

# Многоклассовая модель формы со скрытыми переменными<sup>\*†</sup>

А. Н. Кириллов, М. И. Гавриков, Е. М. Лобачева,  
А. А. Осокин, Д. П. Ветров

В данной работе рассматриваются модели формы объектов на изображении: бинарная и многоклассовая модели Больцмана. Предлагается новый алгоритм обучения многоклассовой модели формы Больцмана, для применения которого достаточно неполной разметки данных, а именно: бинарной разметки и задания семян, указывающих приближенное расположение частей объектов.

**Ключевые слова:** модель формы Больцмана, многоклассовая модель формы Больцмана, графические модели, EM-алгоритм.

## 1. Введение

Модели формы объекта играют важную роль в различных задачах компьютерного зрения: сегментации [6, 9], восстановлении части изображения [8], детектировании объектов [7]. В задачах сегментации изображений модели формы позволяют учесть априорные знания о форме объекта, что значительно улучшает качество решения [11].

Существует множество способов задавать форму объекта [9, 3, 12], однако большинство из них либо недостаточно гибкие, либо требуют сложной дополнительной разметки обучающей выборки. Недавно

---

\*Статья является расширенной версией статьи [2], опубликованной на конференции ГрафиКон'2014.

†Работа выполнена при финансовой поддержке РФФИ (код проекта 15–31–20596), а также в рамках совместного проекта Майкрософт–МГУ (РПД 2015–077 (ДП836230) «Совместные исследования по компьютерному зрению в России»).

предложенная генеративная модель формы SBM [10] позволяет задавать форму объекта с помощью глубокой сети определенной структуры. В [11] модель SBM обобщается на многоклассовый случай, в котором объект состоит из нескольких частей, каждой из которых ставится в соответствие свой класс. Полученная многоклассовая модель MSBM обладает большей выразительной способностью по сравнению с бинарной, так как настроить модель на вариации формы отдельных частей объекта обычно проще, однако для ее обучения требуются данные с полной многоклассовой разметкой.

В данной работе мы предлагаем метод обучения MSBM, использующий лишь бинарные маски объектов и семена их частей (точки на изображении, в окрестностях которых лежат пиксели заданного класса). Для обучения модели с такими входными данными нами вводится графическая модель, задающая взаимосвязь между параметрами и переменными MSBM, бинарной и многоклассовой разметкой и семенами, а также предлагается основанная на ней EM-подобная процедура обучения. При этом семена частей могут размечаться как вручную, так и автоматически с помощью детектора частей объектов (part-based detector, [12]). Использование детектора позволяет обучать MSBM на основе лишь исходных изображений и их бинарных разметок.

На этапе экспериментов мы сравнили MSBM, обученную нашим алгоритмом, с существующими аналогами на двух наборах данных и показали, предложенная модель превосходит SBM по качеству и сопоставима с MSBM, обученной стандартным способом. Также показано, что модель, обученная нашим алгоритмом, лучше MSBM, для которой многоклассовая разметка получена из бинарной маски и семян на основе простых эвристических предположений.

## 2. Нотация

Пусть нам дано  $D$  центрированных и масштабированных изображений с объектами одного типа (рис. 1а, тип: лошади в профиль). За  $P$  обозначим число частей объекта, а  $N$  — число пикселей на изображении. Пусть  $B = \{\mathbf{b}^d\}_{d=1,\dots,D}$  — множество бинарных разметок изображений (рис. 1б), где  $\mathbf{b}^d \in \{0, 1\}^N$ , а  $M = \{\mathbf{m}^d\}_{d=1,\dots,D}$  — множество многоклассовых разметок (рис. 1в), где  $\mathbf{m}^d \in \{0, \dots, P\}^N$ , 0 соответствует фону. Также будем использовать бинарные переменные

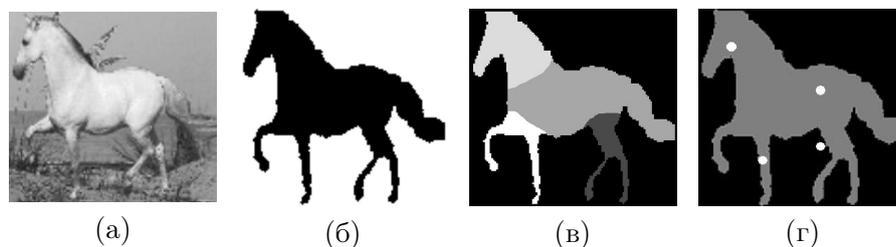


Рис. 1. (а) — изображение, (б) — бинарная разметка  $\mathbf{b}$ , (в) — многоклассовая разметка  $\mathbf{m}$ , (г) семена  $\mathbf{s}$  для 4 частей: головы, передних ноги, задних ноги, крупа.

$m_{ip}^d \in \{0, 1\}$  для обозначения принадлежности пикселя  $i$  на изображении  $d$  классу  $p$  и векторные переменные  $\mathbf{m}_p^d = \{m_{ip}^d\}_{i=1, \dots, N}$ . Введем функцию  $f_{coord}(i)$ , ставящую в соответствие номеру пикселя его координаты на изображении и множество семян  $S = \{\mathbf{s}^d\}_{d=1, \dots, D}$ , где  $\mathbf{s}^d$  — это упорядоченный набор семян для изображения  $d$ , каждое семя задается своими координатами на изображении.

### 3. Модели формы Больцмана

Модель формы Больцмана SBM [10] — модель бинарной формы, представляющая собой трехслойную глубинную модель Больцмана (DBM, [4]) с дополнительными ограничениями. Наблюдаемый слой, соответствующий изображению, делится на 4 равные части с перекрытием и каждая часть соединяется только со своим подмножеством переменных на первом скрытом слое. Более того, веса для каждой из этих частей одинаковы. Такие ограничения позволяют существенно уменьшить число параметров глубинной модели, что позволяет избежать переобучения, ускорить процесс настройки параметров, а также обучать модель на меньшем объеме данных. Будем считать, что эти ограничения неявно учтены в соответствующей матрице весов. Архитектуры DBM и SBM показаны на рис. 2.

