European Spatial Data Research

Deep Learning for Remote Sensing – EduServ Course

Sébastien Lefèvre, Loic Landrieu

Mars 2020

Presentation outline

Outline

In these slides, we will get insight on:

- Motivation for Deep Learning on Raster Imagery
- Semantic Segmentation with Fully-convolutional Networks
- Multi-modal Classification
- Object Detection
- Change Detection and Multi-temporal Analysis
- Classification of Hyperspectral Data
- Deep Learning on SAR Data
- Resources

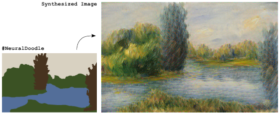
Presentation Layout

Lots of common application in everyone's life:

Self-driving cars

Facebook's facial recognition

Google Translate... images



Painting like Monet!

Figure: Applications of Deep Learning

So... what can we do in remote sensing?

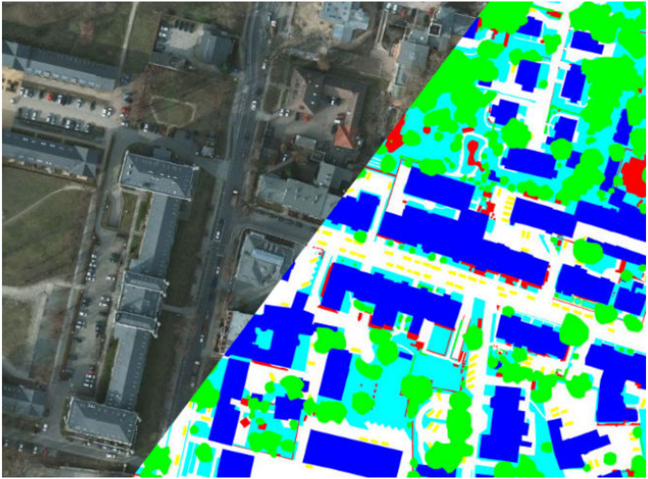


Figure: Can we understand and translate images to cartography?

Motivation for Deep Learning on Raster Imagery

Data Evolution	*Feature Extraction*	*Classification*
resol. −	Pixel-based filtering	Manual modeling, thresholding
resol. +	Textures, local features	Distribution estimates (GMMs, etc.)
data +		Learning-based classifiers (SVMs, ensemble methods: random forests, boosting): high-dimensional, non-linear, complex
resol. ++	Complex features (object modeling)	active learning, latentSVM
data ++	Patchs of pixels, filter banks	Deep neural networks (RBM, RCNN)

resolution, quantity

Figure: An history of classification in remote sensing

- Imagery resolution improves ⇝ **smaller and smaller objects** can be detected (obviously), but image content also gets a **fine-grained description**;
- Reference data increases ⇝ more powerful **machine learning** algorithms can be used;

- In some specific cases, standard machine learning approaches or sensor-based heuristics are well enough...
- but Convolutional Neural Networks make good **generic classifiers**

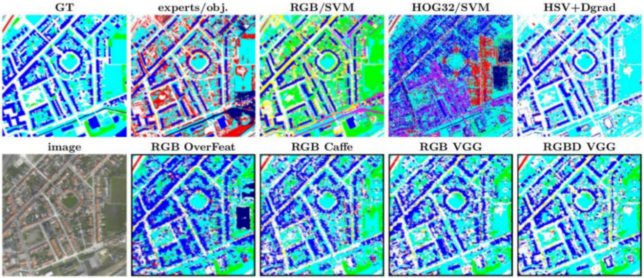


Figure: Benchmarking old-school vs. deep learning methods (Campos-Taberner et al., 2016)

