|  |  |
| --- | --- |
|  | **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, искусственный интеллект и системы управления |
| КАФЕДРА | Системы обработки информации и управления |

**Домашнее задание**

**по дисциплине «Методы машинного обучения»**

Выполнила: Маркова А.В.

Группа: ИУ5-23М

Проверил: Гапанюк Ю.Е.

Москва, 2022 г.

HaGRID — набор данных изображений распознавания жестов HAnd

В этой статье мы представляем огромный набор данных Ha-GRID (HAnd Gesture Recognition Image Dataset) для систем распознавания жестов рук. Этот набор данных содержит 552 992 образца, разделенных на 18 классов жестов. Аннотации состоят из ограничительных коробок рук с гез-турными надписями и разметками ведущих стрелок. Предлагаемый набор данных позволяет создавать системы HGR, которые могут быть использованы в обслуживании видеоконференций, системах домашней автоматизации, автомобильном секторе, услугах для людей с нарушениями речи и слуха и т. д. Мы особенно сосредоточены на интеграции с устройствами для управления ими. Именно поэтому все 18 выбранных жестов являются функциональными, привычными для окружающих людей, и могут быть стимулом к совершению каких-то действий. Кроме того, мы использовали краудсорсинговые платформы для сбора набора данных и учитывали различные параметры для обеспечения разнообразия данных. Мы описываем проблемы использования существующих наборов данных HGR для нашей задачи и предоставляем их подробный обзор. Кроме того, предлагаются исходные условия для задач обнаружения рук и классификации жестов. HaGRID и предварительно обученные модели находятся в открытом доступе.

Знакомство

Использование жестов в человеческом общении играет важную роль [[1](#_bookmark20)]: жесты могут усиливать состояния эмоционально или полностью заменять их. Более того, распознавание жестов рук (HGR) может быть частью взаимодействия человека и компьютера. Такие системы имеют широкий спектр реальных применений в автомобильной среде [[2](#_bookmark21)], [[3](#_bookmark22)], системах домашней автоматизации [[4](#_bookmark23)], огромном разнообразии видео/потоковые платформы и др. [[5](#_bookmark24)], [[6](#_bookmark27)]. Кроме того, система может быть частью виртуального помощника или сервиса для активных пользователей языка жестов слабослышащие и речевые люди [[7](#_bookmark28)], [[8](#_bookmark29)]. Эти области требуют, чтобы система работала в режиме онлайн и была надежной к задним планам, сценам, объектам и условиям освещения.



Рисунок 1. 18 классов жестов, включенных в HaGRID («inv.» является аббревиатурой от «перевернутый»).

В этой статье мы представляем набор данных HaGRID для проектирования систем HGR. Он содержит более полумиллиона изображений, разделенных на 18 классов жестовых знаков (рисунок [1](#_bookmark0)), которые не ориентированы на язык. Такие жесты выбираются n для обозначения системы управления устройством и выполняют одну семиотическую функциональную роль [[18](#_bookmark39)]. Семиотические жесты помогают общаться между людьми и, в наших случаях, они используются для человеко-компьютерного взаимодействия. В разделе [3](#_bookmark6) будет описано, как проектировать динамические жесты с использованием выбранных статических жестов, т.е. создавать эргодические жесты (другая функциональная роль, которая соответствует способности манипулировать объектами) политическими жестами. Небольшой лексикон функциональных жестов в наборе данных задуман для снижения сложности системы HGR и избежания чрезмерной когнитивной нагрузки на пользователя устройства [[19](#_bookmark40)]. Необходимо иметь комфортно разработанные действия при использовании систем управления жестами. Все представленные жесты были выбраны как наиболее полезные для этой цели[[20](#_bookmark41)]. Мы также добавили дополнительный класс с образцами естественных движений рук и назвали его «без жеста». Все изображения различаются по фону, освещению, сцене и объектам. Эта гетерогенность достигается за счет использования двух краудсорсинговых платформ, а именно Яндекс.Толока[3](#_bookmark3) и ABC Elementary[4](#_bookmark4). Все примеры в наборе данных имеют высокое разрешение и собраны в RGB.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Набор данных | Образцы | Классы | Предметы | Сцены | Резолюция | Окружающая среда | Непрерывный |
| Американский язык жестов, 1998 [[9](#_bookmark30)] | 2,500 | 40 | 1 | 1 | 320 × 243 | - | без переходного состояния |
| Кембридж Хэнд ГД, 2007 [[10](#_bookmark31)] | 900 | 9 | 2 | 1 | 320 × 240 | статический | с переходным состоянием |
| ChaLearn Мультимодальный GD, 2013 [[11](#_bookmark32)] | 13,858 | 20 + 1 | 27 | 2 | 640 × 480 | статический | не непрерывный |
| ЧаУарн ИзогД, 2016 [[12](#_bookmark33)] | 47,933 | 249 | 21 | 15 | 240 × 320 | статический | не непрерывный |
| nvGesture, 2016 [[13](#_bookmark34)] | 1,532 | 25 + 1 | 20 | 1 | 320 × 240 | статический, динамический | с переходным состоянием |
| ЭгоГестурия, 2018 [[14](#_bookmark35)] | 24,161 | 83 + 1 | 50 | 6 | 640 × 480 | статический, динамический | с переходным состоянием |
| АрАСЛ, 2018 [[15](#_bookmark36)] | 54,049 | 32 | 40 | - | 64 × 64 | статический | не непрерывный |
| IPN Рука, 2020 [[16](#_bookmark37)] | 4,218 | 13 + 1 | 50 | 28 | 640 × 480 | статический, динамический | без переходного состояния |
| РУКИ, 2021 [[17](#_bookmark38)] | 12,000 | 29 | 5 | 5 | 960 × 540 | статический | не непрерывный |
| ХаГРИД, 2022 | **552,992** | 18 + 1 | **34,730** | ≥ **34 730** | **1920 × 1080** | статический | не непрерывный |

