Supplementary material: Deep Part-Based Generative Shape Model with Latent Variables

Alexander Kirillov[1]
alexander.kirillov@tu-dresden
Mikhail Gavrikov[2]
gavrmike@gmail.com
Ekaterina Lobacheva[4]
elobacheva@hse.ru
Anton Osokin[3]
anton.osokin@inria.fr
Dmitry Vetrov[4]
vetrovd@yandex.ru

[1] TU Dresden, Dresden, Germany
[2] Rubbles, Moscow, Russia
[3] INRIA – École Normale Supérieure, Paris, France
[4] National Research University Higher School of Economics (HSE), Moscow, Russia

1 E-step with unknown seeds

Here, we consider the E-step when the position of some seeds is unknown. Let $\boldsymbol{s}^u, u = 1,\dots,U$, be a subset of the parts that do not have seeds. In this case, instead of $p(\boldsymbol{m},\boldsymbol{h}^1,\boldsymbol{h}^2|\boldsymbol{b},\boldsymbol{s},\boldsymbol{\theta})$ in equation (8), we approximate

$$p(\boldsymbol{m},\boldsymbol{h}^1,\boldsymbol{h}^2,\boldsymbol{s}^u|\boldsymbol{b},\boldsymbol{s}^{\{1,\dots,P\}\setminus u},\boldsymbol{\theta}),$$

using factorized approximation on $\boldsymbol{m}$, $\boldsymbol{h}^1$, $\boldsymbol{h}^2$ and $\boldsymbol{s}^u$ with the following factorization:

$$q^d(\boldsymbol{m},\boldsymbol{h}^1,\boldsymbol{h}^2,\boldsymbol{s}^u) = \prod_{i=1}^{I} q_i^d(m_i) \prod_{j=1}^{J} q_j^d(h_j^2) \prod_{k=1}^{K} q_k^d(h_k^2) \prod_{t\in u} q_t^d(s_t)$$

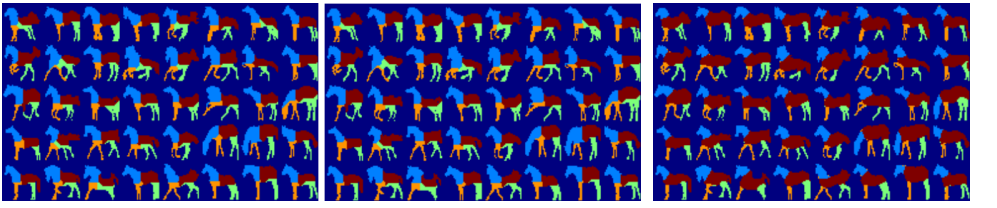
Using (11) we obtain:

$$\hat{q}_t^d(s_t = s) \propto \exp\left(\sum_{i=1}^{I} q_i^d(m_i = t)\|s - f_{\text{coord}}(i)\|_2^2\right).$$

Getting $q_t^d(s_t)$ requires computing of $\hat{q}_t^d(s_t = s)$ for all coordinates of pixels s. We found that it is possible to replace $q_t^d(s_t)$ with delta function to avoid computational complexity, since experiment results did not change. This operation is equivalent to setting seed s_t to the position of s_t^*, where

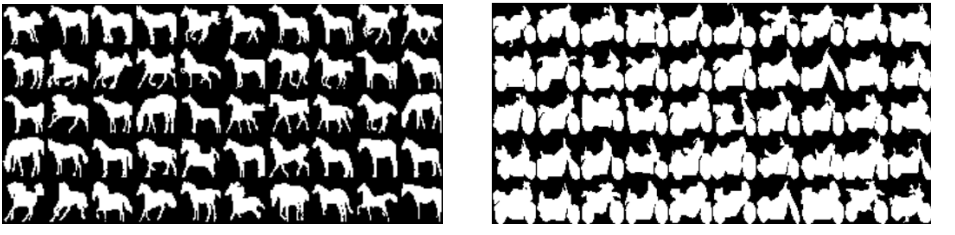
$$s_t^* = \arg\min_s \left(\sum_{i=1}^{I} q_i^d(m_i = t)\|s - f_{\text{coord}}(i)\|_2^2\right).$$

Equations (13), (14) and (15) to calculate $q_i^d(m_i), q_j^d(h_j^1)$ and $q_k^d(h_k^2)$ remain unchanged.



