Learning and scoring Point Process models for object detection in satellite images

Jules Mabon, Mathias Ortner, Josiane Zerubia

Learning and scoring Point Process models for object detection in satellite images

1st Jules Mabon
Inria, Université Côte d'Azur
Sophia-Antipolis, France

2nd Mathias Ortner
Airbus Defense and Space
Toulouse, France

3rd Josiane Zerubia
Inria, Université Côte d'Azur
Sophia-Antipolis, France

***Abstract*—In this paper we propose a joint Point Process and CNN based method for object detection in satellite imagery. The Point Process allows building a lightweight interaction model, while the CNN allows to efficiently extract meaningful information from the image in a context where interaction priors can complement the limited visual information. More specifically, we present matching parameter estimation and result scoring procedures, that allow to take into account object interaction. The method provides good results on benchmark data, along with a degree of interpretability of the output. The code will be available at github.com/Ayana-Inria/**

***Index Terms*—Object Detection, Point Process, Convolutional Neural Network, Energy Based Model, Remote Sensing**

I. Introduction

For over 20 years of study [1] small object detection in optical satellite images has remained challenging; objects of interest are only few pixels large (thus lack visual information), and the dense scattering of objects increases the difficulty of separating instances (introducing interactions between neighboring objects). In this paper, we aim at extracting the geometrical configuration of vehicles in images with resolutions around 0.5 m, while leveraging the priors on object interaction to compensate for the lack of visual information.

Most Convolutional Neural Network (CNN) based approaches utilize anchor proposals [2], [3] or heatmaps [4] to locate objects; failing to consider interactions other than through a post-processing step. Approaches such as [5] model interactions through a cascade of Transformers, at a great complexity cost.

On the other hand, Point Process (PP) [6] methods allow building lightweight interaction models and jointly solve for the local and interaction contribution of a detection. The previous approaches applied to microscopy [7] or remote sensing data [8], [9] rely on contrast measures to assert the correspondence of points with the image; those fail when objects and backgrounds are varied.

With our proposed approach, we leverage the feature extraction capabilities of CNN within the PP framework as introduced in our previous works [10]. More specifically in this paper, we focus on the scoring for the output detection that takes into account interactions (while CNN approaches score objects independently, and previous PP methods use no score, simply computing precision and recall from the output as is). This score has good properties regarding the parameter optimization method utilized, and allows for interpretable results.

Thanks to BPI France (LiChiE contract) for funding, and to the OPAL infrastructure from Université Côte d'Azur for providing computational resources and support.

II. Point Process

Point Processes consider configurations of points (a finite non-ordered set $\mathbf{y}$ of elements of $\mathcal{S} \times \mathcal{M}$) as realizations of a random variable Φ in the set of all possible configurations $\mathcal{Y} = \bigcup_{n=0}^{\infty} (\mathcal{S} \times \mathcal{M})^n$ (with an arbitrary amount of points). Space $\mathcal{S}$ corresponds to the image space, and $\mathcal{M}$ to the mark space. A mark can be any random variable from the radius of a circle to a discrete categorization of the object. In our case, a point $y \in Y$ is composed of coordinates y_i, y_j in $\mathcal{S}$, and three marks that describe a rectangle : width y_a, length y_b and angle y_α. We denote $n(\mathbf{y})$ the number of points in a configuration $\mathbf{y}$. As the number of points is an unknown of the problem, Point Processes require specific sampling procedures we will detail later.

A. Point Process density

The law of the random variable is defined by its density h relative to the uniform Point Process [6]. The density derives from an energy U, through a Gibbs density :

$$h(\mathbf{y}|\mathbf{X}) = \frac{1}{Z} \exp(-U(\mathbf{y}, \mathbf{X}, \theta)), \tag{1}$$

with Z an intractable normalizing constant. The energy measures the compatibility between the image $\mathbf{X}$ and the configuration $\mathbf{y}$ for a given set of parameters θ; the lower the energy the higher the compatibility (see Energy Based Models [11]).

