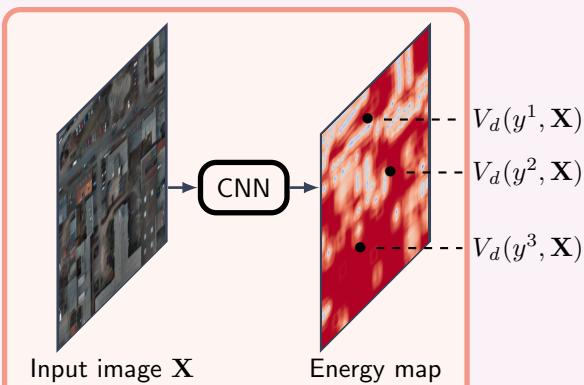


Learning Point Processes and Convolutional Neural Networks for Object Detection in Satellite Images

Performing object detection in remote-sensing images combining Convolutional Neural Network features with geometric priors on configurations within an Energy Based Model.

Energy map from a CNN (computed once per image)

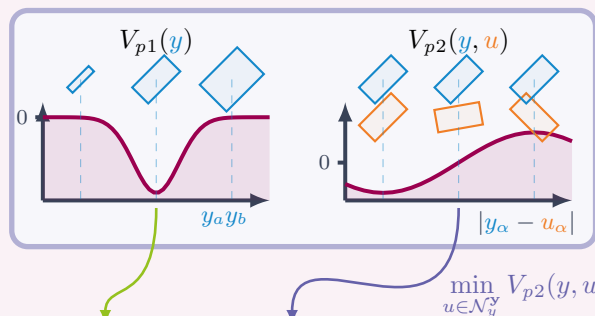


$V_d(y^1, \mathbf{X})$
 $V_d(y^2, \mathbf{X})$
 $V_d(y^3, \mathbf{X})$

Energy Based Model

measures coherence between image \mathbf{X} and configuration \mathbf{y}
 $p(\mathbf{y}|\mathbf{X}) \propto \exp(-U(\mathbf{y}, \mathbf{X}, \theta))$

Priors on configurations (area, size, alignment, overlap ...)



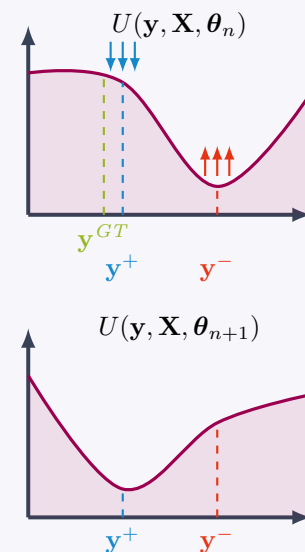
$\min_{u \in \mathcal{N}_y^y} V_{p2}(y, u)$

$$U(\mathbf{y}, \mathbf{X}, \theta) = \sum_{y \in \mathbf{y}} w_d V_d(y, \mathbf{X}) + w_{p1} V_{p1}(y) + w_{p2} V_{p2}(y, \mathcal{N}_y^y) + \dots$$

Data term **Prior term 1** **Prior term 2**
on point y on point y and neighborhood \mathcal{N}_y^y

Parameter Estimation

Finding θ such that
 $\mathbf{y}^{GT} = \arg \min(\mathbf{y}) U(\mathbf{y}, \mathbf{X}, \theta)$



Contrastive Divergence:

$\mathbf{y}^+ \sim \mathbf{y}^{GT} + \mathcal{N}(0, \sigma)$
 $\mathbf{y}^- \sim \exp(-U(\cdot, \mathbf{X}, \theta_n))$
 $\mathcal{L} = U(\mathbf{y}^+, \mathbf{X}, \theta_n) - U(\mathbf{y}^-, \mathbf{X}, \theta_n)$
 $\theta_{n+1} \leftarrow \theta_n + \nabla_{\theta_n} \mathcal{L}$

Sampling Configurations

Markov chain $\mathbf{y}_t \sim \exp\left(-\frac{U(\cdot, \mathbf{X}, \theta)}{T_t}\right)$
 $\mathbf{y}_t \xrightarrow{T_t \rightarrow 0} \arg \min_{\mathbf{y}} U(\mathbf{y}, \mathbf{X}, \theta)$

Jump

Birth: $\mathbf{y}' = \mathbf{y} \cup \{y\}, y \sim \exp(-V_d)$

Death: $\mathbf{y}' = \mathbf{y} \setminus \{y\}, y \sim \mathcal{U}(\mathbf{y})$

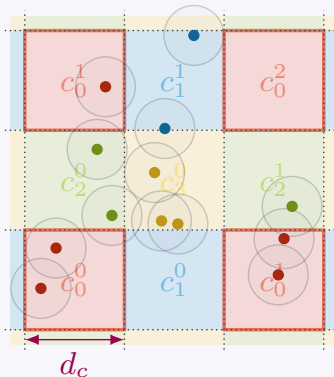
Diffusion

$\mathbf{y}_{t+1} \leftarrow \mathbf{y}_t + \gamma \nabla_{\mathbf{y}_t} U(\mathbf{y}_t, \mathbf{X}, \theta) + \mathcal{N}(0, 2\gamma T_t)$

Energy gradient Noise

Parallel sampling

Same color cells sampled in parallel



Results Interpretability

Novel scoring function
taking into account interactions

$$s(\mathbf{y}|\mathbf{y} \setminus \{y\}, \mathbf{X}) = s_{data}(\mathbf{y}|\mathbf{X}) s_{prior}(\mathbf{y}|\mathbf{y} \setminus \{y\})$$

detection score coherence w.r.t. the image internal coherence

