A Gentle introduction to Remote Sensing Analysis via Deep Learning techniques

Dino Ienco
Research Scientists,
IRSTEA – UMR TETIS - Montpellier,
dino.ienco@irstea.fr

Outline

Introduction to Remote Sensing, Opportunities & Challenges

Data Fusion & Remote Sensing

Deep Learning (DL) for Remote Sensing

Conclusions

Introduction

Nowadays, many earth observation satellite missions exist:

- Sentinel [Senti]
- LandSat-8 [LandSat]
- SPOT 6/7[Spot]
- …

Acquired images have different:

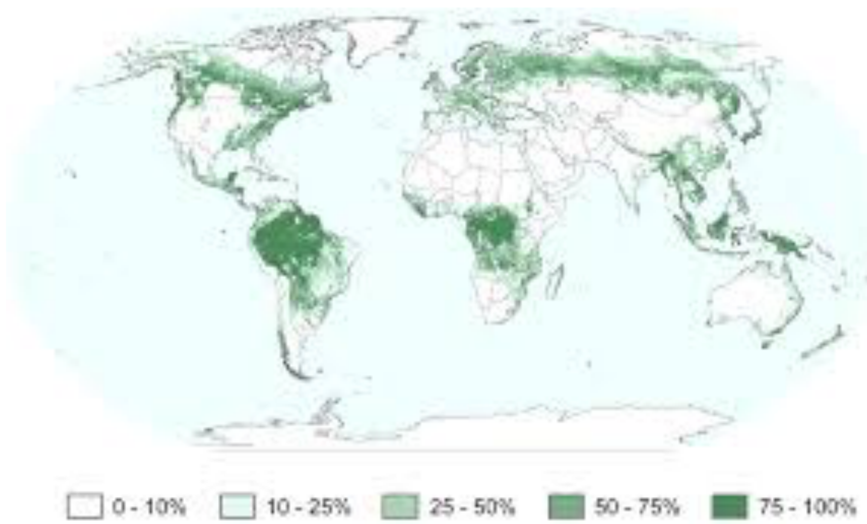
- spatial resolution (0.5 – 30 meters)
- radiometric content (spectral bands)
- temporal resolution (every 5 – 365 days)

HUGE quantity of Satellite Images Describing Earth Phenomena at different scales

Why EOD is an Opportunity

Earth Observation Data can have practical influence on different domains:

Continental Surface analysis



Precise Agriculture

Climate Changes Analysis

Biodiversity Monitoring

Why EOD is an Opportunity

Analyze, Mining and Exploit EOD data can also improve practices on:

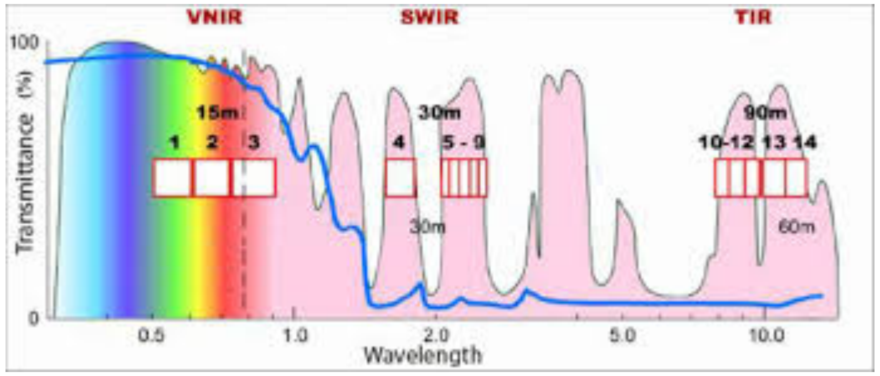
- Forestry characterization
- Lithological classification and mineral mapping
- Food Risk prevention
- Environmental monitoring
- Urban development
- Wildlife and Habitat Monitoring

This is why Satellite imagery analytics is becoming more IMPORTANT

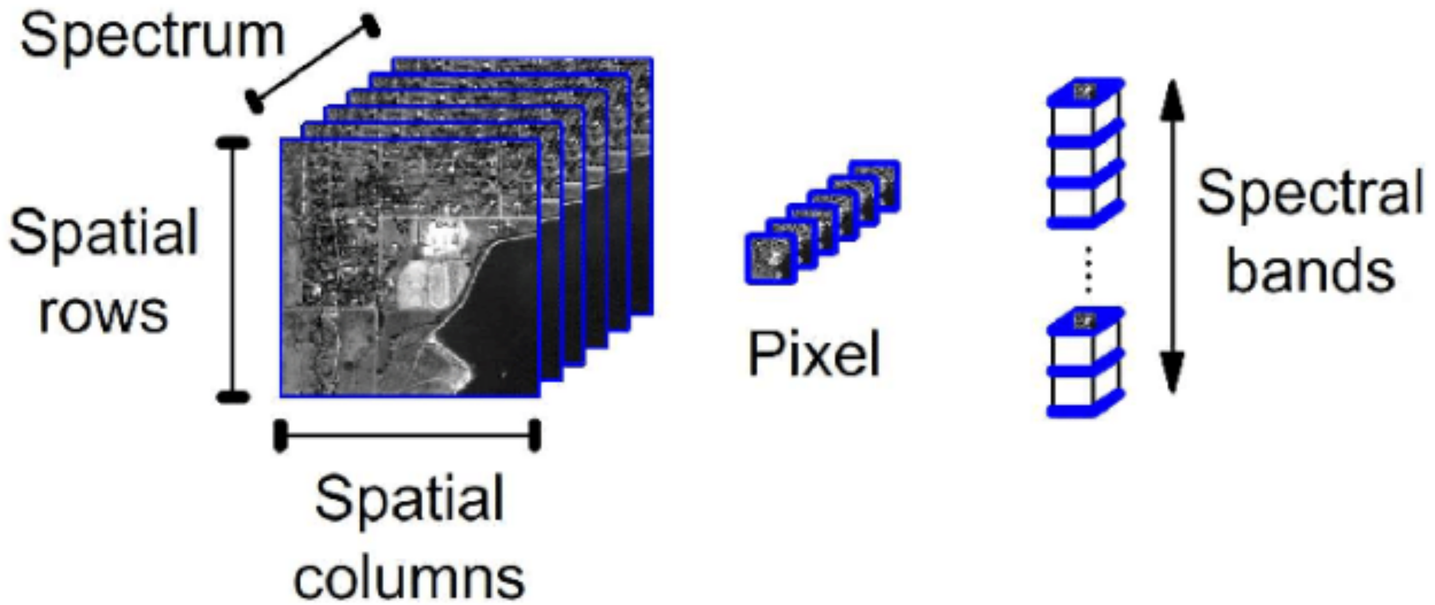
Introduction to RS

A Satellite Image:

A data cube that describes a spatial area by means of several spectral bands



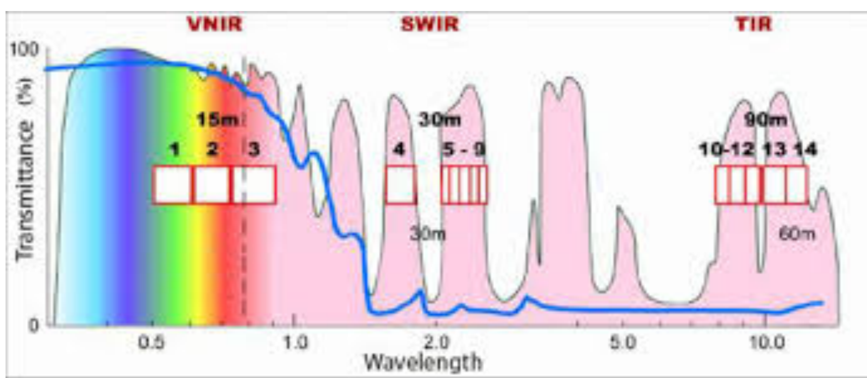
Spectral Bands



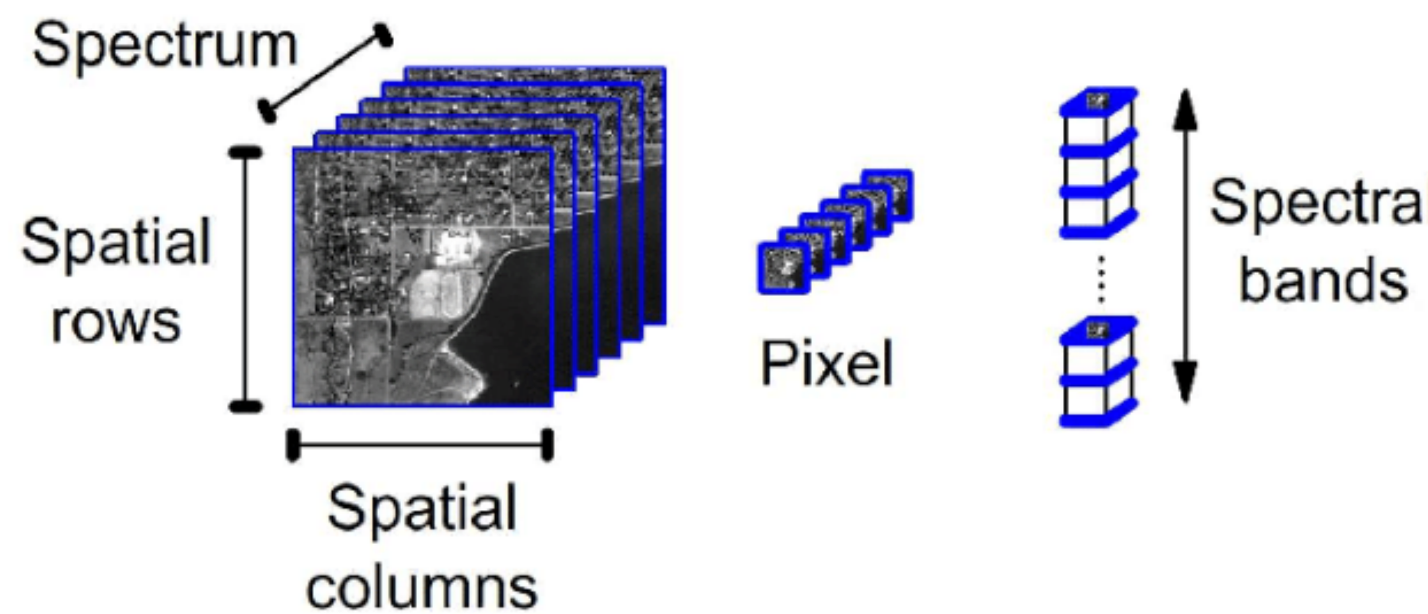
Introduction to RS

A Satellite Image:

A data cube that describes a spatial area by means of several spectral bands



Spectral Bands



Type of information:

- Optical Images (Multi-Spectral / Hyperspectral)
- Radar Images (phase, amplitude, etc…)
- LIDAR (point clouds)
- Etc…

Introduction to RS

EOD allows also to collect Very High Resolution Images (VHR) i.e. Spot6/7 (at 1.5m), Pléiades (.5m), WorldView3 (.3m) at Low Temporal Frequency (once or twice per

VHSR data are useful to obtain fine resolution information to characterise spatial pattern and spatial texture

