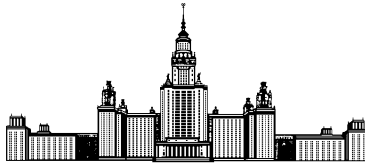


Московский государственный университет имени М. В. Ломоносова



Факультет Вычислительной Математики и Кибернетики

Кафедра Математических Методов Прогнозирования

## ДИПЛОМНАЯ РАБОТА СТУДЕНТА 517 ГРУППЫ

### «Применение $k$ -значных глубинных сетей для задания априорного распределения на форму объекта в задаче сегментации изображений»

Выполнил:

студент 5 курса 517 группы

*Гавриков Михаил Игоревич*

Научный руководитель:

к.ф.-м.н., доцент

*Ветров Дмитрий Петрович*

Заведующий кафедрой

Математических Методов

Прогнозирования, академик РАН

\_\_\_\_\_ Ю. И. Журавлёв

К защите допускаю

К защите рекомендую

«\_\_\_\_\_» \_\_\_\_\_ 2014 г.

«\_\_\_\_\_» \_\_\_\_\_ 2014 г.

Москва, 2014

## Аннотация

В задаче семантической сегментации изображений есть необходимость задания ограничений высокого порядка. Примером ограничений подобного рода является форма сегментируемого объекта. Форма объекта обычно задается бинарной маской, в тоже время зачастую объект имеет ярко выраженные части. В данной работе рассматривается задача построение модели формы с  $k$ -значными переменными, где объект состоит из  $k$  частей, и делается попытка отказаться на этапе обучения от полной разметки данных. Такой подход позволит значительно расширить область применения подобного рода моделей. В качестве дополнительной информации вместо полной разметки используются семена, задающие окрестности частей объекта. В части экспериментов приведено сравнение новой модели с уже имеющимися и показано, что новая модель сопоставима по качеству с моделью, обученной на полной разметке и лучше, чем бинарная модель.

# Содержание

<b>1</b>	<b>Введение</b>	<b>3</b>
<b>2</b>	<b>Обзор литературы</b>	<b>5</b>
2.1	Ограниченная машина Больцмана . . . . .	5
2.2	Глубинная машина Больцмана . . . . .	10
2.3	Усеченная глубинная машина Больцмана . . . . .	12
2.4	Многозначные видимые переменные . . . . .	13
<b>3</b>	<b>Постановка задачи</b>	<b>15</b>
<b>4</b>	<b>Многоклассовая модель формы с семенами</b>	<b>16</b>
4.1	Обозначения . . . . .	16
4.2	Совместное распределение . . . . .	17
4.3	Обучение модели формы с семенами . . . . .	19
<b>5</b>	<b>Модель детектора для автоматического получения семян</b>	<b>22</b>
<b>6</b>	<b>Эксперименты с моделью с семенами</b>	<b>23</b>
6.1	Метрики качества . . . . .	23
6.2	Результаты экспериментов . . . . .	25
6.3	Выводы . . . . .	27
<b>7</b>	<b>Заключение</b>	<b>28</b>
	<b>Список литературы</b>	<b>29</b>

# 1 Введение

Классический подход для решения задачи семантической сегментации изображений заключается в использовании случайных марковских полей, где вершины графа соответствуют пикселям изображения, а ребра соответствуют штрафу за отнесению пикселей, соединяемых ребром, к разным классам. Эффективный алгоритм, решающий данную задачу, умеет работать только с графами, максимальная клика в которых имеет малую степень. Такое ограничение на граф ведет к тому, что данный подход умеет эффективно работать только с локальными ограничениями. В тоже время зачастую в задаче семантической сегментации существует глобальное ограничение. Примером такого глобального ограничения может служить форма сегментируемого объекта. При попытке учесть такое ограничение возникают две задачи: как задавать форму объекта и как сегментировать объект используя его форму. В данной работе рассматривается первый вопрос: как задавать форму объекта.

Прежде чем переходить к вопросу, что такое форма объекта, рассмотрим пример. Пусть в качестве объекта у нас будет выступать лошадь, расположенная на изображении в профиль. Все лошади имеют определенную структуру: голова, туловище, 4 ноги и хвост, где ноги могут накладываться друг на друга, хвост может как присутствовать так и отсутствовать, причем хвост и голова не могут находиться рядом и так далее. Таких ограничений может быть очень много, и задавать их в явном виде не представляется возможным, поэтому будем обучать некоторую модель на множестве всех форм, которая сама учтет все эти ограничения. Теперь перейдем к вопросу, что же такое модель формы.

Под моделью формы будем понимать вероятностное распределение на множестве изображений, в котором изображения с интересующим нас объектом должно иметь высокую вероятность, а остальные низкую. В простом варианте форма интересующего нас объекта представима в виде бинарного изображения с классом фона и классом объекта. В работе же будет использоваться многоклассовая разметка, где класс объекта состоит из частей, и каждая часть имеет свою метку. Ключевым недостатком использования многозначных переменных является необходимость иметь полную разметку данных, которую, как правило, приходится получать вручную. Актуальность данной работы прежде всего в том, что в ней делается успешная

попытка получения многоклассовой модели формы не требующая полной разметки обучающих данных.

Будем задавать форму объекта при помощи генеративной вероятностной графической модели. Такое представление позволит сохранить знание о все возможных формах объектов, получать вероятности конфигураций, а также генерировать новые конфигурации, которых не было в исходном наборе данных. В качестве такой модели будем использовать разновидности машин Больцмана (Boltzmann Machines).

## 2 Обзор литературы

В этой секции мы подробно рассмотрим различные варианты машин Больцмана. При рассмотрении каждой машины Больцмана будем прежде всего отвечать на два вопроса: каким образом генерировать выборку из данного распределения (машины Больцмана это прежде всего генеративные модели) и каким образом осуществляется обучение модели.

Впервые машины Больцмана появились в [2], с тех пор они получили большое распространение и сейчас находят все больше и больше применений, одним из таких применений как раз и является задание модели формы. Самой простой разновидностью машин Больцмана является ограниченная машина Больцмана (Restricted Boltzmann Machine, [3]). Усложненным вариантом является глубинная машина Больцмана (Deep Boltzmann Machine, [4]) и еще более сложным вариантом является усеченная глубинная машина Больцмана (Shape Boltzmann Machine, [6]). Все эти модели первоначально строились из предположения, что исходные данные на видимом слое бинарные, тем не менее они все легко обобщаются на случай многозначных переменных, непрерывных на отрезке или непрерывных на всей оси [3]. Все приведенные выше модели представляют из себя графические модели с разным числом слоев и разной связностью между слоями.

### 2.1 Ограниченная машина Больцмана

На рисунке 1а представлена ограниченная машина Больцмана. Как уже было сказано выше, ограниченная машина Больцмана является самой простой моделью. Она представима в виде двудольного графа, где все вершины одной доли связаны со всеми вершинами другой доли. Нижний слой будем называть наблюдаемыми переменными и будем обозначать за  $\mathbf{v} \in \{0, 1\}^N$ , где  $N$  — число переменных на видимом слое, а верхний слой будем называть скрытыми переменными и будем обозначать за  $\mathbf{h} \in \{0, 1\}^M$ , где  $M$  — число переменных на скрытом слое. Энергия ограниченной машины Больцмана имеет следующий вид:

$$E(\mathbf{v}, \mathbf{h}; \theta) = - \sum_{i=1}^N v^i b_v^i - \sum_{j=1}^M h^j b_h^j - \sum_{i=1}^N \sum_{j=1}^M v^i W^{ij} h^j = -\mathbf{v}^T \mathbf{b}_v - \mathbf{h}^T \mathbf{b}_h - \mathbf{v}^T \mathbf{W} \mathbf{h},$$

где  $\mathbf{W}$  — матрица парных потенциалов размерами  $N$  на  $M$ ,  $\mathbf{b}_v, \mathbf{b}_h$  — вектора унарных потенциалов для видимых и скрытых переменных и  $\theta = \{\mathbf{W}, \mathbf{b}_v, \mathbf{b}_h\}$ .

Вероятность в этой модели вводится на основе распределения Гиббса. Оно обладает свойством, что имеет максимальную энтропию в классе распределений с энергией такого вида. Тогда вероятность примет вид:

$$P(\mathbf{v}, \mathbf{h}; \theta) = \frac{1}{Z(\theta)} e^{-E(\mathbf{v}, \mathbf{h}; \theta)}, \quad Z(\theta) = \sum_{\mathbf{v} \in \{0,1\}^N} \sum_{\mathbf{h} \in \{0,1\}^M} e^{-E(\mathbf{v}, \mathbf{h}; \theta)}$$

где  $Z(\theta)$  — нормировочная константа.