SBM с наблюдаемым слоем  $\mathbf{b}$  и двумя скрытыми  $\mathbf{h}^1, \mathbf{h}^2$  задается с помощью распределения Гиббса:

$$p(\mathbf{b}, \mathbf{h}^1, \mathbf{h}^2 | \boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \exp(-E(\mathbf{b}, \mathbf{h}^1, \mathbf{h}^2 | \boldsymbol{\theta})), \quad (1)$$

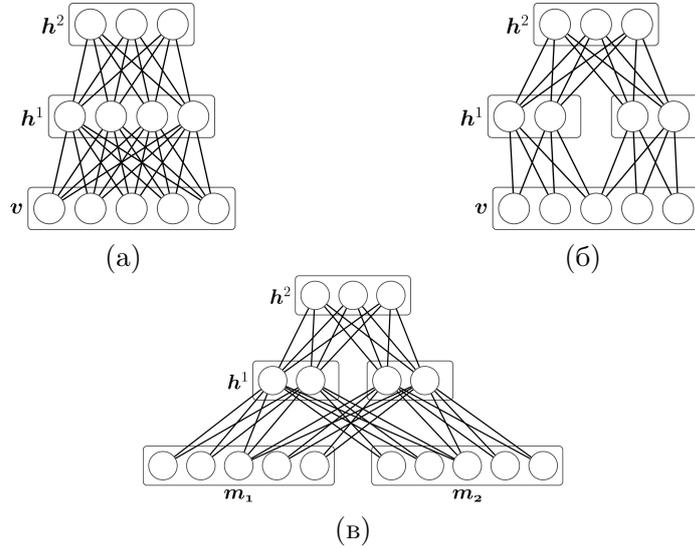


Рис. 2. Архитектуры глубоких моделей Больцмана: (а) — DBM с 2 скрытыми слоями, (б) — одномерная SBM с делением первого скрытого слоя на 2 части, (в) — одномерная MSBM с делением первого скрытого слоя на 2 части и с 2 частями у объекта.

где  $Z(\theta) = \sum_{\mathbf{b}, \mathbf{h}^1, \mathbf{h}^2} \exp(-E(\mathbf{b}, \mathbf{h}^1, \mathbf{h}^2 | \theta))$  — нормировочная константа,  $\theta = \{\mathbf{a}, \mathbf{c}^1, \mathbf{c}^2, \mathbf{W}^1, \mathbf{W}^2\}$  — набор параметров модели, а энергия модели выглядит следующим образом:

$$E(\mathbf{b}, \mathbf{h}^1, \mathbf{h}^2 | \theta) = \sum_{i=1}^I a_i b_i + \sum_{i=1}^I \sum_{j=1}^J b_i W_{i,j}^1 h_j^1 + \sum_{j=1}^J c_j^1 h_j^1 + \sum_{j=1}^J \sum_{k=1}^K h_j^1 W_{j,k}^2 h_k^2 + \sum_{k=1}^K c_k^2 h_k^2. \quad (2)$$

Ограничения на равенство групп весов между частями изображения и подмножествами переменных первого скрытого слоя, описанные выше, заложены в матрицу  $\mathbf{W}_1$ .

Многочлаовая модель формы Больцмана MSBM [11] является обобщением SBM на многоклассовый случай, в котором объект состоит из нескольких частей, каждой из которых соответствует своя

метка класса. MSBM задается аналогично с помощью распределения Гиббса:

$$p(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2 | \boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \exp(-E(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2 | \boldsymbol{\theta})), \quad (3)$$

где  $Z(\boldsymbol{\theta}) = \sum_{\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2} \exp(-E(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2 | \boldsymbol{\theta}))$  — нормировочная константа,  $\boldsymbol{\theta} = \{\mathbf{a}, \mathbf{c}^1, \mathbf{c}^2, \mathbf{W}^1, \mathbf{W}^2\}$  — набор параметров модели, а энергия записывается следующим образом:

$$E(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2 | \boldsymbol{\theta}) = \sum_{i=1}^I \sum_{p=1}^P a_{i,p} m_{i,p} + \sum_{i=1}^I \sum_{j=1}^J \sum_{p=1}^P m_{i,p} W_{i,j,p}^1 h_j^1 + \\ + \sum_{j=1}^J c_j^1 h_j^1 + \sum_{j=1}^J \sum_{k=1}^K h_j^1 W_{j,k}^2 h_k^2 + \sum_{k=1}^K c_k^2 h_k^2. \quad (4)$$

В параметры  $\mathbf{W}_1$  заложены аналогичные ограничения на равенство весов, как и в случае SBM.

Наиболее эффективная процедура обучения DBM [5] основана на максимизации логарифма правдоподобия  $\log p(\mathbf{b} | \boldsymbol{\theta})$  с помощью EM-алгоритма. Она состоит из двух основных этапов:

- 1) Предобучение. На этом этапе каждый слой сети обучается отдельно с помощью стохастического градиентного метода.
- 2) Окончательная настройка весов. Здесь используется вариационный EM-алгоритм для максимизации нижней границы логарифма правдоподобия, которому в качестве инициализации подаются результаты предыдущего этапа.

Модели SBM и MSBM обучаются по аналогичной схеме с учетом ограничений на веса  $W_1$ . Подробное описание алгоритма обучения моделей Больцмана приведено в [5].

Важным ограничением стандартной процедуры обучения глубоких моделей является то, что для ее применения требуются данные с полной разметкой, получение которой, особенно в многоклассовом случае, очень трудозатратная задача. В данной работе мы ослабляем это ограничение для многоклассового случая, а именно, предлагаем процедуру обучения MSBM по данным с неполной разметкой — бинарным маскам объектов и семенам их частей.

#### 4. Совместная модель

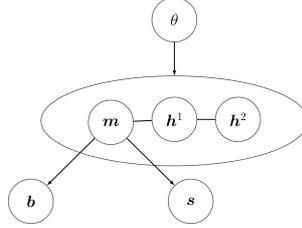


Рис. 3. Графическая модель  $p(\mathbf{b}, \mathbf{s}, \mathbf{m}, \mathbf{h}^1, \mathbf{h}^2 | \theta)$ .

