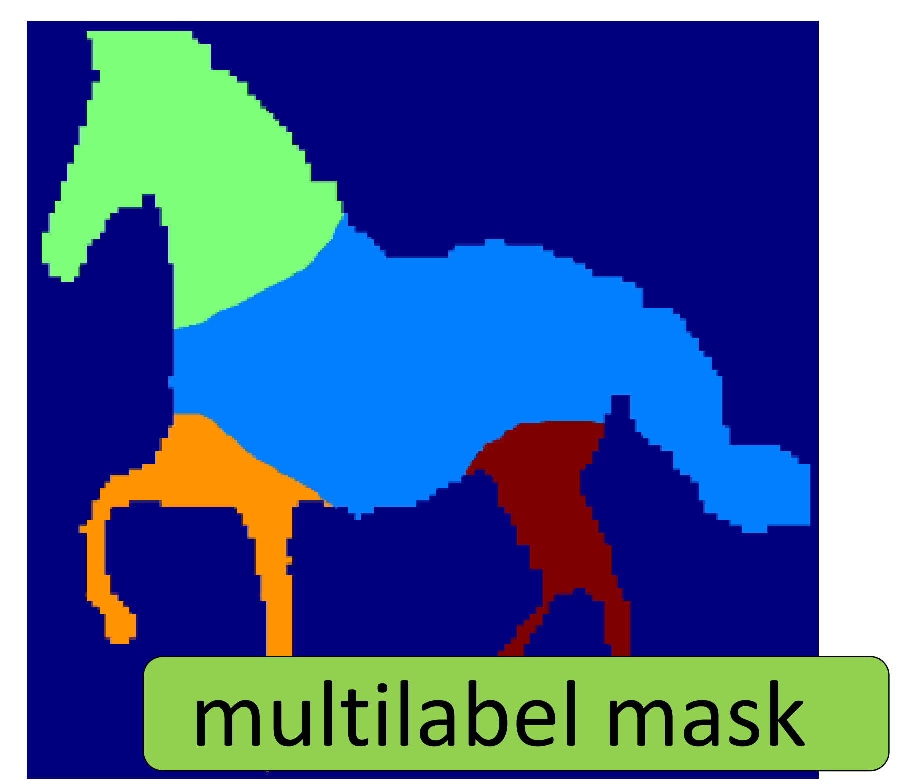
Deep Part-Based Generative Shape Model with Latent Variables

Alexander Kirillov, Mikhail Gavrikov, Ekaterina Lobacheva
Anton Osokin, Dmitry Vetrov