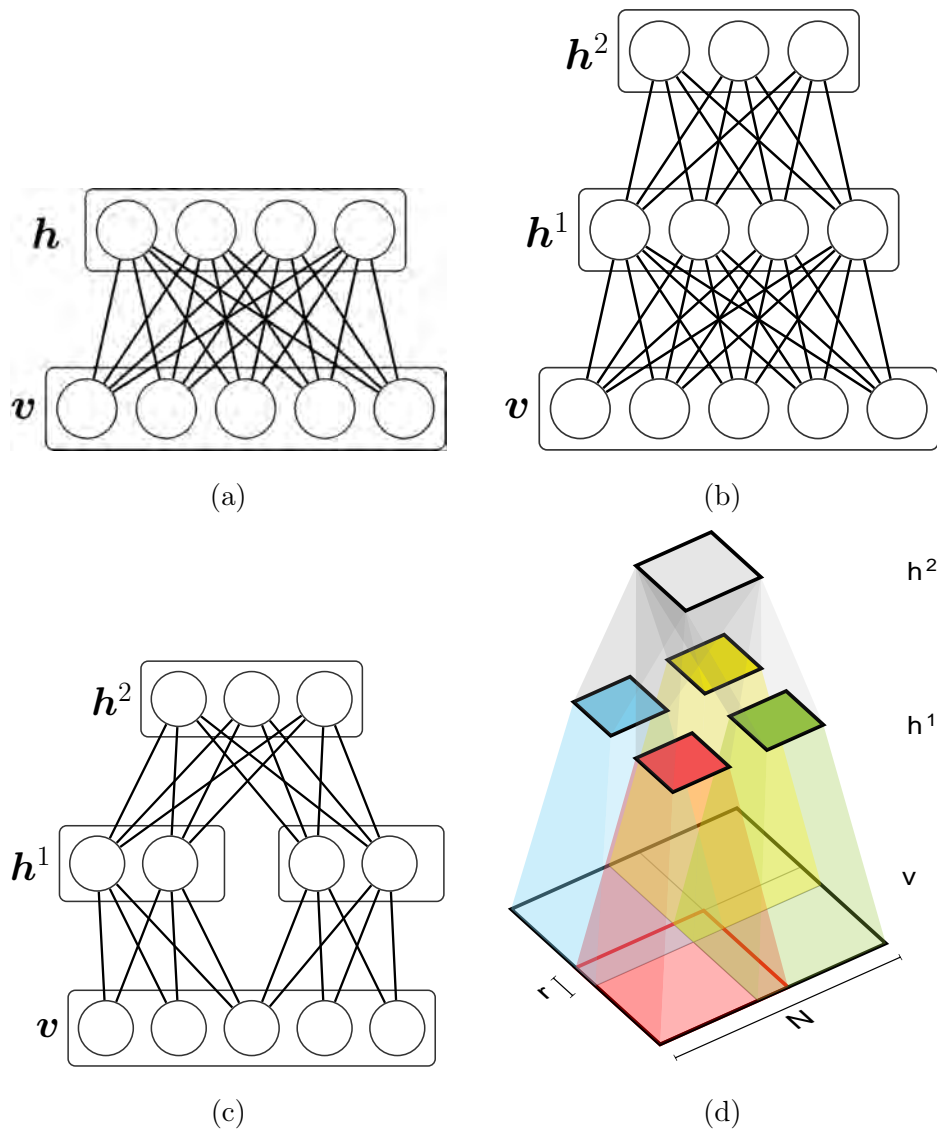


Рис. 1: Различные варианты машин Больцмана, на (a) — ограниченная, (b) — глубинная, (c) — усеченная глубинная (d) — усеченная глубинная для изображений.

Машины Больцмана являются генеративными моделями, это означает, что прежде всего надо уметь генерировать выборку из данного распределения. Но перед тем как переходить к вопросу генерации выборки, рассмотрим основное свойство модели ограниченной машины Больцмана: условной независимости переменных одного слоя при фиксации переменных другого слоя, это свойство следует из графического представления модели. Из него можно легко вывести сами формулы:

$$P(v^i = 1 | \mathbf{h}; \theta) = \sigma\left(\sum_{j=1}^M W^{ij} h^j + b_v^i\right) \quad (1)$$

$$P(h^j = 1 | \mathbf{v}; \theta) = \sigma\left(\sum_{i=1}^N W^{ij} v^i + b_h^j\right) \quad (2)$$

где  $\sigma(x) = \frac{\exp(x)}{1 + \exp(x)}$ .

Основным способом генерации выборки является метод марковских цепей Монте-Карло (Monte-Carlo Markov Chain, МСМС). Задается некоторая начальная инициализация переменных, неограничивая общности, будем считать, что заданы видимые переменные. Затем идет цикл некоторое фиксированное число итераций, где поочередно генерируются сначала скрытые переменные из распределения, заданного в формуле (2), затем переменные видимого слоя (1). Распределения из формул (1, 2) представляют по сути обычные бернулевские испытания, для генерации выборки из которых достаточно уметь генерировать выборку из равномерного распределения. Генерацию из распределения бернулли с вероятностью успеха  $p$  будет обозначать  $\mathbf{Be}(p)$ . Данный алгоритм так же имеет другое название процедура генерация Гиббса (Gibbs Sampling) и полная версия алгоритма представлена ниже:

**Вход:**  $K, \mathbf{v}^{start}$

**Выход:**  $\mathbf{v}, \mathbf{h}$

- 1:  $\mathbf{v} = \mathbf{v}^{start}$
- 2: для  $k = 1, \dots, K$
- 3: для  $j = 1, \dots, M$
- 4:  $h^j \sim \mathbf{Be}(\sigma(b_h^j + \sum_{i=1}^N W^{ij} v^i))$
- 5: для  $i = 1, \dots, N$
- 6:  $v^i \sim \mathbf{Be}(\sigma(b_v^i + \sum_{j=1}^M W^{ij} h^j))$



Существуют так же и другие способы генерации выборки, на основе метода Метрополиса-Хастингса, но этот метод обладаем существенным минусом, он требует больше затрат по памяти и по времени работы, поэтому его в данной работе не рассматриваем.

Теперь перейдем к другому важному вопросу: обучение модели. Для обучения ограниченной машины Больцмана будем максимизировать логарифм неполного правдоподобия:

$$\begin{aligned} \log P(\mathbf{v}; \theta) &= \log \sum_{\mathbf{h} \in \{0,1\}^M} P(\mathbf{v}, \mathbf{h}; \theta) = - \sum_{i=1}^N b_v^i v^i - \\ &\sum_{i=1}^N \log \left( 1 + \exp \left( \sum_{j=1}^M v^i W^{ij} + b_h^j \right) \right) - \log Z(\theta) \end{aligned}$$

Максимизация данного функционала происходит на основе стохастического градиентного подъема. Производные по первым двум суммам легко вычисляются. Рассмотрим производную  $\log Z(\theta)$  по некоторому параметру  $\phi$ :

$$\begin{aligned} \frac{\partial \log Z(\theta)}{\partial \phi} &= \frac{1}{Z(\theta)} \frac{\partial Z(\theta)}{\partial \phi} = \frac{1}{Z(\theta)} \sum_{\mathbf{v} \in \{0,1\}^N} \sum_{\mathbf{h} \in \{0,1\}^M} \frac{\partial \exp(-E(\mathbf{v}, \mathbf{h}; \theta))}{\partial \phi} = \\ &= \frac{1}{Z(\theta)} \sum_{\mathbf{v} \in \{0,1\}^N} \sum_{\mathbf{h} \in \{0,1\}^M} e^{-E(\mathbf{v}, \mathbf{h}; \theta)} \frac{\partial (-E(\mathbf{v}, \mathbf{h}; \theta))}{\partial \phi} = \\ &= \mathbb{E}_{\mathbf{v}, \mathbf{h}} \frac{\partial (-E(\mathbf{v}, \mathbf{h}; \theta))}{\partial \phi} \end{aligned}$$

