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)

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.

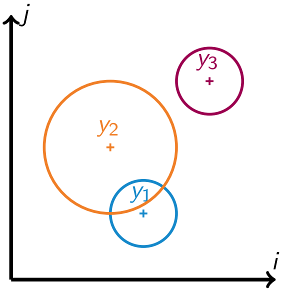
02

Point process: definition

Marked point process

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

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

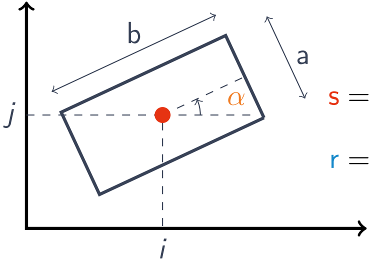


	i	j	r
y_1 : (	2	1	0.5)
y_2 : (	1.5	2	1)
y_3 : (	3	3	0.5)

Parametrization

Oriented rectangle

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



$$s = \frac{a+b}{2}$$

$$r = \frac{a}{b}$$

Point process density

Y is the realization of a random variable of density h

Point process density (Gibbs)

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

Simplified energy model

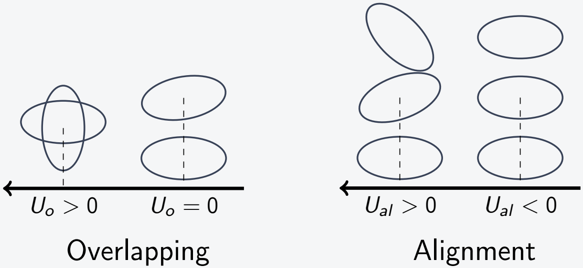
$$U(Y, X) = \sum_{y \in Y} U_{data}(y, X) + U_{prior}(y, \mathcal{N}_y) \tag{2}$$

X: image, $\mathcal{N}_y$: neighborhood of y in Y

Data and prior terms

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

Prior term



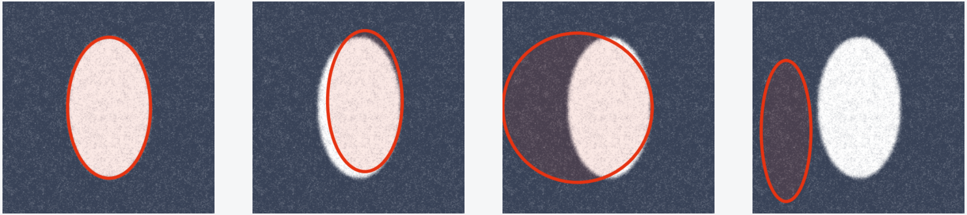
Or others:

- Size
- Shape
- Dynamics
- ...

Data and prior terms

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

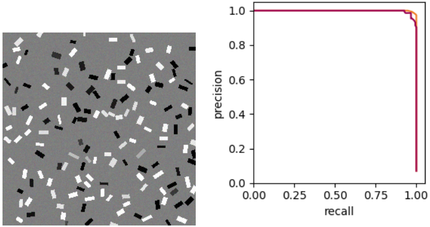
Data term



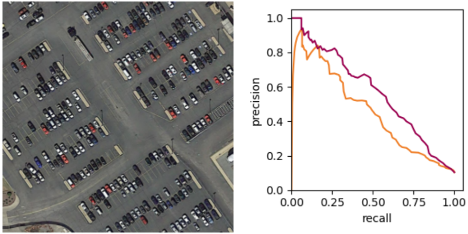
$U_{data} < 0$ $\quad$ $U_{data} \simeq 0$ $\quad$ $U_{data} > 0$ $\quad$ $U_{data} \gg 0$

Contrast measures as data term

Synthetic data



DOTA sample

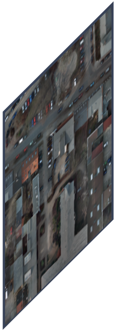


Precision-recall for the T-test measure[2] (red) and for a measure based on image gradient[3] (orange).

[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.

