

Исследование точности детектирования лиц по изображениям в зависимости от расы и пола при помощи каскадов Хаара

Е.А. Рудинская¹, Р.А. Парингер¹

¹Самарский национальный исследовательский университет имени академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

Аннотация. Распознавание предметов на изображениях очень часто применяется в современной жизни. Существует множество факторов, от которых во многом зависит точность детектирования и распознавания. Принадлежность человека к определенному полу или расе – одно из них. Данная статья посвящена разработке методологии, которая позволила бы оценить выбранный алгоритм на универсальность, а также оценке влияния обучающей выборки на результат детектирования. Была рассмотрена устойчивость результатов детектирования каскадов Хаара к набору данных, основанном на биологических признаках человека. Особое внимание уделялось поиску необходимых этапов методологии, поиску значений параметров алгоритма, подбору обучающей выборки и получению результатов с большей точностью детектирования. Самым детектируемым каскадами Хаара классом оказался EuropeanMan с результатом 11/12. Лучший результат детектирования показали каскады «frontalface_alt», «frontalface_alt2» и «eye_tree_eyeglasses» с точностью 5/6.

1. Введение

В настоящее время в связи со стремительным развитием различных систем наблюдений актуальна задача детектирования объектов, в частности - лиц, на изображениях или видео (например, в охранных системах слежения) [1, 2, 3]. Огромный вклад в этой области был внесен методом Виолы-Джонса – детектором лица, который способен находить лица в режиме реального времени с высокой точностью [4, 5].

Каскады относятся к семейству универсальных алгоритмов, которые могут быть обучены на детектирование любого объекта на изображении при наличии необходимого набора данных. С учетом современных возможностей и новых методов задача детектирования лиц становится еще более привлекательной и осуществимой. Именно поэтому область детектирования лиц является перспективной и востребованной. В частности, его используют для нахождения человека и слежения за ним [6]. В Facebook также используется алгоритм детектирования для обнаружения лиц на изображениях и их распознавания.

Информация о количестве людей требуется во многих системах пропускного контроля различного рода учреждениях, например, аэропортах, метро, в учреждениях автоматического учета числа посетителей и многих других системах наблюдений [7, 8]. Все чаще можно слышать о применении того или иного метода для решения разных задач [9, 10]. Также детектирование используют в качестве первого шага при решении задачи распознавания лиц на изображениях [11], т.к. это ускоряет работу метода. Многие исследования направлены на улучшение алгоритма, например, на увеличение точности или производительности, но не всегда уделяют достаточное внимание изучению причины возникновения каких-либо недостатков.

Зачастую можно услышать о новых достижениях нейронных сетей, новых методах и подходах к решению той или иной задачи. Ознакомившись со статьями, можно прийти к выводу, что очень часто возникает ситуация ложного распознавания, вызванная некорректным детектированием, что потенциально может привести к критическим последствиям.

Одним из первых этапов разработки алгоритма машинного зрения является этап обучения, когда алгоритм тренируется работать и выполнять поставленную задачу на определенном обучающем наборе данных. Однако, мало кто учитывает, что от корректности этого набора данных будет зависеть финальный результат. Например, корректно или некорректно подобранный набор данных естественным образом влияет на поведение алгоритма в дальнейшем. Ложные распознавания происходят из-за факторов, которые не были учтены при подготовке обучающей выборки. К таким факторам можно отнести различное освещение, другой угол поворота фотографии, задний фон, принципиально разное качество фотографии и другие факторы. Поэтому при обучении очень важно внимательно и обдуманно подходить к составлению обучающего набора данных.

Сама проблема влияния обучающей выборки на алгоритмы и их позиционирование не до конца изучена, плохо раскрыта и никем не учитывается. Так, если алгоритм детектирования лиц обучался на каком-то одном наборе данных, который состоял из преимущественно представителей, например, европеоидной расы, то корректнее было бы его называть алгоритмом детектирования европеоидных лиц. Т.е. из-за того, что недостаточное внимание уделяется обучающей выборке, то алгоритм не может позиционироваться универсальным. Его можно позиционировать только как алгоритм, выявляющий основные черты именно обучающей выборки. Значит можно предположить, что для предотвращения ложных распознаваний и повышения точности выбранного метода, требуется правильно подобранная обучающая выборка.