Тогда итоговые производные примут вид:

$$\frac{\partial P(\mathbf{v}; \theta)}{\partial W^{ij}} = \mathbb{E}_{\mathbf{v}, \mathbf{h}} v^i h^j - v^i \sigma \left( \sum_{j=1}^M v^i W^{ij} + b_h^j \right) \quad (3)$$

$$\frac{\partial P(\mathbf{v}; \theta)}{\partial b_v^i} = \mathbb{E}_{\mathbf{v}, \mathbf{h}} v^i - v^i \quad (4)$$

$$\frac{\partial P(\mathbf{v}; \theta)}{\partial b_h^j} = \mathbb{E}_{\mathbf{v}, \mathbf{h}} h^j - \sigma \left( \sum_{j=1}^M v^i W^{ij} + b_h^j \right) \quad (5)$$

Вычисление данных математических ожиданий NP-трудно, поэтому оценивать его будем на основе метода монте карло: генерируем точки из данного распределения

и усредняем значения функции под математическим ожиданием по ним:

$$\mathbb{E}_{\mathbf{v}, \mathbf{h}} \frac{\partial(-E(\mathbf{v}, \mathbf{h}; \theta))}{\partial \phi} \approx \frac{1}{K} \sum_{k=1}^K \frac{\partial(-E(\mathbf{v}^k, \mathbf{h}^k; \theta))}{\partial \phi}$$

Вектора  $\mathbf{v}^k, \mathbf{h}^k$  — получаются применением процедуры генерации гиббса на  $k$ -ом шаге. В статье [3] экспериментально показано, что для ограниченной машины Больцмана достаточно одного шага ( $K = 1$ ), если в качестве инициализации видимых переменных использовать обучающую выборку.

У ограниченной машины Больцмана, как и у других машин Больцмана есть недостаток: NP-трудно посчитать нормировочную константу, а значит невозможно посчитать оптимизируемый функционал. В статье [5] описывается приближенный способ подсчета нормировочной константы, тем не менее предложенная процедура работает очень долго и использовать ее на практике для оценки оптимизируемого функционала на каждой итерации невозможно, поэтому на практике используют фиксированное число итераций и фиксированную величину шаг по градиенту. Так же на практике используют регуляризатор вида: квадрат евклидовой нормы параметров модели умноженный на некоторый коэффициент. Регуляризатор данного вида легко добавляется в формулы с градиентами, поэтому явно приводить их не будем. Приведем алгоритм обучения:

**Вход:**  $K, \mathbf{v}_i, i = \overline{1, D}, NumIter, \epsilon$

**Выход:**  $\theta$

- 1: Инициализуем случайно  $\theta$ .
- 2: для  $Iter = 1, \dots, NumIter$
- 3: Перемешать объекты
- 4: для  $i = 1, \dots, D$
- 5: Применить процедуру генерации Гиббса с инициализацией  $\mathbf{v}_i$ , сделав  $K$  шагов и запомнить результат
- 6: Сделать шаг по градиенту:  $\theta_{new} = \theta_{old} + \epsilon \nabla \theta$ , где математические ожидания в формулах для градиентов оцениваются по выборке сгенерированной на предыдущем шаге. Градиенты вычисляются по формулам (3, 4, 5).

## 2.2 Глубинная машина Больцмана

Глубинная машина Больцмана отличается от ограниченной тем, что у нее не два слоя, а три или больше. Рассмотрим для простоты модель с тремя слоями. Нижний слой будем называть видимым, а два верхних — скрытыми. Иллюстрация к данной модели приведена на рисунке 1b. Первый скрытый слой будем обозначать за  $\mathbf{h}^1 \in \{0, 1\}^M$ , второй —  $\mathbf{h}^2 \in \{0, 1\}^P$  и параметры модели как и раньше будем обозначать за  $\theta = \{\mathbf{W}_1, \mathbf{W}_2, \mathbf{b}_v, \mathbf{b}_1, \mathbf{b}_2\}$ . Энергия и вероятность в этой модели задается аналогичным способом:

$$E(\mathbf{v}, \mathbf{h}^1, \mathbf{h}^2; \theta) = -\mathbf{v}^T \mathbf{W}_1 \mathbf{h}^1 - \mathbf{h}^{1T} \mathbf{W}_2 \mathbf{h}^2 - \mathbf{v}^T \mathbf{b}_v - \mathbf{h}^{1T} \mathbf{b}_1 - \mathbf{h}^{2T} \mathbf{b}_2$$

$$P(\mathbf{v}, \mathbf{h}^1, \mathbf{h}^2; \theta) = \frac{\exp(-E(\mathbf{v}, \mathbf{h}^1, \mathbf{h}^2; \theta))}{Z(\theta)}$$

$$Z(\theta) = \sum_{\mathbf{v} \in \{0,1\}^N} \sum_{\mathbf{h}^1 \in \{0,1\}^M} \sum_{\mathbf{h}^2 \in \{0,1\}^P} e^{-E(\mathbf{v}, \mathbf{h}^1, \mathbf{h}^2; \theta)}$$

Для генерации выборки из этого распределения будем использовать процедуру генерацию Гиббса, но она будет незначительно отличаться от описанной выше. Здесь используется тот факт, что видимый слой и второй скрытый становятся условно независимы при фиксации первого скрытого. Тогда формула условной вероятности видимого при условии первого скрытого слоя останется такой же, как и в формуле (1). Таким образом получаются следующие формулы вероятностей:

$$P(h^{1j} = 1 | \mathbf{v}, \mathbf{h}^2; \theta) = \sigma\left(\sum_{i=1}^N v^i W_1^{ij} + \sum_{k=1}^P h^{2k} W_2^{jk} + b_1^j\right)$$

$$P(h^{2k} = 1 | \mathbf{h}^1; \theta) = \sigma\left(\sum_{j=1}^M h^{1j} W_2^{jk} + b_2^k\right)$$

$$P(v^i = 1 | \mathbf{h}; \theta) = \sigma\left(\sum_{j=1}^M W_1^{ij} h^j + b_v^i\right)$$

Процедура генерации Гиббса проиллюстрирована на рисунке 2. Если сравнивать процедуру генерации Гиббса для ограниченной машины Больцмана с процедурой для глубинной модели, то отличие лишь в том, что теперь надо генерировать из распределения бернулли переменные для еще одного слоя, и несколько поменялась зависимость первого скрытого слоя от остальных.

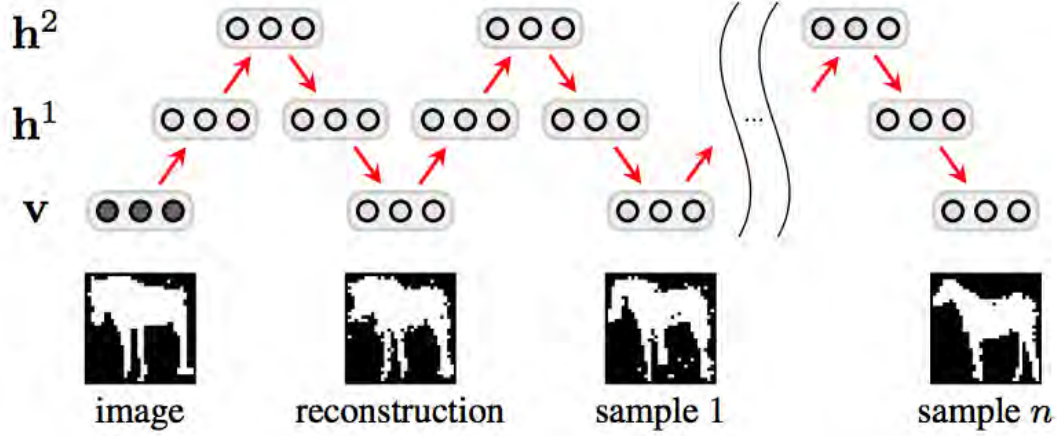


Рис. 2: Иллюстрация к процедуре генерации выборки для глубоких моделей Больцмана

Обучение глубокой машины Больцмана сильно отличается. В данном случае мы не можем в явном виде просуммировать по скрытым переменным, поэтому будем использовать так называемый вариационный EM-алгоритм. Где на E-шаге будем искать некоторое распределение  $q(\mathbf{h}^1, \mathbf{h}^2)$  такое, что оно в некотором смысле является наиболее близким к распределению  $P(\mathbf{h}^1, \mathbf{h}^2 | \mathbf{v}; \theta)$ , а на M-шаге будем:

$$\sum_{d=1}^D \mathbb{E}_{q_d(\mathbf{h}^1, \mathbf{h}^2)} \log P(\mathbf{v}_d, \mathbf{h}^1, \mathbf{h}^2; \theta) \rightarrow \max_{\theta}$$

