

Методы анализа и структурированного распознавания лиц в естественных условиях

В.В. Дергачев, А.А. Александров

Ростовский Государственный Университет Путей Сообщения, Ростов-на-Дону

Аннотация: Настоящая статья посвящена исследованию существующих методов анализа и обнаружения лиц в естественных условиях. Распознавание лица на основании анализа изображения является одной из основных проблем в реализациях компьютерного зрения. Системы эталонных измерений позволяют определять и развивать различные методы обнаружения лица. Были проведены практические эксперименты и рассмотрены алгоритмы распознавания лиц, представленные научными школами и коммерческими разработками, а также проанализированы результаты оценок на всех тестовых наборах, основываясь на данных графиков истинно положительных значений, для оценки качества и скорости использования наборов данных для обнаружения лиц. При использовании детальной оценки, были проанализированы эффективность алгоритмов в различных аспектах в отношении нескольких атрибутов. Сравнивая производительность с различной степенью точности, определены преимущества и недостатки алгоритмов при использовании в различных сценариях работы.

Ключевые слова: компьютерное зрение, детектор лица, обнаружение лиц, распознавание лиц, анализ изображений, алгоритмы, методы обнаружения, эталонные измерения.

Введение

Существующие методы поиска и обнаружения лиц в визуальных наборах данных находят все более широкое применение в реальных приложениях, но по-прежнему они имеют значительное количество недостатков и используют устаревшие наборы данных. Новый набор данных для обнаружения лиц Multi-Attribute Labelled Faces (Мульти-атрибутно-помеченные лица, далее MALF) содержит 5250 изображений, собранных из Интернета, на которых отмечены около 12000 лиц.

Набор данных MALF имеет две главных особенности. Во-первых, это крупнейший набор данных для оценки обнаруженных лиц в естественных условиях с определением нескольких атрибутов лица, что повышает производительность системы в целом. Во-вторых, MALF позволяет оценить истинные характеристики алгоритмов на практике и может быть использован для оценки критериев, которые основываются на скорости отклика и на

относительно низкой частоте ложных срабатываний. Помимо большого набора данных для оценки обнаружения лиц, также представлен сравнительный анализ алгоритмов структурированного обнаружения лиц и их показатели.

Распознавание лица на основании анализа изображения играет важную роль и является одной из основных проблем в реализациях компьютерного зрения. Сейчас наблюдается значительный рост в разработке различных приложений использующих распознавание лиц в работе, от традиционной идентификации лица до поиска, обозначения и проверки по различным базам данных; разработчики таких приложений рассчитывают на точную и эффективную систему обнаружения лиц. Популярные детекторы, такие как метод Виола-Джонса [1] и его вариации (например, Vector Boosting [2]) добились удовлетворительных результатов на более ранних наборах данных, такие как CMU-MIT. Однако, как доказано на работах [3, 4], методы, основанные на принципе Виола-Джонса по-прежнему далеки от совершенства. В последнее время были предложены многие новые методы обнаружения лиц, такие, как DPM, CNN и Multiple Channel Features, которые позволяют повысить производительность обнаружения лиц. Наряду с данными исследованиями в распознавание лиц большие усилия, также приложили и коммерческие компании, такие как Google, Facebook и Face++. Однако, не удастся установить, какой среди всех вышеперечисленных подходов, лучший, и как улучшить их для использования в реальных приложениях. Таким образом, системы эталонных измерений позволяют определять и развивать различные методы обнаружения лица. К сожалению, еще не хватает объективной информации о методах обнаружения лиц в естественных условиях из-за следующих трех причин:

Первая, заключается в том, что существующие на сегодняшний день детекторы лиц не поддерживают детализированный анализ результатов, что

значительно затрудняет определение корреляции между причинами возникновения различных типов ошибок. Существующие критерии обнаружения лица имеют геометрическую рамку (или геометрический эллипс) результата, и соответственно они могут показать только общий результат обнаружения лица на весь тестовый набор. С другой стороны такие детали изображения, как позы, очки, выражения лица могут оказывать существенное влияние на результат обнаружения.

Во-вторых, существующие системы тестирования обнаружения лица не отражают реальный мир. На сегодняшний день наиболее широко используются системы оценки определения лиц, разработанные в первые годы возникновения систем распознавания, такие как MIT + CMU. Другой широко используемый тест обнаружения лица Fddb, алгоритм которого основывается на коллекции изображений, расположенных на новостных порталах. Однако лица и позы людей в этой коллекции изображений, как правило изображены фронтально. В последнее время для получения набора тестовых изображений широко используются изображения из социальных сетей: Flickr, Facebook и Google+, поскольку изображение лиц более разнообразны. Одной из последних крупномасштабных коллекций лиц – это база данных AFLW [5], которая выпускается с подробными описаниями и ориентирами. Вместе с тем некоторые изображения в этой базе данных не описаны, что делает её менее пригодной, для применения в качестве эталона обнаружения лица.