B. Papangelou conditional intensity

The reference Poisson Point Process has an intensity λ that is either constant or depends on the location[1]. The density in (1) implies the intensity is now a function of the location and neighborhood of a point. Thus, the Papangelou conditional intensity $\lambda(\cdot;\cdot)$ [6] associated to a Point Process Φ, is introduced as:

$$\lambda(y;\mathbf{y})dy = p\left(N_\Phi(dy) = 1 | \Phi \cap (dy)^c = \mathbf{y} \cap (dy)^c\right), \tag{2}$$

[1]For any compact $A \subseteq \mathcal{S}$, $\mathbb{E}[n(\Phi)] = \lambda|A|$ if λ is constant.

i.e. the infinitesimal probability to find a point in region dy around $y \in \mathcal{S}$, given the configuration $\mathbf{y}$ outside dy (i.e. $(dy)^c$). This conditional density is what allows modeling interaction between points of the PP.

When $y \in \mathbf{y}$, the Papangelou conditional intensity can be computed from the energy function U as:

$$\begin{aligned}\lambda(y; \mathbf{y} \setminus \{y\}) &= \frac{h(\mathbf{y}|\mathbf{X})}{h(\mathbf{y} \setminus \{y\}|\mathbf{X})} \\ &= \exp\left(U(\mathbf{y} \setminus \{y\}, \mathbf{X}, \theta) - U(\mathbf{y}, \mathbf{X}, \theta)\right).\end{aligned} \tag{3}$$

III. Energy model

We define our energy function in (1) as a sum of energy terms V_e ($e \in \xi$, the set of energy terms) for each point y in the configuration:

$$U(\mathbf{y}, \mathbf{X}, \theta) = \sum_{y \in \mathbf{y}} \underbrace{\sum_{e \in \xi} w_e V_e(y, \mathbf{X}, \mathcal{N}_y^{\mathbf{y}}, \theta)}_{= V(y, \mathbf{X}, \mathcal{N}_y^{\mathbf{y}}, \theta)}. \tag{4}$$

The weight parameters $w_e \in \theta$ encode the relative importance of each energy term. We distinguish between two kinds of energy terms: data terms, written $V_e(y, \mathbf{X}, \theta)$, that measure the compatibility of a point against the image; and prior terms, written $V_e(y, \mathcal{N}_y^{\mathbf{y}}, \theta)$, that measure the coherence of a point itself or against its neighborhood $\mathcal{N}_y^{\mathbf{y}}$. The energy of a point is denoted $V(y, \dots)$.

A. Interpreting CNN outputs as data terms

As in our previous work [10], we build data terms by reinterpreting the outputs of a CNN as energies that can be plugged into the energy model (4).

a) Position term: CNN based models such as [4] use a heatmap to find object key points (in our case: centers). Supposing that the CNN outputs a center probability map as $p(y_i, y_j|\mathbf{X}) = \sigma(\widehat{\mathbf{Z}}_{pos}[y_i, y_j])$ where[2] $\widehat{\mathbf{Z}}_{pos} \in \mathbb{R}^{H \times W}$ is the output tensor for an input image of size $(H, W, 3)$, we reinterpret the output as an energy as follows:

$$V_{pos}(y, \mathbf{X}, \theta) = \ln(1 + \exp(-\widehat{\mathbf{Z}}_{pos}[y_i, y_j] + t_{pos})), \tag{5}$$

with $t_{pos} \in \theta$ a threshold parameter.

b) Mark terms: Similarly, for each mark $\kappa \in \{a, b, \alpha\}$, supposing we have a CNN trained to perform pixel classification into n_κ discrete classes for mark κ. The CNN model provides the probability of the pixel at y_i, y_j to belong to class c_k ($k = 1, \dots, n_\kappa$) as:

$$p\left(c_k | y_i, y_j, \mathbf{X}\right) = \frac{\exp\left(\widehat{\mathbf{Z}}_\kappa^{y_i, y_j}[k]\right)}{\sum_{k'=1}^{n_\kappa} \exp\left(\widehat{\mathbf{Z}}_\kappa^{y_i, y_j}[k']\right)}, \tag{6}$$

i.e. using the Softmax operation. Then, as in [12], we reinterpret the Softmax input as an energy:

$$V_\kappa(y, \mathbf{X}) = -\widehat{\mathbf{Z}}_\kappa^{y_i, y_j}\left[c^\kappa(y_\kappa)\right] + \ln\left(\sum_{k=1}^{n_\kappa} \exp\left(\widehat{\mathbf{Z}}_\kappa^{y_i, y_j}[k]\right)\right), \tag{7}$$

where $c^\kappa(y_\kappa)$ is the class corresponding to value y_κ.