Подробное описание вариационного EM-алгоритма можно найти в [7, стр. 466]. Приведем лишь краткие выкладки. На E-шаге решаем задачу минимизации дивергенции Кульбака-Лейблера по первому аргументу:

$$\text{KL}(q(\mathbf{h}^1, \mathbf{h}^2) \| P(\mathbf{h}^1, \mathbf{h}^2 | \mathbf{v}; \theta)) \rightarrow \min_q$$

Эта задача перестанет быть NP-трудной, если искать  $q$  в полностью факторизованном семействе распределений, то есть когда каждая случайная переменная становится независимой от других ( $q(\mathbf{h}^1, \mathbf{h}^2) = \prod_{j=1}^M \prod_{k=1}^P q^1(h_j^1) q^2(h_k^2)$ ). Тогда данная задача минимизации решается методом покоординатного спуска, проинициализировавшись некоторым случайным образом:

$$q(h_j^1 = 1) = \sigma(b_1^j + \sum_{i=1}^N v^i W_1^{i,j} + \sum_{k=1}^P W_{j,k}^2 q_k^2(h_k^2 = 1))$$

$$q(h_k^2 = 1) = \sigma(b_2^k + \sum_{j=1}^M W_{j,k}^2 q_j^1(h_j^1 = 1))$$

Распишем M-шаг для одного объекта  $\mathbf{v}$ :

$$\begin{aligned} \mathbb{E}_{q(\mathbf{h}^1, \mathbf{h}^2)} \log P(\mathbf{v}, \mathbf{h}^1, \mathbf{h}^2; \theta) &= \sum_{i=1}^N \sum_{j=1}^M v^i W_1^{ij} q_1(h^{1j} = 1) + \\ &\sum_{k=1}^P \sum_{j=1}^M q_2(h^{2k} = 1) W_2^{jk} q_1(h^{1j} = 1) + \sum_{k=1}^P q_2(h^{2k} = 1) b_2^k + \\ &\sum_{j=1}^M q_1(h^{1j} = 1) b_1^j + \sum_{i=1}^N v^i b_v^i - \log Z(\theta) \end{aligned}$$

Как уже было показано выше: производная по логарифму нормировочной константы превращается в математическое ожидание, которое мы опять будем оценивать через метод монте-карло, генерируя выборку через процедуру генерации Гиббса. На M-шаге мы не будем делать полную максимизацию, а лишь несколько градиентных шагов. Как и в ограниченной машине Больцмана будем вводить регуляризатор и использовать фиксированное число шагов и неизменяемую величину шага по градиенту.

## 2.3 Усеченная глубинная машина Больцмана

На рисунках 1c, 1d представлены усеченные глубинные машины Больцмана. Основное отличие этих моделей от глубинной машины Больцмана в том, как связаны переменные видимого слоя и переменные первого скрытого слоя. Рассмотрим сначала вид связей для усеченной глубинной машины Больцмана, а уже потом вариант адаптированный для изображений.

Разделим видимые переменные на две равные части, возможно пересекающиеся и обозначим их за  $V_1, V_2$ . Также разделим переменные первого скрытого слоя на две равные непересекающиеся части и обозначим их за  $H_1, H_2$ . Если раньше были все связи между  $V_1, V_2$  и  $H_1, H_2$ , то теперь оставим только  $\{(x, y) | x \in V_1, y \in H_1\}$  и  $\{(x, y) | x \in V_2, y \in H_2\}$ . То есть мы уменьшили число весов между видимым и первым скрытым слоем практически в два раза (если пересечение множеств  $V_1, V_2$  будем пустым, то будет в точности в два раза). Обозначим матрицу связей для ребер  $\{(x, y) | x \in V_1, y \in H_1\}$  за  $\mathbf{W}_1^1$ , а для ребер  $\{(x, y) | x \in V_2, y \in H_2\}$  за  $\mathbf{W}_1^2$ . Отождествив

матрицы  $\mathbf{W}_1^1$  и  $\mathbf{W}_1^2$ , мы уменьшим число весов еще в два раза. Полученная модель и называется усеченной глубинной машиной Больцмана. Ясно, что можно разбивать видимый слой и первый скрытый не на две равных части, а на большее их количество.

Для модели применяемой к картинкам будем делать все тоже самое, но будем иметь в виду, что работаем с изображениями. Пусть на видимом слое у нас изображением размерами  $N$  на  $N$ , тогда разделим весь видимый слой сеткой  $2$  на  $2$  с пересечениями шириной в  $r$  пикселей, как показано на рисунке 1d. Верхний слой тоже разделим на 4 равные части. И опять оставим только веса между соответствующими блоками, отождествив переменные в оставшихся блоках между собой.

Для удобства записи энергии по прежнему будем представлять в том же виде, как и в глубинной машине Больцмана, просто неявно будем подразумевать, что часть весов в матрице  $\mathbf{W}_1$  занулено и отождествлено.

## 2.4 Многозначные видимые переменные

До этого момента, мы рассматривали модели, где подразумевалось, что видимые переменные бинарные. Теперь рассмотрим модель с  $K$ -значными переменными на видимом слое. Такое изменение повлияет только на вид связей между первым скрытым слоем и видимым слоем. Видимые переменные будем обозначать за  $\mathbf{m} \in \{0, \dots, K\}^N$ . Также будем использовать и другое представление  $\mathbf{m} \in \{0, 1\}^{K \cdot M}$ , где из того, что  $m^i = p$  следует, что  $m^{ip} = 1$  и  $\sum_{p=0}^{K-1} m^{ip} = 1, i = \overline{1, N}$ . На каждый класс пикселя видимого слоя должны быть свои веса с первого скрытого слоя и будем их обозначать за  $\mathbf{W}_1^p, p = \overline{0, K-1}$ , тоже самое произойдет и с унарными потенциалами и обозначать их будем за  $\mathbf{b}_v^p, p = \overline{0, K-1}$ . Тогда энергия будет иметь следующий вид:

$$E(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2; \theta) = - \sum_{p=0}^{K-1} \sum_{i=1}^N \sum_{j=1}^M W_1^{ijp} m^{ip} h^{1j} - \sum_{p=0}^{K-1} \sum_{i=1}^N m^{ip} b_v^{ip} - \sum_{j=1}^M h^{1j} b_1^j - \sum_{k=1}^P h^{2k} b_2^k - \sum_{j=1}^M \sum_{k=1}^P h^{1j} h^{2k} W_2^{jk}$$

Исходя из нового вида энергии и ограничение вида  $\sum_{p=0}^{K-1} m^{ip} = 1, i = \overline{1, N}$  произойдут некоторые изменения в процедуре генерации Гиббса и на этапе обучения.

Приведем для краткости только, то что поменяется:

$$p(m^i = p | \mathbf{h}^1; \theta) = \frac{\exp(\sum_{j=1}^K W_1^{ijp} h^{1j} + b_v^{ip})}{\sum_{p'=0}^{K-1} \exp(\sum_{j=1}^K W_1^{ijp'} h^{1j} + b_v^{ip'})}$$

Генерировать случайные числа из такого дискретного распределения можно на основе метода обратной функции, то есть генерируем случайное число из равномерного распределения и смотрим на какой из отрезком функции распределения, принимающей  $K + 1$  значение попадает сгенерированная точка.

$$p(h^{1j} = 1 | \mathbf{m}, \mathbf{h}^2; \theta) = \sigma\left(\sum_{p=0}^{K-1} \sum_{i=1}^N W_1^{ijp} m^{ip} + b_1^j \sum_{k=1}^P W_2^{jk} h^{2k}\right)$$

На этапе обучения также будем использовать вариационный EM-алгоритм, где по сравнению с бинарной глубинной машиной Больцмана на E-шаге поменяется формула итерационного пересчета для  $q_1(h^{1j})$ , приведем ее:

$$q(h_j^1 = 1) = \sigma\left(b_1^j + \sum_{p=0}^{K-1} \sum_{i=1}^N m^{ip} W_1^{ijp} + \sum_{k=1}^P W_{j,k}^2 q_k^2(h_k^2 = 1)\right)$$

В статье [10] показано, что многоклассовая модель по ряду метрик «ведет себя лучше», чем бинарная модель, так же в статье показано, что для набора данных с лошадьми Weizmann Horse Dataset([8]), оптимальное разбиение на число частей в усеченной глубинной машине Больцмана составляет 4, то есть сеткой два на два. Главным недостатком многоклассовых машин Больцмана, является то, что для их обучения необходимо иметь полные размеченные данные. С практической точки зрения это означает, что это разметку необходимо делать руками.

