

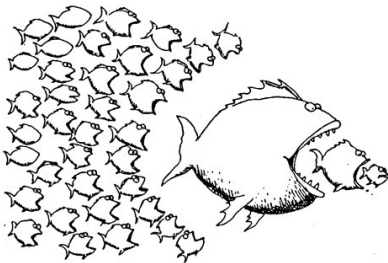
Функции похожести (similarity functions) гипотез в задаче оценки позы человека с использованием порождающих моделей

Дарья Вальтер
Научный руководитель А.С.Конущин

НИС Машинное обучение

10 октября 2016 г.

Порождающие модели vs. дискриминативные модели



Дискриминативные модели (top-down approach):

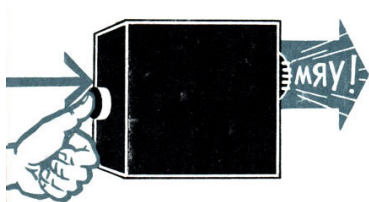
- поиск закономерностей в данных
- позволяют найти точечную оценку целевой функции по наблюдаемому изображению

Порождающие модели (bottom-up approach):

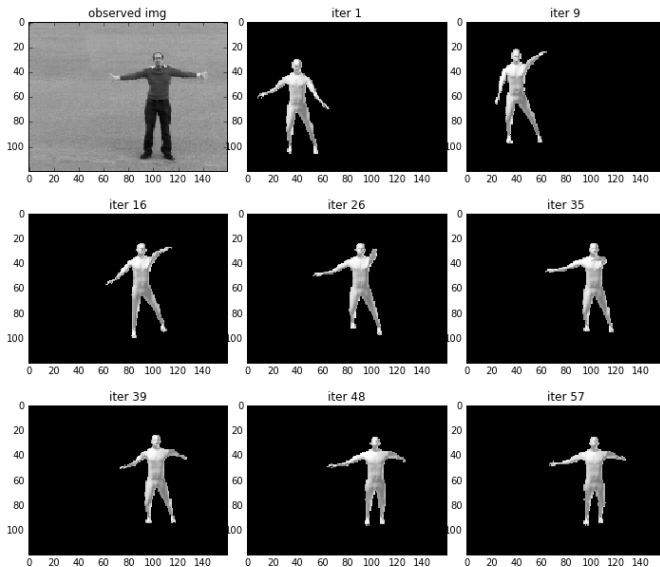
- моделирование гипотез
- позволяют оценить правдоподобие наблюдаемого изображения заданном параметрическом семействе сцен

Недостатки порождающих и дискриминативных моделей

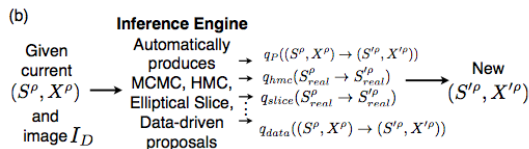
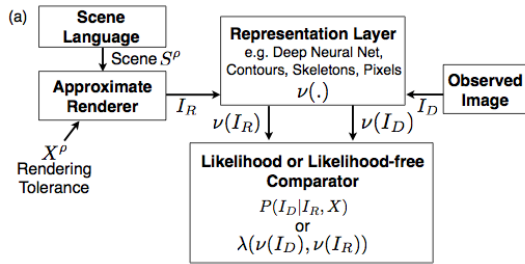
- Дискриминативные модели:
 - Потребность в выборках большого размера
 - Переобучение
 - Медленная скорость обучения
 - Сложность интерпретации
- Порождающие модели:
 - Моделирование каждой отдельной задачи
 - Медленная скорость сходимости



Пример сходимости порождающей модели



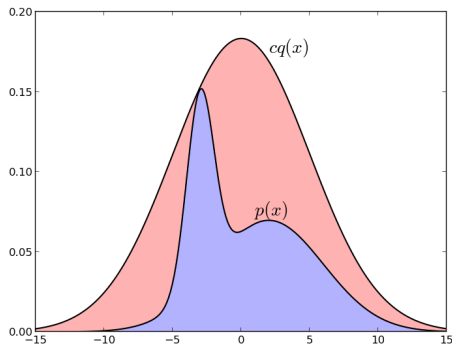
Компоненты порождающей модели



- Язык описания сцены
- Генератор гипотез
- Движок рендеринга
- Уровень признакового представления изображений
- Функция сравнения признаков описаний изображений

Простейший генератор гипотез

Сэмплирование с отклонением (rejection sampling)



Пусть задано распределение $q(x)$, которое приближает распределение $p(x)$

$\exists c : \forall x \quad cq(x) > p(x)$.

