

Московский государственный университет имени М. В. Ломоносова



Факультет Вычислительной Математики и Кибернетики
Кафедра Математических Методов Прогнозирования

ДИПЛОМ

**Локально-глобальные вероятностные модели на основе
объединения марковских случайных полей и больцмановских
моделей формы для сегментации изображений**

Выполнил:

студентка 5 курса 517 группы

Лобачева Екатерина Максимовна

Научный руководитель:

к.ф.-м.н., доцент

Ветров Дмитрий Петрович

Москва, 2014

Содержание

1 Введение	2
2 Постановка задачи	2
3 Обзор литературы	3
3.1 ShapeVM	4
4 Объединение MRF и ShapeVM	5
4.1 Модель	6
4.2 Обучение	7
4.3 Задача вывода	8
4.4 Разрешение изображения	10
4.5 Выбор λ	11
5 Детектирование и сегментация	12
6 Эксперименты	14
7 Заключение	18
Список литературы	20

1 Введение

В данной работе рассматривается задача сегментации определенного объекта (лошади, автомобиля и т.д.) на изображении при условии, что объект на нем расположен в центре и занимает значительную его часть. Для ее решения предлагается модель сегментации, включающая в себя классическую парно-сепарабельную функцию энергии и потенциалы высокого порядка, задающие форму объекта. В качестве потенциалов высокого порядка используется одна из моделей глубинного обучения (deep learning) — больцмановская модель формы (Shape Boltzmann Machine). Предлагаемая модель сегментации позволяет попиксельно сегментировать изображения высокого качества с использованием глубинной сети, настроенной на изображениях с низким разрешением. В работе приведены результаты экспериментов, которые показывают существенный прирост качества по сравнению с использованием классической схемы сегментации с унарными и парными потенциалами. Также предлагается схема объединения полученной модели сегментации с детектированием объектов на изображении, которая позволяет сегментировать изображения произвольного вида.

2 Постановка задачи

В данной работе рассматривается задача сегментации изображения на два класса (объект и фон) со следующими особенностями:

- изображение центрировано и масштабировано, то есть объект находится в центре и занимает значительную часть изображения,
- известен тип объекта, например, лошадь, мотоцикл и т.д.,
- возможно объект занимает положение определенного вида, например, лошадь стоит головой в правую сторону.

Далее будем называть эту задачу задачей сегментации изображений частного вида. За основу при ее решении берется модель случайного марковского поля (MRF) [1]. Обозна-

чим узлы марковского поля за V , а систему их соседства за \mathcal{E} . Тогда задачу сегментации изображения можно рассматривать как задачу минимизации парно-сепарабельной энергии следующего вида:

$$E_{MRF} = (1 - C) \sum_{i \in V} \phi_i(v_i) + C \sum_{(i,j) \in \mathcal{E}} \phi_{ij}(v_i, v_j), \quad (1)$$

где $v_i \in \{0, 1\}$, $\forall i \in V$ и соответствуют отдельным пикселям изображения, $\phi_i(v_i)$ — унарные потенциалы, $\phi_{ij}(v_i, v_j)$ — парные потенциалы, а $C \in [0, 1]$ — коэффициент, отвечающий за соотношения влияния парных и унарных потенциалов.

Далее к данной энергии добавляется часть, отвечающая за потенциалы высоких порядков, а именно за модель формы объекта. Энергия в общем виде записывается следующим образом:

$$E = (1 - \lambda)E_{MRF} + \lambda E_s, \quad (2)$$

где E_s — член, отвечающий за потенциалы высоких порядков, λ — параметр, отвечающий за силу влияния модели формы на сегментацию.

3 Обзор литературы

В работе [3] были предложены Deep Boltzmann Machine (DBM), которые успешно применяются в последние годы в задачах, связанных с моделированием плотности распределения сложного вида. В частности, с помощью DBM можно моделировать распределение на форму объектов в пространстве изображений, однако у такого подхода есть существенный недостаток: модели DBM имеют очень много настраиваемых параметров, поэтому для их обучения необходимы большие объемы данных. Для обучения модели формы на основе DBM необходимы размеченные данные, то есть бинарные сегментации изображений, но получение большой выборки отсегментированных изображений — трудоемкая задача. Вследствие данного недостатка в качестве модели формы авторами [4] была предложена модель Shape Boltzmann Machine (ShapeBM), представляющая собой модификацию модели DBM с двумя слоями скрытых переменных, в которой количество параметров сильно

уменьшено за счет априорных представлений о структуре задачи моделирования формы. Именно эта модель формы будет использоваться в данной работе.

