Сравнение некоторых методов решения задачи детектирования лиц на изображениях

Е.А. Долотов[1], В.Д. Кустикова[1]
dolotov.evgeniy@gmail.com|valentina.kustikova@itmm.unn.ru
[1] Институт информационных технологий, математики и механики
Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, Россия

Рассматривается задача детектирования лиц на изображениях. Проводится сравнение некоторых методов решения задачи, основанных на построении сверточных нейронных сетей. Сравниваемые методы относятся к числу передовых, поскольку демонстрируют одни из лучших результатов на широко известном наборе данных Face Detection Data Set and Benchmark (FDDB). Разрабатываются программные реализации методов, воспроизводятся результаты детектирования лиц на FDDB. Эксперименты показывают сравнимые значения показателей качества с опубликованными на официальной странице FDDB. Оценивается качество детектирования лиц на более сложных данных, входящих в состав набора WIDER FACE. Проводится анализ полученных результатов.

Ключевые слова: *детектирование лиц, глубокое обучение, сверточные нейронные сети, Face Detection Data Set and Benchmark (FDDB), WIDER FACE.*

Comparison of some methods for solving the problem of face detection in images

E.A. Dolotov[1], V.D. Kustikova[1]
dolotov.evgeniy@gmail.com|valentina.kustikova@itmm.unn.ru
[1] Institute of Information Technologies, Mathematics and Mechanics,
Lobachevsky State University of Nizhni Novgorod, Nizhni Novgorod, Russia

We consider the problem of face detection in images. We make a comparison of some methods, based on convolutional neural networks. The compared methods are among the advanced, as they demonstrate the best results on the widely known Face Detection Data Set and Benchmark (FDDB) data set. We develop a software implementation of these methods and reproduce the results of face detection on FDDB. The experiments show comparable values of the quality measurements with those published on the official FDDB web page. We assess and analyse the quality of face detection on more complex data included in the WIDER FACE dataset.

Keywords: *face detection, deep learning, convolutional neural networks, Face Detection Data Set and Benchmark (FDDB), WIDER FACE.*

1. Введение

Задача детектирования лиц на изображениях является одной из классических задач компьютерного зрения. Несмотря на простоту формулировки, задача является достаточно сложной в силу изменчивости визуального образа лиц. Изменчивость обусловлена вариативностью внешнего вида, масштаба и ракурса объекта, а также степенью его освещенности.

Цель настоящей работы состоит в том, чтобы оценить качество детектирования лиц с использованием известных методов [9, 12, 15, 18], основанных на обучении сверточных нейронных сетей, на наборе данных WIDER FACE [17], который содержит разнообразные изображения лиц, полученные в реальных условиях.

Работа построена следующим образом. Вначале дается общая схема работы каждого из рассматриваемых методов детектирования лиц. Приводится краткое описание программных реализаций. Выполняется сравнение качества на наборе данных Face Detection Data Set and Benchmark (FDDB) [3]. Оценивается качество детектирования лиц на более сложных данных, входящих в состав набора WIDER FACE [17]. Проводится анализ полученных результатов.

2. Задача детектирования лиц

Задача детектирования лиц состоит в том, чтобы определить положение всех лиц людей на изображении. Положение определяется прямоугольником, окаймляющим границы лица.

3. Сравниваемые методы

Существует большое количество различных методов детектирования лиц [1, 4, 8, 14, 20]. В настоящее время наилучшее качество показывают алгоритмы, основанные на применении сверточных нейронных сетей (Convolutional Neural Network, CNN). Различия алгоритмов данной группы, как правило, обнаруживаются в архитектурах сетей и параметрах их обучения [9, 14]. Наряду с этим, могут отличаться методы обработки признаков [4, 12], а также подходы к решению проблемы поиска лиц разного масштаба [8].

Множество методов, основанных на применении сверточных нейронных сетей, можно разделить на две группы:

− Методы, использующие выход предварительно обученной сверточной сети в качестве признакового описания изображения [12]. Построенное признаковое описание далее используется для обучения традиционных классификаторов.

− Методы, предоставляющие на выходе нейронной сети полную информацию о расположении лиц, областей или окаймляющих лица прямоугольников [9].

Рассмотрим более подробно четыре метода детектирования лиц, основанные на применении сверточных нейросетей и демонстрирующие одни из лучших результатов на широко известном наборе данных Face Detection Data Set and Benchmark (FDDB) [3].