Moscow
2016

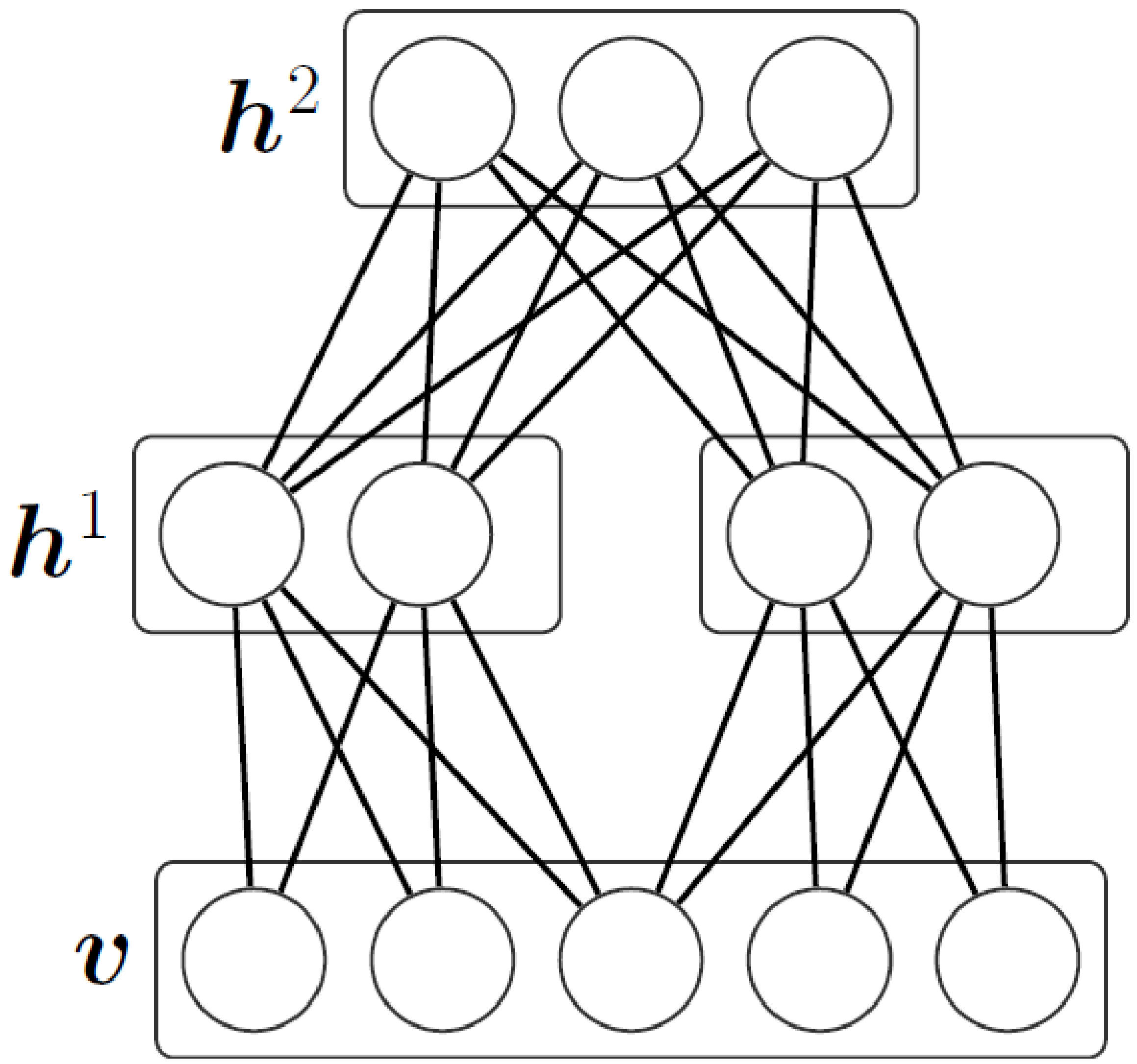
Shape Models

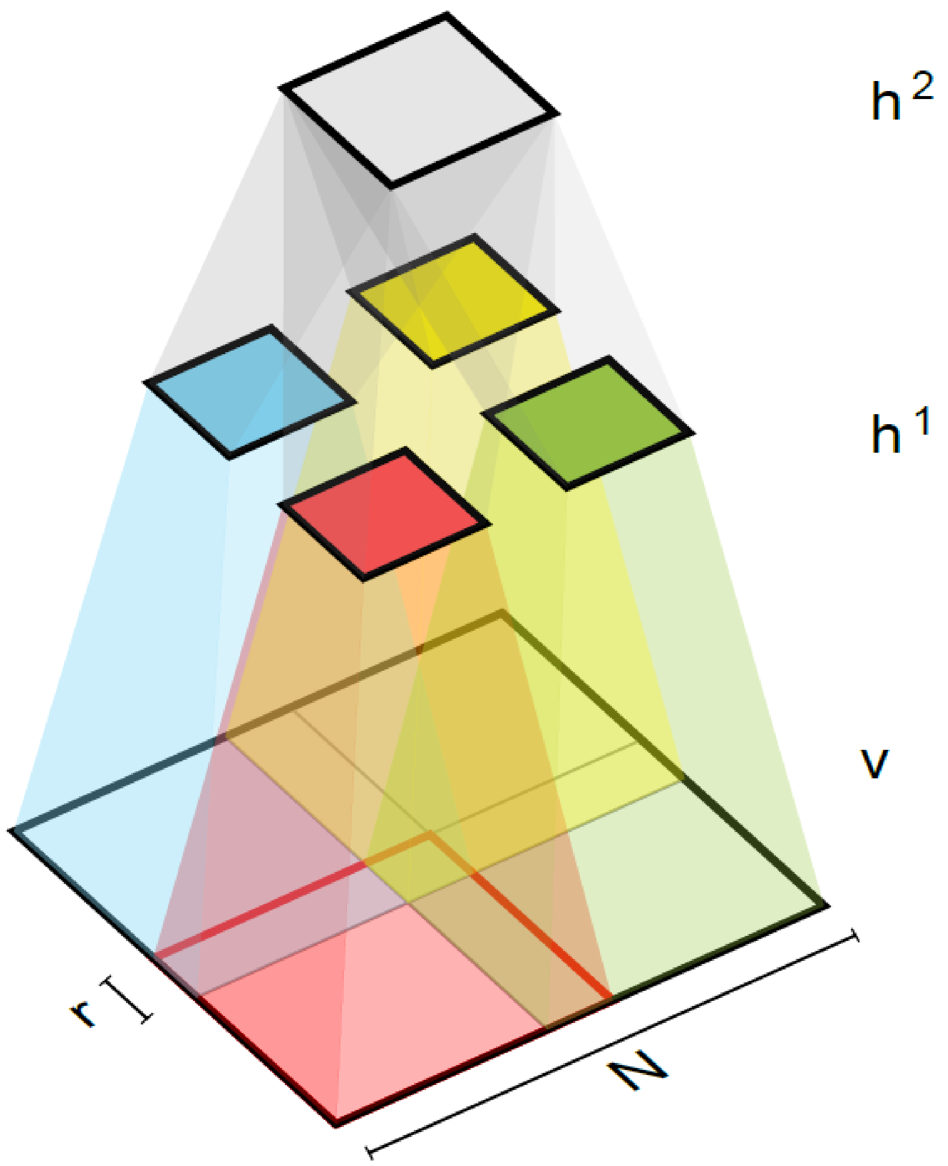


Useful for:

- segmentation,
- inpainting,
- detection,
- …

Shape Boltzmann Machine (SBM)

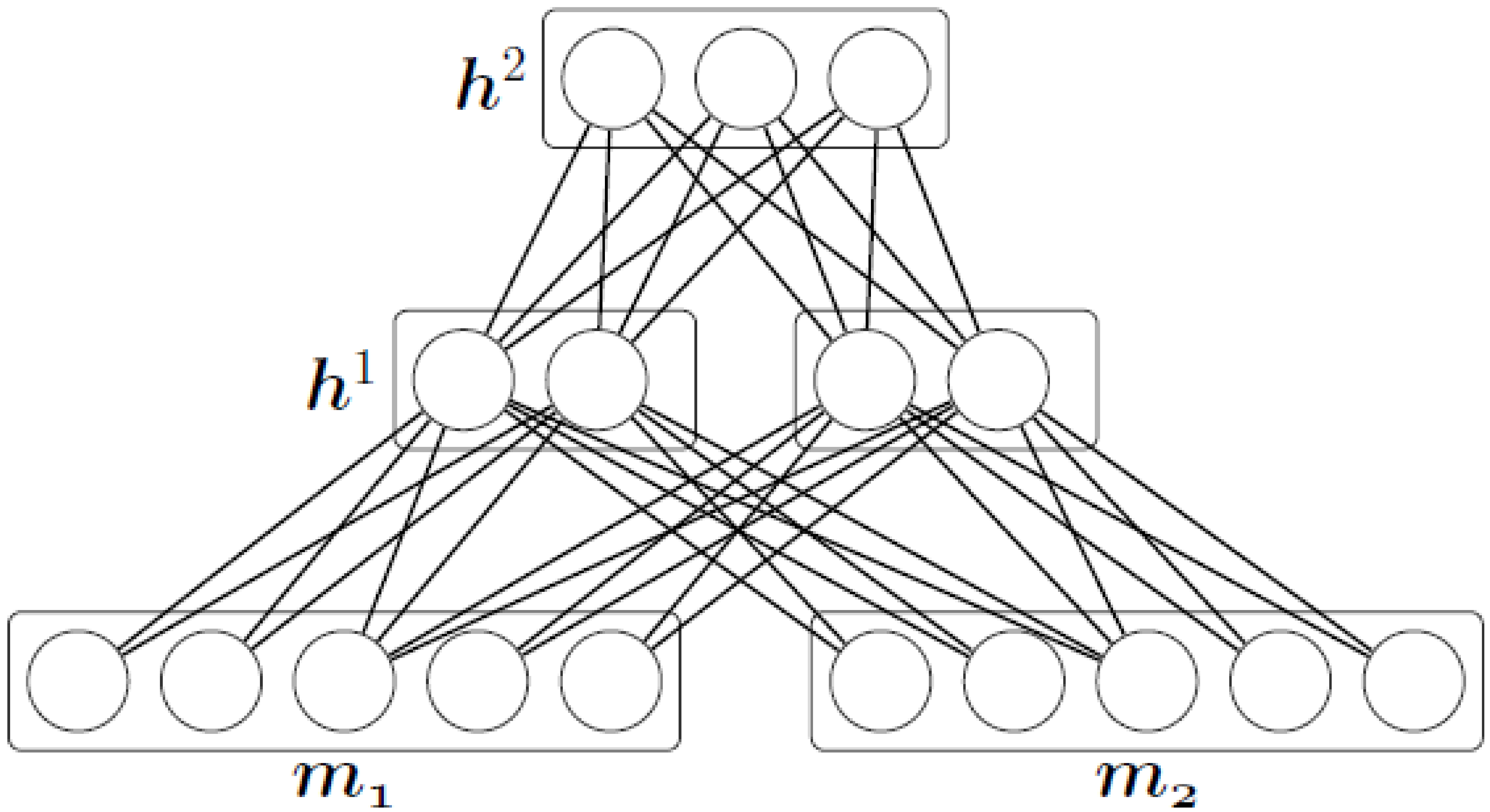




$$p(\boldsymbol{b}, \boldsymbol{h}^1, \boldsymbol{h}^2 \mid \boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \exp\left(-E(\boldsymbol{b}, \boldsymbol{h}^1, \boldsymbol{h}^2 \mid \boldsymbol{\theta})\right)$$

$$E(\boldsymbol{b}, \boldsymbol{h}^1, \boldsymbol{h}^2 \mid \theta) = \boldsymbol{a}^T\boldsymbol{b} + \boldsymbol{b}^T W^1 \boldsymbol{h}^1 + \boldsymbol{c}^{1T}\boldsymbol{h}^1 + \boldsymbol{h}^{1T} W^2 \boldsymbol{h}^2 + \boldsymbol{c}^{2T}\boldsymbol{h}^2$$

Multinomial SBM (MSBM)



$$p(\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2 \mid \boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \exp\left(-E(\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2 \mid \boldsymbol{\theta})\right)$$

$$E(\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2 \mid \theta) = \sum_{p=1}^{P} \boldsymbol{a}_p^T \boldsymbol{m}^p + \sum_{p=1}^{P} \boldsymbol{m}^{pT} W_p^1 \boldsymbol{h}^1 + \boldsymbol{c}^{1T} \boldsymbol{h}^1 + \boldsymbol{h}^{1T} W^2 \boldsymbol{h}^2 + \boldsymbol{c}^{2T} \boldsymbol{h}^2$$

