



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ

КАФЕДРА СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ И УПРАВЛЕНИЯ

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ НА ТЕМУ:

Улучшенный алгоритм YOLOV4 на основе распознавания
жестов

Студент ИУ5-22М
(Группа)

Шань Сяньлун
(Подпись, дата)

Шань Сяньлун
(И.О.Фамилия)

Руководитель

(Подпись, дата)

(И.О.Фамилия)

Консультант

(Подпись, дата)

(И.О.Фамилия)

2022 г.

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

УТВЕРЖДАЮ
Заведующий кафедрой _____ ИУ5 _____
(Индекс)

(И.О.Фамилия)
« __ » ____ 20 ____ г.

З А Д А Н И Е
на выполнение научно-исследовательской работы

по теме Улучшенный алгоритм YOLOV4 на основе распознавания жестов

Студент группы ИУ5-2М
Шань Сяньлун (Фамилия, имя, отчество)

Направленность НИР (учебная, исследовательская, практическая, производственная, др.)
Учебная, практическая
Источник тематики (кафедра, предприятие, НИР) Кафедра

График выполнения НИР: 25% к ____ нед., 50% к ____ нед., 75% к ____ нед., 100% к ____ нед.

Техническое задание

В данной работе предлагается новый алгоритм YOLOV4 для распознавания жестов,
который сравнивается с традиционными сетями YOLOV4 и Faster R-CNN. Подчеркивая
свое превосходство

Оформление научно-исследовательской работы:

Расчетно-пояснительная записка на ____ листах формата А4.
Перечень графического (иллюстративного) материала (чертежи, плакаты, слайды и т.п.)

Дата выдачи задания « ____ » _____ 20 ____ г.

Руководитель НИР

(Подпись, дата) (И.О.Фамилия)

Студент

Шань Сяньлун
(Подпись, дата) (И.О.Фамилия)

Примечание: Задание оформляется в двух экземплярах: один выдается студенту, второй хранится на кафедре.

Содержание

1. Введение	3
2. Суть проблемы, Существующие решения и их проблемы	4
3. Решения - Улучшенная сеть YOLOV4	5
3.1 Введение в алгоритм YOLOV4	6
3.2 Улучшенный алгоритм YOLOV4	8
3.2.1 Изменение разрешения исходной карты характеристик	9
3.2.2 Кластеризация K-means корректирует априорный размер кадра 10	
4. Экспериментальные результаты и анализ	12
4.1 Подготовка набора данных о жестах	12
4.2. Показатели оценки алгоритма	13
4.3 Результаты экспериментов и анализ распознавания жестов на стандартных и самодельных наборах данных жестов	14
4.3.1 Результаты распознавания улучшенного алгоритма для небольших целевых жестов	14
4.3.2 Результаты идентификации трех алгоритмов на стандартном наборе данных	16
4.3.3 Результаты идентификации трех алгоритмов на наборе данных домашнего производства	18
5. Заключение	22
Список использованных источников	24

1. Введение

В нашей повседневной жизни язык тела является незаменимым способом общения.

Многие коммуникативные сообщения можно выразить с помощью определенных поз рук и движений тела, что упрощает общение. В отличие от традиционного взаимодействия с помощью клавиатуры и мыши, взаимодействие на основе языка тела свободно от ограничений устройства и позволяет распространить взаимодействие с экрана на D-пространство, что удобно и эффективно.

В настоящее время соматосенсорная технология широко используется во многих областях, таких как игры, здравоохранение и образование, а также появляются приложения для взаимодействия человека и компьютера, основанные на соматосенсорной технологии. Данная статья посвящена изучению распознавания жестов.

Распознавание жестов - популярная тема исследований в области взаимодействия человека и компьютера. Технология распознавания жестов на основе зрения позволяет компьютеру понимать смысл действий пользователя, делая взаимодействие человека и компьютера более естественным и удобным.

В данной работе улучшен классический одноэтапный алгоритм YOLOV4 для обнаружения целей и проведено сравнение с оригинальным алгоритмом YOLOV4 и двухэтапным алгоритмом Faster R-CNN на наборе данных жестов домашнего производства. Найден алгоритм глубокого обучения, который больше подходит для распознавания жестов.

2.Суть проблемы, Существующие решения и их проблемы

Жесты являются частью общения людей в повседневной жизни. С развитием технологий и разработкой интеллектуальных устройств жесты постепенно используются как естественный и интуитивный способ взаимодействия с людьми и машинами, и широко применяются в симуляторах боевых действий, аэрокосмической промышленности, умных домах, безопасном вождении и т.д. Распознавание жестов - это новая технология, которая использует устройства сбора информации для захвата различных жестов и интерпретации человеческих жестов с помощью распознавания образов или других методов. В системе "человек-компьютер При взаимодействии человека и компьютера, для того чтобы компьютер мог точно распознавать жесты, понимать При взаимодействии человека и компьютера важно повысить точность распознавания жестов, чтобы компьютеры могли точно распознавать жесты, понимать, что они представляют, и выполнять соответствующие действия. При взаимодействии человека и компьютера важно повысить точность распознавания жестов, чтобы компьютер мог точно распознавать жесты, понимать их значение и выполнять соответствующие действия. В настоящее время исследования в области распознавания жестов Существует три основных типа исследований в области распознавания жестов: распознавание жестов на основе внешних аппаратных устройств распознавание жестов на основе внешних аппаратных устройств, распознавание жестов на основе компьютерного зрения и распознавание жестов на основе глубоких технологий распознавание жестов на основе внешних аппаратных устройств, распознавание жестов на основе компьютерного зрения и распознавание жестов на основе глубокого обучения.