В-третьих, современные тесты обнаружения лиц не описывают «истинный» результат. Очень часто системы распознавания лиц используются только для сравнения производительности академических алгоритмов, например, MIT + CMU и Fddb. Однако, как говорилось ранее, существует большой разрыв между доступными в настоящее время научными решениями и коммерческими системами (например, Google Picasa)

и онлайн коммерческого API (например, Face ++). И как следствие сравнивая только академические методы и игнорируя коммерческие невозможно гарантировать высокое качество распознавания в реальных приложениях.

Коллекции изображений лиц

Наиболее влиятельной задачей распознавания лиц в этом десятилетии можно назвать задачу Pascal VOC Challenge. Новизна состоит в классификации, детектировании и сегментации. В задаче обнаружения, набор данных содержит описания объектов 20 различных классов, в то время как каждое описание включает в себя не только координаты ограничивающие области, но и кроме того определяет следующие характеристики: «ориентация», «окклюзию», «усеченность» и «трудоемкость».

Эти атрибуты определены для селективного тренинга и игнорируются во время оценки. В задаче вводится ряд новых методов оценки: Boostrapping, AP и ранг, и в дальнейшем они нормализуются по точности. В последнее время ImageNet challenge расширяет задачи Pascal VOC включая в них расширение категорий и количества изображений. Калифорнийский технологический институт широко использует набор данных пешеходов для обнаружения пешеходов. Учитывая особые случаи обнаружения пешеходов с точки зрения транспортного средства, Piotr разрабатывает стратегии и руководящие принципы в области сбора данных и их описании.

Другие наиболее часто используемые наборы данных для обнаружения лица: AFW и Fddb. AFW данных содержит 205 изображений, собранных с Flickr и 468 помеченных лица. Описание включают прямоугольную рамку, 6 атрибутов и углы позы. Fddb набор содержит 2845 изображений с 5171 лицами, каждое лицо помечается заранее определенным эллипсом вместо прямоугольной рамки. База данных изображений и описаний этих двух наборов данных разделены, и разработчики могут проводить оценку своих методов, а производительность ранжируется в соответствии с

определенными кривыми (Precision-Recall, AFW, ROC, FDDB) на всем множестве испытаний. В таблице №1 приведены коллекции изображений.

Таблица № 1

Коллекции изображений

Название набора изображений	Количество изображений	Количество лиц	Свойства	Описания
CMU/MIT	125	483	Серые полутона/ Фронтальные	6 ориентиров
CMU	208	441	Серые полутона/ Фронтальные /Профильные	6/9 ориентиров
AFW	205	468	Цветные, в природе	6 ориентиров / Граничная квадратная область / просмотр углов
FDDB	2846	5171	Серые полутона/ Цветные, в природе	Граничная эллиптическая область
MALF	5250	11931	Цветные, в природе	Граничная квадратная область / 5 Атрибутов

Сравнительный анализ методов и алгоритмов распознавания лиц

Необходимо отметить, что в существующих алгоритмах оценки обнаружений объектов, оценивающие по нескольким категориям, существуют некоторые ограничения и поэтому невозможно использовать обычный метод для оценки их производительности. Такая же проблема существует и при оценке качества обнаружения лица. Однако, используя алгоритм детализированной оценки можно оценить конкретный атрибут алгоритмов выделив один или два суб-набора оценок, которые, показывают

преимущества алгоритма в определенных аспектах. Алгоритм позволяет определять разные типы ошибок, при обнаружении объекта, и рассчитать количество баллов при обнаружении лица, однако данная оценка не показывает истинное значение, т.к. используются разные стратегии определения ограничивающего прямоугольника на разных наборах данных, а также из-за разницы масштабирования между аннотацией набора данных и результата на выходе детектора.

При оценке обнаружения лица в качестве метрик, наиболее часто используются два графика, Precision-Recall и ROC, а также широко используется два числовых показателя: среднее значение точности (Average Precision value) и площадь под кривой (Area Under Curve). В реальных приложениях не редко возникают проблемы связанные с быстродействием алгоритма и высоким уровне точности. Большинство из этих метрик оценивают скорость, а лишь потом оценивают точность.

