

## **СИСТЕМА РАСПОЗНАВАНИЯ ЛИЦ ПО ВИДЕО ИЗОБРАЖЕНИЮ В РЕАЛЬНОМ МАСШТАБЕ ВРЕМЕНИ ПРИ НАЛИЧИИ МАСКИ НА ЛИЦЕ**

Белорусский государственный университет, г. Минск, Беларусь

Статья посвящена исследованию ключевых технологии и разных подходов к задаче обнаружения и идентификации лиц в реальном масштабе времени при наличии маски на лице. Реализован алгоритм детектирования и распознавания лиц людей с маской на лице по цифровым оптическим изображениям. Создана система распознавания лиц, входящих в здание на основе веб камеры Logitech C270 с разрешением 1280x720 и частотой кадров 30.

Обнаружение лиц – чрезвычайно важная область компьютерного зрения, необходимая для распознавания лиц, анализа настроений, видеонаблюдения и многих других областей. Обнаружение лиц на изображениях в режиме реального времени сопряжено с множеством трудностей, включая окклюзию, значительные вариации масштаба, различные условия освещения, различные позы лица, богатые выражения лица и т. д. Основная проблема обнаружения лиц в реальных сценариях заключается в производительности обнаружения. Кроме того, дополнительные трудности при решении этой задачи вносит наличие маски на лице. Решению этих проблем посвящено множество работ и был достигнут большой прогресс в разработке глубоких сверточных нейронных сетей (CNN), решающих эту задачу. Показатели средней точности (AP) для сложного набора данных WIDER FACE за последние годы улучшились с 40% до 90%. В работе описаны разные подходы как к задаче обнаружения, так и к задаче идентификации лиц при наличии маски на лице. Реализован алгоритм обнаружения и распознавания лиц для задачи детектирования входящих людей в здание.

Для произвольного изображения цель обнаружения лиц состоит в том, чтобы определить, есть ли на изображении какие-либо лица, и, если они есть, то найти местоположение изображения и размер каждого лица.

Для решения задачи обнаружения объектов возможно использование каскадных классификаторов на основе признаков Хаар – это эффективный метод обнаружения объектов, предложенный Полом Виолой и Майклом Джонсом в статье [1]. Это подход основан на машинном обучении, где каскадная функция обучается из множества изображений с целевым объектом и без него (позитивные и негативные изображения, соответственно). Затем она используется для обнаружения объектов на других изображениях.

Авторы методики YOLO (You only look once) [2] поставили обнаружение и классификацию объекта как единую регрессионную задачу, начиная с пикселей изображения и заканчивая ограничивающими прямоугольниками и вероятностями классов. Архитектура YOLO “смотрит” на изображение один раз, чтобы предсказать, какие объекты и где они находятся. Она тренируется на полных изображениях и напрямую оптимизирует производительность обнаружения. Эта унифицированная модель имеет несколько преимуществ по сравнению с традиционными методами обнаружения объектов.

Система распознавания R-CNN [3] состоит из трех модулей. Первый генерирует независимые от категории предложения региона. Эти предложения определяют набор обнаруженный кандидатов, доступных для нашего детектора. Второй модуль представляет собой большую сверточную нейронную сеть, которая извлекает вектор признаков фиксированной длины из каждого региона. Третий модуль представляет собой набор классических линейных SVM. В качестве CNN-сети использовалась преобученная архитектура AlexNet.

Для распознавания лиц применялся наиболее перспективный подход, основанный на использовании нейронных сетей. Наилучшие результаты в области распознавания лиц (по результатам анализа публикаций) показала Convolutional Neural Network (CNN) или сверточная нейронная сеть, которая является логическим развитием идей таких архитектур НС, как

когнитрона и неокогнитрона. Успех обусловлен возможностью учета двумерной топологии изображения, в отличие от многослойного перцептрона. Отличительными особенностями CNN являются локальные рецепторные поля (обеспечивают локальную двумерную связность нейронов), общие веса (обеспечивают детектирование некоторых черт в любом месте изображения) и иерархическая организация с пространственным сэмпингом (spatial subsampling). Благодаря этим нововведениям CNN обеспечивает частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям.

Распознавание лица в маске – это задача распознавания закрытого лица с предварительным знанием скрытой области целевого лица. Маска закрывает область носа, рта и щек. Свободными областями могут быть глаза, брови и лоб. Таким образом, система распознавания лица в маске должна эффективно сосредоточиться на анализе признаков, которые можно извлечь из областей, включая глаза, брови и лоб, которые не закрыты маской, субъекта.

Детектирование наличия на лице маски является более тривиальной задачей чем распознавание и детектирование, так как по сути представляет из себя обычную задачу классификации. Поэтому в качестве модели по распознаванию маски на лице была выбрана простая сеть, состоящая из трех последовательно соединенных сверточных слоев. После детектирования наличия маски на лице, изображение передается в сеть (FaceNet), которая выделит вектор признаков размерностью 512. Далее полученный вектор признаков сравнивается с векторами признаков лиц из списка известных лиц для сравнения. В реализации использовалась норма Фробениуса вектора разницы между векторами, если разница между векторами больше, чем пороговое значение, то лицо считается не распознанным.

$$\|A\|_F = [\sum_{i=1}^N (y_i - \hat{y}_i)^2]^{1/2},$$

где  $y_i$  - вектор признаков лица из списка уже известных, а  $\hat{y}_i$  - неопознанный вектор.

