***Д. Д. Пономарева***

*Вологодский государственный университет*

### ИССЛЕДОВАНИЕ МОДЕЛЕЙ ДЕТЕКЦИИ ЛИЦ

Детекция лиц является одной из наиболее активно развивающихся областей компьютерного зрения и машинного обучения, играя ключевую роль в различных приложениях, таких как системы безопасности, распознавание лиц, дополненная реальность и взаимодействие человека с компьютером. С увеличением объемов данных и улучшением вычислительных мощностей, методы детекции лиц претерпели значительные изменения, от простых алгоритмов на основе признаков до сложных нейронных сетей. На данный момент существует множество различных моделей по данной технологии. В данной статье содержится анализ некоторых моделей, их преимущества и недостатки, а также их сравнительная характеристика.

### Методы детекции лиц, MTCNN, SSD, RetinaFace, Haar Cascades, HOG.

### Введение

### Методы детекции лиц — это технологии и алгоритмы, предназначенные для автоматического обнаружения лиц на изображениях или в видеопотоках. Эти методы позволяют идентифицировать и локализовать лица, определяя координаты области, содержащей лицо. В данной статье мы рассмотрим подходы к детекции с двум сторон. Сначала рассмотрим современные подходы к детекции лиц (MTCNN, SSD, RetinaFace), а потом классические (Haar Cascades и HOG).

### MTCNN

### MTCNN (Multi-Task Cascaded Convolutional Network) — это каскадная нейронных сеть, состоящая из 3 сетей: P-Net, R-Net и O-net. Каждая сверточная сеть использует данные для входа с выхода предыдущей сети, тем самым увеличивая точность прогноза[1]. Работа сети состоит из 3 этапов:

### Предварительное обнаружение (после работы P-net получаем координаты ограничивающих прямоугольников предполагаемых лиц).

### Уточнение (области с отсутствием лиц отсекаются R-net и кроме того добавляется достоверность к оставшимся областям).

### Локализация (O-net убирает прямоугольники с низкой достоверностью и добавляет координаты пяти ключевых лицевых точек (глаза, нос, уголки рта)).

### *Преимущества*:

### *Многозадачность:* предварительное обнаружения лица, уточнение ограничивающих прямоугольников и определение ключевых точек лица.

### ***Детектирование лиц при различных условиях съемки****:* Например, при плохой освещённости, наклоне или повороте головы, неполного нахождения лица в кадре или прикрытого ладонью.

### *Высокая точность*: Благодаря каскадной архитектуре, позволяющей поэтапно уточнять прогноз, достигается высокая точность детекции.

### *Недостатки:*

### Скорость работы: работает более медленно в сравнении с другими современными методами из-за своей многоуровневой архитектуры, кроме того при обработке сложных изображений ей потребуются большие вычислительные ресурсы.

**SSD**

SSD (Single-Shot Multibox Detection) представляет собой глубокую сверточную нейронную сеть, основной особенностью которой является возможность классификации объектов и предсказание их координат за один проход [2]. Архитектура состоит из нескольких сверточных слоев, которые на разных уровнях извлекают признаки из изображения. Для обнаружения объектов разного размера в работе SSD применяется набор якорей(«anchor boxes») - прямоугольников, разных размеров, разбивающих поле зрения изображения на много прямоугольных участков и далее для каждого из них на каждом этапе выполняется классификация (наличие объекта определенного класса) и регрессии (предсказании координат границ объекта). Это позволяет одновременно обрабатывать и определять классы объектов и их местоположение. Когда предсказания получены, то устраняются дубликаты, это делается с помощью метода NMS (non-maximum suppression), который отсекает перекрывающие ограничивающие рамки.

### *Преимущества*:

### ***Высокая скорость и эффективность:*** SSD выполняет обнаружение за один проход по входному изображению, что делает его подходящим для приложений реального времени.

### ***Поддержка различных масштабов:* использования нескольких уровней для предсказания объектов позволяет эффективно обнаруживать лица, которые находятся на различных расстояниях и под разными углами.**

### ***Высокая точность детекции:*** Для этого модель делает предсказания в разных масштабах на картах признаков и явно разделяет предсказания по соотношению сторон.

*Недостатки:*

1. *Чувствительность к масштабу объектов на изображении:*поскольку дополнительные слои предназначены для обнаружения объектов в разных масштабах, SSD-моделям может быть сложно распознавать объекты, которые значительно меньше или больше объектов в обучающем наборе данных[4].