Simple Deep Learning Baseline

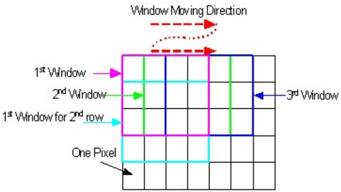
3D	*Algorithm*	Imp. surf.	Build.	Low veg.	Tree	Car	Clutter	Boat	Water	Overall acc. %	Cohen κ
★	Expert	58.97	63.87	74.55					92.39	$\varnothing$	$\varnothing$
	RGB/SVM	53.89	53.53	50.32	32.97	24.02	13.75	12.12	98.52	60.77	0.52
★	RGB*d*/SVM	14.51	67.79	38.03	27.43	7.15	1.12	14.58	98.45	50.76	0.41
★	RGB*d*I/SVM	60.86	69.01	57.12	38.12	11.59	20.49	15.04	94.42	63.83	0.56
	HOG32/SVM	28.94	43.17	48.77	27.32	30.24	17.39	12.61	88.02	52.45	0.41
	HOG16/SVM	39.52	38.45	35.65	29.99	21.93	16.13	13.52	80.02	49.4	0.36
	HSV/SVM	71.60	46.97	68.38	0.12	0.00	13.71	0.00	92.14	70.16	0.60
★	HSVDGr/SVM	73.30	70.85	68.75	0.17	0.00	17.11	0.00	92.37	73.60	0.65
	SOM							51.45		$\varnothing$	$\varnothing$
	DtMM					48.46				$\varnothing$	$\varnothing$
	RGB OverFeat/SVM	55.86	63.34	59.48	64.44	36.03	28.31	41.51	92.07	67.97	0.59
	RGB Caffe/SVM	62.32	62.66	63.23	60.84	31.34	32.49	46.57	95.61	71.06	0.63
	RGB VGG/SVM	63.18	64.66	63.60	66.98	31.46	43.68	51.92	95.93	72.36	0.64
★	RGB*d* VGG/SVM	66.02	74.26	65.04	66.94	32.04	44.96	50.61	96.31	74.77	0.67
★	RGBd^+ VGG/SVM	67.66	72.70	68.38	78.77	33.92	45.6	56.10	96.50	76.56	0.70
★	RGBd^+ trained AlexNet	79.10	75.60	78.00	79.50	50.80	63.40	44.80	98.20	83.32	0.78

Figure: Benchmarking old-school vs. deep learning methods (Campos-Taberner et al., 2016)

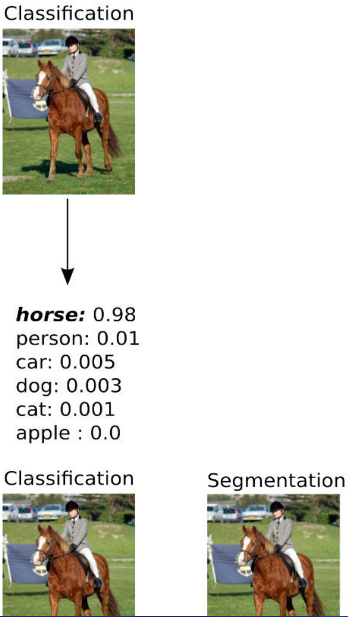
Presentation Layout

1. Deep Learning on Raster Imagery
   - Motivation for Deep Learning on Raster Imagery
   - Semantic Segmentation with Fully-Convolutional Networks
   - Multi-modal Semantic Segmentation
   - Object Detection
   - Change Detection
   - Classification of Hyperspectral Data
   - Deep learning on SAR Data

Remote Sensing images are pretty large!

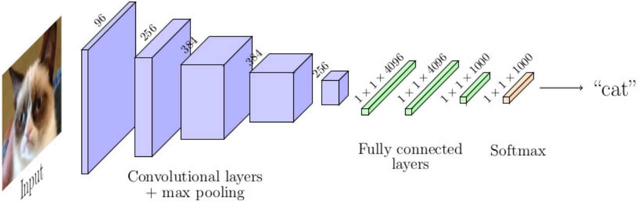


- Extract patches over the image using a sliding window with overlap
- Train on images with ground-truth
- Apply on test patches (by averaging all possible predictions per pixel)



