Combination of MRF and ShapeBM for Image Labeling

Alexander Kirillov, Dmitry Vetrov

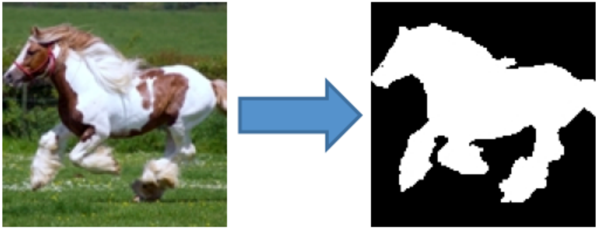
arhipisk@gmail.com

2013

Table of Contents

Problem definition

Given: image and information about the type of object this image contains (horse, car, etc.).
Required: two classes pixels labeling — object and background.



State-of-art method: Markov Random Fields (MRF)

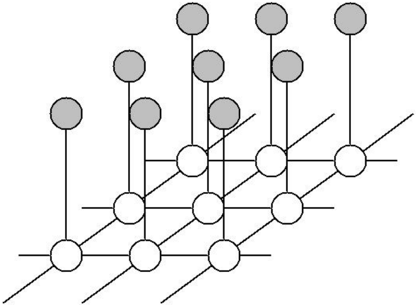
Markov Random Fields

Let $G(\mathbf{v}, \mathbf{e})$ be a graph, $\mathbf{y} \in \{0, 1\}^{|\mathbf{v}|}$ be a vector of labels (0 – background, 1 – object).

Pairwise MRF Energy:

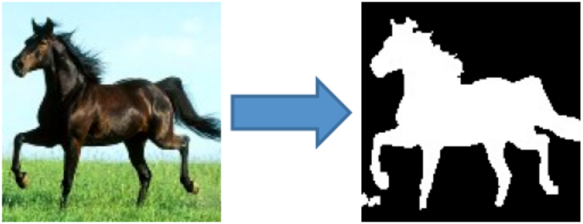
$$E(\mathbf{y}) = \sum_{v \in \mathbf{v}} \phi_v(y_v) + \sum_{(v,u) \in \mathbf{e}} \phi_{vu}(y_v, y_u). \tag{1}$$

The goal is to find $y^* = \operatorname{argmin}_{\mathbf{y}} E(\mathbf{y})$.



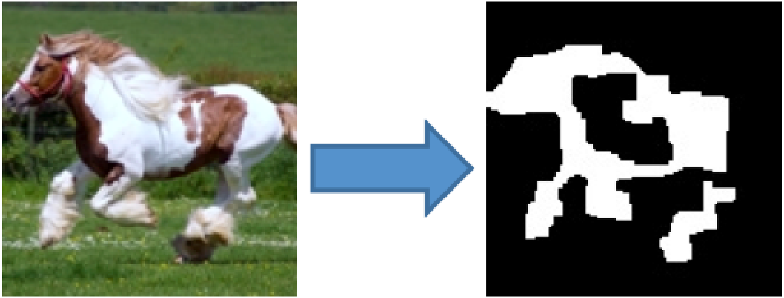
MRF advantages

- Unary potentials $\phi_i(v_i)$ are defined by the colour model which is tuned on images from labeled dataset. They define probability of each pixel being a part of object or background
- Binary potentials $\phi_{i,j}(v_{i,j})$ penalize labeling where a boundary between an object and a background lays across similar pixels.
- Effective inference method — graph cuts



MRF disadvantages

- MRF doesn't take into account the shape of the object
- Only local interdependencies: cannot use global constraints, e.g. shape



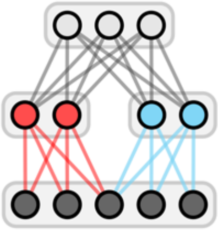
ShapeBM [S. M. Ali Eslami et al. 2012]

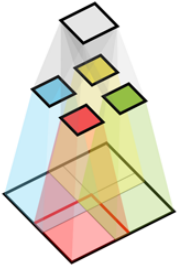
Shape Boltzmann Machine (ShapeBM) — a type of Deep Boltzmann Machine with some weights set to zero.