Training of SBM and MSBM

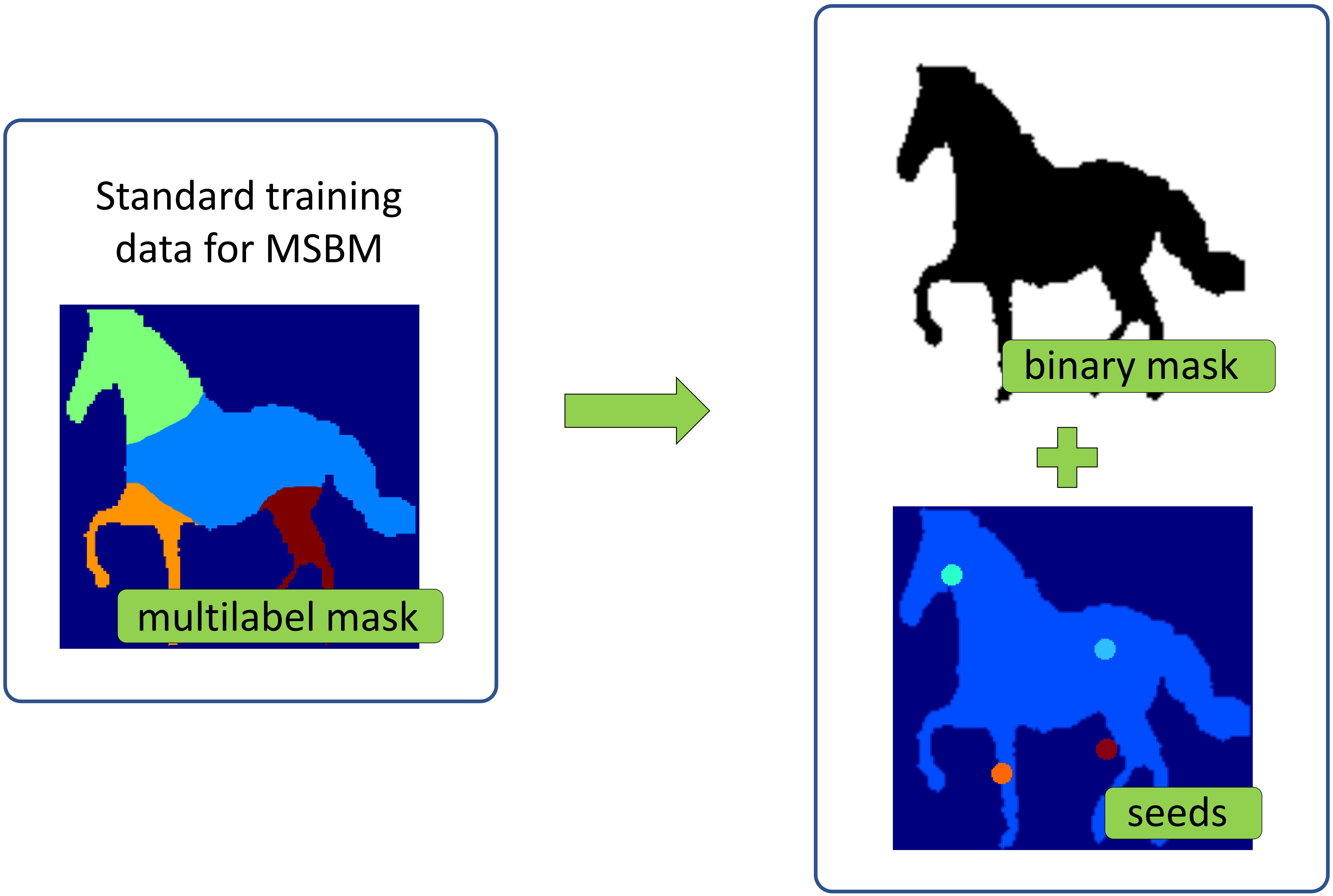
Variational EM-algorithm

$$\log p(\boldsymbol{b}/\boldsymbol{m} \mid \boldsymbol{\theta}) \to \max_{\theta}$$

- ***b/m* –** observed variable
- ***h*¹, *h*² –** hidden variables
- $\boldsymbol{\theta}$ – MSBM parameters

We need fully annotated data!

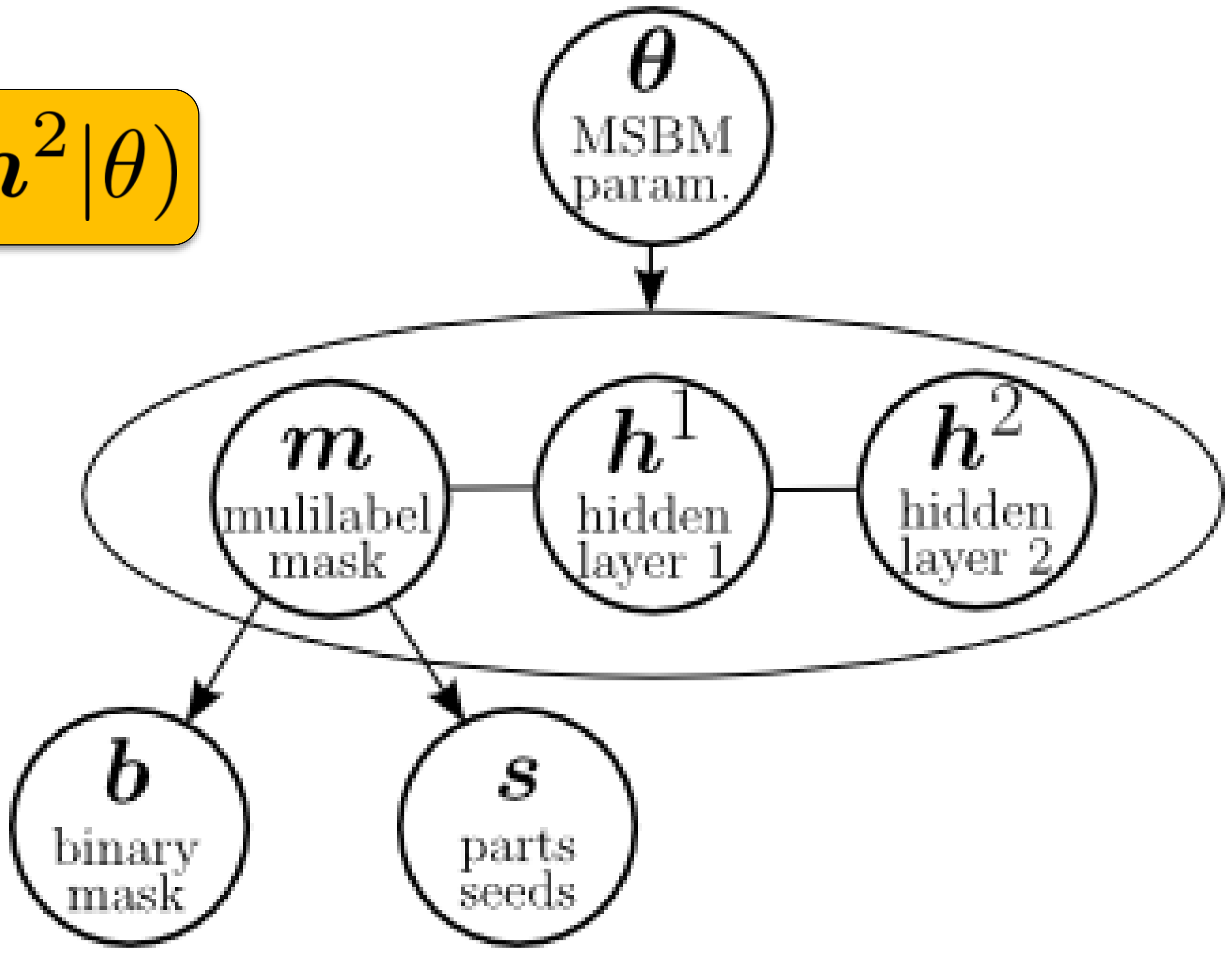
Our model


Standard training
data for MSBM
multilabel mask
binary mask
seeds

Our model

$$p(\boldsymbol{b}, \boldsymbol{s}, \boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2|\theta) = p(\boldsymbol{b}|\boldsymbol{m})\, p(\boldsymbol{s}|\boldsymbol{m})\, p(\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2|\theta)$$