В соответствии с методом, используемым в [2], построен график, который показан на рис.1 для истинно положительных значений FPPI. FPPI подходит для измерения точности обнаружения каждого образа на каждом изображении, а шкала показывает низкую производительность по FPPI.

Используя метод, описанный в работе [6], на графике отмечена True Positive Rate (далее TPR) точка кривой изображения в лог пространстве False Positive Per Image (далее FPPI). Показатель FPPI подходит для измерения точности обнаружения объекта каждого образа используемый логарифмическую шкалу учитывающий производительность при низких скоростях FPPI. Например, в системах видео-видеонаблюдения, где большое количество мишеней обнаружения в каждом изображении, TPR при 100 FPPI хорошо подходит для оценки алгоритма; в то время как в приложениях для мобильных устройств, для большей информативности необходимо использовать TPR при 10-2 FPPI. Для сравнения производительности между

различными алгоритмами, используется числовой показатель mean-miss rate (среднее время отклика), определенный в [6]. Время отклика рассчитывается, как среднее значение 9 истинно-положительных точек [7], равномерно выбранных между 10⁻² FPPI и 10⁻¹ FPPI в лог пространстве.

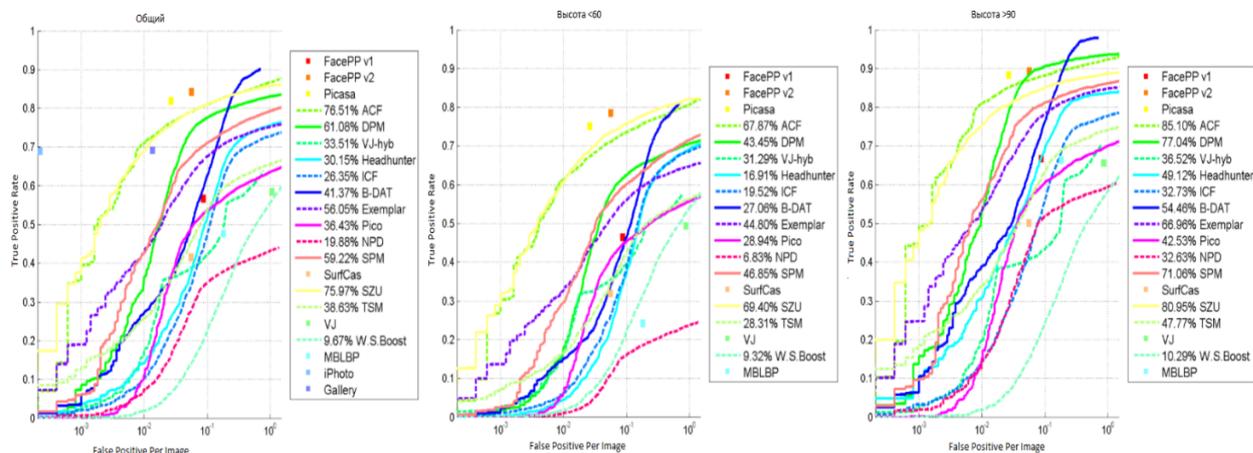


Рис. 1. – График истинно положительных значений

Для того, чтобы оценить качество и скорость использования MALF, рассмотрены алгоритмы распознавания лиц, представленные научными школами и коммерческими разработками, а также проанализированы результаты. Что касается коммерческого программного обеспечения (Google Picasa, Apple iPhoto и Windows Photo Gallery), они не предоставляют конкретные результаты вычисления ограничительной рамки. Вместо этого вручную рассчитаны истинно-позитивные и ложных срабатывания на всем тестовом наборе. Что же касается Face ++, использовался бесплатный онлайн API v1. При использовании академических алгоритмов использовались алгоритмы, представленные непосредственно авторами или свободно доступные исходные программы, с открытым исходным кодом [8]. Также собиралась информация обучающих данных и параметров реализации для всестороннего описания. Оцениваемые алгоритмы реализовывают разные подходы, рассмотрим их.

Алгоритмы DPM и B-DAT, используют деформируемую модель области изображения, которая является довольно популярным методом для

обнаружения объекта. Алгоритмы VJ-hyb, Pico, NPD, SurfCas используют метод Виолы – Джонса. Дополнительное использование функции каналов для улучшения обнаружения обрамляющей рамки искомого изображения методом Виолы – Джонса используют реализации ACF, Headhunter, ICF. FacePP v2 – использует хорошо зарекомендовавший себя метод сверхточной нейронной сети [9]. Обобщенные данные представлены в таблице №2.

