Point process and CNN for small object detection in satellite images

Jules Mabon, Inria, Université Côte d'Azur
In collaboration with **Mathias Ortner** (Airbus DS) and **Josiane Zerubia** (Inria)

Journées Geosto-MIA, 2022

01

Introduction

Goals

- Detection and vectorization of objects in satellite images

Challenges

- small sized objects at 50 cm/pixel
- Visually diverse environment and objects
- Variable density
- Priors on interactions

Image from the DOTA[1]dataset

[1]Xia *et al.*, "DOTA: A Large-Scale Dataset for Object Detection in Aerial Images," 2018.

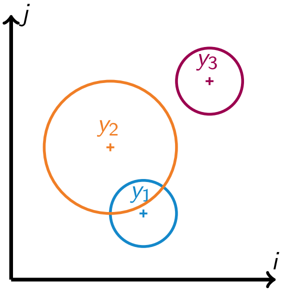
02

Point process: definition

Marked point process

- configuration of points $Y = \{y_0, \dots, y_n\}$
- $y \in S \times M$, $S \subset \mathbb{R}^2$ the image space, M the mark space
- $Y \in \mathcal{P}$, $\mathcal{P} = \bigcup_{n=0}^{\infty} (S \times M)^n$
- Y : realization of a random variable in $\mathcal{P}$

Configuration $\{y_1, y_2, y_3\}$

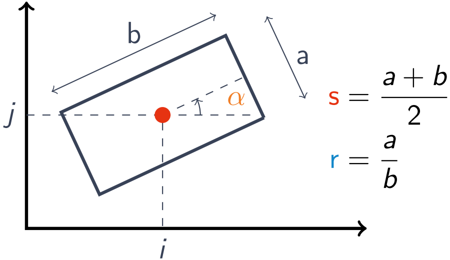


	i	j	r
y_1 : (	2	1	0.5)
y_2 : (	1.5	2	1)
y_3 : (	3	3	0.5)

Parametrization

Oriented rectangle

- $M = R^{+} \times [0, 1] \times [0, \pi]$
- $y = (y_i, y_j, y_s, y_r, y_\alpha)$



Point process density

Y is the realization of a random variable of density h

Point process density (Gibbs)

$$h(Y|X) \propto \exp(-U(Y,X)) \tag{1}$$

Point process measure

Measure of the point process,

$$\nu(B) = \int_B h(Y|X)\mu(dY) \tag{2}$$

$\mu(\cdot)$ the measure of the Poisson point process, $B \subset S$

Point process density

Y is the realization of a random variable of density h

Point process density (Gibbs)

$$h(Y|X) \propto \exp(-U(Y,X)) \tag{1}$$

Building the energy model

We want U such that

$$Y^+ = \underset{Y \in \mathcal{P}}{\operatorname{argmin}}\, U(Y,X) \tag{2}$$

X: image, Y^+: ground truth for image X, $\mathcal{N}_y$: neighborhood of y in Y

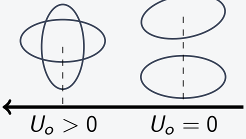
Basic energy model

$$U(Y,X) = \sum_{y \in Y} U_{data}(y,X) + U_{prior}(y, \mathcal{N}y)$$

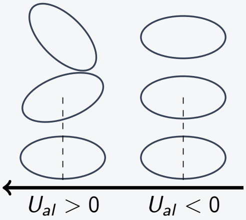
Basic energy model

$$U(Y,X) = \sum_{y \in Y} U_{data}(y,X) + U_{prior}(y,\mathcal{N}_y)$$

Prior term



Overlapping



Alignment

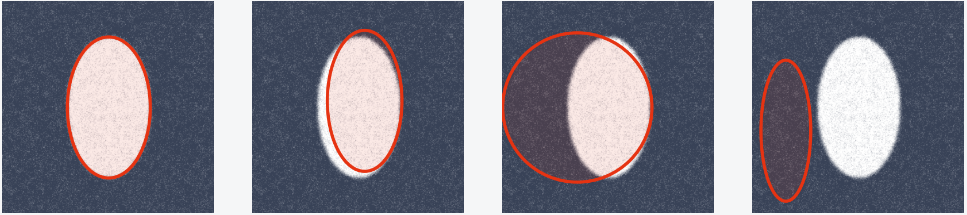
Or others:

- Size
- Shape
- Dynamics
- ...