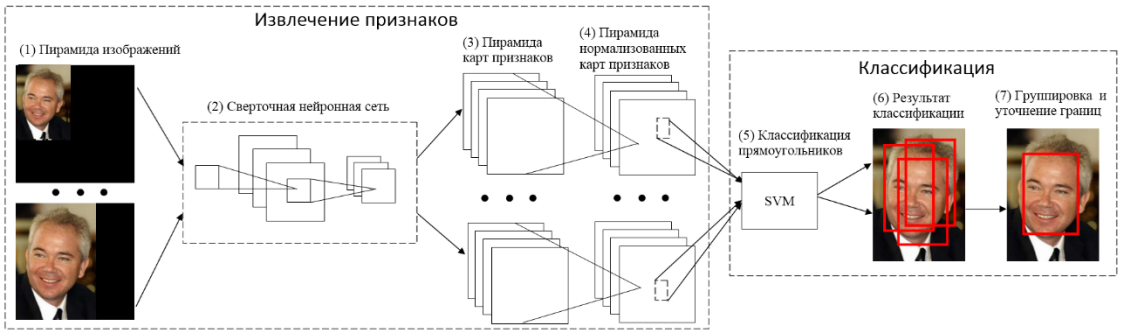

Рис. 1. Схема работы метода, основанного на построении модели деформируемых частей.

3.1 Метод, основанный на построении модели деформируемых частей

В алгоритме детектирования лиц [12] (A Deep Pyramid Deformable Part Model for Face Detection, DPMFD) для исходного изображения строится пирамида изображений, состоящая из разных масштабов одного изображения (рис. 1). Затем изображение с каждого уровня подается на вход сверточной нейронной сети, которая получена из сети AlexNet [10] посредством удаления последних полносвязных слоев. После обработки пирамиды изображений формируется соответствующая пирамида карт признаков, состоящая из такого же количества уровней, как и пирамида изображений. Каждый вектор признаков пирамиды признаков нормализуется. Далее осуществляется проход скользящим окном с единичным шагом по нормализованным картам признаков, извлекаются все прямоугольные области фиксированного размера. Полученные карты признаков преобразуются в вектора, которые затем классифицируются с помощью обученной линейной машины опорных векторов. При этом расстояние до гиперплоскости, разделяющей два класса, принимается за достоверность принадлежности к классу лиц. После этого этапа каждому лицу на изображении может соответствовать несколько прямоугольников. Поэтому необходимо объединить такие прямоугольники, а затем уточнить их границы, используя алгоритм, описанный в [5].

3.2 Метод, основанный на полностью сверточной сети, с малым количеством весов

В большинстве алгоритмов детектирования лиц применяются очень глубокие сверточные нейронные сети, которые имеют большое количество весов (AlexNet [10], VGG-16 [13], ResNet [6]). Метод [15] (FastCNN), описанный в данном разделе, доказывает, что эту задачу можно решить с помощью сверточной нейронной сети со значительно меньшим количеством слоев и весов.

Сначала обучается сеть, решающая задачу бинарной классификации «лицо/не лицо». Сеть принимает на вход трехканальное изображение в формате RGB с разрешением 32×32 пикселя. Затем изображение последовательно обрабатывается с помощью семи сверточных слоев. После каждого сверточного слоя применяется активационная функция PReLu [7], которая является кусочно-линейной, что позволяет делать нелинейные преобразования с выходом предыдущего слоя и в тоже время быстро вычислять производную. Выход сети после каждого сверточного слоя имеет меньший размер, это свойство позволяет решать задачу классификации без применения слоев другого типа. Выходом последнего сверточного слоя является пара чисел, каждое из которых определяет достоверность принадлежности к одному из двух классов. Сеть с такой архитектурой имеет примерно в 800 раз меньше весов, чем сеть AlexNet. Такое значительное уменьшение количества весов позволяет намного сократить время работы детектора, а также сократить время, необходимое для обучения сети.

Так как описанная сеть является полностью сверточной, то она может быть применена к изображению произвольного размера. В результате всех вычислений на выходе сети получаются две матрицы, которые описывают достоверности принадлежности конкретной области изображения к одному из классов. Для уменьшения количества ложных срабатываний детектора каждая из этих матриц усредняется с окрестностью 3×3. После чего по любой из этих двух матриц можно определить размеры и положения окаймляющих прямоугольников.

