M3Fusion : Un modèle d'apprentissage profond pour la fusion de données satellitaires Multi-{Echelles/Modalités/Temporelles}

Paola Benedetti, Ms.c (paola.benedetti@irstea.fr)

Dino Ienco (dino.ienco@irstea.fr)

Raffaele Gaetano (raffaele.gaetano@cirad.fr)

Kenji Osé (kenji.ose@irstea.fr)

Ruggero Pensa (ruggero.pensa@unito.it)

Stephane Dupuy (stephane.dupuy@cirad.fr)

Outline

Earth Observation Data Fusion Challenge

M3F: Spatio-Temporal Data Fusion via Deep Learning

Data Description

Experimental Settings

Results & Findings

Conclusions and Future Works

Earth Observation Data Fusion Challenge

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)

Earth Observation Data Fusion Challenge

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)

Main Challenge

How to exploit the complementarity of Satellite information? i.e. How to leverage such heterogeneity for Land Cover?

Earth Observation Data Fusion Challenge

Different Data fusion scenario [Schmitt16]:

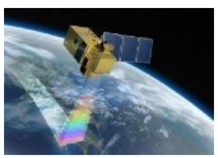
- Single-Sensor Data Fusion
- Multiple-Sensor Data Fusion
- Temporal Data Fusion
- Machine Learning-Based Data Fusion
- And so on….

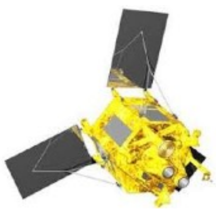
[Schmitt16] M. Schmitt and X. X. Zhu, "Data Fusion and Remote Sensing: An ever-growing relationship". IEEE Geoscience and Remote Sensing Magazine 4(4): 6--23, 2016

Earth Observation Data Fusion Challenge

Different Data fusion scenario [Schmitt16]:

- Single-Sensor Data Fusion
- Multiple-Sensor Data Fusion
- Temporal Data Fusion
- Machine Learning-Based Data Fusion
- And so on….





In our work we focus our effort on the fusion between:

- **Sentinel 2** (S2) Time Series
- **SPOT6** Very High Spatial Resolution image

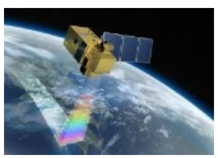
Optical Multiple-**Sensor**/Multi-**Temporal**/Multi-**Scale**

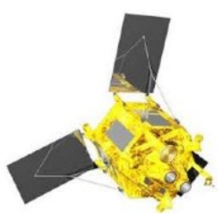
[Schmitt16] M. Schmitt and X. X. Zhu, "Data Fusion and Remote Sensing: An ever-growing relationship". IEEE Geoscience and Remote Sensing Magazine 4(4): 6--23, 2016

Earth Observation Data Fusion Challenge

Different Data fusion scenario [Schmitt16]:

- Single-Sensor Data Fusion
- Multiple-Sensor Data Fusion
- Temporal Data Fusion
- Machine Learning-Based Data Fusion
- And so on….





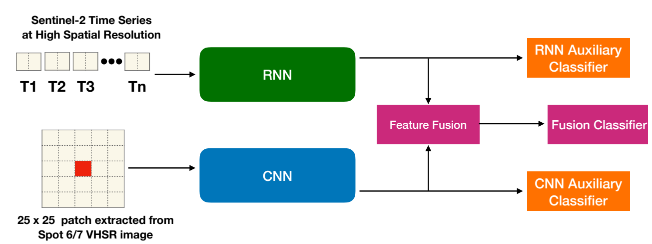
In our work we focus our effort on the fusion between:

- **Sentinel 2** (S2) Time Series
- **SPOT6** Very High Spatial Resolution image

Optical Multiple-**Sensor**/Multi-**Temporal**/Multi-**Scale**

To this end, we conceive a Deep Learning approach leveraging:

- Convolutional Neural Network (SPOT6)
- Recurrent Neural Network (S2 Time Series)