Basic energy model

$$U(Y,X) = \sum_{y \in Y} U_{data}(y,X) + U_{prior}(y,\mathcal{N}y)$$

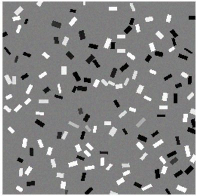
Data term

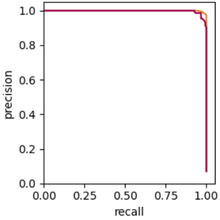


$U_{data} < 0$ $\quad$ $U_{data} \simeq 0$ $\quad$ $U_{data} > 0$ $\quad$ $U_{data} \gg 0$

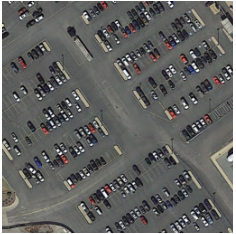
Contrast measures as data term

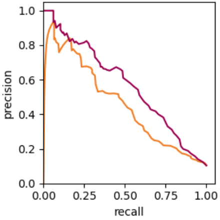
Synthetic data





DOTA sample





Precision-recall for the T-test measure[2] (red) and for a measure based on image gradient[3] (orange).

---

[2] Lacoste *et al.*, "Point processes for unsupervised line network extraction in remote sensing," 2005.
[3] Kulikova *et al.*, "Extraction of Arbitrarily-Shaped Objects Using Stochastic Multiple Birth-and-Death Dynamics and Active Contours," 2010.

Ayana Inria

03

Data term from a CNN

$$U(Y,X) = \sum_{y \in Y} U_{data}(y,X) + U_{prior}(y, \mathcal{N}y)$$

Data term

$$U_{data}(y,X) = U_{pos}(y,X) + \sum_{k \in \{s,r,\alpha\}} U_k(y,X)$$

Inferring the energy map[4]

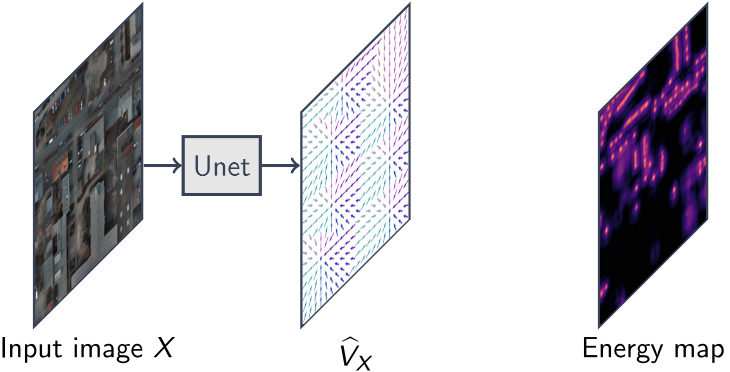
Input image X

[4] Mabon *et al.*, "Processus ponctuels marqués et réseaux de neurones convolutifs pour la détection d'objets dans des images de télédétection," 2022.

Inferring the energy map[4]

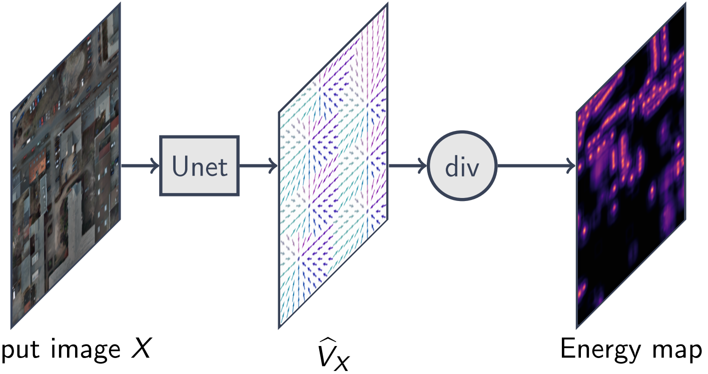
[4] Mabon *et al.*, "Processus ponctuels marqués et réseaux de neurones convolutifs pour la détection d'objets dans des images de télédétection," 2022.

Inferring the energy map[4]



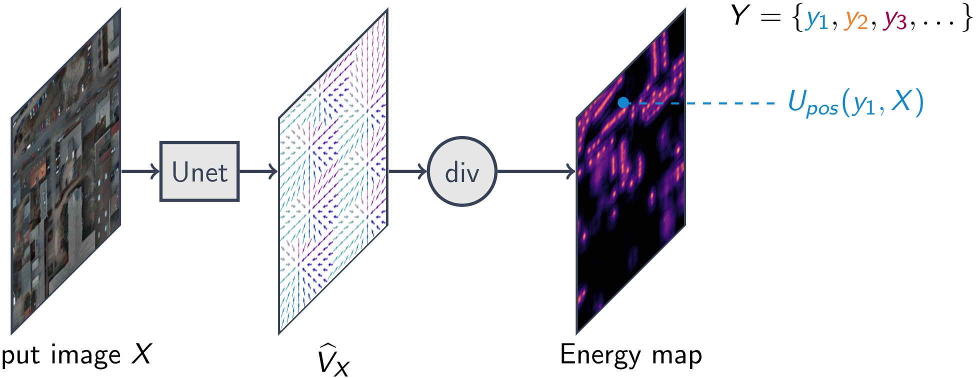
[4]Mabon *et al.*, "Processus ponctuels marqués et réseaux de neurones convolutifs pour la détection d'objets dans des images de télédétection," 2022.