3.3 Метод, основанный на применении нестандартной функции потерь

При обучении нейронной сети для вычисления ошибки, как правило, используется метрика L_2 – евклидово расстояние между вектором, полученным на выходе нейронной сети, и вектором из обучающего набора данных. Этот подход имеет несколько недостатков. Во-первых, при использовании такой функции ошибки параметры оптимизируются независимо. Во-вторых, функция не является нормированной. Это приводит к тому, что ошибка на один пиксель на прямоугольнике большого размера обеспечивает такой же вклад, как и ошибка на прямоугольниках меньшего размера. В результате детектор обращает больше внимания на крупные объекты и игнорирует мелкие. Чтобы исправить эту ситуацию, необходимо использовать пирамиду из разных масштабов изображения, что негативно сказывается на времени работы детектора.

Авторы метода [18] (UnitBox), описанного в настоящем разделе, вводят новую функцию потерь, которая лишена этих недостатков. Данный подход основан на том, что для каждого пикселя изображения окаймляющий прямоугольник может быть описан с помощью четырехмерного вектора, где каждая компонента представляет собой расстояние от пикселя до верхней, нижней, левой и правой границ прямоугольника. Ошибку определения параметров прямоугольника предлагается вычислять с помощью метрики Intersection over Union (IoU), равной отношению площади пересечения прямоугольника, полученного в результате детектирования, и прямоугольника из разметки к площади их объединения. Использование такой функции ошибки позволяет оптимизировать все параметры прямоугольника одновременно. Также, несмотря на величину прямоугольника, значение величины IoU лежит в отрезке [0, 1], что позволяет обучать нейронную сеть на объектах разного масштаба и выполнять детектирование объектов на изображении без дополнительного масштабирования.

Основой метода детектирования является широко известная глубокая сверточная нейронная сеть VGG-16 [13], из которой удалены все полносвязные слои и добавлены две новые ветки (рис. 2). Первая ветка служит для получения бинарной маски изображения

(одноканальное изображение, состоящие из нулей и единиц: 0 – фон, 1 – лицо), а вторая ветка отвечает за построение границ окаймляющих прямоугольников, которые представляются четырехмерными векторами.

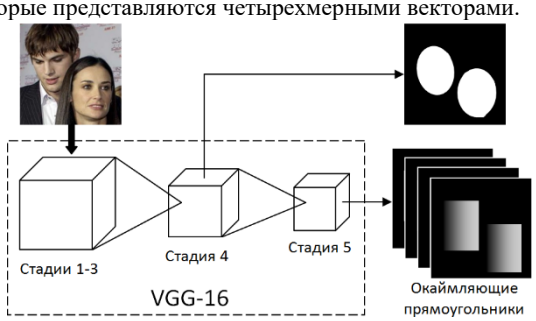

Рис. 2. Схема сверточной нейросети в методе UnitBox.

Для получения бинарной маски с помощью нейронной сети VGG-16 в конец четвертой стадии сети добавляются три слоя. Первые два – это сверточный слой с шагом в один пиксель и размером ядра 512×3×3×1 и слой линейной интерполяции для масштабирования выхода сверточного слоя до размера изображения, поступающего на вход сети. Третий слой обрезает изображение, полученное после линейной интерполяции, до размеров оригинального изображения. На основании полученного одноканального изображения вычисляется сигмоидальная кросс-энтропия, характеризующая ошибку построения бинарной маски. Для получения окаймляющих прямоугольников добавляются четыре дополнительных слоя, среди которых сверточный слой с шагом 1 и размером ядра 512×3×3×4, слой линейной интерполяции и слой, позволяющий обрезать выход после линейной интерполяции до размеров исходного изображения. На четвертом слое применяется функция ReLU, чтобы сделать выход сети неотрицательным. Затем вычисляется функция ошибки IoU. Общая ошибка сети является взвешенной суммой ошибок на обеих ветках.

К выходу первой ветки сети применяется пороговая функция для получения бинарной маски. На полученной бинарной маске находятся все компоненты связности. Из второй ветки сети извлекаются четырехмерные вектора, находящиеся на той же позиции, что и центры компонент связности. Четырехмерные вектора преобразуются в результирующие окаймляющие прямоугольники.

