Classification of spatially enriched pixel time series with convolutional neural networks

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

[*] Université de Paris

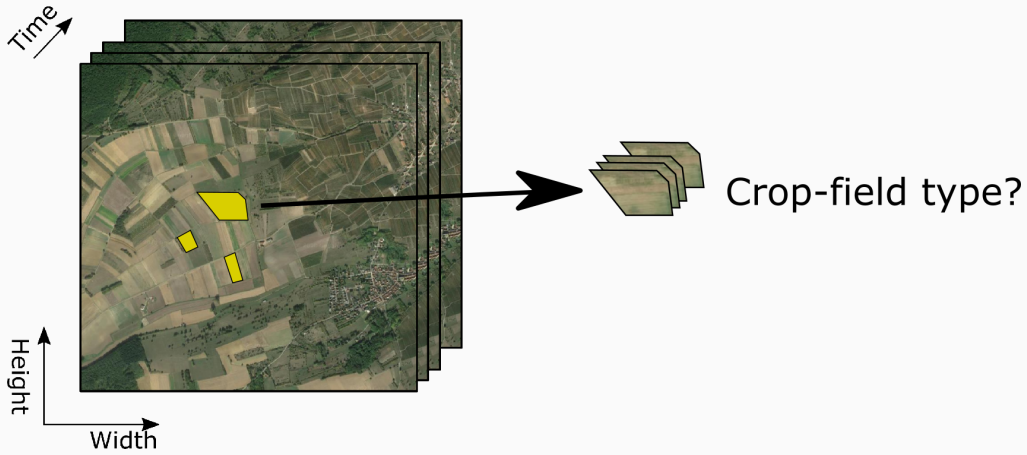
[+] Université de Strasbourg

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

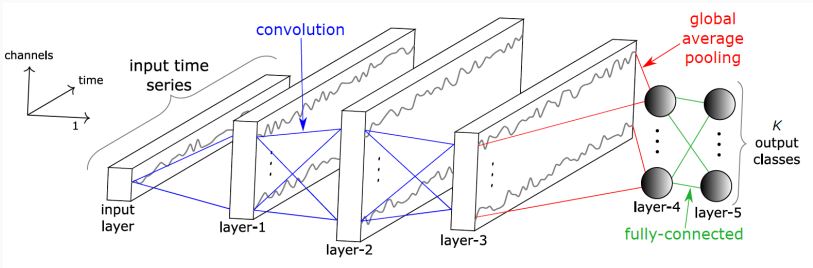
[1] J. Inglada et al. Operational high resolution land cover map production at the country scale using satellite image time series. RS 2017.

[2] C. Weber and A. Puissant. Urbanization pressure and modeling of urban growth: example of the tunis metropolitan area. RSE 2003.

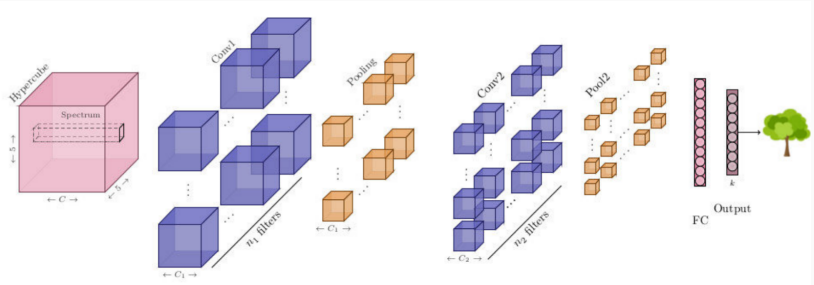
[3] M. Chelali et al. Image time series classification based on a planar spatio-temporal data representation, VISAPP 2020.

Deep neural network for SITS analysis

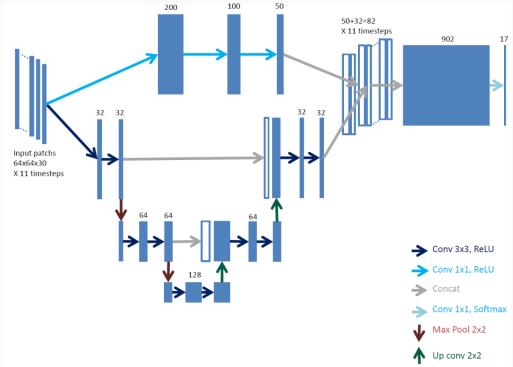
- Temporal pixel analysis
  - Deep neural network methods: LSTM, TempCNN
  - Recurrence plot
- Hybrid strategy
  - 2D CNN + 1D CNN
- Image sequences analysis
  - 3D CNN



(a) TempCNN[1]



(b) Convolution 3D[2]



(c) Hybride UNet[3]

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

[2] N. Audebert, et al. Deep learning for classification of hyperspectral data: A comparative review. IGRSL, 2019.