Inferring the energy map[4]



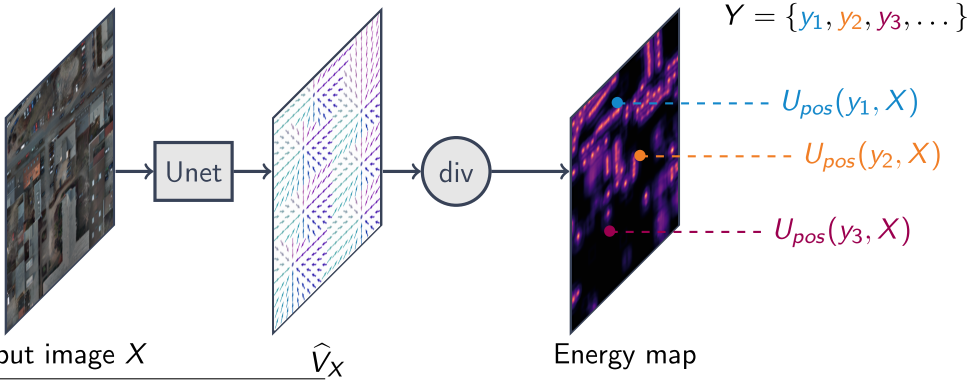
[4]Mabon *et al.*, "Processus ponctuels marqués et réseaux de neurones convolutifs pour la détection d'objets dans des images de télédétection," 2022.

Inferring the energy map[4]



[4] Mabon *et al.*, "Processus ponctuels marqués et réseaux de neurones convolutifs pour la détection d'objets dans des images de télédétection," 2022.

Inferring the energy map[4]



[4]Mabon *et al.*, "Processus ponctuels marqués et réseaux de neurones convolutifs pour la détection d'objets dans des images de télédétection," 2022.

Position energy term

Position energy term

$$U_{pos}(y, X) = -\mathrm{div}(\widehat{V}_X)[y_i, y_j] \tag{3}$$

- Inference of a vector map $\widehat{V}_X$ to better separate instances
- $A[y_i, y_j]$: interpolated value of map A at position $[y_i, y_j]$
- $a, b \in \mathbb{R}$, energy model parameters (learned)
- $\widehat{V}_X$ pre-computed : $U_{pos}(y, X)$ defined $\forall y \in S \times M$

Learning the position energy

Using labeled data X, Y^+ we minimize

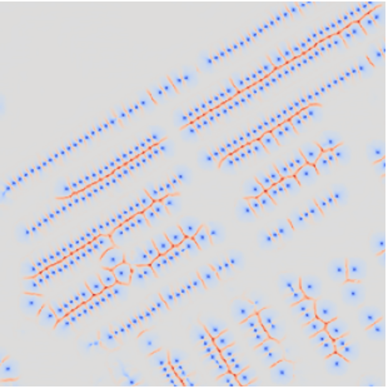
Cost function

$$\mathcal{L}_{pos}(\widehat{V}_X, Y^+) = \mathrm{MSE}(\widehat{V}_X, V_{Y^+}) \tag{4}$$

- MSE : Mean Squared Error
- V_{Y^+} : Vector field built from the ground truth Y_{GT}
- Data augmentation: patch sampling, rotation, hue/contrast/luminosity variations etc.

Position energy term: example

Image X



$\mathrm{div}(\widehat{V}_X)$
blue<0, red>0

Marks energy term: energy tensor[5]

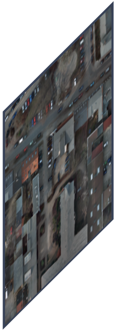
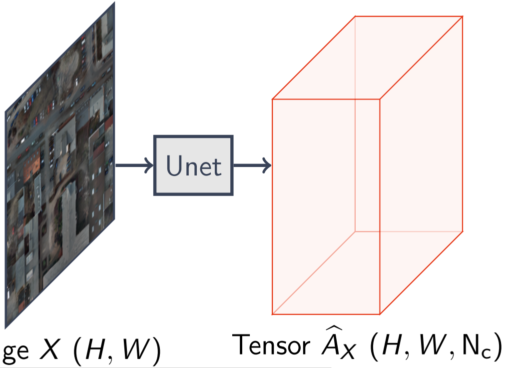


Image X (H, W)