[2]With $\widehat{\mathbf{Z}}_{pos}[y_i, y_j]$ denoting the value of tensor $\widehat{\mathbf{Z}}_{pos}$ at coordinates y_i, y_j.

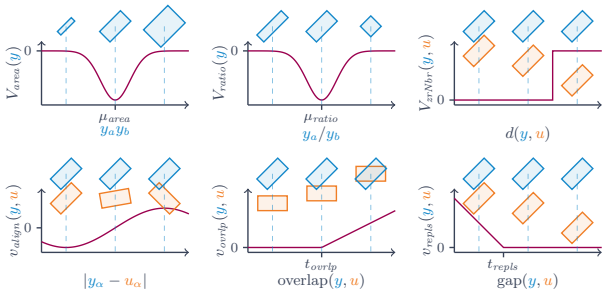

Fig. 1. Illustration of some energy priors.

B. Prior terms

To build the prior and interaction model we combine a set of simple energy functions (that will form more complex behaviors when combined in (4)). We illustrate some priors used in Figure 1, with $V_e(y)$ an energy term computed over a single object y, and $v_e(y, u)$ an interaction energy between objects y and u, aggregated for each y with a $\max$ or $\min$ over its neighbors. Notations μ_e and t_e represent model parameters. Further details can be found in [10].

C. Learning the energy model

To estimate the parameters θ of the energy model (4), we aim to maximize the likelihood of parameters relative to the annotated data $\mathcal{D}$ [11]. To bypass intractable integrals, [13] propose the Contrastive Divergence approach: i.e., within a stochastic gradient descent scheme, minimizing the loss:

$$\mathcal{L}\left(\theta_n, \mathbf{y}^+, \mathbf{y}^-\right) = U(\mathbf{y}^+, \mathbf{X}, \theta_n) - U(\mathbf{y}^-, \mathbf{X}, \theta_n) + \gamma R_V, \tag{8}$$

where R_V is a regularization term that avoid exploding energies (average of all $V(y)$), $\mathbf{y}^+$ is a configuration close to ground truth, and $\mathbf{y}^- \sim \exp(U(\cdot, \mathbf{X}, \theta_n))$. We detail the procedure adapted from [14] to Point Processes in [10].

D. Sampling with Jump Diffusion

The Point Process is sampled by building a Markov chain $(\mathbf{y}_t)_{t=1,\dots}$ which converges towards the stationary density h.

a) Birth and Death: Birth and Death moves [15] allow adding and removing points. For our model we leverage the precomputed tensors $\widehat{\mathbf{Z}}_{pos}$ and $\widehat{\mathbf{Z}}_\kappa^{y_i, y_j}$ to propose more relevant points in space [10].

b) Diffusion: To modify the current configuration $\mathbf{y}_t$ at a fixed number of points, we leverage Diffusion dynamics [16]:

$$\mathbf{y}_{t+1} \leftarrow \mathbf{y}_t + -\beta \nabla_{\mathbf{y}_t} U(\mathbf{y}_t) + \sqrt{2T_t} w_t,\ w_t \sim \mathcal{N}(0, \beta), \tag{9}$$

with T_t the temperature at time t and β the gradient step.

c) Simulated annealing: When looking for the *best fitting configuration* for a given image $\mathbf{X}$, i.e. the configuration that minimizes the energy U, we use simulated annealing[3]: i.e. simulate a chain of stationary density h/T_t with $T_{t+1} = 0.998 T_t$.

[3]For a logarithmic temperature decrease the chain is proven to converge towards the global minimum. Geometric decrease approximates the latter while providing faster inference times.