3.1 ShapeBM

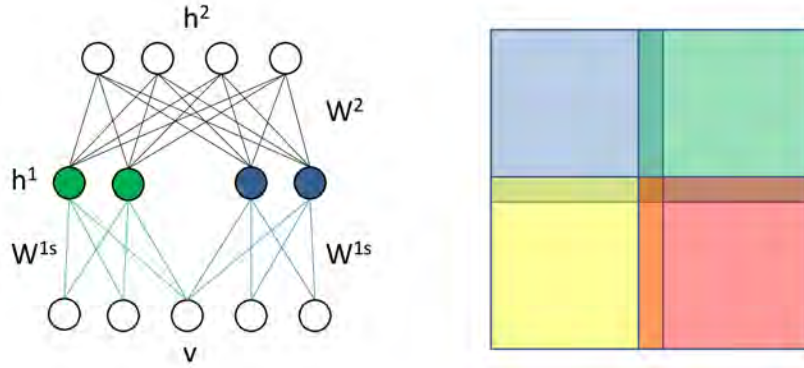


Рис. 1: Граф марковской сети для Shape Boltzmann Machine.

Остановимся подробнее на том, что представляет собой модель формы ShapeBM. ShapeBM — это марковская сеть с тремя слоями переменных, один из которых называется наблюдаемым, а два других — скрытыми. Ее граф показан на рисунке 1, слева. Энергия такой сети записывается следующим образом:

$$E(\mathbf{v}, \mathbf{h}^1, \mathbf{h}^2 | \Theta) = -\mathbf{v}^T W^1 \mathbf{h}^1 - (\mathbf{h}^1)^T W^2 \mathbf{h}^2 - \mathbf{v}^T \mathbf{b}^v - (\mathbf{h}^1)^T \mathbf{b}^{h^1} - (\mathbf{h}^2)^T \mathbf{b}^{h^2}, \quad (3)$$

где \mathbf{v} — наблюдаемые переменные, $\mathbf{h}^1, \mathbf{h}^2$ — скрытые переменные, $\Theta = \{W^1, W^2, b^v, b^{h^1}, b^{h^2}\}$ — настраиваемые параметры модели.

Наблюдаемые переменные сети \mathbf{v} соответствуют пикселям изображения, причем рассматриваются изображения одинакового размера $N \times N$. Переменные первого скрытого слоя \mathbf{h}^1 отвечают за локальные свойства формы (например, гладкость), поэтому:

1. не все \mathbf{h}^1 связаны со всеми \mathbf{v} ; наблюдаемые переменные \mathbf{v} делятся на четыре пересекающихся подмножества $\mathbf{v}^i, i \in \{1, 2, 3, 4\}$, которые соответствуют четвертям изображения с заступом в b пикселей (см. рис. 1, справа), а скрытые переменные

\mathbf{h}^1 делятся на четыре непересекающихся подмножества $\mathbf{h}^{1i}, i \in \{1, 2, 3, 4\}$; соответственно все переменные из \mathbf{h}^{1i} соединяется со всеми переменными из \mathbf{v}^i ;

2. веса между всеми парами множеств \mathbf{h}^{1i} и $\mathbf{v}^i, i \in \{1, 2, 3, 4\}$ считаются одинаковыми и обозначаются $W^{1s}, .$

Таким образом учитывается, что наиболее сильные связи имеются между близко расположенными пикселями, а также, что локальные свойства формы одинаковы для любой части изображения. Переменные второго скрытого слоя \mathbf{h}^2 отвечают за глобальные свойства формы (например, класс объекта, если мы обучаем модель формы на многоклассовой выборке, или положение объекта: лошадь в галопе или стоит), поэтому они связаны со всеми переменными первого скрытого слоя, как и в обычной DBM.

Обучение модели ShapeBM аналогично обучению модели DBM со следующей модификацией: граф на наблюдаемых переменных и переменных первого скрытого слоя представляет собой не одну большую Restricted Boltzmann Machine (RBM) [3], а 4 раза повторенную RBM меньшего размера. Соответственно, эта RBM меньшего размера на этапе предобучения обучается на совокупной выборке всех четвертей изображений обучающей выборки, а на этапе точной настройки учитывается равенство весов между группами \mathbf{h}^{1i} и $\mathbf{v}^i, i \in \{1, 2, 3, 4\}$. Полный алгоритм обучения ShapeBM представлен в [4].

Для определения приближенного распределения на скрытые переменные в ShapeBM при известных значениях наблюдаемых можно использовать вариационный вывод [5]. Также с помощью вариационного вывода можно подсчитать оценку на ненормированный логарифм неполного правдоподобия набора наблюдаемых переменных $p(\mathbf{v}|\Theta)$.

4 Объединение MRF и ShapeBM

В данной работе предлагается модель сегментации изображения частного вида, основанная на объединении классической модели сегментации путем минимизации парно-сепарабельной энергии MRF и модели формы объекта, задаваемой с помощью ShapeBM.

При сегментации изображений с помощью классической схемы возможно учесть толь-

ко локальные ограничения, такие как гладкость границы, цвет объекта и т.д. Глобальные же ограничения, например форма объекта, не вписываются в такую схему, их нужно задавать дополнительно путем введения некоторых потенциалов высокого порядка. В данной работе ограничения на форму объекта вводятся с помощью модели ShapeBM. ShapeBM задает вероятностное распределение в пространстве форм объектов, параметры которого можно настраивать по обучающей выборке.