### 3 Постановка задачи

Задачей данной работы является построение модели формы, которая на этапе обучения не требует полной разметки данных, но в результате получается многоклассовая модель, которая по качеству должна быть лучше, чем бинарная модель и сопоставима с многоклассовой.

Отказавшись от любой дополнительной информации, кроме как разметки объекта, мы не сможем обучить многоклассовую модель. Поэтому встает вопрос выбора, что нам будет известно об объекте, кроме как бинарной разметки. В данной работе в качестве дополнительной информации были выбраны семена, отвечающие в некотором смысле за центр класса, за которое отвечает семя. Подробнее это описывается в следующей секции, где вводится предложенная модель. Так же в данной работе будет рассмотрен вариант автоматического получения семян для изображений.



## 4 Многоклассовая модель формы с семенами

В данной секции подробно описывается модель формы с семенами, которая позволяет обучать многоклассовую машину Больцмана без использования полной многоклассовой разметки. В качестве дополнительной информации используем семена, их количество соответствует числу частей объекта и каждое семя отвечает за конкретную часть объекта. Рассматривается вариант обучения, в котором для каждого объекта известны всего его семена, а так же вариант, когда для части объектов информация о семенах полностью или частично отсутствует. Модель формы с семенами позволяет обучать многоклассовую усеченную глубинную машину Больцмана не требуя при этом полной разметки данных.

### 4.1 Обозначения

Пусть имеется  $D$  изображений,  $\mathbf{b}^d$  — бинарная маска объекта  $D$ -го изображения,  $\mathbf{m}^d$  — многоклассовая разметка изображения, где каждый класс отвечает за какую-то часть объекта. Пусть  $P$  — число частей объекта, тогда  $\mathbf{m}_i^d \in \{0, \dots, P\}$ , где 0 отвечает за фон. За  $\mathbf{s}_p^d$  обозначим координаты семени, отвечающего за  $p$ -ую часть объекта, для изображения  $d$ , то есть явно подразумевается, что для каждой части объекта задано одно семя. Под семенем будем понимать точку, в окрестности которой располагается интересующая нас часть объекта. Индекс  $d$  у семени  $\mathbf{s}^d$  будем опускать. Для наглядной иллюстрации смысла введенных обозначений рассмотрим рисунок 3. Заметим, что семя не обязательно должно принадлежать объекту. Так же для дальнейших рассуждений нам понадобится функция  $f_{coord}(i)$ , которая принимает на вход номер пикселя  $i$ , а возвращает его координаты на изображении  $(x_i, y_i)$ . В таком случае запись  $\|\mathbf{s}_{m_i} - f_{coord}(i)\|^2$  означает, что считаем квадрат евклидовой нормы между положением  $i$ -го пикселя на изображении и семенем, к которому принадлежит  $i$ -ий пиксель.

Как и до этого за  $\mathbf{h}^1, \mathbf{h}^2$  обозначим скрытые переменные многоклассовой усеченной глубинной машины Больцмана, а за  $\theta$  все ее параметры.

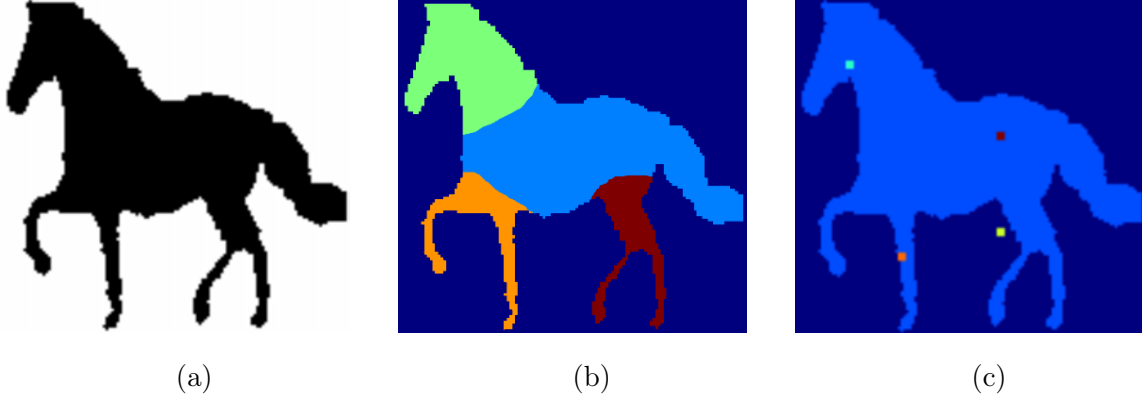


Рис. 3: Иллюстрация к введенным обозначениям, (a) — Бинарная разметка (обозначается за  $\mathbf{b}$ ), (b) — Многоклассовая разметка (обозначается за  $\mathbf{m}$ ), (c) — Бинарная разметка вместе с семенами (положение семян обозначается за  $\mathbf{s}_p$ ).

## 4.2 Совместное распределение

Рассмотрим теперь совместное распределение на все переменные:  $p(\mathbf{b}, \mathbf{s}, \mathbf{m}, \mathbf{h}^1, \mathbf{h}^2|\theta)$ . И введем дополнительное предположение о взаимосвязях переменных: бинарная маска объекта и координаты семян условно независимы при фиксированной многоклассовой разметке. Графическое изображение связей случайных переменных представлено на рисунке 4. Такое предположение берется из простой идеи: семена и бинарная маска объекта однозначно выражается через многоклассовую разметку. Таким образом верно следующее:

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

Теперь рассмотрим каждый из множителей в отдельности. Под распределением  $p(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2|\theta)$  будем понимать многоклассовую усеченную глубинную машину Больцмана, то есть:

$$E(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2; \theta) = - \sum_{p=0}^{K-1} \sum_{i=1}^N \sum_{j=1}^M W_1^{ijp} m^{ip} h^{1j} - \sum_{p=0}^{K-1} \sum_{i=1}^N m^{ip} b_v^{ip} - \sum_{j=1}^M h^{1j} b_1^j - \sum_{k=1}^P h^{2k} b_2^k - \sum_{j=1}^M \sum_{k=1}^P h^{1j} h^{2k} W_2^{jk}$$

$$P(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2|\theta) = \frac{\exp(-E(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2; \theta))}{Z(\theta)},$$

где в матрице связей  $\mathbf{W}_1$  некоторые веса занулены, а некоторые отождествлены, как и рассматривалось выше.

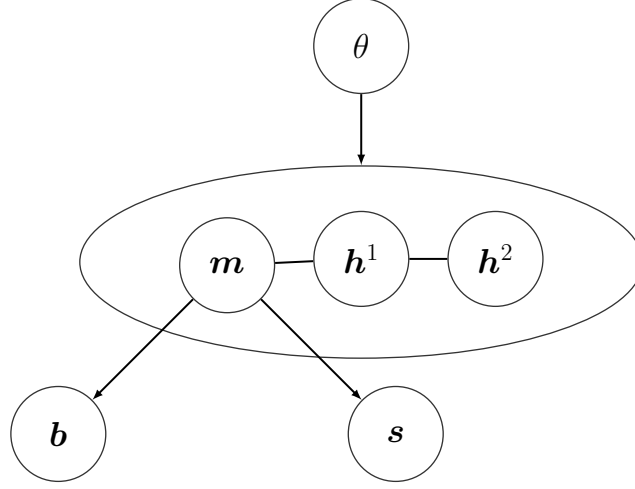


Рис. 4: Иллюстрация взаимосвязей переменных в модели формы с семенами.

Также очевидно, что  $i$ -ый пиксель бинарной маски зависит лишь от  $i$ -го пикселя многоклассовой разметки, и распределение  $p(\mathbf{b}|\mathbf{m})$  тогда представимо в виде  $\prod_i p(b_i|m_i)$ . Вероятность  $p(b_i|m_i)$  для всех  $i$ , в свою очередь является вырожденной, в том смысле что не имеет стохастической природы, а представима в виде детерминированной функции и задавать будем следующим образом:

$$p(b_i = x|m_i) = \begin{cases} 1 & \text{Если } x = 0 \text{ и } m_i = 0 \\ 1 & \text{Если } x = 1 \text{ и } m_i \neq 0 \\ 0 & \text{Иначе} \end{cases} \quad (6)$$

Или через скобки Айверсона:

$$p(b_i = 0|m_i) = [b_i = 0][m_i = 0] + [b_i \neq 0][m_i \neq 0]$$

Определим вероятность  $p(\mathbf{s}|\mathbf{m})$ . В основу этого распределения положим следующие два предположения:

1.  $i$ -ый пиксель объекта «тянет на себя» семя  $s_{m_i}$
2. для  $i$ -го пикселя объекта верно, что он независим со всеми  $s_j$ , где  $j \neq m_i$

Исходя из представленных предположений введем вероятность:

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

Параметр  $\sigma$  будем считать фиксированным и заданным заранее.