В последние годы глубокое обучение набрало огромные обороты в области компьютерного зрения, например, в классификации изображений и обнаружении

целей. По сравнению с традиционными методами распознавания жестов, методы на основе глубокого обучения позволяют избежать ручного извлечения признаков и обеспечивают лучшие адаптивные возможности модели. В настоящее время алгоритмы обнаружения целей на основе глубокого обучения делятся на две категории: первая - алгоритмы обнаружения целей на основе регрессии, представленные YOLOV3, такие как SSD и YOLOV4, которые не генерируют регионы-кандидаты и напрямую используют сеть обнаружения для генерации информации о местоположении и категории цели, что имеет более высокую скорость обнаружения и отвечает требованиям реального времени; вторая - алгоритмы обнаружения целей на основе предложения региона, представленные Fast R - CNN, такие как SPP - Net и Faster R - CNN. Другой тип - это алгоритм обнаружения цели на основе предложения региона, представленный Fast R - CNN, такой как SPP - Net , Faster R - CNN и т.д., который должен быть сначала отнесен к регионам-кандидатам, что делает скорость обнаружения медленнее и не удовлетворяет требованиям реального времени. При распознавании жестов точность распознавания жестов и скорость распознавания должны быть как можно выше. Поэтому в данной работе для распознавания жестов выбран YOLOV4, чтобы обеспечить хороший баланс между точностью и скоростью распознавания. Если бы оригинальный алгоритм использовался непосредственно для распознавания жестов, возникли бы следующие проблемы.

1.Невозможность эффективного обнаружения жестов цели разного масштаба из-за различных размеров человеческих целей.

2.Из-за схожести различных жестов, это приводит к проблеме частичного обнаружения пропущенных жестов, а также ложного обнаружения, и скорость распознавания не высока.

3.Решения - Улучшенная сеть YOLOV4

В связи с вышеуказанными проблемами, в данной работе для

решения проблемы меньшего количества данных о жестах используется увеличение данных, а также усовершенствование YOLOV4 путем изменения разрешения исходной карты признаков, чтобы она могла обнаружить больше признаков, корректировка априорного размера кадра с помощью алгоритма кластеризации K-средств для повышения способности обнаружения жестов разного масштаба, дальнейшее повышение точности распознавания жестов и проведение экспериментов на стандартном наборе данных, а также на самодельном наборе данных о жестах.

3.1 Введение в алгоритм YOLOV4

YOLOV4 - это алгоритм сквозного обнаружения целей. YOLOV4 сначала извлекает входные признаки через магистральную сеть Darknet53. особенности входного изображения через магистральную сеть Darknet53 и выводит 19×19 , 38×38 , 76×76 масштабные характеристики для сети обнаружения. Сеть обнаружения регрессирует эти три Сеть обнаружения регрессирует эти три признака и использует алгоритм подавления не максимального значения для удаления более низких доверительных значений Сеть регрессирует три признака и использует алгоритм подавления не максимального значения, чтобы удалить кадры предсказания с низким доверием и сохранить кадры предсказания с высоким доверием и использовать их в качестве кадров обнаружения цели. Класс и положение цели определяются с помощью алгоритма не максимального подавления, который удаляет кадры предсказания с низким доверием и сохраняет кадры с более высоким доверием в качестве кадров обнаружения цели. Выход класса и местоположения используется в качестве конечного обнаружения цели Результат полученных категорий и позиций используется в качестве окончательного результата обнаружения цели.

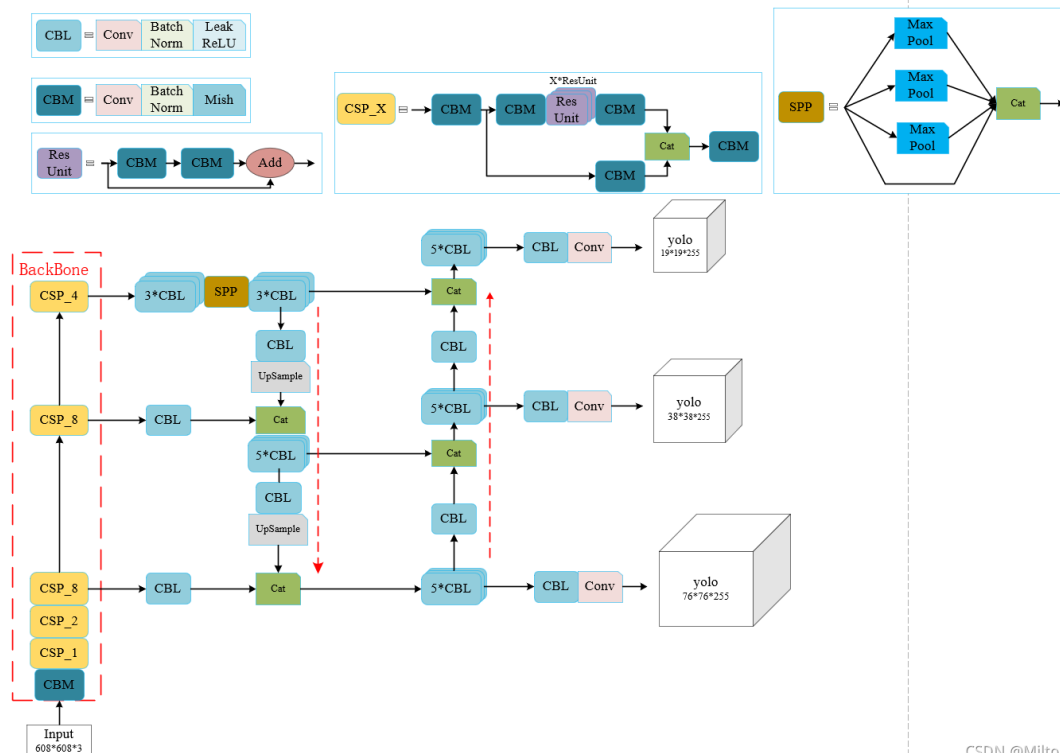
Алгоритм YOLOV4 основан на оригинальной архитектуре обнаружения целей YOLOV3, и вводит некоторые методы оптимизации