4.1 Модель

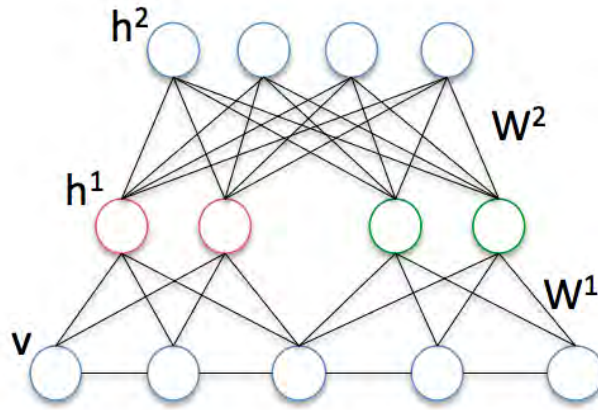


Рис. 2: Граф марковской сети для модели сегментации.

Перейдем непосредственно к рассмотрению модели сегментации. Граф для такой модели показан на рисунке 2. Он представляет собой граф для ShapeBM, в котором переменные на наблюдаемом слое соединены между собой в сетку, как в MRF. Энергия такой сети записывается следующим образом:

$$E(\mathbf{v}, \mathbf{h}^1, \mathbf{h}^2, \theta|I, \Theta) = (1 - \lambda)E_{MRF}(\mathbf{v}, \theta|I) + \lambda E_{ShapeBM}(\mathbf{v}, \mathbf{h}^1, \mathbf{h}^2|\Theta), \quad (4)$$

$$E_{MRF}(\mathbf{v}, \theta|I) = (1 - C) \sum_{i \in V} \phi_i(v_i, \theta|I) + C \sum_{(i,j) \in \mathcal{E}} \phi_{ij}(v_i, v_j|I), \quad (5)$$

$$E(\mathbf{v}, \mathbf{h}^1, \mathbf{h}^2|\Theta) = -\mathbf{v}^T W^1 \mathbf{h}^1 - (\mathbf{h}^1)^T W^2 \mathbf{h}^2 - \mathbf{v}^T \mathbf{b}^v - (\mathbf{h}^1)^T \mathbf{b}^{h^1} - (\mathbf{h}^2)^T \mathbf{b}^{h^2}, \quad (6)$$

где θ — параметры модели цвета, $\Theta = \{W^1, W^2, b^v, b^{h^1}, b^{h^2}\}$ — параметры модели формы,

I — цвета пикселей изображения, а на веса W^1 накладываются ограничения, соответствующие строению ShapeVM: обнуление и приравнивание определенных компонент. Под моделью цвета здесь понимается некоторое распределение вероятностей принадлежности каждого пикселя к объекту или к фону на основании его цвета. Неполное правдоподобие для наблюдаемых переменных в такой модели записывается следующим образом:

$$p(\mathbf{v}|I, \Theta) = \frac{1}{Z(I, \Theta)} \sum_{\mathbf{h}^1, \mathbf{h}^2, \theta} \exp(-E(\mathbf{v}, \mathbf{h}^1, \mathbf{h}^2, \theta|I, \Theta)). \quad (7)$$

Потенциалы MRF. В качестве модели цвета в данной работе используются две смеси нормальных распределений в пространстве цветов: одна для объекта и одна для фона. Число компонент в этих смесях фиксировано, а веса компонент и параметры каждой из них являются параметрами модели цвета. Унарные потенциалы выглядят следующим образом:

$$\phi_i(v_i, \theta|I) = -[v_i = 1] \log(I_i|\theta^1) - [v_i = 0] \log(I_i|\theta^0), \quad (8)$$

где θ^k — смесь для k -ого класса, I_i — цвет i -ого пикселя. Парные потенциалы выглядят следующим образом:

$$\phi_{ij}(v_i, v_j|I) = [v_i \neq v_j] \exp(-\|I_i - I_j\|_{RGB}^2). \quad (9)$$

4.2 Обучение

Обучение описанной модели сегментации распадается на два независимых этапа:

1. максимизация логарифма неполного правдоподобия обучающей выборки изображений $V = \{\mathbf{v}_i\}_{i=1}^N$ по параметрам модели формы:

$$\log p(V|I, \Theta) = \log \prod_{n=1}^N p(\mathbf{v}_n|I, \Theta) = - \sum_{n=1}^N E(\mathbf{v}_n|I, \Theta) \rightarrow \max_{\Theta}; \quad (10)$$

аналогично обучению ShapeVM;

2. обучение начального приближения на параметры модели цвета θ_0 по обучающей выборке изображений; аналогично случаю сегментации без модели формы.

На обоих этапах может использоваться одна и та же обучающая выборка из отсегментированных изображений.

