

# Применение активного обучения к методу Виола-Джонса поиска объектов на изображении

Александр Жуковский  
Московский физико-  
технический институт  
(научно  
исследовательский  
университет)  
[zhukovsky@phystech.edu](mailto:zhukovsky@phystech.edu)

Дмитрий Николаев  
Институт проблем  
передачи информации  
имени  
А. А. Харкевича РАН  
[dimonstr@iitp.ru](mailto:dimonstr@iitp.ru)

Сергей Усилин  
Московский физико-  
технический институт  
(научно  
исследовательский  
университет)  
[usilin.sergey@gmail.com](mailto:usilin.sergey@gmail.com)

## Аннотация

*В работе описан алгоритм активного обучения классификатора Виола – Джонса, предназначенного для поиска объектов на изображениях. Для выбора примеров, подлежащих ручной разметке, предлагается критерий, опирающийся на сильное изменение степени уверенности по сравнению с предыдущей итерацией обучения.*

*Проведен эксперимент активного обучения классификатора для задачи поиска лиц на паспортах. Активное обучение классификатора проходило на малых обучающих выборках и сравнивалось с классификатором, построенным на полноценной выборке.*

## 1. Введение

В последнее время поиск и распознавание объектов на цифровых изображениях и в видеопотоке из стадии лабораторных исследований переросли в стадию использования в промышленных масштабах в основном в системах, где необходим интеллектуальный анализ данных. Например, почти в каждой системе видеонаблюдения используются методы поиска объектов для автоматического выделения важных видео-интервалов, так называемых «экстренных» ситуаций.

Современные методы распознавания практически без исключений опираются на механизмы обучения машин. Привлекательность этого подхода очевидна: результат может быть достигнут без привлечения специалистов в предметной области и без проведения долгосрочных работ.

На практике зачастую используются методы статистического обучения, позволяющие по заранее подготовленной и размеченной оператором выборке положительных и отрицательных примеров автоматически построить классификатор исследуемого объекта.

Однако на пути классических статических алгоритмов есть труднопреодолимые сложности. Дело в том, что точность работы такого классификатора сильно зависит от качества подготовленной тренировочной выборки. В идеальном случае обучающий набор должен быть репрезентативен относительно генеральной совокупности изображений, соответствующих искомому объекту. Следовательно, необходимо огромное количество размеченных оператором обучающих изображений. Чтобы достаточно плотно «заселить»  $n$ -мерное пространство необходимо иметь экспоненциальное от  $n$  число примеров. Фактически мы говорим о так называемом «проклятии размерности», прямом следствии статистической теории восстановления зависимостей по эмпирическим данным, разработанной В.Н. Вапником и А.Я. Червоненкисом в конце 60-х – начале 70-х [1].

Построение генеральной выборки в общем случае является очень сложной, даже нереальной задачей. Однако она вполне решается в некотором частном случае, который представляет собой точно поставленную задачу распознавания. Для каждой четко поставленной задачи поиска объекта требуется адаптированная именно под нее обучающая выборка, состоящая только из частей изображения, на котором и проводится поиск. Например, при постановке задачи поиска лиц на паспортах, обучающая выборка будет содержать в себе всевозможные фрагменты паспорта. Ценность не относящегося к поставленной задаче примера для обучения будет либо ничтожно мала, либо вызовет некоторую неточность в классификации.

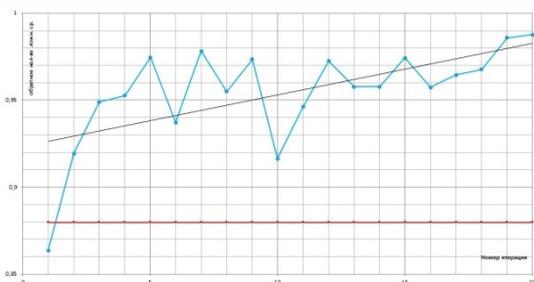


Рисунок 1. Доля правильно распознанных примеров (красная линия – каскад, построенный на полноценной выборке, синяя – каскады, полученные в ходе активного обучения)