В предыдущей статье [12] каскады Хаара комбинировались для увеличения точности детектирования и выяснения более удачных комбинаций. Было выяснено, что при комбинировании каскадов нужно учитывать особенности каждого из них (например, какие признаки используются для поиска, значения параметров и т.д.), чтобы детектирование было более качественным. Также было необходимо учитывать возможность пересечения найденных областей, чтобы не посчитать одно и то же лицо дважды. Однако все приведенные варианты оптимизации не позволили добиться результатов, которые обычно должны соответствовать каскадам. Попытка выяснения причины такого поведения привела к данному исследованию.

Ознакомившись со статьей [11], в которой описывался эксперимент с нейронными сетями и людьми разных биологических признаков (при этом авторы использовали каскады Хаара для первоначального детектирования необходимых областей), было принято решение провести аналогичное исследование, но с более фундаментальным методом - при помощи каскадов Хаара, предназначенным для детектирования лиц, а не их распознавания.

Целью данного исследования являлось постановка эксперимента на одном конкретном детекторе лиц, которое позволило бы сформулировать методологию проверки адекватности работоспособности этого детектора.

2. Методология

Для исследования алгоритма на предмет корректности работы с различными входными данными необходимо:

1. Сформулировать проблематику исследования
2. Выбрать алгоритм, при помощи которого будет проводиться исследование
3. Подготовить правильную выборку, соответствующую проблеме, и определить ожидаемые результаты
4. Провести исследование влияния параметров работы алгоритма на точность детектирования подготовленной выборки
5. Провести эксперимент

6. Вычислить значение ошибки, для возможности определения вхождения результатов в доверительный диапазон значений, что позволит выяснить поведение алгоритма на данной выборке
7. Обработать результаты эксперимента

3. Каскады Хаара

Так как проблема исследования уже была определена, то в соответствии со вторым пунктом методологии для проведения исследований были использованы каскады Хаара.

Каскад Хаара — способ обнаружения объектов на изображении, основанный на машинном обучении, идея которого была предложена в статье за авторством Пола Виолы (Paul Viola) и Майкла Джонса (Michael Jones). [4, 5] Принимая на вход изображение, обученный каскад Хаара определяет, есть ли на нем искомый объект, т.е. выполняет задачу классификации, разделяя входные данные на два класса (есть искомый объект, нет искомого объекта). Правильно обученный каскад Хаара имеет хорошую скорость выполнения классификации, а также неплохую устойчивость к отклонениям разного рода. [13, 14]

Основные признаки каскадов:

Масштабный фактор (scalefactor) – параметр, который определяет размер изображения при каждом его шаге. Он используется для создания масштабной пирамиды (представление изображений таким образом, чтобы мы могли обнаружить как маленькие, так и большие лица, используя одно и то же окно обнаружения). От его значения зависит тщательность исследования каждой области.

Размер скользящего окна (minsize) – параметр, определяющий минимальный возможный размер объекта, т.е. области меньше его игнорируются. Этот размер устанавливается самостоятельно.

Параметр, определяющий количество возможных соседей (minNeighbors) – отвечает за количество соседей, которое может иметь каждый прямоугольник. Этот параметр влияет на качество обнаруженных лиц: чем больше значение, тем меньше обнаружений, но выше качество. Оптимальные значения для него 3-6.

4. Результаты исследования

В соответствии с третьим пунктом методологии для проведения исследований была составлена выборка из 600 различных фотографий, взятых из общедоступного набора изображений лиц IMDB-WIKI [15] и классифицированных по половому признаку и расовой принадлежности на 6 составляющих. Были использованы фотографии с изображением женщин и мужчин негроидной, монголоидной и европеоидной рас. Все фотографии были разделены на 6 классов по 100 фотографий в каждой в зависимости от принадлежности биологическим признакам: MongolWoman, MongolMan, EuropeanWoman, EuropeanMan, BlackWoman, BlackMan. При проведении эксперимента использовались самостоятельные каскады Хаара с двумя постоянными параметрами: скользящим окном квадратной формы с длиной стороны 30 пикселей и параметром, определяющим количество возможных соседних окон равного 5.

Четвертым шагом было необходимо определить правильные параметры работы алгоритма: при обработке результатов верным детектированием считалось только то значение, которое соответствовало числу необходимых детектирований, приведенных в таблице 1. Если же результат отличался даже на единицу (допустим, был детектирован один глаз вместо двух), то детекция считалась неверной.

Далее проводим сам эксперимент – пятый пункт методологии.

Первый этап заключался в оценке ошибки детектирования объектов на изображениях с использованием каждого из готовых каскадов Хаара для каждой группы классификации при определенном значении масштабного фактора, которое варьировались между 1,01 и 1,5 с шагом 0,01.

