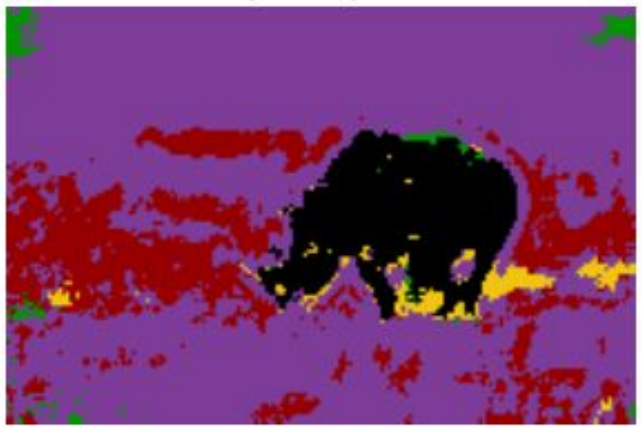
Multiscale Conditional Random Fields for Image Labeling

Simon Evain & Hugues Tavenard & Guillaume Hocquet

original

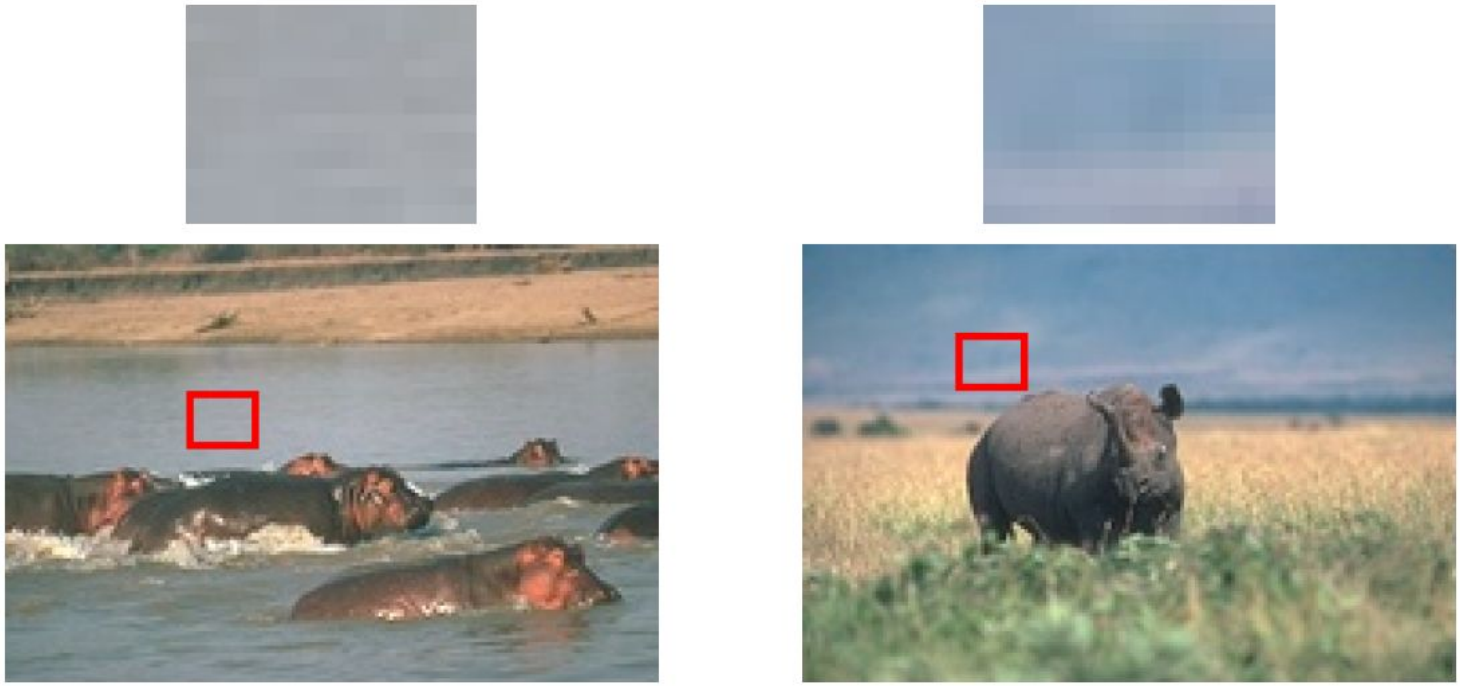
ground truth

initial (MLP) : 0.778



final : 0.949

General Idea



“Which one is water ? Which one is the sky ?”