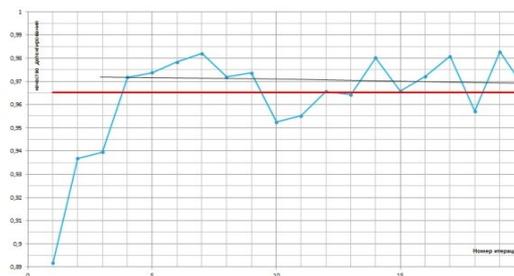


Рисунок 2. Качество распознавания (красная линия – каскад, построенный на полноценной выборке, синяя – каскады, полученные в ходе активного обучения)

Каждый пример, в свою очередь, может предоставлять различное количество информации об особенностях изображения. Разумно составлять выборки только из примеров, сложных для классификатора, а значит предоставляющих наибольшее количество информации и имеющий большее значение. Однако выделение сложных обучающих образцов, предоставляющих максимум информации для обучения, является крайне трудной задачей для оператора.

Для решения части ограничений можно, основываясь на методе Блума и Митчела [2], использовать дополнительный классификатор, обучаемый параллельно и независимо от основного. Данный метод, примененный к поиску автомобилей в видеопотоке, описан Левином, Виолой и Фройндом [3]. Основным улучшением классификатора является уменьшение количества ложных срабатываний. Однако для серых изображений оно уменьшается всего в 1.5 раза и метод целесообразно использовать только для видеоряда с вычтенным фоном, где количество ложных срабатываний уменьшается в 5.5 раз.

Также для решения ограничений статистического обучения предлагается использовать активное обучение. При первоначальном обучении классификатора появляется возможность определить сложные для него примеры – те, которые он проклассифицировал неуверенно. Эти примеры размечаются оператором и добавляются к начальной выборке. Таким образом, выборка расширяется сложными для обучения и адаптированными под задачу примерами. После переобучения классификатора на обновленной выборке можно ожидать качественное улучшение классификатора.

Похожий метод был предложен в статье Абрамсона и Фройнда [4], где он был применен к поиску пешеходов в видеопотоке. Критерий неуверенности классификатора, описанный в

статье, состоит в попадании значения, с которым распознал данный объект с незамеченной выборки в окрестность границы между классами. Авторами были использованы более слабые нежели Хаар-подобные простые классификаторы Абрамсона-Стюкса [5], разделяющие зоны рассматриваемых пикселей, и более сложные объекты для поиска. В результате работы получен процент правильных положительных срабатываний, составляющий ~70%, при довольно малой доле неправильных положительных срабатываний.

В данной работе для активного обучения предлагается использовать только образцы, величина уверенности распознавания которых сильно изменяется от раза к разу. В отличие от разбора только граничных примеров, данный подход позволяет получить сведения об образцах, резко меняющих (из-за возмущения в начальной выборке) класс, к которому они отнесены классификатором, не попадая в граничную зону.

## 2. Алгоритм Виола-Джонса

Алгоритм был представлен Виолой и Джонсом в 2003 году [6]. Он показал возможность поиска объектов в режиме реального времени. На текущий момент он является, пожалуй, самым широко распространенным методом для поиска объектов на изображениях и в видеопотоке. Первоначально была показана применимость алгоритма к поиску лиц на изображениях, но алгоритм зарекомендовал себя и для поиска объектов с жесткой геометрией.

Алгоритм сводит задачу детектирования к задаче классификации каждого региона изображения: для каждого окна изображения, взятого со всевозможными сдвигами, ориентацией и масштабом проверяется при помощи заранее обученного классификатора гипотеза о его принадлежности целевому классу.