обработки данных, магистральной сети, обучения сети, функции активации и других аспектов, так что модель достигает наилучшего соответствия с точки зрения скорости и точности обнаружения. Усовершенствования YOLOV3 заключаются в следующем.

(1) Магистральная сеть YOLOV4 использует сеть CSPDarknet53. CSPDarknet53 имеет на пять модулей CSP больше, чем магистральная сеть Darknet53 YOLOV3. Регуляризация Drop -block использовалась для случайного удаления количества нейронов для упрощения структуры сети.

(2) Чтобы получить большее поле восприятия для кадра предсказания и эффективно отделить значимые признаки в данных, YOLOV4 добавляет модуль SPP после структуры CSPDark-net53. Структура SPP может эффективно увеличить диапазон восприятия признаков магистральной сети и значительно отделить информацию о значимых признаках.

3) YOLOV4 использует модуль FPN + PAN. В отличие от FPN YOLOV3, YOLOV4 добавляет пирамиду признаков снизу вверх после FPN, которая содержит две структуры PAN. Эта улучшенная структура заменяет единую структуру FPN в YOLOV3, где слой FPN передает сильные семантические признаки сверху вниз, а пирамида признаков передает сильные признаки локализации снизу вверх, что позволяет использовать улучшенную структуру FPN + PAN для агрегирования параметров на разных уровнях детекторов, дополнительно улучшая возможности извлечения признаков и, следовательно, возможности обнаружения цели.



CSDN @MiltonY

Сетевая структура yolov4

3.2 Улучшенный алгоритм YOLOV4

Алгоритм YOLOV4, обученный на общих данных PASCAL VOC плохо распознает жесты после обучения на наборе данных PASCAL, потому что в наборе данных PASCAL имеется 20 моделей классификации, которые не включают модель для жестов. Это связано с тем, что в наборе данных PASCAL есть 20 моделей классификации, которые не включают модель жеста, и поэтому жест не может быть хорошо обнаружен. Для решения этой проблемы Это объясняется тем, что в наборе данных PASCAL имеется 20 моделей классификации, которые не содержат моделей жестов. Структура сети YOLOV4 оптимизирована на основе оригинального алгоритма YOLOV4, а разрешение оригинальной карты признаков изменено. Алгоритм кластеризации K-means также используется для изменения размера предыдущего кадра Алгоритм кластеризации K - средств также используется для изменения размера предыдущего кадра, чтобы сеть могла быть более чувствительной к мелким целям и избежать проблемы человеческих Сеть более чувствительна к мелким целям,

чтобы избежать ложного обнаружения жестов рук из-за различных размеров целей и схожих жестов рук. Алгоритм кластеризации K-средств также используется для изменения размера предыдущего кадра, чтобы сеть была более чувствительна к мелким целям, избегая ложных и пропущенных обнаружений жестов из-за различных размеров целей человеческих рук и похожих жестов, улучшая обнаружение различных жестов и повышая точность определения жестов. Сеть также способна распознавать различные жесты и повышать точность распознавания жестов.

3.2.1 Изменение разрешения исходной карты характеристик

YOLOV4 использует сеть Darknet53 для извлечения признаков из входного изображения путем пятикратного понижения дискретизации изображения для получения пяти различных масштабов карт признаков. Неглубокие карты признаков с большим разрешением могут содержать больше информации о пикселях, т.е. подробную информацию, такую как контуры, края, текстуры и т.д., и точную локализацию цели, в то время как глубокие карты признаков с малым разрешением содержат больше семантической информации для классификации объектов. Если мелкие признаки объединяются с глубокими признаками, область принятия классификационной сети может быть расширена, при этом изучается более глобальная информация. Карта мелких признаков может улучшить семантическое представление модели, тем самым улучшая способность сети различать различные цели. Для того чтобы повысить скорость распознавания различных жестов, карты признаков, извлеченные моделью YOLOV4, в данной работе модифицируются и объединяются. Во-первых, тримасштаба оригинальных карт признаков 19×19 , 38×38 и 76×76 модифицируются в два больших масштаба 38×38 и 76×76 , и модель YOLOV4 предсказывает цели под этими двумя масштабами. Блок-схема показана на рисунке 1

**Рисунок 1 Блок-схема для изменения разрешения исходной карты
характеристик**

**Fig. 1 The flow chart of changing the resolution of the original feature map
original feature map**



3.2.2 Кластеризация K-means корректирует априорный размер кадра

Предложенный априорный кадр преобразует обнаружение цели в две задачи, которые решаются одновременно, т.е. наличие или отсутствие цели в фиксированной сетке и задача определения расстояния между предсказанным кадром и реальным кадром, что делает обнаружение цели гораздо более точным и быстрым, поскольку связь скользящего окна в других сетях отбрасывается. Размер априорной рамки важен для сети обнаружения целей YOLOV4. Оригинальная априорная рамка YOLOV4 получена путем кластеризации на

