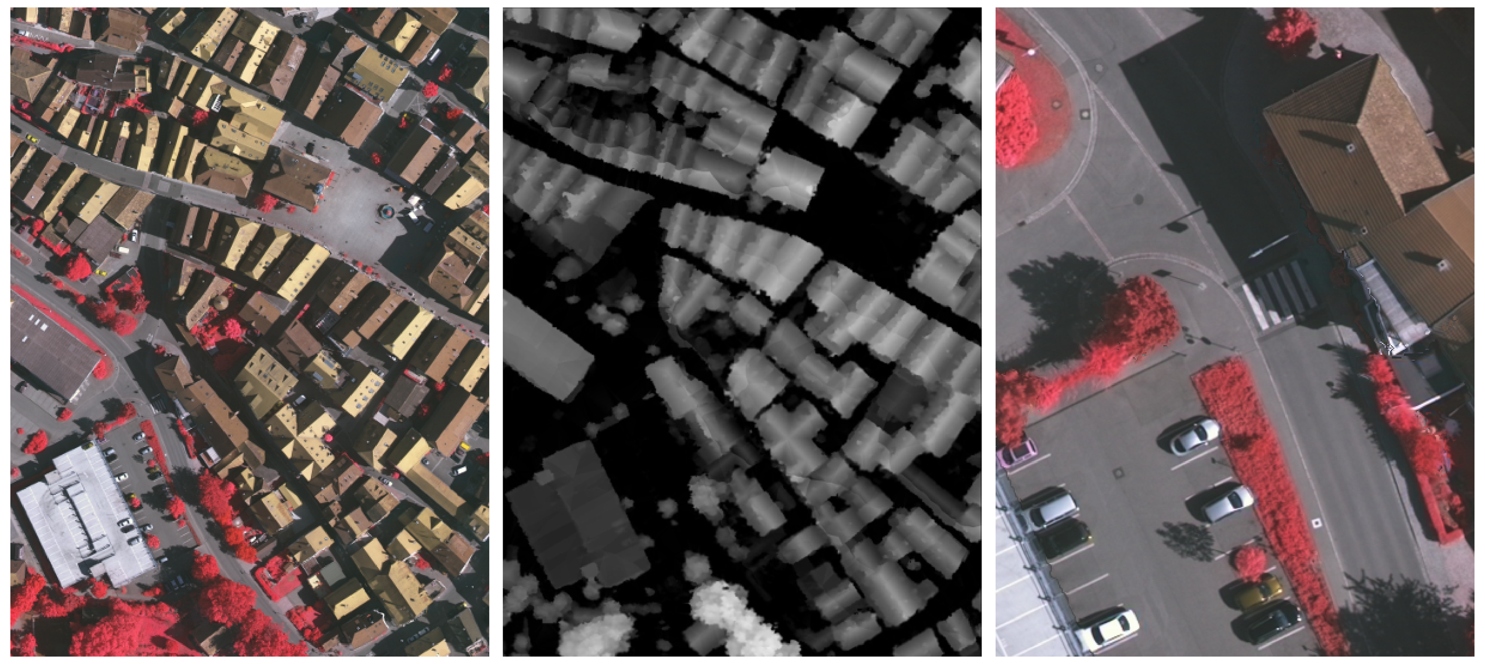
Deep learning for aerial data cartography

Nicolas Audebert[1,2], Bertrand Le Saux[1], Sébastien Lefèvre[2]
ONERA/DTIM[1], IRISA[2]