Таблица 1. Динамика роста выборок и результаты активного обучения

Итер.	Полож. выборка	Отриц. выборка	Кол-во срабат.	Кол-во пропусков	Качество детект.	Кол-во ложн. сраб.	Доля пол. ср.	Ф-ция стоим.
max	-	-	-	0	1,00	0	1,00	1,00
Полн.	269	281	490	3	0,97	69	0,86	0,97
1	25	15	264	12	0,89	36	0,86	0,89
2	38	30	273	7	0,94	22	0,92	0,94
3	54	36	412	7	0,94	21	0,95	0,94
4	61	42	570	3	0,97	27	0,95	0,97
5	76	49	592	3	0,97	<b>15</b>	0,97	0,97
6	90	54	574	<b>2</b>	<b>0,98</b>	36	0,94	<b>0,98</b>
7	97	60	646	<b>2</b>	<b>0,98</b>	<b>14</b>	<b>0,98</b>	<b>0,98</b>
8	110	67	666	3	0,97	30	0,95	0,97
9	117	73	642	3	0,97	17	0,97	0,97
10	119	80	753	5	0,96	63	0,92	0,95
11	131	89	744	5	0,96	40	0,95	0,96
12	144	96	656	4	0,96	18	0,97	0,97
13	152	102	732	4	0,96	31	0,96	0,96
14	160	110	757	<b>2</b>	<b>0,98</b>	32	0,96	0,98
15	171	117	778	4	0,96	20	0,97	0,97
16	178	125	776	3	0,97	33	0,96	0,97
17	190	135	791	<b>2</b>	<b>0,98</b>	28	0,96	<b>0,98</b>
18	200	141	742	5	0,96	24	0,97	0,96
19	214	147	706	<b>2</b>	<b>0,98</b>	<b>10</b>	<b>0,99</b>	<b>0,98</b>
20	225	152	727	4	0,96	<b>9</b>	<b>0,99</b>	0,97

Детектором является усиливающаяся по сложности последовательность (каскад) сильных классификаторов. Сильный классификатор строится алгоритмом машинного обучения AdaBoost [7] как линейная комбинация слабых классификаторов. Слабый классификатор основывается на величину хаар-подобного признака. Используемые в работе хаар-подобные признаки представляют собой два смежных прямоугольника одинакового размера на изображении, расположенные вертикально или горизонтально. Величину признака составляет разница между суммами значений пикселей содержащихся в этих прямоугольниках.

## 2.1 Автоматическое обучение каскада классификаторов

Для обучения каскада «сильных классификаторов» необходимо 3 выборки: положительная, отрицательная и тренировочная. Первые две выборки используются для обучения очередного сильного классификатора (стадии каскада). Последняя (тренировочная) выборка состоит из изображений частей паспортов, не содержащих лица, и используется для генерации отрицательной выборки для обучения следующего сильного классификатора. После обучения одного сильного классификатора, он добавляется к каскаду и проводится поиск

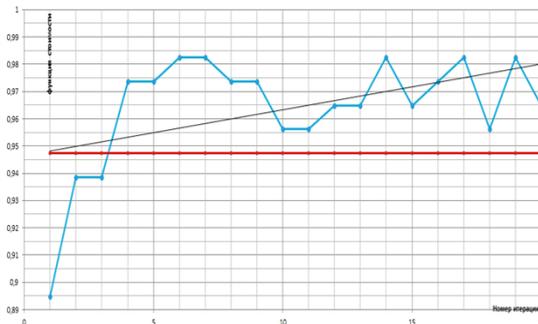


Рисунок 3. Функция стоимости (красная линия – каскад, построенный на полноценной выборке, синяя – каскады, полученные в ходе активного обучения)

объектов на тренировочной выборке. Если детектор ошибся и нашел объекты, то обучается следующая стадия каскада. Для нее положительная и тренировочная выборки остаются теми же, а отрицательную составляют все срабатывания на тренировочной выборке.

## 2.2 Активное обучение детектора

Активное обучение детектора представляет собой полуавтоматическое интерактивное обучение.

При поиске объектов на неразмеченной выборке часто встречаются образцы, сложные для классификации. Идея активного обучения состоит в том, чтобы предложить эти образцы для классификации оператору, расширить ими обучающие выборки и переобучить следующий классификатор на новых выборках. После нескольких таких итераций обучения, ожидается увидеть классификатор, построенный на относительно малой обучающей выборке и превосходящий по свойствам классификатор, однократно обученный на большой выборке.

Активное обучение основывается на величинах распознавания окон. При поиске объектов, классификатор ставит в соответствие каждому участку изображения величину уверенности, то есть число, показывающее насколько данный участок похож на искомый объект. Для разделения классов используется пороговое значение: если величина распознавания больше некоторого значения, то участок относится к целевому классу, иначе – нет. Для разных классификаторов величина для одного окна может принимать разные значения. И если, в ходе активного обучения данное значение сильно изменилось по сравнению с предыдущим классификатором (больше, чем на заданное максимальное число), то

окно считается распознанным неуверенно. Изменение могло произойти из-за обнаружения обучающим алгоритмом нового общего свойства у элементов обучающей выборки.