3.4 Метод, основанный на Faster R-CNN

Алгоритм [9] состоит из двух модулей: полностью сверточная нейронная сеть, которая используется для извлечения признаков из изображения и нахождения предполагаемых прямоугольников, и детектор на основе R-CNN [5], который использует эти прямоугольники. Указанные модули объединяются в одну нейронную сеть (Faster R-CNN, рис. 3).

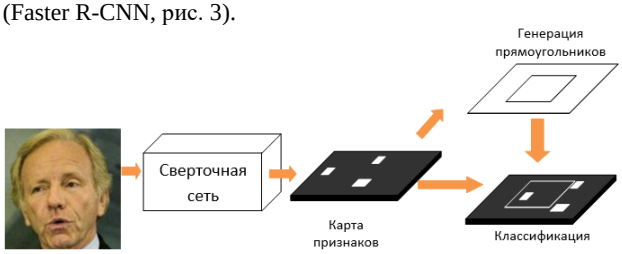

Рис. 3. Схема работы Faster R-CNN.

На первом этапе изображение произвольного размера подается на вход нейронной сети, которая обеспечивает выделение прямоугольников для дальнейшей классификации (Region Proposal Network, RPN). Эта сеть является полностью сверточной нейронной сетью, которая получена из сети VGG-16 путем удаления последних полностью связных слоев и добавлением одного или нескольких полностью сверточных слоев. Добавленные сверточные слои используются непосредственно для определения предполагаемых прямоугольников. Добавленные слои обходятся скользящим окном по карте признаков, полученной из сети VGG-16, и для каждой позиции окна извлекается вектор признаков малой размерности. Вектора признаков подаются на вход двум новым полностью связным слоям. Один из этих слоев используется для уточнения границ прямоугольника, а другой для классификации объекта, расположенного внутри этого прямоугольника.

В каждой позиции скользящего окна одновременно могут рассматриваться несколько прямоугольников, которые обладают разным размером или разным соотношением сторон. С помощью такого подхода на изображении одного масштаба можно найти лица разной величины.

После того как предполагаемые прямоугольники, содержащие лица, получены, их необходимо классифицировать. Обычно для этого используется другая сеть, которая обучается отдельно. Но в данном случае предлагается использовать для классификации часть слоев из сети RPN. Это позволяет значительно сократить время детектирования, а также упрощает процесс обучения модели, так как модель представляет собой единую нейросеть, которую можно обучать с помощью метода обратного распространения ошибки (Back Propagation). После применения сети на выходе получается набор прямоугольников. Каждому лицу на изображении могут соответствовать несколько различных прямоугольников. Поэтому завершающим этапом детектирования является объединение прямоугольников.

4. Программная реализация

Программная реализация описанных методов [20] выполнена с использованием языка программирования C++ на базе библиотек OpenCV и Caffe (рис. 4).

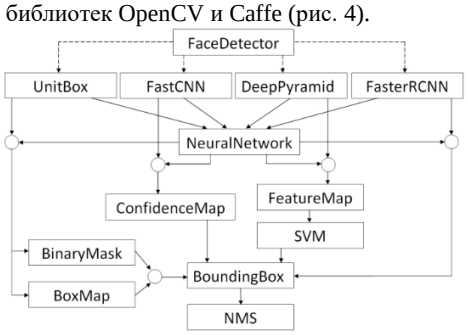

Рис. 4. Схема компонент системы.

1. *FaceDetector*. Базовый класс детектора лиц. *UnitBox*, *FastCNN*, *DeepPyramid*, *FasterRCNN* – наследники класса *FaceDetector*, каждый реализует соответствующий метод.
2. *NeuralNetwork*. Модуль для работы с нейронными сетями на базе библиотеки глубокого обучения Caffe [2].
3. *SVM*. Реализует метод скользящего окна и обеспечивает классификацию областей, накрываемых окном, с помощью метода опорных векторов. Использует реализацию метода из библиотеки компьютерного зрения OpenCV [11].
4. *BoxMap*, *FeatureMap*, *BinaryMask*. Содержат примитивы для представления окаймляющих прямоугольников через четырехмерные вектора, для представления пирамиды карт признаков и для представления бинарной маски.
5. *BoundingBox*. Класс, содержащий информацию об окаймляющих прямоугольниках.
6. *NMS*. Содержит реализацию нескольких стратегий объединения прямоугольников.