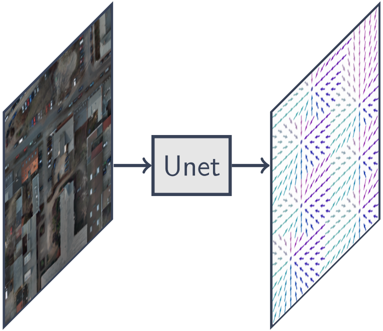
03

Inferring the energy map



Input image X

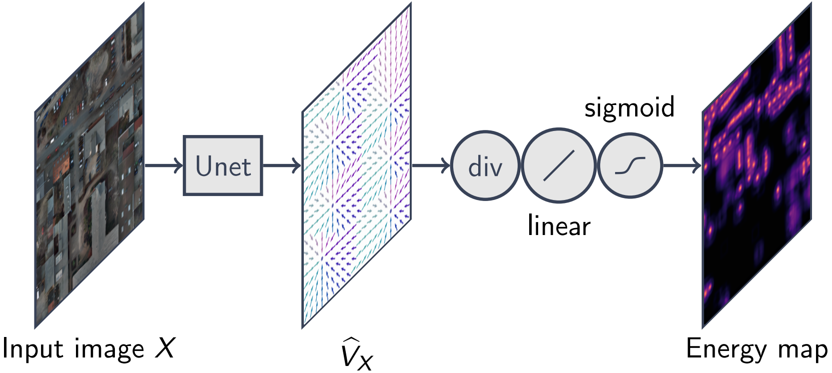
Inferring the energy map



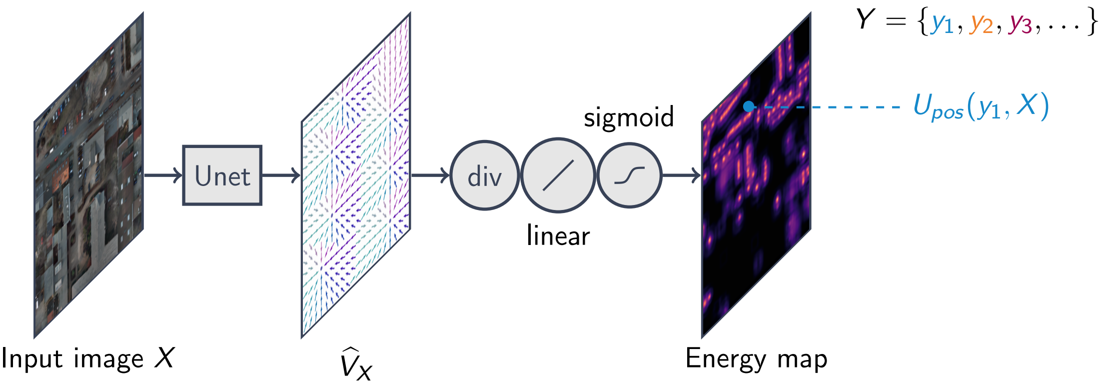
Input image X

$\widehat{V}_X$

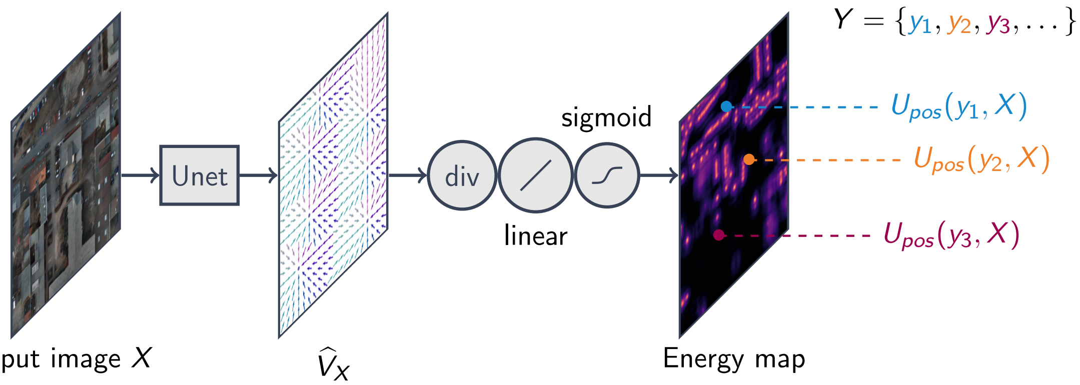
Inferring the energy map



Inferring the energy map



Inferring the energy map



Position energy term

Position energy term

$$U_{pos}(y, X) = -2\sigma \left(a.\mathrm{div}(\widehat{V}_X) + b \right) [y_i, y_j] + 1 \quad (3)$$