Полученные результаты представлены на рисунке 1, где вертикальная ось – отношение количества фотографий, где не было ни одной детекции, к количеству всех изображений, а горизонтальная – значение масштабного фактора.

Таблица 1. Ожидаемые значения результатов детектирования.

Название каскада	eye	eye_tree_eyeglasses	frontalcatface_extended	Frontalface	frontalface_alt	frontalface_alt2	frontalface_alt_tree	frontalface_default	lefteye_2splits	lowerbody	profileface	smile	righteye_2splits
Ожидаемые результаты распознавания	2	2	1	1	1	1	1	1	1	1	1	1	1

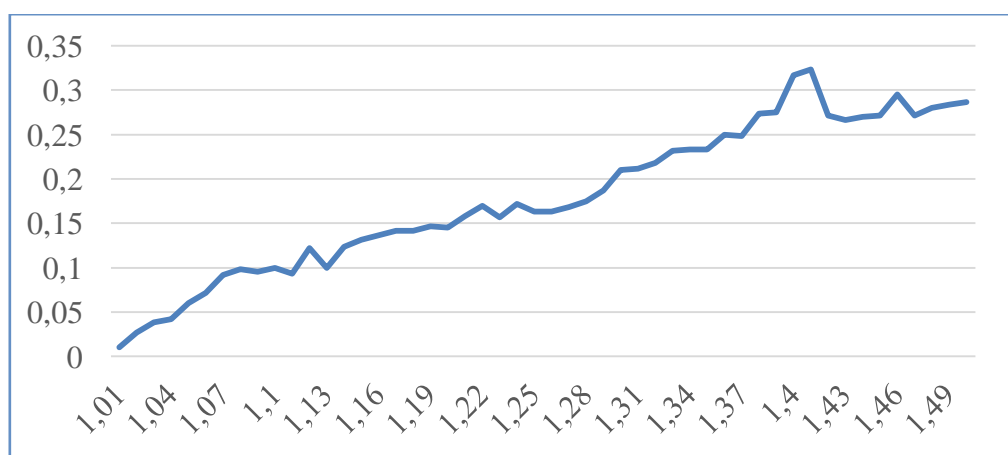


Рисунок 1. Доля ложных детекций в зависимости от значения масштабного фактора.

Анализируя результаты первого эксперимента, можно сделать выводы:

- Чем меньше значение масштабного фактора, тем больше верных детекций. Соответственно максимальное число верных детектирований было достигнуто при значении масштабного фактора равного 1,01.

Второй этап эксперимента заключался в исследовании точности детектирования лиц по каждому изображению в отдельности. Были использованы те же значения масштабного фактора и необходимые значения полученных результатов, что были представлены выше. Считалось, что изображение может детектироваться каскадами, если хотя бы при одном значении параметра один из самостоятельных каскадов давал верную детекцию изображения. Ниже, в таблице 2, приведены результаты по количеству детектирований. В первой строке представлены все значения детекций, которые были получены при проведении эксперимента, а во второй строке – количество фотографий, у которых было соответственное количество детекций.

Исходя из этой таблицы можно сделать выводы:

- Максимальное число изображений было достигнуто при детектировании лиц семью различными каскадами, на втором месте – восьмью.
- Из 600 изображений осталось лишь одно, которое не смог детектировать ни один каскад. Оно приведено ниже на рисунке 2
- Максимально достигнутое количество детекций одной фотографии равняется 12 (из возможных 13)

Таблица 2. Число детекций по изображениям.

Значение полученных детекций	0	1	2	3	4	5	6	7	8	9	10	11	12
Количество изображений	1	15	23	28	63	80	83	118	103	48	27	9	2



Рисунок 2. Изображение, которое не смог детектировать ни один каскад.

Далее из одной большой таблицы с результатами мы сделали таблицы, в которых считали количество верных детекций каждым каскадом, но уже не по отдельным изображениям, а по каждому из шести классов при определенном значении масштабного фактора.

Устойчивым каскадом к биологическим признакам можно назвать тот, значение ошибки которого будет отличаться не больше, чем стандартное отклонение от среднего значения (шестой пункт методологии).

После получения результатов, среднее значение по каждому каскаду находилось по формуле

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n x_j,$$

где \bar{X} – среднее значение, n – общее количество значений, x_j – каждый элемент.

Далее находим сигму

$$\sigma(X) = (D(X))^{1/2}, \text{ где } D(X) = M(X^2) - (M(X))^2$$

Здесь $D(X)$ – дисперсия, а $M(X)$ – математическое ожидание.