From CNNs to Fully-Convolutional Networks

AlexNet CNN for classification (Krizhevsky et al., 2012)



Fully-convolutional AlexNet for semantic segmentation (Long et al., 2015)

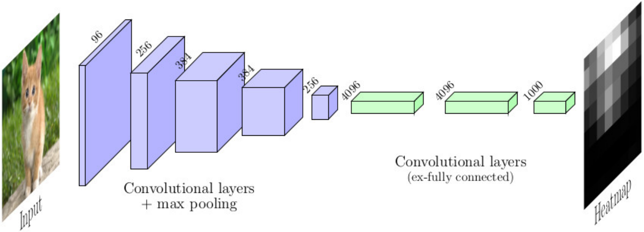
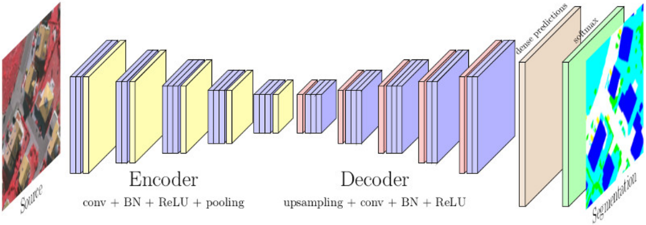
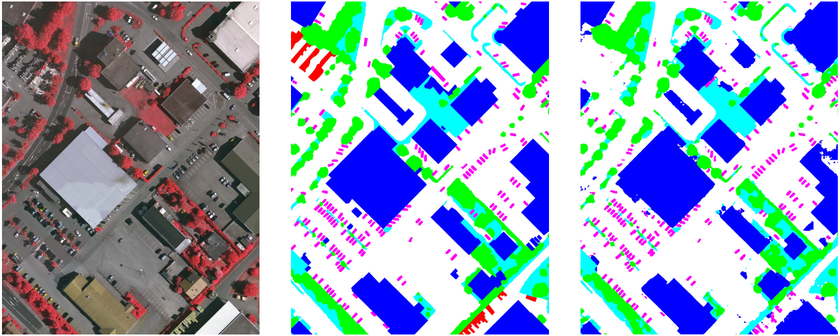


Figure: From CNNs which include fully-connected layers to Fully-Convolutional Networks (FCNs)

Fully-Convolutional Networks



- SegNet: A deep convolutional Encoder-Decoder architecture for Image Segmentation (Bandrinarayanan et al., 2016)
- And today: U-Net (Ronneberger et al., 2015), Hourglass (Newell et al. 2016))



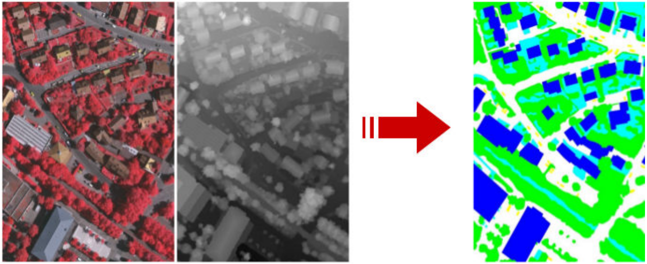
IR/R/G 10cm/pixel Ground-truth SegNet

Figure: SegNet semantic segmentation (Audebert et al., 2017a)

Presentation Layout

How to automatically generate maps from aerial imagery?

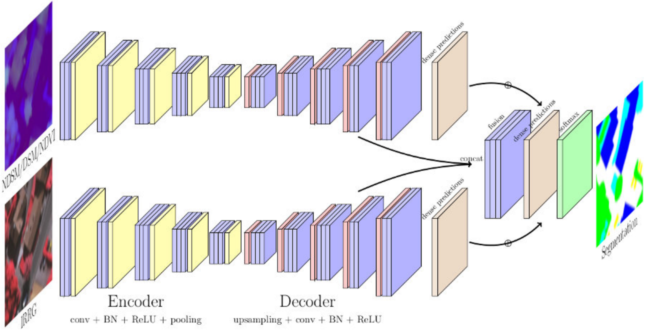
- Non-standard (yet complimentary) imagery: multispectral, LiDAR...
- Auxiliary data: existing open-source maps (OpenStreetMap, etc.)