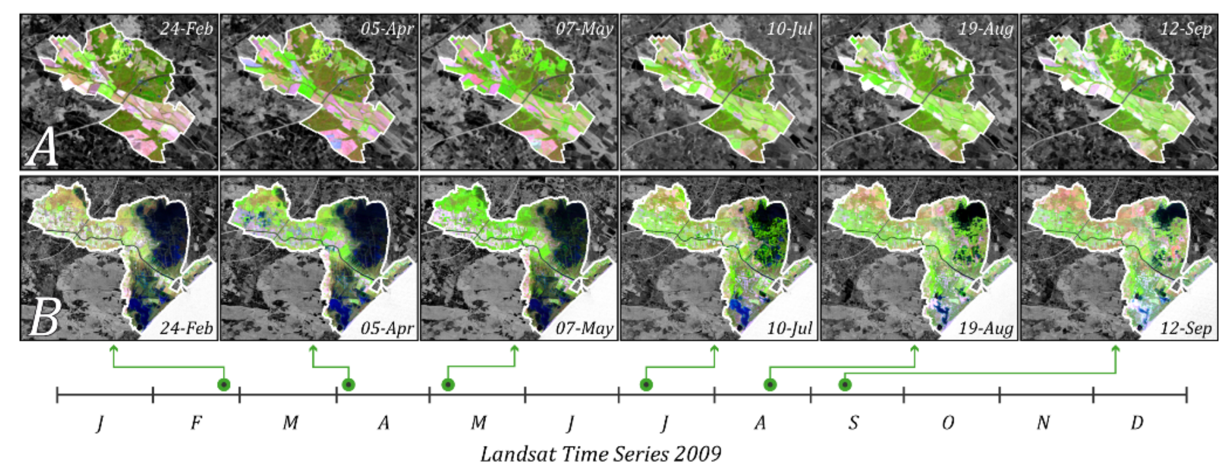
Introduction to RS

EOD allows also to collect Very High Resolution Images (VHR) i.e. Spot6/7 (at 1.5m), Pléiades (.5m), WorldView3 (.3m) at Low Temporal Frequency (once or twice per

VHSR data are useful to obtain fine resolution information to characterise spatial pattern and spatial texture

EOD allows to collect Satellite Image Time Series (SITS) at High Spatial Resolution (Sentinel ~10m) and High Temporal Frequency (every 5/10 days)

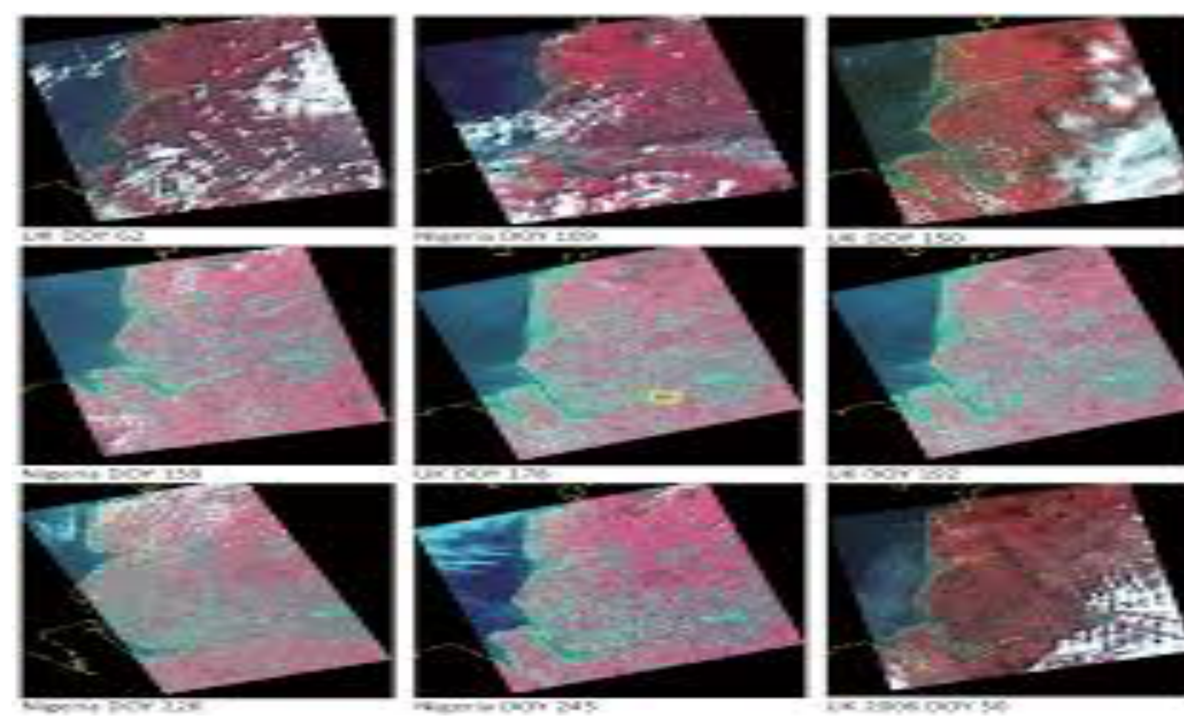
The same geographical area is observed



SITS data are useful to analyze spatio-temporal phenomena (trends and changes) over the time

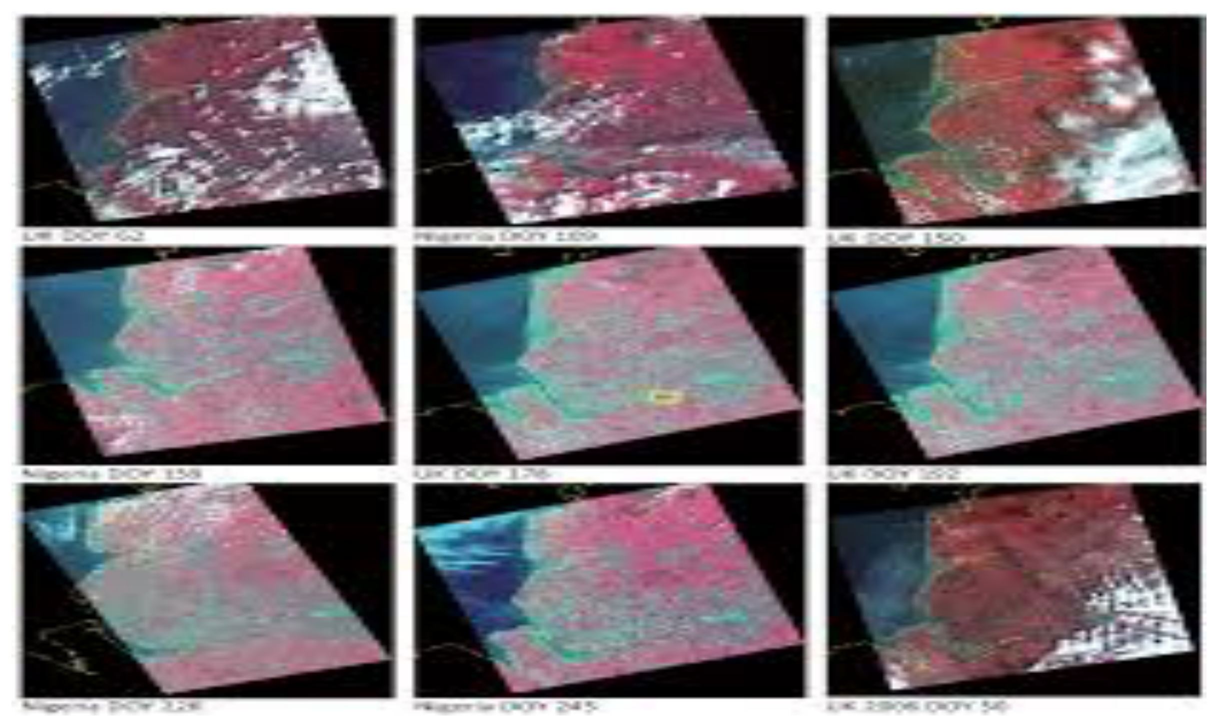
Challenges in EOD

Satellite imagery analytic is
challenging due to EOD diversity



Challenges in EOD

Satellite imagery analytic is challenging due to EOD diversity

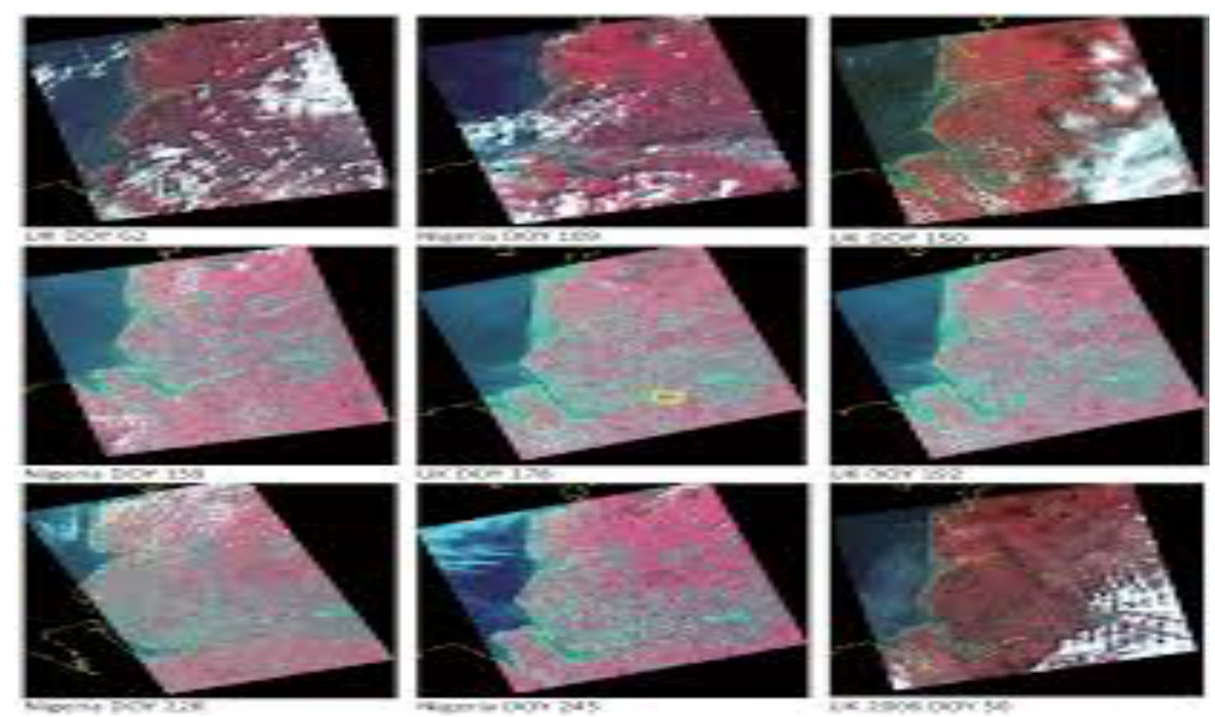


Heterogeneity:

- Spatial (Different resolutions)
- Temporal (time steps not always constant)
- Acquisition Sensor (Optical Images, Radar Images, DEM, LiDAR, etc..)
- Very High Spatial vs High Temporal Resolution (VHR & SITS)

Challenges in EOD

Satellite imagery analytic is challenging due to EOD diversity



Heterogeneity:

- Spatial (Different resolutions)
- Temporal (time steps not always constant)
- Acquisition Sensor (Optical Images, Radar Images, DEM, LiDAR, etc..)
- Very High Spatial vs High Temporal Resolution (VHR & SITS)

…and also Data Quality (from pre-processing to information extraction)

…and Ground Truth (or annotation) to build predictive models.