аннотации набора данных VOC. Из-за разнообразия объектов в наборе данных VOC соотношение сторон сильно варьируется, и априорный размер кадра также сильно варьируется. Однако в наборе данных жестов из-за разного расстояния руки человека от камеры или разного размера руки человека прямое использование исходного априорного кадра не отвечает требованиям распознавания. Для повышения точности распознавания жестов в данной работе для кластеризации набора данных снова используется алгоритм кластеризации K-средств для получения размера априорных кадров. Кластеризация набора меченых данных сначала выполняется путем увеличения числа центров кластеров, а затем коэффициент пересечения IoU между центрами кластеров и предсказанными и реальными кадрами используется в качестве параметра сходства алгоритма кластеризации K - means. Формула расчета расстояния следующая

$$d = 1 - \text{IoU} \quad (1)$$

где: IoU представляет собой отношение пересечения предсказанного кадра с истинным кадром. Уравнение (1) делает результирующий кадр предсказания более похожим на целевой кадр истинного маркера в определенной степени.

Для набора данных по жестам кластерный анализ был повторно проведен с использованием вышеуказанного метода, и были получены априорные ячейки (62, 70), (83, 132), (114, 93), (186, 128), (254, 127) и (374, 404), соответственно. Размер априорной рамки обратно пропорционален размеру карты признаков, при этом меньшие априорные рамки на больших картах признаков для меньших целей и большие априорные рамки на меньших картах признаков для больших целей. Таким образом, меньшие карты признаков 38×38 используют большие априорные рамки (186, 128), (254, 127), (374, 404), а большим картам признаков 76×76 назначаются меньшие априорные рамки (62, 70), (83, 132), (114, 93). Подробная информация представлена в таблице 1.

Таблица 1 Улучшенные априорные размеры рамки YOLOV4

Масштаб карты характеристик	Априорная коробка
38 x 38	(186, 128), (254, 127), (374, 404)
76 x 76	(62, 70), (83, 132), (114, 93)

4. Экспериментальные результаты и анализ

4.1 Подготовка набора данных о жестах

Набор данных языка жестов, используемый в этом эксперименте, представляет собой стандартный набор данных языка жестов, а также самодельный набор данных языка жестов. Стандартный набор данных жестов - это Sign Language DigisDataset, цифровой набор данных языка жестов, подготовленный учениками средней школы Anadolu в Анкаре, Турция, по следующему URL: <https://github.com/ardamavi/Sign-Lan-guage-Digits-Dataset>. Десять жестов были сделаны каждым учеником, представляя числа от 0 до 9. Этот набор данных был взят в том же контексте.

Набор данных по жестам домашнего изготовления был получен с помощью самой камеры. В качестве представителей были выбраны пять жестов рук: рука _ладонь, рука _ ван, рука _ кулак, рука _ палец и рука _ ок. Камера использовалась для съемки изображений пяти жестов с разных расстояний и под разными углами для восьми экспериментаторов.

Для того чтобы обогатить набор данных жестов и получить лучшие результаты обучения, в данной работе используются случайные преобразования для улучшения набора данных жестов, включая переворот изображения, поворот по часовой стрелке, поворот против часовой стрелки и преобразование масштаба для расширения набора данных жестов с первоначальных 600 изображений пяти жестов до 5000 изображений.

Набор данных жестов (как стандартных, так и самодельных) был создан с использованием формата набора данных VOC2007. С помощью программы

аннотирования изображений Labe-Img изображения в формате JPEGImages были аннотированы вручную, жесты на изображениях были обрамлены минимальным внешним прямоугольником, были получены граничная область жеста и метка класса жеста, и был создан соответствующий XML-файл, что завершило преобразование формата набора данных VOC2007. После этого случайным образом были сформированы обучающее валидационное и тестовое множества в соотношении 8:2, а затем проведено обучение и тестирование модели.

4.2. Показатели оценки алгоритма

Для каждого жеста я получу разные оценки уверенности и оценки позиционирования целевой рамки. Уверенность - это точность, с которой YOLOV4 предсказывает категорию целевого объекта. Точность локализации целевого кадра - это степень совпадения предсказанного кадра, выведенного моделью YOLOV4, с реальным кадром (я считаю локализацию целевого кадра точной, если значение IOU больше 0,7, в противном случае она не является точной).

В качестве метрик оценки для распознавания жестов используются Precision (P) и Recall (R). Precision - это доля всех правильно идентифицированных образцов ко всем идентифицированным образцам, а Recall - доля всех правильно идентифицированных образцов ко всем образцам, которые должны быть идентифицированы, рассчитывается как

$$\begin{cases} P = \frac{TP}{TP + FP} \times 100\%, \\ R = \frac{TP}{TP + FN} \times 100\%, \end{cases} \quad (2)$$

Прогнозы делаются на основе партии тестовых данных, и результаты можно разделить на четыре категории.

TP (True Positive): исходный класс является положительным, а предсказанный результат - положительным. (Правильный прогноз - положительный)

FP (False Positive): исходный отрицательный класс предсказывается как положительный. (Ложное срабатывание)

TN(True Negative): изначально отрицательный класс, предсказание отрицательное. (Правильный прогноз - отрицательный)

FN(False Negative): изначально положительный класс, предсказанный как отрицательный. (Ложноотрицательный)

Где: TP - количество правильно идентифицированных изображений жестов, т.е. истинно положительных; FP - количество неправильно идентифицированных изображений жестов, т.е. ложноположительных; FN - количество пропущенных изображений жестов, т.е. ложноотрицательных. В теории, чем ближе точность и отзыв предсказания данных к 1, тем лучше модель предсказания.