Multi-modal Semantic Segmentation

How to automatically generate maps from aerial imagery?

- Dual-stream networks which translate data to representations (coding)...
- Fusion and decoding from representations to maps



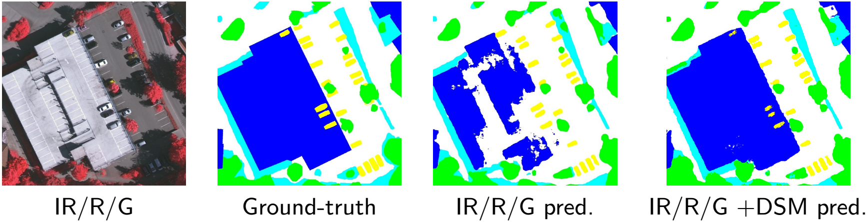


Figure: Multimodal semantic segmentation results (Audebert et al., 2017a)

- Sensor-based information helps!

Multi-modal Semantic Segmentation

How to automatically generate maps from aerial imagery using auxiliary maps?

- Incorporate information to guide the process, FuseNet: (Hazirbas et al., 2016)
- Faster training and better results!

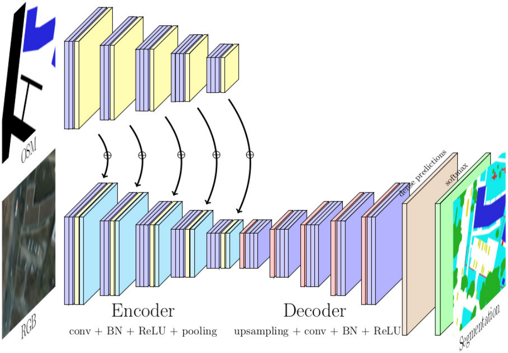


Figure: Joint learning of RGB imagery and OSM (Audebert et al., 2017b)

Multi-modal Semantic Segmentation

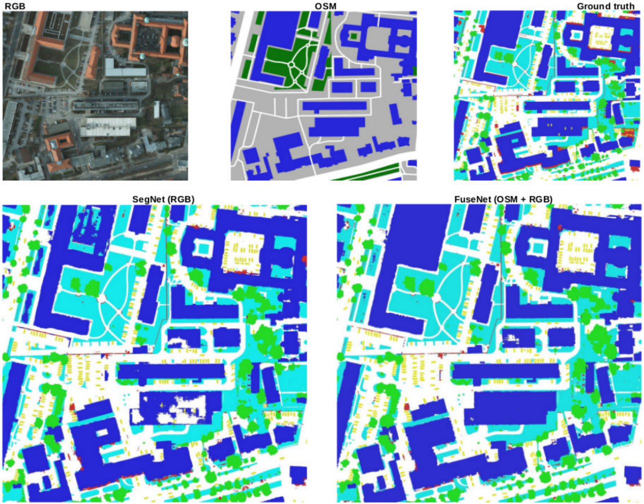
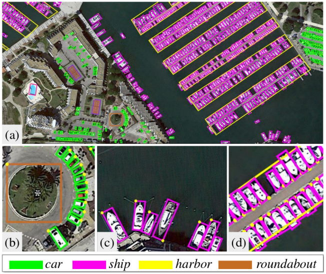


Figure: RGB+OSM semantic segmentation results (Audebert et al., 2017b)

Presentation Layout



(a)

(b)

Figure: Object detection with bounding boxes (a) or by instance mask (b)