[3] D. Ienco, et al. Land cover classification via multitemporal spatial data by deep recurrent neural networks. IGRSL, 2017.

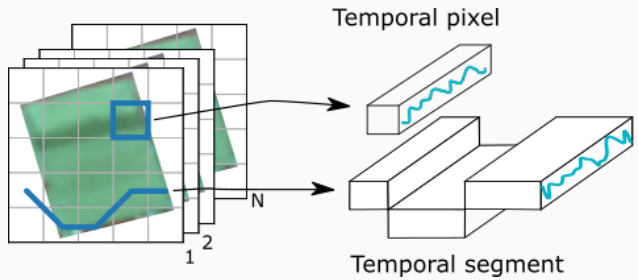
Proposed approach: Enriching temporal pixels with spatial information

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

Data representation:

- Segment construction is based on Random Walk method

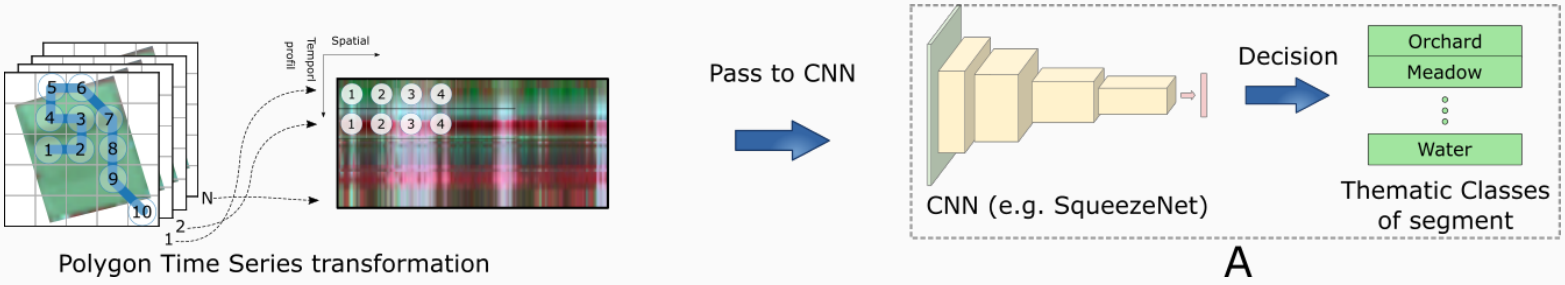


Proposed approach: Enriching temporal pixels with spatial information

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:

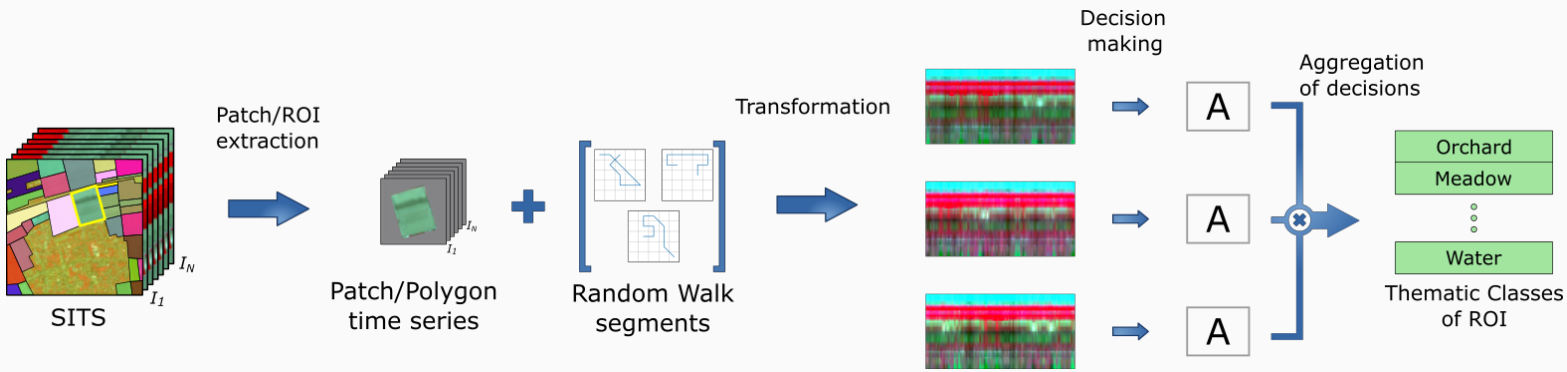


Proposed approach: Enriching temporal pixels with spatial information

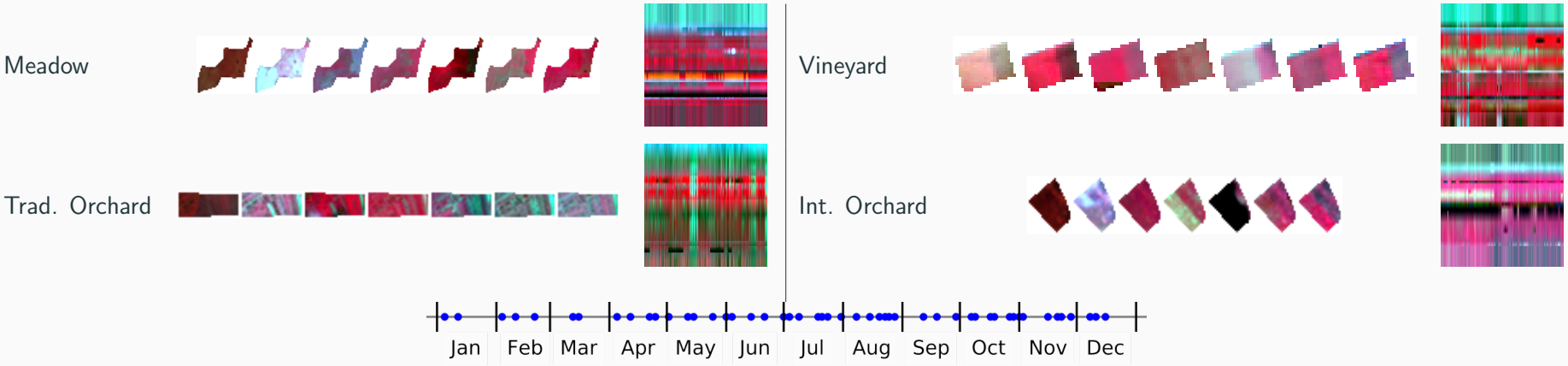
Strategy:

- 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:



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

Lengths L of the segments	percentage	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**

