Сравнение параметрически оптимизированных реализаций метода Виолы–Джонса и MTCNN

А. Д. Егоров[1], А. Ф. Идиятуллин[2], А. Д. Закиров[3]
Национальный исследовательский ядерный университет МИФИ
[1]ADEgorov@mephi.ru, [2]almaz.idiyatullin7@gmail.com, [3]arthur.zakirovd@gmail.com

***Аннотация*. Нахождение лиц на изображении – классическая задача компьютерного зрения. Решение этой задачи необходимо для использования в широком круге приложений, в первую очередь связанных с обеспечением элементов управления (доступ к данным по лицу человека) и контроля (системы контроля безопасности периметра). Наиболее популярные методы нахождения лиц объектов делятся на два класса: классические (типовой пример – метод Виолы–Джонса) и нейросетевые (типовой пример – каскадная свёрточная нейронная сеть). Существует большое количество сравнений методов нахождения лиц, однако они зачастую не учитывают потенциал для оптимизации по параметрам рассматриваемых реализаций. В данной работе проводится сравнение параметрически оптимизированных реализаций типовых методов каждого из классов. Оптимизация проводится по предложенному в работе алгоритму. Показано, что оптимизированный метод Виолы–Джонса проигрывает MTCNN на 20–50 % с точки зрения метрики качества, но выигрывает в 7–14 раз по скорости обработки одного кадра. Показано, что без использования оптимизации алгоритмов по параметрам невозможно добиться хорошего качества и производительности, так как среднее значение метрики качества и производительности заметно ниже, чем оптимальные.**

Ключевые слова: компьютерное зрение, метод Виолы–Джонса, каскадная свёрточная нейронная сеть, оптимизация, производительность

I. Общая постановка задачи и проблематизация

Область работы с различными методами компьютерного зрения для поиска лиц хорошо разработана [1]. Для многих ключевых методов обнаружения лиц приведено достаточно большое количество модификаций, которые позволят добиться необходимого качества работы для практического использования. Однако, в «коробочных» решениях модификации не используются, а практическая реализация модификаций осложнена. Наибольшее качество работы «коробочных» решений обычно означает их малую производительность. Однако, например, в работах [2, 3] показано, что за счёт оптимизации параметров реализации Метода Виолы–Джонса можно добиться заметного прироста производительности с минимальной потерей качества. Аналогичных исследований для наиболее известной нейросетевой модели для поиска лиц – MTCNN – не проводилось, однако, например, в работе [4] упоминается, что параметры влияют на результаты эксперимента. В работах [5, 6] приводится сравнение различных вариантов методов поиска лиц, однако не уделено внимание возможностям, связанным с параметрической оптимизацией данных «коробочных» решений.

В данной работе изучаются возможности параметрической оптимизации методов Виолы–Джонса и MTCNN, а также проводится сравнение значения метрики качества на основании собственного критерия качества и производительности параметрический оптимизированных методов поиска лиц.

II. Краткое описание сравниваемых методов

A. Метод Виолы–Джонса

Метод Виолы–Джонса состоит из нескольких ключевых компонент [7]: преобразование анализируемого изображения в интегральное представление; использование предобученных каскадов для нахождения лиц; обучение каскадов с помощью бустинговых алгоритмов. В основе метода Виолы–Джонса важную роль играет так называемая «пирамида» изображений: на каждом цикле работы метода размер изображения (w пикселей в ширину и h пикселей в высоту) уменьшается в размере в конкретное количество раз [8], которое обозначим sf_{Vj}. На sf_{Vj} накладывается следующее условие: $sf_{Vj} > 1$.

То есть: $w_i = w_{i-1} / sf_{Vj}$; $h_i = h_{i-1} / sf_{Vj}$. Для работы задаётся минимальный размер исследуемого изображения ms_{Vj}, размер измеряется в пикселях.

B. Каскадная свёрточная нейронная сеть MTCNN

Каскадная свёрточная нейронная сесть для поиска лиц MTCNN состоит из трёх последовательных свёрточных нейронных сетей [9]: первая сеть подбирает потенциальные окна-кандидаты то есть, окна в которых может находиться лицо; вторая сеть фильтрует найденные окна, оставляет только те, нахождения лица в которых наиболее вероятно; третья сужает область поиска через определение места нахождения ключевых точек. При этом, для работы сети так же используется пирамида изображений. На каждой итерации размер изображения уменьшается с коэффициентом $sf_{MT\ CNN} < 1$ по следующей формуле: $w_i = w_{i-1} * sf_{MTCNN}$; $h_i = h_{i-1} * sf_{MTCNN}$. Для работы задаётся минимальный размер исследуемого изображения ms_{MTCNN}.