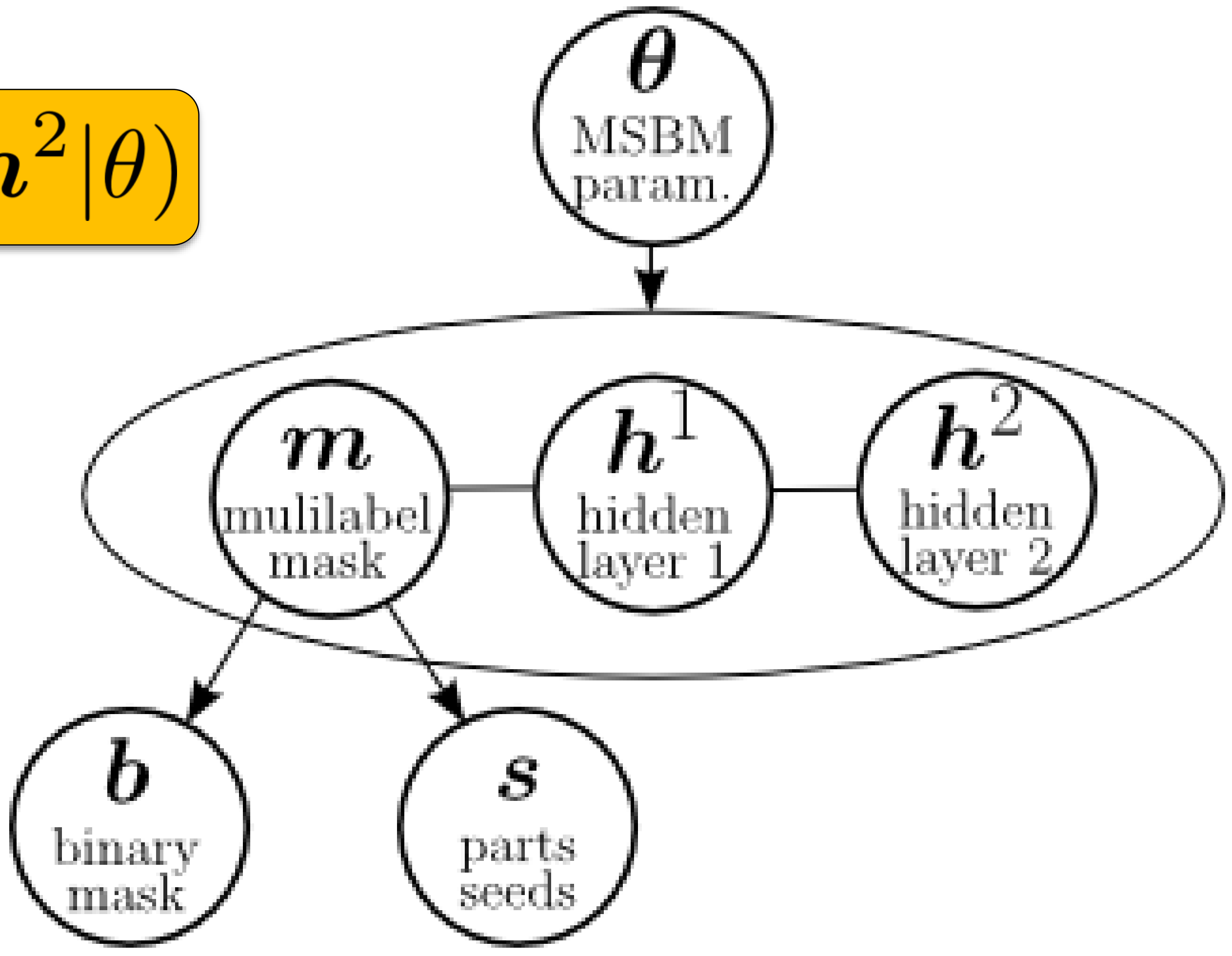
Variables: $\boldsymbol{b}$, $\boldsymbol{s}$ are observed, $\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2$ are hidden



Our model

$$p(\boldsymbol{b}, \boldsymbol{s}, \boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2 | \theta) \; = \; p(\boldsymbol{b}|\boldsymbol{m})\, p(\boldsymbol{s}|\boldsymbol{m})\, p(\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2 | \theta)$$

Variables: $\boldsymbol{b}$, $\boldsymbol{s}$ are observed, $\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2$ are hidden



$$p(\boldsymbol{b}|\boldsymbol{m}) = \prod_i p(b_i|m_i)$$
$$= \prod_i \left([b_i = 0][m_i = 0] + [b_i \neq 0][m_i \neq 0]\right)$$

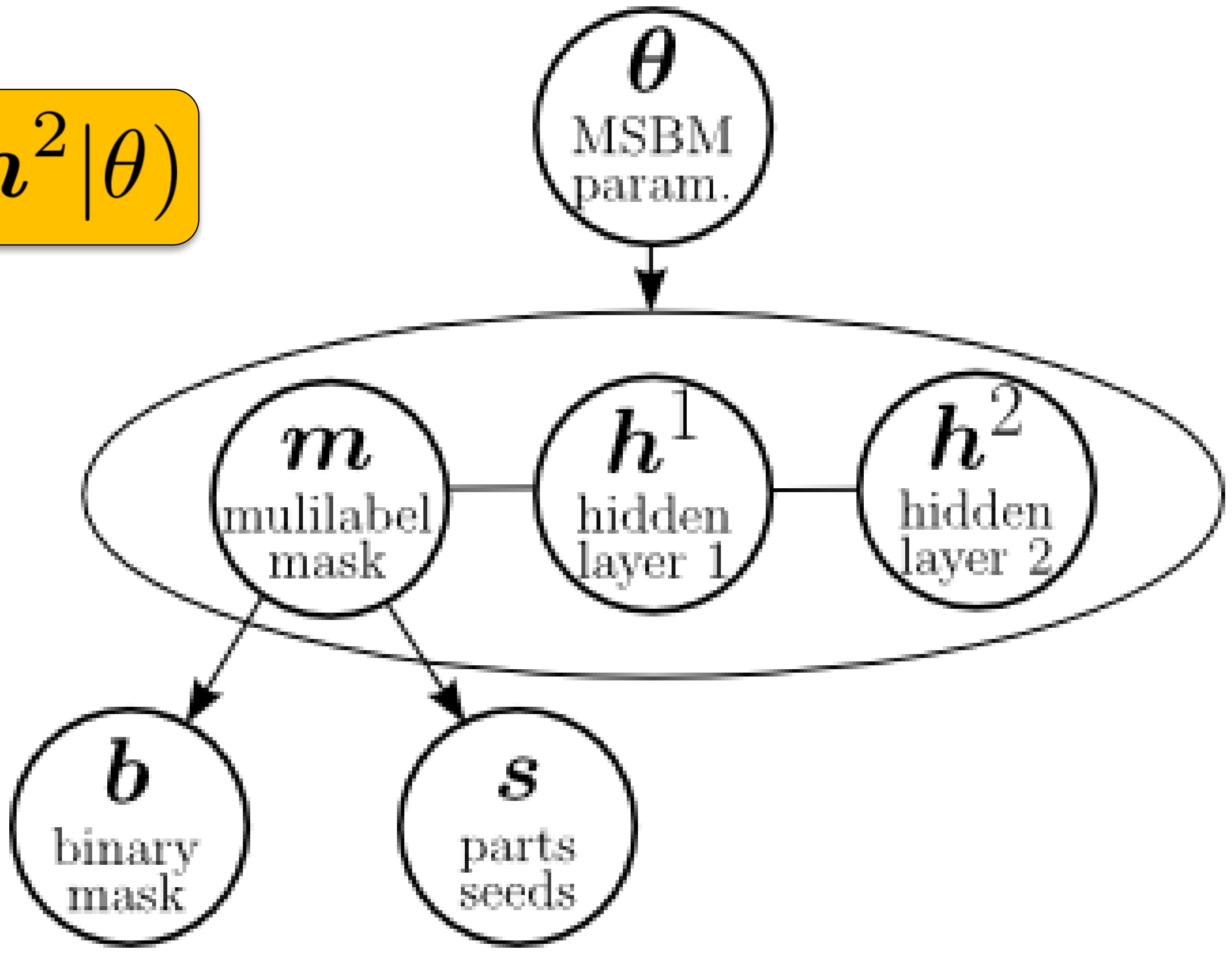
$\Leftrightarrow$ $m_i \neq 0 \Rightarrow b_i = 1$
$m_i = 0 \Rightarrow b_i = 0$

If a pixel belongs to any part of an object, then it belongs to the object with probability 1, otherwise this pixel belongs to the background.