- 1 сэмплируется $x \sim q(x)$
- 2 берётся случайное число $u \sim U_{[0, cq(x)]}$
- 3 вычисляется $p(x)$
- 4 если $p(x) > u$, то x принимается, иначе x отклоняется.

Более эффективные методы: Алгоритм Метрополиса-Гастингса, сэмплирование по Гиббсу

Порождающие модели. Оценка правдоподобия

- Модель сходится к ответу, дающему максимум правдоподобия

Оценка правдоподобия (Log-likelihood comparator):

$$\mathcal{L} = \log p(I_D | I_R, \lambda) = \log (p(S^\rho) p_\epsilon(\lambda(\nu(I_D), \nu(I_R))))$$

$$p(S^\rho) = \prod_{i=1}^n p_i(s_i)$$

I_D - наблюдаемое изображение

I_R - изображение-гипотеза

$\lambda(\nu(I_D), \nu(I_R))$ - similarity function двух изображений

s_1, \dots, s_n - реализация параметров сцены

$P_\epsilon \sim \mathcal{N}(0, \sigma)$ - распределение отклонения от наблюдаемого изображения, σ предполагается близкой к нулю

- "Самописные" дескрипторы
 - Карты краев. Детектор Canny
 - Гистограммы ориентированных градиентов. Дескриптор SIFT
- С использованием методов МО
 - Нейросетевые признаки
 - Нейросетевые структуры для обучения similarity function

Методы сравнения изображений

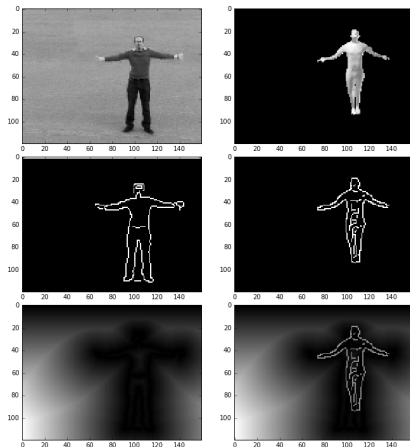
Детектор краев изображения Canny

Алгоритм Canny:

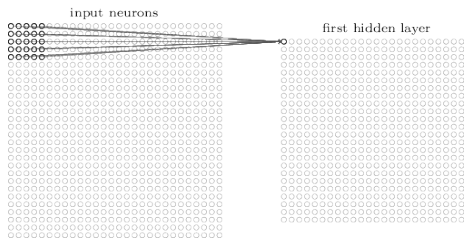
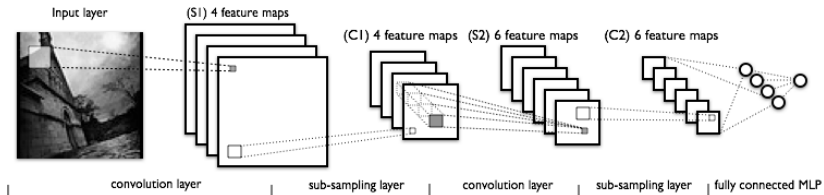
- 1 Убрать шум (фильтр Гаусса)
- 2 Рассчитать градиент изображения
- 3 Выделить области локальных максимумов

Для подсчета метрики:

- 1 К карте краев наблюдаемого изображения применить фильтр distance transform (карта расстояний)
- 2 Наложить карту краев гипотезы на полученное изображение

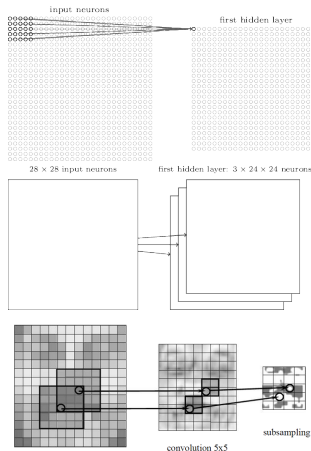


Сверточные нейросети



- Неполносвязная: легче обучать
- Локальное рецептивное поле (local receptive field)
- Разделенные веса (shared weights)
- Pooling

Сверточные нейросети



Разделенные веса:

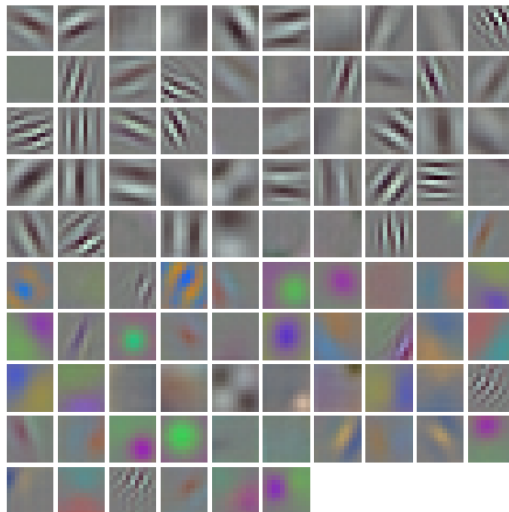
- Каждый скрытый нейрон имеет одинаковый набор весов и баейс
- Активация j, k -ого скрытого нейрона на $(i + 1)$ -ом слое :

$$a_{j,k}^{(i+1)} = \sigma(b + \sum_{l=0}^4 \sum_{m=0}^4 w_{l,m} a_{j+l,k+m}^{(i)})$$

- Для разных feature maps веса и баейсы разные

Pooling: инвариантность

Пример фильтров



Методы сравнения изображений

Нейросетевые признаки, l2 метрика

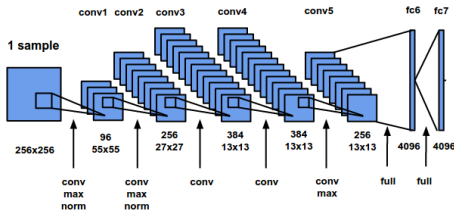
- Метрика: евклидово расстояние между нейросетевыми признаками

- с последнего сверточного слоя ($A \times B \times B$)

$$l_2(x, y) = \sqrt{\sum_{i=1}^A \sum_{j=1}^B \sum_{k=1}^B (x_{ijk} - y_{ijk})^2}$$

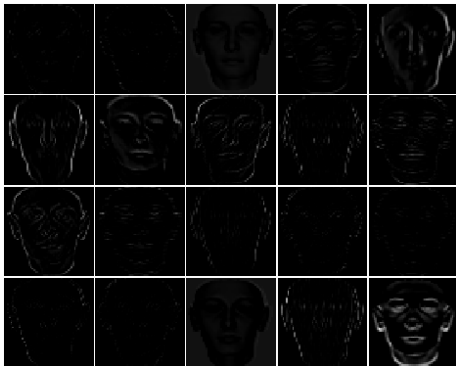
- с полносвязных слоев

$$l_2(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

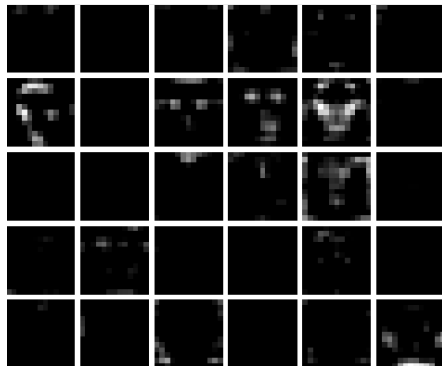


Архитектура сверточной нейросети AlexNet

Нейросетевые признаки (feature maps)



Признаки с 1 сверточного слоя

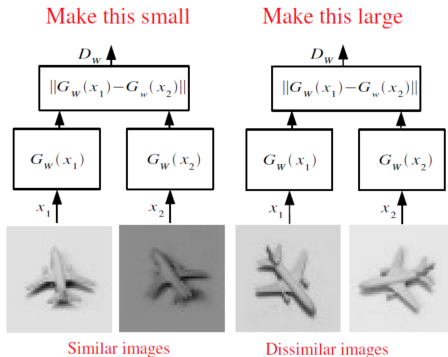
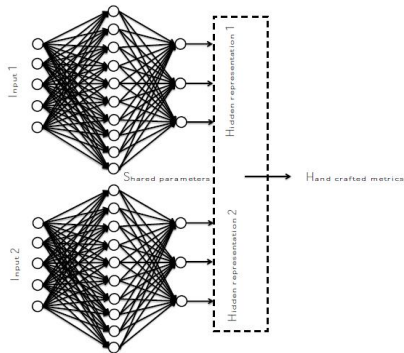


Признаки с 5 сверточного слоя

Методы сравнения изображений

Siamese нейросеть

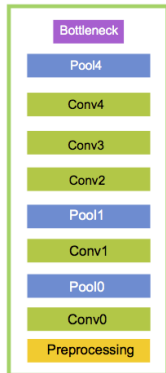
- Нейросетевая Siamese архитектура позволяет обучать similarity function



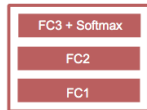
MatchNet архитектура

$$L = -\frac{1}{n} \sum_{i=1}^n (1 - y_i) \log(1 - \hat{y}_i) + y_i \log(\hat{y}_i) \rightarrow \min$$