4.3 Задача вывода

Сегментация изображения с помощью уже обученной модели сегментации состоит в нахождении состояния наблюдаемых переменных \mathbf{v} , обладающего минимальной энергией при фиксированных Θ , C , λ . Так как в модели сегментации присутствуют скрытые переменные, то для решения данной задачи можно воспользоваться EM-алгоритмом [6]:

- E-шаг:

Для пересчета распределения на значения скрытых переменных $\mathbf{h}^1, \mathbf{h}^2$ при фиксированных значениях наблюдаемых \mathbf{v} , а также параметрах модели цвета θ , применяется вариационное приближение в полностью факторизованном семействе:

$$q(\mathbf{h}^1, \mathbf{h}^2) = \prod_j q_j(h_j^1) \prod_k q_k(h_k^2), \quad (11)$$

$$KL(q(\mathbf{h}^1, \mathbf{h}^2) \| p(\mathbf{h}^1, \mathbf{h}^2 | \mathbf{v}^*, \Theta, \theta^*)) \rightarrow \min_{q(\mathbf{h}^1, \mathbf{h}^2)}, \quad (12)$$

где KL — дивергенция Кульбака-Лейблера, \mathbf{v}^*, θ^* — текущие значения наблюдаемых переменных и параметров модели цвета соответственно.

В итоге для пересчета распределения на значения скрытых переменных применяется итерационный процесс, на каждой итерации которого применяются следующие формулы:

$$q_j(h_j^1) = \frac{1}{1 + \exp(-\sum_i W_{ij}^1 v_i - \sum_k W_{jk}^2 h_k^2)}, \quad (13)$$

$$q_k(h_k^2) = \frac{1}{1 + \exp(-\sum_j W_{jk}^2 h_j^1)}. \quad (14)$$

- M-шаг:

На данном шаге решается следующая задача оптимизации:

$$\mathbb{E}_{q(\mathbf{h}^1, \mathbf{h}^2)} E(\mathbf{v}, \mathbf{h}^1, \mathbf{h}^2, \theta | I, \Theta) \rightarrow \min_{\mathbf{v}, \theta}. \quad (15)$$

Для ее решения используется метод покоординатного спуска, то есть функционал минимизируется итерационно, каждая итерация заключается в поочередной оптимизации по \mathbf{v} и по θ . В данной работе на M -шаге производится только одна итерация покоординатного спуска.

- M_1 -шаг:

Пересчет наблюдаемых переменных \mathbf{v} при фиксированных распределениях на значения скрытых $\mathbf{h}^1, \mathbf{h}^2$, а также параметрах модели цвета θ , производится путем минимизации следующего функционала:

$$\mathbb{E}_{q(\mathbf{h}^1, \mathbf{h}^2)} E(\mathbf{v}, \mathbf{h}^1, \mathbf{h}^2, \theta^* | I, \Theta) \rightarrow \min_{\mathbf{v}}, \quad (16)$$

где θ^* — текущее значение параметров модели цвета. Для решения задачи минимизации используется алгоритм нахождения минимального $S - T$ -разреза в графе.

- M_2 -шаг:

Пересчет параметров модели цвета θ при фиксированных распределениях на значения скрытых переменных $\mathbf{h}^1, \mathbf{h}^2$ и значениях наблюдаемых переменных \mathbf{v} производится путем минимизации следующего функционала:

$$\mathbb{E}_{q(\mathbf{h}^1, \mathbf{h}^2)} E(\mathbf{v}^*, \mathbf{h}^1, \mathbf{h}^2, \theta | I, \Theta) \rightarrow \min_{\theta}, \quad (17)$$

где \mathbf{v}^* — текущее значение наблюдаемых переменных.

В качестве начального приближения на наблюдаемые переменные \mathbf{v} для EM-алгоритма предлагается использовать результаты минимизации энергии без части, отвечающей за форму объекта, то есть функционала E_{MRF} по \mathbf{v} при фиксированной инициализации на параметры модели цвета θ_0 .

Итоговой сегментацией изображения является результирующее значение \mathbf{v}^* .

Объединение шагов M1 и M2. Шаги M1 и M2 в текущей схеме вывода можно объединить в один шаг оптимизации с помощью метода OneCut [7]. Этот метод предназначен

для оптимизации функционала упрощенного вида с унарными и парными потенциалами одновременно по наблюдаемым переменным и параметрам модели цвета за один $S - T$ -разрез в графе.

Упрощенный функционал в методе OneCut не содержит члена, отвечающего за баланс объема объекта и фона, поэтому при использовании этого метода необходима добавка к функционалу, задающая каким-то образом этот баланс. В данной работе в качестве такой добавки на E-шаге выступает $E_{ShapeBM}$.