IV. Scoring using point interactions

Classical CNN based object detection models for object detection (such as [2], [4], [17]) yield a confidence score $s(y) \in \mathbb{R}$ for each proposed object y in the image. This confidence score is often interpreted, for each detection, as proportional to the probability of proposed element y to be a true positive, $s(y) \propto p(y|\mathbf{X})$. Applying a score (or confidence) threshold t_s gives a set of detections, for which metrics such as precision and recall can be computed by matching the detections with the ground truth. This allows adapting the threshold according to the need of the application; i.e. some applications may require few false positive (high precision) while others require less missed detections (high recall). In order to assess the performance independently of the threshold selection, the Average Precision (AP) metric sums up the performance of a model as the area under the precision-recall curve.

Previous Point Process approaches [7], [18], [19] only compute simple metrics such as precision, recall or F1 score for the configuration given by the sampling procedure, as no score is associated to each object detection.

A. Papangelou intensity as score

With our PP approach, we propose to introduce a scoring function, first to filter the detections given a confidence threshold, second to be able to compare our method to others using the widely used AP metric. Within the PP framework, the probability of one proposed point being an object of interest depends on the rest of the inferred configuration $\hat{\mathbf{y}}$, thus the scoring function reflects it: $s(y|\hat{\mathbf{y}} \setminus \{y\}) \propto p(y|\hat{\mathbf{y}} \setminus \{y\}, \mathbf{X})$. From (2) we have that the Papangelou conditional intensity is proportional to the probability of finding a point $y \in \mathbf{y}$ in a small neighborhood dy knowing the rest of the configuration $\mathbf{y} \setminus \{y\}$. We propose to use the Papangelou conditional intensity as a score :

$$s(y|\mathbf{y} \setminus \{y\}) = \lambda(y; \mathbf{y} \setminus \{y\}) \tag{10}$$

B. Pruning sequence

However, the dependency of the score on the current configuration yields a complication while computing the Average Precision: when applying a threshold t_s to prune the configuration $\mathbf{y}$ into $\mathbf{y}' \subset \mathbf{y}$, for any $y \in \mathbf{y}'$, the score $s(y|\mathbf{y} \setminus \{y\})$ may differ from $s(y|\mathbf{y}' \setminus \{y\})$. With a score of the form $s(y)$, that only depends on y and the image $\mathbf{X}$ — such as those from classical CNN models — the score from one object after pruning is unchanged.

In the PP case, we compute the scores by sequentially removing the lowest scoring point until none is left; i.e., we build a sequence of configurations $\mathbf{y}_1 \supset \mathbf{y}_2 \dots \mathbf{y}_{n(\hat{\mathbf{y}})-1} \supset \mathbf{y}_{n(\hat{\mathbf{y}})} \supset \emptyset$, with $\mathbf{y}_1 = \hat{\mathbf{y}}$, for $n = 1, \dots, |\hat{\mathbf{y}}|$:

$$\mathbf{y}_{n+1} = \mathbf{y}_n \setminus \{y_n\},\ y_n = \underset{y \in \mathbf{y}_n}{\arg\min} \lambda(y; \mathbf{y}_n \setminus \{y\}), \tag{11}$$

$$s(y_n|\mathbf{y}_n \setminus \{y_n\}) = s(y_n|\mathbf{y}_{n+1}) = \lambda(y; \mathbf{y}_{n+1}). \tag{12}$$

Equation (11) provides a pruning order $y_1, \dots, y_{n(\hat{\mathbf{y}})}$ of points in $\hat{\mathbf{y}}$. This ordering allows to plot the precision and recall curve. Indeed, to trace a precision recall-curve, one only requires the sequence of $(\text{Recall}(t_s), \text{Precision}(t_s))$ pairs, which are obtained by sequentially pruning the lowest scoring points. Equation (12) provides a score to each point y_n.