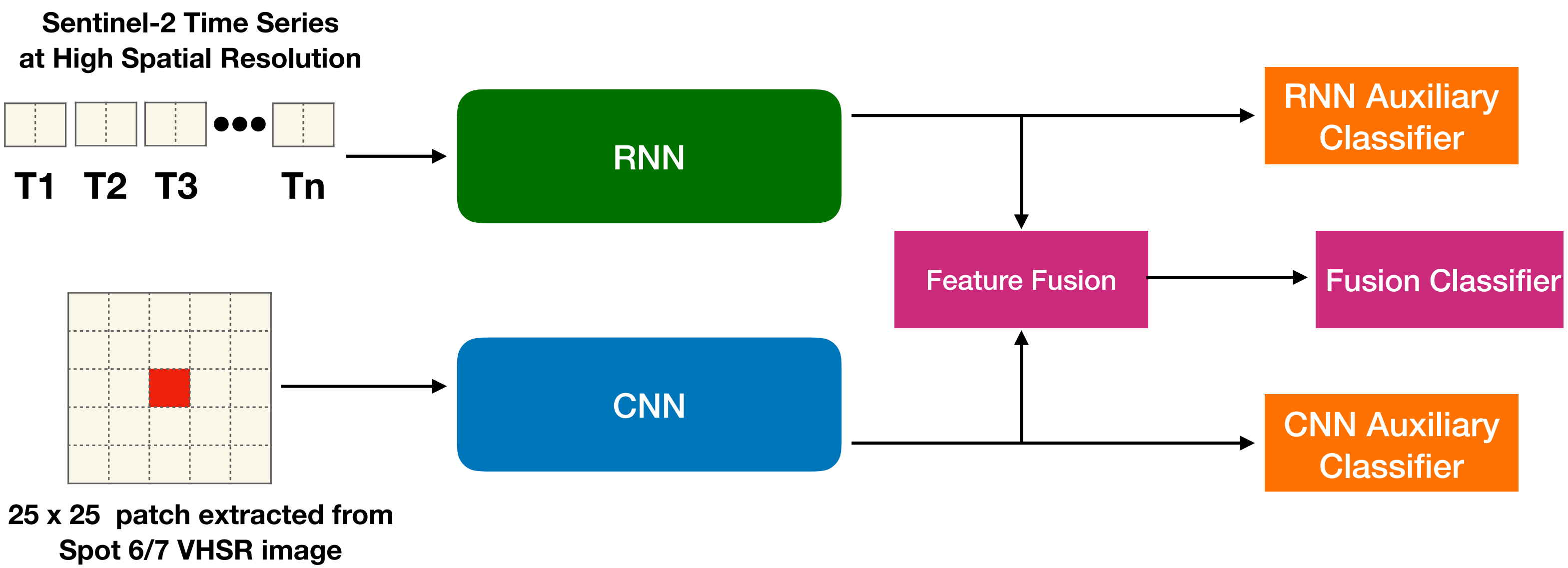


[Schmitt16] M. Schmitt and X. X. Zhu, "Data Fusion and Remote Sensing: An ever-growing relationship". IEEE Geoscience and Remote Sensing Magazine 4(4): 6--23, 2016

AgroParisTech cirad cnrs irstea

M3F: Spatio-Temporal Data Fusion via Deep Learning

M3Fusion: Multi-{Scale/Modality/Temporal} data fusion architecture



An example is described by:
- A (multi-dimensional) pixel time series
- A 25 x 25 patch coming from VHSR image

M3F: Spatio-Temporal Data Fusion via Deep Learning

GRU with Attention - Temporal Component

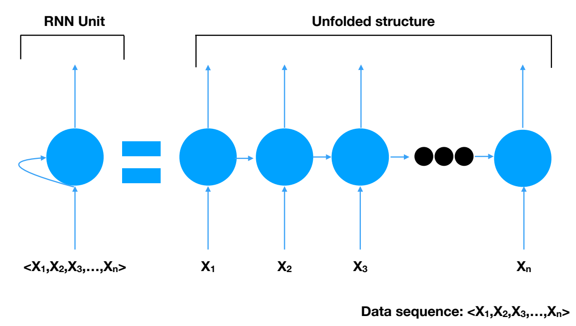
M3F: Spatio-Temporal Data Fusion via Deep Learning

GRU with Attention - Temporal Component

$$z_t = \sigma(W_{zx}x_t + W_{zh}h_{t-1} + b_z)$$
$$r_t = \sigma(W_{rx}x_t + W_{rh}h_{t-1} + b_r)$$
$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tanh(W_{hx}x_t + W_{hr}(r_t \odot h_{t-1}) + b_h)$$

We use DropOut to alleviate overfitting

Gated Recurrent Unit:
- Lighter architecture than LSTM
- Recurrent Unit with gates
- Widely employed in NLP



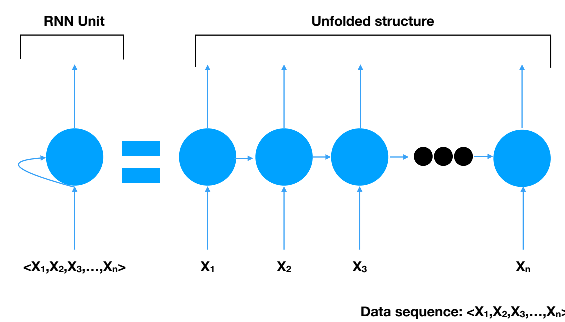
M3F: Spatio-Temporal Data Fusion via Deep Learning

GRU with Attention - Temporal Component

$$z_t = \sigma(W_{zx}x_t + W_{zh}h_{t-1} + b_z)$$
$$r_t = \sigma(W_{rx}x_t + W_{rh}h_{t-1} + b_r)$$
$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tanh(W_{hx}x_t + W_{hr}(r_t \odot h_{t-1}) + b_h)$$

Gated Recurrent Unit:
- Lighter architecture than LSTM
- Recurrent Unit with gates
- Widely employed in NLP

We use DropOut to alleviate overfitting



Attention Mechanism
Combine the information extracted at each timestamps together

$$v_a = tanh(H \cdot W_a + b_a)$$
$$\lambda = SoftMax(v_a \cdot u_a)$$
$$rnn_{feat} = \sum_{i=1}^{N} \lambda_i \cdot h_{t_i}$$