Введем совместное распределение на бинарную маску  $\mathbf{b}$ , многоклассовую маску  $\mathbf{m}$ , семена  $\mathbf{s}$  и скрытые переменные модели  $\mathbf{h}^1, \mathbf{h}^2$  предполагая, что бинарная разметка  $\mathbf{b}$  и семена  $\mathbf{s}$  условно независимы при известной многоклассовой разметке  $\mathbf{m}$ :

$$p(\mathbf{b}, \mathbf{s}, \mathbf{m}, \mathbf{h}^1, \mathbf{h}^2 | \theta) = p(\mathbf{b} | \mathbf{m}) p(\mathbf{s} | \mathbf{m}) p(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2 | \theta). \quad (5)$$

Соответствующая графическая модель представлена на рис. 3. Под распределением  $p(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2 | \theta)$  здесь понимается распределение Гиббса для MSBM (3).

Далее рассмотрим распределения  $p(\mathbf{b} | \mathbf{m})$  и  $p(\mathbf{s} | \mathbf{m})$ . Будем считать, что при известной многоклассовой разметке  $\mathbf{m}$  бинарная метка пикселя  $b_i$  зависит только от его же многоклассовой метки  $m_i$ , причем если пиксель принадлежит какой-либо части объекта, то он точно принадлежит объекту, иначе он принадлежит фону. Таким образом совместное распределение  $p(\mathbf{b} | \mathbf{m})$  задается следующим образом:

$$p(\mathbf{b} | \mathbf{m}) = \prod_{i=1}^N p(b_i | m_i) = \prod_{i=1}^N ([b_i = 0][m_i = 0] + [b_i \neq 0][m_i \neq 0]). \quad (6)$$

Также будем считать, что на расположение семян все пиксели объекта влияют независимо друг от друга, а пиксели фона не влияют вовсе. Более того, на расположение семени  $s_{m_i}$ , отвечающего за часть объекта  $m_i$ , влияют только пиксели, относящиеся к этой части. Таким образом распределение  $p(\mathbf{s} | \mathbf{m})$  пропорционально произведению  $\prod_{i: m_i \neq 0} p(s_{m_i} | m_i)$ . Каждый фактор этого произведения зададим в виде нормального распределения:

$$p(s|m) \propto \prod_{i:m_i \neq 0} \mathcal{N}(s_{m_i} | f_{coord}(i), \sigma^2) \propto \prod_{i:m_i \neq 0} \exp \left\{ -\frac{\|s_{m_i} - f_{coord}(i)\|_2^2}{2\sigma^2} \right\}, \quad (7)$$

где  $\sigma$  — внешний параметр модели. Таким образом, мы предполагаем, что пиксель  $i$  «тянет на себя» себя  $m_i$ .

## 5. EM-алгоритм

В данном разделе опишем предлагаемый нами алгоритм обучения MSBM по данным с неполной разметкой. Для обучения параметров модели  $\theta$  будем с помощью EM-алгоритма решать следующую задачу максимизации:

$$\log P(B, S | \theta) = \sum_{d=1}^D \log p(\mathbf{b}^d, \mathbf{s}^d | \theta) \rightarrow \max_{\theta}. \quad (8)$$

При этом будем считать переменные  $\mathbf{b}$ ,  $\mathbf{s}$  наблюдаемыми, а  $\mathbf{m}$ ,  $\mathbf{h}^1$ ,  $\mathbf{h}^2$  — скрытыми.

На E-шаге будем находить вариационную оценку апостериорного распределения на скрытые переменные в семействе полностью факторизованных распределений

$$q^d(\mathbf{m}, \mathbf{H}) = \prod_{i=1}^I q_i^d(m_i) \prod_{j=1}^J q_j^d(h_j^1) \prod_{k=1}^K q_k^d(h_k^2) \quad (9)$$

минимизируя дивергенцию Кульбака-Лейблера:

$$\min_{q^d} \text{KL} \left( q^d(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2) \| p(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2 | \mathbf{b}^d, \mathbf{s}^d, \theta) \right). \quad (10)$$

На M-шаге будем обновлять параметры модели  $\theta$  решая следующую задачу максимизации:

$$\max_{\theta} \sum_{d=1}^D \left[ \sum_{\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2} q^d(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2) \log p(\mathbf{b}^d, \mathbf{s}^d, \mathbf{m}, \mathbf{h}^1, \mathbf{h}^2 | \theta) \right] \quad (11)$$

Далее рассмотрим оба шага подробнее.

### 5.1. E-шаг

На этом шаге необходимо решить задачу оптимизации (10) в полностью факторизованном семействе (9). Для этого мы будем использовать стандартный результат из [13, с. 466]:

$$q_z^d(m) \propto \exp\left(\mathbb{E}_{r \neq z} \log p(\mathbf{b}^d | \mathbf{m}) p(\mathbf{s}^d | \mathbf{m}) p(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2 | \theta)\right) \quad (12)$$

где индекс  $z$  задает фактор из (9) и  $\mathbb{E}_{r \neq z}$  — математическое ожидание по всем факторам из (9) кроме  $z$ . Используя (12) мы получаем следующее выражение для ненормированной вероятности:

$$\hat{q}_i^d(m_i = p) = [b_i^d \neq 0][m_i \neq 0] \exp\left(-\frac{1}{2\sigma^2} \left\| s_{m_i}^d - f_{coord}(i) \right\|_2^2 + a_{i,m_i} + \sum_j W_{i,j,m_i}^1 q_j(h_j^1 = 1)\right) + [b_i^d = 0][m_i^d = 0]. \quad (13)$$

Учитывая, что  $m_i \in \{0, 1, \dots, P\}$ , получаем нормированную вероятность:

$$q_i^d(m_i = p) = \frac{\hat{q}_i^d(m_i = p)}{\sum_{p'} \hat{q}_i^d(m_i = p')}, \quad p \in \{0, 1, \dots, P\}. \quad (14)$$

Формулы пересчета для  $q_j^d(\mathbf{h}_j^1)$  и  $q_k^d(\mathbf{h}_k^2)$  получаются аналогичным образом:

$$q_j^d(h_j^1 = 1) = \sigma(c_j^1 + \sum_{i,p=1}^P q_i(m_i = p) W_{i,j,p}^1 + \sum_k W_{j,k}^2 q_k(h_k^2 = 1)), \quad (15)$$

$$q_k^d(h_k^2 = 1) = \sigma(c_k^2 + \sum_j q_j(h_j^1 = 1) W_{j,k}^2), \quad (16)$$

где  $\sigma(x) = 1/(1 + e^{-x})$ .