Таблица 1. Основные параметры указанных наборов данных жестов рук. + 1 в третьем столбце означает, что набор данных содержит дополнительный класс «без жеста». Количество сцен в последней строке не может быть меньше количества предметов. Обратите внимание, что HaGRID состоит как минимум из 90% изображений FullHD.

Остальная часть документа структурирована следующим образом: в разделе 2 мы рассмотрим связанную работу, которая описывает некоторые из существующих наборов данных HGR; раздел 3 предоставляет весь процесс создания набора данных; в разделе 4 представлены несколько моделей , обученных на наборе данных и экспериментальных результатах. В заключение опишем основные направления будущей работы.

Связанная работа

Существует не менее 50 наборов данных распознавания жестов рук. Они различаются по количеству образцов, их разрешению, количеству классов, наличию отрицательных образцов, гомогенности сцен, расстоянию между камерой и каждым объектом, прикладным задачам и присутствию человека в кадре. В данной статье рассматриваются наборы данных в контексте решения задачи классификации. Обратите внимание, что некоторые из описанных наборов данных предназначены для решения дополнительных задач, таких как обнаружение и оценка позы. Язык жестов является одной из категорий стилей жестов [[21](#_bookmark42)], и к нему относится набор данных американского языка жестов (ASL) [[9](#_bookmark30)]. Набор данных ASL включает 2 500 изображений с 40 классами жестов из 1 субъекта и только из 1 сцены. Отдельного внимания заслуживает главная особенность наборов данных для языка жестов – подход от первого лица к сбору HGR dataset. Как и набор данных ASL, набор данных EgoGesture [[14](#_bookmark35)] также имеет эту функцию. Он содержит 83 класса жестов из 50 предметов и включает в себя примеры с динамическими средами. Мультимодальный набор данных о жестах ChaLearn [[11](#_bookmark32)] содержит около 15.000 изображений и может быть полезен для решения задач по es-тимации. Его использование ограничено тем фактом, что все его образцы жестов взяты из итальянского словаря жестов. Набор данных Cambridge Hand Gesture Dataset [[10](#_bookmark31)] включает 900 последовательностей изображений из 9 классов жестов рук без человека в кадре. Набор данных nvGesture [[13](#_bookmark34)] относится к другому типу стиля жестов – манипулятивным жестам. Он был разработан специально для управления автомобильными устройствами. Поэтому в нем содержится всего 1 сцена – человек, который сидит в машине. Этот набор данных разделен на 26 классов, включая дополнительный класс «без жеста». СТРЕЛКА IPN [[16](#_bookmark37)] содержит более тысяч изображений с 50 предметов. Основные преимущества набора данных IPN Hand заключаются в том, что он содержит непрерывные ges-tures без переходных состояний (такие же, как набор данных ASL) и образцы с естественными движениями рук для нацеливания на гес-туры. Однако набор данных ограничен классами жестов, потому что он был собран для взаимодействия с бесконтактными экранами. Один из самых больших наборов данных, ChaLearn IsoGD [[12](#_bookmark33)], основан на ранее упомянутом наборе данных ChaLearn Multi-modal ges- ture. Он включает в себя 9 различных областей жестов, включая язык тела, жестикуляции, выполняемые для сопровождения речи, и другие.