(a) (b) (c)

Figure 1: Multilabel segmentations: (a) segment the object region into parts given the seeds using naive Euclidean distance (Euc1), (b) segment the object region into parts given the seeds using Euclidean distance on the path lying inside the object (Euc2), (c) manual segmentations.



(a) (b)

Figure 2: (a) – Samples of the cropped and rescaled images from Weizmann dataset; (b) – Samples of the cropped and rescaled images from Caltech-101 motorbikes dataset.

2 Multilabel samples

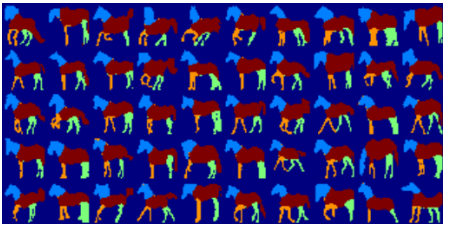
Different types of the multilabel segmentations for Weizmann dataset are shown on fig. (1).

3 Datasets

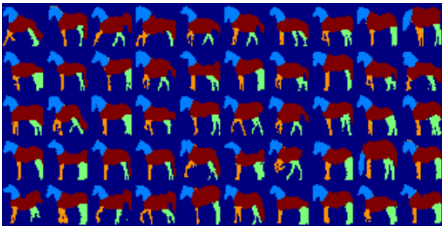
We perform all the experiments on the two datasets: the Weizmann horse dataset and the Caltech-101 motorbikes.

The Weizmann horse dataset contains 327 images of horses and a binary mask for each image. The images from the dataset were cropped and rescaled using a 2-step rescaling procedure with delotation in the middle. This procedure allows to obtain smooth and realistic horses' shapes with 32×32 pixels resolution i.e. see fig. 2a. We split the dataset into training set with 277 shapes and test set with 50 horses by random selection.

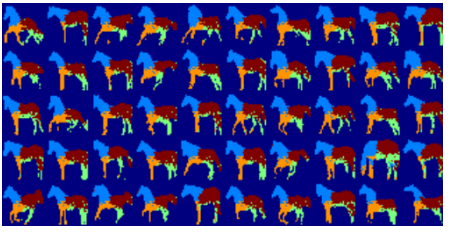
The Caltech-101 motorbikes dataset contains 798 images of motorbikes and a binary mask for each image. The images from the dataset were cropped and rescaled to 40×40 pixels resolution i.e. see fig. 2b. We split dataset into training set with 600 shapes and test set with 198 motorbikes by random selection.



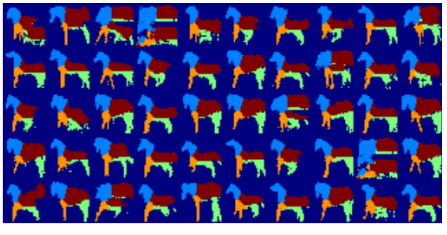
(a)



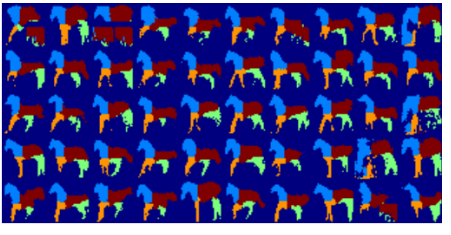
(b)



(c)



(d)



(e)

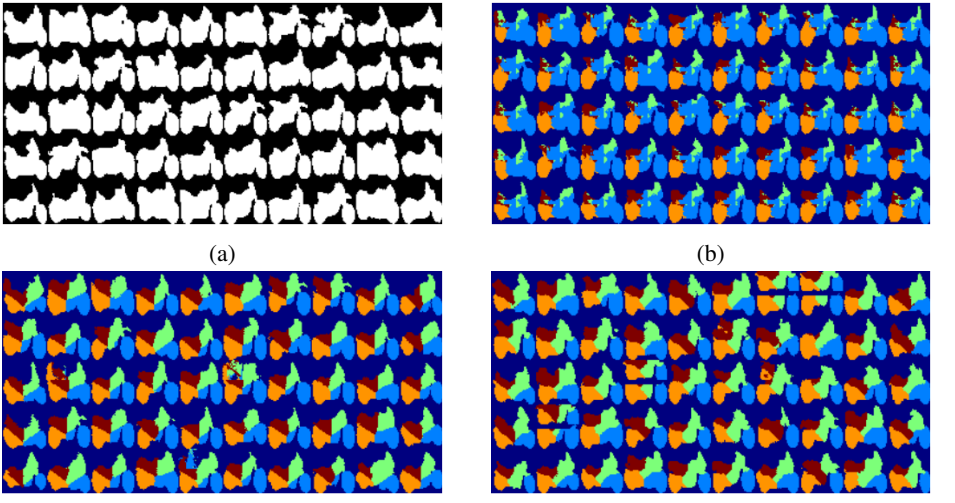
Figure 3: (a) – ground-truth labelings of parts obtained manually; (b) – samples from the MSBM trained using the full ground truth; (c) – samples from the MSBM trained using our method that is aware of only binary masks and part seeds; (d) – samples from the MSBM trained using multilabel segmentations obtained by Euc1; (e) – samples from the MSBM trained using multilabel segmentations obtained by Euc2. The corresponding cells of (b), (c), (d) and (e) are generated from the same initialization.