Our model

$$p(\boldsymbol{b}, \boldsymbol{s}, \boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2 | \theta) = p(\boldsymbol{b}|\boldsymbol{m})\, p(\boldsymbol{s}|\boldsymbol{m})\, p(\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2 | \theta)$$

Variables: $\boldsymbol{b}$, $\boldsymbol{s}$ are observed, $\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2$ are hidden

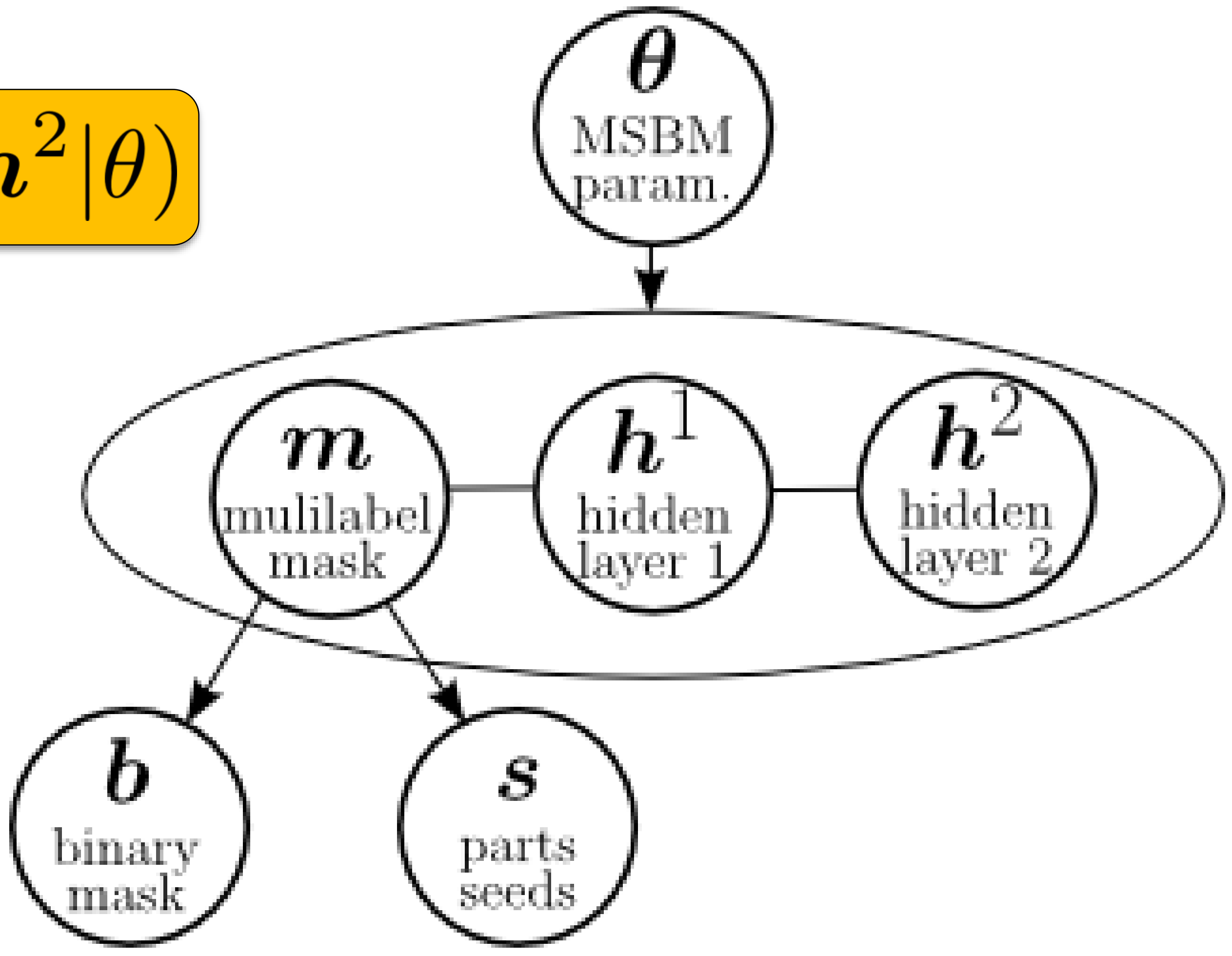


$$p(s|m) \propto \prod_{i: m_i \neq 0} \mathcal{N}\left(s_{m_i} | f_{coord}(i), \sigma^2\right)$$

- Each pixel impacts the seeds independently
- Background pixels do not impact the seeds
- Each pixel draws the corresponding seed closer

Our model

$$p(\boldsymbol{b}, \boldsymbol{s}, \boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2 | \theta) \;=\; p(\boldsymbol{b}|\boldsymbol{m})\, p(\boldsymbol{s}|\boldsymbol{m})\, p(\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2 | \theta)$$



Variables: $\boldsymbol{b}$, $\boldsymbol{s}$ are observed, $\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2$ are hidden

$$p(\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2 \mid \boldsymbol{\theta}) = \exp\left(-E_{\mathrm{MSBM}}(\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2 \mid \boldsymbol{\theta})\right) / Z(\boldsymbol{\theta})$$

Probability model for MSBM

Training: Variational EM-algorithm

$$\log P(B, S \mid \boldsymbol{\theta}) = \sum_{d=1}^{D} \log p(\boldsymbol{b}^d, \boldsymbol{s}^d | \boldsymbol{\theta}) \to \max_{\theta}$$

E-step:

$$\min_{q^d} \mathrm{KL}\left(q^d(\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2) \| p(\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2 | \boldsymbol{b}^d, \boldsymbol{s}^d, \boldsymbol{\theta})\right)$$

s.t. $q^d(\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2)$ – from fully factorized family

Variational inference:

$$q_i^d(m_i = p) \propto [b_i^d \neq 0][m_i \neq 0] \exp\left(-\frac{1}{2\sigma^2} \left\| s_{m_i}^d - f_{coord}(i) \right\|_2^2 + a_{i,m_i} + \sum_j W_{i,j,m_i}^1 q_j(h_j^1 = 1)\right) + [b_i^d = 0][m_i^d = 0],$$

$$q_j^d(h_j^1 = 1) = \sigma(c_j^1 + \sum_{i,p=1}^{P} q_i(m_i = p) W_{i,j,p}^1 + \sum_k W_{j,k}^2 q_k(h_k^2 = 1)),$$

$$q_k^d(h_k^2 = 1) = \sigma(c_k^2 + \sum_j q_j(h_j^1 = 1) W_{j,k}^2)$$

Standard MSBM training procedure

Training: Variational EM-algorithm

$$\log P(B, S \mid \boldsymbol{\theta}) = \sum_{d=1}^{D} \log p(\boldsymbol{b}^d, \boldsymbol{s}^d | \boldsymbol{\theta}) \to \max_{\theta}$$

E-step:

$$\min_{q^d} \mathrm{KL}\left(q^d(\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2) \| p(\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2 | \boldsymbol{b}^d, \boldsymbol{s}^d, \boldsymbol{\theta})\right)$$

s.t. $q^d(\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2)$ – from fully factorized family

M-step:

$$\max_{\theta} \sum_{d=1}^{D} \left[\sum_{\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2} q^d(\boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2) \log p(\boldsymbol{b}^d, \boldsymbol{s}^d, \boldsymbol{m}, \boldsymbol{h}^1, \boldsymbol{h}^2 | \theta) \right]$$