В дополнение к этому, время обнаружения также используется в качестве метрики оценки. Время обнаружения выражается путем нахождения среднего времени, которое требуется каждому алгоритму для вывода одного изображения в тестовом наборе выводов. Чем меньше время обнаружения, тем быстрее работает алгоритм.

4.3 Результаты экспериментов и анализ распознавания жестов на стандартных и самодельных наборах данных жестов

Для того чтобы проверить эффективность усовершенствованного алгоритма в данной работе, были проведены сравнительные эксперименты алгоритма Faster R-CNN, оригинального алгоритма YOLOV4 и усовершенствованного алгоритма YOLOV4 на стандартном наборе данных жестов и наборе данных самодельных жестов, соответственно.

4.3.1 Результаты распознавания улучшенного алгоритма для небольших целевых жестов

После улучшения размера предыдущего кадра результаты распознавания небольших целевых жестов, находящихся на большом расстоянии от камеры, показаны на рис. 3. Целевой кадр представляет собой место, где находится

распознанный жест, первая часть текста указывает на результат распознавания жеста, а вторая часть - на уровень доверия к жесту, а рис. 4-6 аналогичны. Анализ результатов распознавания небольших целевых жестов показан в таблице 2, где уровень уверенности для нераспознанных жестов и неправильно распознанных жестов, а также местоположение целевого поля обозначены знаком "×".

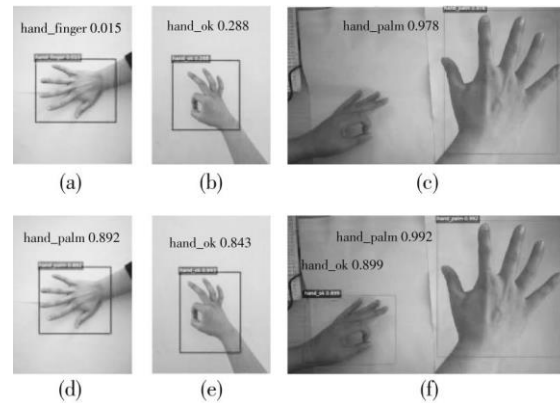


Рисунок 3 Экспериментальные результаты улучшенного априорного кадра для распознавания жестов мелких целей

Первый ряд (рисунки (a), (b) и (c)) на рисунке 3 показывает результаты распознавания жестов до улучшения, а второй ряд (рисунки (d), (e) и (f)) показывает соответствующие результаты распознавания жестов после улучшения априорного кадра.

Таблица 2 Анализ результатов распознавания жестов мелких целей

Жесты	Содержание теста	Оригинальный алгоритм	Улучшенные алгоритмы
Hand_palm	Результаты идентификации	Hand_finger	Hand_palm
	Уверенность	X	0.892
	Позиционирование целевого кадра	X	Точность
Hand_ok	Результаты идентификации	Hand_ok	Hand_ok
	Уверенность	0.288	0.943
	Позиционирование целевого кадра	Точность	Точность
Hand_ok	Результаты идентификации	Не определено	Hand_ok
	Уверенность	X	0.899
	Позиционирование целевого кадра	X	Точность

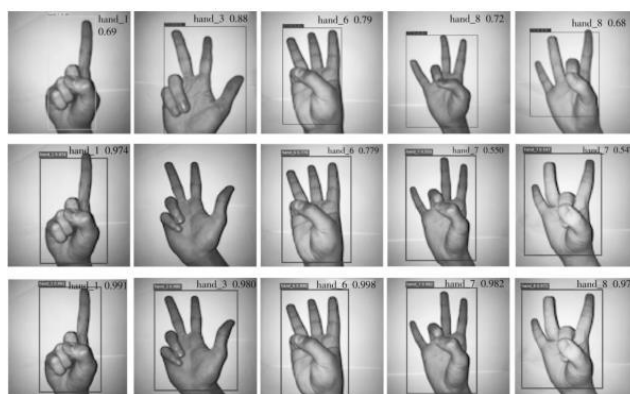
Hand_palm	Результаты идентификации	Hand_palm	Hand_palm
	Уверенность	0.978	0.992
	Позиционирование целевого кадра	Неточность	Точность

Анализ в таблице 2 показывает, что алгоритм был улучшен для точного определения небольших целей и избежания проблемы пропущенного обнаружения, а уровень доверия был увеличен.

4.3.2 Результаты идентификации трех алгоритмов на стандартном наборе данных

Жесты были обнаружены на стандартном наборе данных жестов с помощью модифицированного алгоритма YOLOV4 и алгоритма Faster R-CNN. Некоторые результаты обнаружения жестов показаны на рис. 4, анализ результатов распознавания жестов приведен в таблице 3, а сравнение скорости распознавания жестов тремя алгоритмами показано в таблице 4. На рисунке используется целевое поле для обозначения местоположение жеста, а в верхнем левом углу поля - жест категория, к которой принадлежит жест, и уровень доверия, который указывает на вероятность того, что жест будет распознан правильно. Уровень доверия указывает на вероятность правильного распознавания жеста, но текст размыт, поэтому он снова отмечен в правом верхнем углу изображения. Это еще раз отмечено в правом верхнем углу изображения.