Heterogeneity & Data Fusion

Due to the huge amount of different sensors, today available,
Data Fusion is a very **important and hot topic** in Remote Sensing Community

[Schmitt16] M. Schmitt and X. X. Zhu, "Data Fusion and Remote Sensing: An ever-growing relationship". IEEE Geoscience and Remote Sensing Magazine 4(4): 6--23, 2016

Heterogeneity & Data Fusion

Due to the huge amount of different sensors, today available,
Data Fusion is a very **important and hot topic** in Remote Sensing Community

Data Fusion process (for Remote Sensing Data) [Schmitt16] :

- **MATCHING AND COREGISTRATION** (i.e. align together sources via coordinate transformations and unit adjustments)

- **FUSION BY ESTIMATION** (the step in which data are really fused together):
  - Combine multiple images covering the same area to **reduce uncertainty**
  - Combine together multiple images with **complementary** spatio/spectral **information**
  - Combine images with **shared information**: i.e. combine multiple VHR images for 3-D reconstruction

[Schmitt16] M. Schmitt and X. X. Zhu, "Data Fusion and Remote Sensing: An ever-growing relationship". IEEE Geoscience and Remote Sensing Magazine 4(4): 6--23, 2016

Heterogeneity & Data Fusion

Fusion can happen at different levels

[Schmitt16]

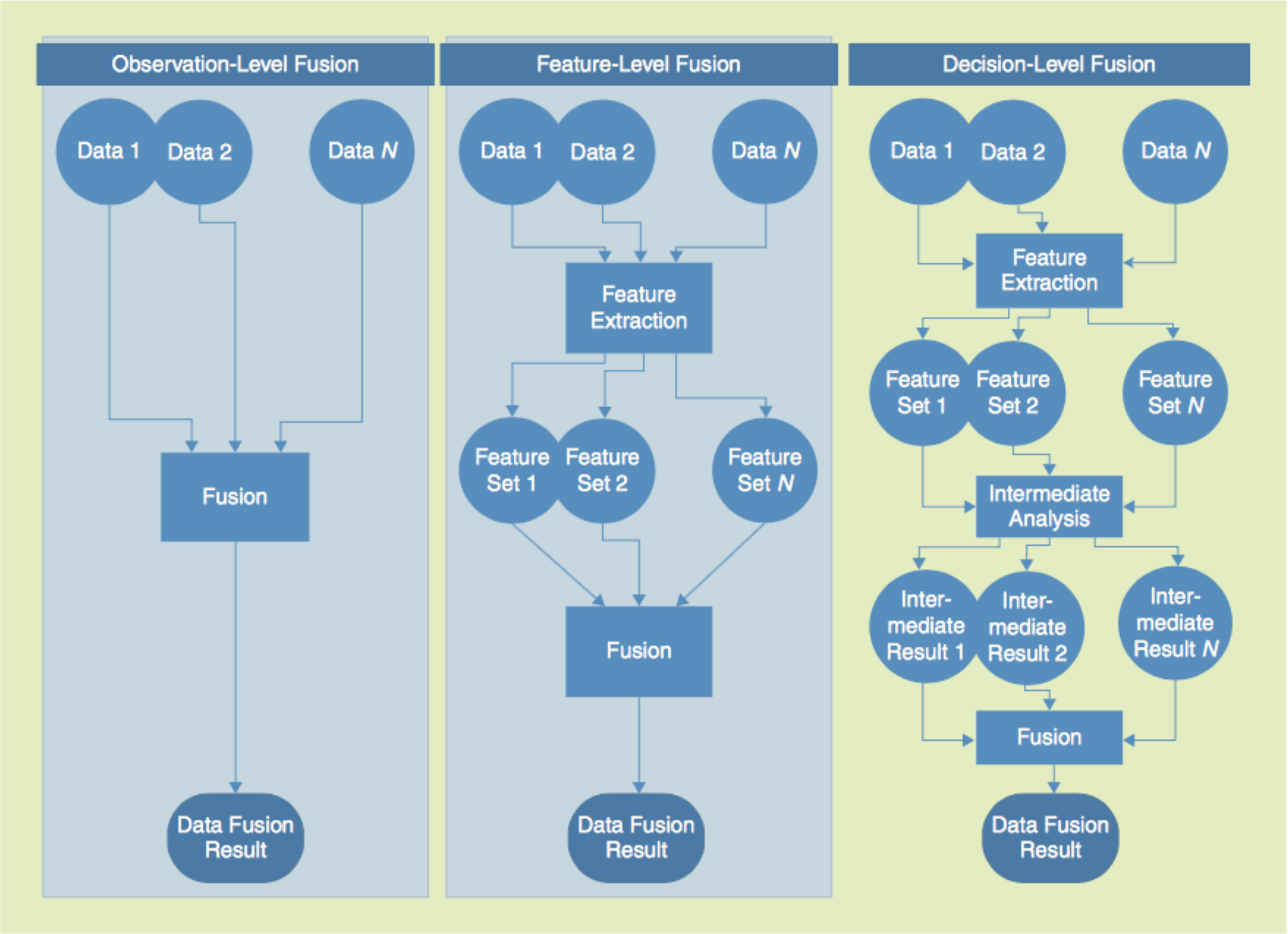


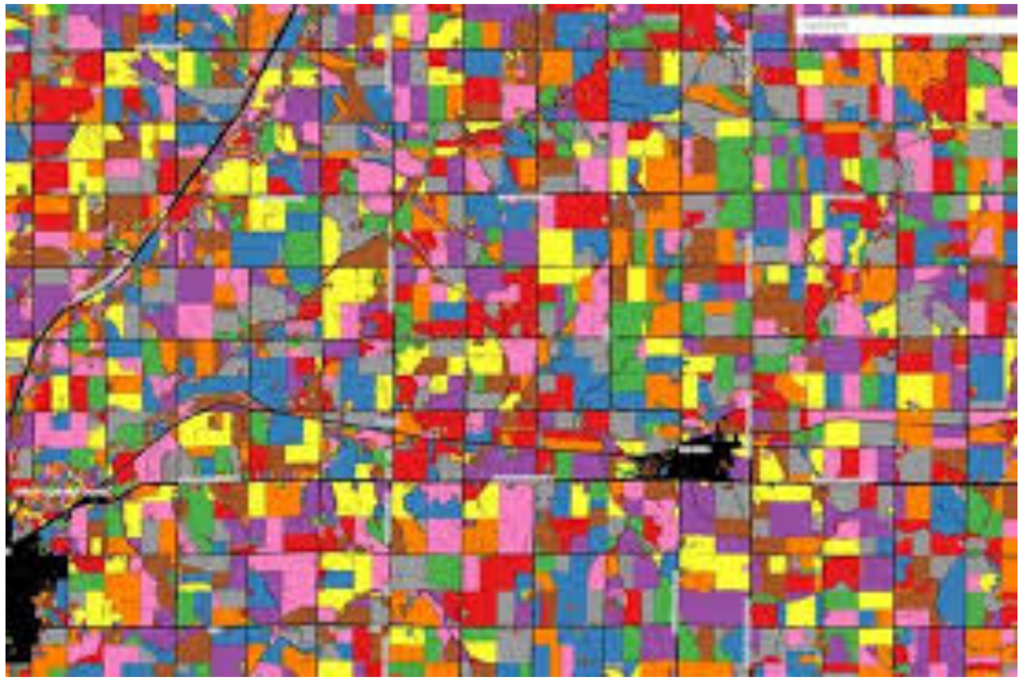
FIGURE 3. The three types of data fusion are compared side by side: observation level, feature level, and decision level.

[Schmitt16] M. Schmitt and X. X. Zhu, "Data Fusion and Remote Sensing: An ever-growing relationship". IEEE Geoscience and Remote Sensing Magazine 4(4): 6--23, 2016

Current Trends in RS

Nowadays, Machine Learning techniques are a standard tool in Remote Sensing analytics [Holloway18]:

- Deal with huge amount of data
- Automatically build predictive methods
- Group together similar areas
- Detect Objects of Interest



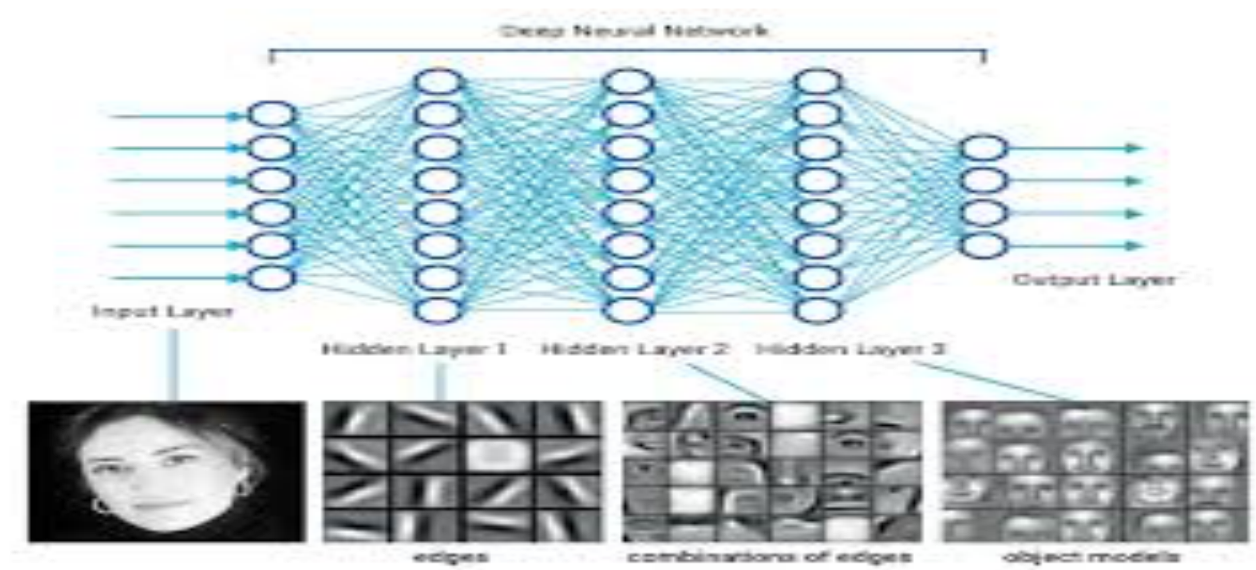
[Holloway18] J. Holloway, K. Mengersen: Statistical Machine Learning Methods and Remote Sensing for Sustainable Development Goals: A Review. Remote Sensing 10(9): 1365 (2018)

[LeCun15] Y. LeCun, Y. Bengio and G. Hinton. “Deep Learning” In Nature 52(8): 436-444 (2015).

Current Trends in RS

Nowadays, Machine Learning techniques are a standard tool in Remote Sensing analytics [Holloway18]:
- Deal with huge amount of data
- Automatically build predictive methods
- Group together similar areas
- Detect Objects of Interest

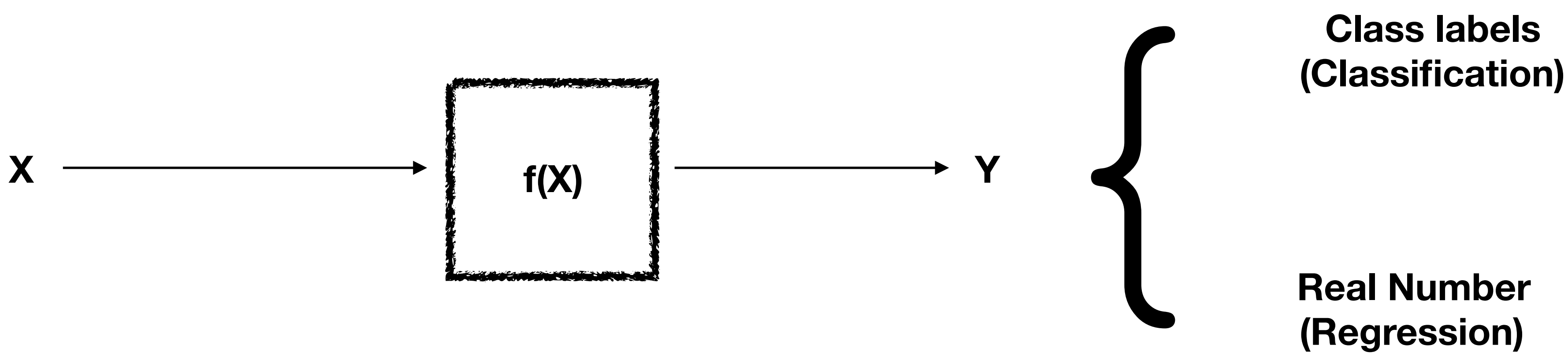
Recent Trends 'Deep Learning Methods' [LeCun15] :
- Inspired by human brain
- Layers architecture
- Applications in different domains:
  + Speech Recognition
  + Image Recognition
  + Natural Language Processing



[Holloway18] J. Holloway, K. Mengersen: Statistical Machine Learning Methods and Remote Sensing for Sustainable Development Goals: A Review. Remote Sensing 10(9): 1365 (2018)

[LeCun15] Y. LeCun, Y. Bengio and G. Hinton. "Deep Learning" In Nature 52(8): 436-444 (2015).