Согласно [13] общая схема E-шага представляется собой итеративную процедуру с формулами пересчета (13)–(14), (15), (16). Стоит отметить, что (15) и (16) практически совпадают с аналогичными формулами для вариационного приближения на E-шаге в классической схеме обучения DBM [1]. Разница состоит лишь в том, что вместо известной многоклассовой разметки мы используем результаты (13)–(14).

Описанная процедура E-шага может быть обобщена на случай, когда часть семян неизвестна. Обозначим за  $\mathbf{s}^u, u = 1, \dots, U$  подмножество частей объекта, семена для которых не заданы. В данном случае вместо аппроксимации  $p(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2 | \mathbf{b}, \mathbf{s}, \theta)$  в (10) мы будем аппроксимировать

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

в следующем факторизованном семействе:

$$q^d(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2, \mathbf{s}^u) = \prod_{i=1}^I q_i^d(m_i) \prod_{j=1}^J q_j^d(h_j^2) \prod_{k=1}^K q_k^d(h_k^2) \prod_{t \in u} q_t^d(s_t). \quad (17)$$

Используя (12) получаем:

$$\hat{q}_t^d(s_t = s) \propto \exp \left( \sum_{i=1}^I q_i^d(m_i = t) \|s - f_{\text{coord}}(i)\|_2^2 \right). \quad (18)$$

При этом для получения  $q_t^d(s_t)$  необходимо подсчитать  $\hat{q}_t^d(s_t = s)$  для всех возможных координат семян  $s$ . Для уменьшения вычислительной сложности данной процедуры мы предлагаем заменить  $q_t^d(s_t)$  на дельта-функцию, что эквивалентно следующей схеме задания координат семян:

$$s_t = \arg \min_s \left( \sum_{i=1}^I q_i^d(m_i = t) \|s - f_{\text{coord}}(i)\|_2^2 \right). \quad (19)$$

Формулы пересчета (14), (15), (16) остаются неизменными.

## 5.2. M-шаг

На M-шаге мы обновляем параметры модели  $\theta$  решая задачу максимизации (11). Распределение  $q(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2)$  при этом уже известно нам из результатов работы E-шага. Используя (5) можно переписать задачу оптимизации (11) следующим образом:

$$\begin{aligned} \max_{\theta} \sum_{d=1}^D \left( \sum_{i=1}^I \sum_{p=1}^P a_{i,p} q_i(m_i^d = p) + \sum_{i=1}^I \sum_{j=1}^J \sum_{p=1}^P q_i(m_i^d = p) W_{i,j,p}^1 q_j(h_j^1 = 1) + \right. \\ \left. + \sum_{j=1}^J c_j^1 q_j(h_j^1 = 1) + \sum_{j=1}^J \sum_{k=1}^K q_j(h_j^1 = 1) W_{j,k}^2 q_k(h_k^2 = 1) + \right. \\ \left. + \sum_{k=1}^K c_k^2 q_k(h_k^2 = 1) - \log Z(\theta) \right). \quad (20) \end{aligned}$$

Данную задачу оптимизации можно решать приближенно с помощью стохастического градиентного метода [1]. Заметим, что описанная процедура обучения имеет ту же вычислительную сложность, что и стандартная процедура обучения MSBM [11].

## 6. Использование детектора частей

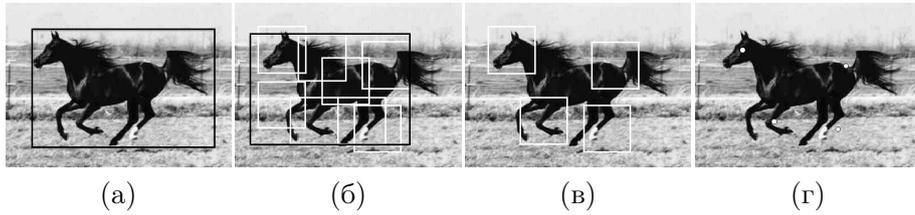


Рис. 4. (а) изображение с выделенным ограничивающим прямоугольником, (б) результат применения детектора, (в) выбранные части, (г) семена для выбранных частей.

Предлагаемая в данной работе процедура обучения MSBM использует данные с бинарной разметкой и выделенными семенами частей объектов. Семена на изображениях можно получать как посредством ручной разметки, так и автоматически с помощью детектора частей [12]. Процедура ручной разметки семян значительно проще получения полной многоклассовой разметки, поэтому даже в этом случае предлагаемая процедура обучения требует менее трудоемкой разметки обучающих данных.

Остановимся подробнее на автоматическом получении семян с помощью детектора частей. Для обучения такого детектора требуются изображения с размеченными ограничивающими прямоугольниками для объектов (рис. 4а). Заметим, что если для обучающей выборки изображений уже известна бинарная разметка, то получение ограничивающих прямоугольников из нее не составляет труда. На этапе обучения детектор сам выделяет части на изображении, имеющие сходную структуру на обучающей выборке, причем это именно части всего изображения, не обязательно объекта. Однако из обученных частей изображения можно легко выбрать части объекта. На этапе применения к тестовому изображению детектор выделяет ограничивающие прямоугольники для объекта и обученных частей изображения (рис. 4б), из которых для дальнейшего использования выбираются только части объекта (рис. 4в), центры которых и принимаются за семена (рис. 4г).

## 7. Эксперименты

В данном разделе мы сравним MSBM, обученную нашим алгоритмом, с SBM на бинарной задаче и с MSBM, обученной классическим способом, на многоклассовой задаче. Также мы рассмотрим MSBM, полная многоклассовая разметка для обучения которой с помощью стандартного алгоритма получена из бинарных масок и семян с помощью простых эвристик.

**Данные.** Все сравнение будем проводить на двух наборах данных: лошади [6] и мотоциклы [14].

Набор изображений лошадей состоит из 327 изображений, для которых даны их бинарные маски. Мы образали и отмасштабировали все изображения, приведя их к размеру  $32 \times 32$  пикселя (рис. 5а). Далее мы случайным образом разделили данные на обучающую выборку из 277 изображений и тестовую — из 50. Для применения стандартного метода обучения MSBM мы вручную сделали многоклассовую разметку данного набора (рис. 5в).

Набор изображений мотоциклов состоит из 798 изображений, для которых даны их бинарные маски. Мы образали и отмасштабировали все изображения, приведя их к размеру  $40 \times 40$  пикселей (рис. 5б).