C. Contrastive divergence loss and Papangelou intensity

On one hand the energy model is trained by minimizing the loss function in (8) derived from the likelihood maximization of the parameters regarding the annotated data. On the other we evaluate the performance of the inferred configuration with the scoring method in (12) sourced from the Papangelou intensity. Here we show that while the two are derived differently, minimization of the loss function leads to good properties on the score function.

Here we consider a simplified loss with only the two energy terms (as $\gamma \simeq 0$). Denoting the energy change induced by the move from configuration $\mathbf{y}$ to $\mathbf{x}$ as $\Delta U(\mathbf{y} \to \mathbf{x}) = U(\mathbf{x}) - U(\mathbf{y})$, we have :

$$\mathcal{L}(\theta, \mathbf{y}^+, \mathbf{y}^-) = \Delta U(\mathbf{y}^- \to \mathbf{y}^+). \tag{13}$$

Similarly, the Papangelou intensity can be rewritten as such:

$$\lambda(u; \mathbf{y}) = \exp(\Delta U(\mathbf{y} \cup \{u\} \to \mathbf{y})) \tag{14}$$

a) Single point addition: Thus, for a simple negative sample $\mathbf{y}^- = \mathbf{y}^+ \cup \{u\}$ in which we add a non-valid point u to $\mathbf{y}^+$, we have:

$$\mathcal{L}(\theta, \mathbf{y}^+, \mathbf{y}^-) = \log(\lambda(u; \mathbf{y}^+)). \tag{15}$$

This leads into the expected behavior: minimizing the loss $\mathcal{L}$ leads to minimizing the score of non-valid point u. The same stands for the removal of a valid point $y \in \mathbf{y}^+$, and maximizing its score.

b) Arbitrary sequence of moves: This is also valid for the generic case where $\mathbf{y}^-$ is generated from an arbitrary sequence of additions or removal of points from $\mathbf{y}^+$ (a translation/rotation/scaling can be viewed as removal then addition). This defines a sequence $(\mathbf{y}_k)_{k=0,\dots,n}$ of n configurations as:

$$\forall k = 1, \dots, n,\ \mathbf{y}_k = \begin{cases} \mathbf{y}_{k-1} \setminus \{y_k\} & \text{if } y_k \in \mathbf{y}^+ \\ \mathbf{y}_{k-1} \cup \{y_k\} & \text{otherwise,} \end{cases} \tag{16}$$

with $\mathbf{y}_0 = \mathbf{y}^+$, $\mathbf{y}^- = \mathbf{y}_n$, and y_k elements of either $\mathcal{S} \times \mathcal{M}$ or $\mathbf{y}^+$ Without loss of generality we can reorder the sequence to match the pruning order defined in (11). The energy change for one move is given as:

$$\Delta U(\mathbf{y}_{k-1} \to \mathbf{y}_k) = \begin{cases} \log(\lambda(y_k; \mathbf{y}_{k-1} \setminus \{y_k\})) & \text{if } y_k \in \mathbf{y}^+ \\ -\log(\lambda(y_k; \mathbf{y}_{k-1})) & \text{otherwise.} \end{cases} \tag{17}$$

As we have (by definition) $\Delta U(\mathbf{x} \to \mathbf{x}'') = \Delta U(\mathbf{x} \to \mathbf{x}') + \Delta U(\mathbf{x}' \to \mathbf{x}'')$, the loss is given as:

$$\mathcal{L} = \sum_{y_k \notin \mathbf{y}^+} \log(\underbrace{\lambda(y_k; \mathbf{y}_{k-1})}_{\text{(a)}}) - \sum_{y_k \in \mathbf{y}^+} \log(\underbrace{\lambda(y_k; \mathbf{y}_{k-1} \setminus \{y_k\})}_{\text{(b)}}). \tag{18}$$

By ordering the $y_k, \mathbf{y}_k$ to match the pruning order in (11) each $\lambda(y_k; \dots)$ can be matched to their respective score:

- (18)(a) corresponds to non-valid points added to $\mathbf{y}^+$, their score is minimized as the loss is decreased;
- (18)(b) corresponds to valid points removed from $\mathbf{y}^+$, their score is increased as the loss is minimized.

Hereby we showed that minimization of the loss at a configuration level leads to the expected results on object scores.