Context

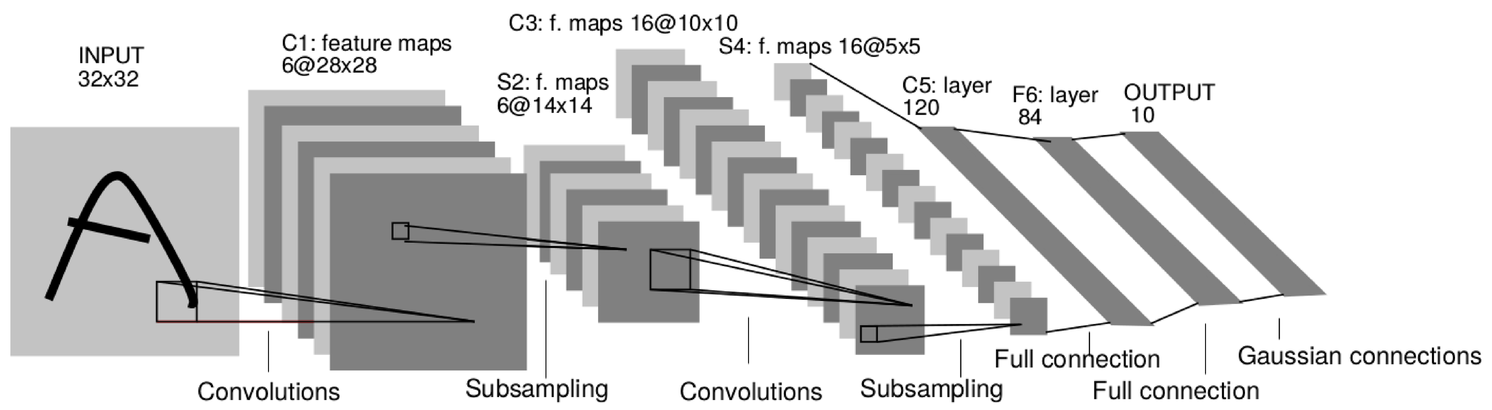
Remote sensing

Aerial and satellite images :
- huge quantities of data
- heterogeneous sensors (RGB, SAR, Lidar, ...)
- multiscale (vehicles, buildings, ...)



Deep learning

Deep convolutional neural networks (CNN) :
- Artificial neural networks : statistical model for machine learning
- CNN (*convolutional neural networks*) = state-of-the-art for computer vision
- Powerful models for classification of RGB data (AlexNet, GoogLeNet, ...)



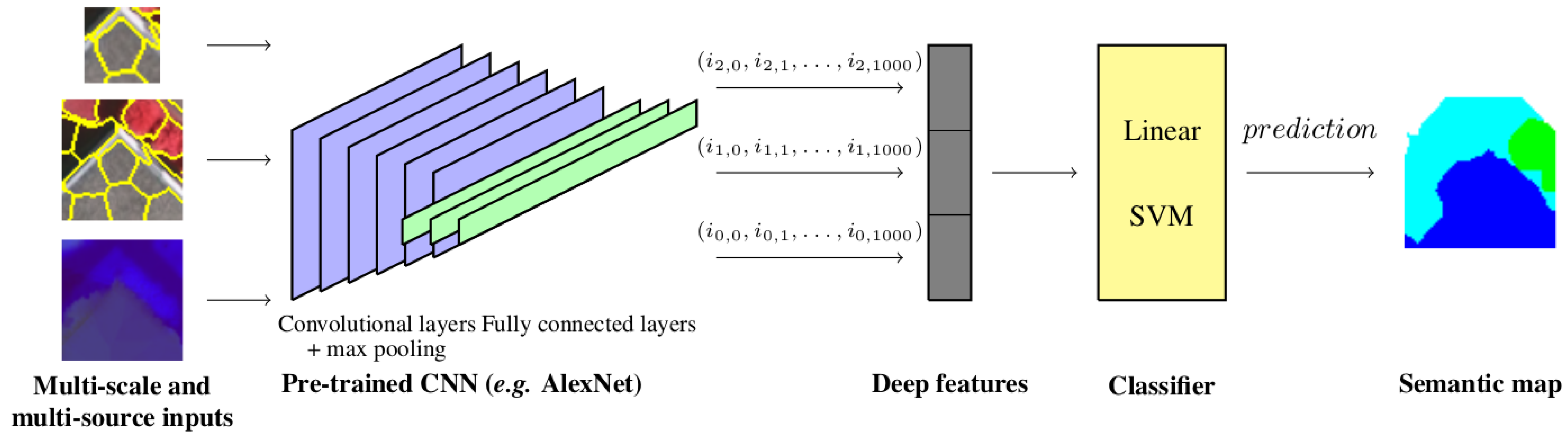
LeNet architecture from "Gradient-Based Learning Applied to Document Recognition", LeCun Y. *et al* (1998)

⇒ **Can we extend deep networks models for Earth Observation ?**

Semantic labeling framework

Classification pipeline

4 steps : (i) Preprocessing - unsupervised **segmentation** (sliding window/superpixels) ⇒ (ii) **feature extraction** using a deep CNN (*deep features*) on each segmented patch ⇒ (iii) linear SVM **classification** (one-vs-all) ⇒ (iv) semantic map **reconstruction**