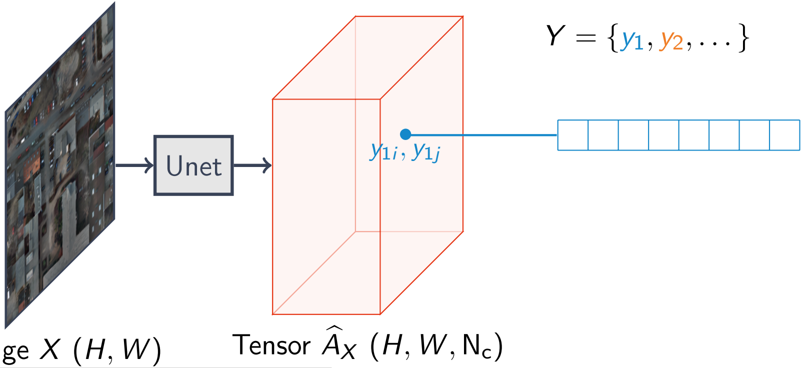
[5]Mabon *et al.*, "Processus ponctuels marqués et réseaux de neurones convolutifs pour la détection d'objets dans des images de télédétection," 2022.

Marks energy term: energy tensor[5]



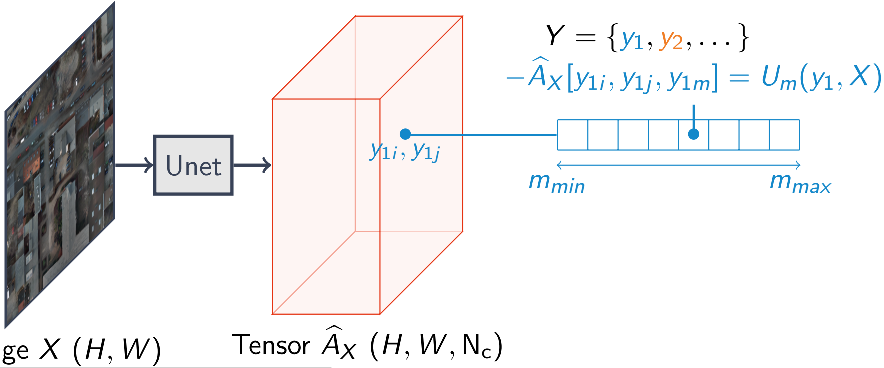
[5] Mabon *et al.*, "Processus ponctuels marqués et réseaux de neurones convolutifs pour la détection d'objets dans des images de télédétection," 2022.

Marks energy term: energy tensor[5]



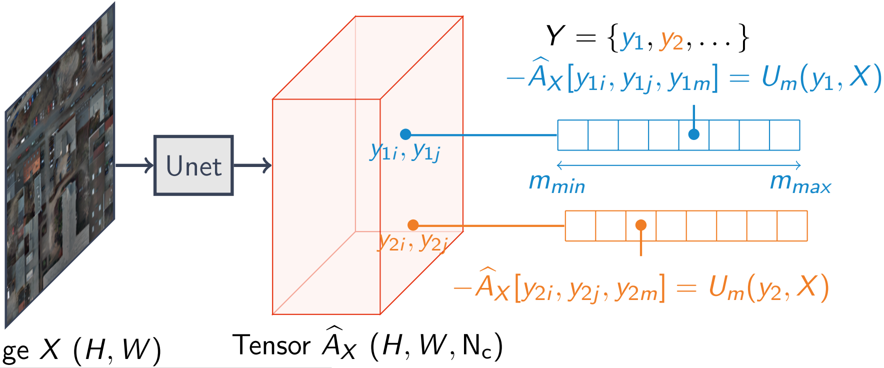
[5] Mabon *et al.*, "Processus ponctuels marqués et réseaux de neurones convolutifs pour la détection d'objets dans des images de télédétection," 2022.

Marks energy term: energy tensor[5]



[5] Mabon *et al.*, "Processus ponctuels marqués et réseaux de neurones convolutifs pour la détection d'objets dans des images de télédétection," 2022.

Marks energy term: energy tensor[5]



[5] Mabon *et al.*, "Processus ponctuels marqués et réseaux de neurones convolutifs pour la détection d'objets dans des images de télédétection," 2022.

Marks energy term

Energy on marks

For each mark m

$$U_m(y, X) = -\widehat{A}_X[y_i, y_j \text{remap}(y_m)] \tag{5}$$

- $A[i, j, k]$: interpolated value of tensor A at coordinates $[i, j, k]$
- $\text{index}(y_m)$: index corresponding to the discretization of values from m in N_c intervals from m_{min} to m_{max}
- $\widehat{A}_X$ pre-computed : $U_m(y, X)$ defined $\forall y \in S \times M$
- $\text{remap}(y_m) = N_c \frac{y_m - m_{min}}{m_{max} - m_{min}}$

Learning the mark energy term

Cost function

$$\mathcal{L}_m(\widehat{A}_X, Y_{GT}) = \frac{1}{|P|}\sum_{p\in P} \mathrm{CE}(\mathrm{Softmax}(\widehat{A}_X[p]), \mathrm{Softmax}(A_{Y_{GT}}[p])) \quad (6)$$

