Classification of spatially enriched pixel time series with convolutional neural networks

Mohamed Chelali*, Camille Kurtz*, Anne Puissant+ and Nicole Vincent*

* Université de Paris
+ Université de Strasbourg

ICPR – January 10-15, 2020
Introduction: Context and objectives

- Satellite Image Time Series (SITS): $2D + t$ data
- Thematic usage of SITS
  - Land use mapping $^1$
  - Territories management $^2$
  - Change detection

Agricultural crop-field classification $^3$

Focus on spatial and temporal domains during analysis

---

$^1$ J. Inglada et al. Operational high resolution land cover map production at the country scale using satellite image time series. RS 2017.


Deep neural network for SITS analysis

- Temporal pixel analysis
  - Deep neural network methods: LSTM, TempCNN
  - Recurrence plot
- Hybrid strategy
  - 2D CNN + 1D CNN
- Image sequences analysis
  - 3D CNN

---


---

M Chelali et al — Classification of spatially enriched pixel time series with convolutional neural networks
Strategy:

- Enriching temporal pixels with spatial information by using 1D segments leading to a planar representation
- A 2 step approach
  - Multi-segments extraction in the 2D domain
    - Definition of temporal segments
    - Classification of the segments
  - Decision process at crop-field level

Data representation:

- Segment construction is based on Random Walk method
Proposed approach: Enriching temporal pixels with spatial information

Strategy:
- Enriching temporal pixels with spatial information by using 1D segments leading to a planar representation
- A 2 step approach
  - Multi-segments extraction in the 2D domain
    - Definition of temporal segments
    - Classification of the segments
  - Decision process at crop-field level

Segment labeling:
Proposed approach: Enriching temporal pixels with spatial information

Strategy:
- Enriching temporal pixels with spatial information by using 1D segments leading to a planar representation
- A 2 step approach
  - Multi-segments extraction in the 2D domain
    - Definition of temporal segments
    - Classification of the segments
  - Decision process at crop-field level

Decision process:

Decision making → A

Aggregation of decisions

Thematic Classes of ROI

Orchard
Meadow
Water
Experimental study: Agricultural crop-field classification from SITS

- Satellite Image Time Series from SENTINEL-2 in year 2017
  - 50 images; Tile 32ULU (Zone of Grand Est - Alsace); Correction at level 2A
- Ground truth: European delineation of agricultural crop-fields

Proposed method scores

<table>
<thead>
<tr>
<th>Lengths $L$ of the segments</th>
<th>percentage</th>
<th>OA</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random pixels(100)</td>
<td>10</td>
<td>91.97</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>91.72</td>
<td>2.72</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>92.29</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>91.75</td>
<td>0.61</td>
</tr>
<tr>
<td>RW(50)</td>
<td>10</td>
<td>91.07</td>
<td>2.53</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>93.80</td>
<td>1.57</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>94.06</td>
<td>1.44</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>94.80</td>
<td>1.57</td>
</tr>
<tr>
<td>RW(100)</td>
<td>10</td>
<td>92.50</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>93.20</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>94.21</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>94.64</td>
<td>0.80</td>
</tr>
</tbody>
</table>

TempCNN scores

<table>
<thead>
<tr>
<th>Nb filt.</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
<th>256</th>
<th>512</th>
<th>1024</th>
</tr>
</thead>
<tbody>
<tr>
<td>OA</td>
<td>92.18</td>
<td>92.82</td>
<td>92.77</td>
<td>92.82</td>
<td>92.12</td>
<td>92.98</td>
<td>92.77</td>
</tr>
<tr>
<td>STD</td>
<td>0.62</td>
<td>0.59</td>
<td>0.85</td>
<td>0.68</td>
<td>1.22</td>
<td>0.89</td>
<td>1.01</td>
</tr>
</tbody>
</table>
Attention mechanism for visual explanations

Strategy:

- Use of GradCAM++\(^1\) to get the most used region of the planar representation
- Temporal Attention: Choose a significant temporal range
- Spatial Attention: Explaining the usefulness of the spatial space

\(^1\)A. Chattopadhyay, A. Sarkar, P. Howlader, V. N. Balasubramanian, Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks, in: WACV, Procs., 2018
Attention mechanism: Temporal attention

Method
1 - Capture the temporal attention, thanks to the thresholding process
2 - Restriction of the temporal domain defines masks used to classify data

<table>
<thead>
<tr>
<th>New temporal range [0; 120]</th>
</tr>
</thead>
<tbody>
<tr>
<td>224 dates</td>
</tr>
<tr>
<td>OA</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>RW(100)</td>
</tr>
</tbody>
</table>

(d) Vineyard  (e) Meadow  (f) Int. orchard  (g) trad. orchard

(h) Vineyard  (i) Meadow  (j) Int. orchard  (k) trad. orchard
Attention mechanism: Spatial attention

1. Capture the maximum spatial attention of the pixel $p$.
2. Attribute the maximum spatial attention of $p$ from $N_{seg}$ planar representations to $p$ position.
3. Create a semantic map by attributing a color for each maximal attention of $C$ classes.

Vineyards  Meadows  Traditional orchards  Intensive orchards
Conclusion and perspectives

Conclusion

- Proposition of a spatio-temporal representation of image time series
  - Use of a 2D CNN to learn spatio-temporal features
  - Use of a pretrained model, e.g. trained on ImageNet dataset
- Proposition of an attention mechanism
  - To analyze the temporal domain
  - To generate a semantic segmentation

Perspectives

- Integrate the attention during the training
- Involve the framework in another domain, such as gait recognition in video.
Classification of spatially enriched pixel time series with convolutional neural networks

Mohamed Chelali*, Camille Kurtz*, Anne Puissant+ and Nicole Vincent*

* Université de Paris
+ Université de Strasbourg

ICPR – January 10-15, 2020