5. Критерий оценки качества детектирования

Для оценки качества детектирования с помощью разработанной программной реализации используется показатель, равный отношению площади пересечения прямоугольника, полученного в результате детектирования, и прямоугольника из разметки к площади их объединения (Intersection over Union, IoU). Таким образом, считается, что лицо обнаружено правильно, если данный показатель превышает некоторый порог. В противном случае принимается, что лицо не обнаружено. Срабатывание алгоритма детектирования на области, где лицо отсутствует, считается ложным срабатыванием.

На основании приведенного показателя осуществляется построение ROC-кривой, которая отражает зависимость количества ложных срабатываний алгоритма детектирования (false positive) от точности детектирования (true positive rate).

Построение ROC-кривых обеспечивается с помощью инструментов, предоставляемых разработчиками наборов данных, используемых в ходе апробации.

6. Тренировочные и тестовые данные

В процессе проведения экспериментов для обучения детекторов использовался набор данных WIDER FACE [17], который состоит из 32203 изображений, содержащих 393703 лица, максимальное и минимальное разрешение которых составляет 851×1295 и 1×1 пикселей соответственно.

Изображения разделены на 61 класс (спортивные мероприятия, демонстрации, интервью и т.д.). Набор содержит лица, достаточно сложные для детектирования – с большим перекрытием, в масках, с размытием, малого размера. Тестирование проводится на валидационном множестве набора WIDER FACE, который разделен на несколько частей с различными уровнями сложности, а также на наборе данных Face Detection Data Set and Benchmark (FDDB) [3], который состоит из 2845 изображений, содержащих 5171 лицо, максимальное и минимальное разрешение которых составляет 398×589 и 8×13 пикселей.

7. Результаты экспериментов

Результаты, полученные на данных FDDB, приведены на рисунке (рис. 5). Построенные ROC-кривые говорят о сравнимости результатов. При одинаковом количестве ложных срабатываний точность детектирования в среднем отличается не более чем на 1.5% для каждого из методов, что свидетельствует о корректности разработанных реализаций.

С использованием разработанных программных реализаций проводится оценка качества на наборе данных WIDER FACE. Результаты для четырех описанных методов, а также для методов [8, 20], демонстрирующих лучшие результаты, приведены на рис. 9.

На наборе WIDER FACE лучшие результаты показывает метод, основанный на применении Faster R-CNN. При этом на подмножестве с низким уровнем сложности методы Faster R-CNN и UnitBox имеют близкие результаты к лучшим результатам, опубликованным на официальной странице набора [16], в то время, как на наборах со средней и высокой сложностью они значительно отстают от этих методов. Данный факт объясняется тем, что они хуже распознают лица малой величины (рис. 6).

Методы FastCNN и DPMFD на наборах со средней и низкой сложностью показывают намного худшие результаты, но на наборе с высокой сложностью результаты всех четырех методов близки, что во многом говорит об ориентированности методов на набор данных FDDB.

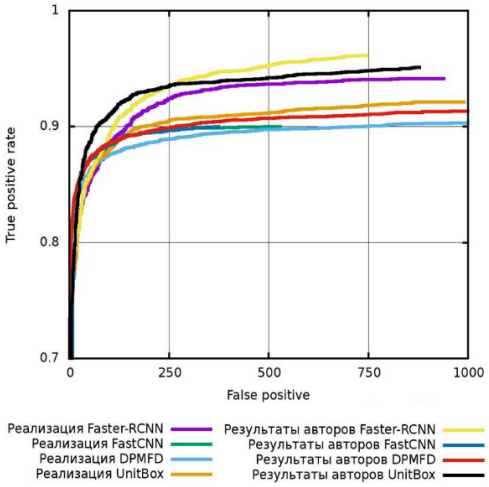

Рис. 5. Результаты работы на FDDB.

Рис. 6. Пример работы Faster R-CNN. Красным отмечены прямоугольники из разметки, зеленым результат поиска.

Тестирование выполняется на двух системах с различными конфигурациями. Конфигурация 1: Ubuntu, Intel(R) Core i5-2430M @ 2.4GHz, 6 GB. Конфигурация 2: CentOS, Intel(R) E5-2660 @ 2.2GHz, GPU: Nvidia Kepler K20, 64 GB. Средние времена работы методов на изображениях из базы WIDER FACE представлены в таблице 1.