4 Comparison of MSBMs trained differently

Fig. (3) shows samples generated by the MSBMs trained differently on Weizmann horses dataset. Fig. (4) shows samples generated by the SBM and the MSBMs trained differently on Caltech-101 motorbikes dataset.

5 Shape completions

Fig. (5) presents result of "imputation score" experiment for Weizmann horses dataset and for Caltech-101 motorbikes.

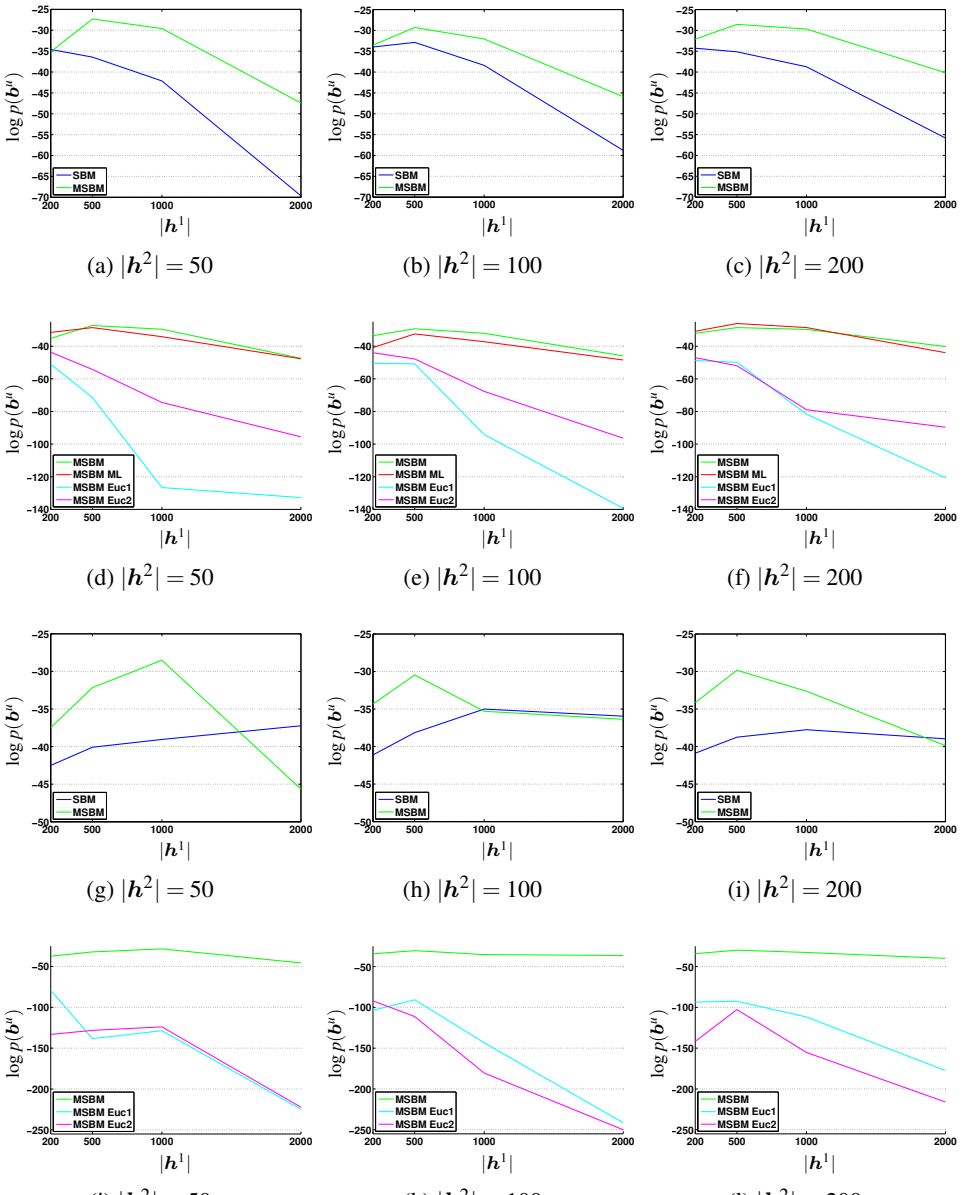


(a) (b) (c) (d)