HANDS [[17](#_bookmark38)] и ArASL [[15](#_bookmark36)] являются наиболее подходящими наборами данных для нашего приложения. Они состоят из жестов, похожих на наши, и потенциально могут быть использованы нами. Как бы то ни было, HANDS содержит только 15 жестов, некоторые из которых выполняются обеими руками в общей сложности 29 уникальных классов. Большинство из них мало отличаются друг от друга, что усложняет использование системы HGR. ArASL подписан для арабского языка жестов и содержит около 54 000 изображений, разделенных на 32 класса жестовых знаков, соответствующих арабским буквам. Все егомаги i находятся в оттенках серого с 64 × 64 измерениями. MediaPipe Hands [[22](#_bookmark43)] мог бы быть хорошим вариантом, если бы он не был собран только у 18 пользователей с ограниченными вариациями в фоновом режиме. Кроме того, он не является общедоступным.

Некоторые рассмотренные наборы данных используют свой словарь жестов для конкретных областей, таких как язык жестов, автомобильный сектор и другие. Несмотря на то, что большинство наборов данных охватывают несколько типов жестов [[23](#_bookmark44)], этого недостаточно для наших целей. HaGRID предназначен для управления устройствами или приложениями для устройств. Все жесты набора данных вызывают определенные ассоциации из-за их предполагаемого значения. Это позволяет жестам решать определенные проблемы, такие как лайк / не нравится, воспроизведение / остановка записи, включение / выключение звука, управление регулируемой шкалой (например, шкала громкости) и т. Д. Кроме того, пользователь может комбинировать некоторые жесты для создания нового жеста, который не включён в набор данных (например, жесты «кулак» и «ладонь» могут быть использованы для создания ges-ture «захвата»). Все жесты показаны на рисунке  [1](#_bookmark0).

Приватные домены и не имеющие для нас необходимых жестов, коллекция HaGRID была мотивирована множеством других причин. Во-первых, разрешение других изображений HGR не соответствует разрешению камер на конкретных устройствах (например, SberPortal[5](#_bookmark7) и Sber-Top [6](#_bookmark8)), которое является FullHD. Во-вторых, другим наборам данных не хватает степени в сценах и условиях освещения, которые необходимы нейронным сетям для хорошего обобщения. Наконец, система HGR требует входов с разных расстояний к камере. Обратите внимание, что не все наборы данных доступны для общественности.

Изтаблицы 1 видно, что предлагаемый набор данных Ha-GRID является самым большим по количеству выборок. Он имеет самые высокие оценки разнообразия по предметам и сценам, что помогает избежать внутриклассовой изменчивости. Две последние колонки подтверждают сравнение между HaGRID и описанными наборами данных HGR в соответствии с такими характеристиками, как тип среды и наличие непрерывных жестов.

Набор данных HaGRID

Потребность в сочетании таких характеристик, как (1) изображения с высоким разрешением, (2) неоднородность по всей сцене изображения, предметы, их возраст и пол, освещение, расстояние от камеры до объектов и (3) количество образцов стали мотивацией для создания HaGRID. Набор данных содержит около полумиллиона RGB-изображений FullHD (1920 × 1080) с 18 жестами и классом «без жестов». В нем не менее 34 730 уникальных сцен. Имейте в виду, что предлагаемый набор данных содержит некоторые жесты в двух положениях: передней и тыльной стороне руки. Это позволяет выполнять динамические жесты с помощью двух статических жестов. Например, с помощью ges-tures "stop" и "stop inverted" вы можете спроектировать динамические жесты "swipe up" ("stop thumbs down", т.е. "stop" ro- ted на 180 градусов, как начало строки и "stop in-verted" в качестве конца) и "swipe down" ("stop" как начало строки и "stop inverted thumbs down", т.е. "stop in- перевернутый» повернутый на 180 градусов, как конец). Кроме того, вы можете получить еще 2 динамических жеста, «свайп вправо» и «свайп влево», с увеличением поворота на 90 градусов. Все примеры де-подписанных динамических жестов показаны на рисунке [2](#_bookmark10).

В дополнение к классификации жестов, HaGRID может использоваться для решения проблемы обнаружения рук (каждое изображение имеет *n* соответствующих ограничительных рамок для *n* рук в кадре) и двух задач двоичной классификации: (1) жест / не жест и (2) правая/левая ведущая рука[7](#_bookmark9). На рисунке [3](#_bookmark11) приведен пример разметки для одного образца в наборе данных.

