

Применение сверточных нейронных сетей для поиска ключевых точек лица человека

В.Ф. Кассым¹

¹Московский физико-технический институт (государственный университет)

В данной работе рассматривается задача автоматической локализации ключевых точек на изображении лица человека, таких как уголки глаз и рта, кончик носа и другие. Пример таких ключевых точек приведен на рис. 1. Решение данной проблемы может стать основой для большого количества задач, связанных с детектированием лиц, идентификацией, анализом эмоций, чтением по губам.

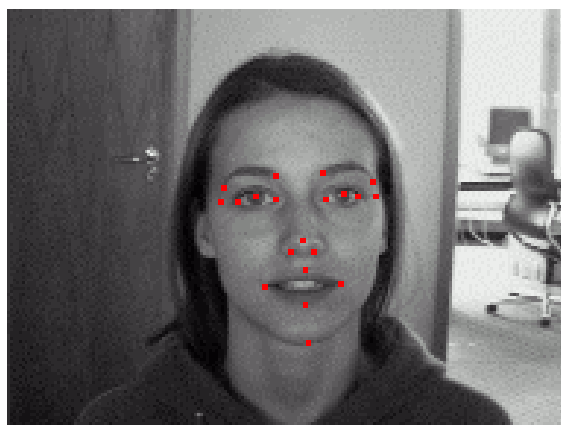


Рис. 1 Ключевые точки лица человека

Задача локализации ключевых точек лица, как и многие другие задачи компьютерного зрения, сталкивается с такими трудностями как изменение освещения, позы и выражения лица, частичное перекрытие объекта. В связи с этим возникает необходимость в извлечении особенностей исследуемого изображения, в той или иной степени инвариантных к данным изменениям.

Для решения данной задачи был предложен подход, основанный на использовании сверточных нейронных сетей [1]. В архитектуру данного вида нейронных сетей заложены априорные знания о предметной области, а именно: локальная корреляция пикселей изображений и инвариантность к смещению признаков. Таким образом, извлекаемые обученной нейронной сетью признаки, удовлетворяют описанным выше требованиям. Также стоит заметить, что глобальные зависимости между ключевыми точками лица строятся в процессе обучения и неявно используются в процессе локализации.

Основная структура сверточной нейронной сети была представлена в работе [1]. Она состоит из следующих типов слоев:

- слой свертки;
- слой подвыборки;
- слой полносвязной многослойной нейронной сети.

Рассмотрим слой из $N \times N$ нейронов после которого расположен слой свертки. Сверточный слой состоит из набора различных фильтров ω размера $m \times m$. Размер слоя нейронов после свертке будет равен $(N - m + 1) \times (N - m + 1)$, где значение индуцированного локального поля v_{ij} будет связано с входным слоем следующим соотношением:

$$v_{ij} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \omega_{ab} Y_{(i+a)(j+b)}. \quad (1)$$

Слой подвыборки, который следует за сверточным слоем уменьшает размерность полученного слоя нейронов. Обычно для каждой области размера $k \times k$ на выходе данного слоя берется максимальное значение в данной области. Так для входного слоя размера $N \times N$ выходной слой будет иметь размер $\frac{N}{k} \times \frac{N}{k}$.

Также для решения проблемы переобучения часто используют исключаящий слой, подробно описанный в [2]. Суть данного слоя заключается в “отключении” входящего нейрона с некоторой вероятностью. Также использование данного слоя может привести к ускорению обучения.

В данной работе было проведено моделирование сверточных нейронных сетей, конфигурация которых приведены в табл.1, табл. 2 и табл. 3.

#	Тип слоя	Описание
0	Входной	Матрица интенсивностей пикселей изображения
1	Сверточный	32 фильтра размера 3x3
2	Подвыборки	Размер области 2x2
3	Слой исключения	Вероятность исключения 0.1
4	Сверточный	64 фильтра размера 2x2
5	Подвыборки	Размер области 2x2
6	Слой исключения	Вероятность исключения 0.2
7	Сверточный	128 фильтра размера 2x2
8	Подвыборки	Размер области 2x2
9	Слой исключения	Вероятность исключения 0.3
10	Скрытый	1000 нейронов
11	Слой исключения	Вероятность исключения 0.5
12	Скрытый	1000 нейронов
13	Выходной	30 нейронов

Таблица 1 Конфигурация сверточной нейронной сети #1

#	Тип слоя	Описание
0	Входной	Матрица интенсивностей пикселей изображения
1	Сверточный	32 фильтра размера 5x5
2	Подвыборки	Размер области 2x2
3	Слой исключения	Вероятность исключения 0.1
4	Сверточный	64 фильтра размера 3x3
5	Подвыборки	Размер области 2x2
6	Слой исключения	Вероятность исключения 0.2
7	Сверточный	128 фильтра размера 2x2
8	Подвыборки	Размер области 2x2
9	Слой исключения	Вероятность исключения 0.3
10	Скрытый	1000 нейронов
11	Слой исключения	Вероятность исключения 0.5
12	Скрытый	1000 нейронов
13	Выходной	30 нейронов