Figure 4: (a) – samples from the SBM; (b) – samples from the MSBM trained using our method that is aware of only binary masks and part seeds; (c) – samples from the MSBM trained using multilabel segmentations obtained by Euc1; (d) – samples from the MSBM trained using multilabel segmentations obtained by Euc2. The corresponding cells of (a), (b), (c) and (d) are generated from the same initialization.

6 The shape generation from the seeds

Fig. (6) presents result of "generation from seeds" experiment for Weizmann horses dataset and for Caltech-101 motorbikes.



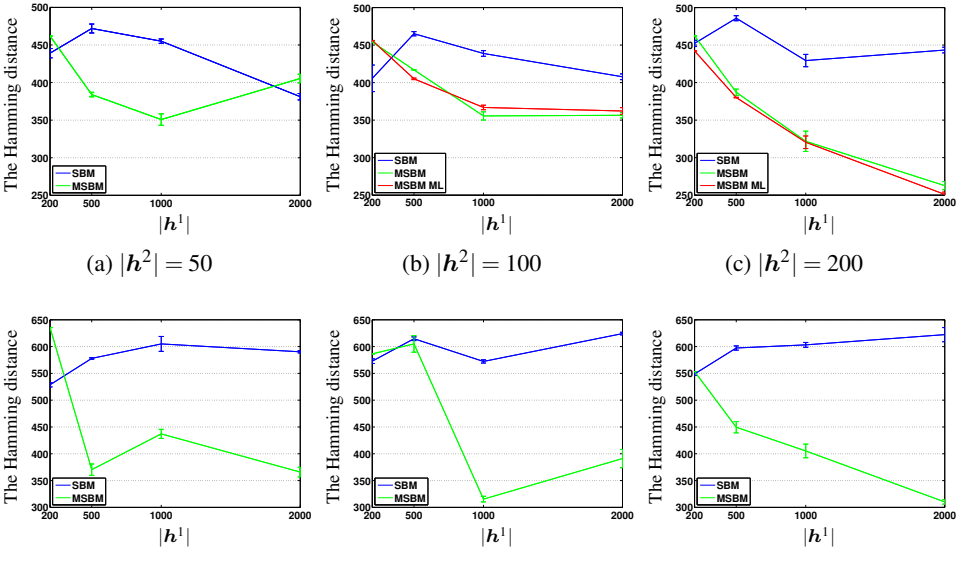
(a) $|\boldsymbol{h}^2| = 50$ (b) $|\boldsymbol{h}^2| = 100$ (c) $|\boldsymbol{h}^2| = 200$

(d) $|\boldsymbol{h}^2| = 50$ (e) $|\boldsymbol{h}^2| = 100$ (f) $|\boldsymbol{h}^2| = 200$

(g) $|\boldsymbol{h}^2| = 50$ (h) $|\boldsymbol{h}^2| = 100$ (i) $|\boldsymbol{h}^2| = 200$

(j) $|\boldsymbol{h}^2| = 50$ (k) $|\boldsymbol{h}^2| = 100$ (l) $|\boldsymbol{h}^2| = 200$

Figure 5: Imputation scores for the different models with different number of units on the hidden layers for Weizmann dataset – (a)-(f) and for Caltech-101 motorbikes – (g)-(l) (the higher the better). The blue lines correspond to the SBM model, the red ones – to the MSBM trained using the fully annotated ground truth, the green ones – trained with our EM-based method, the cyan ones – to the MSBM trained using the multilabel segmentations obtained by Euc1, the magenta ones– to the MSBM trained using the multilabel segmentations obtained by Euc2.



(a) $|h^2| = 50$ (b) $|h^2| = 100$ (c) $|h^2| = 200$

(d) $|h^2| = 50$ (e) $|h^2| = 100$ (f) $|h^2| = 200$

Figure 6: The Hamming distance between the test shapes and shapes generated by the SBM and the MSBM from the seeds for Weizmann dataset – (a), (b) and (c) and for Caltech-101 motorbikes – (d), (e) and (f) (the lower the better).