Краудсорсинг

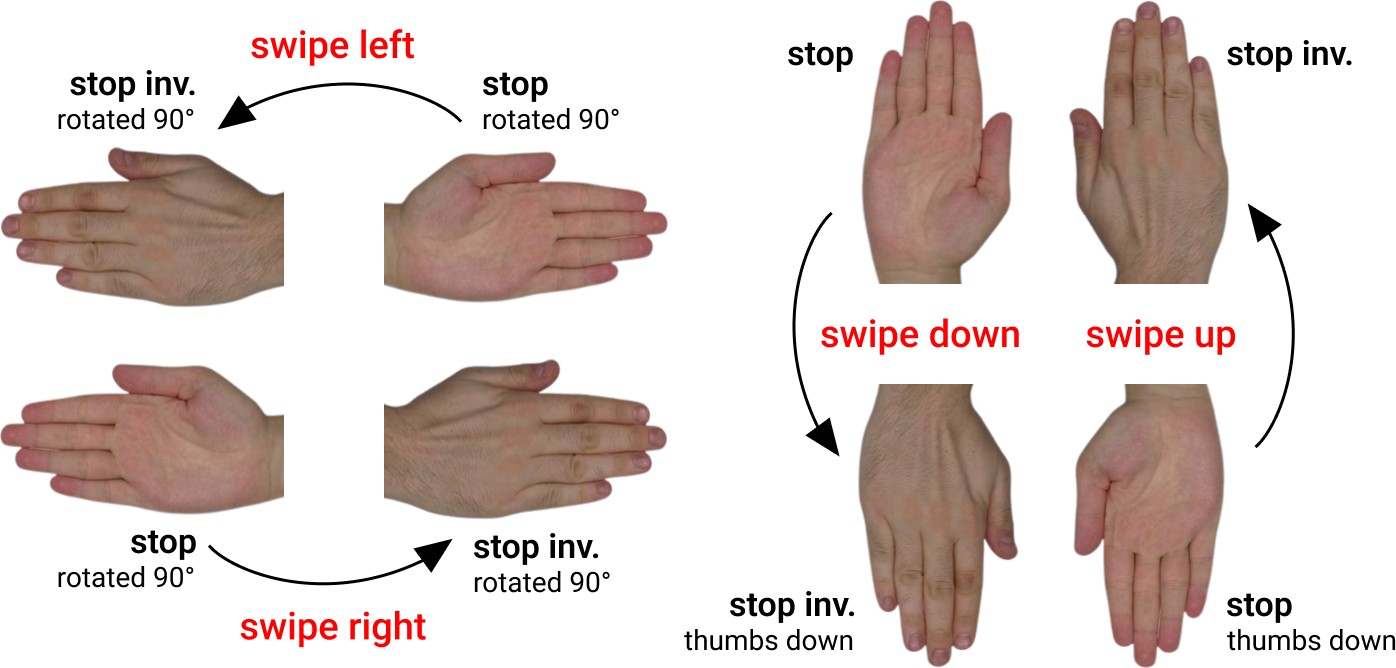


Рисунок 2. Примеры разработанных свайпов.



Рисунок 3. Пример образца и его аннотации.

Набор данных собирался в 4 этапа: (1) этап сбора изображений жестов, называемый майнингом, (2) этап проверки, где проверяются правила майнинга и некоторые условия, (3) фильтрация неподходящих изображений, (4) этап аннотации для разметки ограничительных рамок и ведущих стрелок. Этап классификации встроен в конвейеры интеллектуального анализа и проверки путем разделения pools для каждого класса жестов. Для завершения этих этапов мы используем две российские краудсорсинговые платформы: Яндекс.Толока (1, 2 и 4 шага) и ABC Elementary (3 и 4 шага). Отметим, что задания на платформе ABC Elementary выполняются только сотрудниками компании – это необходимо на этапе фильтрации. Использование двух платформ на этапе аннотации позволяет повысить достоверность итоговой аннотации, поскольку два разных аннотатора делают разметку изображений. Платформы аналогично организированы. Таким образом, приведенное ниже описание справедливо для обоих из них. Подробная информация о каждом из шагов приведена ниже.

Добыча. Задача работников толпы состояла в том, чтобы сфотографировать себя с помощью конкретного жеста, указанного в описании задачи. Определяем следующие критерии: (1) аннотатор должен находиться на расстоянии 0,5 – 4 метра от камеры, и (2) рука с жестом (т.е. ведущая рука) должна быть полностью в кадре. Пример майнинга задача находится на рисунке [6](#_bookmark52) Приложения [А](#_bookmark51). Иногда испытуемые получают задание сделать фотографию в условиях низкой освещенности или против яркого источника света, чтобы сделать нейронную сеть устойчивой к экстремальным условиям. Мы блокировали исполнителей за быстрые ответы (менее 3 секунд) на этапе майнинга и все остальные. Все полученные изображения также были проверены на наличие дублей с помощью сравнения хэшей изображений [[24](#_bookmark45)]. Задания по добыче сопровождались инструкциями с предупреждением о дальнейшей публикации фотографий рабочих толпы.