4.4 Разрешение изображения

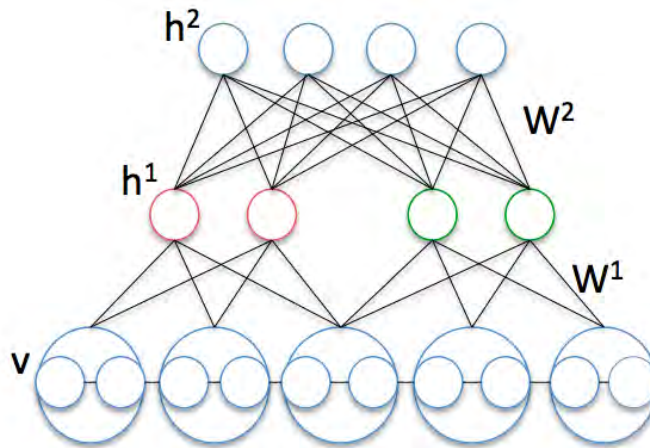


Рис. 3: Различные разрешения для MRF и ShapeBM.

Необходимый для обучения ShapeBM объем данных сильно зависит от количества переменных на наблюдаемом и скрытых слоях. В связи с этим, не представляется возможным обучать модель формы с большим разрешением на наблюдаемом слое, которое бы соответствовало размерам реальных изображений. Также стоит отметить, что разрешение наблюдаемого слоя в ShapeBM фиксировано, тогда как реальные изображения могут быть произвольной формы и размеров. Для решения этих проблем модель сегментации была модифицирована таким образом, что разрешение MRF и разрешение наблюдаемого слоя ShapeBM стали не обязательно совпадающими. Граф модифицированной модели

представлен на рисунке 3, а энергия записывается следующим образом:

$$E(\mathbf{v}, \mathbf{h}^1, \mathbf{h}^2, \theta | I, \Theta) = (1 - \lambda)E_{MRF}(\mathbf{v}, \theta | I) + \lambda \mathbb{E}_{\mathbf{v}^s} E_{ShapeBM}(\mathbf{v}^s, \mathbf{h}^1, \mathbf{h}^2 | \Theta), \quad (18)$$

где за \mathbf{v} обозначены переменные, соответствующие разрешению MRF, а за \mathbf{v}^s — соответствующие разрешению наблюдаемого слоя модели формы. При этом:

$$v_i^s \in \{v_j | v_j \in fields(i)\} \quad (19)$$

где функция $fields(i)$ возвращает множество переменных MRF, в соответствие которым в ShapeBM ставится переменная v_i^s . В данной работе функция $fields(i)$ задавалась равномерной сеткой необходимого разрешения на исходном изображении.

Данная модификация никак не влияет на процедуру обучения всей модели, однако влияет на процедуру вывода: на M-шаге изменится вид минимизируемого функционала. После E-шага часть минимизируемого функционала, соответствующую $(v^s)_k$ можно выписать в следующем виде:

$$\left(\mathbb{E}_{q(h^1, h^2)} \mathbb{E}_{\mathbf{v}^s} - \lambda ((h^1)^T (W^1)^T + b_v^T) \right)_k (v^s)_k. \quad (20)$$

Однако от этого представления легко перейти к следующему:

$$\sum_{i \in fields(k)} \left(\mathbb{E}_{q(h^1, h^2)} \frac{-\lambda}{|fields(k)|} ((h^1)^T (W^1)^T + b_v^T) \right)_k \cdot (v)_i. \quad (21)$$

В итоге на M-шаге оптимизируется функционал вида:

$$(1 - \lambda)E_{MRF}(\mathbf{v}, \theta | I) + \sum_k \sum_{i \in fields(k)} \left(\mathbb{E}_{q(h^1, h^2)} \frac{-\lambda}{|fields(k)|} ((h^1)^T (W^1)^T + b_v^T) \right)_k \cdot (v)_i, \quad (22)$$

который не зависит от переменных \mathbf{v}^s и может быть оптимизирован с помощью метода градиентного спуска по схеме, аналогичной описанной ранее.

4.5 Выбор λ

Параметр λ значительно влияет на результаты сегментации, оптимальное его значение сильно варьируется от изображения к изображению. В связи с этим, предлагается автоматически выбирать этот параметр по следующей логике:

1. подсчитать начальное приближение на наблюдаемые переменные \mathbf{v} для EM-алгоритма \mathbf{v}' ;
2. подсчитать оценку на ненормированный логарифм неполного правдоподобия $p(\mathbf{v}'|\Theta)$ для ShapeVM;
3. если эта оценка ниже среднего, значит \mathbf{v}' «не похожа» на объект заданного типа в терминах ShapeVM, а значит влияние модели формы нужно усилить, то есть взять большое λ , так как на изображении заведомо находится объект заданного типа;
4. если же эта оценка выше среднего, значит \mathbf{v}' «похожа» на объект заданного типа в терминах ShapeVM, а значит нет необходимости использовать большое λ .

На практике, значение средней оценки $p(\mathbf{v}'|\Theta)$, а также зависимости оптимального λ от оценки $p(\mathbf{v}'|\Theta)$ для конкретного изображения нужно обучать на дополнительной обучающей выборке.