В качестве источника видео использовалась веб камера Logitech C270 с разрешением 1280x720 и частотой кадров 30. Получаемое изображение переводится в черно-белый формат и уменьшается в разрешении на 25%. Все эти шаги сделаны с целью ускорения работы детектора положения лиц. С точки зрения точности, алгоритмы нахождения координат лица на изображении, основанные на нейронных сетях, показывают высокую точность и устойчивость к различного рода помехам (очки, маска, плохое позиционирование лица и т.п.). Однако высокая скорость их работы достигается при запуске данных алгоритмов на графических картах. Так как решение хотелось сделать универсальным, не привязываясь к наличию на ноутбуке видеокарты одного производителя (Nvidia), то выбор пал на алгоритмы классического компьютерного зрения, а именно HOG, так как HOG применяется в промышленных решениях детектирования объектов в случаях, когда скорость обработки имеет важное значение.

Структурная схема работы алгоритма распознавания лиц людей при наличии маски на лице представлена на рисунке 1. Пример работы разработанного программного приложения, реализующего созданный алгоритм распознавания, представлен на рисунке 2. Результат реализованного и распознанного изображения выводится в отдельном окне, найденное лицо обводится ограничивающим прямоугольником и подписывается распознанным из базы именем. Созданный алгоритм способен распознавать лица вне зависимости есть ли на лице маска или нет. Так же на экране приводится информация о наличии маски на распознанном лице и о общем количестве найденных на изображении лиц, по умолчанию выбирается лицо с самой большой площадью ограничивающего треугольника. Следующим шагом было получение координат прямоугольника, ограничивающего местоположение лица, из черно-белого изображения. Как упоминалось выше, для выделения лица, используется алгоритм HOG.

Так как целевым применением алгоритма является распознавание людей, входящих в здание, то скорость изменения картинки будет невысокой. Обработать кадры 30 раз в секунду не имеет смысла, поэтому ограничимся обработкой одного кадра в секунду, что позволит нам ещё больше оптимизировать алгоритм.

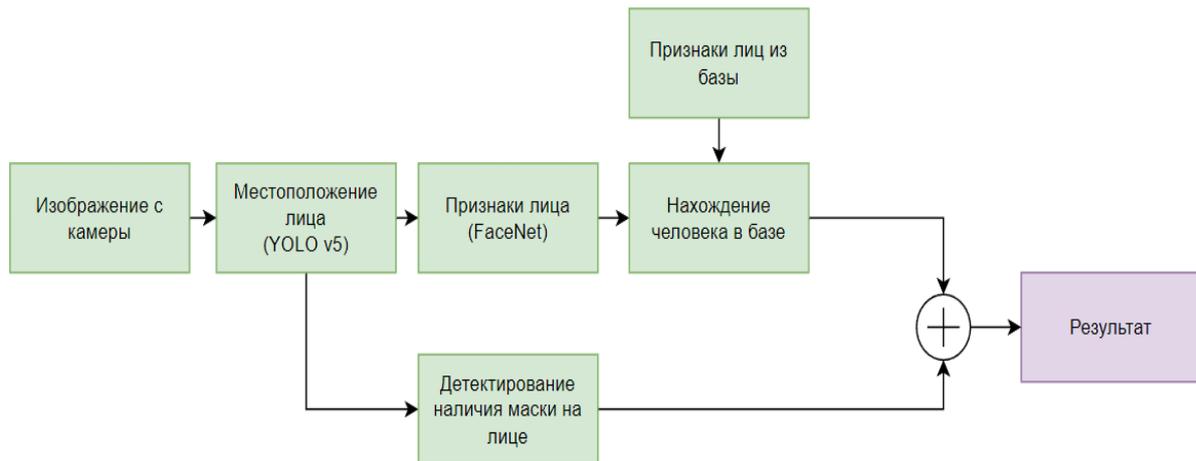


Рисунок 1 – Структурная схема работы алгоритма распознавания лиц людей при наличии маски на лице



Рисунок 2 – Пример работы программного приложения, реализующего созданный алгоритм распознавания

После получения координат лиц, возвращаем изображение в исходное разрешение, увеличивая на 25% и используем их для выделения интересующей нас области, но уже на цветном изображении. Область с распознанным лицом передаем в сверточную сеть, для получения вектора признаков. В данном подходе используется, предобученная на классификацию лиц, архитектура ResNet34, в этом случае, сеть на последних слоях будет выделять абстрактные признаки, трудно интерпретируемые для человека, однако весьма информативные для задачи распознавания.

Таким образом, в результате работы разработаны алгоритм и программное приложение для распознавания лиц людей при наличии маски на лице, основанное на нейронных сетях, обеспечивающее устойчивость к различного рода помехам (очки, маска, плохое позиционирование лица и т.п.). Проведены экспериментальные исследования разработанного приложения для распознавания в реальном масштабе времени лиц людей, входящих в здание Исследования показали хорошую эффективность работы системы и время распознавания составило менее 0.5 с.

#### Список литературы

1. Viola, Paul & Jones, Michael. Rapid Object Detection using a Boosted Cascade of Simple Features. IEEE Conf Comput Vis Pattern Recognit. 2001. – 1. I-511. 10.1109/CVPR.2001.990517.
2. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection. 2016. P.779-788. 10.1109/CVPR.2016.91.
3. Ross Girshick, Microsoft Research. Fast R-CNN, 2015. 10.1109/ICCV.2015.169.