- xView: Objects in Context in Overhead Imagery (Lam et al. 2018) ⇝ `http://xviewdataset.org/`
- DOTA$^{(a)}$: Dataset for Object deTection in Aerial images (Xia et al., 2019) ⇝ `https://github.com/jessemelpolio/Faster_RCNN_for_DOTA`
- Busy Parking Lot UAV Video dataset$^{(b)}$: vehicle instance segmentation (Mou & Zhou 2018) ⇝ `https://www.sipeo.bgu.tum.de/downloads`

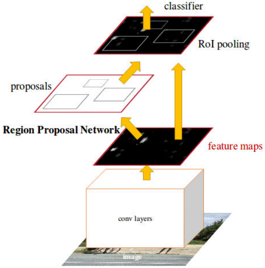


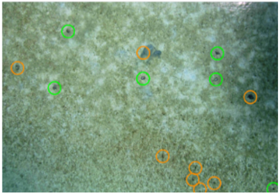
Figure: Faster RCNN (Re et al, 2015)

- Object detection principle combines region proposal by a Region Proposal Network, and region-of-interest classification
- Instance segmentation adds a segmentation branch to output the mask
- R-CNN, Mask-RCNN architectures and more: `https://github.com/facebookresearch/Detectron`

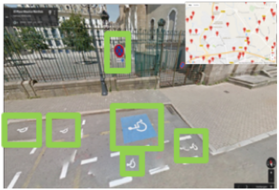
Object Detection

Any generic detection network (e.g. Faster-RCNN) can be fine-tuned to fit user's needs, e.g.:

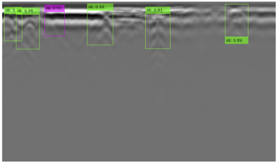
- mapping and monitoring marine turtles from UAVs in environmental surveys
- detecting accessibility signs to map related parking lots
- detect and map buried networks from geophysical data (ground-penetrating radar)



Turtle detection
WIPSEA



Accessibility mapping
(Nassar & Lefèvre, 2019)



Buried network detection
(Pham & Lefèvre, 2018)

Presentation Layout

1. Deep Learning on Raster Imagery
   - Motivation for Deep Learning on Raster Imagery
   - Semantic Segmentation with Fully-Convolutional Networks
   - Multi-modal Semantic Segmentation
   - Object Detection
   - **Change Detection**
   - Classification of Hyperspectral Data
   - Deep learning on SAR Data

How to extend semantic analysis to multi-temporal data ?

- detect changes
- monitor activity in high-revisit rate acquisitions
- focus on specific changes (urban, agriculture, vehicles, industrial activity...)

Date 1

Date 2

Change map

How to extend semantic analysis to multitemporal data ?

- As before, simply concatenating images...
- Or with siamese networks, i.e. dual stream nets which share weights.

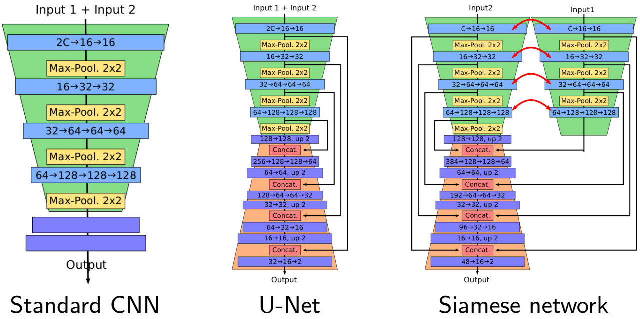


Figure: CNN / FCN architectures for change detection. (Daudt et al., 2018)

Change detection

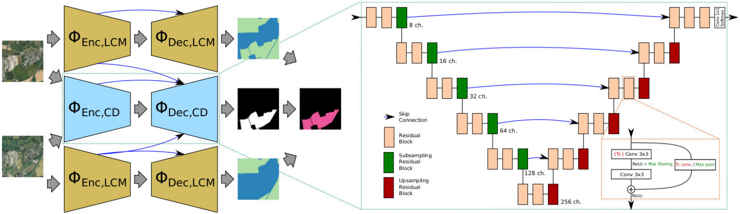
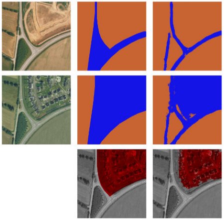


