évolution temporelle ou leur relation avec leur voisinage

Localisation d'objets urbains à partir de sources multiples dont des images aériennes, et analyse de leur état, leur **évolution temporelle** ou leur relation avec leur voisinage



F. Karim and S. Majumdar and al., "LSTM fully convolutional networks for time series classification".  
In *IEEE Access*, 2018.



J. Núñez and R. Cabido and al., "Multiview 3D human pose estimation using improved least-squares and LSTM networks".  
In *Neurocomputing*, 2019.

Je vous remercie  
de votre attention

# Références |



L. Breiman.

Random forests.

*Machine Learning*, 45(1):5–32, 2001.



J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah.

Signature verification using a " siamese" time delay neural network.

In *Advances in neural information processing systems*, pages 737–744, 1994.



M. Cramer.

The dgpf-test on digital airborne camera evaluation - overview and test design.

*Photogrammetrie-Fernerkundung-Geoinformation*, 2010(2):73–82, 2010.



I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio.

Generative adversarial nets.

In *Advances in Neural Information Processing Systems (NIPS)*, pages 2672–2680, 2014.



J. Hoffman, S. Gupta, and T. Darrell.

Learning with side information through modality hallucination.

In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 826–834, Las Vegas, USA, June 2016.



M. Kampffmeyer, A.-B. Salberg, and R. Jenssen.

Urban land cover classification with missing data using deep convolutional neural networks.

In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Fort Worth, USA, 2017.



A. Krizhevský, I. Sutskever, and G. E. Hinton.

ImageNet classification with deep convolutional neural networks.

In *Conference on Neural Information Processing Systems (NIPS)*, pages 1097–1105, Lake Tahoe, USA, Dec. 2012.

# Références ||



Y. LeCun, Y. Bengio, and G. Hinton.

Deep learning.

*Nature*, 52(8):436–444, 2015.



X. Liu, Y. Wang, and Q. Liu.

Psgan: A Generative Adversarial Network for Remote Sensing Image Pan-Sharpening .

In *IEEE International Conference on Image Processing (ICIP)*, pages 873–877, Athens, Greece, October 2018.



H. Noh, S. Hong, and B. Han.

Learning deconvolution network for semantic segmentation.

In *IEEE International Conference on Computer Vision (ICCV)*, pages 1520–1528, Las Condes, Chile, Dec. 2015.



Z. Qiang, Y. Mei-Chen, C. Kwang-Ting, and S. Avidan.

Fast human detection using a cascade of histograms of oriented gradients.

In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 1491–1498, New York, USA, June 2006.



J. W. Rouse Jr, R. H. Haas, J. A. Schell, and D. W. Deering.

Monitoring vegetation systems in the Great Plains with ERTS.

In *Proceedings of the Third ERTS Symposium*, volume 1, pages 309–317. NASA, Dec. 1973.



H. Shen, X. Li, Q. Cheng, C. Zeng, G. Yang, H. Li, and L. Zhang.

Missing information reconstruction of remote sensing data: A technical review.

*IEEE Geoscience and Remote Sensing Magazine*, 3(3):61–85, 2015.



C. Szegedy, W. Liu, Y. Jia, P. Sermanet, et al.

Going deeper with convolutions.

In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, Boston, USA, June 2015.

# Références III



L. Tran, X. Liu, J. Zhou, and R. Jin.

Missing modalities imputation via cascaded residual autoencoder.

In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1405–1414, Honolulu, Hawaii, July 2017.



C. J. Tucker.

Red and photographic infrared linear combinations for monitoring vegetation.

*Remote Sensing of Environment*, 8(2):127–150, 1979.



V. Vapnik.

*The Nature of Statistical Learning Theory*.

Springer-Verlag New York, Inc., New York, NY, USA, 1995.