Classification of spatially enriched pixel time series with convolutional neural networks

Mohamed Chelali[1], Camille Kurtz[1], Anne Puissant[2], Nicole Vincent[1]
[1]Université de Paris, LIPADE, France
[2]Université de Strasbourg, LIVE, France

Objectives

- Spatio-temporal features extraction from image time series such as Satellite Image Time Series (SITS)
- Agricultural crop-fields classification based on SITS

Contributions

- Planar spatio-temporal data representation : from $2D + t$ to $2D$
- Use of a classical $2D$ Convolutionnal Neural Network (CNN) to extract spatio-temporal features from SITS
- Aattention mechanism for visual explanation

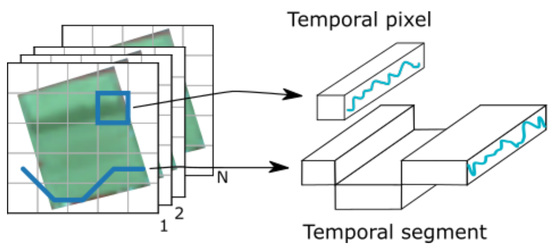
Enriching temporal pixels with spatial information

Strategy:

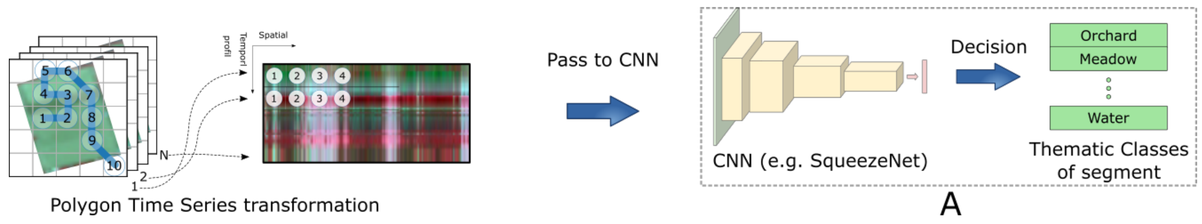
- Enriching temporal pixels with spatial information by using 1D segments leading to a planar representation
- A 2-step approach
  - Multi-segments extraction in the 2D domain
    - Definition of temporal segments
    - CNN-based classification
  - Decision process at crop-field level

Data representation:

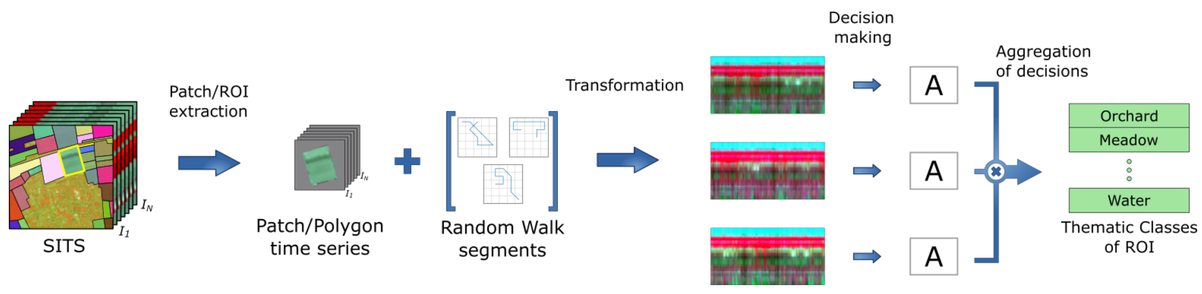
- Segment construction is based on Random Walk method



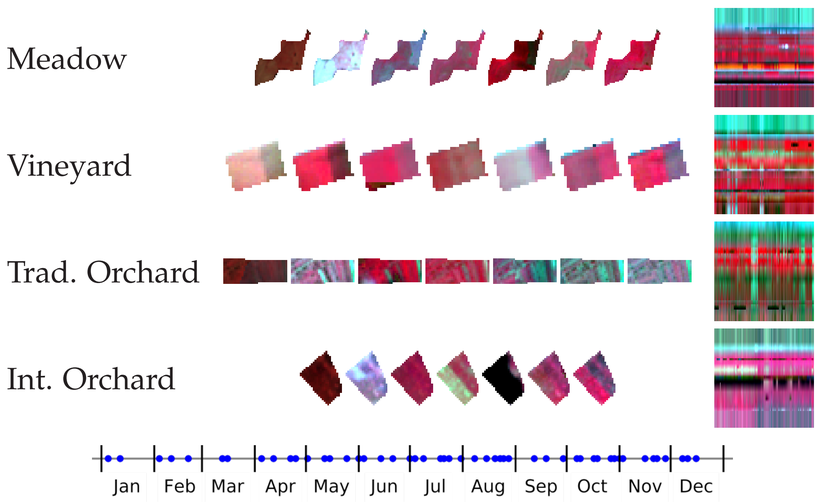
Segment labeling:



Decision process:



Classification of agricultural crop-fields



- 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

Lengths L of the segments	Number of segments used for decision*	OA	STD
Random pixels(100)	10	91.97	1.11
	20	91.72	2.72
	50	92.29	0.60
	70	91.75	0.61
RW(50)	10	91.07	2.53
	20	93.80	1.57
	50	**94.06**	**1.44**
	70	**94.80**	**1.57**
RW(100)	10	92.50	1.05
	20	93.20	0.65
	50	**94.21**	**1.19**
	70	**94.64**	**0.80**
best method of the state of the art – TempCNN[1]		**92.98**	**0.89**

** Number of segments used for decision expressed as a proportion of the area of the crop field*

[1] *C. Pelletier. et al. Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. RS, 2019*

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.

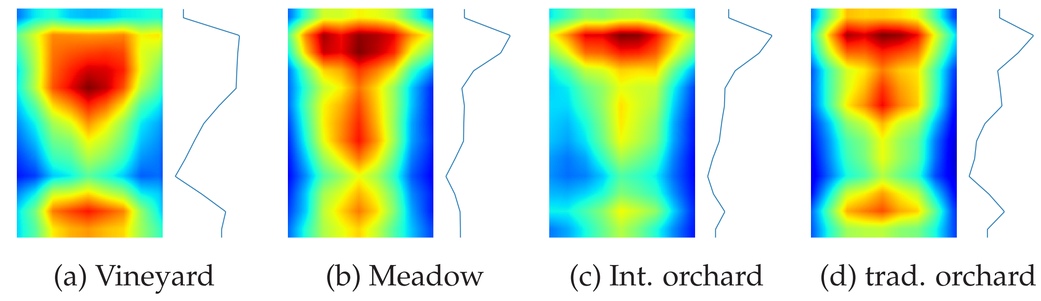
Attention mechanism for visual explanations

Strategy:

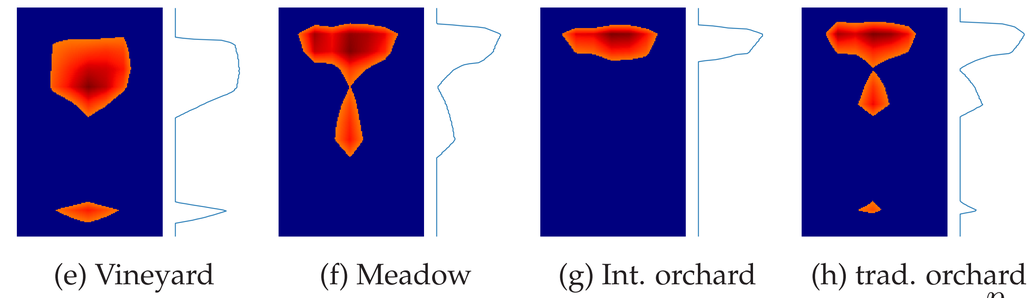
- Use of GradCAM++ to get the most interested region of the planar representation

Temporal attention: Choose a significant temporal range

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



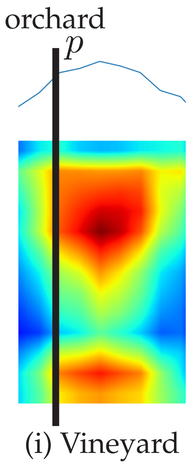
(a) Vineyard (b) Meadow (c) Int. orchard (d) trad. orchard



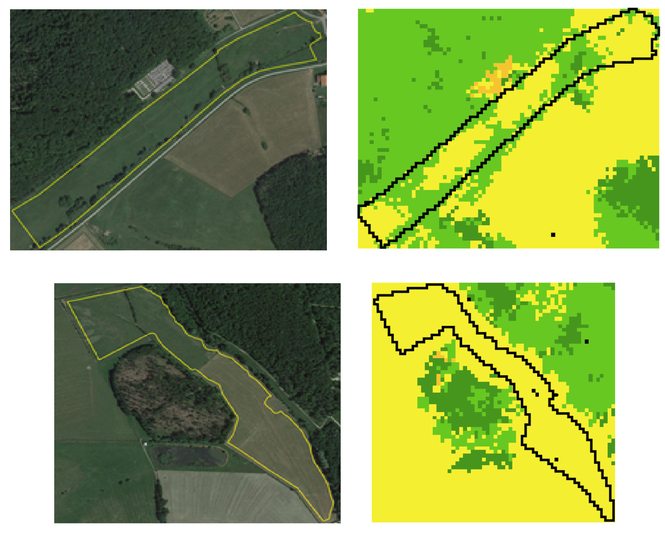
(e) Vineyard (f) Meadow (g) Int. orchard (h) trad. orchard

Spatial attention: Explaining the usefulness of the spatial space

- 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



(i) Vineyard



The French ANR supported this work under Grant ANT-17-CE23-0015