M3F: Spatio-Temporal Data Fusion via Deep Learning

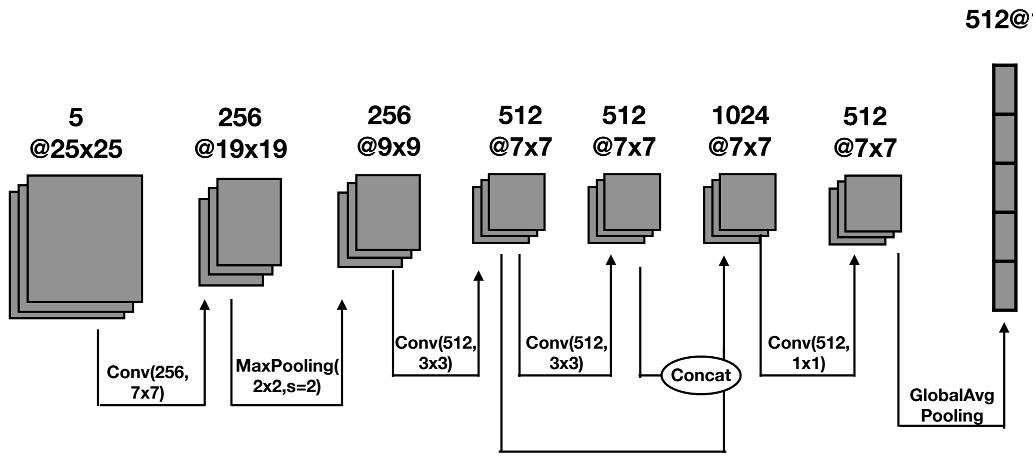
CNN for Spatial Information

M3F: Spatio-Temporal Data Fusion via Deep Learning

CNN for Spatial Information

Relu

Batch Normalization

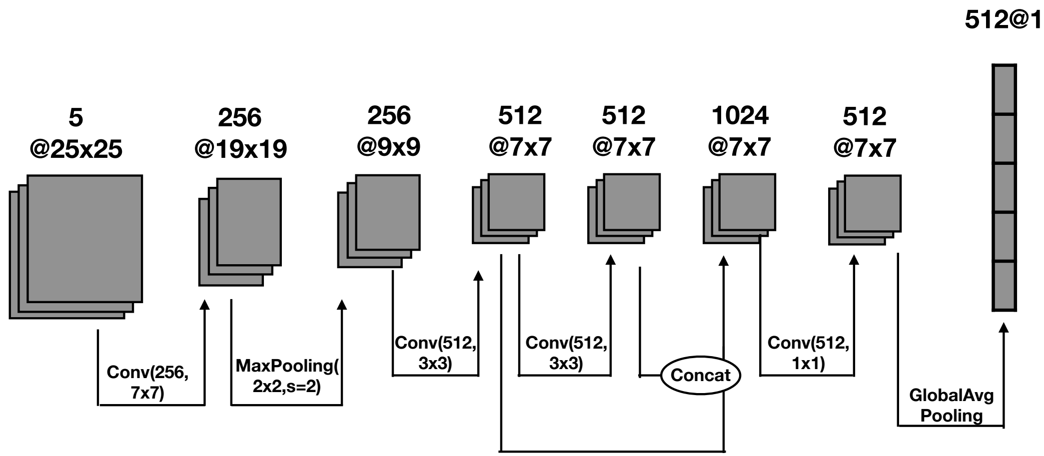


M3F: Spatio-Temporal Data Fusion via Deep Learning

CNN for Spatial Information

Relu

Batch Normalization



Loss to Optimise:

$$
\begin{aligned}
L_{total} &= \alpha_1 * L_{rnn}(rnn_{feat}, W_1, b_1) + \\
&= \alpha_2 * L_{cnn}(cnn_{feat}, W_2, b_2) + \\
&= L_{fus}([cnn_{feat}, rnn_{feat}], W_3, b_3)
\end{aligned}
$$

Categorical Cross-Entropy to compute the Loss of each set of features

M3F: Spatio-Temporal Data Fusion via Deep Learning

Ent-To-End Process from scratch

RNN Module dedicated to Time Series Data **(temporal correlation)**

CNN Module dedicated to VHSR data **(spatial neighbourhood)**

Multi-Scale and Multi-Source data fusion automatically managed by the architecture

Dedicated approach to **fuse** together **Temporal and Non-Termporal** information by **Deep Learning**

[Hou17] S. Hou, X. Liu, Z. Wang: DualNet: Learn Complementary Features for Image Recognition. ICCV 2017: 502-510

M3F: Spatio-Temporal Data Fusion via Deep Learning

Ent-To-End Process from scratch

RNN Module dedicated to Time Series Data **(temporal correlation)**

CNN Module dedicated to VHSR data **(spatial neighbourhood)**

Multi-Scale and Multi-Source data fusion automatically managed by the architecture

Dedicated approach to **fuse** together **Temporal and Non-Termporal** information by **Deep Learning**

Auxiliary Classifiers adapted from [Hou17], the goal is to boost the discrimination power of each set fo features independently

In our context, sources are naturally complementary w.r.t the work proposed in [Hou17]

[Hou17] S. Hou, X. Liu, Z. Wang: DualNet: Learn Complementary Features for Image Recognition. ICCV 2017: 502-510