Рис. 4 Сравнение результатов распознавания жестов алгоритмом Faster R-CNN и улучшенным алгоритмом на стандартном наборе данных



**Таблица 3 Анализ результатов распознавания жестов для трех алгоритмов
на стандартном наборе данных**

Жесты	Содержание теста	Faster R-CNN	Оригинальный алгоритм	Улучшенные алгоритмы
Жест № 1	Результаты идентификации	Hand_1	Hand_1	Hand_1
	Уверенность	0.69	0.974	0.991
	Позиционирование целевого кадра	Неточность	Неточность	Точность
Жест № 3	Результаты идентификации	Hand_3	Не определено	Hand_3
	Уверенность	0.88	X	0.980
	Позиционирование целевого кадра	Неточность	X	Точность
Жест № 6	Результаты идентификации	Hand_6	Hand_6	Hand_6
	Уверенность	0.79	0.779	0.988
	Позиционирование целевого кадра	Неточность	Точность	Точность
Жест № 7	Результаты идентификации	Hand_8	Hand_7	Hand_7
	Уверенность	X	0.550	0.982
	Позиционирование целевого кадра	X	Точность	Точность
Жест № 8	Результаты идентификации	Hand_8	Hand_7	Hand_8
	Уверенность	0.68	X	0.972
	Позиционирование целевого кадра	Неточность	X	Точность

На рисунке 4 показаны результаты частичного распознавания жестов на стандартном наборе данных. В первой строке на рисунке 4 показаны результаты распознавания алгоритма Faster R - CNN, сокращенно F R, во второй - оригинальные результаты распознавания жестов YOLOV4, а в третьей - улучшенные результаты распознавания жестов YOLOV4. Анализ в таблице 3 показывает, что при использовании алгоритма Faster R - CNN и улучшенного алгоритма наблюдаются неточная локализация жестов, пропущенное

обнаружение и ложное обнаружение, а уровень доверия низкий; однако улучшенный алгоритм имеет хороший эффект обнаружения и может точно распознавать различные жесты с высоким уровнем доверия и точной локализацией целевого кадра, а пропущенное обнаружение и ложное обнаружение жестов улучшилось.

Таблица 4 Сравнение скорости распознавания жестов тремя алгоритмами на стандартном наборе данных

Алгоритмы	Коэффициент точности(Precision/ %)	Коэффициент отзыва(Recall/ %)	Время обнаружения / с
Faster R-CNN	89.33	84.70	0.138
Оригинальный алгоритм	92.45	87.6	0.102
Улучшенный алгоритм	97.80	92.45	0.086

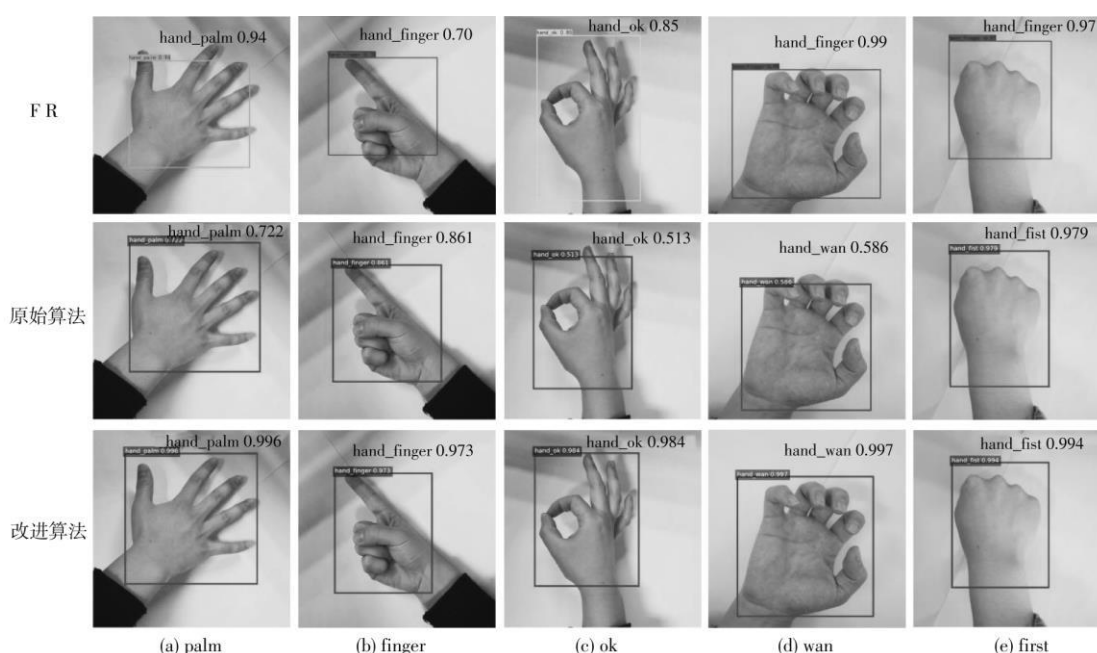
Результаты в таблице 4 показывают, что улучшенный алгоритм YOLOV4 имеет показатель точности 97,80% и показатель запоминания 92,45% на стандартном наборе данных жестов, что на 5,35% выше точности распознавания и на 4,85% выше показателя запоминания по сравнению с алгоритмом до улучшения, и на 8,47% выше точности распознавания и на 7,75% выше показателя запоминания по сравнению с алгоритмом Faster R-CNN. Среднее время обнаружения для одного изображения составляет 0,086 с, что меньше, чем у оригинального алгоритма и Faster R-CNN, и хорошо удовлетворяет требованиям реального времени.