- CE : Cross Entropy, P : set of all pixels from X, $A[p]$: value of tensor A at position p
- $A_{Y_{GT}}$ tensor built from the ground truth Y_{GT}
- Data augmentation
- Y augmented during training : we offset marks randomly with a normal law

Ayana Inria

Priors on configurations

No-overlap prior

$$U_o(y, \mathcal{N}_y) = \max_{\tilde{y} \in \mathcal{N}_y} \left\{ \frac{\mathrm{area}(\tilde{y} \cap y)}{\min\{\mathrm{area}(\tilde{y}), \mathrm{area}(y)\}} \right\} \tag{7}$$

Alignment prior

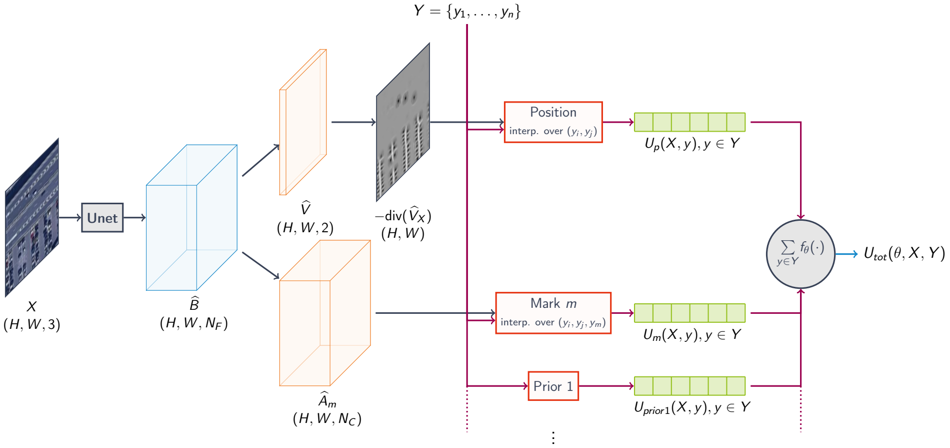
$$U_a(y, \mathcal{N}_y) = \min_{\tilde{y} \in \mathcal{N}_y} \{-|\cos(|y_\alpha - \tilde{y}_\alpha|)|\} \tag{8}$$

Size prior

$$U_s(y) = \max\{s_{min} - \mathrm{area}(y), \mathrm{area}(y) - s_{max}, 0\} \tag{9}$$

hyperparameters $s_{min}, s_{max} \in \mathbb{R}^+$, minimal and maximal sizes

Energy model



Ayana Inria

Energies combination

Multiple energy terms

Energy	Notation
position	$U_{pos}(y, X)$
marks $m \in \{s, r, \alpha\}$	$U_m(y, X)$
size	$U_s(y)$
superposition	$U_s(y, \mathcal{N}_y)$
alignment	$U_a(y, \mathcal{N}_y)$

Total energy

$$U_{tot} = \sum_{y \in Y} f_\theta \left(U_1, \dots, U_k \right)$$

With $f_\theta : \mathbb{R}^k \to \mathbb{R}$
How to combine energy terms ?

Linear combination

Total energy

$$U_{tot}(Y, X, \theta) = \sum_{y \in Y} \theta_1 U_1(y, X) + \cdots + \theta_k U_k(y, \mathcal{N}_y, X) \tag{10}$$

Linear combination

Total energy

$$U_{tot}(Y,X,\theta) = \sum_{y \in Y} \theta_1 U_1(y,X) + \cdots + \theta_k U_k(y,\mathcal{N}_y,X) \quad (10)$$

- Weights $\theta \in \mathbb{R}^k$
- $\widehat{\theta}$ set by trial and error, aiming for $Y^+ = \underset{Y \in \mathcal{P}}{\operatorname{argmin}}\, U_{tot}(Y,X,\widehat{\theta})$
- Calibration needed :

  for $l = 1,\dots,k$, must find $d_l \in \mathbb{R}$ so that $U_l'(y,X) = U_l(y,X) - d_l < 0$ for valid y and > 0 for non-valid y

Learning the energy combination model

New total energy

We introduce a scalar θ_0 to learn the calibration,

$$U_{tot}(Y, X, \theta) = \sum_{y \in Y} \theta_1 U_1(y, X) + \cdots + \theta_k U_k(y, \mathcal{N}_y, X) + \theta_0 \tag{11}$$

Ayana Inria

Learning the energy combination model

Contrastive learning[6]

$Y^+ = \underset{Y \in \mathcal{P}}{\text{argmin}}\, U_{tot}(Y, X, \widehat{\theta})$ translates as local constraint[7] to

$$U_{tot}(Y^+, X, \theta) < U_{tot}(Y^-, X, \theta), \quad Y^- \sim Q(Y^+ \to \cdot) \tag{12}$$

We minimize the loss

$$\mathcal{L}(Y^+, X, \theta) = U_{tot}(Y^+, X, \theta) - U_{tot}(Y^-, X, \theta) \tag{13}$$

- Perturbation kernel Q needs defining

[6] Mabon *et al.*, "CNN-based energy learning for MPP object detection in satellite images," 2022.
[7] Craciun *et al.*, "Joint Detection and Tracking of Moving Objects Using Spatio-temporal Marked Point Processes," 2015.