D. Results interpretability

Due to the decomposition of the total energy into energy terms introduced in (4), the object score can be decomposed similarly:

$$s(y|\mathbf{y} \setminus \{y\}) = \prod_{e \in \xi} s_e(y|\mathbf{y} \setminus \{y\}) \tag{19}$$

with $s_e(y|\mathbf{y} \setminus \{y\}) = \exp(w_e \Delta V_e(\mathbf{y} \to \mathbf{y} \setminus \{y\}))$, the Papangelou intensity obtained by considering the single energy term e. This allows viewing the contribution of each component. Moreover, we propose grouping these contributions into the data and prior contributions to respectively obtain s_{data} and s_{prior} such that the final score is a product of the two: $s(y|\mathbf{y} \setminus \{y\}) = s_{data}(y) s_{prior}(y|\mathbf{y} \setminus \{y\})$.

V. Application

A. Models

In this paper we show results on three CNN based models, and our two PP models: *CNN-PP*$^\blacklozenge$ and *CNN-PP*$^\bigstar$ correspond to our PP+CNN model, while the first only estimates the weights w_e through the estimation method (other parameters set manually); *CNN-LocalMax.* only uses the CNN part of our model with local maxima applied to extract object; *BBA-Vec.* and *YOLOV5-OBB* correspond to [2] and [17]. We provide further insights in the relative complexity of those models in [20].

B. Results

a) Quantitative and qualitative evaluation on benchmark data: We train and evaluate our models on the DOTA [21] dataset, sub-sampled to a 0.5 m resolution (in order to match satellite sensor specifications from Airbus). To assert the noise resilience, we also evaluate the methods on the same data with additive noise. For every model we compute the Average Precision (AP) in Table I: the increased performance from *CNN-LocalMax.* to *BBA-Vec.* and *CNN-PP*$^\bigstar$ shows the PP improves results over the CNN alone. Some results on sample images are shown in Fig. 3; it shows our CNN and PP combination allows for regularization of the resulting configurations

TABLE I

Method	AP_{DOTA}	$AP_{DOTA+noise}$
BBA-Vec.	0.82	0.19
YOLOV5-OBB	0.86	0.10
CNN-LocalMax.	0.86	0.55
CNN-PP$^\blacklozenge$	**0.91**	**0.58**
CNN-PP$^\bigstar$	**0.92**	**0.62**

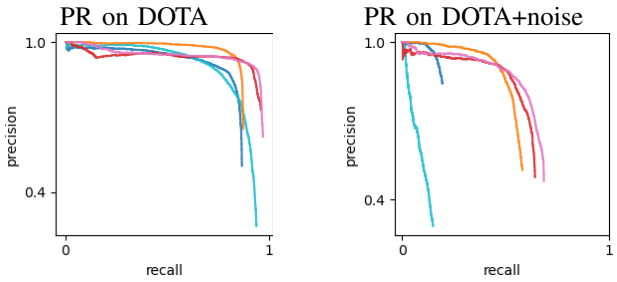

Fig. 2. Precision Recall (PR) curves on DOTA and DOTA+noise evaluation data, with each model colored as in Table I

b) Qualitative evaluation on ADS data: We evaluate the methods on data provided by ADS, at a 0.5 m resolution. As this data is not labeled, models are trained only on the benchmark data presented above. Results are presented in Fig. 4; Qualitatively, our PP model is able to produce regular configurations of vehicles, while missing fewer objects of interest compared to *BBA-Vec.*.

c) Inference interpretability: In Figure 5, we illustrate how the two components of the score s can help analyze the results: green objects correspond to detection with high prior and data scores, while blue detection have a higher data contribution. The few yellow detection correspond to objects with low data score, often located on ambiguous locations.

VI. Conclusion

Here we propose a novel scoring method for our model that utilizes Convolutional Neural Networks within a Point Process framework. This score allows to measure the detection confidence considering the object interactions. We show that, on top of allowing regularization and robustness on the resulting configuration of points, this enables some explainability of the results through the decomposition of the model intro multiple terms.