### 4.3 Обучение модели формы с семенами

Будем считать, что  $\mathbf{b}$ ,  $\mathbf{s}$  это наблюдаемые переменные, а  $\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2$  — скрытые. Поэтому на этапе обучения мы будем максимизировать логарифм неполного правдоподобия:

$$\sum_d \log p(\mathbf{b}^d, \mathbf{s}^d | \theta) \rightarrow \max_{\theta}$$

Для решения подобного рода задач используют вариационный EM-алгоритм, где на E-шаге решается задача:

$$\min_q \text{KL}(q(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2) \| p(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2 | \mathbf{b}, \mathbf{s})) \quad (7)$$

на M-шаге:

$$\max_{\theta} \sum_{d=1}^D [\mathbb{E}_{\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2 \sim q(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2)} \log p(\mathbf{m}^d, \mathbf{h}^1, \mathbf{h}^2 | \theta)] . \quad (8)$$

Задача на E-шаге NP-трудная, поэтому воспользуемся классическим подходом: будем приближать распределение  $p(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2)$  полностью факторизованным распределением  $q(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2) = \prod_i q_i^m(\mathbf{m}_i) \prod_j q_j^1(h_j^1) \prod_k q_k^2(h_k^2)$ . Минимизация дивергенции Кульбака-Лейблера по первому аргументу в семействе полностью факторизованных распределений решается покоординатным спуском, где на каждом пересчитываются вероятности следующим образом (подробное описание откуда получается данная формула можно найти в [7, стр. 466]):

$$\begin{aligned} q_t(x) &\propto \exp(\mathbb{E}_t \log p(\mathbf{b}^d, \mathbf{s}^d, \mathbf{m}^d, \mathbf{h}^1, \mathbf{h}^2 | \theta)) \\ &= \exp(\mathbb{E}_t \log p(\mathbf{b} | \mathbf{m}) p(\mathbf{s} | \mathbf{m}) p(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2 | \theta)) , \end{aligned} \quad (9)$$

где  $t$  — это номер распределения и соответствующей ему переменной в упорядоченной последовательности:  $q_i^m(m_i) \forall i, q_j^1(h_j^1) \forall j, q_k^2(h_k^2) \forall k$ ,  $x$  соответствующая этому номеру переменная, а  $\mathbb{E}_t$  — математическое ожидание по распределению  $q$  без переменной отвечающей номеру  $t$ .

Используя формулу (9), свойства линейности математического ожидания и логарифма произведений получаем:

$$q_i^m(m_i = p) \propto [b_i \neq 0][p \neq 0] \exp\left(-\frac{1}{2\sigma^2} \|s_{m_i} - f_{coord}(i)\|_2^2 + a_{i,m_i} + \sum_j W_{i,j,m_i}^1 q_j(h_j^1 = 1)\right) + [b_i = 0][p = 0] \quad (10)$$

$$q_j^1(h_j^1 = 1) = \sigma(c_j^1 + \sum_{i,p=1}^P m_{i,p} W_{i,j,p}^1 + \sum_k W_{j,k}^2 q_k^2(h_k^2 = 1)), \quad (11)$$

$$q_k^2(h_k^2 = 1) = \sigma(c_k^2 + \sum_j q_j^1(h_j^1 = 1) W_{j,k}^2). \quad (12)$$

Знак пропорциональности в формуле (10) означает, что необходимо посчитать нормировочную константу для каждого  $q_i^m$ : просуммировав по всем  $p$  выражение стоящее в правой части формулы (10).

Теперь подробно рассмотрим M-шаг:

$$\begin{aligned} & \sum_{d=1}^D [\mathbb{E}_{\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2 \sim q(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2)} \log p(\mathbf{m}^d, \mathbf{h}^1, \mathbf{h}^2 | \theta)] = \\ & \sum_{d=1}^D [-\mathbb{E}_{\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2 \sim q(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2)} E(\mathbf{m}^d, \mathbf{h}^1, \mathbf{h}^2 | \theta) - \log Z(\theta)] = \\ & \sum_{d=1}^D \left( \sum_{i=1}^I \sum_{p=1}^P a_{i,p} q_i^m(m_i^d = p) + \sum_{i=1}^I \sum_{j=1}^J \sum_{p=1}^P q_i^m(m_i^d = p) W_{i,j,p}^1 q_j^1(h_j^1 = 1) + \right. \\ & \quad \left. \sum_{j=1}^J c_j^1 q_j^1(h_j^1 = 1) + \sum_{j=1}^J \sum_{k=1}^K q_j^1(h_j^1 = 1) W_{j,k}^2 q_k^2(h_k^2 = 1) + \right. \\ & \quad \left. \sum_{k=1}^K c_k^2 q_k^2(h_k^2 = 1) - \log Z(\theta) \right). \quad (13) \end{aligned}$$

Здесь мы воспользовались свойством линейности математического ожидания, полной факторизованностью распределения  $q(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2)$ , а также тем, что математическое ожидание бернулевской случайной величины совпадает с вероятностью выпадения 1 для этой величины. Стоит также отметить, что в этой формуле неявно подразумевается зависимость  $q(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2)$  от  $d$ .

Максимизация формулы (13) по параметрам  $\theta$  происходит так же как и в случае с усеченной глубинной машины Больцмана на основе стохастического градиентного спуска.

Также возможно обучение с частично заданными семенами. То есть для некоторых объектов часть семян, а может и все отсутствуют. В результате такого требования изменится только E-шаг. Для удобства обозначим за  $\mathbf{s}^o$  — известные семена и за  $O$  множество индексов,  $\mathbf{s}^u$  — неизвестные и  $U$  — индексы неизвестных семян. Если неизвестных семян нету, то задача становится в точности как и описано выше. Поэтому считаем, что множество неизвестных семян непусто. Теперь на E-шаге будем решать задачу:

$$\text{KL}(q(\mathbf{s}^u, \mathbf{m}, \mathbf{h}^1, \mathbf{h}^2) \| p(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2, \mathbf{s}^u | \mathbf{b}, \mathbf{s}^o)) \rightarrow \min_q$$

Будем по тем же причинам рассматривать полную факторизацию распределения  $q$ :  $q(\mathbf{m}, \mathbf{h}^1, \mathbf{h}^2, \mathbf{s}^u) = \prod_i q_i^m(m_i) \prod_j q_j^1(h_j^1) \prod_k q_k^2(h_k^2) \prod_{p \in U} \delta(s_p^u)$ . В данном случае под  $\delta(s_p^u)$  подразумевается дельта-функция, то есть вырожденная функция распределения, принимающая ненулевое значение в одной точке. Задачу минимизации дивергенции Кульбака-Лейблера по первому аргументу решается тем же способом, что и был описан выше. Приведем формулы итерационного пересчета, только лишь для тех членов, которые добавились:

$$s_p^* = \arg \min_s \sum_i q_m(m_i = p) \|s - f_{coord}(i)\|^2, \forall p \in U$$

Остальные формулы необходимые для итерационного пересчета остались без изменения.

В данной секции была рассмотрена модель, которая способна обучать многоклассовую усеченную глубинную машину Больцмана не имея полной разметки данных. В следующих секциях мы рассмотрим вопрос автоматического получения семян и результаты экспериментов.

## 5 Модель детектора для автоматического получения семян

В данной секции рассмотрим способ автоматического получения семян для модели формы с семенами.

В 2010 году Felzenszwalb в статье [9] предложил модель детектора, которая позволяет выделять части объекта. Детектор обучается на основе исходных изображений и ограничивающих прямоугольников, в которых находится интересующий объект. После обучения детектор, который умеет выделять ограничивающий прямоугольник со всем объектом, а так же с его частями. Детектор запоминает расположение частей и их вид на основе часто встречающихся кусков изображений. На рисунке 5 представлена наглядная иллюстрация обученной модели детектора на данных с лошадьми. После применения детектора к изображениям, на выходе мы получаем ограничивающие прямоугольники, среди которых выберем вручную наиболее интересные с точки зрения семантического смысла, и за семена обозначим центры соответствующих прямоугольников.