5 Детектирование и сегментация

С помощью описанной выше модели сегментации изображений частного вида можно обрабатывать только узкий класс изображений. Опишем возможную схему применения предлагаемой модели для любых изображений, на которых объект может находиться в произвольном месте и позиции. Эта схема состоит из трех этапов:

1. детекция объекта на изображении,
2. преобразование части изображения, в которой находится объект, к стандартному для конкретной модели формы виду (например, поворот, растяжение и т.д.),
3. сегментация преобразованной части как центрированного изображения с помощью описанной выше модели сегментации.

В данной работе рассматривается детектор, основанный на выделении частей объекта, описанный в [8]. Он обладает следующими свойствами:

- для обучения необходима слабо размеченная выборка, а именно, на каждом изображении должен быть выделен ограничивающий прямоугольник для объекта;
- определим два понятия:

1. модель в терминах детектора — некоторый подкласс объектов внутри обучающей выборки,
2. части объекта — некоторые составляющие части объектов из обучающей выборки (например, голова лошади или круп);

при обучении детектора фиксируются число моделей и частей, однако типы моделей и частей обучаются автоматически,

- при применении детектора к изображению можно получить тип модели в терминах детектора, к которому относится объект на изображении, а также ограничивающие прямоугольники для всего объекта и для каждой его части.

За часть изображения, содержащую объект, берется ограничивающий прямоугольник для всего объекта. Для преобразования этой части к стандартному виду делается следующее:

1. детектор применяется к изображениям из обучающей выборки для ShapeVM, получают математические ожидания центров ограничивающих прямоугольников для всего объекта и его частей,
2. детектор применяется к сегментируемому изображению, также получают центры ограничивающих прямоугольников для всего объекта и его частей,
3. проводится аффинное преобразование части сегментируемого изображения, содержащей объект, к виду стандартных изображений для ShapeVM, которое переводит центры ограничивающих прямоугольников сегментируемого изображения в соответствующие математические ожидания их центров на обучающей выборке для ShapeVM.

Улучшение модели цвета. Объекты некоторых типов имеют специфические модели цвета. Например, лошади в общем могут быть разных цветов: черные, белые, и т.д., однако конкретная лошадь имеет определенный окрас. Ясно, что модель цвета, обученная на конкретной лошади, будет лучше описывать ее цвет, чем модель, обученная на выборке разных лошадей.

В связи с такими особенностями, имеет смысл обучать начальное приближение на параметры модели цвета не на некоторой обучающей выборке, а на конкретном изображении. Однако для этого необходима дополнительная информация о том, где на изображении находится объект, недостаточно просто знать, что объект центрирован, слишком сильна может быть вариация форм. Такую дополнительную информацию можно извлечь из результатов работы детектора.

Предлагается для каждой модели объекта в терминах детектора подсчитать среднюю сегментацию каждой части объекта по обучающей выборке. После, для подсчета начального приближения на параметры модели цвета для конкретного изображения провести детекцию частей объекта на этом изображении и обучить модель цвета на пикселях этого же изображения, сегментацию на которых задать равной обученной средней сегментации соответствующих частей.

6 Эксперименты

В качестве обучающей выборки взята выборка Вайзмана [9], содержащая 327 центрированных и масштабированных изображений лошадей, развернутых головой в левую сторону. В экспериментах используется обученная авторами [4] модель формы ShapeVM с разрешением наблюдаемого слоя 32 на 32 пикселя, 2000 узлами на первом скрытом слое и 100 — на втором и модель цвета, обученная самостоятельно на той же обучающей выборке.

Для проведения экспериментов была самостоятельно размечена тестовая выборка из 16 изображений с лошадьми размера 128 на 128 пикселей. Она представлена на рисунке 4

Обозначим за \mathbf{v} получаемую разметку, \mathbf{v}^* — истинную. Для оценки качества сегментации выбраны две меры качества:



Рис. 4: Тестовая выборка

1. Доля пикселей, отсегментированных верно (ассигасу):

$$Acc(\mathbf{v}, \mathbf{v}^*) = \frac{\sum_i [v_i = v_i^*]}{|\mathbf{v}|}, \quad (23)$$

где $|\mathbf{v}|$ — общее количество пикселей на изображении.

2. Коэффициент Жаккара или intersection-over-union:

$$IoU(\mathbf{v}, \mathbf{v}^*) = \frac{\sum_i [v_i = 1][v_i^* = 1]}{\sum_i [v_i = 1][v_i^* = 1] + \sum_i [v_i = 0][v_i^* = 1] + \sum_i [v_i = 1][v_i^* = 0]}. \quad (24)$$

Общие схемы. Было проведено сравнение 4 алгоритмов сегментации:

1. MRF с фиксированной цветовой моделью;
2. EM-алгоритм с использованием только шагов E, M_1 ;
3. адаптивный MRF: разрез графа применяется многократно, после каждого раза модель цвета пересчитывается аналогично шагу M_2 ;
4. Полный EM-алгоритм с использованием шагов E, M_1 , M_2 .