III. Оптимизация и сравнение между собой оптимизированных методов

A. *Алгоритм оптимизации и сравнения между собой методов*

Для сравнения между собой оптимизированных методов необходимо последовательно выполнить следующие шаги:

- Выбрать схожие параметры в каждом из рассматриваемых методов.
- Провести оптимизацию параметров для каждого из алгоритмов по заранее выбранному критерию, которые учитывает качество работы (на основе выбранной метрики качества) и производительность.
- Сравнить качество работы оптимизированных методов на основании метрики качества работы и производительности.

В связи со сложной внутренней природой методов возможно единственный вариант для единообразной оптимизации: оптимизация по сетке параметров, которые для работы заложены алгоритмически. Это уже было показано на примере метода Виолы–Джонса [10]. В качестве параметров для перебора предлагается использовать следующую сетку:

- Минимальный размер: от 11 пикселей включительно до 50 пикселей включительно с шагом в 3 пикселя.
- Коэффициент изменения размера: от 0.05 включительно до 0.9 включительно с шагом в 0.05. При этом необходимо учесть, что $sf_{Vj} = 1/sf_{MTCNN}$.
- Для метода Виолы–Джонса – предобученные каскады с признаками Хаара: default; alt, alt_tree, alt2.

В качестве метрики качества предлагается использовать метрику Ван Ризбергена, известную так же как *f*-мера – среднегармоническое точности и полноты [11]. Использование точности или полноты самих по себе кажется недостаточно обоснованным в связи с тем, что при малой полноте всегда получается достаточно большая точность.

Расчёт точности и полноты строится на расчёте положительных, отрицательных и ложноположительных срабатываний. Положительным считается срабатывание в случае, если потенциальное лицо, найденное в случае срабатывания, пересекается с истинным положением лица на 70 или более процентов. Если не одно истинное положение лица не пересекается с найденными в результате срабатывания, то такое лицо считается ненайденным (а срабатывание – отрицательным).

Чтобы в критерии качества для сравнения методов между собой учитывалась производительность, по аналогии с [10] введём следующий критерий:

$$Q = (1 - F) + \frac{t}{t_{\max}}$$

где $t_{\max}$ – максимальное среднее время работы метода при обработке одного кадра (для всех сравниваемых методов) на одинаковом наборе данных, t – среднее время обработки одного кадра в эксперименте с конкретными параметрами, F – значение *f*-меры для эксперимента с конкретными параметрами. Наилучший набор параметров для такого критерия, это такой набор, при котором критерий достигает наименьшего значения, то есть *F*-мера наибольшая, а нормированное время наименьшее. Диапазон изменения значения *f*-меры: [0;1]. При этом единица измерения *f*-меры – пункт. Диапазон изменений значений критерия составляет промежуток (0; 2] и складывается из диапазонов изменений значений (1-*F*) – [0;1]; и нормированного времени (0;1].

Таким образом, на основании критерия *Q* предлагается провести оптимизацию каждого из методов, а затем сравнить оптимизированные значения критериев на различных наборах данных.

B. *Используемые наборы данных*

Ключевой недостаток метода Виолы–Джонса состоит в том, что с его помощью можно хорошо определять только фронтальные лица, то есть лица, которые повёрнуты чётко в камеру. Соответственно, для проведения сравнения необходимо выбрать как минимум два различных набора данных: с преобладающими фронтальными лицами и без преобладания фронтальных лиц.

В качестве набора данных с фронтальными лицами в работе используется набора данных FDDB [12], в качестве второго набора данных в работе используется UFDD [13]. Второй набор данных содержит большое количество сложных примеров, на которых в большинстве случае всех алгоритмы поиска лиц не смогут показать высокую точность.

IV. Полученные результаты

Для каждого из набора данных проведём проверку результатов работы методов для всех возможных параметров. Таким образом, проводится 252 эксперимента для MTCNN и 1008 экспериментов для четырёх различных каскадов метода Виолы–Джонса. В таблицах I, II, III, IV приведены результаты оптимизации методов с указанием значений ключевых параметров, метрики качества (*f*-меры), среднего времени обработки одного кадра в секундах и критерия. Все результаты, в которых значение *F* было невозможно рассчитать, были удалены из рассматриваемой выборки.

A. *FDDB*

Набор данных FDDB содержит 2845 изображений небольшого разрешения, на которых было размечено 5171 лицо [12]. Большая часть лиц на изображениях повёрнута лицом к камере, что позволяет методу Виолы–Джонса показывать относительно высокую эффективность.

ТАБЛИЦА I РЕЗУЛЬТАТ ОПТИМИЗАЦИИ ПО ПАРАМЕТРАМ MTCNN ДЛЯ НАБОРА FDDB (5 ЛУЧШИХ РЕЗУЛЬТАТОВ).