4.3.3 Результаты идентификации трех алгоритмов на наборе данных домашнего производства

Для дальнейшей проверки превосходства усовершенствованной сети в данной работе алгоритм Faster R-CNN был использован для обучения на наборе данных жестов домашнего изготовления. Для обнаружения использовались оригинальный алгоритм, улучшенный алгоритм и алгоритм Faster R-CNN. Для

сравнения были отобраны 10 изображений (5 для простых фонов и 5 для сложных сцен). Результаты распознавания жестов на простом фоне показаны на рисунке 5, а анализ результатов распознавания приведен в таблице 5. Результаты сложных сцен показаны на рис. 6, а анализ результатов приведен в таблице 7.

Рис. 5 Сравнение результатов распознавания жестов алгоритмом Faster R-CNN и улучшенным алгоритмом "до и после" на домашнем наборе данных



На рисунке 5 в первой строке показаны результаты распознавания жестов, полученные с помощью Faster R - CNN, где Faster R - CNN сокращенно обозначается как FR, во второй строке показаны результаты распознавания жестов оригинальным алгоритмом YOLOV4, а в третьей строке – результаты распознавания улучшенным алгоритмом YOLOV4. Анализ результатов распознавания в таблице 5 показывает, что по сравнению с оригинальным алгоритмом YOLOV4 и алгоритмом Faster R - CNN, улучшенный алгоритм в данной статье может точно распознавать различные жесты, а целевой кадр точно позиционируется с высокой достоверностью распознавания, что указывает на то, что производительность улучшенного алгоритма в данной статье лучше, чем у первых двух алгоритмов.

Результаты в таблице 6 показывают, что улучшенный алгоритм YOLOV4

имеет точность 98,57% и запоминание 93,89%, что на 5,24% выше точности распознавания и на 4,57% выше запоминания по сравнению с предыдущим алгоритмом, и на 8,61% выше точности распознавания и на 8,46% выше запоминания по сравнению с алгоритмом Faster R-CNN. Среднее время обнаружения для одного изображения составляет 0,090 с, что меньше, чем у оригинального алгоритма и Faster R-CNN. В целом, видно, что метод, описанный в данной статье, может удовлетворить требования распознавания жестов в реальном времени, сочетая в себе как точность обнаружения, так и скорость обнаружения.

**Таблица 5 Анализ результатов распознавания жестов
тремя алгоритмами в простом контексте**

Жесты	Содержание теста	FR	Оригинальный алгоритм	Улучшенные алгоритмы
palm	Результаты идентификации	Hand_palm	Hand_ok	Hand_palm
	Уверенность	0.74	X	0.998
	Позиционирование целевого кадра	Точность	X	Точность
Finger	Результаты идентификации	Hand_finger	Hand_finger	Hand_finger
	Уверенность	0.82	0.317	0.995
	Позиционирование целевого кадра	Неточность	Точность	Точность
ok	Результаты идентификации	Hand_ok	Hand_ok	Hand_ok
	Уверенность	0.73	0.979	0.997
	Позиционирование целевого кадра	Точность	Точность	Точность
Wan	Результаты идентификации	Hand_wan	Hand_wan	Hand_wan
	Уверенность	0.59	0.996	0.992
	Позиционирование целевого кадра	Точность	Неточность	Точность

Fist	Результаты идентификации	Hand_finger	Hand_fist	Hand_fist
	Уверенность	0.82	0.317	0.995
	Позиционирование целевого кадра	Неточность	Точность	Точность

Таблица 6 Сравнение скорости распознавания жестов тремя алгоритмами для простых сцен

Алгоритмы	Коэффициент точности Precision/ %	Коэффициент отзыва Recall/ %	Время обнаружения / с
Faster R-CNN	89.96	85.43	0.146
Оригинальный алгоритм	93.33	89.32	0.106
Улучшенный алгоритм	98.57	93.89	0.090

Таблица 7 Анализ результатов распознавания жестов тремя алгоритмами в сложных сценариях

Жесты	Содержание теста	Faster R-CNN	Оригинальный алгоритм(Yolov4)	Улучшенные алгоритмы
palm	Результаты идентификации	Hand_palm	Hand_ok	Hand_palm
	Уверенность	0.74	X	0.998
	Позиционирование целевого кадра	Точность	X	Точность
Finger	Результаты идентификации	Hand_finger	Hand_finger	Hand_finger
	Уверенность	0.82	0.317	0.995
	Позиционирование целевого кадра	Неточность	Точность	Точность
ok	Результаты идентификации	Hand_ok	Hand_ok	Hand_ok
	Уверенность	0.73	0.979	0.997
	Позиционирование целевого кадра	Точность	Точность	Точность
Wan	Результаты идентификации	Hand_wan	Hand_wan	Hand_wan
	Уверенность	0.59	0.996	0.992
	Позиционирование целевого кадра	Точность	Неточность	Точность

Fist	Результаты идентификации	Hand_finger	Hand_fist	Hand_fist
	Уверенность	0.82	0.317	0.995
	Позиционирование целевого кадра	Неточность	Точность	Точность