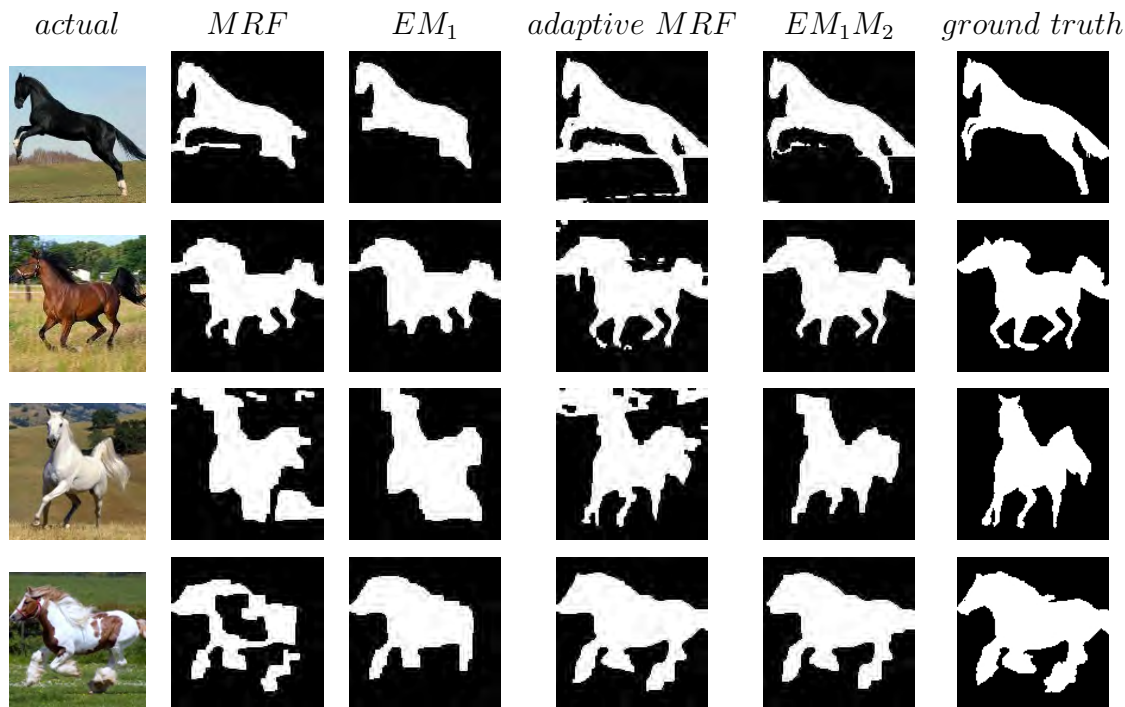


Рис. 5: Сегментация изображений из тестовой выборки

Мера	MRF	EM_1	adaptive MRF	EM_1M_2
Acc	0.8596	0.8723	0.9068	0.9265
IoU	0.6452	0.6883	0.7918	0.8131

Таблица 1: Качество сегментации.

Полученные результаты приведены в таблице 1 и на рисунке 5.

Параметры λ , C , общие для всех изображений, подбирались экспериментальным путем. Приведенные результаты соответствуют их лучшему значению.

Результаты экспериментов говорят о том, что добавление в модель сегментации модели формы и пересчет модели цвета в ходе итераций значительно улучшают результаты сегментации.

Сравнение с существующими методами. Результаты сравнения предлагаемой модели с другими моделями сегментации, использующими каким-либо образом форму объ-

екта, представлены в таблице 2. Заметим, что качество подсчитывалось на разных выборках изображений: для предлагаемой модели — на описанной выше тестовой выборке, а для других — на подмножествах выборки Вайзмана (см. подробнее в соответствующих статьях). Однако, и выборка Вайзмана, и описанная в данной работе тестовая выборка являются случайными выборками центрированных и масштабированных изображений лошадей, повернутых в левую сторону, поэтому с точностью до статистической погрешности, цифры, представленные в таблице 2, можно сравнивать.

Метод	Acc	IoU
EM_1M_2	0.927	0.813
Borenstein et al. [9]	0.936	—
LOCUS [10]	0.931	—
ClassCut [11]	0.862	—
FSA [12]	0.88	—
CHOPPs [13]	0.8867	0.699
MMBM [14]	0.907	0.758

Таблица 2: Сравнение с существующими методами.

Результаты показывают, что качество работы предложенной модели сегментации сравнимо с результатами существующих методов.

OneCut. Замена шагов M_1, M_2 в полном EM-алгоритме на один шаг оптимизации с помощью OneCut не приводит к значительному изменению качества сегментации, однако значительно ускоряет алгоритм: если сегментации одного тестового изображения с помощью полного EM-алгоритма без замены шагов проводится за время порядка 4 секунд, то с использованием OneCut — порядка 1 секунды.