Tun	*SF*	*MinSize*	*F*	*t*	*Q*
MTCNN	0,35	50	0,8984	0,6151	0,3542
MTCNN	0,45	50	0,9057	0,6517	0,362
MTCNN	0,3	38	0,8834	0,6013	0,3635
MTCNN	0,5	26	0,9148	0,6784	0,3638
MTCNN	0,3	50	0,8932	0,6311	0,366

ТАБЛИЦА II РЕЗУЛЬТАТ ОПТИМИЗАЦИИ ПО ПАРАМЕТРАМ МЕТОДА ВИОЛЫ-ДЖОНСА ДЛЯ НАБОРА FDDB (5 ЛУЧШИХ РЕЗУЛЬТАТОВ)

Tun	*SF*	*MinSize*	*F*	*t*	*Q*
alt2	0,85	50	0,7552	0,048	0,2645
alt2	0,85	47	0,7552	0,048	0,2645
alt	0,9	47	0,7677	0,0785	0,2645
alt	0,9	50	0,7677	0,0787	0,2646
alt2	0,9	47	0,7652	0,0781	0,2668

Оптимизированный метод Виолы–Джонса позволяет добиться качества работы в 0.7552 пункта при среднем времени обработки одного кадра в 0.048 секунды. Оптимизированный MTCNN позволяет получить 0.8984 пункта по *f*-мере при среднем времени работы на один кадр в 0.6151 секунды.

B. UFDD

Набор данных UFDD содержит порядка 6424 изображения с 10895 размеченными лицами [13]. Разрешение изображений больше, чем в UFDD. В наборе данных присутствует большое количество помех для качественной работы методов поиска лица: дождь, снег, размытие, засветка и прочее.

ТАБЛИЦА III РЕЗУЛЬТАТ ОПТИМИЗАЦИИ ПО ПАРАМЕТРАМ MTCNN ДЛЯ НАБОРА UFDD (5 ЛУЧШИХ РЕЗУЛЬТАТОВ).

Tun	*SF*	*MinSize*	*F*	*t*	*Q*
MTCNN	0,3	38	0,5827	0,6898	0,5498
MTCNN	0,45	32	0,6068	0,8295	0,5524
MTCNN	0,4	38	0,6033	0,8121	0,5527
MTCNN	0,4	32	0,6014	0,8044	0,553
MTCNN	0,35	38	0,5938	0,7744	0,5549

ТАБЛИЦА IV РЕЗУЛЬТАТ ОПТИМИЗАЦИИ ПО ПАРАМЕТРАМ МЕТОДА ВИОЛЫ–ДЖОНСА ДЛЯ НАБОРА UFDD (5 ЛУЧШИХ РЕЗУЛЬТАТОВ)

Tun	*SF*	*MinSize*	*F*	*t*	*Q*
alt2	0,85	26	0,3201	0,2321	0,7244
alt2	0,85	23	0,3259	0,2707	0,7261
alt2	0,85	32	0,3127	0,2031	0,7263
alt2	0,85	29	0,3127	0,2032	0,7263
alt2	0,9	26	0,3414	0,3655	0,7288

Оптимизированный метод Виолы–Джонса позволяет добиться качества работы 0.3201 пункта при среднем времени обработки одного кадра в 0.2321 секунды. MTCNN показывает заметно лучшие результаты: 0.5827 пункта при среднем времени обработки одного кадра в 0.6869 секунды.

V. ДИСКУССИЯ О ПОЛУЧЕННЫХ РЕЗУЛЬТАТАХ

Целесообразно отдельно обсудить потенциал и результат оптимизации, и результат сравнения методов между собой.

Оптимизированный метод Виолы–Джонса на наборе FDDB позволяет за счёт потери 0.0125 пункта *F*-меры увеличить скорость обработки кадра на 68 %. Для MTCNN же верно обратное замечание: за счёт потери производительности на 10 % можно добиться роста качества на 0.0164. Похожее замечание можно и сделать и для набора UFDD. Выводы по итогам оптимизации метода Виолы–Джонса подтверждают предыдущие работы, например [2, 10] о том, что каскады alt, alt2 эффективнее справляются с поиском лиц, чем default и alt_tree.

Таким образом, оптимизация даёт заметный результат, но для её полноценного использования рекомендуется смотреть несколько лучших (топ-5, топ-10) результатов.

Анализируя результаты сравнения методов между собой в рамках данной работы можно подтвердить выводы работы [5]: MTCNN показывает лучшие результаты в любом случае, при этом производительность метода заметно ниже, чем у метода Виолы–Джонса.