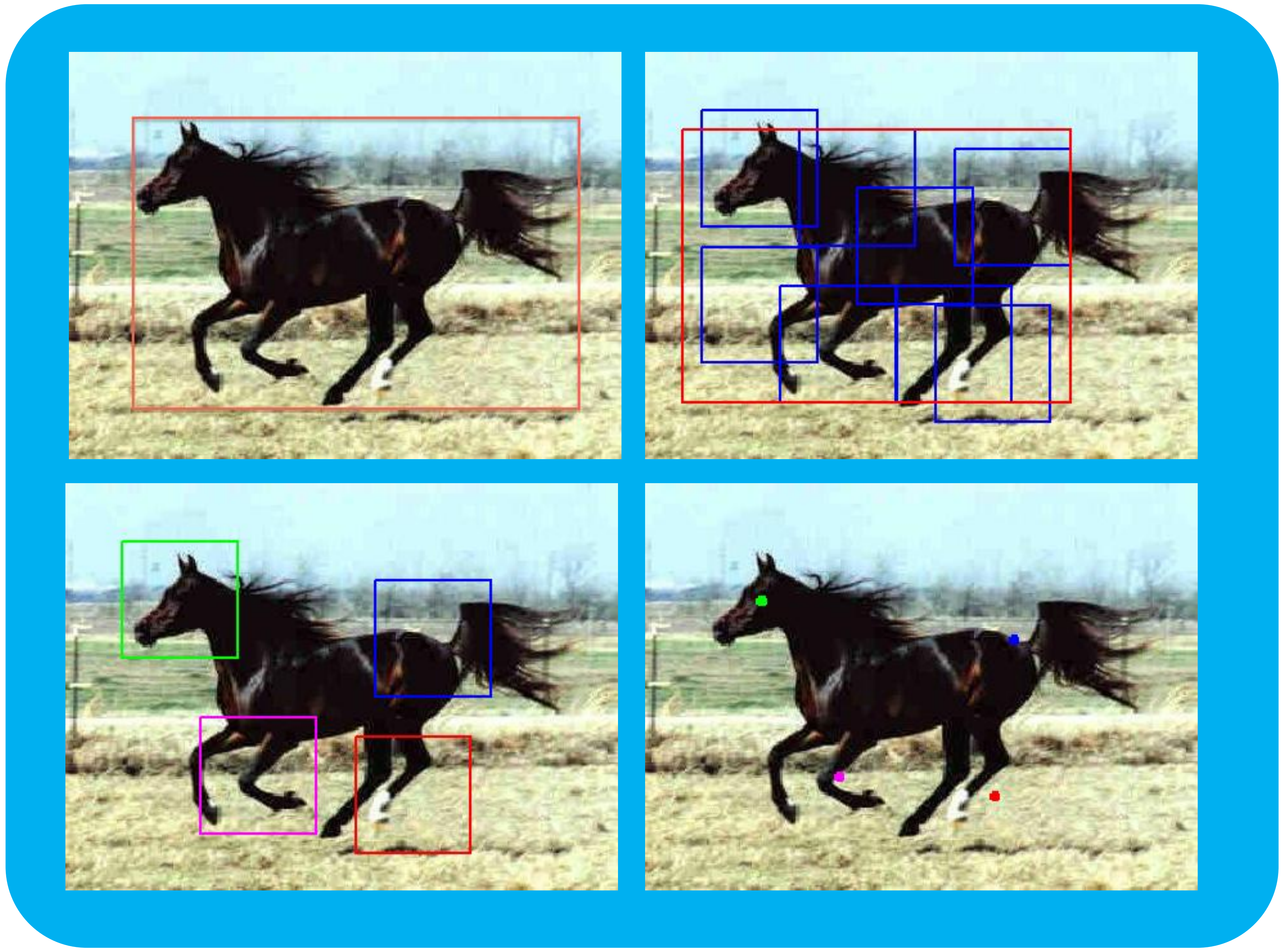
MCMC based stochastic approximation procedure

Seeds Extraction

For training we need
binary masks + seeds

Part-based detector:

- automatically identify parts that remain consistent in all images
- only bounding boxes around each object are required for training



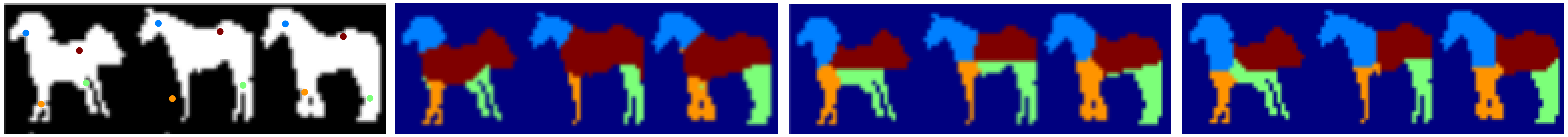
We can train MSBM given only object binary masks

Experiments: Baselines

MSBM The new procedure + binary annotations + automatically extracted seeds

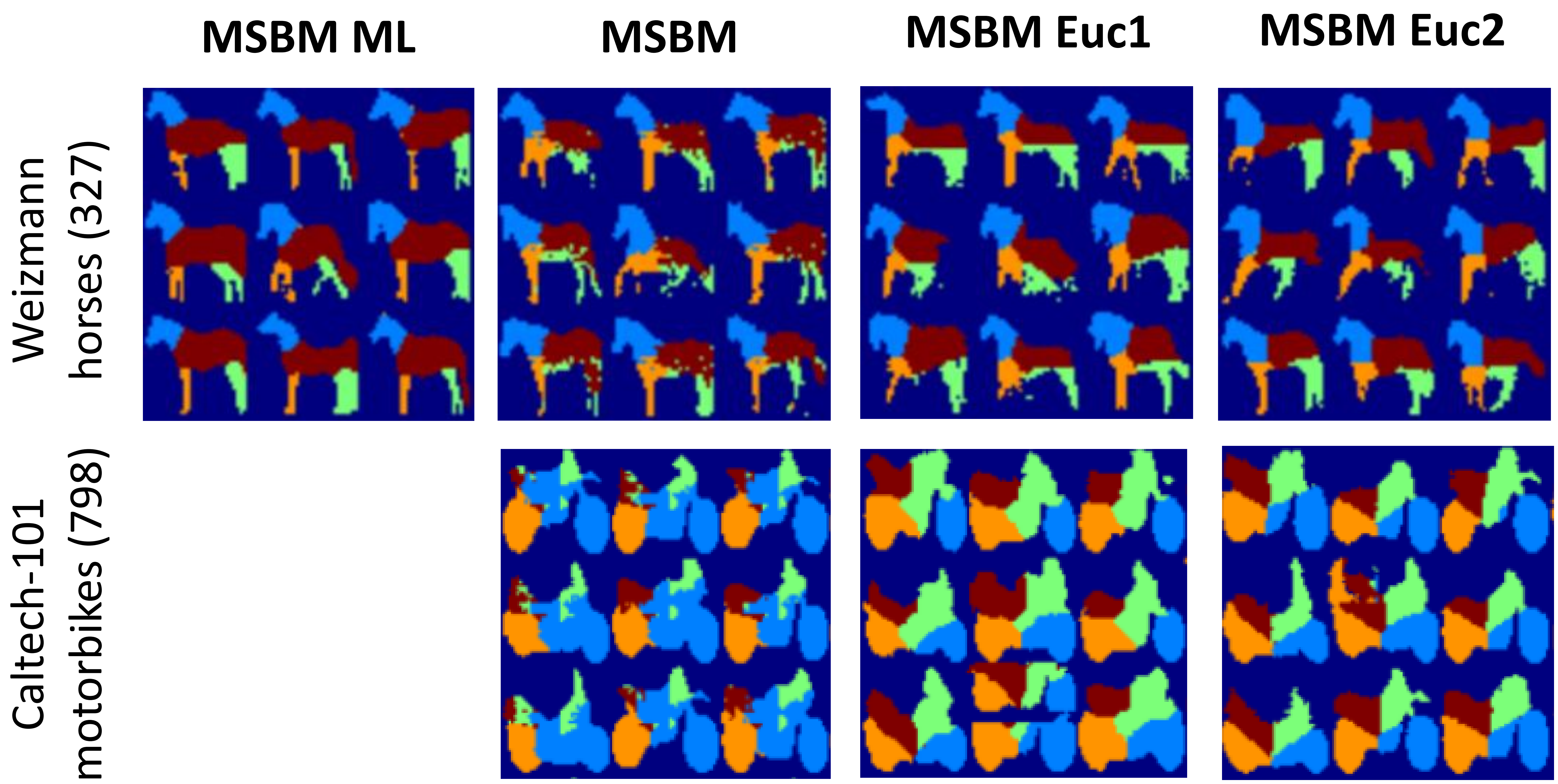
Baselines:

- **SBM:** original method + binary masks
- **MSBM ML**: original method + manually obtained multilabel annotations
- **MSBM Euc1** and **MSBM Euc2**: original method + heuristically obtained multilabel annotations from seeds



binary mask with seeds | manual | Euc1 | Euc2

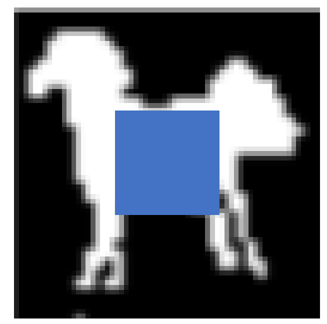
Experiments: Sampling



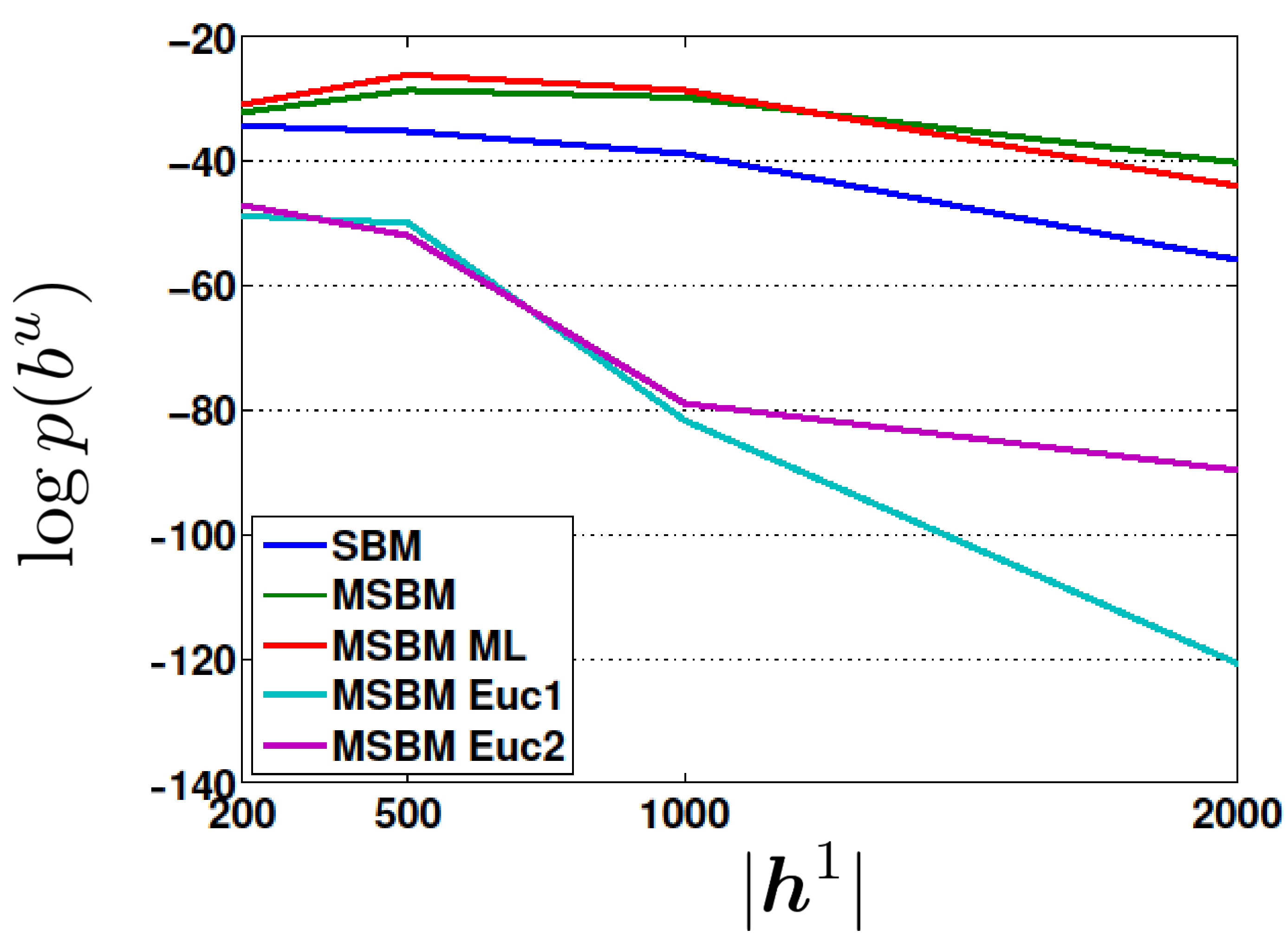
Model size: 1000+100

Experiments: Shape Completion

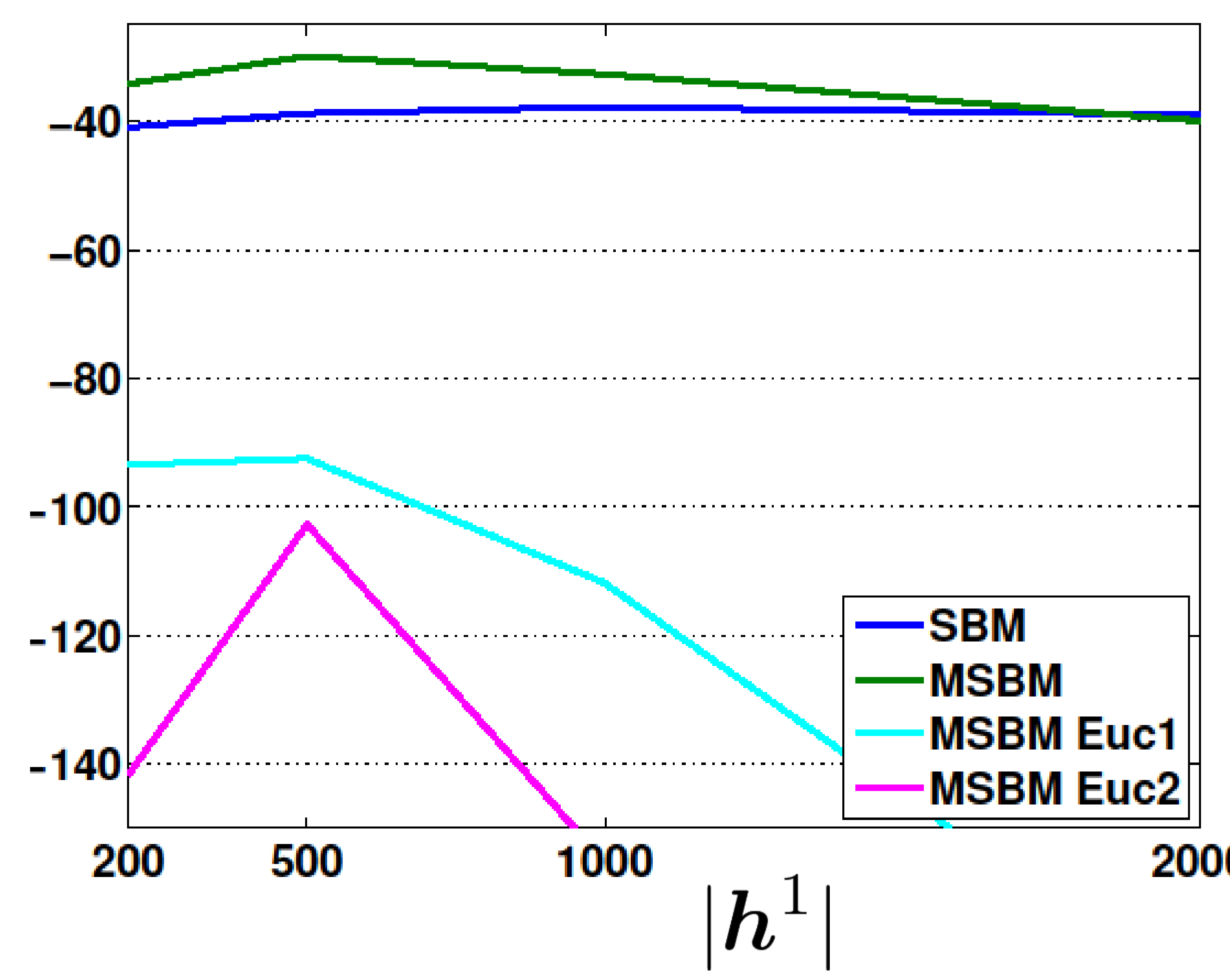
Divide each test image into 9 segments



Infer binary mask of one segment given the mask of the remaining 8 segments



Weizmann horses



Caltech-101 motorbikes

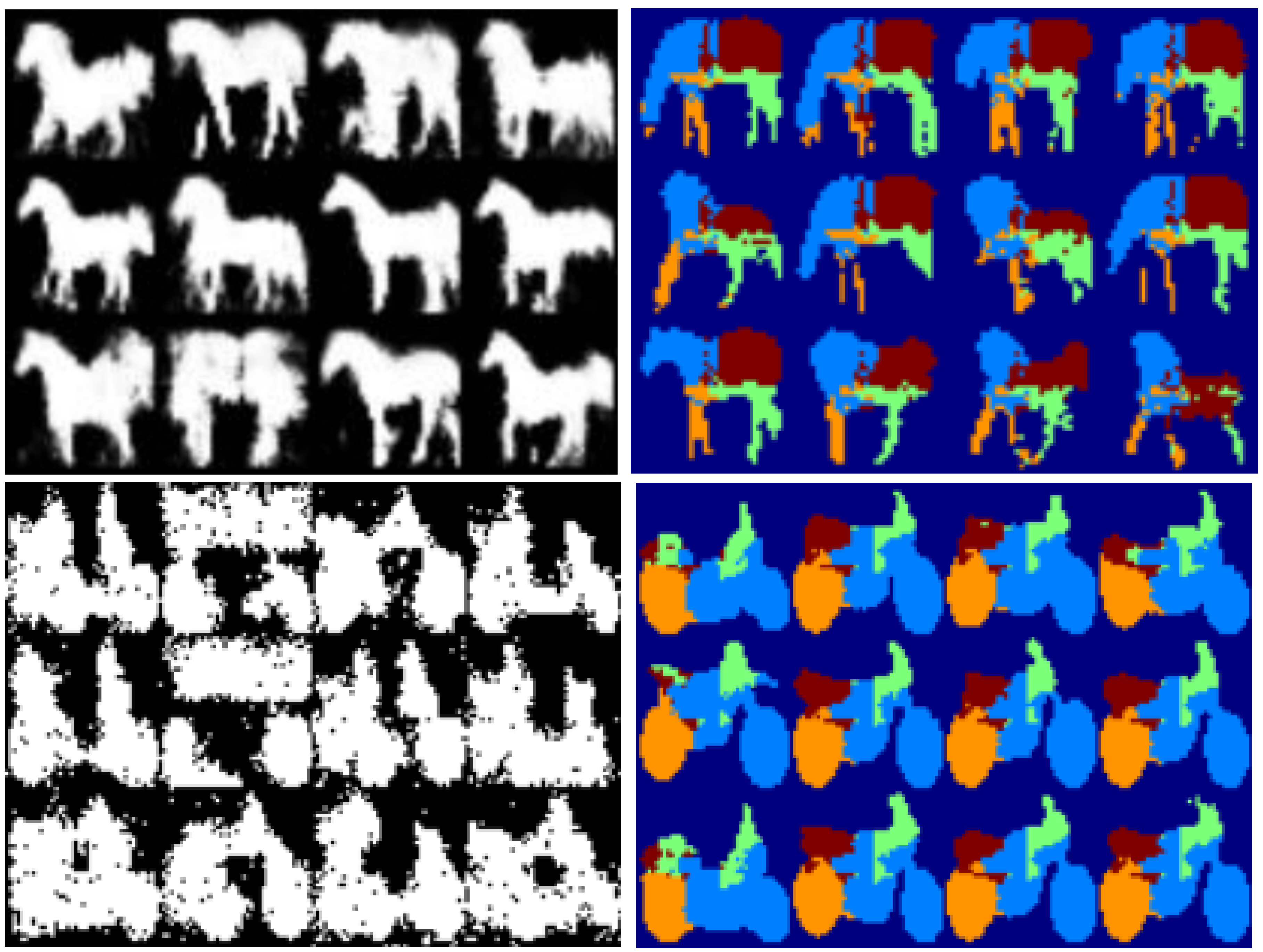