Energy function:

$$E(v, h^1, h^2) = -v^T W^1 h^1 - (h^1)^T W^2 h^2 - b_v^T v - b_{h^1}^T h^1 - b_{h^2}^T h^2$$

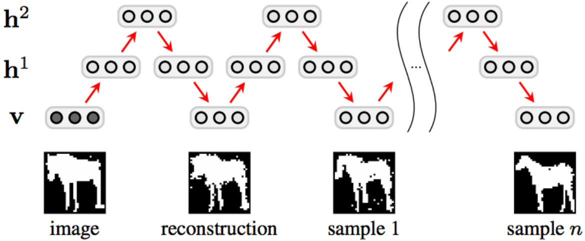
v – visible variables (labeling), h^1, h^2 – first and second hidden layers variables.





ShapeBM

ShapeBM is a model of an object shape, which allows to reconstruct the object from a binary image with an uncompleted or vague object.



ShapeBM also estimates the probability of a binary image resembling a horse.

ShapeBM — pros and cons

+ Stores probability distribution on shapes which is tuned to a train dataset
+ Generates new objects different from a training data
— Not applicable for high resolution images. For 32×32 image the number of nodes is approx. 3000
— Doesn't take into account the interdependences between neighbour pixels. The boundary between the object and the background may lay across two pixels of same color

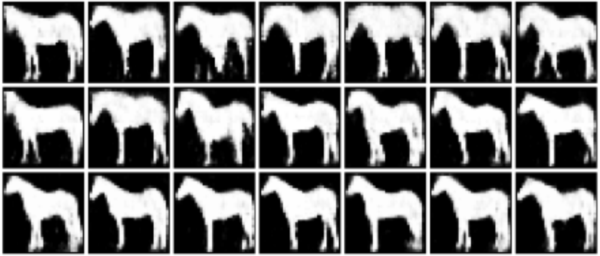


Table of Contents

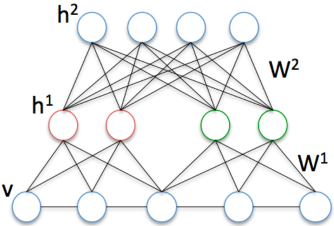
MRF + ShapeBM

The main objective is to build model which can label image containing certain predefined shape with help of local interdependences.

Energy function:

$$E(v, h^1, h^2) = E_{MRF}(v) + \gamma E_{ShapeBM}(v, h^1, h^2) \tag{2}$$

Combination of MRF and ShapeBM:



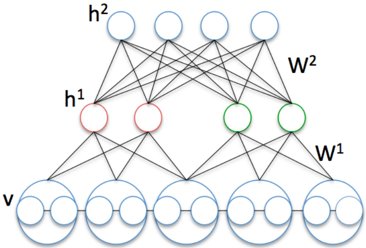
[Fei Chen et al. CVPR 2013]

Different resolution

We use ShapeBM for a coarse shape prior, so low-resolution ShapeBM is enough.
Energy function:

$$E(v, h^1, h^2) = E_{MRF}(v) + \gamma E_{ShapeBM}(v^s, h^1, h^2) \qquad (3)$$

where $v_i^s = \frac{\sum_{j \in fileds(i)} v_j}{|fields(i)|}$.
Combination of high-resolution MRF and low-resolution ShapeBM:



MRF + ShapeBM inference

- During experiments we use per-tuned ShapeBM [S. M. Ali Eslami et al. 2012]. It is pre-tuned on Weismann horse dataset for centered and scaled shapes with 32×32 resolution.
- Unary potentials $\phi_i(\theta, v_i)$ are defined by the colour model – mixtures of Gaussians on RGB and LUV representations of images for object and background (tuned on Weismann dataset subsample).
- Binary potentials $\phi_{i,j}(v_i, v_j) = \exp(-\|v_i - v_j\|^2_{RGB})$