Machine Learning



Machine Learning

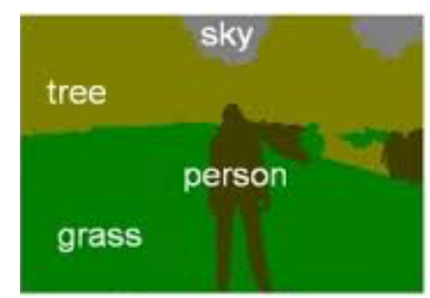
X → f(X) → Y

Y:
- Class labels (Classification)
- Real Number (Regression)

Object Recognition →

{Dog, Cat, Sheep, Bear, Lion, ...}

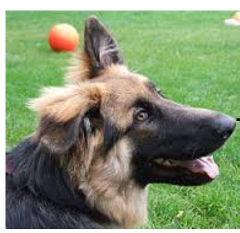
Semantic Segmentation →



Sentiment Classification →

Deep Learning
Learning representation

Traditional Machine Learning systems leverage **feature engineering** to represent the data:
- Text Analysis: Bag of Words
- Image Analysis: Hog (Histogram of Oriented gradient), SIFT (Scale Invariant Feature Transform)



→ **Hand-Crafted Features** → **Simple Trainable Classifier (SVM, RF, NB, …)**

Deep Learning
Learning representation

Traditional Machine Learning systems leverage **feature engineering** to represent the data:
- Text Analysis: Bag of Words
- Image Analysis: Hog (Histogram of Oriented gradient), SIFT (Scale Invariant Feature Transform)

Hand-Crafted Features → **Simple Trainable Classifier (SVM, RF, NB, …)**

Deep Learning approaches **learn internal representations (new features)** without necessity of hand-crafted features

Trainable Feature Extractor → **Trainable Classifier**

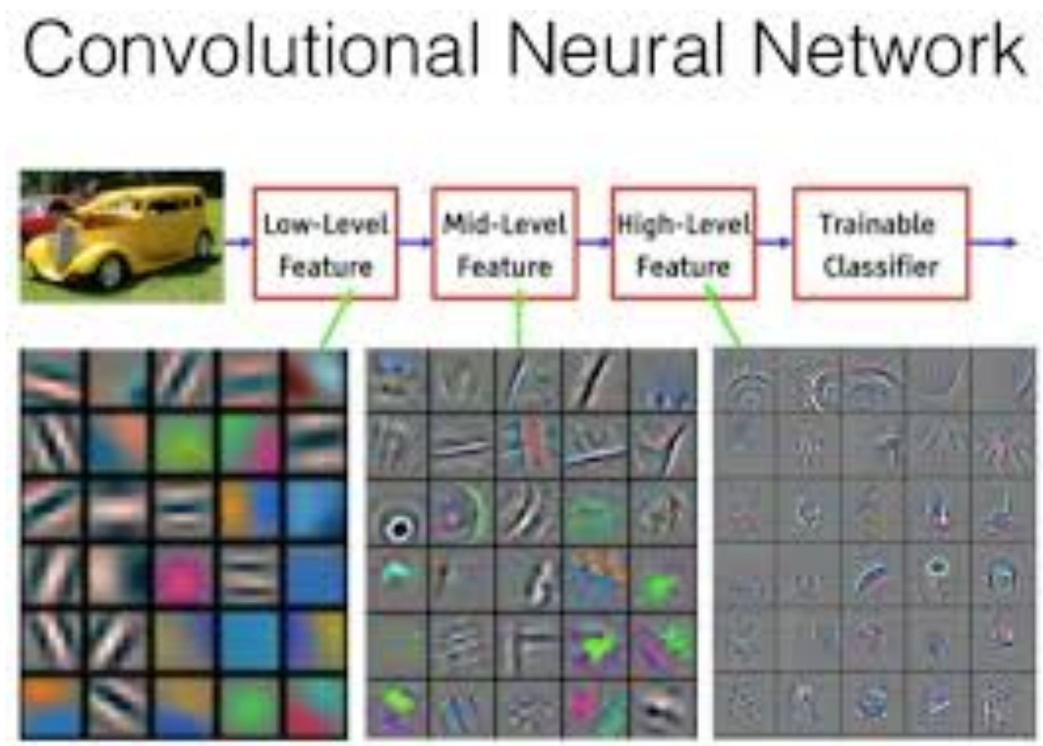
Deep Learning Model

Deep Learning
Learning representation

Deep Learning allows to:

Learn different level of features from low-level to high-level in a kind of hierarchical organisation

Can share the low-level representation for many different tasks

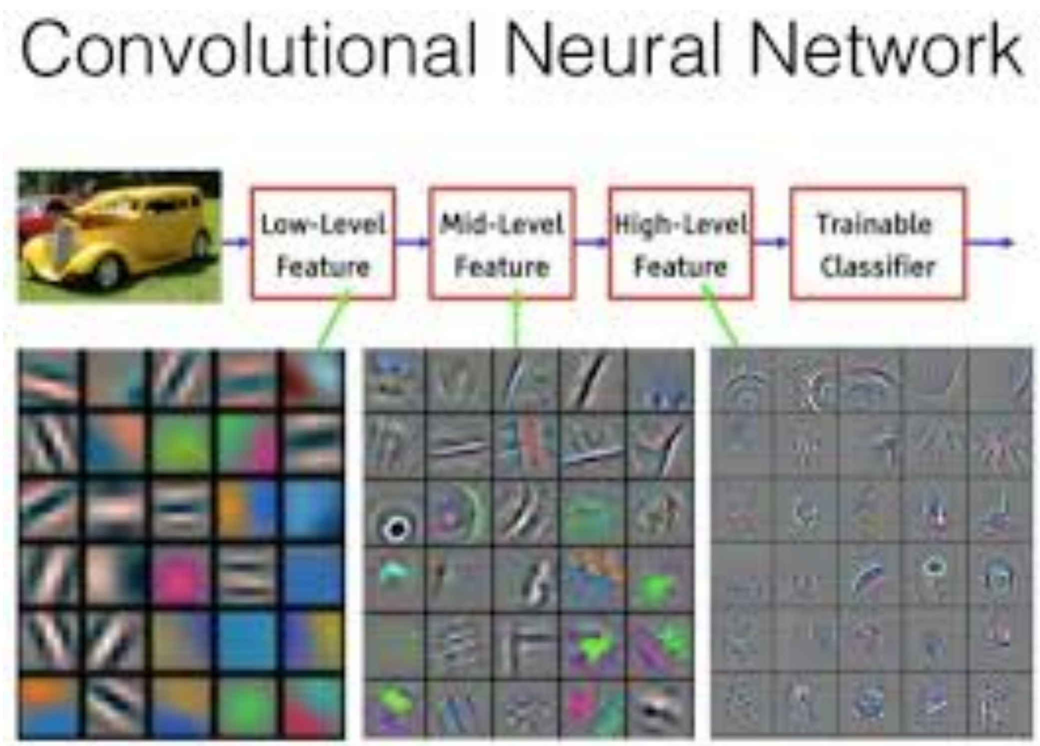


Deep Learning
Learning representation

Deep Learning allows to:

Learn different level of features from low-level to high-level in a kind of hierarchical organisation

Can share the low-level representation for many different tasks



Deep Learning, nowadays, is used in many domains:

Computer Vision (Object Detection and Segmentation, Image SuperResolution, Image Classification)

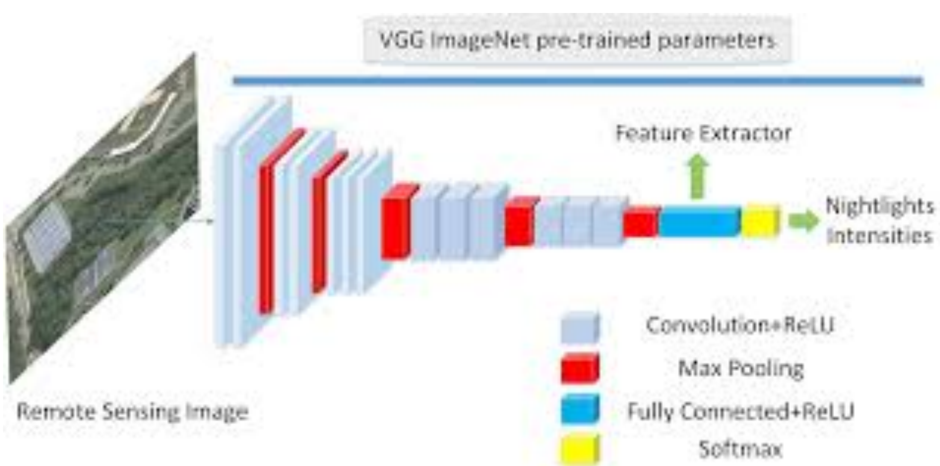
Natural Language Processing (NLP) and Speech

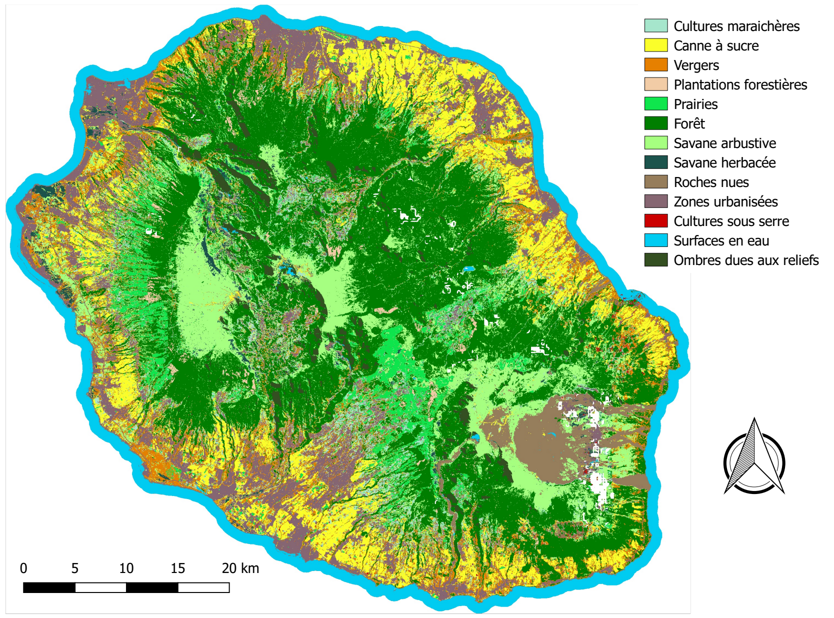
Robotics and AI

Music and arts!