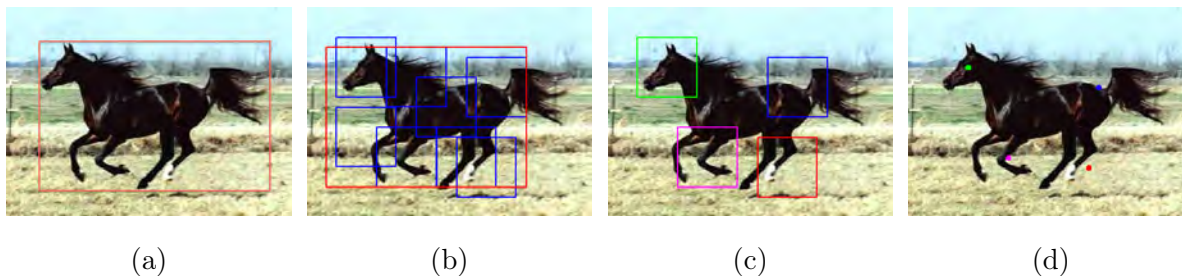


Рис. 5: (a) — исходное изображение с лошадью и ограничивающий прямоугольник для нее, (b) — результат применения детектора. (c) — отобранные ограничивающие прямоугольники, (d) — исходное изображение лошади с полученными семенами.

Заметим, что вместе с моделью детектора наша модель с семенами дает возможность обучать многоклассовую усеченную глубинную машину Больцмана на основе лишь исходных изображений и бинарной разметки.

## 6 Эксперименты с моделью с семенами

Мы рассмотрели новую модель, которой на вход на этапе обучения подается не многоклассовая разметка, а лишь бинарная маска и положение семян, отвечающих за окрестность соответствующей части. После же обучения эту модель можно воспринимать как многоклассовую усеченную глубинную модель Больцмана. Теперь необходимо сравнить ее с бинарное усеченной глубинной моделью Больцмана и с многоклассовой. Для количественного анализа будем использовать две метрики, подробное описание которых представлено ниже. Помимо количественных экспериментов также рассмотрим качественные эксперименты, то есть насколько хороший результат с точки зрения человека.

Эксперименты будем проводить на примере данных Weizmann horse dataset [8]. В данных имеются исходные изображения с лошадьми и бинарная маска с формами. Всего доступно 328 лошадей. Для обучения модели формы центрируем лошадей и отмасштабируем до размера 32 на 32 пикселя. Всех лошадей разобьем на две части: тестовые выборка (50 лошадей) и обучающая выборка (278 лошадей).

Для обучения обычной многоклассовой модели нам пришлось разметить всех лошадей. Получение ручной разметки требует больших временных затрат, и не всегда соответствует действительности в том смысле, что человек может с легкостью допустить ошибку и не заметить ее. Человеческий фактор является еще одним плюсом в пользу использования неполноразмеченных данных.

Для обучения модели формы с семенами, получим семена из детектора, описанного в предыдущей секции.

### 6.1 Метрики качества

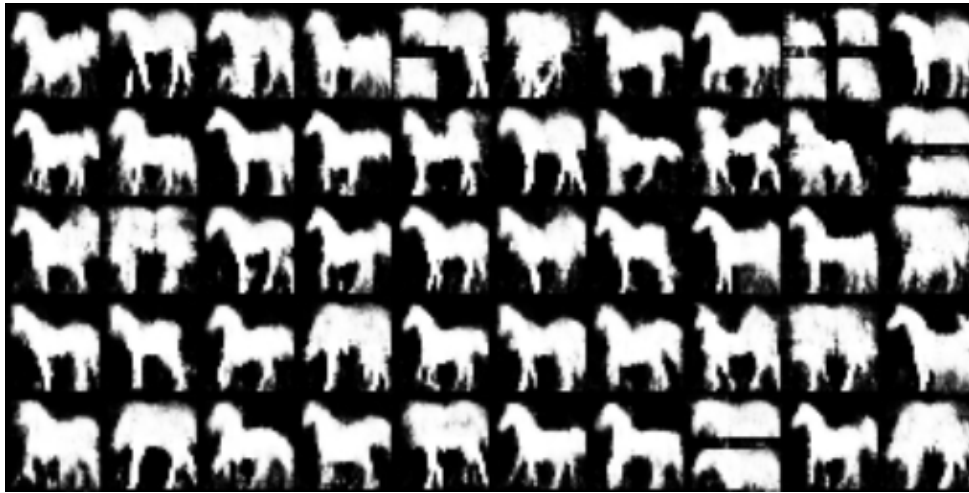
Как уже было сказано выше будем использовать две метрики качества для сравнения моделей. Рассмотрим первую метрику, которая оценивает качество восстановления части изображения по всему остальному изображению. Фиксируем некоторую бинарную маску объекта и семена для него, все пиксели этого изображения делим на два класса случайных переменных: наблюдаемые  $\mathbf{b}^o$  и скрытые  $\mathbf{b}^u$ , тогда будем считать вероятность  $p(\mathbf{b}^u | \mathbf{b}^o)$  для бинарной усеченной глубинной машины Больцмана



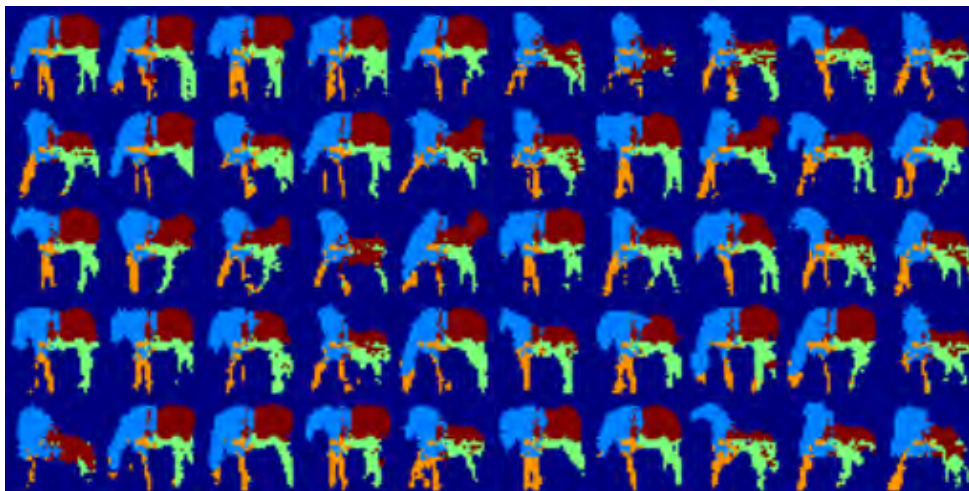
и  $p(\mathbf{b}^u | \mathbf{b}^o, \mathbf{s}^o)$  для многоклассовых моделей, где  $\mathbf{s}^o$  — наблюдаемые семена, то есть семена попавшие в область с наблюдаемой частью. Посчитать эту вероятность в явном виде не удастся поэтому будем приближенно оценивать следующим образом:

$$p(\mathbf{b}^u | \mathbf{b}^o) \approx \sum_{r=1}^R \frac{p(\mathbf{b}^u | \mathbf{h}_b^{1,r})}{R}, \quad p(\mathbf{b}^u | \mathbf{b}^o, \mathbf{s}^o) \approx \sum_{r=1}^R \frac{p(\mathbf{b}^u | \mathbf{h}_m^{1,r})}{R}$$

В свою очередь, набор векторов  $\mathbf{h}_b^{1,r}$  для  $r = \overline{1, R}$  соответствует набору векторов  $\mathbf{h}^1$  на последних  $R$  итерациях при выполнении процедуры генерации выборки для фиксированной части  $\mathbf{b}^o$ . Аналогично делаем и для  $\mathbf{h}_m^{1,r}$ ,  $r = \overline{1, R}$  с учетом того, что информация о многоклассовой разметке у нас хранится как в семенах, так и в модели.



(a)



(b)

Рис. 6: Применение процедуры генерации из семян: (a) — для бинарной модели, (b) — для модели с семенами.

Данную метрику будем измерять на всех тестовых объектах. Для однородности разделим наше изображение на 9 частей, сеткой 3 на 3, предварительно выкинув все граничные пиксели (тогда изображение станет размером 30 на 30 пикселей, что и позволяет разделить изображение сеткой 3 на 3, где каждый из сегментов имеет размер 10 на 10). Поочередно будем обозначать каждый из сегментов сетки как ненаблюдаемые переменные, а остальные как наблюдаемые. В результате измерять будем итоговый средний логарифм данной вероятности, где усреднение ведется одновременно по всем тестовым объектам и по всем сегментам.