Рисунок 6 Результаты распознавания жестов на сложных фонах

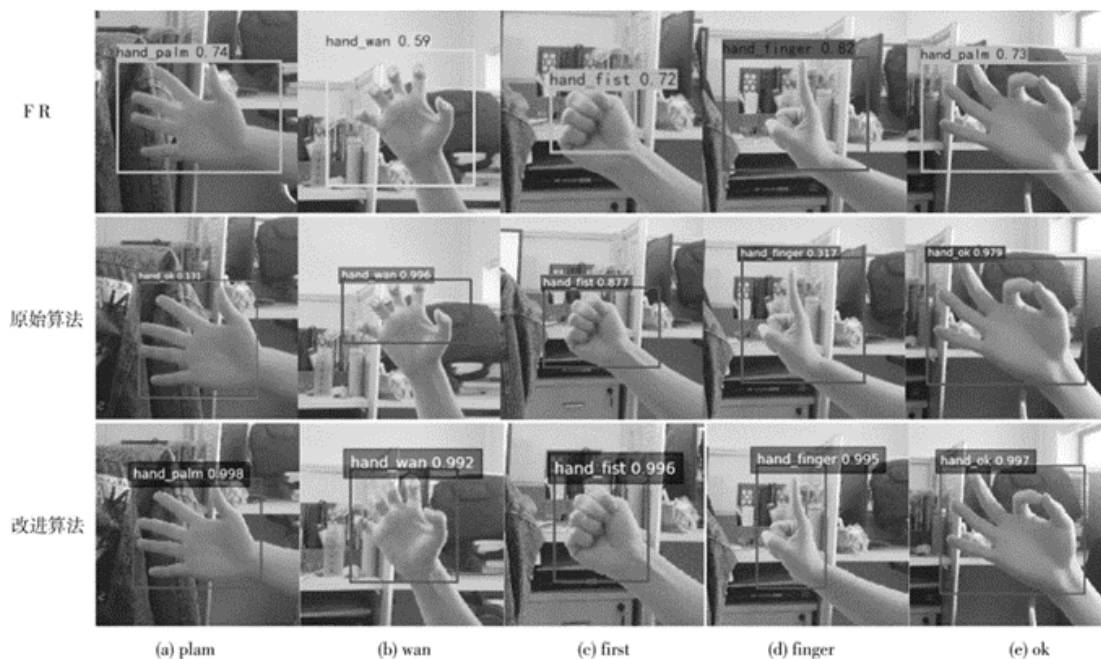


Рисунок 6 Результаты распознавания жестов на сложных фонах На рисунке 6 в первой строке показаны результаты распознавания жестов, полученные с помощью Faster R - CNN, сокращенно F R. Во второй строке показаны результаты распознавания жестов оригинальным алгоритмом YOLOV4, а в третьей - результаты распознавания улучшенным алгоритмом YOLOV4. Как показано в таблице 7, улучшенный алгоритм может точно распознавать различные жесты в сложных сценах, и целевой кадр точно позиционируется, а доверие к жестам высокое по сравнению с алгоритмом Faster R - CNN и оригинальным алгоритмом YOLOV4, что полностью демонстрирует высокую точность и устойчивость улучшенного алгоритма YOLOV4 для задач распознавания жестов.

5.3 Заключение

В данной работе предлагается улучшенный алгоритм взаимодействия жестов на основе YOLOV4, который использует увеличение данных для решения

проблемы малого количества данных о жестах, оптимизируя сетевую модель YOLOV4 и пересчитывая размер предыдущего кадра с помощью алгоритма кластеризации K-средств для повышения способности обнаружения различных жестов. Результаты экспериментов показывают, что предложенный метод превосходит традиционный алгоритм и алгоритм Faster R- CNN по точности распознавания жестов, отвечает требованиям реального времени, позволяет достичь быстрого обнаружения жестов и в то же время может распознавать жесты в сложных сценах.

Список использованных источников

- [1] Silpani D C, Suematsu K, Yoshida K. A Feasibility Study on Hand Gesture Intention Interpretation Based on Gesture Detection and Speech Recognition[J]. Journal of Advanced Computational Intelligence and Intelligent Informatics, 2022, 26(3): 375-381.
- [2] Kong H, Lu L, Yu J, et al. Continuous authentication through finger gesture interaction for smart homes using WiFi[J]. IEEE Transactions on Mobile Computing, 2020, 20(11): 3148-3162.
- [3] Pandey P, Jain V. Hand gesture recognition for sign language recognition: A review[J]. International Journal of Science, Engineering and Technology Research (IJSETR), 2015, 4(3): 464-470.
- [4] Wang S, Jiang K, Chen J, et al. Simple But Effective: Upper-Body Geometric Features for Traffic Command Gesture Recognition[J]. IEEE Transactions on Human-Machine Systems, 2021, 52(3): 423-434.
- [5] LaViola Jr J J. Context aware 3D gesture recognition for games and virtual reality[M]. ACM SIGGRAPH 2015 Courses, 2015: 1-61.
- [6] Wadhawan A, Kumar P. Sign language recognition systems: A decade systematic literature review[J]. Archives of Computational Methods in Engineering, 2021, 28(3): 785-813.
- [7] Quam D L. Gesture recognition with a dataglove[C]. IEEE Conference on Aerospace and Electronics, 1990: 755-760.
- [8] Schlenzig J, Hunter E, Jain R. Vision based hand gesture interpretation using recursive estimation[C]. Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers, 1994, 2: 1267-1271.
- [9] Gupta B, Shukla P, Mittal A. K-nearest correlated neighbor classification for Indian sign language gesture recognition using feature fusion[C]. 2016 International conference on computer communication and informatics (ICCCI), 2016: 1-5.
- [10] Sharma S, Jain S. A static hand gesture and face recognition system for blind

people[C]. 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN). IEEE, 2019: 534-539.

[11] Alzubaidi L, Zhang J, Humaidi A J, et al. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions[J]. Journal of big Data, 2021, 8(1): 1-74.

[12] Guo X, Xu W, Tang W Q, et al. Research on optimization of static gesture recognition based on convolution neural network[C]. 2019 4th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), 2019: 398-3982.

[13] Adithya V, Rajesh R. A deep convolutional neural network approach for static hand gesture recognition[J]. Procedia Computer Science, 2020, 171: 2353-2361.