Learning the energy combination model

Extending to maximum likelihood training

Hinton[8]and Du[9]show that if we sample Y^- as

$$Y^- \sim p_\theta, \quad p_\theta \propto \exp(-U_{tot}(\cdot, X, \theta)) \tag{14}$$

and it minimizes the negative log likelihood

$$\mathcal{L}_{nll}(\theta, \mathcal{S}) = -\log(P(Y_1^+, \dots, Y_N^+ | X_1, \dots, X_N, \theta)) \tag{15}$$

With training set $\mathcal{S} = \{(X_N, Y_N^+), \dots, (X_N, Y_N^+)\}$

[9]Hinton, "Training Products of Experts by Minimizing Contrastive Divergence," 2002.
[9]Du and Mordatch, "Implicit Generation and Modeling with Energy Based Models."

04

Point process simulation

Inferring the best fitting configuration

- Knowing $h(Y|X) \propto \exp(-U(Y,X))$
- We look for $\widehat{Y} = \text{argmin}_{Y \in \mathcal{P}} U_{tot}(Y, X, \widehat{\theta})$

Reversible Jump Markov Chain Monte Carlo[10]

- Simulate $Y_t \sim h/T_t$, with simulated annealing ($T_{t+1} = 0.999\, T_t$)
- Y_t converges towards $\widehat{Y}$

[10] Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," 1995.

Reversible Jump Markov Chain Monte Carlo

- Markov chain with stationary density $Y_t \sim h/T_t$
- Simulated annealing, with geometrical decrease of T ($T_{t+1} = 0.999 T_t$)

[11] Geyer and Møller, "Simulation Procedures and Likelihood Inference for Spatial Point Processes," 1994.

Reversible Jump Markov Chain Monte Carlo

- Markov chain with stationary density $Y_t \sim h/T_t$
- Simulated annealing, with geometrical decrease of T ($T_{t+1} = 0.999\,T_t$)
- Minimal proposition kernel[11]: Birth-Death

[11] Geyer and Møller, "Simulation Procedures and Likelihood Inference for Spatial Point Processes," 1994.

Reversible Jump Markov Chain Monte Carlo

- Markov chain with stationary density $Y_t \sim h/T_t$
- Simulated annealing, with geometrical decrease of T ($T_{t+1} = 0.999\, T_t$)
- Minimal proposition kernel[11]: Birth-Death
- + Birth map defined by the energy map built from the energy maps
- + Translation/rotation/scaling sampled according to densities set by energy maps

[11] Geyer and Møller, "Simulation Procedures and Likelihood Inference for Spatial Point Processes," 1994.

Reversible Jump Markov Chain Monte Carlo

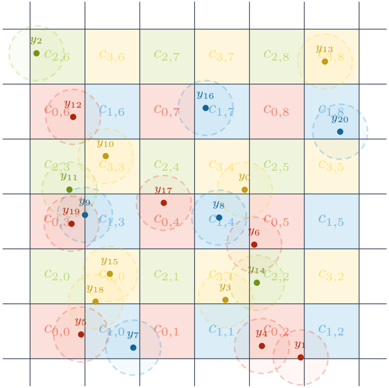
- Markov chain with stationary density $Y_t \sim h/T_t$
- Simulated annealing, with geometrical decrease of $T (T_{t+1} = 0.999 T_t)$
- Minimal proposition kernel[11]: Birth-Death
- + Birth map defined by the energy map built from the energy maps
- + Translation/rotation/scaling sampled according to densities set by energy maps
- a move $Y \to Y'$ is accepted with probability

$$\alpha(Y \to Y') = \min\left\{1, \frac{Q(Y' \to Y)}{Q(Y \to Y')} \exp\left(\frac{U_{tot}(Y, X, \widehat{\theta}) - U_{tot}(Y', X, \widehat{\theta})}{T}\right)\right\} \tag{16}$$

[11] Geyer and Møller, "Simulation Procedures and Likelihood Inference for Spatial Point Processes," 1994.