В случае с набором данных FDDB, MTCNN по критерию Q показывает заметно худшие результаты, несмотря на то, что MTCNN работает качественнее, чем метод Виолы–Джонса на 0.1569 пункта по *F*-мере (на 25 % хуже), MTCNN работает в 14 раз медленнее, что практически не позволяет добиться работы метода в реальном времени. Наибольшей скорости обработки одного кадра с помощью MTCNN можно добиться при значении *F*-меры в 0.2626 пункта, что практически в 3 раза хуже, чем значение F-меры для оптимизированного метода Виолы-Джонса. При этом MTCNN работает в 7 раз медленнее, чем метод Виолы–Джонса.

В случае UFDD ситуация обстоит иным образом. Как было сказано раннее, UFDD сложнее устроен, чем FDDB, что совершенно закономерно приводит к низкому качеству работы метода Виолы–Джонса. Оптимальный результат метода Виолы–Джонса почти в два раза хуже, чем оптимальный результат MTCNN, причём выигрыш в производительности заметно ниже, чем в случае с FDDB – всего в три раза. Это связано с тем, что метод Виолы–Джонса сильно зависит от размера исходного изображения из-за того, что в нём применяется технология скользящего окна, в отличие от MTCNN, в которой изображение любого размера рассматривается целиком. Видно, что оптимальные результаты работы MTCNN по критерию Q имеют заметно меньшие значения, чем оптимальные результаты для метода Виолы–Джонса.

VI. ЗАКЛЮЧЕНИЕ

Оптимизация по параметрам для методов поиска лиц является одним из важных элементов технологии их практического использования. Среднее время обработки одного кадра для MTCNN на наборе FDDB для всей сетки параметров составляет 0.85 секунды, оптимальное – 0.61. Среднее значение *F*-меры для MTCNN 0.83, оптимальное

0.89. Аналогичное сравнение для метода Виолы–Джонса: средняя скорость обработки 0.044 секунды при среднем значении $F - 0.42$. Оптимальное значение производительности: 0.048 секунды. Оптимальное значение F – 0.7552. В случае, если оптимизация не используется, то добиться качественной работы методов можно только случайно.

Сравнение оптимизированных методов между собой подтверждаю ранее проведённые исследования. MTCNN заметно точнее метода Виолы–Джонса, но его не получается на текущий момент использовать в реальном времени.

Список литературы

[1] Zafeiriou S., Zhang C., Zhang Z. A survey on face detection in the wild: past, present and future // Computer Vision and Image Understanding. 2015. V. 138. P. 1-24.

[2] Egorov A. D., Divitskii D.U., Dolgih A.A., Mazurenko G.A. Some cases of optimization face detection methodes on image (Using the Viola-Jones method as an example) // 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus). IEEE, 2018. P. 1075-1078.

[3] Bruce B. R., Aitken J. M., Petke J. Deep parameter optimisation for face detection using the Viola-Jones algorithm in OpenCV // International Symposium on Search Based Software Engineering. Springer, Cham, 2016. P. 238-243.

[4] Kaziakhmedov E. et al. Real-world attack on MTCNN face detection system // 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON). IEEE, 2019. P. 0422-0427.

[5] Guo Y., Wünsche B. C. Comparison of Face Detection Algorithms on Mobile Devices // 2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ). IEEE, 2020. P. 1-6.

[6] Каляшов Е.В. сравнительный анализ систем распознавания лиц, построенных с использованием блоков стандартных архитектур // Информационные технологии и телекоммуникации. 2020. Т. 8. Вып. 3. С. 94–101. DOI 10.31854/2307-1303-2020-8-3-94-101.

[7] Viola P. et al. Robust real-time object detection //International journal of computer vision. 2001. V. 4. №. 34-47. P. 4.

[8] Viola P., Jones M. Rapid object detection using a boosted cascade of simple features // Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001. IEEE, 2001. V. 1. P. I-I.

[9] Zhang K. et al. Joint face detection and alignment using multitask cascaded convolutional networks //IEEE Signal Processing Letters. 2016. V. 23. №. 10. P. 1499-1503.

[10] Egorov A.D. Algorithm for optimization of Viola–Jones object detection framework parameters // Journal of Physics: Conference Series. IOP Publishing, 2018. V. 945. №. 1. P. 012032.

[11] Chan T.F., Vese L.A. Active contours without edges // IEEE Transactions on image processing. 2001. V. 10. №. 2. P. 266-277.

[12] Jain V., Learned-Miller E. Fddb: A benchmark for face detection in unconstrained settings // UMass Amherst technical report, 2010. V. 2. №. 4. P. 5.

[13] Nada H. et al. Pushing the limits of unconstrained face detection: a challenge dataset and baseline results // 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS). IEEE, 2018. P. 1-10.