Таблица 1. Время работы детекторов, с

	Конфигурация 1	Конфигурация 2
Faster R-CNN	4,5	0,39
FastCNN	5,3	0,42
DPMFD	15,4	5,2
UnitBox	3,9	0,36

Все четыре метода показывают хорошие результаты в довольно нестандартных случаях. Детекторы находят лица как на цветных изображениях, так и на изображениях в оттенках серого (рис. 7). Можно предположить, что это связано с тем, что в обучающем наборе изображений встречались изображения того и другого типа. Лица также находятся и на искусственно созданных изображениях (портреты, кадры из мультфильмов и т.д.). Исходя из этих результатов, можно сделать вывод, что с помощью глубоких сверточных сетей из изображения извлекаются признаки высокого уровня, такие, как границы лица, наличие глаз, носа, рта.

Очень часто на реальных изображениях лицо бывает перекрыто другими предметами. Все детекторы показывают возможность детектирования лиц в подобных случаях. Наилучший результат показывает метод,

основанный на Faster R-CNN. Даже при перекрытии около 60% данный метод достаточно точно определяет положение объекта и сохраняет верные пропорции лица (рис. 8). Метод UnitBox хуже справляется с этой задачей, так как определяет положение лица на основе бинарной маски, которая с данном случае получается искаженной.

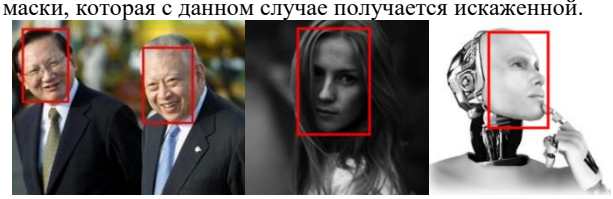

Рис. 7. Пример работы DPMFD на изображениях с лицами различных цветов.

Рис. 8. Пример работы Faster R-CNN при перекрытии лица.

По результатам проведенных экспериментов можно сделать вывод, что среди описанных методов наилучшим качеством детектирования обладает метод, основанный на Faster R-CNN. Данный метод в отличие от других, рассмотренных и реализованных в работе, достаточно хорошо определяет лица небольшого масштаба. Методы UnitBox и Faster R-CNN используют только один масштаб изображения при детектировании, что дает им большой выигрыш во времени по сравнению с двумя другими методами.

8. Заключение

В данной работе рассмотрены четыре метода решения задачи детектирования лиц. Описанные методы показывают, что задачу детектирования лиц можно решать разными способами. Разработана программная реализация этих методов на базе библиотек компьютерного зрения OpenCV [11] и глубокого обучения Caffe [2]. Программная реализация выложена в открытый доступ [20]. Воспроизведены результаты детектирования лиц с использованием разработанной реализации на данных FDDB. Эксперименты показывают сравнимые значения показателей качества с опубликованными на официальной странице FDDB [3]. Проведена оценка качества детектирования на наборе данных WIDER FACE. Методы Faster R-CNN и UnitBox на наборе с низкой сложностью показывают результаты, близкие к наилучшим опубликованным, но на наборах со средней и высокой степень сложности они значительно им проигрывают, что говорит об ориентированности указанных методов на более простой набор данных FDDB.

9. Литература

[1] Barbu A., Lay N., Gramajo G. Face Detection with a 3D Model. – [https://arxiv.org/abs/1404.3596]. – 2015.

[2] Caffe [http://caffe.berkeleyvision.org].

[3] Face Detection Data Set and Benchmark Home [http://vis-www.cs.umass.edu/fddb].

[4] Farfade S., Saberian M., Li L. Multi-view Face Detection Using Deep Convolutional Neural Networks. – [https://arxiv.org/abs/1502.02766]. – 2015.

[5] Girshick R., et al. Rich feature hierarchies for accurate object detection and semantic segmentation // In the Proc. of the Conf. on Computer Vision and Pattern Recognition. – 2014.

[6] He K., et al. Deep Residual Learning for Image Recognition. – [https://arxiv.org/abs/1512.03385]. – 2015.

[7] He K., et al. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. – [https://arxiv.org/abs/1502.01852]. – 2015.

[8] Hu P., Ramanan D. Finding Tiny Faces. – [https://arxiv.org/abs/1612.04402]. – 2017.

