

ИСПОЛЬЗОВАНИЕ СВЁРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧЕ РАСПОЗНАВАНИЯ ЛИЦ ДЛЯ ФОРМИРОВАНИЯ ВЕКТОРА ПРИЗНАКОВ

Савицкий Ю.В.

Научный руководитель: В.Г. Спицын
Томский политехнический университет
mr-l-ik@yandex.ru

Введение

Задача распознавания лиц является одной из самых востребованных практических задач в сфере интеллектуальной обработки изображений.

Актуальность данной работы обусловлена потребностью в разработке алгоритма распознавания лиц в реальных условиях, обеспечивающего высокую степень надёжности результата распознавания. Реализация алгоритма, способного решать поставленную задачу распознавания лиц в режиме реального времени, является неотъемлемым требованием организаций, работающих в сфере обеспечения безопасности.

Большинство алгоритмов распознавания лиц могут быть разделены на следующие этапы:

1. Поиск лица.
2. Нормализация найденного изображения.
3. Извлечение признаков.
4. Классификация.

Данная работа посвящена исследованию применения свёрточных нейронных сетей для извлечения вектора признаков из изображения лица.

Свёрточные нейронные сети

Свёрточная нейронная сеть – особый вид архитектуры искусственных нейронных сетей прямого распространения, разработанный для обработки изображений.

Основным отличием свёрточных нейронных сетей от классических многослойных сетей прямого распространения является использование операции свёртки и операции подвыборки. На рисунке 1 приведена типичная схема архитектуры свёрточной нейронной сети.

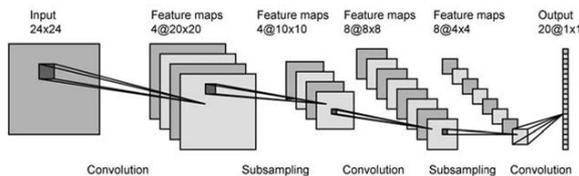


Рис. 1. Схема архитектуры свёрточной нейронной сети

Исходя из рисунка 1, входное изображение представляется как матрица чисел. Операция свёртки представляет собой проход ядром свёртки по матрице входных данных для получения карты признаков. Ядром свёртки является матрица чисел, которая представляет собой фильтр, а картой признаков является матрица, получаемая в

процессе применения ядра свёртки к входным данным. Операция подвыборки заключается в уменьшении размерности, в большинстве случаев путём выбора максимума по нескольким соседним ячейкам карты признаков.

Обычно заключительным слоем свёрточной нейронной сети является полносвязный слой, выход которого будет являться вектором признаков заданного изображения.

Реализация свёрточной сети

На текущий момент существует множество инструментов для работы со свёрточными нейронными сетями, к ним относятся: Caffe, dlib, Theano, Torch7, TensorFlow. В ходе проведённого аналитического обзора было принято решение на текущем этапе использовать библиотеку Caffe.

Библиотека

Caffe разрабатывается командой Berkeley Vision and Learning Center и предназначена для научных разработок в сфере компьютерного зрения [1]. Библиотека написана на языке C++ и имеет API для языков C++, Python, а также Matlab. Задание архитектуры сети производится в текстовом файле на специализированном языке, похожем на JavaScript.

Для обучения сети была выбрана база лиц Casia-WebFace [2], изображения которой были приведены к размеру 160x160. Данный набор содержит порядка 500 000 изображений 10 575 персон.

Архитектура сети и параметры метода обучения были подобраны исходя из рекомендаций, даваемых в литературе, и наблюдений за процессом обучения. Итоговая сеть показана в таблице 1.

Таблица 1. Архитектура нейронной сети

Входное изображение 160x160
Conv1, размер ядра 3x3, кол-во ядер 16, шаг 1 Pool1, размер ядра подвыборки 3x3, шаг 1 Tanh1
Conv2, размер ядра 3x3, кол-во ядер 32, шаг 1 Pool2, размер ядра подвыборки 3x3, шаг 1 Tanh2
Conv3, размер ядра 3x3, кол-во ядер 64, шаг 1 Pool3, размер ядра подвыборки 3x3, шаг 1 Tanh3
Conv4, размер ядра 3x3, кол-во ядер 128, шаг 1 Pool4, размер ядра подвыборки 3x3, шаг 1 Tanh4
FC, размер 2000 Sigm
FC-classification, 10 575

В таблице 1, схематично представляющей нейронную сеть, Conv# - свёрточный слой, Pool# - подвыборка, Tanh# - применение гиперболического тангенса как функции активации, FC-полносвязный слой, Sigm- применение сигмоидной функции, FC-classification-полносвязный слой, используемый для обучения. Итоговый вектор признаков получается из слоя Sigm, его размерность равняется 2000.