Data Description

Reunion Island Study Site:
- Covered Area: 2512 Km2
- French Department located in Indian Ocean

Goal of the Land cover mapping task:
A 13 classes classification

We use two sources of data:
- **Time Series** of Optical Satellite Images (Sentinel-2)
  - Acquired between April 2016 and May 2017
  - 34 images at 10m of resolution

- A **Very High Spatial Resolution** (VHSR) Image SPOT6
  - Acquired in April 2016
  - 1 image at 1.5m of resolution (resampled at 2m)

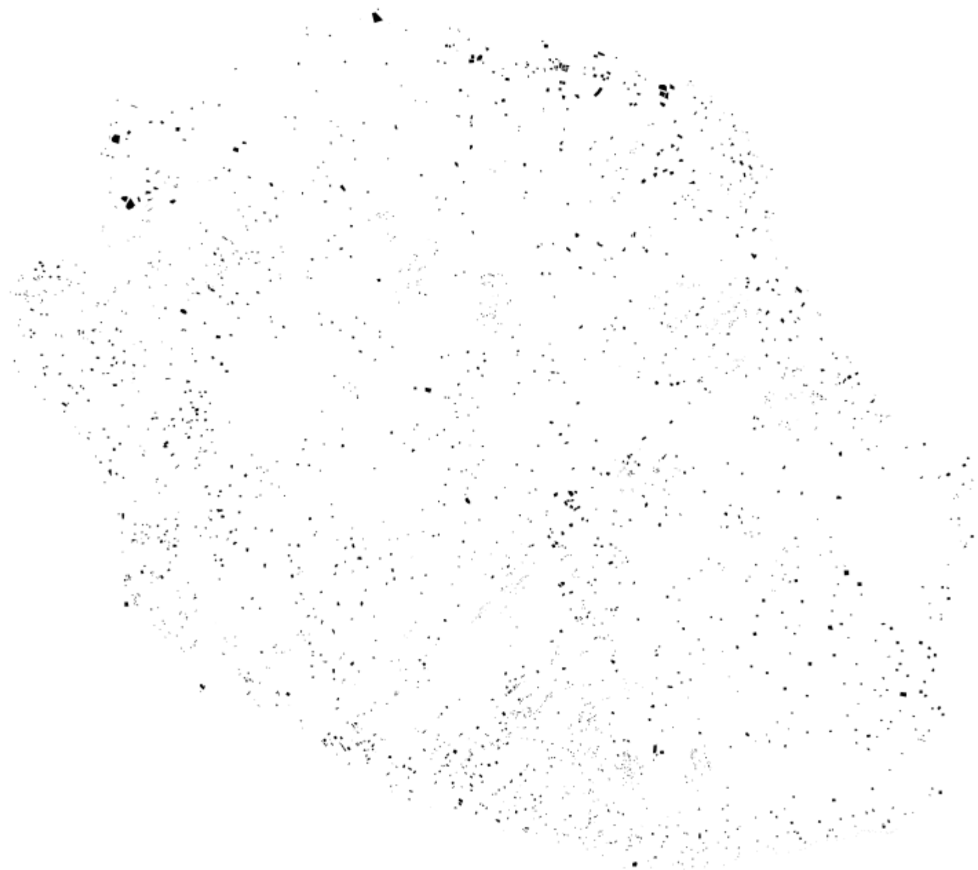
VHSR Image size (in pixels) 33280 X 29565
S2 scene size (in pixels) 6656 x 5913

Data Description

Reunion Dataset Characteristics

Class	Label	# Objects	# Pixels
1	*Crop Cultivations*	380	12090
2	*Sugar cane*	496	84136
3	*Orchards*	299	15477
4	*Forest plantations*	67	9783
5	*Meadow*	257	50596
6	*Forest*	292	55108
7	*Shrubby savannah*	371	20287
8	*Herbaceous savannah*	78	5978
9	*Bare rocks*	107	18659
10	*Urbanized areas*	125	36178
11	*Greenhouse crops*	50	1877
12	*Water Surfaces*	96	7349
13	*Shadows*	38	5230

A total of 322 748 pixels (2656 objects) over 13 classes



Spatial Distribution of Reference Data

Reference Data obtained by:

- RPG data (2014)
- GPS record (June 2017)
- Photo interpretation of VHSR Image

Experimental Settings

Training Dataset:

- N. Objects: 30% of the Original Objects (around 800)
- N. Pixels: 97110

Test Dataset:

- N. Objects: 70% of the Original Objects (around 1800)
- N. Pixels: 225638

Competitor:

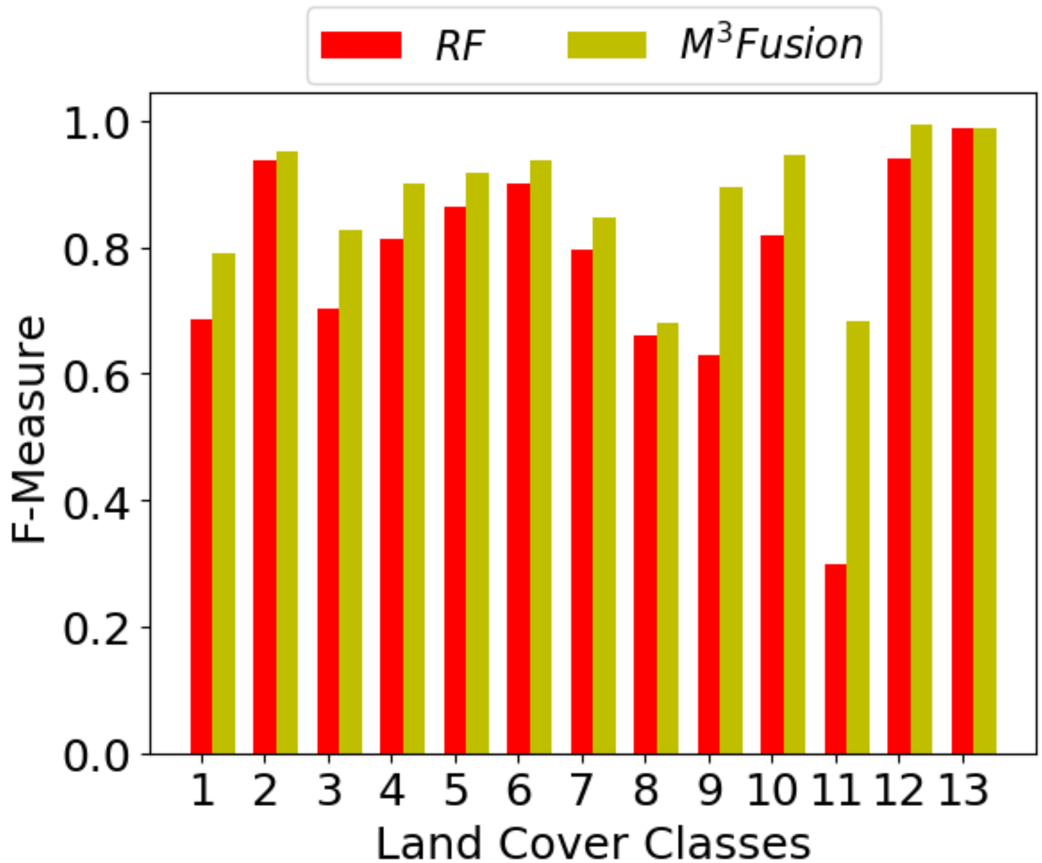
Common Machine Learning Approach (**Random Forest**) applied on the temporal and non temporal data stacked together

Evaluation Measures (On Test Data):

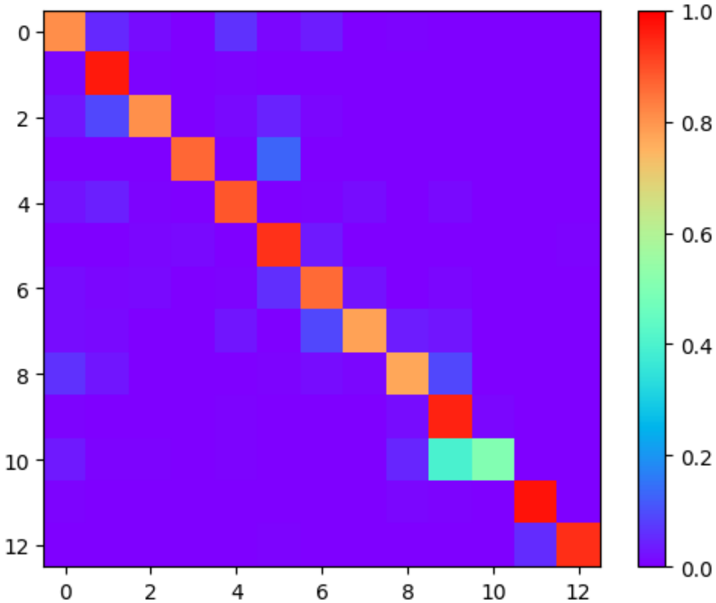
Accuracy (Global Accuracy)
F-Measure (it helps to take into account unbalance class distribution)
Kappa Measure

Comparison with Standard ML method

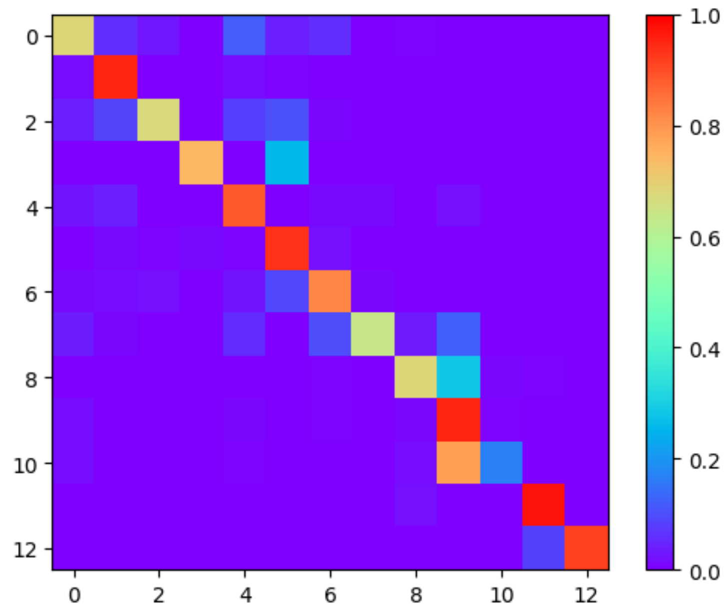
Class	Label
1	*Crop Cultivations*
2	*Sugar cane*
3	*Orchards*
4	*Forest plantations*
5	*Meadow*
6	*Forest*
7	*Shrubby savannah*
8	*Herbaceous savannah*
9	*Bare rocks*
10	*Urbanized areas*
11	*Greenhouse crops*
12	*Water Surfaces*
13	*Shadows*