[9] Jiang H., Learned-Miller E. Face Detection with the Faster R-CNN. – [https://arxiv.org/abs/1606.03473]. – 2016.

[10] Krizhevsky A., Sutskever I., Hinton G. ImageNet classification with deep convolutional neural networks // Advances in Neural Information Processing Systems. – 2012.

[11] OpenCV [http://opencv.org].

[12] Ranjan R., Patel V.M., Chellappa R. A Deep Pyramid Deformable Part Model for Face Detection. – [http://arxiv.org/abs/1508.04389]. – 2015.

[13] Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. – [https://arxiv.org/abs/1409.1556]. – 2015.

[14] Sun X., Wu P., Hoi S. Face Detection using Deep Learning: An Improved Faster RCNN Approach. – [https://arxiv.org/abs/1701.08289]. – 2017.

[15] Triantafyllidou D., Tefas A. A Fast Deep Convolutional Neural Network for Face Detection in Big Visual Data. – [http://arxiv.org/abs/1508.04389]. – 2015.

[16] WIDER FACE: A Face Detection Benchmark Home [http://mmlab.ie.cuhk.edu.hk/projects/WIDERFace].

[17] Shuo Y., et al. WIDER FACE: A Face Detection

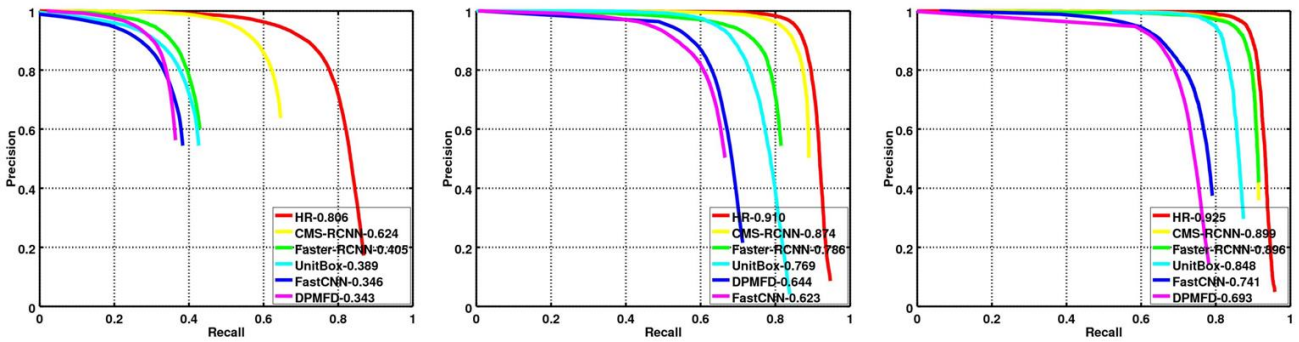

Рис. 9. Результаты работы на валидационной части набора WIDER FACE на изображениях высокой, средней и низкой сложности (слева направо). Показаны результаты методов: HR – Hybrid resolution [8], CMS-RCNN – Contextual Multi-Scale Region-based CNN [20], Faster RCNN – Faster Region-based CNN [9], UnitBox [18], FastCNN [15], DPMFD [12]. Для каждого из методов рядом с названием указано среднее значение точности.

Benchmark. – [https://arxiv.org/abs/1511.06523]. – 2015.

[18] Yu J., Jiang Y., Wang Z., Cao Z., Huang T.. UnitBox: An Advanced Object Detection Network. – INNS. – 2016.

[19] Zhang K., Zhang Zh., Li Z. Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks. – [http://arxiv.org/abs/1604.02878]. – 2016.

[20] Zhu C., et al. CMS-RCNN: Contextual Multi-Scale Region-based CNN for Unconstrained Face Detection. – [https://arxiv.org/abs/1606.05413]. – 2016.

[21] Разработанная программная реализация [https://github.com/DolotovEvgeniy/FaceDetection].

Об авторах

Евгений Долотов, студент Института информационных технологий, математики и механики ННГУ им. Н.И. Лобачевского. Его e-mail dolotov.evgeniy@gmail.com.

Валентина Кустикова, к.т.н., старший преподаватель кафедры Математического обеспечения и суперкомпьютерных технологий Института информационных технологий, математики и механики ННГУ им. Н.И. Лобачевского. Ее e-mail valentina.kustikova@itmm.unn.ru.