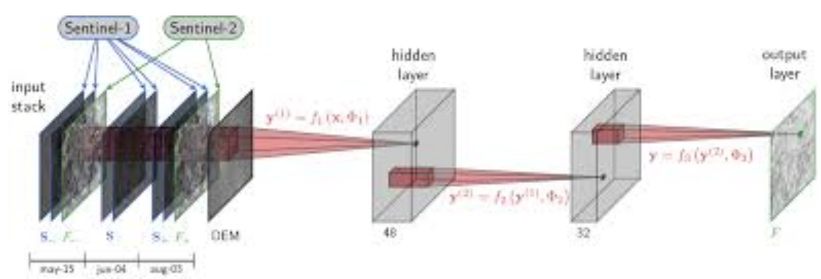
Deep Learning & Earth Observation (EO) Data

Scene Classification



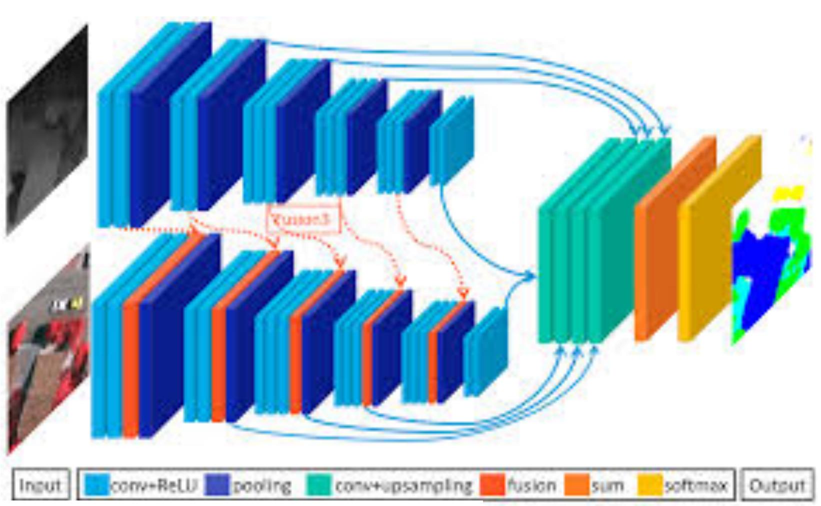


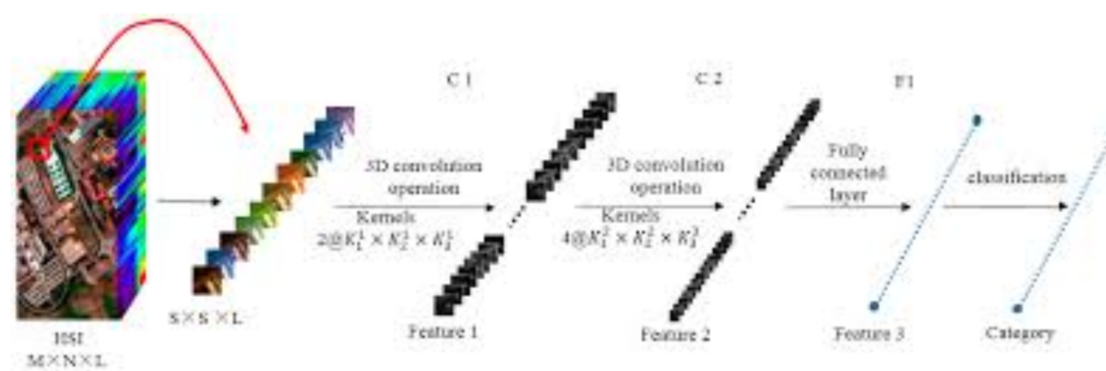
Satellite Image Time Series Analysis



Land Cover Mapping

Remote Sensing Data Fusion





Hyperspectral Classification and Retrieval

Deep Learning & Earth Observation (EO) Data

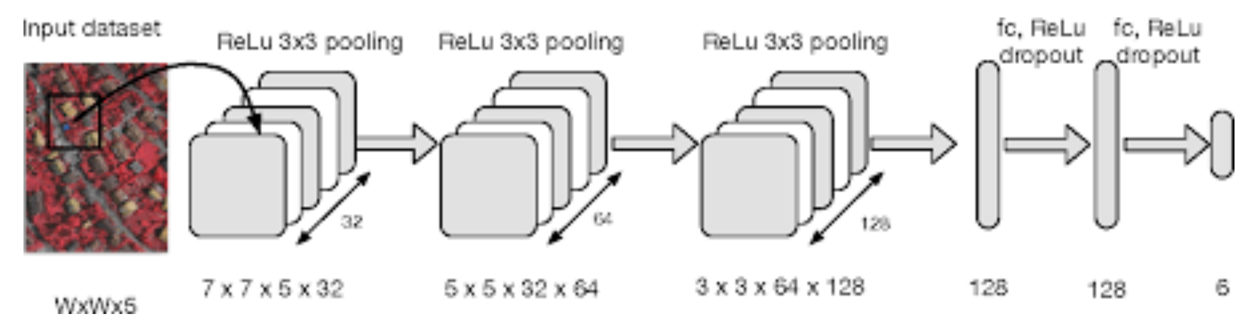
Three popular Deep Learning base blocks in Remote Sensing are:
- Convolutional Neural Network (**CNNs**)
- Recurrent Neural Networks (**RNNs**)
- Convolutional Recurrent Neural Networks (**ConvRNNs**)

Deep Learning & Earth Observation (EO) Data

Three popular Deep Learning base blocks in Remote Sensing are:

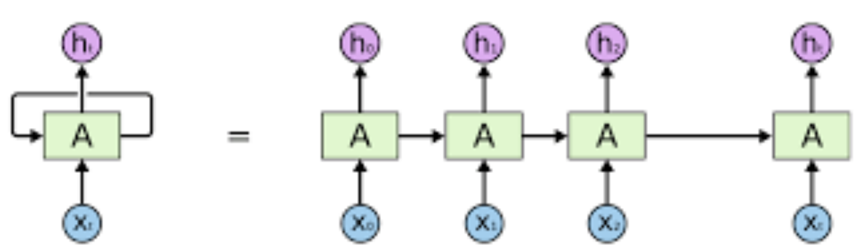
- Convolutional Neural Network (**CNNs**)
- Recurrent Neural Networks (**RNNs**)
- Convolutional Recurrent Neural Networks (**ConvRNNs**)

CNNs



Well suited neural networks to model (mainly) spatial-autocorrelation via Convolution

RNNs



Well suited neural networks to model temporal correlation via recurrent operations

CNNs

CNNs are a special type of neural network whose hidden units are only connected to local receptive field.

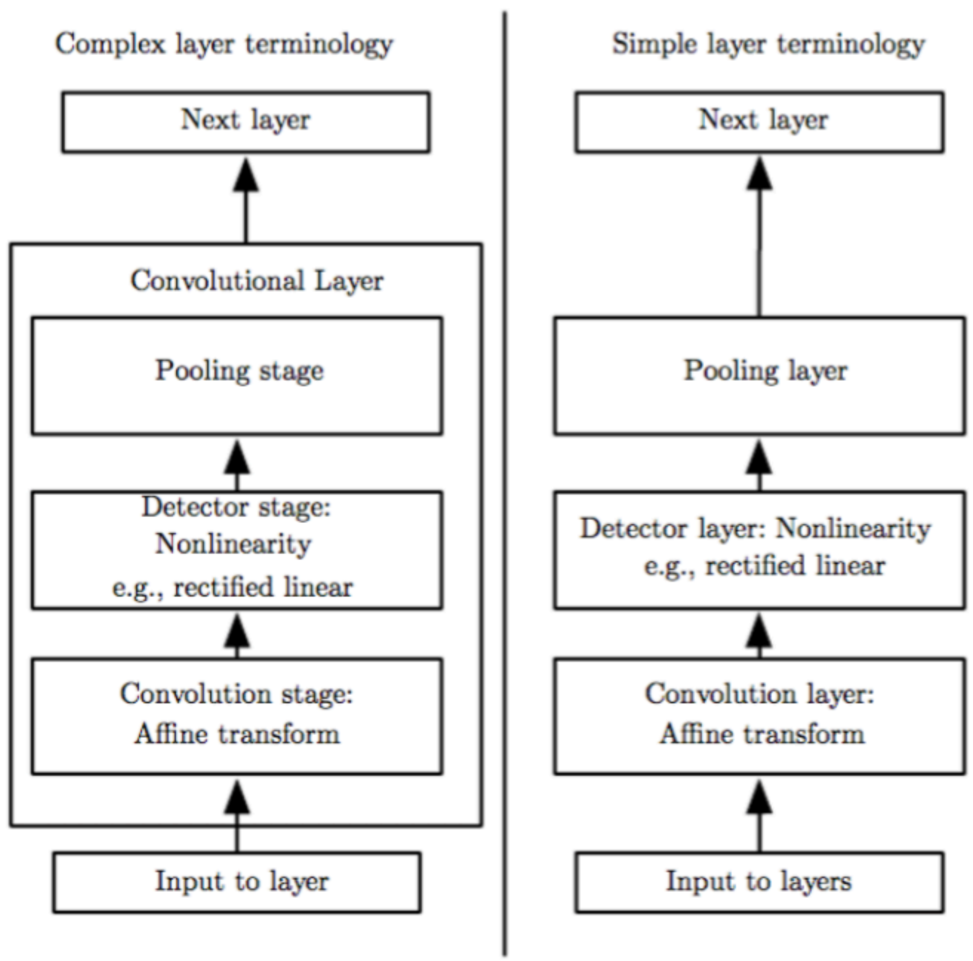
The number of parameters needed by CNNs is much smaller than a Fully Connected counterpart.

Deep Learning & Earth Observation (EO) Data

CNNs

CNNs are a special type of neural network whose hidden units are only connected to local receptive field.

The number of parameters needed by CNNs is much smaller than a Fully Connected counterpart.



CNN has three main stages:

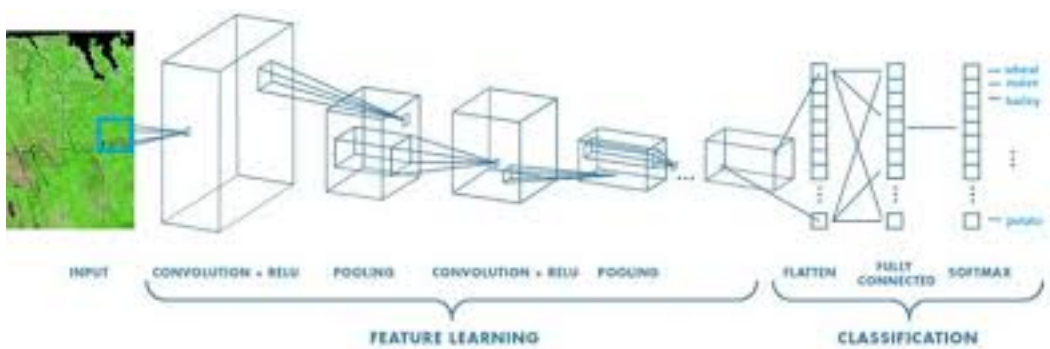
1) Convolution Stage
2) Non-linearity Stage
3) Pooling Stage

Commonly, a normalisation stage is added between 1) and 2)