- Local : Feature-based pixel-wise classification
- Regional & Global : Labeling-based region-wise classification

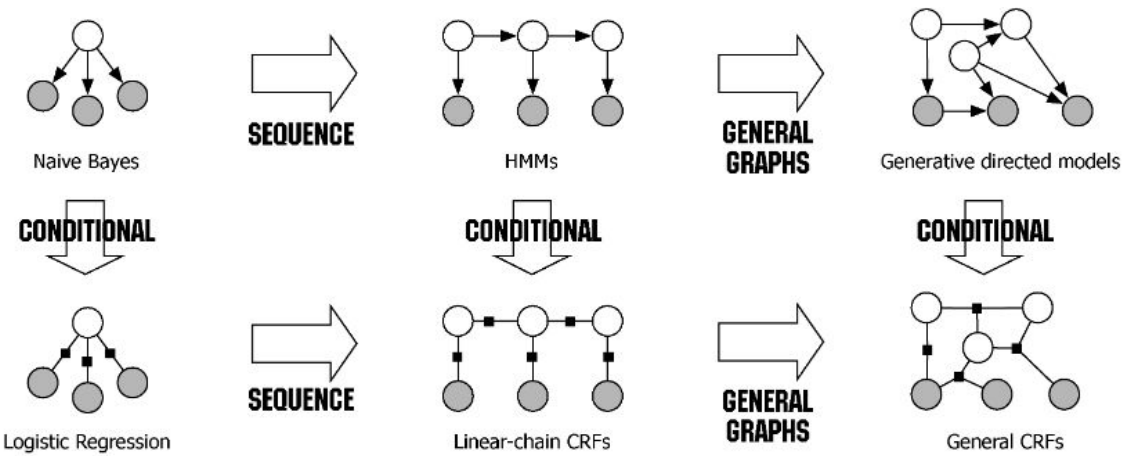
Conditional Random Fields

Definition. *Let $G = (V, E)$ be a graph such that $Y = (Y_v)_{v \in V}$ indexed by the vertices of G (X, Y) is a Conditional Random Field if, when conditioned on X, the random variables Y_v follow the Markov property with respect to the graph :*

$$p(Y_v|X, Y_w, w \neq v) = p(Y_v|X, Y_w, w \sim v)$$

where $w \sim v$ means w and v are neighbors in G. The conditional distribution can be written :

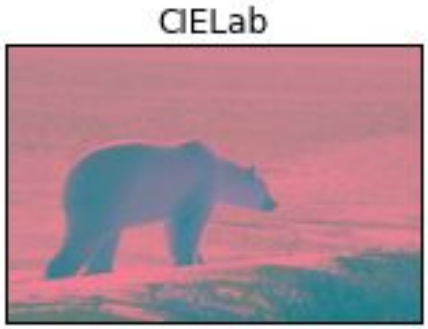
$$p(y|x) = \frac{1}{Z(x)} \prod_{c \in C} \Psi(y_c, x_c)$$

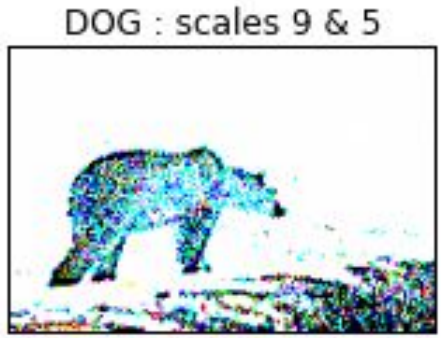


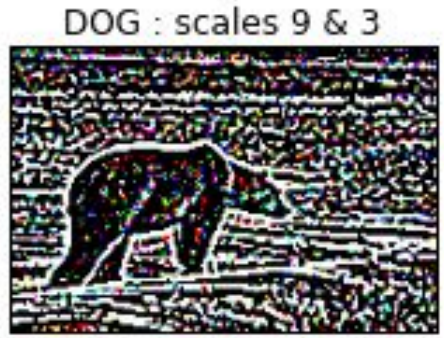
Relation between Conditional Random Fields and other graphical models