1. Ратификация. Мы реализовали этап проверки для достижения высокой достоверности изображений, потому что пользователи пытались обмануть систему на этапе майнинга. Удалось установить, что в течение некоторого времени они отправляли ложную информацию и использовали тактику обмана, для экзаменаe, отправляя шаблонную фотографию из описания задачи, дубликат фотографии из одного из своих предыдущих заданий, третьей стороне картинка, фотография с неправильным жестом или фотография с рукой, которая не полностью находится в кадре. Также есть несколько обнаруженных несчастных случаев, когда работник по ошибке загрузил повернутое изображение или фотографию с недействительным жестом. На рисунке [4](#_bookmark54) указано, сколько данных остается после каждого из этапов.

Целью этапа валидации является подбор корректно выполненных изображений на этапе майнинга, т.е. классификация с классами «правильно», «неправильно», «повернуто», «сторонняя фотография» и «фото не лоад». В набор данных попадают только «правильные» изображения. См. рисунок [6](#_bookmark52) для экзамена по проверочному заданию. Условия освещения и расстояние от камеры до объектов не проверяются из-за неявности этих характеристик. Без сомнения, результирующий набор данных неоднозначен по этим параметрам, поскольку все субъекты делали фотографии в разных местах.

Для каждого изображения на этом этапе мы устанавливаем в системе динамическое перекрытие от 3 до 5 исполнителей, т.е. каждое задание выполнялось не менее чем 3 работниками толпы в зависимости от их мастерства. Основываясь на правиле большинства, некоторые фотографии были отклонены. В противном случае изображение считается правильным, и тогда оно будет передано на стадию фильтрации.

Рабочие должны были завершить учебный пул из 10 заданий и экзаменационный пул с 10 назначениями. Работники толпы должны пройти обучение с точностью не менее 80% и экзаменационный пул с точностью не менее 75%, чтобы получить доступ к экзаменационному пулу и задачам проверки соответственно. Стоит отметить, что основной пул валидации содержит задачи по предварительному избавлению работников толпы от обмана. Те, кто достиг низкой производительности по контрольным задачам, исключаются, а их аннотации отклоняются автоматически.

2. Фильтрация. Изображения детей, людей без одежды и изображения с надписями были удалены из Ha-GRID на этом этапе по этическим соображениям. Этап фильтрации осуществляется только на платформе ABC Elementary наличие сомнительного контента. Мы используем очень строгое правило для стадии фильтрации – 5 рабочих должны фильтровать каждый из них (пример задачи фильтрации показан на рисунке [6](#_bookmark52) Приложения [А](#_bookmark51)). Ответ принимается, если за него положительно проголосовали не менее 4 работников. Подобно этапу валидации, аннотаторы проходят тщательный экзамен, обучение и контрольные задания на этапе фильтрации.

3. Аннотация. На этом этапе работники толпы должны нарисовать красную рамку вокруг жеста на каждом изображении и зеленую рамку вокруг руки без гез-ture, если она полностью находится в кадре (пример показан на рисунке [7](#_bookmark53)). Для их дальнейшего перевода в этикетки нужны разные цвета. Две платформы позволяют сохранять ограничительные рамки в форме COCO при[8](#_bookmark12) с нормированными относительными координатами.

Работники толпы должны были завершить бывший пул с 75% успехом, как и на этапе валидации. Задачи Control используются и на этом этапе. Здесь корректность исполнения проверяется пересечением через объединение (IoU).

Перекрытие аннотаций динамически размещается на Yandex. Толока от 3 до 5 и она всегда равна 5 на платформе ABC Elementary. Это используется для снижения вероятности неправильной разметки. Варианты, которые не были агрегированы после максимального перекрытия, не будут использоваться в наборе данных. Все наценки, количество которых варьируется от 8 до 10, кол-селектируются с двух платформ и агрегируют по двум схемам

жесткая и мягкая агрегация.

Для жесткой агрегации используется следующий алгоритм. Главное правило для этого этапа заключается в том, что мы пропускаем ag-gregation и теряем возможность найти результирующую ограничительную рамку, если один из следующих шагов не удовлетворен (каждый im-age соответствует нескольким разметкам, которые будут сравниваться):

1. Проверьте количество коробок длябу-динга на равенство.
2. Найдите центроиды с помощью сдвига средств для всех ограничительных прямоугольников и объедините ближайшие объекты с группами. Затем проверьте количество полей в каждой группе (не менее 0*.* 7*n* должно оставаться в группе, где *n* - количество наценок).
3. Проверьте метки на ограничительные рамки в каждой группе на предмет равенства.
4. Проверьте IoU для каждой группы и сравните с пороговым значением (по умолчанию 0,7).