Deep Learning & Earth Observation (EO) Data

CNN can be employed to perform:
- Image Classification
- Semantic Segmentation

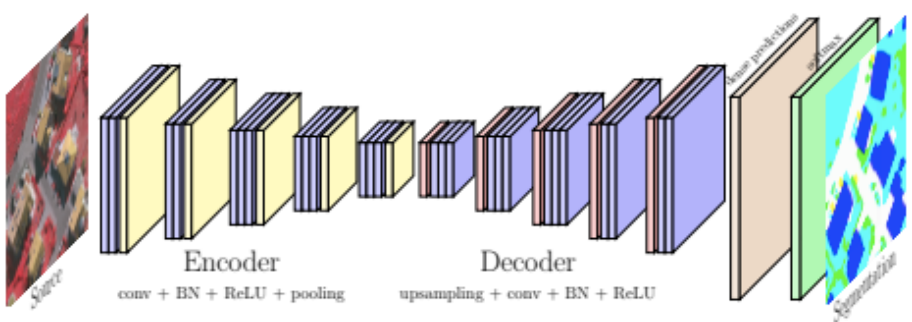
Image Classification



The input is an image and the output is a label for the whole image

In Remote Sensing, the image classification is also employed to perform patch-based classification

Semantic Segmentation



The input is an image, the output is an image with a label for each input pixel

The common architecture for Semantic Segmentation is called AutoEncoder

RNNs

RNNs are a special type of neural network **characterised by recurrent connections**.
The output of the network **at time t** is exploited by the network itself **at time t+1**

Nowadays, two different RNNs model are mainly employed:

- **LSTM** (Long-Short Term Memory)
- **GRU** (Gated Recurrent Unit)

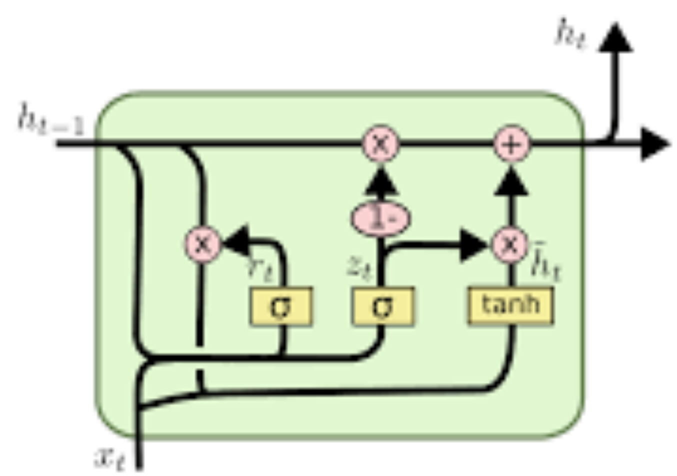
RNNs

RNNs are a special type of neural network **characterised by recurrent connections**.
The output of the network **at time t** is exploited by the network itself **at time t+1**

Nowadays, two different RNNs model are mainly employed:

- **LSTM** (Long-Short Term Memory)
- **GRU** (Gated Recurrent Unit)

Such kind of networks are heavily exploited in Natural Language Processing and Speech Recognition or other kind of 1-D signal

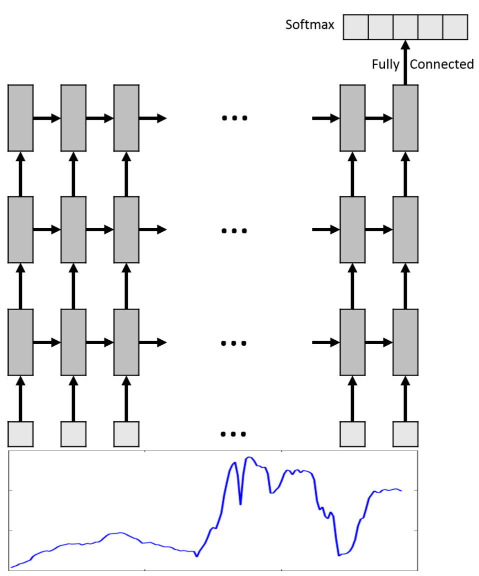


$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$
$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$
$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Deep Learning & Earth Observation (EO) Data

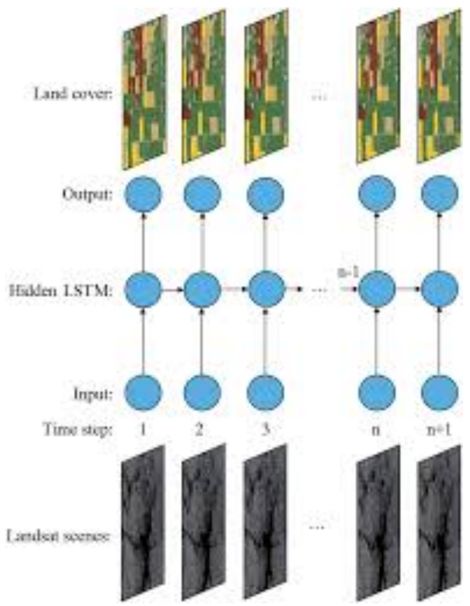
RNN can be employed to perform:
- Signal Classification
- Time Series Analysis

Time Series Classification



The input is a multidimensional Time Series and the output is the classification label

Per-Time classification



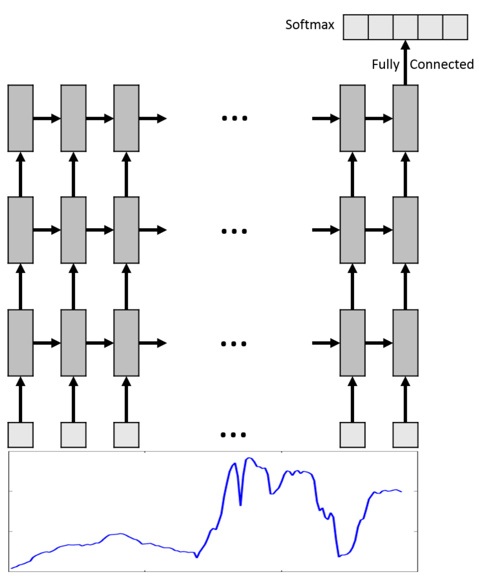
The input is a multidimensional Time Series and the output is a label per timestamps

Deep Learning & Earth Observation (EO) Data

RNN can be employed to perform:

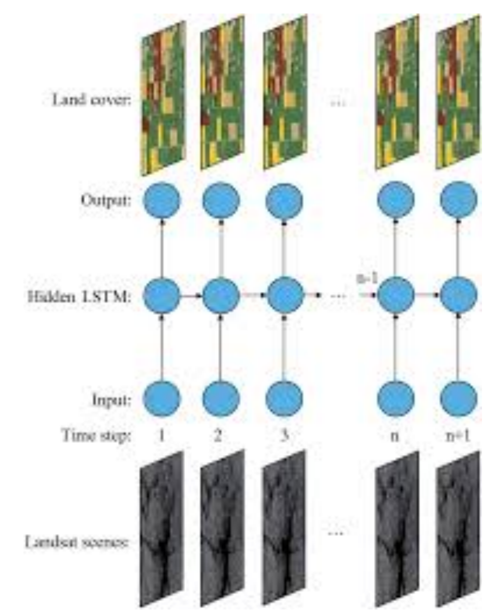
- Signal Classification
- Time Series Analysis

Time Series Classification



The input is a multidimensional Time Series and the output is the classification label

Per-Time classification



The input is a multidimensional Time Series and the output is a label per timestamps

In Remote Sensing, RNN models are especially employed for **Satellite Image Time Series** or **Hyperspectral** data

ConvRNN

ConvRNNs are neural network models that combine Convolutional and Recurrent Neural Network together to manage spatio-temporal information characterised by spatial as well as temporal correlations.

The ConvRNN Unit is a recurrent unit that integrates convolutional filters.

The output of the network **at time t** is exploited by the network itself **at time t+1**

[Shi15] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, W.-c. Woo: Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. NIPS 2015: 802-810

Deep Learning & Earth Observation (EO) Data

ConvRNN

ConvRNNs are neural network models that combine Convolutional and Recurrent Neural Network together to manage spatio-temporal information characterised by spatial as well as temporal correlations.

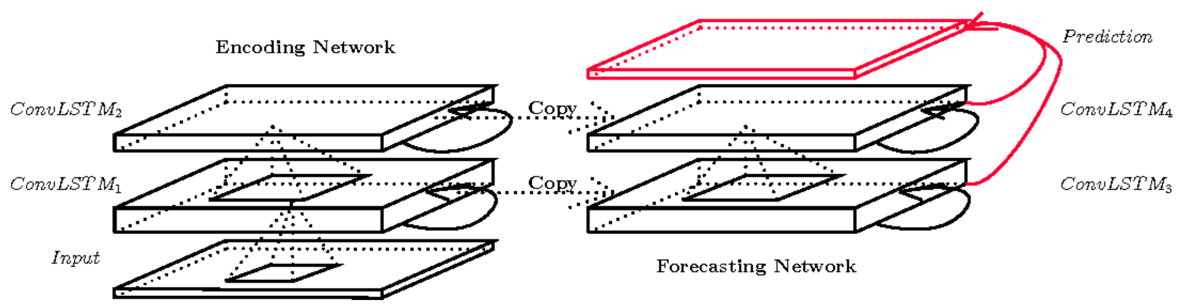
The ConvRNN Unit is a recurrent unit that integrates convolutional filters.

The output of the network **at time t** is exploited by the network itself **at time t+1**

ConvRNN can be derived considering both:

- **LSTM** (Long-Short Term Memory)
- **GRU** (Gated Recurrent Unit)

In which the inner kernel is replaced by convolutional filters



[Shi15]

[Shi15] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, W.-c. Woo: Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. NIPS 2015: 802-810

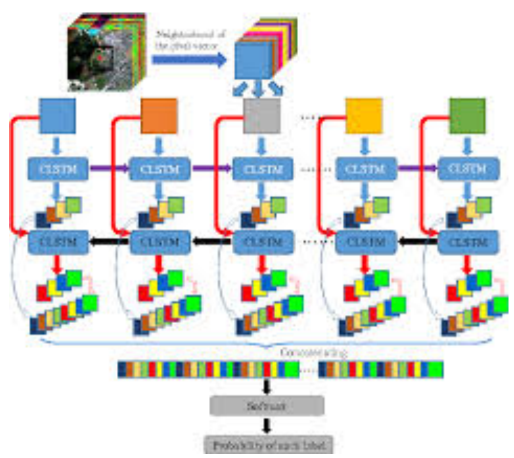
Deep Learning & Earth Observation (EO) Data

ConvRNN

ConvRNN can be employed to perform:

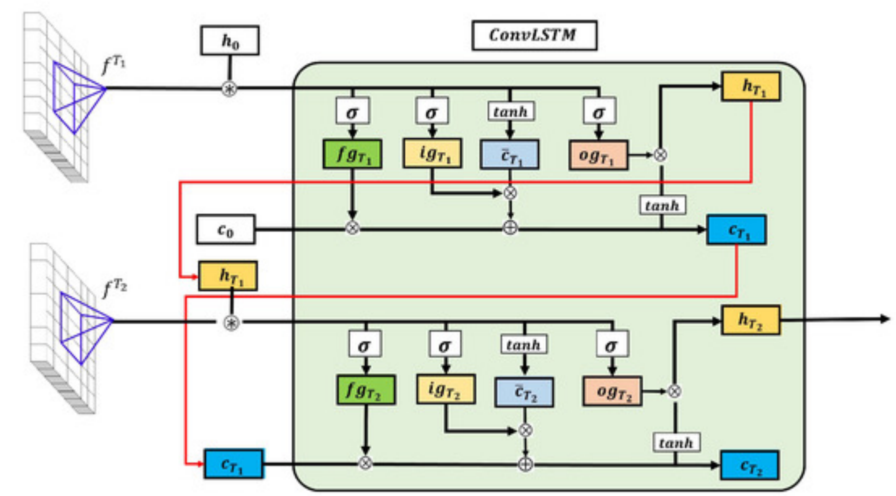
- Semantic Segmentation for Time Series
- Spatio-Spectral Analysis
- Change Detection