EM algorithm for inference

Maximize $p(\theta, v, h^1, h^2) = \exp(-E(\theta, v, h^1, h^2))$

- E-step:
  Variational approximation $q(h^1, h^2) = \prod_i q_i(h^1_1) \cdot \prod_i q_i(h^2_i)$
  with fixed v and θ:

$$KL(q(h^1, h^2) \| p(h^1, h^2 | v, \theta) \rightarrow \min_{q(h^1, h^2)} .$$

  The parameters are recalculated iteratively as follows:

$$h^1_j = \frac{1}{1 + \exp(-\sum_i W^1_{ij} v^s_i - \sum_k W^2_{jk} h^2_k)},$$

$$h^2_k = \frac{1}{1 + \exp(-\sum_j W^2_{jk} h^1_j)},$$

$$h^1_j = \frac{1}{1 + \exp(-\sum_i W^1_{ij} v^s_i - \sum_k W^2_{jk} h^2_k)}$$

EM algorithm for inference

- M_1-step:

$$\mathbb{E}_{q(h^1,h^2)} E(\theta, v, h^1, h^2) \to \min_{v}$$

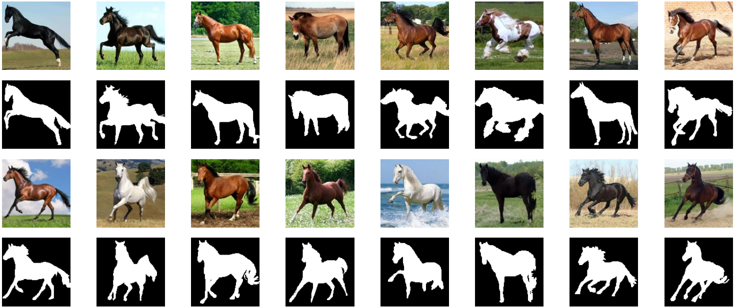
- M_2-step:
  On this step we adjust the colour model to specific input image:

$$\mathbb{E}_{q(h^1,h^2)} E(\theta, v^*, h^1, h^2) \to \min_{\theta}$$

  The v^* is resulted labeling from the M_1 step.

Test sample

As a testing dataset we created a new set with 16 images with 128×128 resolution.
Inference quality measure — mean weighted Hamming distance on our dataset.



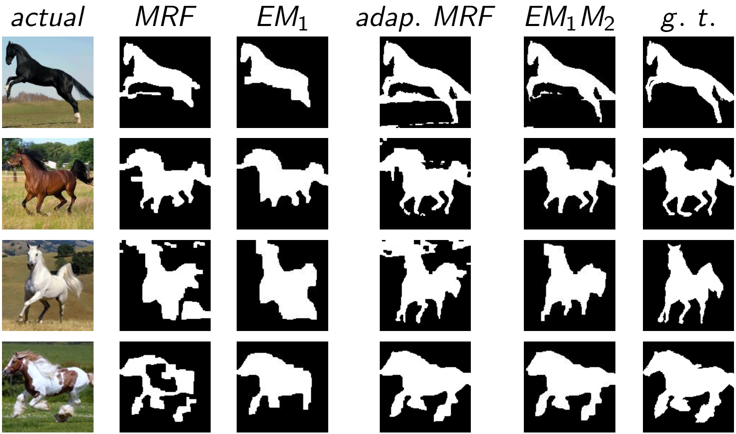
Experiments results

Four inference schemes were compared:

- *MRF* with fixed colour model
- *EM*-algorithm, which uses steps E and M_1
- adaptive *MRF*: re-tuning of colour model after each iteration using new labeling
- *EM*-algorithm, which uses steps E, M_1 and M_2

MRF	EM_1	adaptive *MRF*	$EM_1 M_2$
0.3247	0.2883	0.1790	0.1585

Experiments results



MRF + ShapeBM inference

Energy function:

$$E(v, h^1, h^2) = E_{MRF}(v) + \gamma E_{ShapeBM}(v, h^1, h^2) \tag{4}$$

Minimization of this energy is equivalent to the following system:

$$\min_{v_1} E_{MRF}(v^1) + \gamma \sum_{t \in T} \min_{v_t^2, h_t^1, h_t^2} E_{ShapeBM}(v_t^2, h_t^1, h_t^2) \tag{5}$$

$$s.t. \qquad v^1 = v^2 \tag{6}$$

$$v_t^2 = v^2, t \in T \tag{7}$$

$$h_t^1 = h^1 \in T \tag{8}$$

$$h_t^2 = h^2, t \in T \tag{9}$$

Where T — is a set of subgraphs.