Image Features

Features : Color (CIE Lab), Edge and Texture (Difference-of-Gaussians and Gabor filters)



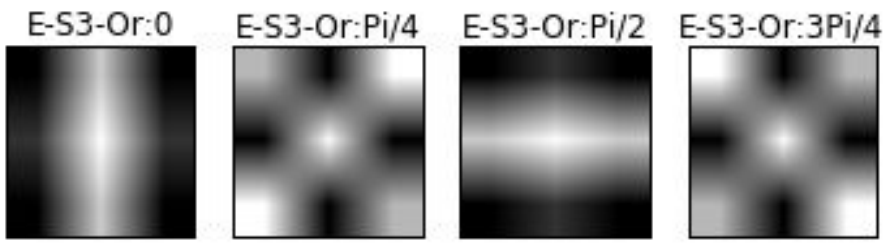


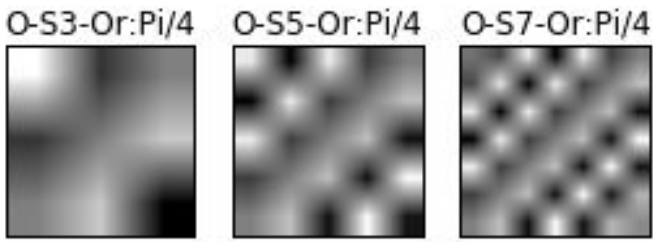


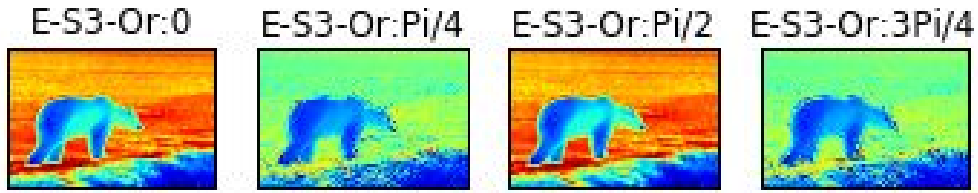
CIELab color space

Difference of Gaussian filters

Gabor filters



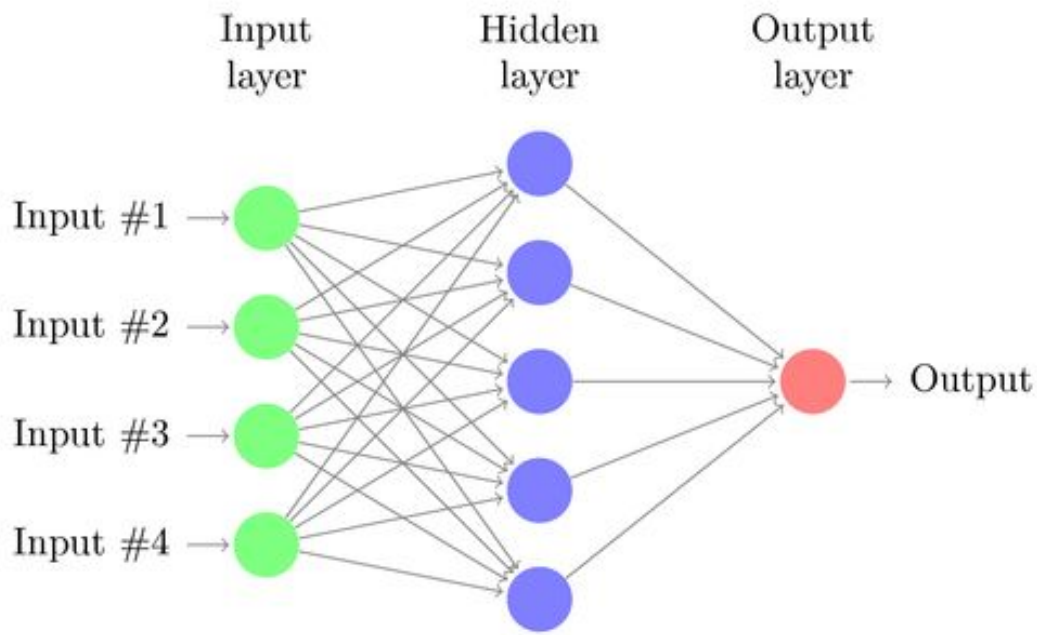




Multilayer Perceptron as a local classifier

One hidden layer MLP trained to minimize the cross entropy by stochastic gradient descent on **118 features**.

$$Loss(\hat{y}, y, W) = -y \ln \hat{y} - (1 - y) \ln (1 - \hat{y}) + \alpha ||W||_2^2$$



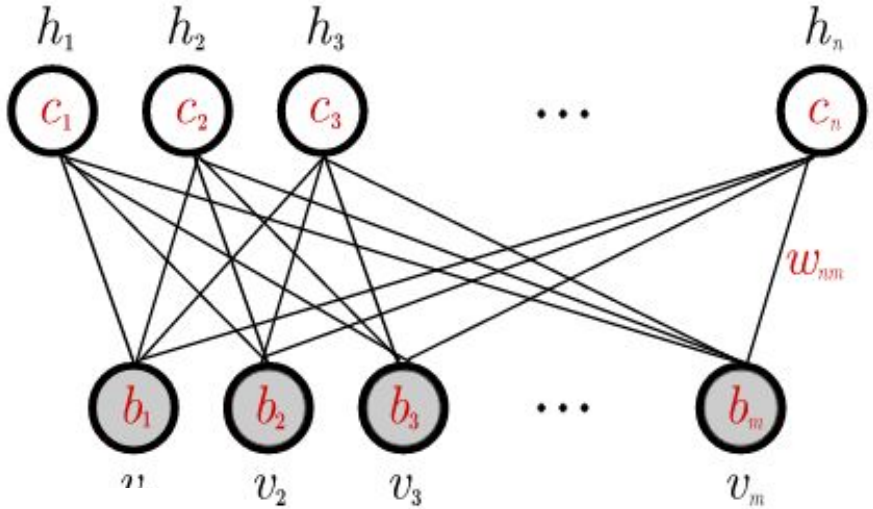
65% accuracy on one dataset, 88% accuracy on the other dataset.

Training is long due to the size of the input.

Restricted Boltzmann Machine as regional classifier

Graphical model with **bipartite undirected graph**
composed of visible and hidden units

$$p(v) = \frac{1}{Z} \prod_{j=1}^{m} \prod_{i=1}^{n} \left(1 + \exp \left(\sum_{j=1}^{m} w_{ij} v_j \right) \right)$$



Trained by **Gradient Descent** using Gibbs Sampling
to maximize the Likelihood :

$$\frac{1}{T} \sum_{v} \frac{\partial \ln \mathcal{L}(W|v)}{\partial w_{ij}} = \langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model}$$

Approximation with **Contrastive Divergence** (only 1 step of sampling)

Inference

$$P(L|X;\theta) = \frac{1}{Z(X)} \prod_i \underbrace{P(l_i|x_i,\lambda)}_{\text{MLP}} \times \prod_{r,a} \underbrace{(1+\exp(w_a^T l_r))}_{\text{RBM}} \times \prod_b \underbrace{(1+\exp(u_b^T L))}_{\text{RBM}}$$

Criterion to maximize : **Maximum Posterior Marginals** (MPM) estimate

$$l_i^* = \arg\max_{l_i} P(l_i|X) \quad \forall i$$

But : Probability of a pixel depends on labeling of other pixels
Solution : Use **Gibbs sampling** on the labels of each pixel one by one

Initialize Gibbs sample with output of MLP

Results

The algorithm can sometimes improve significantly on the MLP but it can also lose details.
At each gibbs sampling iteration, the level of details decreases, meaning that the accuracy can decrease.
When the MLP is wrong, the algorithm cannot improve the result very much.
One step of Gibbs sampling can take up to 35 seconds.