**RetinaFace**

Retina Face - один из самых высокоточных методов детекции лиц, который использует глубокое обучение, в своей основе, как и SSD, имеет сверточную нейронную сеть, которая позволяет извлекать сложные признаки для дальнейшего обнаружения лиц и определения ключевых точек[11]. Но RetinaFace отличается некоторыми усовершенствованиями. Важно отметить, что RetinaFace -специализированная модель, разработанная специально для детектирования лиц, а SSD - как универсальный детектор разных объектов. Улучшению точности в RetinaFace способствует многозадачный подход, когда одновременно выполняется идентификация участка изображения как лицо или нет, определяется положение ограничительной рамки вокруг лица и координаты 5 ключевых точек лица (глаза, уголки рта, нос). На основе извлеченных признаков модель предсказывает вероятность наличия лица в каждой области, а также координаты ключевых точек. Дубликаты удаляются с помощью метода NMS.

*Преимущества:*

1. *Многомасштабность:* возможность обрабатывать изображения с лицами разных размеров.
2. *Многозадачность:* одновременное решение нескольких задач позволяет сразу применять полученную информацию к решению других задач, тем самым повышая эффективность и точность модели.
3. *Очень высокая точность детекции:* высокая точность в сравнении с другими моделями достигается благодаря лежащей в основе глубокой сверточной нейронной сети, многозадачному подходу и многомасштабному анализу.

*Недостатки:*

1. *Вычислительные ресурсы:* являясь сложной моделью, требует ресурсов для обучения и работы.
2. *Скорость работы:* работает несколько медленнее чем более простые методы детекции лиц.

**HOG**

Идея HOG (Histogram of Oriented Gradients) состоит в получении векторного представления изображения и дальнейшей классификации изображения. Вначале для каждого изображения вычисляются градиенты по горизонтали и вертикали, чтобы можно было определять направления яркости для дальнейшего выделения контуров. Далее изображение разбивается на ячейки, для каждой из которых вычисляется HOG (гистограмма ориентированных градиентов), которая показывает распределение направлений градиентов в этой ячейки, после чего они объединяются в блоки. Значения в блоках нормализуются (чтобы уменьшить влияние освещения и контраста) и создается общий HOG - дескриптор, представляющий собой вектор, содержащий информацию о распределении градиентов. Дескриптор подается на вход классификатору, который принимает решение о наличии или отсутствии лица, для обнаружения лиц на изображении используется метод скользящего окна. Как и в методе SSD для устранения дубликатов применяется метод NMS[8].

### *Преимущества*:

### ***Эффективность и простота реализации:* в сравнении с другими более сложными моделями относительно проста в реализации.**

### ***Устойчивость к изменению освещения:* Благодаря нормализации блоков** влияние освещения и контраста уменьшается.

### *Недостатки:*

### ***Чувствительность к перемене ориентации объекта*** *и масштабированию:* модель не всегда эффективно обнаруживает лица, находящиеся под углом или при изменении выражений лица и плохо справляется с лицами, значительно отличающимися от тех, на которых она была обучена.

### *Точность:* в сравнении с другими современными методами имеет невысокую точность детекции.

### *Чувствительность классификатора:* важен выбор параметров, таких как размер ячеек и количество бинов, используемых в гистограмме.

### *Реализация:* для достижения хорошей производительности требуется большой объём обучающих данных, а процесс обучения может быть вычислительно дорогим.

**Haar Cascades**

Метод основан на использовании простых признаков Хаара и каскадной структуры классификаторов. Для обучения необходим набор из положительных (с лицами) и отрицательных изображений (без лиц). Признаки Хаара вычисляются с помощью интегрального изображения и представляют собой простые прямоугольные фильтры, выявляющие определенные текстурные особенности изображения. Классификаторы обучаются на основе алгоритма Adaboost, который выбирает наиболее информативные признаки и формирует из них единый классификатор, и классификаторы объединяются в каскад, состоящий из нескольких уровней, который проверяет наличие лица на изображении. Как и в предыдущих моделях на этапе постобработки используется метод скользящего окна и NMS[10].

