HOW TO DEAL WITH MULTI-SOURCE DATA FOR TREE DETECTION BASED ON DEEP LEARNING

Lionel Pibre[a,e], Marc Chaumont[a,b], Gérard Subsol[a,c], Dino Ienco[d] and Mustapha Derras[e]

[a] LIRMM, Université de Montpellier, [b] Université de Nîmes, [c] CNRS, [d] IRSTEA, [e] Berger-Levrault

2017/11/14

Context

What is our goal?

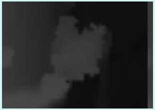
- Detect and localize trees from aerial images

Why?

- Manage trees in cities

How?

- With Deep Learning
- With Multi-source data



LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." Nature 521, no. 7553, pp. 436-444, 2015.

Context

What is the difficulty?

- It is complex to merge several information sources
- Trees are often regrouped and occluded

Some solutions exist[1]

- But not with multi-source data

[1] Yang, Lin, Xiaqing Wu, Emil Praun, and Xiaoxu Ma. "Tree detection from aerial imagery." In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 131-137. ACM, 2009.

Method

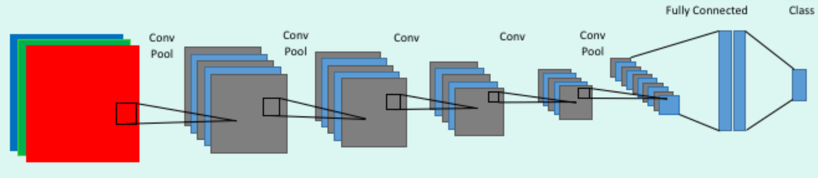


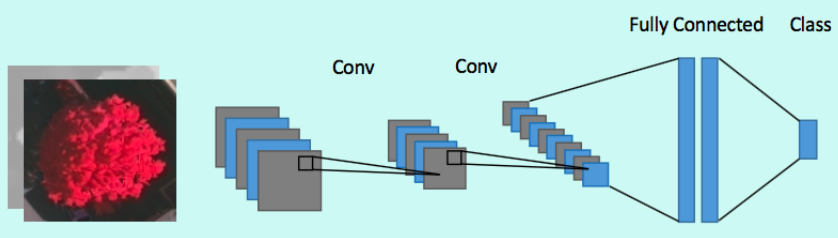
Figure: AlexNet network.

Two methods are tested:

- The Early Fusion
  - Each sensor source is treated as a channel
  - Give it through a classical CNN
- The Late Fusion[2]
  - A subnet for each sensor source

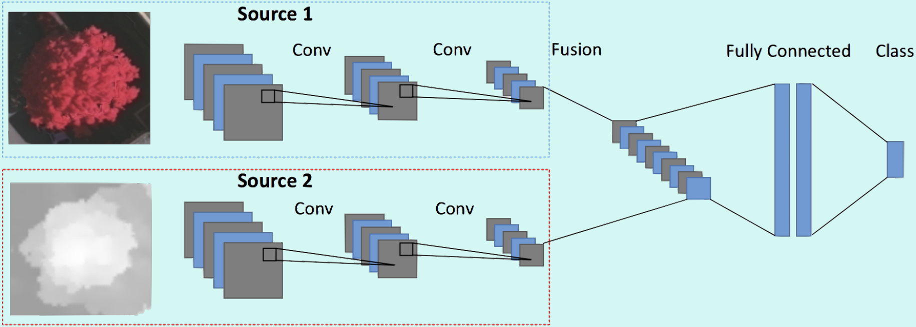
[2] J. Wagner, V. Fischer, M. Herman and S. Behnke, "Multispectral pedestrian detection using deep fusion convolutional neural networks", in *European Symp. on Artificial Neural Networks (ESANN)*, Bruges, Belgium, 2016.

Early Fusion



Early Fusion diagram.

Late Fusion



Late Fusion diagram.