Таблица 2 Конфигурация сверточной нейронной сети #2

#	Тип слоя	Описание
0	Входной	Матрица интенсивностей пикселей изображения
1	Сверточный	32 фильтра размера 7x7
2	Подвыборки	Размер области 2x2
4	Сверточный	64 фильтра размера 4x4
5	Подвыборки	Размер области 2x2
6	Сверточный	128 фильтра размера 3x3
7	Подвыборки	Размер области 2x2
8	Сверточный	256 фильтра размера 2x2
9	Подвыборки	Размер области 2x2
10	Сверточный	512 фильтра размера 3x3
11	Подвыборки	Размер области 3x3
12	Скрытый	1000 нейронов
13	Скрытый	1000 нейронов
14	Выходной	30 нейронов

Таблица 3 Конфигурация сверточной нейронной сети #3

Также проведено моделирование методов активных моделей внешнего вида (MAMBВ)

[3] и локально ограниченных моделей (МЛОМ) [4].

Рассматриваемые методы были смоделированы на языке Python, при помощи библиотек Theano, Lasagne, Nolearn, Менро и OpenCV.

Результаты работы исследованных методов локализации ключевых точек были оценены с помощью корня из среднеквадратичной ошибки (RMSE).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (2)$$

где y_i – точное значение предсказываемой координаты, а \hat{y}_i – ее приближение.

График зависимости RMSE обучения данных нейронных сетей, описанных в таблицах 1, 2 и 3 приведён на рис. 2.

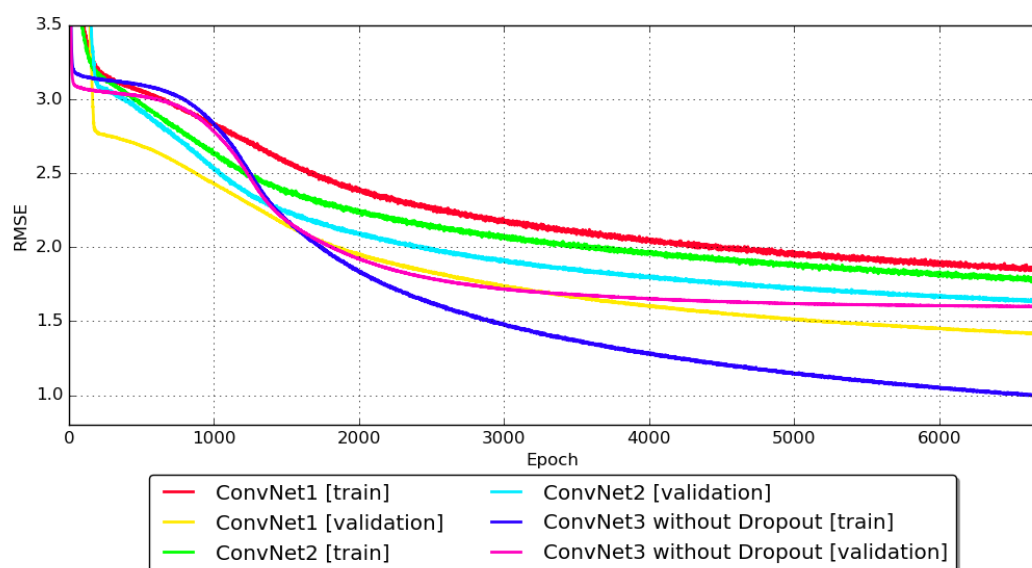


Рис. 2 График зависимости RMSE от эпохи обучения сверточных нейронных сетей

В табл. 4 приведены значения RMSE на тестовой выборке всех исследуемых в данной работе методов поиска ключевых точек. Из таблицы видно, что наименьшую ошибку показали решения на основе сверточных нейронных сетей.

	СНС #1	СНС #2	СНС #3	МАМВВ	МЛОМ
RMSE	2.368	2.302	2.624	6.631	5.053

Таблица 4 Общая таблица RMSE на тестовой выборке для всех исследуемых методов

В данной работе было проведено исследование методов поиска ключевых точек лица человека на изображении. Была показана актуальность данной проблемы как одной из составляющих обширной области практических и теоретических задач.

Были представлены результаты исследования различных методов, в рамках которого было проведено моделирование различных архитектур сверточных нейронных сетей, активных моделей внешности и локально ограниченных моделей.

Полученные результаты исследования были проанализированные и сделаны выводы о преимуществе нейросетевого подхода при решении задачи поиска ключевых точек лиц.

Литература

1. *Le Cun B. B.* et al. Handwritten digit recognition with a back-propagation network //Advances in neural information processing systems. – 1990. С. 396-404.
2. Hinton G. E. et al. Improving neural networks by preventing co-adaptation of feature detectors //arXiv preprint arXiv:1207.0580. – 2012.
3. *Cootes T. F., Edwards G. J., Taylor C. J.* Active appearance models //Computer Vision—ECCV'98. – Springer Berlin Heidelberg, 1998. – С. 484-498.
4. *Cootes T. F., Edwards G. J., Taylor C. J.* Active appearance models //Computer Vision—ECCV'98. – Springer Berlin Heidelberg, 1998. – С. 484-498.