### *Преимущества*:

### ***Скорость:* быстрая работа благодаря каскадной структуре, позволяющей быстро отсекать области без лиц.**

### ***Простота реализации:* метод прост в реализации в сравнении с другими моделями и не требует сложных вычислительных ресурсов**

### *Недостатки:*

### ***Чувствительность к условиям съёмки****:* в сложных сценариях, таких как изменение освещенности, наклон лица, нестандартная поза, эффективность метода может снижаться.

### ***Точность:*** в сравнении с другими современными методами имеет невысокую точность детекции.

### ****Сравнение моделей****

Попробуем сравнить качество моделей детекции лиц. Для этого возьмем данные о качестве, полученные на датасете WIDER FACES[6], одном и самых популярных датасетов, содержащего множество разнообразных изображений для детектирования, где лица людей могут быть частично прикрыты, иметь разный масштаб, освещенность, угол и т.д. Кроме того данные являются размеченными. Датасет делится на три уровня сложности по размеру лиц на изображении: Легкий (больше 300 pixels), Средний (50-300 pixels), Сложный (10-50 pixels). Для определения точности работы алгоритмов детекции лиц обычно используется метрика IOU (Intersection over Union). Для каждого лица проверяется параметр IOU и если он  ≥0.5, то лицо считается верно обнаруженным и помечается как True Positive (TP). Неподходящие векторы, обозначаются как False Positive (FP), необнаруженные лица - False Negative (FN). Для определения качества модели оцениваются метрики **recall, precision, F-score**(**recall**- сколько лиц из всех лиц мы действительно нашли; **precision** - доля лиц которые детектор верно определил; **F-score** - среднее гармоническое precision и recall). Для оценки точности модели важно рассматривать не только метрики, **recall, precision, но и F-score[5]**, поскольку она объединяет в себе информацию о точности и полноте алгоритма. Мы будем пользоваться результатами исследования полученными в работе[9], а также показателями mAP (mean Average Precision) для моделей MTCNN, SSD, RetinaFace на момент 2023 года[[7](https://learnopencv.com/what-is-face-detection-the-ultimate-guide/#RetinaFace-(May-2019))]. Данные из этих исследований представлены в таблице 1.

Таблица 1. Сравнение метрик моделей детекции лиц

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Модель | Recall | Precision | F-score | mAP |
| MTCNN | 0.86 | 0.89 | 0.87 | 0.915 |
| SSD | 0.82 | 0.66 | 0.73 | 0.931 |
| RetinaFace | - | - | - | 0.994 |
| Haar | 0.45 | 0.91 | 0.61 | - |
| HOG | 0.38 | 0.95 | 0.54 | - |

### Согласно этим данным наилучшие результаты показывают модели RetinaFace, MTCNN и SSD, которые демонстрируют лучшие показатели F-score (MTCNN и SSD), а также имеют высокий mAP (mean Average Precision). Классические методы детектирования Haar Cascades и HOG выигрывают в precision, но уступают современным моделям в recall и f-score. Это связано с тем, что методы Haar Cascades и HOG используют относительно простые признаки для классификации в отличие от методов, основанных на сверточных нейронных сетях. Помимо этого, они чувствительны к изменениям освещенности, углам обзора или выражению лица, а также сложному фону или шуму, поскольку не могут эффективно отделить объект от фона. Основным преимуществом у современных методов является то, что они используют многоуровневую архитектуру для извлечения признаков, что позволяет им лучше справляться с разнообразием объектов, обучаться на больших объемах данных и адаптироваться к различным условиям. Кроме того, в современных сетях, таких как RetinaFace используются различные подходы по улучшению работы сверточных нейронных сетей, такие как многозадачность и многомасштабность. Эти улучшения значительно влияют на качество, так показатель mAP у RetinaFace имеет самое высокое значение из рассмотренных сетей. Но у классических методов тоже есть свои плюсы, как мы увидели, алгоритмы с простой реализацией могут работать не достаточно точно, но зато быстро, в отличие от сложных и точных, но медленных моделей, поэтому выбор конкретной модели или метода зависит от требований и ограничений в задаче. Проанализировав анализ работы моделей, ключевые особенности, а также выделив их преимущества и недостатки, была составлена сравнительная характеристика, представленная таблицей 2. Из нее можно сделать вывод, что MTCNN имеет высокую точность детекции и устойчива к изменениям условий съемки, но работает медленно и сложна в реализации. SSD также имеет высокую точность, устойчива к условиям съемки и имеет сложную реализацию, но показывает высокую скорость работы за счет однопроходного подхода, RetinaFace является самой точной, устойчивой к условиям съемки, но сложной из-за улучшений глубоких нейронных сетей и поэтому работает не очень быстро, Haar Cascades - просты в реализации и работают быстро, но не достигают высокой точности и неустойчивы к условиям съемки. HOG демонстрируют несколько лучшие показатели в скорости работы и точности, но уступают MTCNN , SSD и RetinaFace.