The CNN is trained for object classification. No fine-tuning is required.
— We extract the truncated output of the CNN (before the softmax classifier)
— Projects into the CNN representation space: $projection : \mathbb{R}^{w*h*c} \mapsto \mathbb{R}^{1000}$
— The **deep features** are an efficient and compact **descriptor of the image**
— We **concatenate several representations** (from different scales and different sensors) into one vector before classification

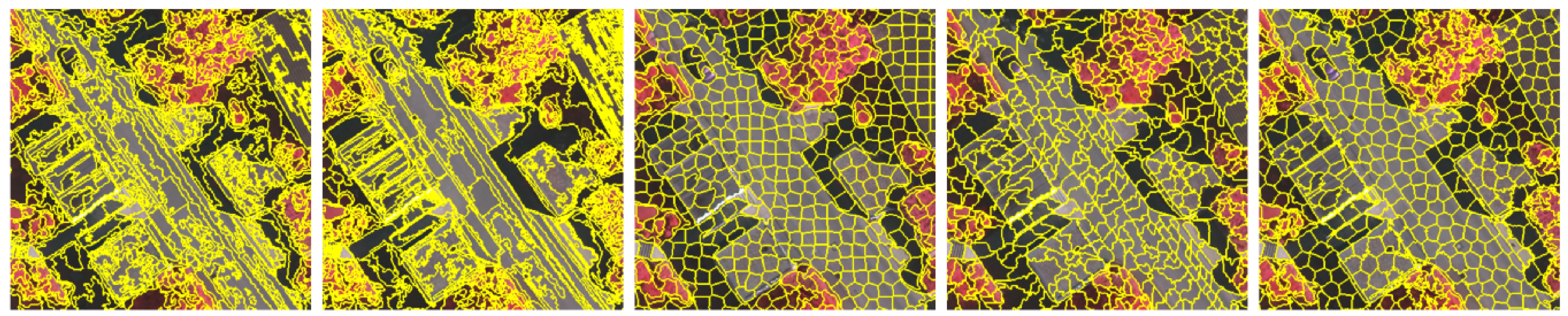
The linear SVM classifier separates the representation space into hyperplanes :
— Defines a classification function : $classify : \mathbb{R}^{1000} \mapsto [0..k]$, k number of classes
— Applied on the deep features : $prediction = classify(projection(input))$
— Requires **supervised training**

Choosing a segmentation algorithm

Classification performance vs. segmentation algorithm

Preprocessing	MRS	HSEG	SLIC[1]	**LSC**[1]	Quickshift[1]	SW[2]
Accuracy (%)	80.53	79.56	82.20	**82.45**	82.05	81.22
F1_car	0.56	0.54	0.54	**0.58**	0.52	0.53

[1] : superpixel based algorithms
[2] : sliding window baseline



(a) MRS (b) HSEG (c) SLIC (d) Quickshift (e) LSC

⇒ **Superpixel** algorithms achieve **better overall accuracy** and are competitive on small object labeling (cars).

References

Romero A. *et al.* (2015). Unsupervised deep feature extraction of hyperspectral images. *TGRS*
Lagrange A. *et al.* (2015). Benchmarking classification of earth-observation data : From learning explicit features to convolutional networks. *IGARSS*
Lefèvre S. *et al.* (2014). Hyperspectral image classification from multiscale description with constrained connectivity and metric learning. *WHISPERS*
Paisitkriangkrai S. *et al.* (2015). Effective semantic pixel labelling with Convolutional Networks and Conditional Random Fields. *CVPR Workshops*
Razavian A. *et al.* (2014). CNN Features off-the-shelf : an Astounding Baseline for Recognition. *CVPR Workshops*

Future work

Integrated end-to-end network for segmentation :
Fully Convolutional Networks (Long *et al.*, 2015) are a promising architecture to perform end-to-end supervised learning on segmentation tasks.
⇒ No need for unsupervised superpixel preprocessing
⇒ Full backpropagation learning

Multisource learning and data fusion :
Vision deep networks assume RGB data : only 3 channels
⇒ Can we extend this to 4 colour channels using infrared information ?
⇒ Can we add a 4th channel with different properties, e.g. the elevation ?
⇒ Can we generalize CNN to deal with multispectral data ? With hyperspectral data ?