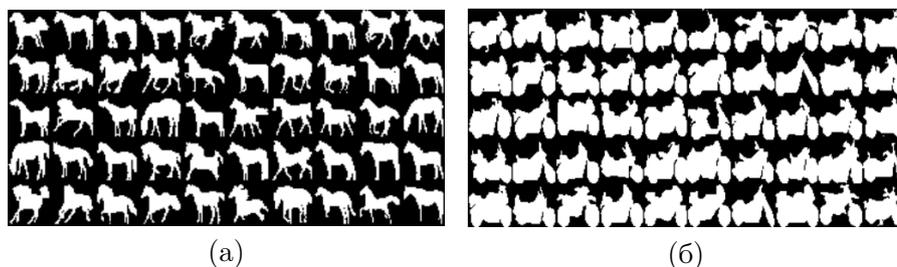


Рис. 5. (а) обрезанные и отмасштабированные формы лошадей; (б) обрезанные и отмасштабированные формы мотоциклов.

Далее мы случайным образом разделили данные на обучающую выборку из 600 изображений и тестовую — из 198. Для данного датасета ручной многоклассовой разметки не существует.

**Алгоритмы для сравнения.** В экспериментах будем сравнивать MSBM, обученную по данным с неполной разметкой с помощью предлагаемого алгоритма, со следующими моделями:

- SBM,
- MSBM, обученной классическим способом по данным с полной многоклассовой разметкой [11] (пример разметки показан на рис. 6в); будем обозначать этот алгоритм MSBM ML,
- MSBM, полная многоклассовая разметка для стандартного обучения которой получена автоматически из бинарной разметки и семян с помощью простых эвристик. Эвристики состоят в следующем: каждый пиксель объекта мы будем относить к той части, семя которой расположено ближе всего к нему по одной из двух метрик. В качестве первой метрики рассматривается евклидово расстояние между координатами пикселя и семени (Euc1), а в качестве второй — длина кратчайшего пути от пикселя к семени внутри границ объекта (Euc2). Примеры полученных многоклассовых разметок приведены на рис. 6.

Для проведения всех экспериментов мы использовали компьютер с процессором i7 2.8GHz CPU и 12GB RAM. Обучение с помощью нашего метода модели MSBM с 2000 переменными на первом скрытом

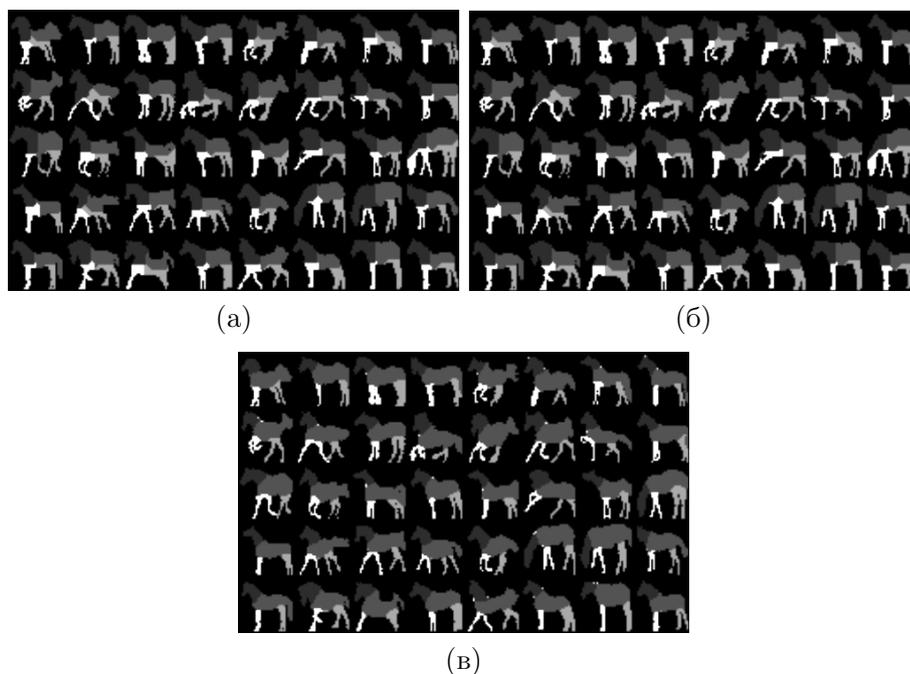


Рис. 6. Многоклассовые разметки: (а) полученная из бинарной разметки и семян с использованием евклидова расстояния между пикселями и семенами (Euc1), (б) полученная из бинарной разметки и семян с использованием кратчайшего пути между пикселями и семенами внутри границ объекта (Euc2), (в) полученная вручную.

слое и 200 — на втором на наборе данных с лошадьми занимало около 2 часов (имплементация на Cython).

### 7.1. Сравнение MSBM, обученных разными способами

В данном эксперименте мы сравним:

- MSBM, обученную классическим способом по данным с полной многоклассовой разметкой, полученной вручную,
- MSBM, обученную предлагаемым алгоритмом,

- MSBM, полная многоклассовая разметка для стандартного обучения которой получена автоматически из бинарной разметки и семян с помощью эвристик.

Обучим все модели со следующим количеством переменных на скрытых слоях: 1000 переменных на первом слое и 100 — на втором. Перекрытие между частями изображения на видимом слое моделей возьмем равным 4 пикселям. Для получения семян для второй и третьей моделей обучим детектор частей объектов [12].

Сравнивать будем генерируемые из этих моделей примеры форм. Генерация будем проводить по стандартной схеме Гиббса [10] с 200 итерациями. На рисунках 7 и 8 представлены результаты генерации из всех моделей, обученных на изображениях лошадей и мотоциклов.