- 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

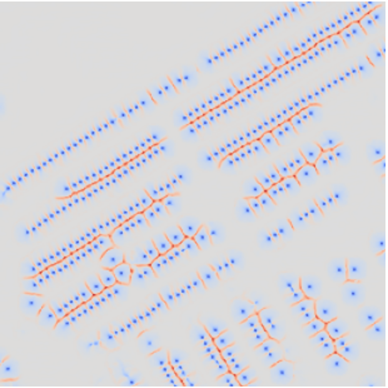
Cost function

$$L_{pos}(\widehat{V}_X, Y_{GT}) = \mathrm{MSE}(\widehat{V}_X, V_{Y_{GT}}) + \mathrm{BCE}(G_{Y_{GT}}, \sigma(a.\mathrm{div}(\widehat{V}_X) + b)) \quad (4)$$

- MSE : Mean Squared Error, BCE : Binary Cross Entropy
- $V_{Y_{GT}}$: Vector field built from the ground truth Y_{GT}
- $G_{Y_{GT}}$: center binary map + Gaussian filter (objects as “blobs”)
- Data augmentation: patch sampling, rotation, hue/contrast/luminosity variations etc.

Position energy term: example

Image X



$\mathrm{div}(\widehat{V}_X)$

Marks energy term: energy tensor

Extracting energies $U_m(y, X)$ from an energy tensor inferred with a Unet[4]

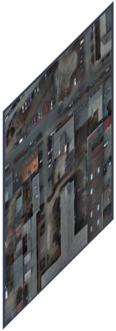


Image X (H, W)

[4] Ronneberger *et al.*, "U-Net: Convolutional Networks for Biomedical Image Segmentation," 2015.

Marks energy term: energy tensor

Extracting energies $U_m(y, X)$ from an energy tensor inferred with a Unet[4]

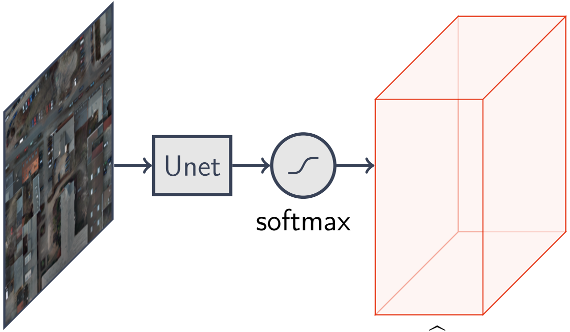
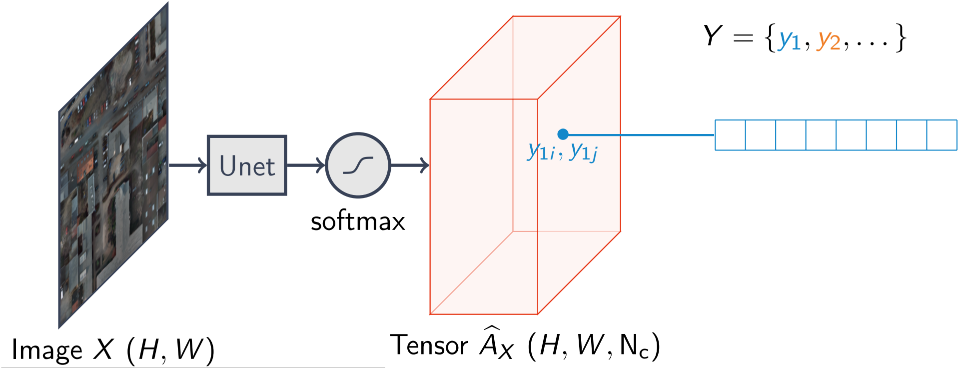


Image X (H, W) Tensor $\widehat{A}_X$ $(H, W, \mathrm{N_c})$

[4] Ronneberger *et al.*, "U-Net: Convolutional Networks for Biomedical Image Segmentation," 2015.