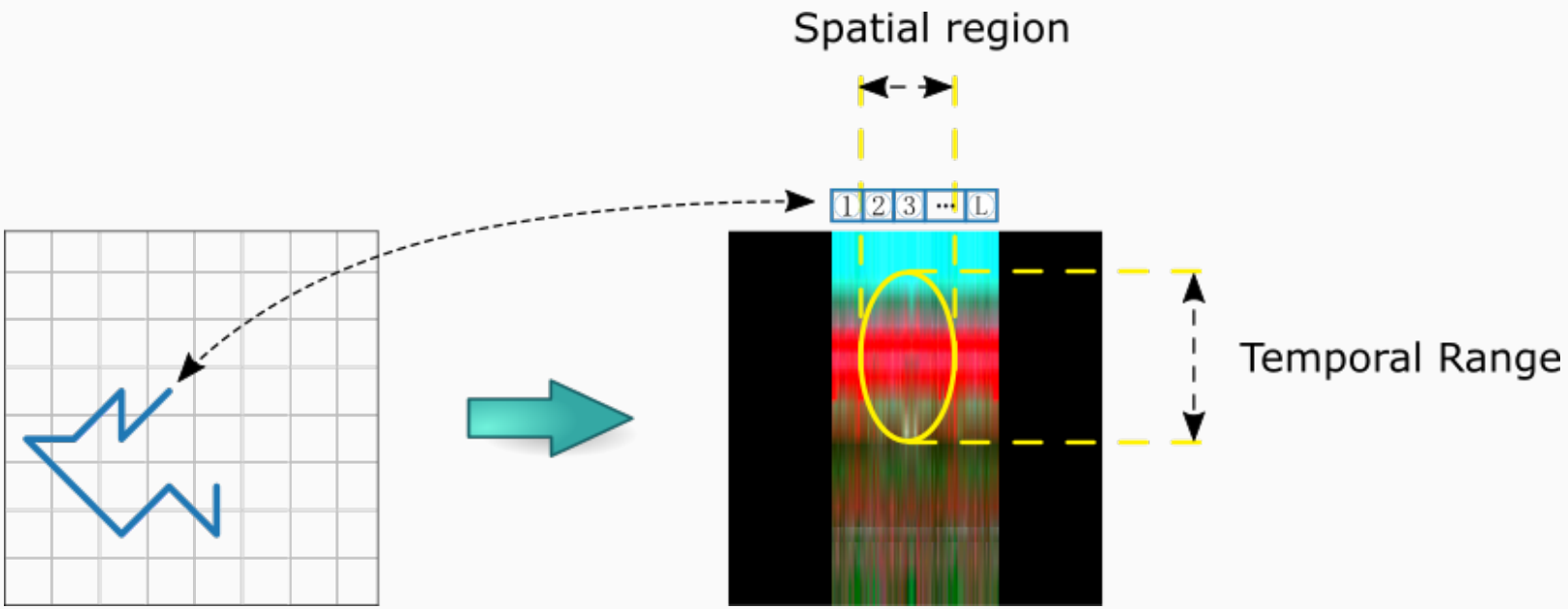
TempCNN scores

Nb filt.	16	32	64	128	256	512	1024
OA	92.18	92.82	92.77	92.82	92.12	**92.98**	92.77
STD	0.62	0.59	0.85	0.68	1.22	**0.89**	1.01

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

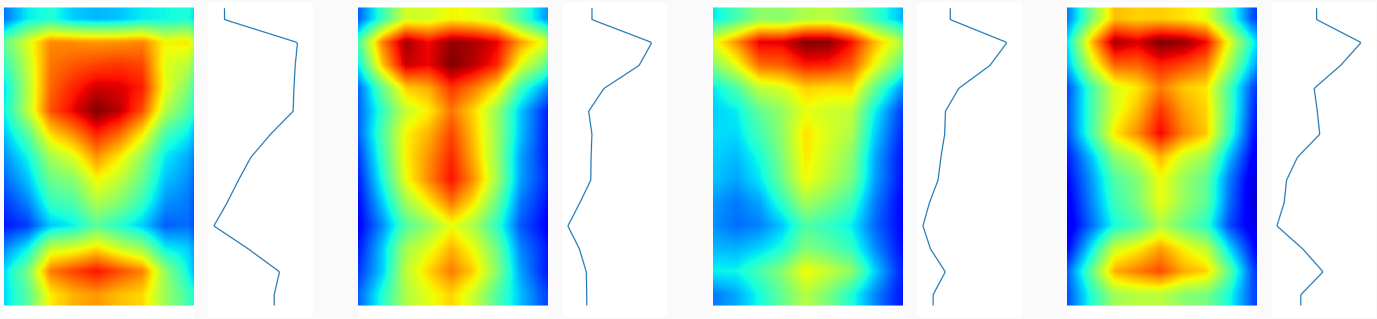
Method

1 - Capture the temporal attention, thanks to the thresholding process

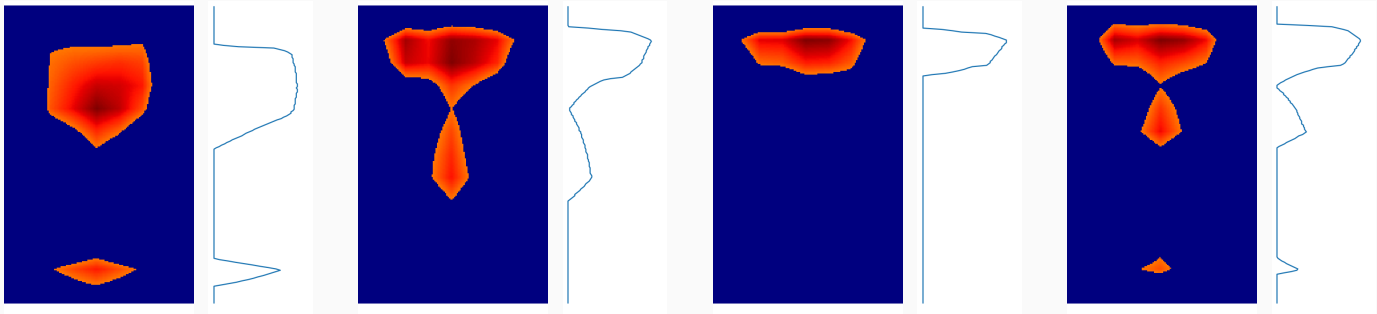
2 - Restriction of the temporal domain defines masks used to classify data