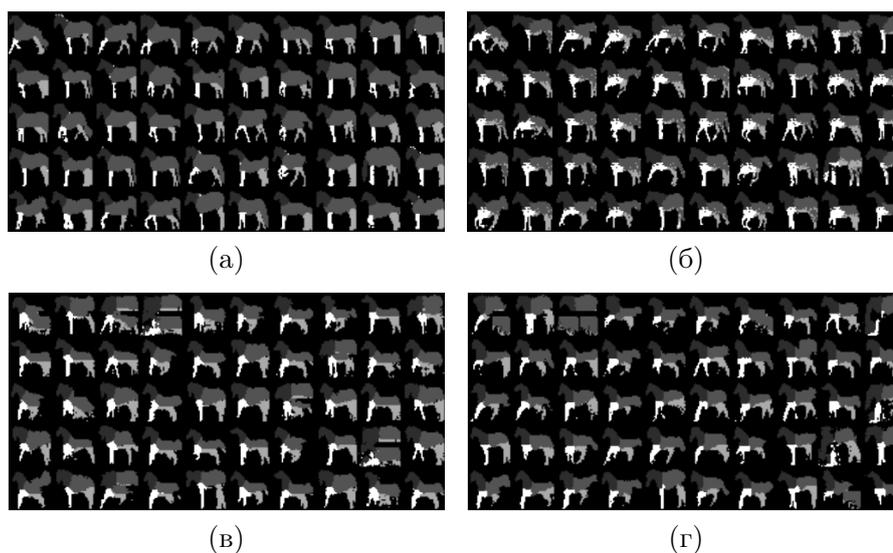


Рис. 7. Результаты генерации лошадей из: (а) — MSBM, обученной классическим способом по данным с полной многоклассовой разметкой, полученной вручную; (б) — MSBM, обученную предлагаемым алгоритмом с использованием только бинарной разметки и семян; (в) — MSBM, полная многоклассовая разметка для стандартного обучения которой получена автоматически с помощью Euc1; (г) — MSBM, полная многоклассовая разметка для стандартного обучения которой получена автоматически с помощью Euc2.

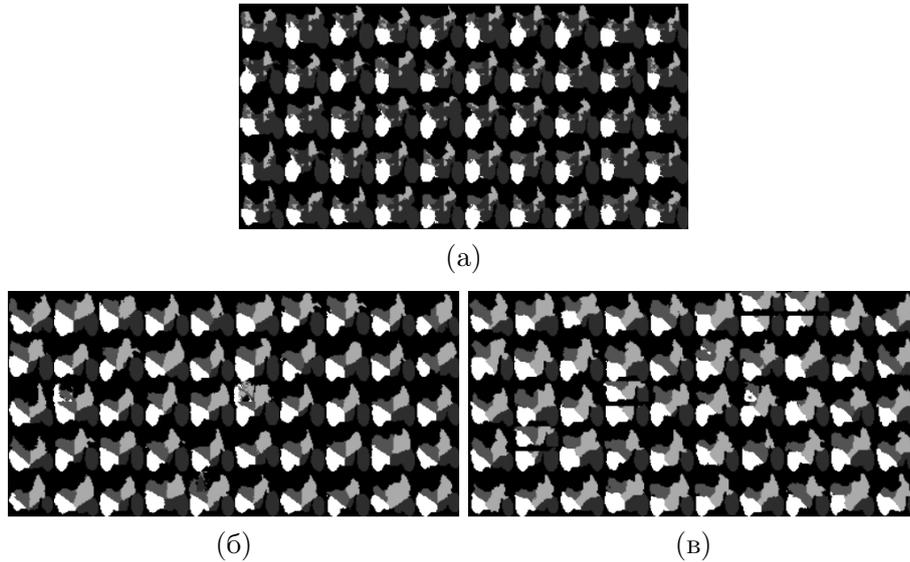


Рис. 8. Результаты генерации мотоциклов из: (а) — MSBM, обученную предлагаемым алгоритмом с использованием только бинарной разметки и семян; (б) — MSBM, полная многоклассовая разметка для стандартного обучения которой получена автоматически с помощью Euc1; (в) — MSBM, полная многоклассовая разметка для стандартного обучения которой получена автоматически с помощью Euc2.

Анализируя полученные результаты, можно заключить, что внешние границы объектов при генерации из MSBM, обученной нашим методом, и MSBM, обученной стандартным методом по данным с ручной многоклассовой разметкой, одинаково гладкие и формы объектов соответствуют их типу (модель MSBM, обученная на лошадях, генерирует новых лошадей). Однако, границы между отдельными частями объектов в нашей модели очень шумные. Это объясняется тем, что на этапе обучения наша модель получает данные о внешних границах объекта, но не о внутренних границах между частями. Данный недостаток не влияет на качество работы в задачах, связанных с генерацией бинарных форм.

Также можно заметить, что форма объектов, генерируемых из модели, для обучения которой многоклассовая разметка получалась автоматически, выглядит значительно хуже, чем для двух других

моделей. Возможной причиной этого является слишком сильное раз-  
нообразии в автоматически получаемых границах между частями,  
которое мешает настройке MSBM.

## 7.2. Восстановление формы

В данном эксперименте сравним все модели, приведенные в пунк-  
те 7, по качеству восстановления части изображения по всему осталь-  
ному изображению.

Рассмотрим бинарную маску объекта с известными семенами.  
Разделим все пиксели этой маски на два класса случайных перемен-  
ных: наблюдаемые  $\mathbf{b}^o$  и скрытые  $\mathbf{b}^u$ , семена при этом также делят-  
ся на два соответствующих класса. В случае с SBM можно оценить  
вероятность  $p(\mathbf{b}^u|\mathbf{b}^o) \approx \sum_{r=1}^R p(\mathbf{b}^u|\mathbf{H}_r)/R$ , а в случае MSBM — веро-  
ятность  $p(\mathbf{b}^u|\mathbf{b}^o, \mathbf{s}^o) \approx \sum_{r=1}^R p(\mathbf{b}^u|\mathbf{H}_r)/R$ , где  $\mathbf{H}_r$ ,  $r = \overline{1, R}$  — резуль-  
таты  $R$  итераций семплирования Гиббса из распределений  $p(\mathbf{H}|\mathbf{b}^o)$  и  
 $p(\mathbf{H}|\mathbf{b}^o, \mathbf{s}^o)$  соответственно.

Будем вычислять данные оценки для всех изображений тестовой  
выборки. Деление пикселей на классы зададим следующим образом:  
разделим изображение на 9 равных частей сеткой 3 на 3 и будем по-  
очередно брать в качестве скрытых переменных один из полученных  
сегментов. Итоговая мера качества представляет собой усредненный  
по разбиениям пикселей и тестовым изображениям логарифм оценен-  
ных вероятностей.