MRF + ShapeBM inference

Using dual decomposition:

$$\min_{v,h^1,h^2}(E(v,h^1,h^2)) \geq \min_{v^1}(E_{MRF}(v^1) + \lambda_0^v v^1) +$$

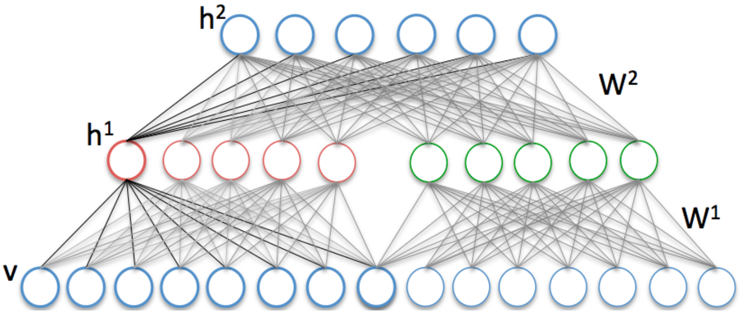
$$+\gamma \sum_{t\in T} \min_{v_t^2,h_t^1,h_t^2}(E_{ShapeBM}(v_t^2,h_t^1,h_t^2) + \lambda_t^v v_t^2 + \lambda_t^{h^2} h_t^2) =$$

$$= Q(\gamma,\lambda_0^v,\lambda_t^v,\lambda_t^{h^2},v^1,v_t^2,h_t^1,h_t^2) \quad (10)$$

$$s.t \qquad \sum_{t\in T}\lambda_t^v + \lambda_0^v = 0 \quad (11)$$

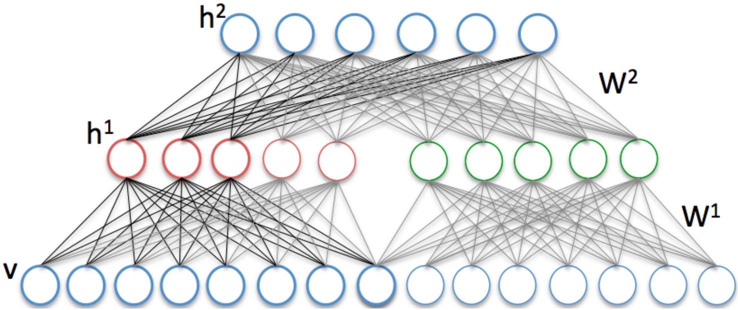
$$\sum_{t\in T}\lambda_t^{h^2} = 0 \quad (12)$$

Subgraph



+ BP for inference
— A huge number of Lagrangian Coefficients

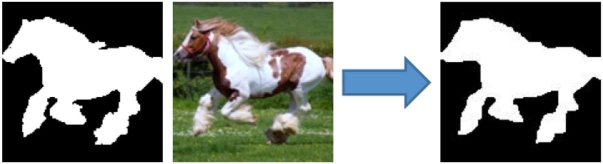
Subgraph



Number of nodes from middle layer in each subgraph is a trade-off between a speed of inference and a number of Lagrangian Coefficients.

Combination benefits

- Train MRF and SBM models independently since they are responsible for different types of dependencies
- Combine high-resolution MRF with low-resolution SBM
- Use dual decomposition framework from inference by splitting SBM into subtrees



Open issue

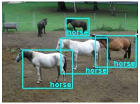
The task is to label arbitrary images of objects using ShapeBM as a shape prior which is tuned on the shapes:

- centered in the middle of the image,
- in the same direction (e.g. horses facing left),
- uniformly scaled relative to the size of the image.

Future work

- Tune ShapeBM to “remember” $2.5D$ shapes, where each pixel has an additional depth value;

- Improve object detector methods using our model;



- Tune ShapeBM for an multi-label segmentation problem [Kae A., Sohn K. et al. 2013].

Table of Contents

References

- Eslami, S.M., Heess, N., Winn, J. The Shape Boltzmann Machine: a Strong Model of Object Shape. CVPR, 2012.
- Chena F. et al. Deep Learning Shape Priors for Object Segmentation. CVPR, 2013
- Kae A., Sohn K. et al. Augmenting CRFs with Boltzmann Machine Shape Priors for Image Labeling. CVPR, 2013