M3Fusion : Un modèle d’apprentissage profond pour la fusion de données satellites 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
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)
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?
Different Data fusion scenario [Schmitt16]:
- Single-Sensor Data Fusion
- Multiple-Sensor Data Fusion
- Temporal Data Fusion
- Machine Learning-Based Data Fusion
- And so on....

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

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)

M3F: Spatio-Temporal Data Fusion via Deep Learning

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

Sentinel-2 Time Series at High Spatial Resolution

25 x 25 patch extracted from Spot 6/7 VHSR image

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

\[ z_t = \sigma(W_{zz}x_t + W_{zh}h_{t-1} + b_z) \]
\[ r_t = \sigma(W_{rr}x_t + W_{rh}h_{t-1} + b_r) \]
\[ h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tanh(W_{hx}x_t + W_{hr}(r_t \odot h_{t-1}) + b_h) \]

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

We use DropOut to alleviate overfitting
M3F: Spatio-Temporal Data Fusion via Deep Learning

**GRU with Attention - Temporal Component**

\[
\begin{align*}
    z_t &= \sigma(W_{zx}x_t + W_{zh}h_{t-1} + b_z) \\
    r_t &= \sigma(W_{rx}x_t + W_{rh}h_{t-1} + b_r) \\
    h_t &= z_t \odot h_{t-1} + (1 - z_t) \odot \tanh(W_{hx}x_t + W_{hr}(r_t \odot h_{t-1}) + b_h)
\end{align*}
\]

We use DropOut to alleviate overfitting

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

**Attention Mechanism**
Combine the information extracted at each timestamps together

\[
\begin{align*}
    v_a &= \tanh(H \cdot W_a + b_a) \\
    \lambda &= \text{SoftMax}(v_a \cdot u_a) \\
    \text{rnn}_{feat} &= \sum_{i=1}^{N} \lambda_i \cdot h_{ti}
\end{align*}
\]
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

Loss to Optimise:

\[ L_{total} = \alpha_1 \ast L_{rnn}(rnn_{feat}, W_1, b_1) + \]
\[ = \alpha_2 \ast L_{cnn}(cnn_{feat}, W_2, b_2) + \]
\[ = L_{fus}([cnn_{feat}, rnn_{feat}], W_3, b_3) \]

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-Temporal information by Deep Learning

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-Temporal information by Deep Learning

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

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

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

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

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

Reference Data obtained by:
• RPG data (2014)
• GPS record (June 2017)
• Photo interpretation of VHSR Image

Spatial Distribution of Reference Data
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

<table>
<thead>
<tr>
<th>Class</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Crop Cultivations</td>
</tr>
<tr>
<td>2</td>
<td>Sugar cane</td>
</tr>
<tr>
<td>3</td>
<td>Orchards</td>
</tr>
<tr>
<td>4</td>
<td>Forest plantations</td>
</tr>
<tr>
<td>5</td>
<td>Meadow</td>
</tr>
<tr>
<td>6</td>
<td>Forest</td>
</tr>
<tr>
<td>7</td>
<td>Shrubby savannah</td>
</tr>
<tr>
<td>8</td>
<td>Herbaceous savannah</td>
</tr>
<tr>
<td>9</td>
<td>Bare rocks</td>
</tr>
<tr>
<td>10</td>
<td>Urbanized areas</td>
</tr>
<tr>
<td>11</td>
<td>Greenhouse crops</td>
</tr>
<tr>
<td>12</td>
<td>Water Surfaces</td>
</tr>
<tr>
<td>13</td>
<td>Shadows</td>
</tr>
</tbody>
</table>

F-Measure

Land Cover Classes

RF

M^3Fusion
Comparison with Standard ML method

<table>
<thead>
<tr>
<th>Class</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Crop Cultivations</td>
</tr>
<tr>
<td>2</td>
<td>Sugar cane</td>
</tr>
<tr>
<td>3</td>
<td>Orchards</td>
</tr>
<tr>
<td>4</td>
<td>Forest plantations</td>
</tr>
<tr>
<td>5</td>
<td>Meadow</td>
</tr>
<tr>
<td>6</td>
<td>Forest</td>
</tr>
<tr>
<td>7</td>
<td>Shrubby savannah</td>
</tr>
<tr>
<td>8</td>
<td>Herbaceous savannah</td>
</tr>
<tr>
<td>9</td>
<td>Bare rocks</td>
</tr>
<tr>
<td>10</td>
<td>Urbanized areas</td>
</tr>
<tr>
<td>11</td>
<td>Greenhouse crops</td>
</tr>
<tr>
<td>12</td>
<td>Water Surfaces</td>
</tr>
<tr>
<td>13</td>
<td>Shadows</td>
</tr>
</tbody>
</table>

M3Fusion

Random Forest
Importance/complementary assessment of the information sources

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>F-Measure</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RF_{TS}$</td>
<td>0.8543</td>
<td>0.8519</td>
<td>0.8258</td>
</tr>
<tr>
<td>$M3F_{TS}$</td>
<td>0.8319</td>
<td>0.8325</td>
<td>0.8033</td>
</tr>
<tr>
<td>$RF_{VHSR}$</td>
<td>0.8237</td>
<td>0.8140</td>
<td>0.7908</td>
</tr>
<tr>
<td>$M3F_{VHSR}$</td>
<td>0.8369</td>
<td>0.8364</td>
<td>0.8677</td>
</tr>
<tr>
<td>RF</td>
<td>0.8716</td>
<td>0.8681</td>
<td>0.8491</td>
</tr>
<tr>
<td>M3F</td>
<td>0.9149</td>
<td>0.9148</td>
<td>0.9000</td>
</tr>
</tbody>
</table>
M3Fusion per-class Analysis

<table>
<thead>
<tr>
<th>Class</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Crop Cultivations</td>
</tr>
<tr>
<td>2</td>
<td>Sugar cane</td>
</tr>
<tr>
<td>3</td>
<td>Orchards</td>
</tr>
<tr>
<td>4</td>
<td>Forest plantations</td>
</tr>
<tr>
<td>5</td>
<td>Meadow</td>
</tr>
<tr>
<td>6</td>
<td>Forest</td>
</tr>
<tr>
<td>7</td>
<td>Shrubby savannah</td>
</tr>
<tr>
<td>8</td>
<td>Herbaceous savannah</td>
</tr>
<tr>
<td>9</td>
<td>Bare rocks</td>
</tr>
<tr>
<td>10</td>
<td>Urbanized areas</td>
</tr>
<tr>
<td>11</td>
<td>Greenhouse crops</td>
</tr>
<tr>
<td>12</td>
<td>Water Surfaces</td>
</tr>
<tr>
<td>13</td>
<td>Shadows</td>
</tr>
</tbody>
</table>

![Graph showing F-Measure for M3Fusion compared to other methods across land cover classes.](image-url)
Map Production via ML methods

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 pixels
Map Details on some particular cases

Detail on a Cloudy Zone
Map Details on some particular cases

Detail on a Cloudy Zone

Detail on Urban Zone
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