Результаты приведены на рис. 9. Сравнение новой модели с SBM и  
разными вариантами MSBM представлены на разных графиках для  
лучшей наглядности. Из графиков видно, что MSBM, использую-  
щая автоматическую многоклассовую разметку, показывает резуль-  
таты даже хуже, чем SBM. Модель, обученная с помощью нашего  
алгоритма, показывает результаты, сравнимые с результатами для  
MSBM, обученной по данным с полной разметкой, и превосходит мо-  
дель SBM.

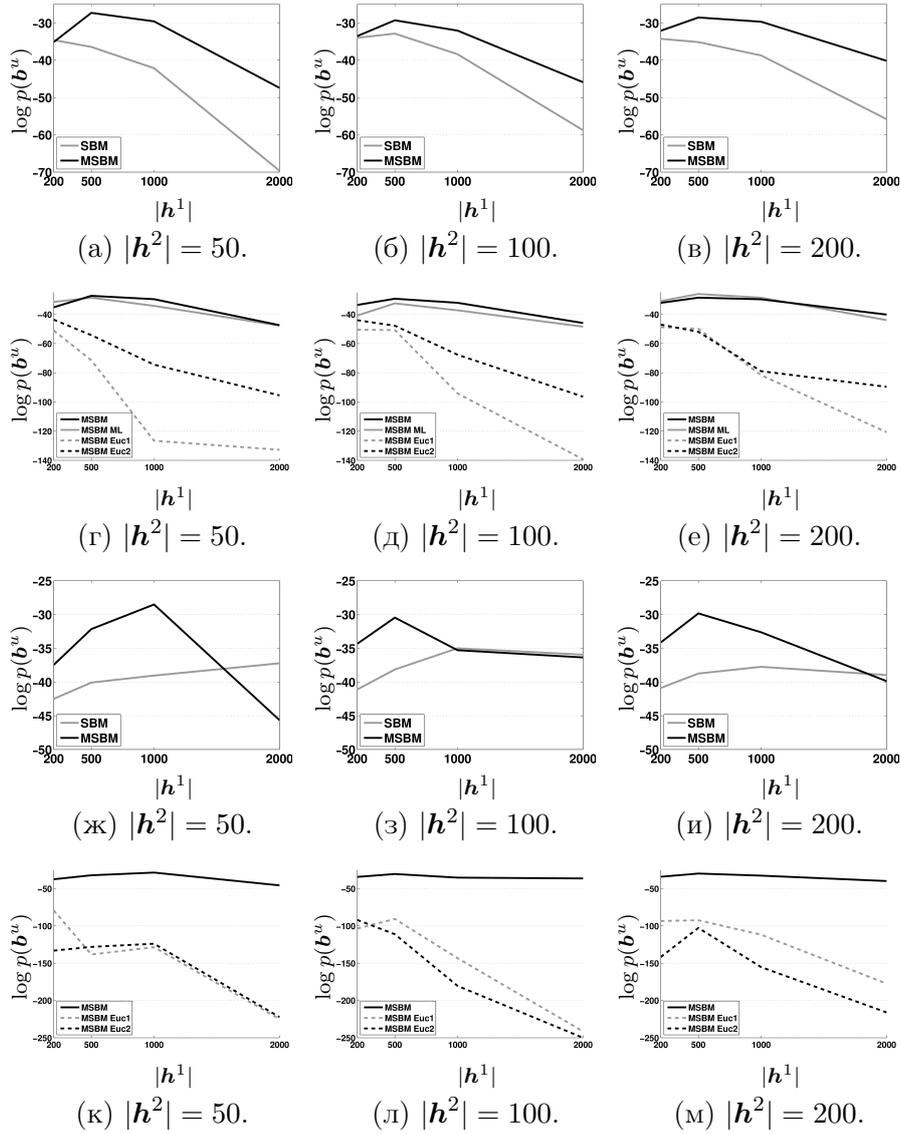


Рис. 9. Качество восстановления формы для разных моделей с различным числом переменных на скрытых слоях для изображений лошадей — (а)–(е) и для изображений мотоциклов — (ж)–(м). Чем выше значение значения меры, тем лучше.

### 7.3. Порождение формы из семян

В последнем эксперименте сравним по качеству порождаемых из семян форм нашу модель с SBM и MSBM, обученной по данным с ручной многоклассовой разметкой. Пусть даны семена  $\mathbf{s}$  и соответствующие им пиксели  $\mathbf{b}^s$ :  $f_{coord}(b_p^s) = s_p \forall p = \overline{1, P}$  принадлежат объекту. Сгенерируем значения скрытых переменных из распределений  $p(\mathbf{H}|\mathbf{b}^s)$  для SBM и из  $p(\mathbf{H}|\mathbf{b}^s, \mathbf{s})$  для MSBM. Далее сгенерируем форму  $\mathbf{b}$  или  $\mathbf{m}$  соответственно. Повторим эти две операции 100 раз и получим итоговые формы объекта. Примеры получаемых форм представлены на рис. 10. Итоговая мера качества представляет собой усредненное по тестовой выборке расстояние Хэмминга между полученной бинарной формой (в многоклассовом случае также бинаризуем результат) и исходной, для которой брались семена. Результаты приведены на рис. 11. Из графиков видно, что, если известны только семена частей объекта, MSBM значительно лучше восстанавливает форму объекта, чем SBM.

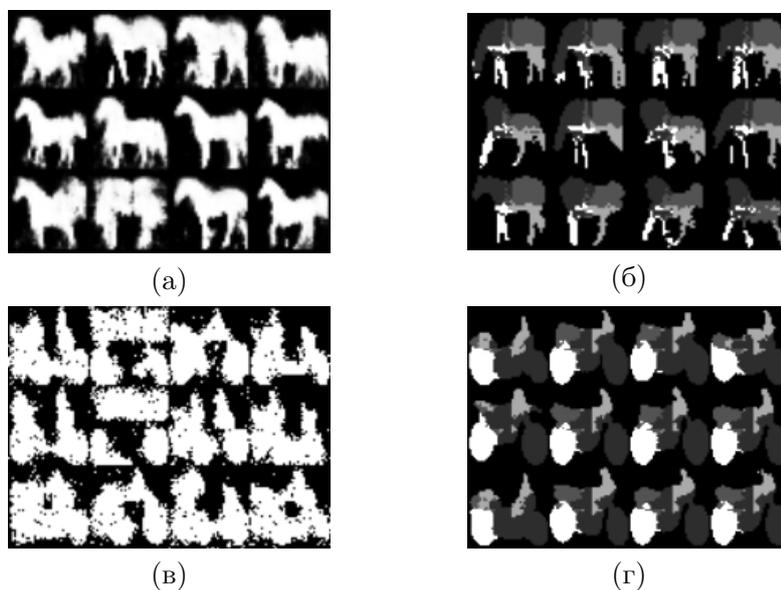


Рис. 10. Порождение формы из семян. (а), (в) формы сгенерированны из SBM; (б), (г) формы сгенерированны из MSBM, обученной нашим алгоритмом.