References

[1] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object Detection in 20 Years: A Survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276, Mar. 2023.

[2] 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 *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, virtual, Jan. 2021, pp. 2149–2158.

[3] Y. Li, Y. Xing, Z. Wang, T. Xiao, Q. Song, W. Li, and J. Wang, "A Framework of Maximum Feature Exploration Oriented Remote Sensing Object Detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023, feature pyramids; coarse location and refinement.

[4] H. Law and J. Deng, "CornerNet: Detecting Objects as Paired Keypoints," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, Sep. 2018, pp. 734–750.

[5] Q. Zeng, X. Ran, H. Zhu, Y. Gao, X. Qiu, and L. Chen, "Dynamic Cascade Query Selection for Oriented Object Detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, Aug. 2023.

[6] M.-C. V. Lieshout, *Markov Point Processes and Their Applications*. London: Imperial College Press, Jul. 2000.

[7] X. Descombes, "Multiple objects detection in biological images using a marked point process framework," *Methods*, vol. 115, pp. 2–8, Feb. 2017.

[8] T. Li, M. Comer, and J. Zerubia, "A Connected-Tube MPP Model for Object Detection with Application to Materials and Remotely-Sensed Images," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. Athens: IEEE, Oct. 2018, pp. 1323–1327.

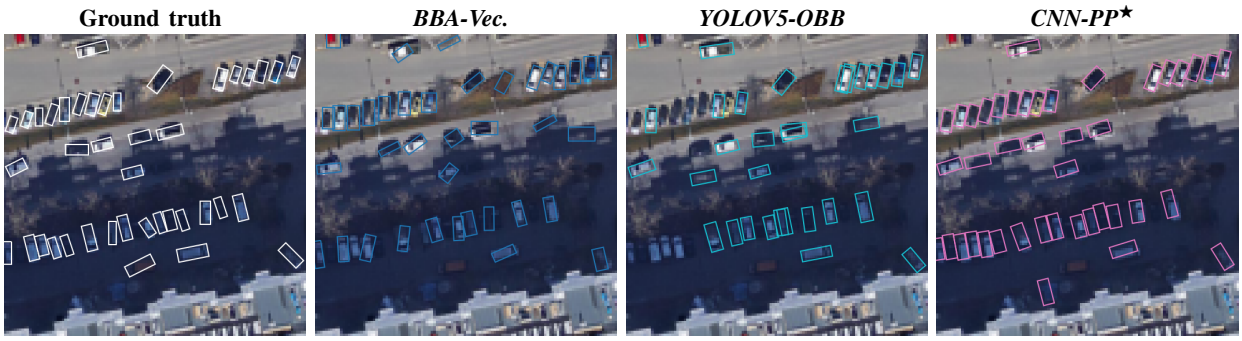

Fig. 3. Samples of detection on the test dataset. The score threshold (to not display low score objects) is set to maximize the $F1$ score for each model.

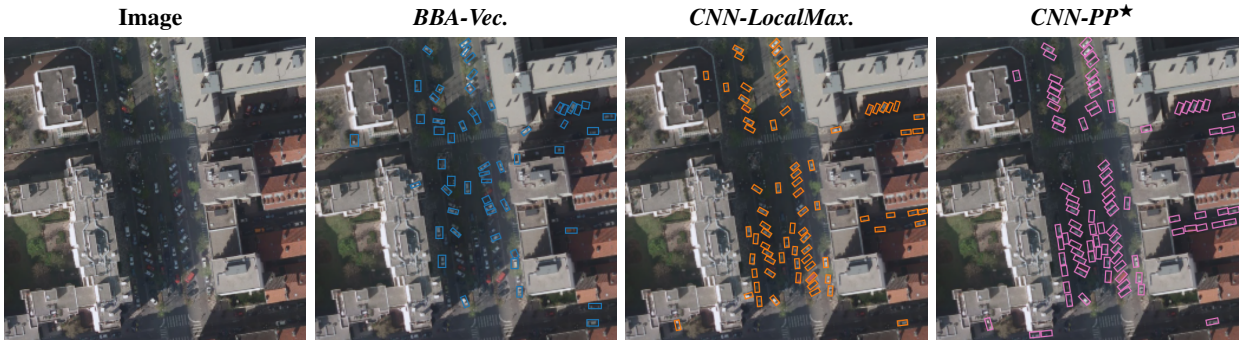

Fig. 4. Samples of detection on the ADS data. The dataset is not annotated. [© Airbus Defense and Space]