Figure: Semantic change detection. (Daudt et al., 2019)

Semantic change detection:

- Joint multi-task learning of semantics and differences with FCNs
- Prediction of land cover and change maps



Siamese networks can work with very different inputs! (e.g. ground vs aerial imagery)

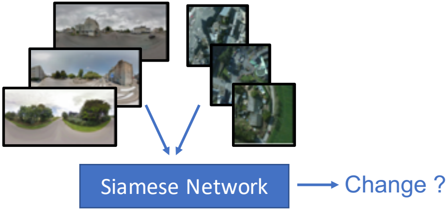


Figure: Multiview change detection (Lefèvre et al. 2017)

Recurrent Neural Networks

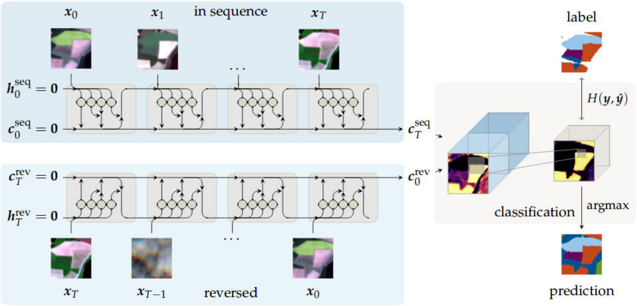


Figure: Multi-Temporal Land Cover Classification with Recurrent Auto-encoders (Russwurm & Köner, 2018)

Presentation Layout

How to extend semantic analysis to hyperspectral imaging (HSI) data ?

- RGB to 100+ bands, image to data cube;
- finer spectral description, out-of-visible;
- lower resolution but finer class discrimination (materials, stressed or healthy vegetation...)

Figure: Hyperspectral data cube: Houston (Texas, USA) – IEEE GRSS IADF TC's Data Fusion Contest 2018

CNN architectures adapted to HSI classification:

- 1D CNN: spectrum classification
- 1D RNN: spectral sequence classification

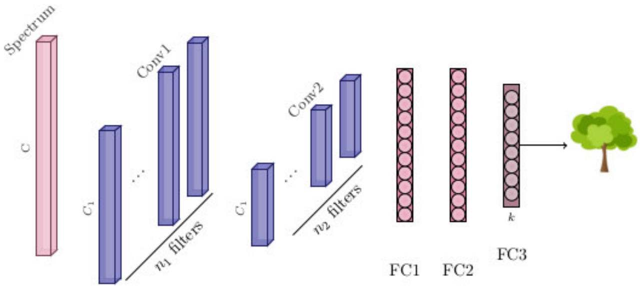


Figure: CNN 1D

CNN architectures adapted to HSI classification:

- Spatial-spectral, 2D+1D approaches
- Reduce to RGB-like data + 2D CNN
- PCA or supervised reduction : alternate 2D and 1D convolutions

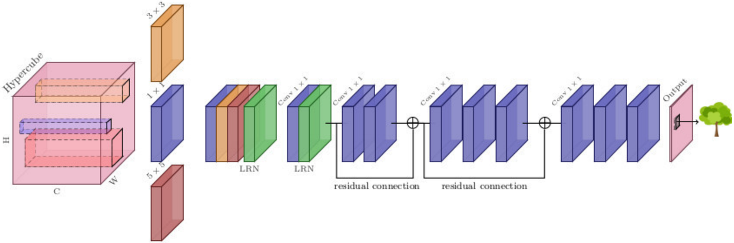


Figure: CNN 2D+1D (Lee et al., 2016)

CNN architectures adapted to HSI classification:

- End-to-end 3D pattern recognition: apply learnable (w, h, B) filters on the hypercube