Comparison with Standard ML method



M3Fusion



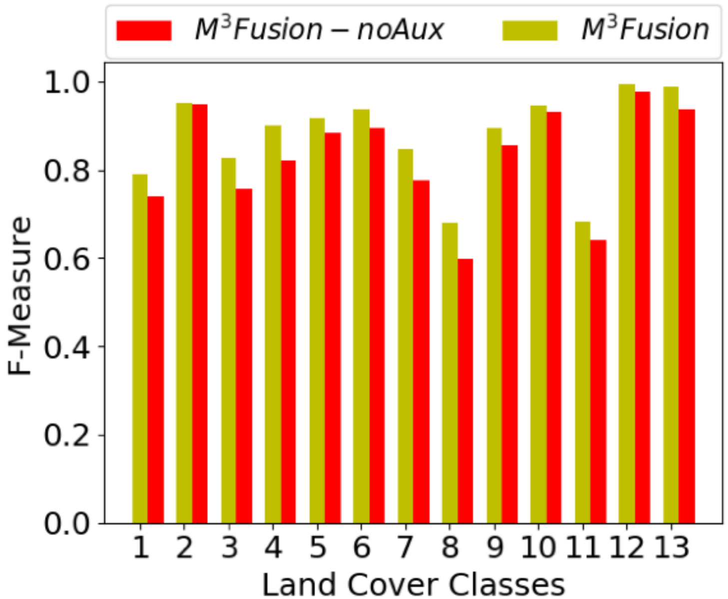
Random Forest

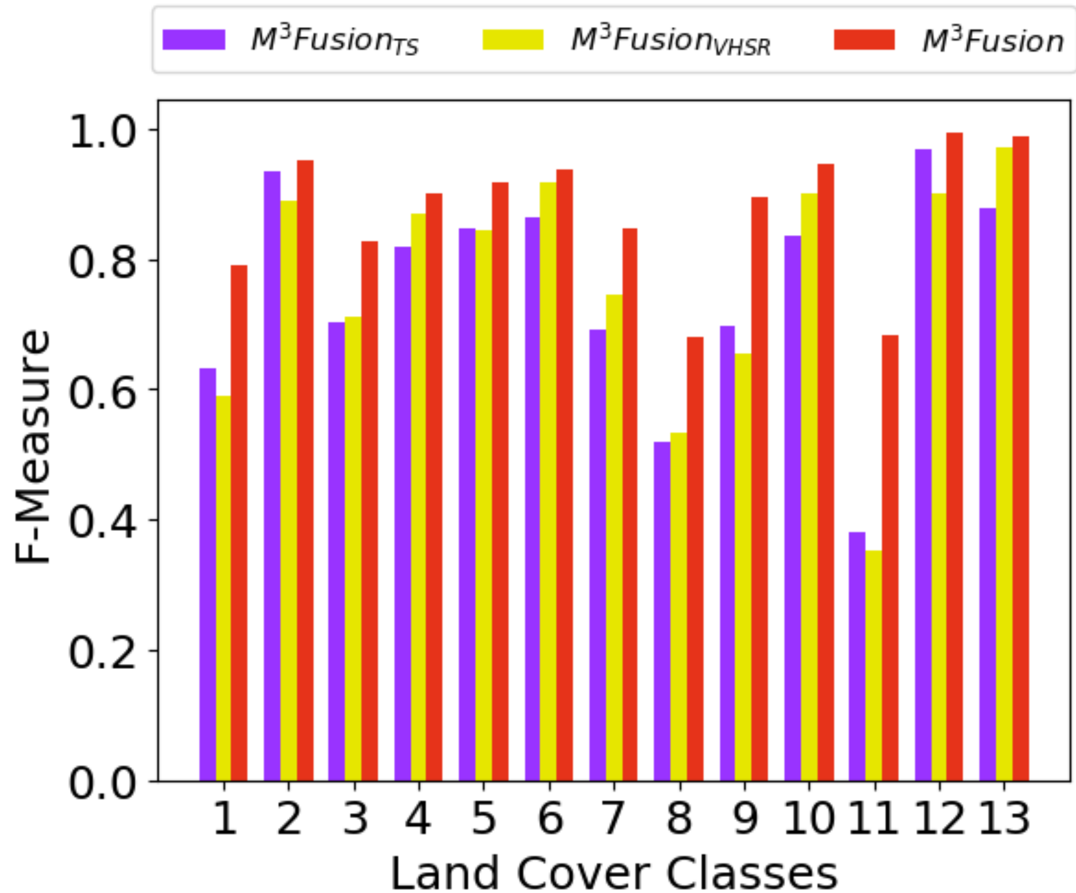
Class	Label
1	*Crop Cultivations*
2	*Sugar cane*
3	*Orchards*
4	*Forest plantations*
5	*Meadow*
6	*Forest*
7	*Shrubby savannah*
8	*Herbaceous savannah*
9	*Bare rocks*
10	*Urbanized areas*
11	*Greenhouse crops*
12	*Water Surfaces*
13	*Shadows*