Таким образом, под внимание оператора попадают примеры, сильно меняющие величину своей уверенности перескакивая из одного класса в другой, минуя пограничную зону. Кроме того, в случае нескольких кластеров объектов на неразмеченной выборке, становится возможным обнаружение не вошедшего в обучающую выборку подмножества.

Активное обучение заканчивается, когда количество сложных примеров становится меньше некоторой заранее выбранной величины или их не становится совсем. Окончательно, описанный метод активного обучения формально представим следующим образом:

1. Задаются параметры, требуемые для обучения классификатора
2. задается  $M$  – максимально допустимое значение изменения величины уверенности
3. Размечаются начальные выборки:
  1.  $P_0$  – положительная обучающая выборка
  2.  $N_0$  – отрицательная обучающая выборка
  3.  $T$  – тренировочная выборка
  4.  $V$  – неразмеченная выборка
4. Обучается классификатор  $H_0$  на  $P_0$ , части  $N_0$  и  $T$
5. Считается величина уверенности  $H_0(v_j)$  для всех элементов из  $V$
6.  $P_1 = P_0; N_1 = N_0$
7.  $i = 0$
8. **делать:**
  1.  $++i$
  2. Обучается классификатор  $H_i$  на  $P_i, N_i$  и  $T$
  3. Считается величина уверенности  $H_i(v)$  для всех элементов из  $V$
  4.  $R_i = \{v \mid H_i(v) - H_{i-1}(v) \geq M, \forall v \in V\}$  – предназначенные для рецензирования пользователем элементы
  5. Если  $R_i \neq \emptyset$ :
    1.  $N_{i+1} = N_i \cup \{v \mid op(v) = 0, \forall v \in R_i\}$
    2.  $P_{i+1} = P_i \cup \{v \mid op(v) = 1, \forall v \in R_i\}$
 где  $op(v): V \rightarrow \{0,1\}$  – рецензия оператора
9. **пока**  $R_i \neq \emptyset$

## 3. Эксперимент

В ходе эксперимента проводилось обучение и поиск лиц на паспортах. Для оценки полученных результатов активного обучения, проводилось сравнение с результатами, полученными при использовании классификатора, обученного на

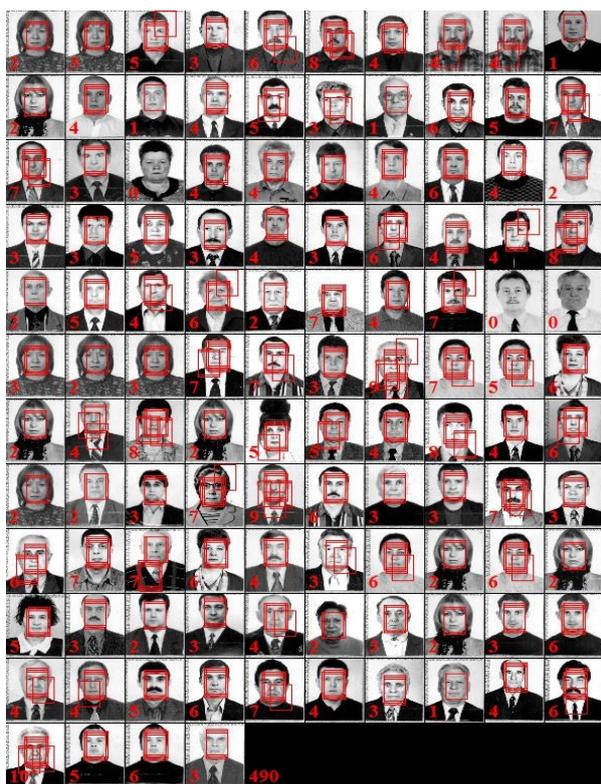


Рисунок 4. Результаты поиска детектором, обученным на полной выборке.