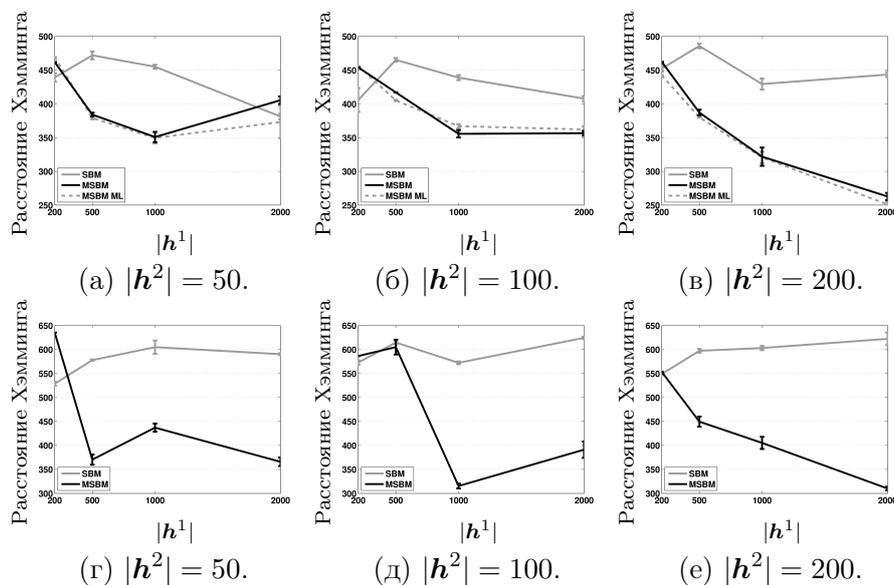


Рис. 11. Порождение формы из семян. Расстояние Хэмминга между сгенерированными из SBM и MSBM бинарными формами и исходными, для которой брались семена (чем ниже значение значения меры, тем лучше). (а), (б), (в) — формы лошадей; (г), (д), (е) — формы мотоциклов.

## 8. Заключение

В данной работе был предложен новый алгоритм обучения многоклассовой модели формы Больцмана MSBM, использующий лишь бинарные маски объектов и семена их частей. Было показано, что семена частей объектов могут быть получены автоматически с помощью детектора частей. Предложенный алгоритм обучения, представляющий собой EM-подобную процедуру, может быть использован как при разметке всех семян частей объектов, так и в случае, когда часть семян неизвестна. Проведенные эксперименты доказывают, что MSBM, обученная предлагаемым алгоритмом, превосходит по качеству SBM и сопоставима с MSBM, обученной стандартным способом. Также было показано, что предлагаемый метод обучения дает лучшие результаты, чем эвристическое получение многоклассовой разметки из бинарной и семян и последующее классическое обучение MSBM.

## Список литературы

- [1] Salakhutdinov R., Larochelle H. Efficient Learning of Deep Boltzmann Machines // Y.W. Teh, D.M. Titterington (eds.) AISTATS. — JMLR.org; 2010. — P. 693–700.
- [2] Кириллов А. Н., Гавриков М. И., Лобачева Е. М., Осокин А. А., Ветров Д. П. Многоклассовая модель формы со скрытыми переменными // 24-я международная конференция по компьютерной графике и зрению, ГрафиКон. — 2014. — С. 155–158.
- [3] Yangel B., Vetrov D. Image Segmentation with a Shape Prior Based on Simplified Skeleton // Energy Minimization Methods in Computer Vision and Pattern Recognition. — 2011. Series: Lecture Notes in Computer Science. Vol. 6819. — P. 247–260.
- [4] Hinton G. E., Osindero S., Teh Y. W. A fast learning algorithm for deep belief nets // Neural Computation. — 2006. Vol. 18, num. 7. — P. 1527–1554.
- [5] Salakhutdinov R., Hinton G. An Efficient Learning Procedure for Deep Boltzmann Machines // Neural Computation. — 2012. Vol. 24, num. 8. — P. 1967–2006.
- [6] Borenstein E. Combining top-down and bottom-up segmentation // IEEE workshop on Perceptual Organization in Computer Vision, CVPR. — 2004. — P. 46.
- [7] Ferrari V., Jurie F., Schmid C. From Images to Shape Models for Object Detection // International Journal of Computer Vision. — 2010. Vol. 87, num. 3. — P. 284–303.
- [8] Chan T. F., Shen J. Non-Texture Inpainting by Curvature-Driven Diffusions (CDD) // J. Visual Comm. Image Rep. — 2001. Vol. 12. — P. 436–449.
- [9] Van Ginneken B., Frangi A. F., Staal J. J., ter Haar Romeny B. M., Viergever M. A. Active shape model segmentation with optimal features // IEEE Transactions on Medical Imaging. — 2002. Vol. 21, num. 8. — P. 924–933.
- [10] Eslami S. M. A., Heess N., Williams C. K. I., Winn J. The Shape Boltzmann Machine: a Strong Model of Object Shape // International Journal of Computer Vision. — 2013.

- [11] Eslami S. M. A., Williams C. A Generative Model for Parts-based Object Segmentation // P. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, K. Q. Weinberger (eds.) *Advances in Neural Information Processing Systems 25*. — 2012. — P. 100–107.
- [12] Felzenszwalb P. F., Girshick R. B., McAllester D., Ramanan D. Object Detection with Discriminatively Trained Part Based Models // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. — 2010. Vol. 32, num. 9. — P. 1627–1645.
- [13] Bishop C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. — Springer-Verlag New York, Inc.: 2006.
- [14] Fei-Fei L., Fergus R., Perona P. One-shot learning of object categories // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. — 2006. Vol. 28 (4). — P. 594–611.