Semantic Segmentation for Time Series



The input is a Time Series of images and the output is the classification label for each pixel

Spatio-Spectral Analysis



The input is an hyper spectral signal with spatial context and the output is a label

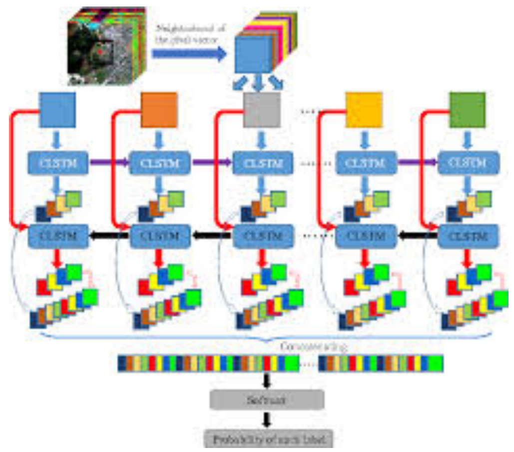
Deep Learning & Earth Observation (EO) Data

ConvRNN

ConvRNN can be employed to perform:

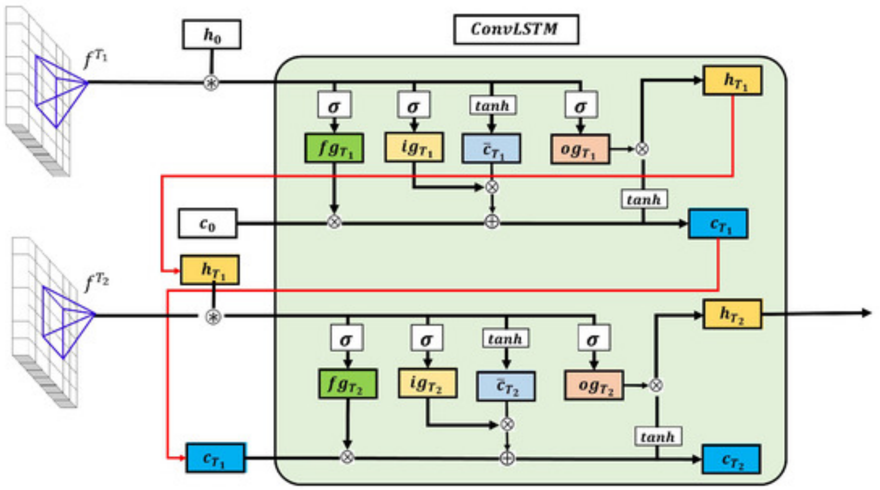
- Semantic Segmentation for Time Series
- Spatio-Spectral Analysis
- Change Detection

Semantic Segmentation for Time Series



The input is a Time Series of images and the output is the classification label for each pixel

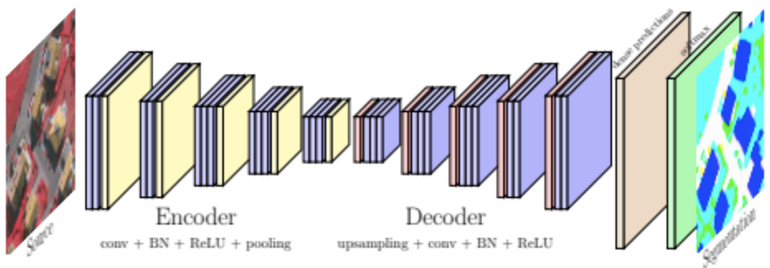
Spatio-Spectral Analysis



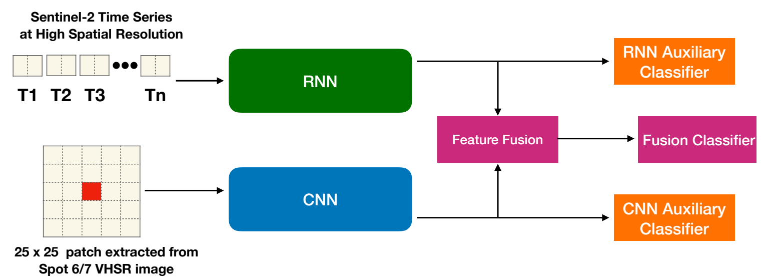
The input is an hyper spectral signal with spatial context and the output is a label

In Remote Sensing, ConvRNN models are especially employed for **Satellite Image Time Series** or **Hyperspectral** data

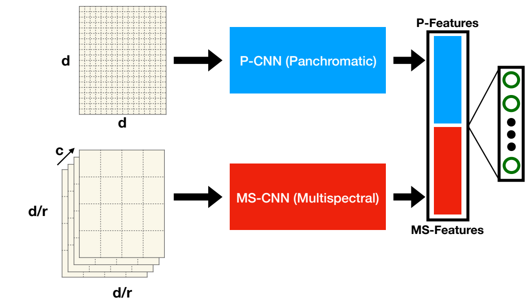
Deep Learning & EO Data Fusion



(Very High Spatial Resolution)
VHSR + DEM [Audebert17]



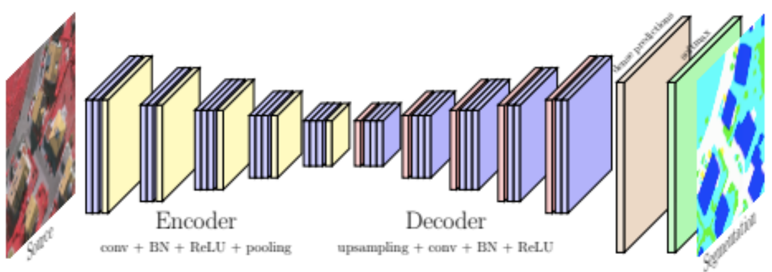
Time Series + VHSR [Benedetti18]



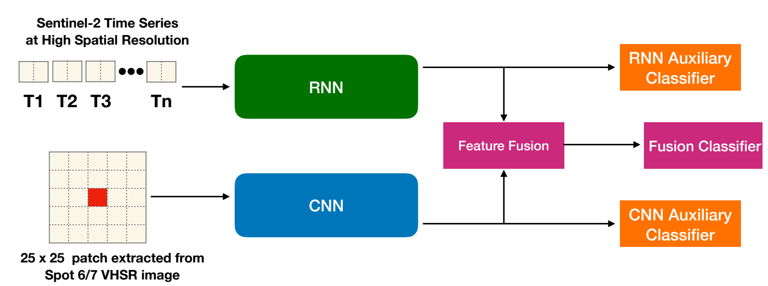
Pan + MS information from VHSR
[Gaetano18]

[Chen17] Y. Chen, C. Li, P. Ghamisi, X. Jia, Y. Gu: Deep Fusion of Remote Sensing Data for Accurate Classification. IEEE GRSL 14(8): 1253-1257 (2017)
[Audebert17] N. Audebert, B. Le Saux, S. Lefèvre: Beyond RGB: Very High Resolution Urban Remote Sensing With Multimodal Deep Networks. ISPRS J. of Photogrammetry and Rem. Sens. 140, 20-32 (2018)
[Benedetti18] P. Benedetti, D. Ienco, R. Gaetano, K. Ose, R. G. Pensa, S. Dupuy: M3Fusion: A Deep Learning Architecture for Multi-{Scale/Modal/Temporal} satellite data fusion. IEEE JSTARS (2018)
[Gaetano18] R. Gaetano, D. Ienco, K. Ose, C. Cresson: MRFusion: A Deep Learning architecture to fuse PAN and MS imagery for land cover mapping CoRR abs/ (2018)
[Ienco19] D. Ienco, R. Gaetano, R. Interdonato, K. Ose and D. Ho Tong Minh: Combining Sentinel-1 and Sentinel-2 time series via RNN for object-based Land Cover Classification. IGARSS (2019).
[Cresson19] R. Cresson, D. Ienco, R. Gaetano, K. Ose and D. Ho Tong Minh: Optical images gap filling with deep convolutional autoencoder. IGARSS (2019).

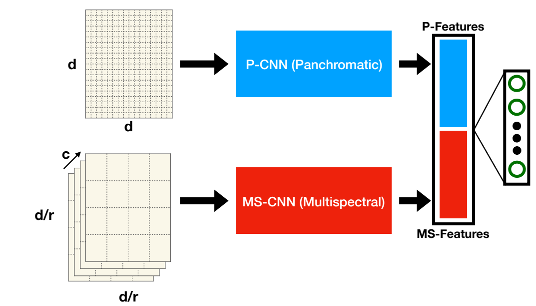
Deep Learning & EO Data Fusion



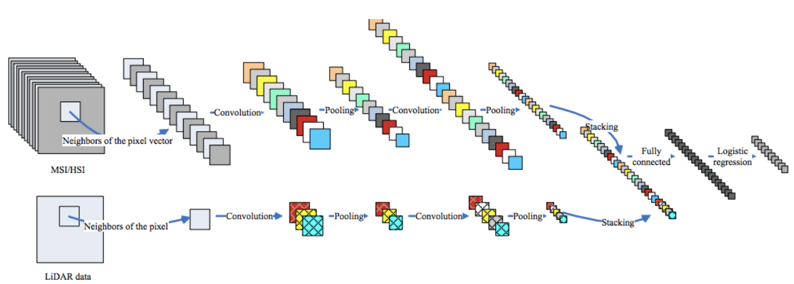
(Very High Spatial Resolution) VHSR + DEM [Audebert17]



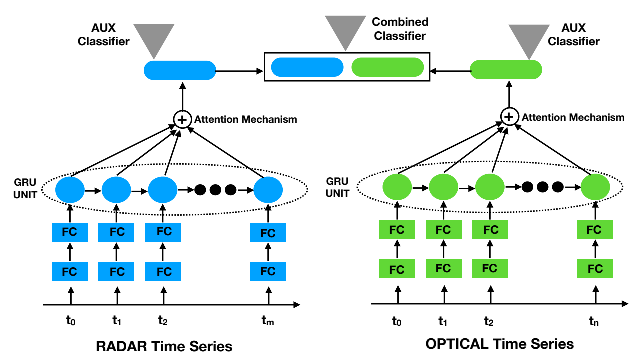
Time Series + VHSR [Benedetti18]



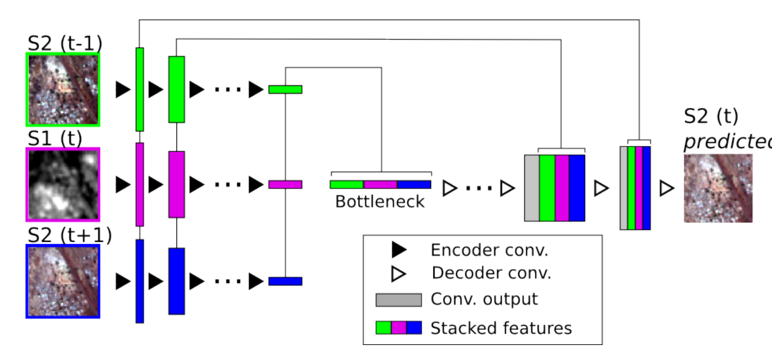
Pan + MS information from VHSR [Gaetano18]



Hyperspectral + DEM [Chen17]



Sentinel1 & Sentinel2 Satellite Image Time Series Classification [Ienco19]

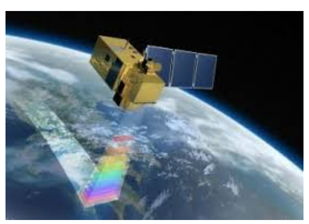


Sentinel1 & Sentinel2 Satellite Image Time Series Restoration [Cresson19]