Speeding up the simulation : parallel sampling

The point process is Markovian: points do not interact beyond a maximum distance.
Verdie and Lafarge[12] show one can split the space into sets where $\alpha(Y \to Y')$ are independent



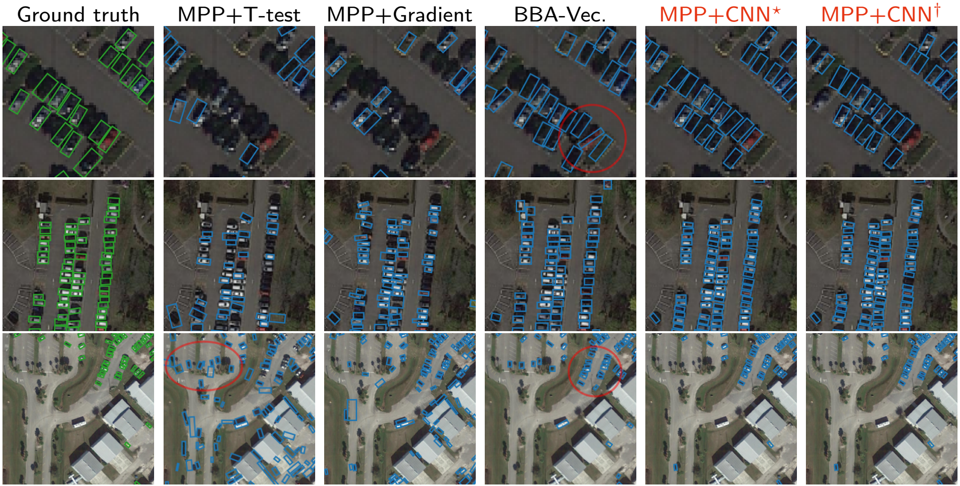
[12] Verdié and Lafarge, "Detecting Parametric Objects in Large Scenes by Monte Carlo Sampling," 2014.

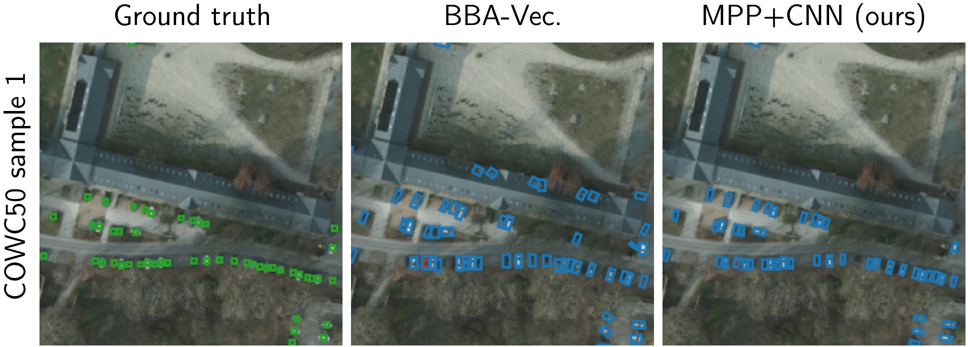
05

Results on remote sensing datasets

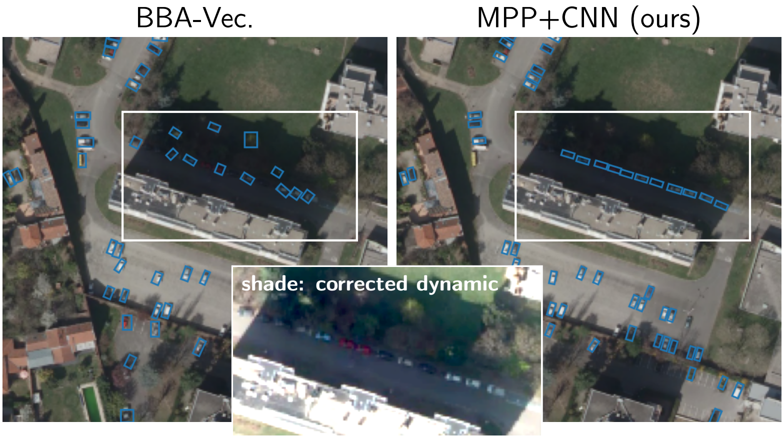
- Images subsampled to 50 cm/pixel
- Compare our MPP+CNN method against BBA-Vec.[13]
- Various datasets :
  - DOTA[14] (labeled with oriented rectangles, training dataset)
  - COWC[15] (labeled with centers)
  - Airbus aerial images (unlabeled)
- MPP+CNN*: manual weights θ, MPP+CNN†: learned weights θ

[13] Yi *et al.*, "Oriented Object Detection in Aerial Images with Box Boundary-Aware Vectors," 2021.
[14] Xia *et al.*, "DOTA: A Large-Scale Dataset for Object Detection in Aerial Images," 2018.
[15] Mundhenk *et al.*, "A Large Contextual Dataset for Classification, Detection and Counting of Cars with Deep Learning," 2016.