Experimental Settings - Database

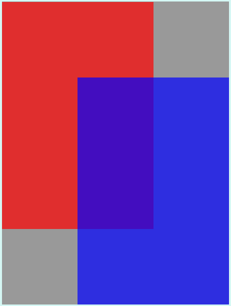
- **Database**: Vaihingen
- **Type of images**: Red, Green and Near-Infrared (RGNIR) and Digital Surface Model (DSM). We also generated Normalized Difference Vegetation Index (NDVI) images (grayscale) from the RGNIR images.

$$NDVI = \frac{NIR - R}{NIR + R} \tag{1}$$

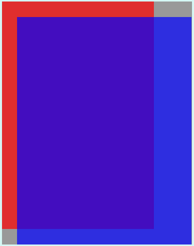
- **Training**: 6,000 **"tree"** thumbnails and 40,000 **"other"** thumbnails. The thumbnail size is 64×64 pixels.
- **Testing**: 20 images of variable size (from 125×150 pixels up to 550×725 pixels) and that contain about hundred trees.

Experimental Settings - Evaluation

$$label = \begin{cases} tree & \text{If} \quad \frac{area(detection \cap ground\ truth)}{area(detection \cup ground\ truth)} > 0.5 \\ \\ not\ tree & \text{If} \quad \frac{area(detection \cap ground\ truth)}{area(detection \cup ground\ truth)} \leq 0.5 \end{cases} \tag{2}$$

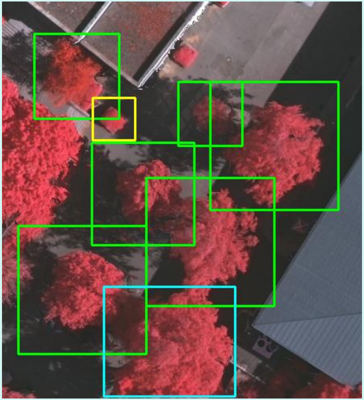


Example when the label will be "not tree".



Example when the label will be "tree".

Experimental Settings - Evaluation



- In green: **True Positives**
- In yellow: **False Positives**
- In blue: **False Negatives**

Experimental Settings - Evaluation

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{3}$$

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \tag{4}$$

$$F - Measure_{max} = \frac{2Recall * Precision}{Recall + Precision} \tag{5}$$

- $TruePositives$: Yeah! we really found a tree
- $FalseNegatives$: Oups, we missed this one
- $FalsePositives$: Oh really? Did you really think THAT was a tree?

Results using one source

Results using **one** source.

Source	RGNIR	DSM	NDVI
F-Measure$_{max}$	60.45%	62.47%	**63.97%**
Recall	57.89%	57.62%	**62.34%**
Precision	63.44%	**68.56%**	67.04%

- The DSM allows to obtain the best precision
- NDVI gives better results than RGNIR and the best F-Measure$_{max}$

Early Fusion and Late Fusion

Results using multi-source data and the **Early Fusion** architecture.

Early Fusion	RGNIR+DSM	NDVI+DSM
F-Measure$_{max}$	67.12%	**75.30%**
Recall	65.40%	**68.37%**
Precision	69.54%	**84.11%**

Results using multi-source data and the **Late Fusion** architecture.

Late Fusion	RGNIR+DSM	NDVI+DSM
F-Measure$_{max}$	62.14%	**72.57%**
Recall	62.54%	**70.99%**
Precision	62.65%	**74.83%**

Discussion Early Fusion and Late Fusion

- From one source to multi-source, we increase the f-measure$_{max}$ by 11%
- No matter the architecture used, NDVI+DSM gives the best results
- The Early Fusion allows us to obtain the best performances
- We have an important increase of the precision when we use the Early Fusion
  - 74% up to 84% with NDVI+DSM
  - 62% up to 69% with RGNIR+DSM
- The recall does not increase with the Early Fusion
- **We decrease the number of False Positives with the Early Fusion architecture**

Complementarity between sources

Results of the correlation between each source.

Sources	RGNIR/DSM		NDVI/DSM	
Correlation	47.86%		48.96%	
Distribution	26.47%	25.66%	28.75%	22.27%

- 50% of the trees are found in both sources
- The remaining 50% is distributed in the two sources and thus shows us the utility of combining several sources

Conclusions

- The Early Fusion gives better performances than the Late Fusion
- NDVI allows us to obtain the best performances
- This highlights the importance of the data that are used to learn a model with a CNN (RGNIR is not enough)
- We show the effectiveness of CNNs in merging different information with a performance gain exceeding 10%

THE
END