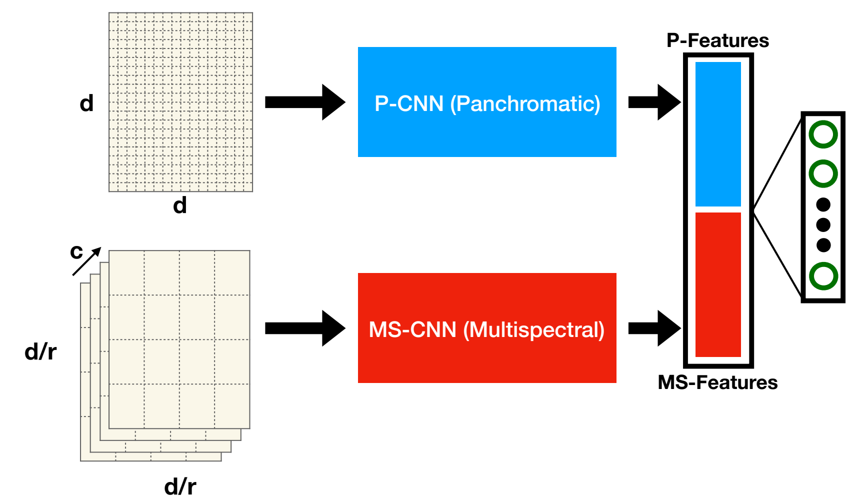
[Chen17] Y. Chen, C. Li, P. Ghamisi, X. Jia, Y. Gu: Deep Fusion of Remote Sensing Data for Accurate Classification. IEEE GRSL 14(8): 1253-1257 (2017)
[Audebert17] N. Audebert, B. Le Saux, S. Lefèvre: Beyond RGB: Very High Resolution Urban Remote Sensing With Multimodal Deep Networks. ISPRS J. of Photogrammetry and Rem. Sens. 140, 20-32 (2018)
[Benedetti18] P. Benedetti, D. Ienco, R. Gaetano, K. Ose, R. G. Pensa, S. Dupuy: M3Fusion: A Deep Learning Architecture for Multi-{Scale/Modal/Temporal} satellite data fusion. IEEE JSTARS (2018)
[Gaetano18] R. Gaetano, D. Ienco, K. Ose, C. Cresson: MRFusion: A Deep Learning architecture to fuse PAN and MS imagery for land cover mapping CoRR abs/ (2018)
[Ienco19] D. Ienco, R. Gaetano, R. Interdonato, K. Ose and D. Ho Tong Minh: Combining Sentinel-1 and Sentinel-2 time series via RNN for object-based Land Cover Classification. IGARSS (2019).
[Cresson19] R. Cresson. D. Ienco. R. Gaetano, K. Ose and D. Ho Tong Minh: Optical images gap filling with deep convolutional autoencoder. IGARSS (2019).

Deep Learning & EO Data Fusion

MRFusion: A DL approach to fuse PAN and MS for LC mapping



Single-Sensor Data Fusion on SPOT6:
- **Panchromatic Image (1.5m)**
- **Multi-Spectral Image (6m)**

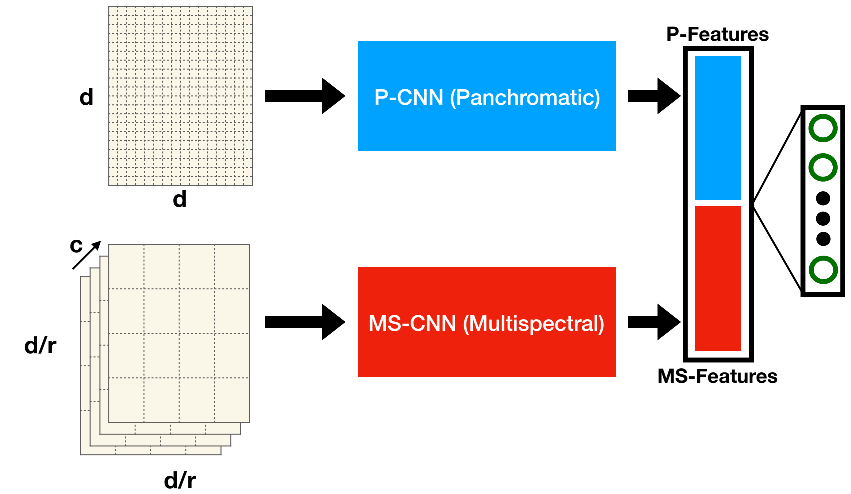


d = patch size on the PAN image
r = spatial ratio between PAN and MS (i.e. in SPOT6 is 4)
c = number of channels in the MS image

Deep Learning & EO Data Fusion

MRFusion: A DL approach to fuse PAN and MS for LC mapping

Single-Sensor Data Fusion on SPOT6:
- **Panchromatic Image (1.5m)**
- **Multi-Spectral Image (6m)**



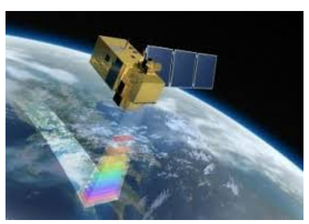
d = patch size on the PAN image
r = spatial ratio between PAN and MS (i.e. in SPOT6 is 4)
c = number of channels in the MS image

Reunion Island Dataset:
- Spot6 image
- 13 Land Cover Classes
- PAN Image 44374 x 39422
- MS Image 11094 x 9856

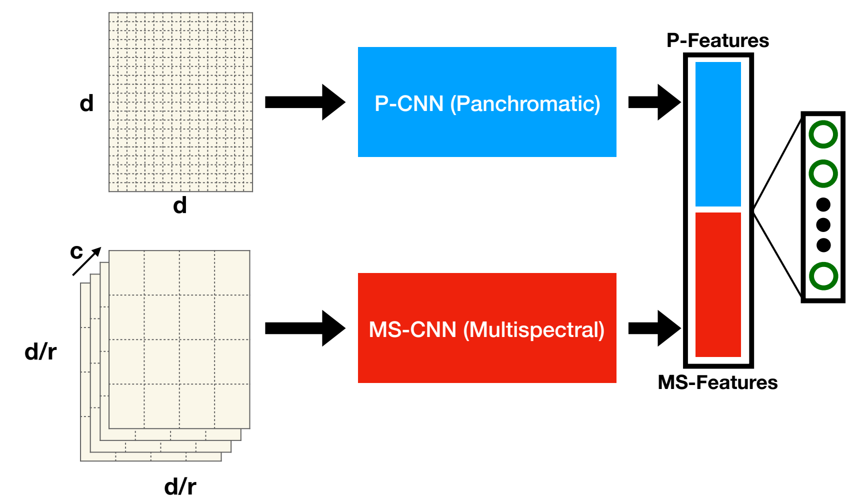
Class	Label	# Objects	# Pixels
1	*Crop Cultivations*	168	50061
2	*Sugar cane*	167	50100
3	*Orchards*	167	50092
4	*Forest plantations*	67	20100
5	*Meadow*	167	50100
6	*Forest*	167	50100
7	*Shrubby savannah*	173	50263
8	*Herbaceous savannah*	78	23302
9	*Bare rocks*	107	31587
10	*Urban areas*	125	36046
11	*Greenhouse crops*	49	14387
12	*Water Surfaces*	96	2711
13	*Shadows*	38	11400

Deep Learning & EO Data Fusion

MRFusion: A DL approach to fuse PAN and MS for LC mapping



Single-Sensor Data Fusion on SPOT6:
- **Panchromatic Image (1.5m)**
- **Multi-Spectral Image (6m)**



d = patch size on the PAN image
r = spatial ratio between PAN and MS (i.e. in SPOT6 is 4)
c = number of channels in the MS image

Reunion Island Dataset:
- Spot6 image
- 13 Land Cover Classes
- PAN Image 44374 x 39422
- MS Image 11094 x 9856

Class	Label	# Objects	# Pixels
1	*Crop Cultivations*	168	50061
2	*Sugar cane*	167	50100
3	*Orchards*	167	50092
4	*Forest plantations*	67	20100
5	*Meadow*	167	50100
6	*Forest*	167	50100
7	*Shrubby savannah*	173	50263
8	*Herbaceous savannah*	78	23302
9	*Bare rocks*	107	31587
10	*Urban areas*	125	36046
11	*Greenhouse crops*	49	14387
12	*Water Surfaces*	96	2711
13	*Shadows*	38	11400

	Accuracy	*F-Measure*	*Kappa*
RF(PIXEL)	24.87 ± 0.2	23.66 ± 0.2	0.1719 ± 0.0024
RF(PATCH)	72.22 ± 1.31	71.53 ± 1.4	0.6943 ± 0.0144
CNN_{PS}	74.49 ± 1.20	74.25 ± 1.24	0.7195 ± 0.0131
DMIL	69.40 ± 1.11	69.34 ± 1.12	0.6637 ± 0.0121
MRFusion	**79.65** ± 0.87	**79.56** ± 0.91	**0.7764** ± 0.0096

Reunion Island Results

Conclusion and Directions

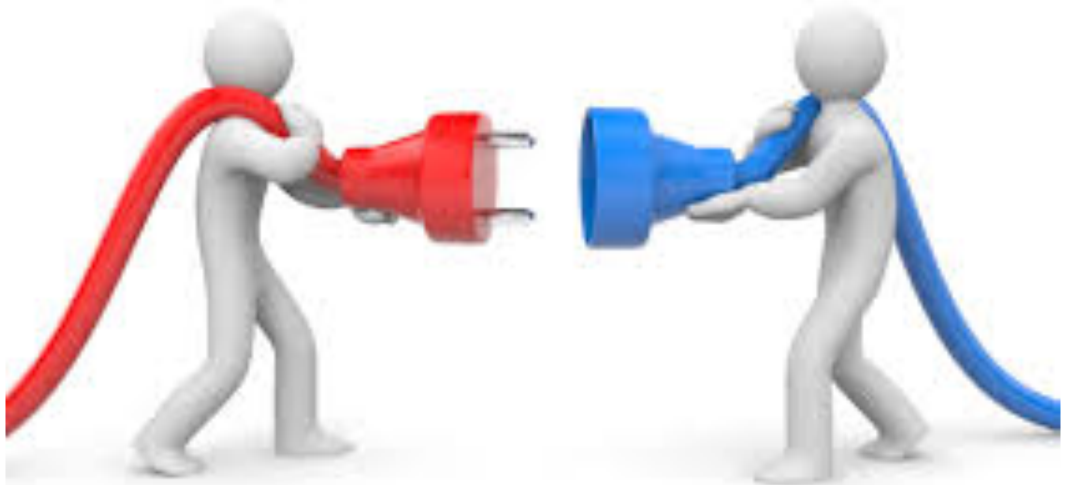
Deep Learning seems a **promising tool for task-driven multi source data fusion** in Remote Sensing.

Most of the literature in Remote Sensing & Deep Learning exploits (almost) directly results from Computer Vision but… Remote Sensing has some peculiarity (multi-scale, multi-sensors, multi-temporal, sparsely annotated data, etc…) -> **Necessity for ad-hoc architectures.**

In operational cases, when predictive analysis need to be deployed, **some sensors can be damaged or unavailable**. How to develop methods capable to work on **misaligned (between training and test) information sources** is mandatory in Remote Sensing.

Many efforts were done in creating and developing **physical-based models and now?** Data-Driven models seem overpass previous work but…**How to combine physical-based and (DL) data-driven models** in Remote Sensing is a promising direction. **Data Assimilation can be an answer.**

Thank you for your Attention



tetis
TERRITOIRE ENVIRONNEMENT TÉLÉDÉTECTION
INFORMATION SPATIALE

Thank you for your Attention

Questions

AgroParisTech cirad cnrs irstea AGENCE NATIONALE DE LA RECHERCHE ANR #DigitAg INVESTISSEMENTS D'AVENIR