$|\boldsymbol{h}^2| = 200$

Experiments: Shape Generation From the Seeds

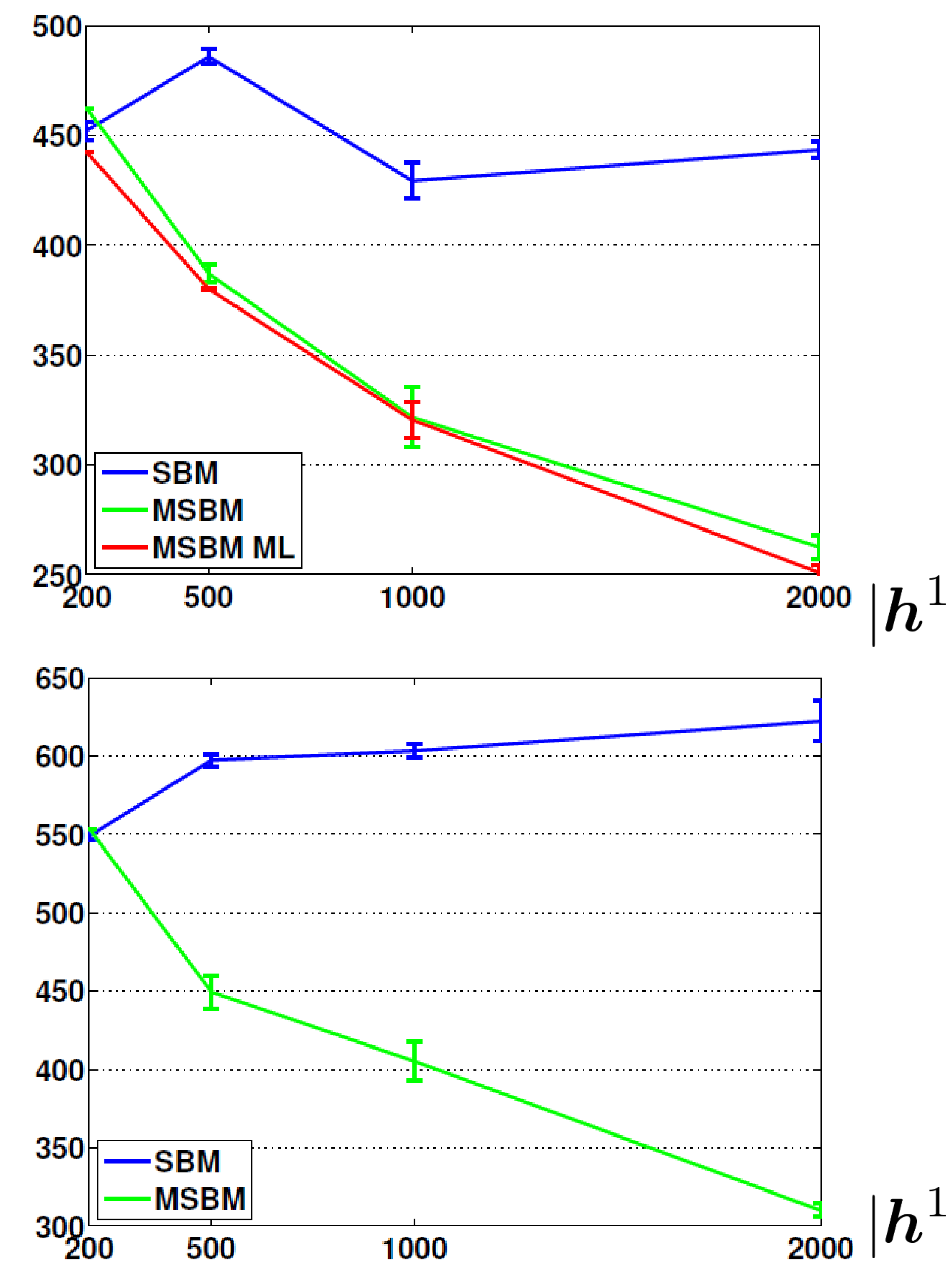
Seeds of object parts → Generate shape with MCMC

Weizmann horses

Caltech-101 motorbikes



SBM samples MSMB samples



Hamming distance between test and generated shapes

Conclusions

Contributions:

- A joint probabilistic model of a binary mask, a multilabel mask, and object seeds.
- A training procedure allowing to train a multilabel model given only binary mask and seeds that can be obtained automatically.

Results:

- MSBM trained by new procedure outperforms SBM in the tasks related to binary shapes and is very close to the original MSBM in terms of quality of multilabel shapes.