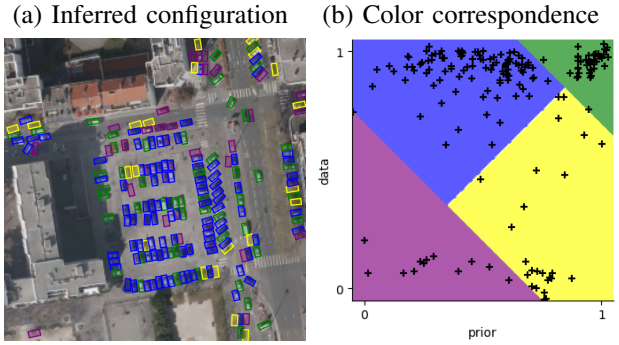

Fig. 5. Inferred configuration on an ADS data sample (a), colored according to their prior/data scores (b)(yellow: $s_{prior} > s_{data}$; blue: $s_{prior} < s_{data}$; purple: low s; green: high s). Each point in (b) corresponds to a detection in s_{data},s_{prior} space (log values scaled to $[0, 1]$).

[9] M. Ortner, X. Descombes, and J. Zerubia, "A Marked Point Process of Rectangles and Segments for Automatic Analysis of Digital Elevation Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 1, pp. 105–119, Jan. 2008.

[10] J. Mabon, M. Ortner, and J. Zerubia, "CNN-Based Energy Learning for MPP Object Detection in Satellite Images," in *2022 IEEE 32nd International Workshop on Machine Learning for Signal Processing (MLSP)*, Aug. 2022, pp. 1–6.

[11] Y. LeCun, S. Chopra, R. Hadsell, M. Ranzato, and F. J. Huang, "A Tutorial on Energy-Based Learning," *Predicting structured data*, p. 59, 2006.

[12] W. Grathwohl, K.-C. Wang, J.-H. Jacobsen, D. Duvenaud, M. Norouzi, and K. Swersky, "Your classifier is secretly an energy based model and you should treat it like one," in *International Conference on Learning Representations (ICLR)*, virtual, Sep. 2019.

[13] G. Hinton, "Training Products of Experts by Minimizing Contrastive Divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, Aug. 2002.

[14] Y. Du and I. Mordatch, "Implicit generation and modeling with energy based models," in *Advances in Neural Information Processing Systems (NeurIPS)*, H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Vancouver, Canada: Curran Associates, Inc., Dec. 2019.

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

[16] U. Grenander and M. I. Miller, "Representations of Knowledge in Complex Systems," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 56, no. 4, pp. 549–581, 1994.

[17] X. Yang and J. Yan, "On the Arbitrary-Oriented Object Detection: Classification Based Approaches Revisited," *International Journal of Computer Vision*, vol. 130, no. 5, pp. 1340–1365, May 2022.

[18] T. Li, M. Comer, and J. Zerubia, "An Unsupervised Retinal Vessel Extraction and Segmentation Method Based On a Tube Marked Point Process Model," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, May 2020, pp. 1394–1398.

[19] T. T. Pham, S. Hamid Rezatofighi, I. Reid, and T.-J. Chin, "Efficient Point Process Inference for Large-Scale Object Detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, Jun. 2016, pp. 2837–2845.

[20] J. Mabon, "Learning stochastic geometry models and convolutional neural networks. Application to multiple object detection in aerospatial data sets," PhD thesis, Université Côte d'Azur, Dec. 2023.

[21] G.-S. Xia, X. Bai, J. Ding, Z. Zhu, S. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang, "DOTA: A Large-Scale Dataset for Object Detection in Aerial Images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, USA, Jun. 2018, pp. 3974–3983.