Marks energy term: energy tensor

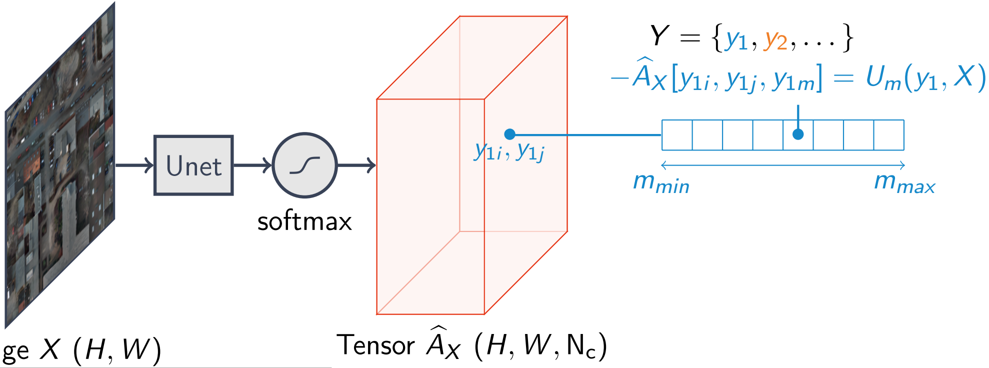
Extracting energies $U_m(y, X)$ from an energy tensor inferred with a Unet[4]



[4] Ronneberger *et al.*, "U-Net: Convolutional Networks for Biomedical Image Segmentation," 2015.

Marks energy term: energy tensor

Extracting energies $U_m(y, X)$ from an energy tensor inferred with a Unet[4]

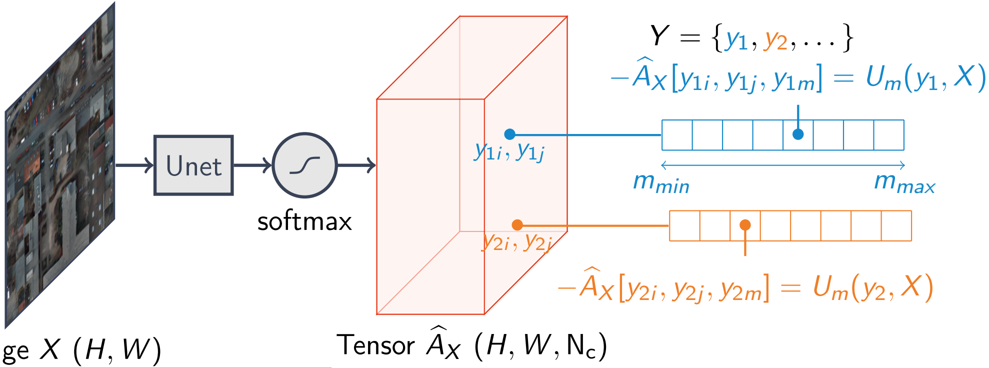


[4]Ronneberger *et al.*, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” 2015.

Ayana Inria

Marks energy term: energy tensor

Extracting energies $U_m(y, X)$ from an energy tensor inferred with a Unet[4]



[4] Ronneberger *et al.*, "U-Net: Convolutional Networks for Biomedical Image Segmentation," 2015.

Marks energy term

Energy on marks

For each mark m

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

- $A[i, j, k]$: 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 $\forall y \in S \times M$
- $\text{index}(y_m) = \left\lfloor N_c \frac{y_m - m_{min}}{m_{max} - m_{min}} \right\rfloor$

Learning the mark energy term

Cost function

$$L_m(\widehat{A}_X, Y_{GT}) = \frac{1}{|P|} \sum_{p \in P} \mathrm{CE}(\widehat{A}_X[p], A_{Y_{GT}}[p]) \tag{6}$$

- 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

Priors on configurations

Non-superposition 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

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

$$U_{tot} = \sum_{y \in Y} f\left(U_1, \ldots, U_k\right)$$

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

Linear combination

Total energy

$$U_{tot}(Y, X, \theta) = \sum_{y \in Y} U_1(y, X) + \theta_2 U_2(y, \mathcal{N}_y, X) + \cdots + \theta_k U_k(y, \mathcal{N}_y, X) \quad (10)$$

Linear combination

Total energy

$$U_{tot}(Y, X, \theta) = \sum_{y \in Y} U_1(y, X) + \theta_2 U_2(y, \mathcal{N}_y, X) + \cdots + \theta_k U_k(y, \mathcal{N}_y, X) \quad (10)$$