Importance/complementary assessment of the information sources

	Accuracy	*F-Measure*	*Kappa*
RF_{TS}	0.8543	0.8519	0.8258
$M3F_{TS}$	0.8319	0.8325	0.8033
RF_{VHSR}	0.8237	0.8140	0.7908
$M3F_{VHSR}$	0.8369	0.8364	0.8677
RF	0.8716	0.8681	0.8491
M3F	**0.9149**	**0.9148**	**0.9000**

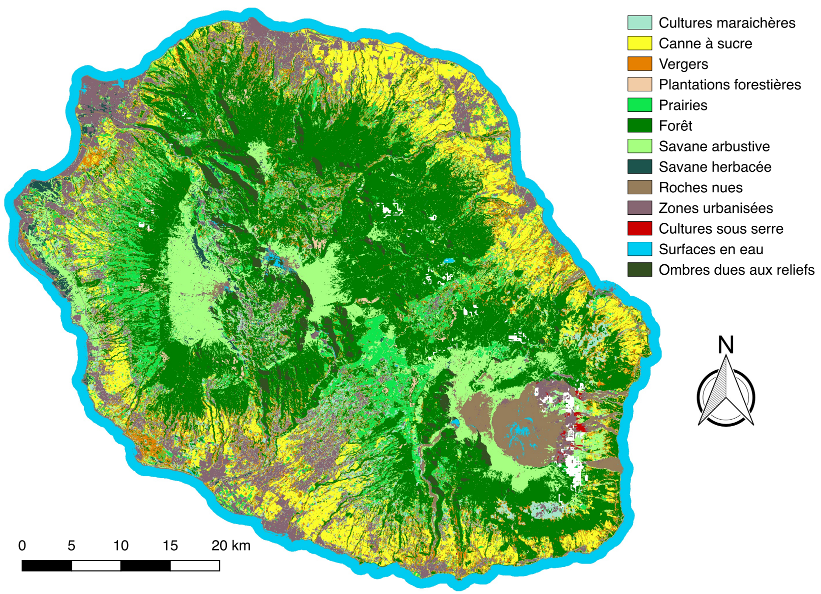
M3Fusion per-class Analysis





Class	Label
1	*Crop Cultivations*
2	*Sugar cane*
3	*Orchards*
4	*Forest plantations*
5	*Meadow*
6	*Forest*
7	*Shrubby savannah*
8	*Herbaceous savannah*
9	*Bare rocks*
10	*Urbanized areas*
11	*Greenhouse crops*
12	*Water Surfaces*
13	*Shadows*

Map Production via ML methods

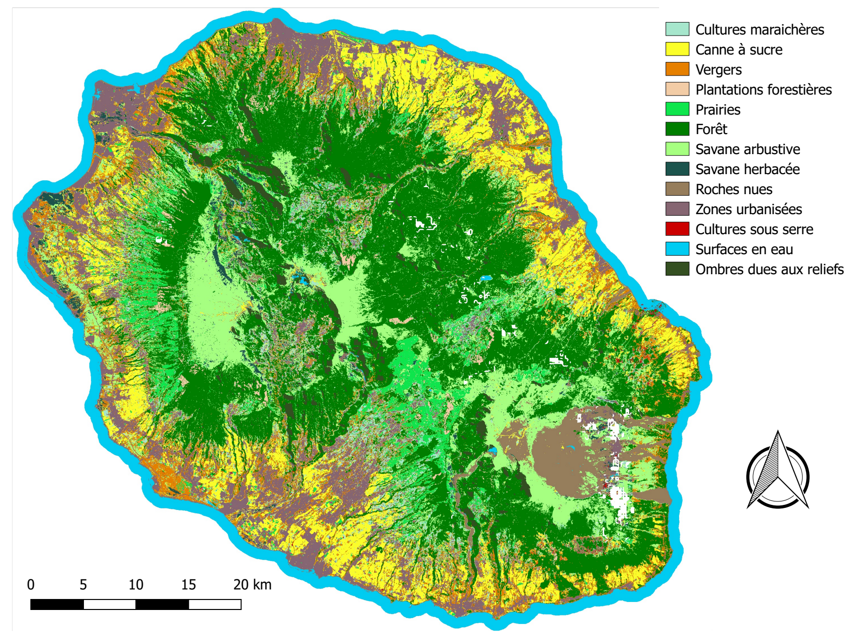


Random Forest Map

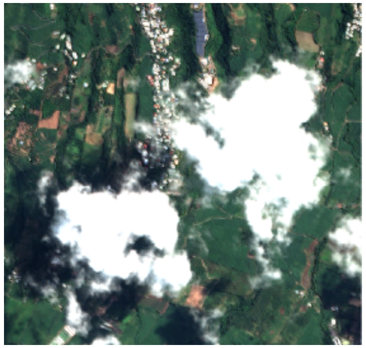
A Map involves around 27M pixels

- **M3Fusion takes about 15h to train (on training data)**
- **M3Fusion takes about 9h to produce the map on 27M**