Автоматический выбор λ . В ходе экспериментов было подтверждено предположение о том, что существует зависимость между оценкой на ненормированный логарифм неполного правдоподобия $p(\mathbf{v}'|\Theta)$ для ShapeBM (\mathbf{v}' — результат сегментации изображения без

использования модели формы) и оптимальным значением параметра λ : чем оценка выше, тем оптимальное значение λ ниже.

Модель цвета по результатам детекции. Приведем результаты экспериментов для одной модели объекта в терминах детектора. Обученные средние сегментации частей для этой модели показаны на рисунке 6.

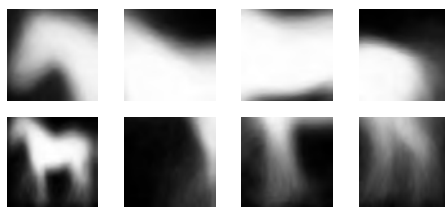


Рис. 6: Средние сегментации частей объекта.

Под выбранную модель объекта в терминах детектора подошло только 11 изображений из тестовой выборки, сравнение качества сегментации на них представлено в таблице 3.

Color model	MRF	adaptive MRF	EM_1M_2
general	0.3118	0.1542	0.0943
from detection	0.2237	0.0899	0.0743

Таблица 3: Ошибка на 11 изображениях тестовой выборки при обучении модели цвета с использованием результатов детектирования.

Результаты показывают, что использование результатов детектирования для обучения начального приближения на модель цвета значительно улучшает качество сегментации центрированных изображений.

7 Заключение

Обозначим итоги:

1. Построена теоретически обоснованная модель сегментации центрированных изображений любого разрешения, основанная на объединении MRF и ShapeVM и использующая адаптивный пересчет цветовых моделей.
2. Рассмотрены различные модификации полученной модели: применение OneCut, автоматический выбор параметра, отвечающего за баланс между MRF и ShapeVM.
3. Предложена схема применения полученной модели для сегментации обычных нецентрированных изображений с помощью детектора объектов, а также рассмотрен вариант улучшения модули сегментации центрированных изображений с использованием результатов детектирования.
4. Проведены эксперименты, подтверждающие преимущество построенной модели сегментации центрированных изображений перед MRF и MRF с пересчетом цветовой модели, а также подтверждающие необходимость использования процедуры пересчета модели цвета в данной модели. Также проведены эксперименты, показывающие результаты применения всех предложенных улучшений и модификаций данной модели.
5. Проведено сравнение полученной модели с другими существующими моделями сегментации изображений, использующими каким-либо образом форму объекта. Показано, что качество работы предложенной модели сегментации сравнимо с результатами существующих методов.

Список литературы

- [1] Kindermann, Ross; Snell, J. Laurie. Markov Random Fields and Their Applications // American Mathematical Society, ISBN 0-8218-5001-6, 1980.
- [2] Casella, George; George, Edward I. Explaining the Gibbs sampler // The American Statistician 46 (3): 167–174, 1992.
- [3] R.Salakhutdinov, G.Hinton. An Efficient Learning Procedure for Deep Boltzmann Machines // MIT Technical Report MIT-CSAIL-TR-2010-037, 2010.
- [4] S.M.Ali Eslami, Nicolas Heess, John Winn. The Shape Boltzmann Machine: a Strong Model of Object Shape // IEEE Conference on Computer Vision and Pattern Recognition, 2012.
- [5] Christopher M. Bishop. Pattern recognition and Machine Learning // Springer, 2012, 462 - 490.
- [6] Christopher M. Bishop. Pattern recognition and Machine Learning // Springer, 2012, 450 - 455.
- [7] Meng Tang, Lena Gorelick, Olga Veksler, Yuri Boykov. Grabcut in One Cut // International Conference on Computer Vision, 2013.
- [8] P.F.Felzenszwalb, R.B.Girshick, D.McAllester, D.Ramanan. Object Detection with Discriminatively Trained Part Based Models // IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010.
- [9] Eran Borenstein, Eitan Sharon, Shimon Ullman. Combining Top-Down and Bottom-Up Segmentation // IEEE Conference on Computer Vision and Pattern Recognition, 2004.
- [10] John Winn and Nebojsa Jojic. LOCUS: Learning object classes with unsupervised segmentation // International Conference on Computer Vision, 2005.

- [11] Bogdan Alexe, Thomas Deselaers, Vittorio Ferrari. ClassCut for unsupervised class segmentation // European Conference on Computer Vision, 2010.
- [12] S.M. Ali Eslami, Christopher K.I. Williams. Factored Shapes and Appearances for Parts-based Object Understanding // British Machine Vision Conference, 2011.
- [13] Y. Li, D. Tarlow, R. Zemel. Exploring compositional high order pattern potentials for structured output learning // IEEE Conference on Computer Vision and Pattern Recognition, 2013.
- [14] Jimei Yang, Simon Sáfár, Ming-Hsuan Yang. Max-Margin Boltzmann Machines for Object Segmentation // IEEE Conference on Computer Vision and Pattern Recognition, 2014.