Если все этапы жесткого конвейера успешно пройдены, конечные наценки усредняются. Таким образом, результат агрегирования для каждого изображения состоит из списка ограничительных рамок и их меток. Если все разметки изображений не имеют ограничительных рамок, то изображение удаляется из набора данных.  Пример жесткой агрегации и схему конвейера см. в верхней части рисунка [8](#_bookmark55) и слева на рисунке 9 соответственно.

Конвейер мягкой агрегации используется в случае сбоя алгоритма жесткой агрегации . Мягкая агрегация может быть понята

Все сотрудники нашей компании осведомлены о прогиби о передаче персональных данных третьим лицам и в качестве предварительной обработки бытьпередней жесткой агрегацией. Он не требует таких условий, как одинаковое количество коробок и ручек. Этапы мягкой агрегации, следующие:

1. Уберите близкие к точкам поля (т.е. слишком маленькие, вероятно, нарисованные наполовину), дубликаты в пределах одного варианта из n.
2. Находят центроиды такими же, как при жесткой агрегации.
3. Удалите «выбросы» (т.е. коробки, которые не были включены в группу).
4. Проверьте метки для каждой ограничительной рамки в группе. Цель этого состоит в том, чтобы найти правильную метку в группе и изменить правильные метки на правильные. Метка считается правильной, если она соответствует 70 процентам ответов испытуемых или более.
5. Проверьте количество ограничительных полей на равенство в каждой группе: если количество полей в группе меньше n, то мы пытаемся добавить недостающие поля, агрегируя оставшиеся, которые должны составлять не менее 70% от n. Совокупный результат помещается вместо отсутствующего ящика.
6. Еще раз предпринята попытка жесткой агрегации.
7. Если это не удается, мы итеративно устраняем некоторые из n вариа-муравьев разметки (70% n должны остаться) и пытаемся жестко агрегировать остальные.

См. нижнюю часть рисунка [8](#_bookmark55) и справа на рисунке [9](#_bookmark56) пример мягкой агрегации и его схему конвейера, респективно. Нет, мы не платим работнику, если его вариант разметки неправильный [9](#_bookmark14) (т.е. в разметке отсутствует коробка или есть лишняя коробка или коробки неправильного цвета).

Мы также попросили испытуемых на платформе ABC Elementary аннотировать ведущую руку для каждого изображения, чтобы решить задачу определения стороны руки. Причина этого в том, чтобы уметь интерпретировать динамические жесты статическими. Как обсуждалосьвыше, динамические жесты «свайп вверх» и «свайп вниз» могут быть показаны одной рукой, в то время как жесты «свайп вправо» и «свайп влево» трудно показать без использования второй руки. Если горизонтальные статические жесты «стоп» и «стоп инвертированный» показаны левой рукой, то это динамический жест «свайп вправо», в противном случае – «свайп влево». Ведущие ручные этикетки могут быть использованы для разработки двух жестов, отличающихся от одного, как в [[17](#_bookmark38)]. Например, справа «три» и слева «три» могут быть два разных жеста «правая тройка» и «левая тройка».

Характеристики набора данных

Размер HaGRID составляет примерно 716 ГБ – он включает в себя более 550 тысяч изображений, разделенных на 18 классов жестов: «звонок », «неприязнь», «кулак», «четыре», «нравится», «немой»,«ок», «один», «пальма», «мир», «мир перевернутый», «камень», вверх перевернутый» (показан на рисунке [1](#_bookmark0)). Каждый класс жестов содержит более 30 000 RGB-изображений с высоким разрешением [10](#_bookmark15). Каждое изображение имеет ограничительную рамку с жестом. Кроме того, некоторые im-ages имеют ограничительную рамку «без жеста», если в кадре есть секундная стрелка. Этот дополнительный класс содержит 123 589 sam- ples.

Dataset содержит 34 730 уникальных лиц и, по крайней мере, столько уникальных сцен. Испытуемыми являются люди от 18 до 65 лет. Соотношение женщин и мужчин составляет около 27 к 20. Набор данных был собран в основном в помещении со значительными различиями в освещении, включая искусственное и естественное освещение. Кроме того, набор данных включает в себя изображения, сделанные в экстремальных условиях, таких как облицовка и задняя часть окна. Также испытуемые должны были демонстрировать жесты на разных расстояниях от камеры (рисунок [4](#_bookmark13)).