Доверительный диапазон значений вычислялся по формуле:

$$\bar{X} - \sigma(X) < \mu < \bar{X} + \sigma(X)$$

После нахождения проводился анализ: какие каскады давали более стабильный результат.

В таблице 3 представлена таблица при значении масштабного фактора равного 1.01, где указано количество верных детектирований.

Таблица 3. Результаты верных детектирований каждого самостоятельного каскада по отдельным расовым признакам.

Название класса	eye	eye_tree_eyeglasses	frontalcaiface_extended	frontalface	frontalface_alt	frontalface_alt2	frontalface_alt_tree	frontalface_default	lefteye_2splits	profileface	righteye_2splits	smile
MongolWoman	16	41	35	38	81	73	49	56	22	70	22	5
MongolMan	24	30	27	47	80	77	51	69	22	61	27	9
BlackWoman	29	42	23	29	81	78	43	68	19	55	26	10
BlackMan	23	14	16	20	62	59	22	53	24	58	19	12
EuropeanWoman	18	40	38	41	87	79	63	58	18	63	25	7
EuropeanMan	24	32	32	38	81	77	43	60	22	59	20	2

- Зеленым цветом выделены те результаты, которые соответствуют доверительному диапазону значений
- Из таблицы можно видеть, что каскад «smile» лучше остальных детектировал BlackMan, но его значение выходит за пределы доверительного диапазона
- Максимальные значения были получены при помощи каскадов «frontalface_alt» и «frontalface_alt2»
- MongolWoman и MongolMan имеют примерно одинаковый процент детектирования.
- EuropeanWoman идет на первом месте по среднему числу верных детектирований, на втором - MongolWoman и MongolMan, самыми мало детектируемые – BlackMan.

Результаты оценивались по среднему значению отдельно взятого класса.

Исходя из такой оценки результатов эксперимента, можно предположить, что каскады обучались преимущественно на представителях женского пола европеоидной расы, т.к. они больше остальных верно детектировались. Из-за большего сходства с ними на втором месте шли мужчины и женщины монголоидной расы, после них мужчины европеоидной и женщины негроидной расы, а на последнем месте – мужчины негроидной расы.

После был введен критерий оценивания для ранжирования отдельных классов и каскадов. Верным результатом является только тот, который вошел в доверительный интервал. В итоге получаем: на первом месте по верным детекциям идет класс EuropeanMan с результатом 11 из 12 возможных; потом MongolWoman, MongolMan, BlackWoman и EuropeanWoman имеют 9 из 12, а вот класс BlackMan всего лишь 3 из 12. Из каскадов лучше всех с задачей детектирования справились: «eye_tree_eyeglasses», «frontalface_alt» и «frontalface_alt2» с результатом 5/6, меньше всего детектирований дал каскад «frontalface_default» - 1/2, а оставшиеся каскады работали одинаково с результатом 2/3.

5. Выводы

Была предложена и опробована методология, которая позволяет выяснить универсальность выбранного метода, в частности проверить каскады на устойчивость результатов каскадов при входных данных с изображениями людей разных биологических признаков. Каскады Хаара нельзя позиционировать как универсальный детектор, поскольку они плохо детектируют BlackMan. Можно предположить, что скорее всего это вызвано проблемой с обучающей выборкой, например, было взято недостаточное количество фотографий представителей негроидной расы. Поэтому необходимо корректно составлять обучающую выборку, которая будет соответствовать поставленной задаче и соответствующим образом в дальнейшем будет позиционировать разработанный алгоритм. В ходе эксперимента было выяснено, что чем меньше значение масштабного фактора, тем больше верных детекций. Лучше всего начинать детектирование при значениях масштабного фактора в пределах 1,01 – 1,1. При составлении обучающей выборки надо подходить с максимальной ответственностью и предусматривать различные варианты исхода событий. При обучении алгоритма детектирования лиц необходимо брать довольно большое количество изображений с людьми разных возрастов, полов, рас и других биологических признаков. Также необходимо предусматривать предварительную обработку фотографий, чтобы не было размытости и неточности во входных данных. Было выяснено, что самым детектируемым классом оказался EuropeanMan с результатом 11/12, а для первоначального детектирования лучше всего подходят «frontalface_alt», «frontalface_alt2» и «eye_tree_eyeglasses», которые имеют результат точности 5/6.

6. Благодарности

Работа выполнена в рамках государственного задания по теме FSSS-2020-0017, при частичной финансовой поддержке Российского фонда фундаментальных исследований № 19-29-01135.