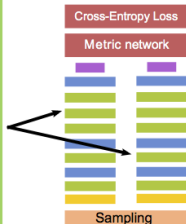
A: Feature network



B: Metric network



C: MatchNet in training



$$y_i = \begin{cases} 1 & \text{match} \\ 0 & \text{nomatch} \end{cases}$$

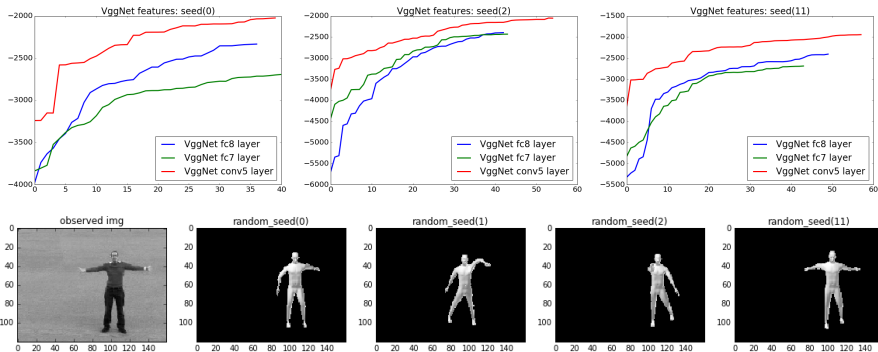
\hat{y}_i - приближение вероятности класса 1

$$\text{Softmax} : \hat{y}_i = \frac{e^{v_1(x_i)}}{e^{v_0(x_i)} + e^{v_1(x_i)}}$$

$v_1(x_i), v_0(x_i)$ - активации нейронов на FC3

Результаты: Нейросетевые признаки, l2 метрика

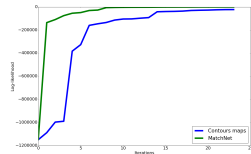
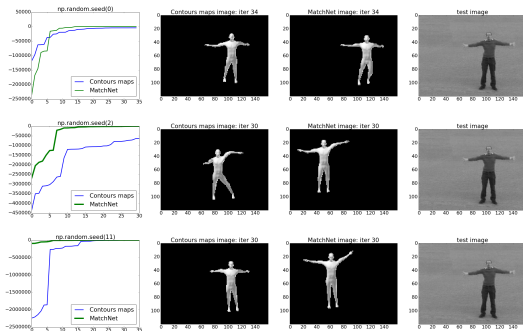
Графики значений Log-likelihood от итерации



- Наибольшие значения Log-likelihood для conv5
- Низкая скорость сходимости (50 итерация)
- Низкая точность






Результаты: Сравнение Canny и Siamese метрик

Графики значений Log-likelihood от итерации








- Оценка правдоподобия растет быстрее всего при использовании Siamese метрики

Библиография

-  A. Krizhevsky, I. Sutskever è G. E. Hinton, "Imagenet classification with deep convolutional neural networks", *Advances in neural information processing systems*, 2012, с. 1097-1105.
-  Tejas D. Kulkarni (MIT), Pushmeet Kohli (MSR Cambridge, UK), Joshua B. Tenenbaum (MIT), Vikash Mansinghka (MIT), "Picture: A Probabilistic Programming Language for Scene Perception"
-  John Canny. "A computational approach to edge detection". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-8(6):679–698, Nov. 1986.
-  Lowe D. G. "Object recognition from local scale-invariant features"// *Proc. Intl. Conference on Computer Vision*. – 1999. – P. 1150–1157.
-  Philipp Fischer, Alexey Dosovitskiy, Thomas Brox. "Descriptor Matching with Convolutional Neural Networks: a Comparison to SIFT". arXiv pre-print arXiv:1405.5769 (2014)

Библиография

-  Dosovitskiy, A., Springenberg, J.T., Brox, T.: "Unsupervised feature learning by augmenting single images pre-print, arXiv:1312.5242v3 (2014)
-  Sergey Zagoruyko, Nikos Komodakis. "Learning to Compare Image Patches via Convolutional Neural Networks arXiv:1504.03641 [cs.CV] (2015)
-  Han Xufeng, Leung Thomas, Jia Yangqing, Sukthankar Rahul, Berg Alexander. C., "MatchNet: Unifying Feature and Metric Learning for Patch-Based Matching *CVPR*, 2015
-  Karen Simonyan, Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition arXiv:1409.1556 [cs.CV], 2014
-  Y. Jia. Caffe: An open source convolutional architecture for fast feature embedding. <http://caffe.berkeleyvision.org/>, 2013.

Спасибо за внимание!