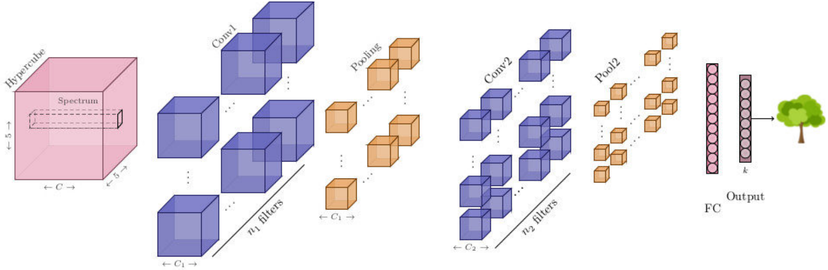


Figure: CNN 3D (Audebert et al, 2019)

HSI classification with CNNs:

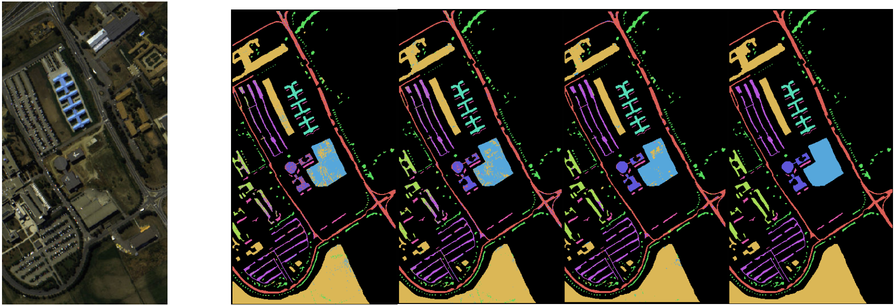


Figure: Comparison of HSI classifications with various CNNs (Audebert et al, 2019)

Presentation Layout

Deep learning on SAR Data

How to extend deep learning processing to SAR data ?

- Specific physics, different from optical images: intensity + phase;
- 🙂 High resolution, "cloud-free" images;
- 🙁 Presence of "speckle" and changing appearance depending on the angle of view.

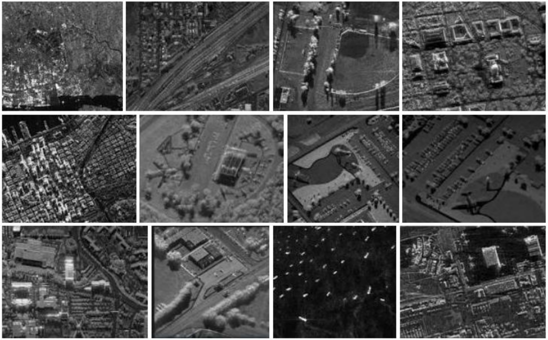


Figure: SAR image examples

Despeckling:

- Inspired by denoising Denoising Auto-Encoders (Vincent et al., 2010);
- Auto-encoder learn to reconstruct the image from itself
- Denoising / despeckling autoencoders learn to reconstruct the image from the image with added speckle

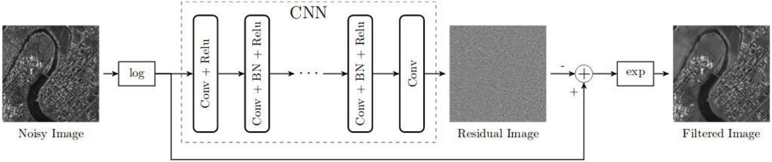


Figure: Despeckling auto-encoder (Chierchia et al., 2017)

- Usually straightforward (but do they miss something?);
- Complex valued CNN for processing intensity + phase images (Haensch & Hellwich, 2010) (Zhang et al., 2010)