полноценной выборке. Для обучения полноценного классификатора использовалась выборка, состоящая из порядка 300 положительных примеров и 300 отрицательных. Тренировочная выборка состояла из 25 страниц паспортов, не содержащих лиц. Тестовая выборка содержала 114 изображений страниц паспортов, содержащих по одному лицу. Окончательный каскад состоял из 14 сильных классификаторов, с последним, состоящим из 140 слабых. Начальные положительная и отрицательная выборки для активно обучаемого каскада составляли соответственно 25 и 30 примеров. Обучаемый каскад содержал 11 сильных классификаторов. Проводилось 20 итераций обучения классификатора и разметки участков, изменивших свое значение более чем на  $\frac{1}{2}$ . Количество размечаемых оператором примеров (включая не проклассифицированные оператором) в ходе одной итерации было порядка 30.

#### 4. Результаты эксперимента

В ходе активного обучения использовались несколько оценок работы классификаторов.



Рисунок 5. Результаты поиска детектором, полученным на 19 итерации активного обучения

Первая состояла в оценке доли ложных положительных срабатываний – когда классификатор обнаруживает объект там, где его нет на самом деле. Для ее оценки использовалась дополняющая до единицы величина ( $p$ ), равная доле правильных срабатываний. Как показывает рисунок 1, график хотя и довольно дисперсен, однако имеет тенденцию к увеличению.

Следующей величиной было качество распознавания ( $q$ ), показывающее, какую часть объектов классификатор нашел на неразмеченной выборке (рисунок 2). Качество распознавания также довольно дисперсно, но после нескольких стадий выходит на постоянный уровень, лежащий немного выше уровня детектора, построенного на большой выборке.

Но стоимость пропуска объекта и срабатывания на ложном примере не равноценны. В промышленных системах ложное срабатывание может отсечь на следующих классификаторах, а пропущенный объект уже не восстановить. Кроме того, при разметке, отрицательные примеры генерируются автоматически, а положительные – вручную. Все это увеличивает стоимость пропуска объекта. Поэтому, для окончательной оценки эффективности активного обучения, вводится функцию стоимости разметки:

$$v(x) = \frac{c_1 * q(x) + c_2 * p(x)}{c_1 + c_2} \quad (3)$$

Выбор констант для функции стоимости зависит от системы разметки примеров и желаемого качества распознавания. Рассмотрим пример, когда при разметке одного изображения оператор обозначает все местоположения положительных примеров, а отрицательными считаются все остальные участки, не пересекающиеся с заданной окрестностью положительных. В таком случае, положительные примеры размечаются вручную, а отрицательные - автоматически, причем их число значительно превышает количество положительных. Следовательно, и стоимость разметки положительного примера намного больше чем отрицательного. Учитывая желание получить классификатор, не пропускающий положительные примеры, можно заключить, что вес пропуска положительного примера на несколько порядков превышает вес ложного срабатывания. Для рассмотренного случая график функции стоимости с константами  $c_1 = 100$ ,  $c_2 = 1$  показан на рисунке 3.

Пример результатов распознавания для каскада, обученного на полной выборке, показан на рисунке 4, для 19-ой итерации активного обучения, показавшей лучшие результаты, - на рисунке 5.

В таблицу 1 сведены динамика роста обучающих выборок в ходе активного обучения, общее количество срабатываний, количество пропусков положительных примеров и посчитанное по нему качество детектирования  $q$ , количество ложных срабатываний и посчитанная по нему доля положительных срабатываний  $p$  и окончательная функция стоимости  $v$ .

Максимально достигнутый результат в ходе активного обучения для качества распознавания составляет 98,2% (на 6, 7, 14 и 19 итерациях) против 97,4% у эталонного каскада. Причем, обучающие выборки для построения классификатора существенно меньше (у 7-й стадии: 157 против 550). Для доли положительных срабатываний максимум у 19-й и 20-й стадий (98,5% и 98,7% соответственно) против 86% у каскада, обученного на полной выборке.

Таким образом, классификаторы, построенные при активном обучении на малых выборках, показывают лучший результат, чем однократно обученный на большой выборке классификатор.

## 5. Обсуждение

Прежде всего, отметим, что не имеет значения увеличение сложности процедуры обучения детектора, так как она является единичной операцией. Гораздо важнее построить качественный классификатор и получить выборки, на которых он построен для последующего улучшения.

Следует также заметить, что построенные в ходе активного обучения классификаторы не подвержены переобучению, так как не все изображения тестовой выборки были рецензированы.