Ground truth
MPP+T-test
MPP+Gradient
BBA-Vec.
MPP+CNN*
MPP+CNN†


Ground truth
BBA-Vec.
MPP+CNN (ours)
COWC50 sample 1

Airbus data, difficult example



Effects of fewer training data

Training on 100% and 25% of the training dataset

BBA-Vec. | BBA-Vec. (25%) | MPP+CNN | MPP+CNN (25%)

DOTA : Metrics

F1@t: F1 metric for IOU threshold t
Higher t requires a finer match between GT to the prediction

Method	Training data	F1@0.25	Pr@0.25	Rc@0.25	F1@0.5	Pr@0.5	Rc@0.5
BBA-Vec.[16]	100%	**0.68**	0.63	0.74	0.48	0.43	0.53
BBA-Vec.	50%	0.58	0.55	0.62	0.23	0.22	0.25
BBA-Vec.	25%	0.52	0.51	0.54	**0.13**	0.12	0.15
MPP+CNN	100%	**0.66**	0.56	0.79	0.42	0.32	0.64
MPP+CNN	50%	0.57	0.46	0.75	0.31	0.22	0.52
MPP+CNN	25%	0.55	0.48	0.64	**0.34**	0.26	0.49

[16] Yi *et al.*, "Oriented Object Detection in Aerial Images with Box Boundary-Aware Vectors," 2021.

06

Conclusion & perspectives

Contributions

- Likelihood terms learned with CNNs, replacing contrast measures.
- Propose a method to learn the relative weights of energies.
- Results equivalent to SOTA (metrics wise) while having more spatial coherence/regularization thanks to added priors.

Perspectives

- Working on applying this model to time series where the priors on dynamics are stronger

Publications & Code

Publications

See julesmabon.com/publications or team.inria.fr/ayana/publications-hal

- Processus ponctuels marqués et réseaux de neurones convolutifs pour la détection d'objets dans des images de télédétection, *GRETSI*, 2022, Nancy
- Point process and CNN for small object detection in satellite images, *SPIE, Image and Signal Processing for Remote Sensing XXVIII*, 2022, Berlin
- CNN-based energy learning for MPP object detection in satellite images, *IEEE, International workshop on machine learning for signal processing*, 2022, Xi'an (virtual)

Code

Available at github.com/Ayana-Inria/MPP_CNN_RS_object_detection

Thank you !

Any questions ?

Bibliography I

[1] G.-S. Xia *et al.*, "DOTA: A large-scale dataset for object detection in aerial images," in *Proc. CVPR*, 2018.

[2] C. Lacoste, X. Descombes, and J. Zerubia, "Point processes for unsupervised line network extraction in remote sensing," *IEEE TPAMI*, vol. 27, no. 10, pp. 1568–1579, 2005.

[3] M. S. Kulikova, I. H. Jermyn, X. Descombes, E. Zhizhina, and J. Zerubia, "Extraction of arbitrarily-shaped objects using stochastic multiple birth-and-death dynamics and active contours," in *Proc. Computational Imaging VIII*, SPIE, 2010.

[4] J. Mabon, M. Ortner, and J. Zerubia, "Processus ponctuels marqués et réseaux de neurones convolutifs pour la détection d'objets dans des images de télédétection," in *GRETSI 2022 - XXVIIIème Colloque Francophone de Traitement du Signal et des Images*, Nancy, France, Sep. 2022. [Online]. Available: `https://hal.inria.fr/hal-03715337`.

[5] ——,"Cnn-based energy learning for mpp object detection in satellite images," in *IEEE International workshop on machine learning for signal processing*, 2022.

Bibliography II

[6] P. Craciun, M. Ortner, and J. Zerubia, "Joint detection and tracking of moving objects using spatio-temporal marked point processes," in *Proc. IEEE Winter Conf. on Applications of Computer Vision*, 2015.

[7] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, 2002.

[8] Y. Du and I. Mordatch, "Implicit Generation and Modeling with Energy Based Models," *NeurIPS 2019*, p. 11,

[9] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.

[10] C. J. Geyer and J. Møller, "Simulation procedures and likelihood inference for spatial point processes," *Scandinavian journal of statistics*, pp. 359–373, 1994.

[11] Y. Verdié and F. Lafarge, "Detecting parametric objects in large scenes by Monte Carlo sampling," *International Journal of Computer Vision*, vol. 106, no. 1, pp. 57–75, 2014.

Bibliography III

[12] J. Yi, P. Wu, B. Liu, Q. Huang, H. Qu, and D. Metaxas, "Oriented object detection in aerial images with box boundary-aware vectors," in *Proc. IEEE Winter Conf. on Applications of Computer Vision*, 2021.

[13] T. N. Mundhenk, G. Konjevod, W. A. Sakla, and K. Boakye, "A Large Contextual Dataset for Classification, Detection and Counting of Cars with Deep Learning," in *Proc. ECCV*, 2016.

Fin !