Таблица №2

Обобщенные данные списков алгоритмов

Название	Автор	Разработчик	Данные для обучения	Источник	Параметры
iPhoto	н/а	Apple	н/а	iPhoto	Версия 9.6
FacePP_v1	н/а	Megvii	н/а	Онлайн API	Версия от 2014.10.08
FacePP_v2	н/а	Megvii	н/а	Submission	
Picasa	н/а	Google	н/а	Google Picasa	Версия 3.7
Gallery	н/а	Microsoft	н/а	Windows Photo Gallery	Версия 16.4.3528.331
ACF	B. Yang, J. Yan, Z. Lei	CBSR & NLPR, Chinese Academy of Sciences	БД AFLW	Предоставлено разработчиком	multi_scale feature, 6 views, imresize 3x
DPM	M. Mathias, R. Benenson M. Pedersoli	iMinds & MPI Informatics	AFLW, Pascal набор лиц	Открытый исходный код	threshold = -0.5 imresize 2x
Exemplar	H. Li, Z. Lin, J. Brandt, X. Shen, G. Hua	Stevens Institute of Tech. & Adobe Research	15, 832 изображений лиц 12, 732 изображений без лица	Предоставлено разработчиком	



VJ-hyb	Y. Gavini	VIT University	н/а	Предоставлено разработчиком	Viola-Jones based hybrid detector
Headhunter	M. Mathias, R. Benenson M. Pedersoli	iMinds & MPI Informatics	AFLW, Pascal набор лиц	Открытый исходный код Doppia	Headhunter model
ICF	M. Mathias, R. Benenson M. Pedersoli	iMinds & MPI Informatics	AFLW, Pascal набор лиц	Открытый исходный код Doppia	Headhunter-baseline model
B-DAT	J. Deng, J. Yang, D. Wang, S. Yan, G. Liu, Q. Liu	NUIST	AFLW, дополнительный набор изображений ~1 млн. лиц	Предоставлено разработчиком	Improved DPM, context, alignment, imresize 2x
Pico	N. Markus, M. Frljak, I. S. Pandzic, J. Ahlberg, R. Forchheimer	University of Zagreb	~20k фронтальных изображений лиц	Предоставлено разработчиком, Открытый исходный код GitHub	scale factor: 1.075, stride factor: 0.05
NPD	S. Liao, A.K. Jain, S.Z. Li	CBSR & NLPR, Chinese Academy of Sciences	FDDDB	Предоставлено разработчиком	

SPM	Ahmed EL-Barkouky, Ahmed Shalaby, Ali Mahmoud, Aly Farag	CVIP Lab, University of Louisville	Helen & Fddb	Предоставлено разработчиком	
SurfCas	J. Li, T. Wang, Y. Zhang	Intel Labs China	н/а	Открытый исходный код	model_type = 1, minsz = 8
SZU	S. Yu	Shenzhen University	н/а	Предоставлено разработчиком	
TSM	X. Zhu, D. Ramanan	University of California, Irvine	Multi-PIE	Открытый исходный код	face_p146_small, threshold = -2
VJ	P Viola, M.J. Jones	Microsoft Research, Redmond Mitsubishi Electric Research Laboratory	н/а	Открытый исходный код OpenCV	haarCascade models: frontalface-default & profileface
W.S.Boost	Z. Kalal, J. Matas, K. Mikolajczyk	University of Surrey Czech Technical University	н/а	Открытый исходный код	Модели: фронтальных и профильных изображений
MBLBP	L. Zhang, R. Chu	CBSR & NLPR, Chinese Academy of Sciences	н/а	Предоставлено разработчиком	

Выводы

Рассмотрим результаты оценки на всем тестовом наборе основываясь на данных графиков истинно положительных значений. Из рис.1, видно, что среди коммерческих систем, FacePP v2 и Picasa удается достичь хороших показателей с высоким уровнем TPR при низких FPPI. Iphoto показывает

средние результаты при очень низких FPPI. FacePP v1 и Gallery показывают относительно плохие результаты. Что касается академических алгоритмов, то трудно сказать, какой из них работает лучше. С точки зрения времени отклика, B-DAT обладает достаточно сильным преимуществом над другими. Тем не менее, он показывает значительно худшую производительность при меньших FPPI. С точки зрения общей производительности, ACF и SZU показывают лучшие результаты.