При активном обучении, оператору все же предоставляется для разметки большое количество примеров. Есть несколько техник для автоматического отсева заведомо негативных примеров. К примеру, для фронтальных лиц, можно воспользоваться свойством их симметричности относительно построенного классификатора. Тогда при поиске двумя классификаторами - прямым и зеркально отображенным на всех лицах он даст утвердительный результат, а на несимметричных отрицательных примерах, прошедших один из них, второй даст отрицательный ответ.

Также возникает ситуация, когда на изображении есть один положительный объект и классификатор неуверенно распознал несколько близких по расположению к нему зон на объекте, которые оператор пометил как положительные. Тогда, для того чтобы улучшить качество положительной выборки из этих объектов выбирается один, который является «центром масс» всех положительных примеров на изображении и только он добавляется к положительной выборке. В случае, когда объекты имеют разный размер, таким способом из них выбирается «оптимальный».

Выявление сложных примеров может проходить не только по одному критерию. Вполне возможно использование как предложенного критерия для рецензии, предложенного Абрамсоном и Фройндом [4] критерия, так и любого другого.

Метод может найти применение в полуавтоматическом поиске объектов. Например, подключение метода к интерфейсу, позволяющему производить разметку положения объектов на рассматриваемых изображениях, окруженную некоторой окрестностью, за пределами которой не находится ни одного объекта позволит производить разметку только на изображениях, плохо проклассифицированных одним из

классификаторов. Последовательное обучение, как ожидается, приведет к построению классификатора, качественно решающего поставленную задачу. Еще, показ оператору до рецензирования результатов поиска построенным классификатором на размечаемой выборке и разрешение выбирать оператору изображения выборки для рецензирования позволит сконцентрировать внимание обучающего алгоритма на сложных изображениях.

## 6. Заключение

В работе описан алгоритм активного обучения статистического классификатора Виола-Джонса. Активное обучение может быть использовано для улучшения детекторов за счет расширения обучающих выборок сложными примерами из неразмеченной зоны. При построении детектора за счет усложнения процедуры возможно существенное уменьшение начальных выборок. Кроме того, алгоритм позволяет построить лучший классификатор на меньшем количестве обучающих примеров.

## 7. Список литературы

- [1] Вапник В.Н., Червоненкис А.Я. *Теория распознавания образов*, М.: Наука, 1974
- [2] A. Blum and T. Mitchell, *Combining labeled and unlabeled data with co-training*, Workshop on Computational Learning Theory, 1998
- [3] A. Levin, P. Viola, and Y. Freund, *Unsupervised improvement of visual detectors using co-training*, In International Conference on Computer Vision, 2003
- [4] Y. Abramson and Y. Freund, *Active learning for visual object recognition*, Technical report, UCSD, 2004
- [5] Y. Abramson B. Steux, *YEF Real-Time Object Detection*, IJISTA, 2007
- [6] P. Viola and M. Jones, *Robust Real-time Object Detection*, International Journal of Computer Vision, 2001
- [7] Yoav Freund, Robert E. Schapire *A decision-theoretic generalization of on-line learning and an application to boosting*, Journal of Computer and System Sciences, no. 55, 1997
- [8] Ashish Kapoor, Gang Huaz, Amir Akbarzadeh, Simon Baker, *Which Faces to Tag: Adding Prior Constraints into Active Learning*, International Conference on Computer Vision, 2009
- [9] K. Levi and Y. Weiss, *Learning Object Detection from a Small Number of Examples: The*

- Importance of Good Features*, Computer Vision and Pattern Recognition Workshop, 2004
- [10] Peter Roth, Helmut Grabner, Danijel Skocaj, Horst Bischof and Ales Leonardis, *Conservative Visual Learning for Object Detection with Minimal Hand Labeling Effort*, Proc. of DAGM, 2005
- [11] Robin Hewitt, Serge Belongie, *Active Learning in Face Recognition: Using Tracking to Build a Face Model*, Computer Vision and Pattern Recognition Workshop, 2006
- [12] Thuy Thi Nguyen, Nguyen Dang Binh, and Horst Bischof, *Efficient Boosting-Based Active Learning for Specific Object Detection Problems*, International Journal of Electrical and Electronics Engineering 3:3, 2009
- [13] Сергей Усилин, Дмитрий Николаев, Василий Постников, *Идентификация изображений документов методами, предназначенными для поиска лиц*, ИТИС, 2010