Вторую метрику мы обозначим как качество восстановления изображения по сегментам. Для этого зафиксируем семена и при помощи процедуру вариационного вывода будем искать оптимальные  $q^1(\mathbf{h}^1)$ ,  $q^2(\mathbf{h}^2)$  после этого сгенерируем новые  $\mathbf{m}$  для многоклассовой модели или  $\mathbf{b}$  для бинарной модели. После многократного повторения данной процедуры получится некоторое бинарное изображение или многоклассовое изображение в зависимости от модели, итоговое изображение будем называть генерацией из семян. Метрика измеряет расстояние хемминга между получившимся бинарным изображением и тестовым изображением, для которого были взяты семена.

## 6.2 Результаты экспериментов

Эксперименты проводились для различных структур моделей: менялось число скрытых переменных на обоих слоях. Параметры обучения: величина регуляризации, размер выборки, для которой делать стохастический градиентный шаг, величина градиентного шага были предварительно оценены качественно и полученные значения согласуются с рекомендациями, предложенными в статье [3]. Так же параметр  $R$  из первой метрики был выбран из соображений, что с ростом  $R$  значение метрики должно стабилизироваться, что и было установлено на практике и значение параметра было зафиксировано 30.

На рисунках 7, 8 представлены результаты количественных экспериментов. Из них можно сделать вывод, что модель с семенами обладает качеством лучшим чем, бинарная модель и сопоставима по качеству с многоклассовой моделью. Так же посмотрим на изображения, получаемые в результате применения генерации из семян,

на рисунке 6. Качественно видно, что модель с семенами лучше справляется с задачей, в том смысле, что внешний контур и очертания объекта больше похожи на лошадей, данные изображения были получены на моделях с 1000 скрытых переменных на первом слое, и 100 скрытых переменных на втором слое.

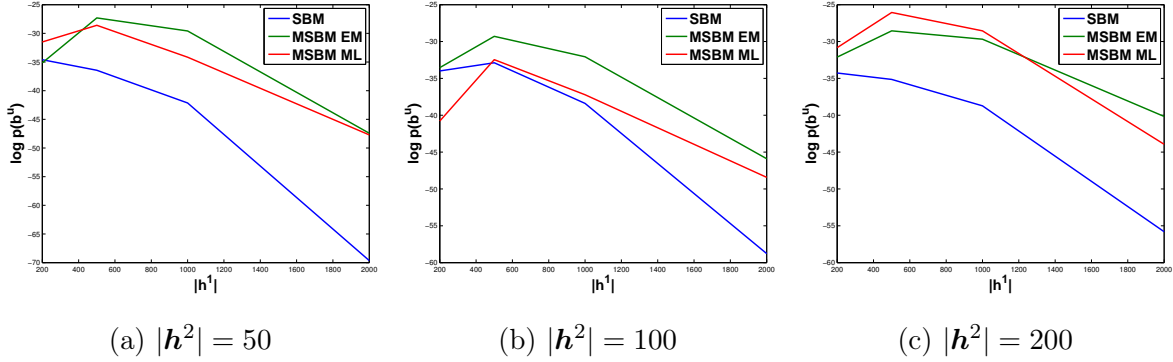


Рис. 7: Графики сравнения качества для различных структур и моделей по метрике, оценивающей качество восстановления части изображения по всему остальному. Большие значения лучше. Модель с семенами обозначена как «MSBM EM», «MSBM ML» — это многоклассовая усеченная глубинная машина Больцмана, обученная на полной разметке и «SBM» — это соответственно бинарная усеченная глубинная машина Больцмана.

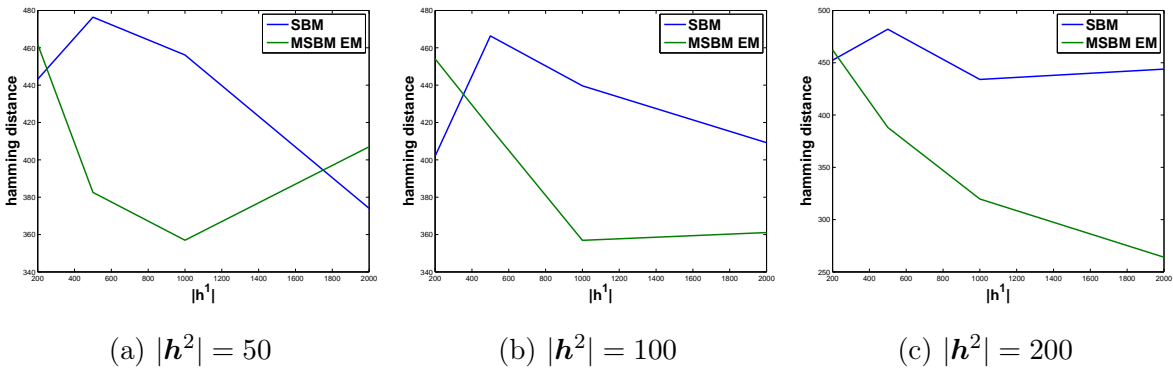


Рис. 8: Графики сравнения качества для различных структур и моделей по метрике, оценивающей качество восстановления изображения только по семенам. Меньшие значения лучше. Модель с семенами обозначена как «MSBM EM» и «SBM» — это соответственно бинарная усеченная глубинная машина Больцмана.

### 6.3 Выводы

В данной работе мы продемонстрировали, что наша модель формы с семенами, обученная на бинарных масках объекта и семенах частей работает лучше, чем обычная бинарная модель формы и сопоставима по качеству с многоклассовой моделью формы. Сравнение с бинарной моделью формы проводилось на основе двух метрик. Одна метрика имеет смысл качества восстановления формы на основе лишь информации о семенах. Другая метрика имеет смысл качества восстановления части изображения по всему остальному. Для сравнения модели с семенами с обычной многоклассовой усеченной глубинной машиной Больцмана использовалась только вторая метрика, по причине того, что иногда семена полученные из детекторов плохо сопоставимы с ручной разметкой. Для получения адекватных результатов сравнивались модели с разной архитектурой. На всех архитектурах результат был одним и тем же. Так же были представлены качественные результаты, по которым тоже видно преимущество нашей модели, хотя в абсолютном смысле оно не идеально.

## 7 Заключение

В данной работе была разработана новая модель, которая позволяет обучать многоклассовую усеченную глубинную машину Больцмана на основе данных о бинарной маске объектов и семенах, описывающих части объекта. Предложенная модель, а также модели машин Больцмана были реализованы. В разделе экспериментов показано, что данная модель превосходит или сопоставима по приведенным метрикам с имеющимися моделями. Актуальность данной работы в том, что в задаче семантической сегментации есть необходимость задавать ограничения, и примером такого ограничения может являться форма. Одним из способов задать форму на сегодняшний день, это использовать многоклассовую усеченную глубинную машину Больцмана, на этапе обучения которой требуются полноразмеченные данные. В работе же делается попытка отказаться от полной разметки, а использовать некоторую легкодоступную информацию: семена. Дальнейшее развитие данной работы связано с отказом от полной бинарной маски объекта, и в использовании только информации об ограничивающих прямоугольниках совместно с некоторым сегментатором или детектором.

## Список литературы

- [1] Andrew Kae , Kihyuk Sohn , Honglak Lee , Erik Learned-Miller. Augmenting CRFs with Boltzmann Machine Shape Priors for Image Labeling, 2013.
- [2] Ackley, D., Hinton, G., and Sejnowski, T. (1985). A Learning Algorithm for Boltzmann Machines. *Cognitive Science*, 9(1):147-169.
- [3] Geoffrey Hinton, Department of Computer Science, University of Toronto. A Practical Guide to Training Restricted Boltzmann Machines, 2010.
- [4] Ruslan Salakhutdinov, Geoffrey Hinton. An Efficient Learning Procedure for Deep Boltzmann Machine.
- [5] Radford M. Neal, Annealed Importance Sampling, 1998.
- [6] S. M. Ali Eslami, K. I. Williams, NIPS 2012. A Generative Model for Parts-based Object Segmentation.
- [7] Bishop, Christopher M., *Pattern Recognition and Machine Learning (Information Science and Statistics)*, 2006.
- [8] Вайзмановский набор изображений с лошадьми <http://www.msri.org/people/members/eranb/>.
- [9] P. F. Felzenszwalb et al., Part-Based Object Detector, 2010.
- [10] Eslami et al, NIPS 2012, A Generative Model for Parts-based Object Segmentation.