Необходимо обратить внимание, что при использовании не больших изображений лица в MALF размером 20x20 довольно сложная задача для многих алгоритмов. При этом коэффициент масштабирования сильно оказывает влияние на производительность [10]. Поэтому произведено выравнивание размеров изображений: “маленькие” к размеру, 60X60, а “большие” к размеру 90X90. Из кривых на графиках видно «рис.1», что на подмножестве “маленьких” изображений падение производительности алгоритмов на основе DPM (DPM и B-DAT) больше по сравнению с теми, что основаны на методе каналов. Рассматривая же подмножество с “большими” изображениями лица, ситуация имеет обратную тенденцию. Учитывая разницу между этими двумя подходами, можно сделать вывод, что при отсутствии влияния функции ухудшающей производительность вызванную низким разрешением, алгоритмы основанные на DPM показывают более высокую производительность.

Таким образом, при использовании детальной оценки, проанализирована эффективность алгоритма в различных аспектах в отношении нескольких атрибутов. Сравнивая производительность с различной степенью точности, можно определить преимущества и недостатки алгоритмов при использовании в различных сценариях работы. Использование MALF может так же служить ориентиром, для оценки качества обнаружения лица, и предоставляет дополнительную глубокую и

всестороннюю диагностику и рекомендации по улучшению алгоритмов распознавания образов.

Работа выполнена при финансовой поддержке РФФИ, проекты 16-01-00597-а, 17-07-00620-а.

Литература

1. Implementing the Viola-Jones Face Detection Algorithm // URL: etd.dtu.dk/thesis/223656/ep08_93.pdf (accessed: 17.07.17).
 2. C. Huang, «Vector boosting for rotation invariant multi-view face detection,» Tenth IEEE International Conference on Computer Vision, т. 1, pp. 446 - 453, 10 2005.
 3. V. J. a. E. G. Learned-Miller, «Fddb: A benchmark for face detection in unconstrained settings.,» UMass Amherst Technical Report, 2010., 11 p.
 4. C. Z. a. Z. Zhang, «A survey of recent advances in face detection.,» Microsoft Research, 2010., 50 p.
 5. P. W. P. R. a. H. B. M. Kostinger, «2011 IEEE International Conference,» Annotated facial landmarks in the wild. In Computer Vision Workshops (ICCV Workshops), 2011., pp. 2144 – 2151.
 6. C. W. B. S. P. P. P. Doll'ar, «Pedestrian detection,» PAMI, p. 34, 2012.
 7. Земцов А.Н. Анализ математических подходов к идентификации лиц // Инженерный вестник Дона, 2017, №3. URL: ivdon.ru/ru/magazine/archive/N3y2017/4258.
 8. P. Viola and M.J. Jones, «Rapid Object Detection using a Boosted Cascade of Simple Features», proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2001), 2001, 8 p.
 9. Пучков Е.В. Сравнительный анализ алгоритмов обучения искусственной нейронной сети // Инженерный вестник Дона, 2013, №4. URL: ivdon.ru/ru/magazine/archive/n4y2013/2135.
-

10. P. Viola and M.J. Jones, «Robust real-time face detection», International Journal of Computer Vision, vol. 57, no. 2, 2004., pp.137–154.

References

1. Implementing the Viola-Jones Face Detection Algorithm. URL: etd.dtu.dk/thesis/223656/ep08_93.pdf (accessed: 17.07.17).

2. C. Huang, «Vector boosting for rotation invariant multi-view face detection,» Tenth IEEE International Conference on Computer Vision, т. 1, pp. 446 - 453, 10 2005.

3. V. J. a. E. G. Learned-Miller, «Fddb: A benchmark for face detection in unconstrained settings.,» UMass Amherst Technical Report, 2010., 11 p.

4. C. Z. a. Z. Zhang, «A survey of recent advances in face detection.,» Microsoft Research, 2010, 50 p.

5. P. W. P. R. a. H. B. M. Kostinger, «2011 IEEE International Conference,» Annotated facial landmarks in the wild. In Computer Vision Workshops (ICCV Workshops), 2011., pp. 2144 – 2151.

6. C. W. B. S. P. P. P. Doll'ar, «Pedestrian detection:,» PAMI, p. 34, 2012.

7. Zemcov A.N. Inženernyj vestnik Dona (RUS), 2017, №3. URL: ivdon.ru/ru/magazine/archive/N3y2017/4258.

8. P. Viola and M.J. Jones, «Rapid Object Detection using a Boosted Cascade of Simple Features», proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2001), 2001., 8 p.

9. Puchkov E.V. Inženernyj vestnik Dona (RUS), 2013, №4. URL: ivdon.ru/ru/magazine/archive/n4y2013/2135.

10. P. Viola and M.J. Jones, «Robust real-time face detection», International Journal of Computer Vision, vol. 57, no. 2, 2004., pp.137–154.