Таблица 2. Сравнительная характеристика моделей детекции лиц

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Модель | Скорость | Точность | Сложность реализации | Устойчивость к условиям съемки |
| MTCNN | + | +++ | + | ++ |
| SSD | +++ | +++ | +++ | ++ |
| RetinaFace | + | ++++ | +++ | ++ |
| Haar Cascades | +++ | - | - | +- |
| HOG | + | + | - | +- |

- низкая

+- умеренная

+ средняя

+ + хорошая

++ + высокая

++++ очень высокая

### Заключение

В данной статье были рассмотрены основные методы детекции лиц, а именно MTCNN, SSD, RetinaFace, Haar Cascades и HOG, описаны алгоритмы работы этих методов, проанализированы их преимущества и недостатки, а в качестве подведения итогов всей работы проведён краткий сравнительный анализ, выявлены их зависимости от некоторых характеристик. Результаты представлены в таблице 2.

### Литература

### MTCNN algorithm and code understanding-joint learning of face detection and face alignment [электронный ресурс]. // URL: https://www.programmerall.com/article/5714682723/ (дата обращения 27.11.24).

### Snehal Joshi. Why does the choice of object detection models matter? [электронный ресурс]./ Snehal Joshi. // URL: <https://www.hitechbpo.com/blog/top-object-detection-models.php> (дата обращения 27.11.24)

### Kukil . Mean Average Precision (mAP) in Object Detection [электронный ресурс]/ Kukil //URL: <https://learnopencv.com/mean-average-precision-map-object-detection-model-evaluation-metric/#Datasets-and-Model-Evaluation-Competitions> (дата обращения 27.11.24)

### Single shot detectors (SSDs) [электронный ресурс].// URL: <https://www.baeldung.com/cs/ssd> (дата обращения 27.11.24)

### F-score: what are Accuracy, Precision, Recall and F1score? [электронный ресурс]. //URL: <https://klu.ai/glossary/accuracy-precision-recall-f1> (дата обращения 27.11.24)

### WIDER face dataset [электронный ресурс]. // URL: <https://datasets.activeloop.ai/docs/ml/datasets/wider-face-dataset/> (дата обращения 27.11.24)

### What is Face Detection? Ultimate Guide 2023 + Model Comparison [электронный ресурс ].//URL: <https://learnopencv.com/what-is-face-detection-the-ultimate-guide/#RetinaFace-(May-2019)> (дата обращения 27.11.24)

### Гистограмма направленных градиентов [электронный ресурс ].// URL: <https://ru.wikipedia.org/wiki/Гистограмма_направленных_градиентов> (дата обращения 27.11.24)

### Жилкин, Ф.И. Анализ решений задачи детекции лиц на изображениях в сфере киберкриминалистики : Курсовая работа. - М: Санкт-Петербург, 2020.

### Face Detection using Haar Cascades [электронный ресурс].// URL: <https://docs.opencv.org/4.x/d2/d99/tutorial_js_face_detection.html> (дата обращения 27.11.24)

### David Cochard. RetinaFace: A Face Detection Model for High Resolution Images [электронный ресурс]./ David Cochard // URL: <https://readmedium.com/retinaface-a-face-detection-model-designed-for-high-resolution-6c3900771a01> (дата обращения 27.11.24)

### *D. D. Ponomareva*

### *Vologda State University*

### INVESTIGATION OF FACE DETECTION MODELS

### Face detection is one of the most actively developing areas of computer vision and machine learning, playing a key role in various applications such as security systems, facial recognition, augmented reality and human-computer interaction. With increasing amounts of data and improved computing power, face detection methods have undergone significant changes, from simple feature-based algorithms to complex neural networks. At the moment, there are many different models for this technology. This article contains an analysis of some models, their advantages and disadvantages, as well as their comparative characteristics.

### Face detection methods, MTCN, SSD, RetinaFace, HaarCascades, HOG.