- Weights $\theta \in \mathbb{R}^{k-1}$
- $\widehat{\theta}$ set by trial and error
- Calibration needed :

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

04

Point process simulation

Objective

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

Reversible Jump Markov Chain Monte Carlo[5]

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

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

05

Results on remote sensing datasets

- Images subsampled to 50 cm/pixel
- Compare our MPP+CNN method against BBA-Vec.[6]
- Various datasets :
  - DOTA[7] (labeled with oriented rectangles, training dataset)
  - COWC[8] (labeled with centers)
  - Airbus aerial images (unlabeled)

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

Ground truth | BBA-Vec. | MPP+CNN (ours)

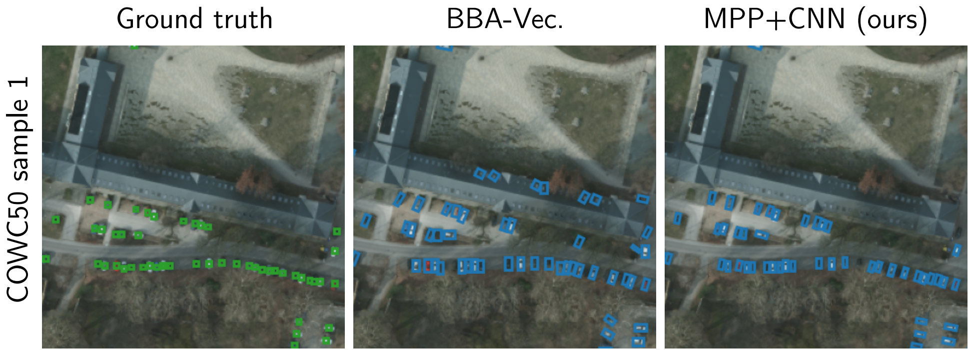
DOTA50 sample 1

Ayana Inria

Ground truth | BBA-Vec. | MPP+CNN (ours)

DOTA50 sample 3

Ayana Inria



Ayana Inria

DOTA : Metrics

Method	Training data	F1@0.25	Pr@0.25	Rc@0.25	F1@0.5	Pr@0.5	Rc@0.5
BBA-Vec.[9]	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

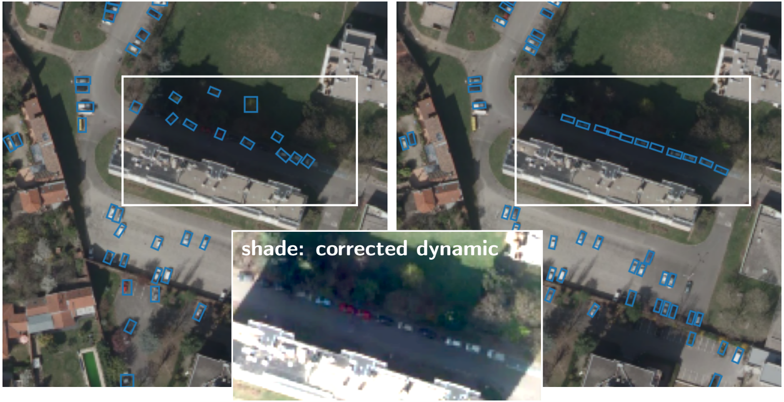
[9] Yi *et al.*, "Oriented Object Detection in Aerial Images with Box Boundary-Aware Vectors," 2021.

Effects of fewer training data

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

Airbus data, difficult example

BBA-Vec. MPP+CNN (ours)



06

Ayana Inria

Conclusion & perspectives

Contributions

- Likelihood terms learned with CNNs, replacing contrast measures.
- Results equivalent to SOTA (metrics wise) while having more spatial coherence/regularization thanks to added priors.

Perspectives

- While energies weights are set manually here, one can learn these parameters [10]
- Working on applying this model to time series where the priors on dynamics are stronger

[10] Mabon *et al.*, "CNN-based energy learning for MPP object detection in satellite images," 2022.

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] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. MICCAI*, 2015.

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

Bibliography II

[6] 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.

[7] 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.

[8] J. Mabon, M. Ortner, and J. Zerubia, "Cnn-based energy learning for mpp object detection in satellite images," in *IEEE International workshop on machine learning for signal processing*, 2022.

Thank you !

Any questions ?