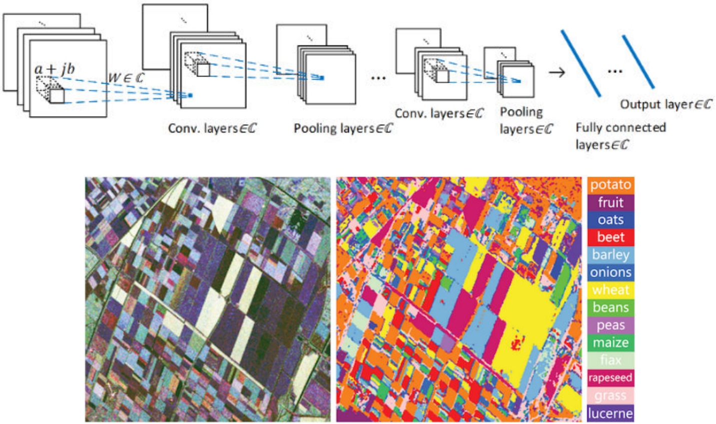
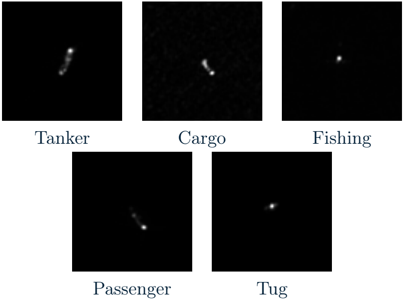


Figure: SAR classification (Zhang et al., 2017) and (Zhu et al., 2017)

Object characterization for SAR Data

Similarly to optical imagery, SAR imagery can be exploited in many ways, e.g.

- ship analysis from Sentinel-1: detection, length estimation, classification



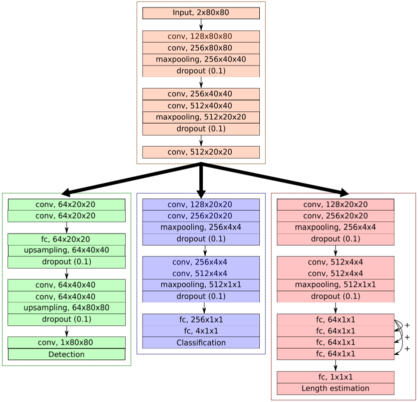


Figure: Dechesne et al., 2019

Public resources

Good reads:

- **Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art**, *Zhang, Zhang and Du*, IEEE Geosci. and Rem. Sens. Mag, 6 (2) June 2016
- **Deep learning in remote sensing: A comprehensive review and list of resources**, *Zhu, Tuia, Mou, Xia, Zhang, Xu, and Fraundorfer*, IEEE Geosci. and Rem. Sens. Mag, 5 (4) Dec. 2017
- **Deep Learning for Classification of Hyperspectral Data: A Comparative Review**, *Audebert, Le Saux and Lefèvre*, IEEE Geosci. and Rem. Sens. Mag., 7 (2) June 2019

Public resources

Toolbox:

- DeepNetsForEO (`https://github.com/nshaud/DeepNetsForEO`): python code for semantic segmentation of aerial / satellite imagery
- DeepHyperX (`https://github.com/nshaud/DeepHyperX`): python toolbox for classification of hyperspectral imagery (spectral, spatial-spectral and 3D convolutions)

Public resources

Public datasets:

- ISPRS datasets: semantic labeling, reconstruction $\rightsquigarrow$ `https://www.isprs.org/data/`
- IEEE GRSS Data Fusion Contests: `http://www.grss-ieee.org/community/technical-committees/data-fusion/data-fusion-contest/`
- IEEE GRSS: hyperspectral datasets with standard train/test splits (DFC2018, Pavia, Indian Pines) $\rightsquigarrow$ `http://dase.grss-ieee.org/`
- INRIA Aerial Semantic labeling dataset: buildings $\rightsquigarrow$ `https://project.inria.fr/aerialimagelabeling/`
- XView: objects in aerial images $\rightsquigarrow$ `http://xviewdataset.org/`
- DOTA: Detecting Objects in Aerial images $\rightsquigarrow$ `https://captain-whu.github.io/DOTA/dataset.html`