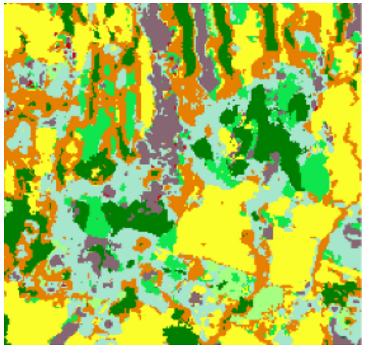
M3Fusion Map



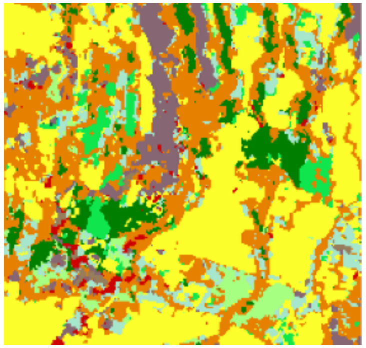
Map Details on some particular cases



RF

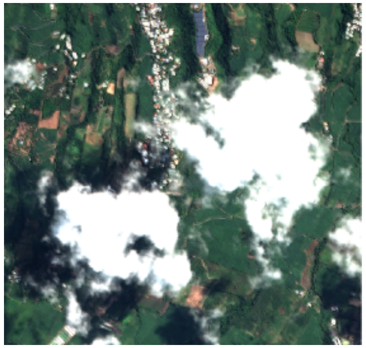


M3F

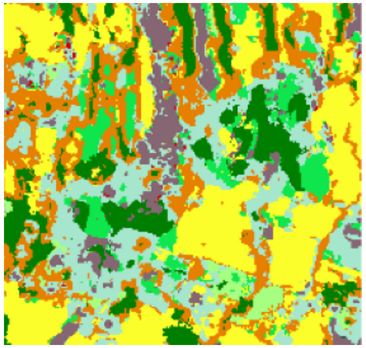


Detail on a Cloudy Zone

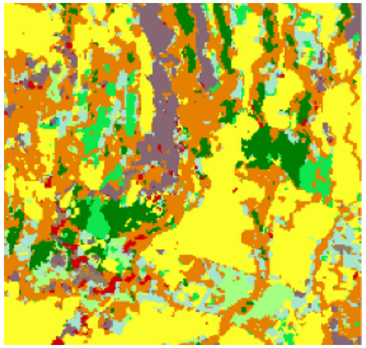
Map Details on some particular cases



RF

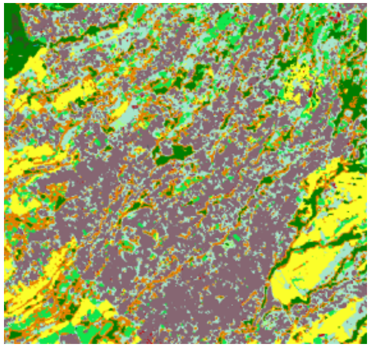


M3F

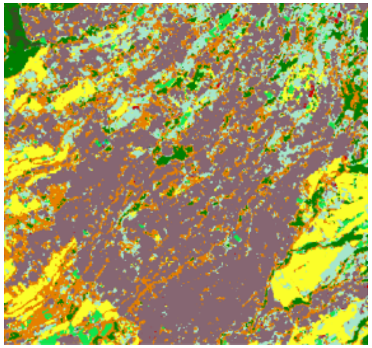


Detail on a Cloudy Zone

RF



M3F



Detail on Urban Zone

Conclusion and Future Works

A deep architecture to merge Multi-Scale and Multi-Temporal Data

To our knowledge, this is the first DL methods to make this kind of Data Fusion (S2/SPOT6/7)

Performance results underline the effectiveness of the proposed approach

Conclusion and Future Works

A deep architecture to merge Multi-Scale and Multi-Temporal Data

To our knowledge, this is the first DL methods to make this kind of Data Fusion (S2/SPOT6/7)

Performance results underline the effectiveness of the proposed approach

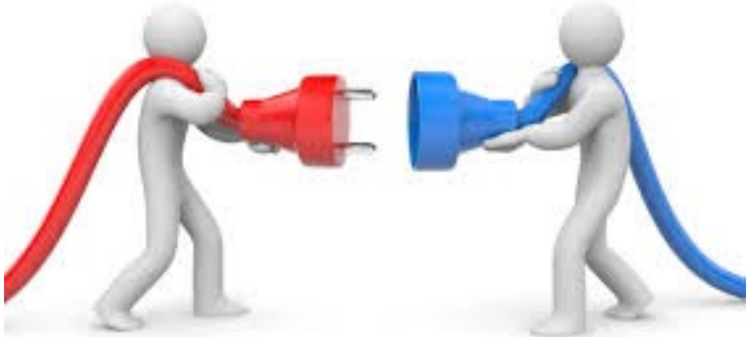
Future Works

Introduce Sentinel-1 time series data

Manage multi-resolution information without down(up)sampling (i.e. Sentinel 2 bands at 20m)

In depth study of the fusion process performed by our model

Thank you for your Attention



Thank you for your Attention

Questions