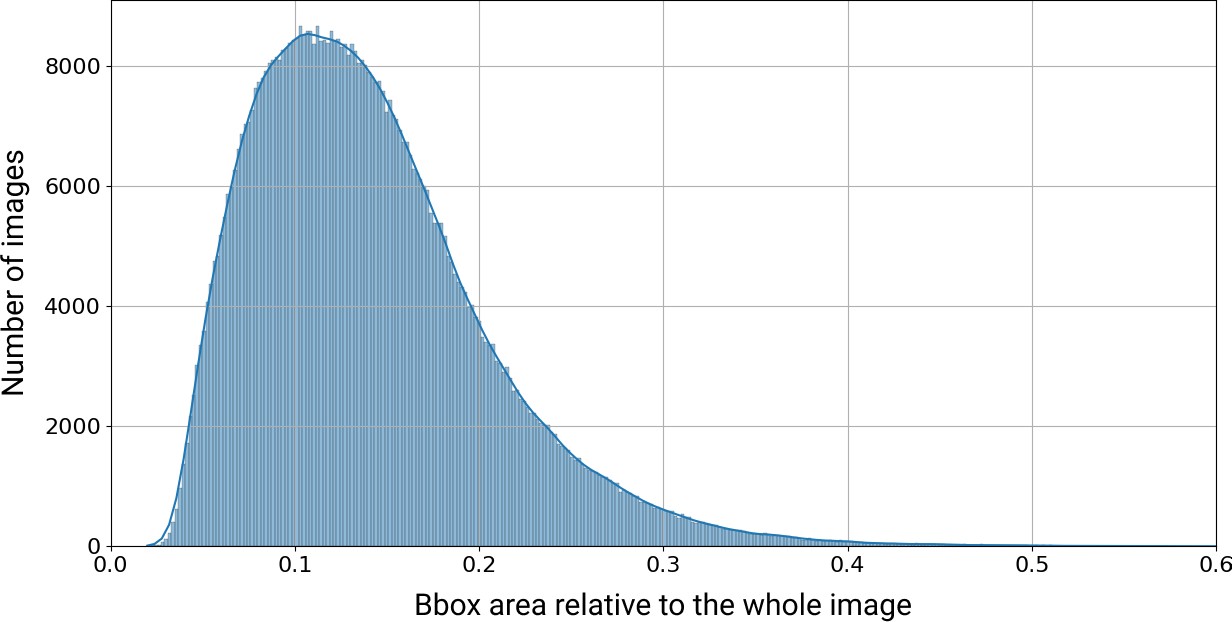


Рисунок 4. Распределение области ограничительной рамки относительно всего изображения. Коробки занимают от 5 до 25% изображения, что означает, что camera в основном находится далеко от человека.

Разделение набора данных

Данные были разделены на обучающие (92%) и тестовые (8%) наборы по субъектам, в среднем 28 300 изображений на класс (всего 509 323) и 2 400 изображений на класс (43 669 до таль), соответственно. Подробная информация о количестве изображений приведена в таблице [2](#_bookmark16). Число испытуемых в учебных и тестовых наборах равно 31 952 и 2 778, соответственно. Кроме того, в файл аннотации добавлен хэш анонимного идентификатора пользователя . Это позволит пользователю набора данных самому разделить набор данных train-val. Поскольку размер набора данных большой, мы разработали небольшую версию (100 образцов на класс) HaGRID с аннотациями для предварительного просмотра по ссылке для вашего удобства.

Эталонная оценка

Чтобы проверить эффективность набора данных, оцениваются 8 популярных архитектур для 2 задач HGR: де-текция рук и классификация жестов рук. В качестве детектора мы выбрали SS-DLite [[27](#_bookmark48)] и набор из 7 архитектур , состоящих из «стоп», «стоп перевернутый», «три», «три2», «два вверх», «два вверх», «два выбора друого варианта. как ResNets [[25](#_bookmark46)]. Однако это может быть важным преимуществом для больших моделей, например, визуальных трансформаторов [[26](#_bookmark47)].

|  |  |  |  |
| --- | --- | --- | --- |
| Жест | Поезд + Вал | Тест | Итог |
| звать | 28,193 | 2,436 | 30,629 |
| неприязнь | 28,537 | 2,551 | 31,088 |
| кулак | 27,764 | 2,506 | 30,270 |
| Четыре | 28,880 | 2,492 | 31,372 |
| любить | 27,721 | 2,415 | 30,136 |
| немой | 28,971 | 2,438 | 31,409 |
| Хорошо | 27,999 | 2,391 | 30,390 |
| Один | 28,444 | 2,462 | 30,906 |
| ладонь | 28,326 | 2,427 | 30,753 |
| мир | 28,303 | 2,477 | 30,780 |
| мир перевернутый | 27,864 | 2,278 | 30,142 |
| скала | 27,782 | 2,413 | 30,195 |
| остановка | 27,963 | 2,400 | 30,363 |
| стоп инвертированный | 28,857 | 2,469 | 31,326 |
| Три | 28,015 | 2,408 | 30,423 |
| три2 | 27,789 | 2,332 | 30,121 |
| два вверх | 29,679 | 2,545 | 32,224 |
| два перевернутых вверх | 28,236 | 2,229 | 30,465 |
| **всего изображений** | **509,323** | **43,669** | **552,992** |
| без жеста | 112,740 | 10,849 | 123,589 |
| **всего коробок** | **622,063** | **54,518** | **676,581** |

*h*

Таблица 2. Разделение HaGRID на тренировочные и тестовые наборы. Обратите внимание, что количество изображений не равно количеству ограничительных рамок, поскольку изображение может содержать две руки и, следовательно, две коробки.