В качестве алгоритма обучения был выбран метод Adam с рекомендуемыми параметрами из [3]. Параметр $learningrate = 0,0001$, $weightdecay = 0,0001$, длительность обучения 300 000 итераций.

Тестирование

Тестирование обученной сети было решено проводить на базе LabeledFacesintheWild (LFW) [4]. Данная база содержит 13233 изображений 5749 персон. Данный набор формировался специально для проведения тестирования алгоритмов распознавания лиц. LFW содержит специальный тестовый список из 6000 пар изображений, разделённый на 10 частей, по 300 совпадающих пар и по 300 несовпадающих для перекрёстной проверки.

Для тестирования изображения были приведены к размеру 160x160. В качестве метрики было использовано расстояние Евклида. Между получаемыми с помощью нейронной сети векторами находилось расстояние, которое сравнивалось с порогом допущения совпадения пары векторов. В случае если расстояние было меньше порога допущения, пара считалась совпадающей, в противном случае - несовпадающей. Для сравнения результатов работы сети, была построена кривая зависимости истинно-положительных срабатываний от ложно-положительных срабатываний, при изменении порога допущения, приведённая на рисунке 2.

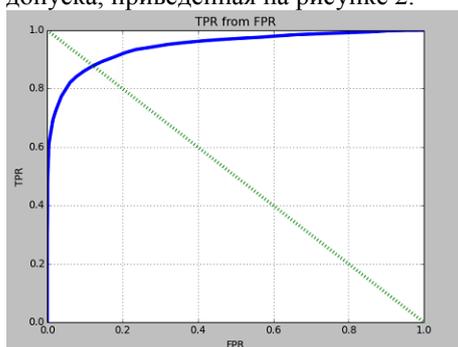


Рис. 2. Зависимость доли истинно-положительных (TPR) от доли ложно-положительных (FPR) срабатываний

В качестве точечной оценки принято использовать значение доли истинно-положительных срабатываний, лежащее на диагонали, показанной на рисунке зелёной пунктирной линией. Чем выше данное значение,

тем качественней в смысле распознавания считается алгоритм.

Для обученной сети данное значение равно 0,878. Данный результат превосходит результаты, приведенные в [5] - 0,8445 и в [6] - 0,8554 для алгоритмов, не использующих нейронные сети. Кроме того, данный результат превосходит результат 0,876 приведённый в [7], использующий ансамбль из 16 свёрточных сетей и метод JoinBayesian как меру сходства.

Заключение

В результате проделанной работы была построена архитектура свёрточной нейронной сети. Предложенная свёрточная нейронная сеть была обучена с использованием библиотеки Caffe на базе Casia-WebFace. Обученная сеть была протестирована на базе лиц LFW. Полученные результаты были соотнесены с результатами других алгоритмов, использующих и не использующих нейронные сети. Результаты тестирования показали, что используемый подход к построению архитектуры и обучению нейронных сетей является перспективным, и может быть использован в разработке алгоритмов распознавания лиц.

В дальнейших работах планируется уделить больше внимания изучению влияния параметров метода обучения на результаты работы сети, влияние качества обучающей выборки, а также возможность использования других функций активации.

Список использованных источников

1. Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, Trevor Darrell Caffe: Convolutional Architecture for Fast Feature Embedding // arXiv preprint arXiv:1408.5093. - 2014.
2. Dong Yi, Zhen Lei, Shengcai Liao and Stan Z. Li Learning Face Representation from Scratch // arXiv preprint arXiv:1411.7923. - 2014.
3. D. Kingma, J. Ba Adam: A Method for Stochastic Optimization // arXiv preprint arXiv:1412.6980. - 2015.
4. Gary B. Huang, Marwan Mattar, Honglak Lee, Erik Learned-Miller Learning to Align from Scratch // Advances in Neural Information Processing Systems. - 2012.
5. N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar Attribute and Simile Classifiers for Face Verification // IEEE International Conference on Computer Vision. - 2009.
6. Cao Z, Yin Q, Tang X, Sun J Face recognition with learning-based descriptor // Computer Vision and Pattern Recognition. - 2010. - C. 2707-2714.
7. Hu G, Yang Y, Yi D, Kittler J, Christmas W, Li SZ, Hospedales T. When face recognition meets with deep learning: an evaluation of convolutional neural

networks for face recognition // IEEE International
Conference on Computer Vision. - 2015. - С. 142-
150.