New temporal range [0; 120]

	224 dates		time restriction to [0,120]	
	OA	STD	OA	STD
RW(100)	93.00	2.44	94.0	2.54

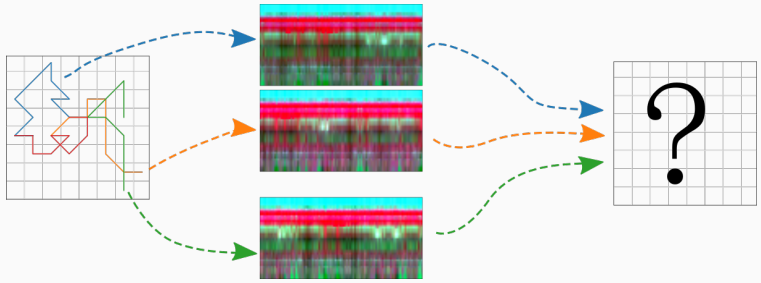


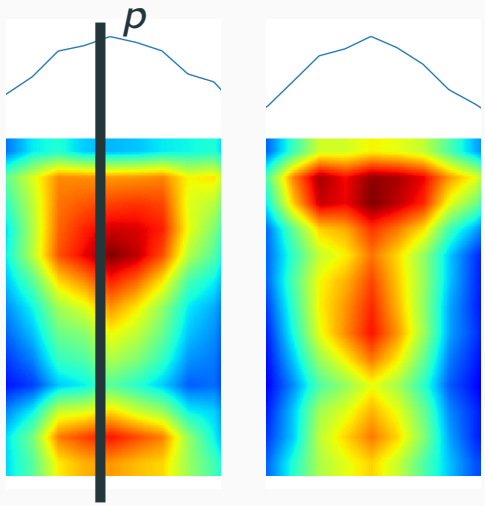
(d) Vineyard **(e)** Meadow **(f)** Int. orchard **(g)** trad. orchard



(h) Vineyard **(i)** Meadow **(j)** Int. orchard **(k)** trad. orchard

Attention mechanism: Spatial attention



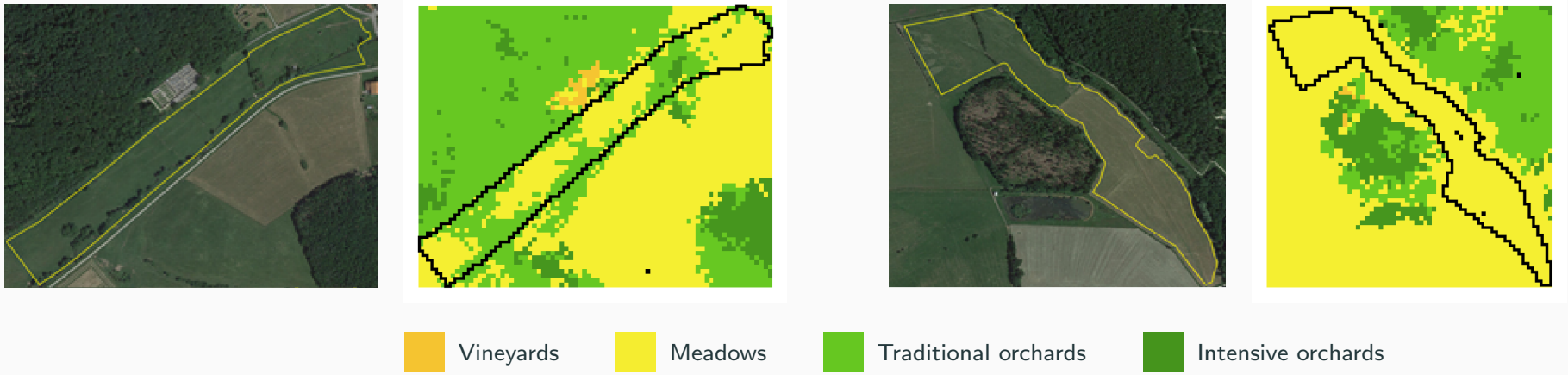


(l) Vineyard **(m)** Meadow

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



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