7. Литература

- [1] Sirovich, L. Low dimensional procedure for characterization of human faces / L. Sirovich, M. Kirby // *Journal of the Optical Society of America A*. – 1987. – Vol. 4. – P. 519.
- [2] Turk M. Face recognition using eigenfaces / M. Turk, A. Pentland // *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. – 1991. – P. 586-591.
- [3] Zhao, W. Face Recognition: A Literature Survey / W. Zhao, R. Chellappa, A. Rosenfeld, P.J. Phillips // *ACM Computing Surveys*. – 2003. – Vol. 35(4). – P. 399-458. DOI: 10.1145/954339.954342.
- [4] Viola, P. Rapid Object Detection using a Boosted Cascade of Simple Features / P. Viola, M.J. Jones // *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2001.
- [5] Viola, P. Robust real-time face detection / P. Viola, M.J. Jones // *International Journal of Computer Vision*. – 2004. – Vol. 57(2). – P. 137-154.
- [6] Arlazarov, V.V. MIDV-500: a dataset for identity document analysis and recognition on mobile devices in video stream / V.V. Arlazarov, K. Bulatov, T. Chernov, V.L. Arlazarov // *Computer Optics*. – 2019. – Vol. 43(5) – P. 818-824. DOI: 10.18287/2412-6179-2019-43-5-818-824.
- [7] Гонсалес, Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – Изд-во: Техносфера, 2005. – 1072 с.
- [8] Форсайт, Д. Компьютерное зрение. Современный подход / Д. Форсайт, Ж. Понс. – Изд-во: Вильямс, 2004. – 928 с.
- [9] Bezmaternykh, P.V. U-Net-bin: hacking the document image binarization contest / P.V. Bezmaternykh, D.A. Ilin, D.P. Nikolaev // *Computer Optics*. – 2019. – Vol. 43(5). – P. 825-832.
- [10] Stengcai, L. A Fast and Accurate Unconstrained Face Detector / L. Stengcai, K. Jain Anil, Z. Li Stan // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2016. – Vol. 38(2) – P. 211-223.
- [11] Sarfaraz, M. Prediction of Human Ethnicity from Facial Images Using Neural Networks / M. Sarfaraz, Sh. Gupta, A. Wajid, S. Gupta, A. Musher // *A Novel Cluster Algorithms of Analysis and Predict for Brain Derived Neurotrophic Factor (BDNF) Using Diabetes Patients*. – 2018. – P. 217-226.
- [12] Рудинская, Е.А. Разработка технологии детектирования лиц с использованием комбинаций каскадов Хаара / Е.А. Рудинская, Р.А. Парингер // V международная конференция и молодежная школа «Информационные технологии и нанотехнологии» (ИТНТ) – Самара: Новая техника, 2019. – С. 6-12.
- [13] Калиновский И.А. Обзор и тестирование детекторов фронтальных лиц / И.А. Калиновский, В.Г. Спицын // *Компьютерная оптика*. – 2016. – Т. 40, №1 – С. 99-111. DOI: 10.18287/2412-6179-2016-40-1- 99-111.
- [14] Анастасов, А.В. Применение каскада Хаара для обнаружения регистрационных государственных знаков транспортных средств на изображении / А.В. Анастасов // XII Международная научно-практическая конференция студентов, аспирантов и молодых учёных «Молодёжь и современные информационные технологии» - Томск, 2014. – Т. 2. – С. 126-127.
- [15] Rothe, R. IMDB-WIKI – 500k+ face images with age and gender labels [Electronic resource]. – Access mode: <https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/>.

Study of a face detection accuracy based on race and gender using Haar cascades

E.A. Rudinskaya¹, R.A. Paringer¹

¹Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

Abstract. Object recognition in images is very often used in modern life. Many factors greatly affect the accuracy of detection and recognition. Belonging to a particular gender or race is one of them. This article is devoted to the development of a methodology that would allow evaluating the selected algorithm for universality, as well as evaluating the impact of the training sample on the detection result. The stability of the results of detecting Haar cascades to a data set based on human biological characteristics was considered. Special attention was paid to finding the necessary stages of the methodology, searching for the values of the algorithm parameters, selecting a training sample and obtaining results with greater detection accuracy. The class most detected by Haar cascades was EuropeanMan with a score of 11/12. The best detection result was shown by the cascades "frontalface_alt", "frontalface_alt2" and "eye_tree_eyeglasses" with an accuracy of 5/6.