ResNet-18, ResNet-152 [[25](#_bookmark46)], ResNeXt-50, ResNeXt- 101 [[28](#_bookmark49)], MobileNetV3 маленький, MobileNetV3 большой [[29](#_bookmark50)] и ViT-B/32 [[26](#_bookmark47)] в качестве классификаторов.

Настройка эксперимента обнаружения

Модель SSDLite с большой магистралью MobileNetV3 [[29](#_bookmark50)] использовалась для решения проблемы обнаружения рук.

Из-за большого размера набора данных модель была обучена с нуля, используя SGD с импульсом 0,9 и весовым затуханием 0,0005 в качестве оптимизатора. Скорость обучения начинается от 0,005 с шагом затухания скорости обучения в 3 эпохи. Детектор достиг наивысшей метрики в эпоху 67.

Настройка классификационного эксперимента

Все классификаторы обучаются на обрезанных изображениях со шкалой ran-dom от 1 до 2. Все приведенные ниже метрики были рассчитаны на тестовом наборе, который содержит 43 669 изображений (около 2 400 на класс жестов[11](#_bookmark17)).

В каждую модельный костяк были добавлены две головки: для классификации жестов рук на 19 классов и для двоичной классификации ведущих рук в 2 класса. Модель предварительно диктует класс жеста (или «без жеста») и сторону руки для каждой руки, чтобы упростить конвейер. Суммируем выходные данные функции cross-entropy Loss для соответствующих задач.

В таблице [3](#_bookmark18) представлены результаты оценки выбранных модельных архитектур для решения трех задач – детекции рук, жестов и задач классификации ведущих рук. Лучший классифицированный жестпо результату получается моделью ResNeXt-101 и лучший результат классификации рук ResNet-152. Несмотря на снижение мет-рик , архитектуры MobileNetV3 являются наиболее подходящими мод-элами для нас из-за их легкости и скорости.

Нарисунке [5](#_bookmark19) представлена матрица путаницы наилучшего результата классификации жестов рук, полученного ResNeXt- 101. Можно сделать вывод, что все классыхорошо отделены друг от друга. Как и ожидалось, существуют пары похожих жестов, например, «нравится» и «звонок», «три» и «четыре», «стоп» и «ладонь», которые порождают несколько ошибок в прогнозе. Проблема уменьшения количества ошибок с классом «без жеста» будет решена путем добавления в набор данных образцов с естественными движениями рук, похожими на целевые жесты .

Демо-версию системы распознавания жестов с одним из самых легких и быстрых классификаторов MobileNetV3 large можно найти в нашем репозитории.

Обсуждение

**Модели.** У нас не было цели обучать лучших мод-элов с учетом набора данных, так как в этой статье предлагалось представить HaGRID и конвейер его создания. Полученные результаты являются отправной точкой для исследования возможностей набора данных влиять на производительность системы HGR. Но даже сейчас метрики показывают высокое качество набора данных.

Одним из направлений следующих исследований является добавление дополнений данных для включения редких случаев жестикуляции. Необходимо попробовать такие дополнения , как трансляция и вращение. Мы ожидаем, что эти манипуляции усложнят задачу, но в то же время снизят ограничения на использование

Метрика

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Модель | Размер модели (МБ) | Параметры (М) | Время вывода (мс) |  | | |
| Жесты | Ведущая рука | карта |
| РесНет-18 | 44.8 | 11.19 | 64.5 | 98.72 | 99.27 | - |
| РесНет-152 | 233.7 | 58.19 | 411.5 | 99.11 | **99.45** | - |
| ResNeXt-50 | 92.5 | 23.02 | 136.1 | 98.99 | 99.39 | - |
| ResNeXt-101 | 348.2 | 86.79 | 581.4 | **99.28** | 99.28 | - |
| Мобильная сетьV3 маленький | 8.7 | 2.13 | 8.9 | 96.78 | 98.28 | - |
| Мобильная сетьV3 большой | 22.0 | 5.46 | 16.9 | 97.88 | 98.58 | - |
| ВиТ-Б/32 предварительно обученный | 353.1 | 88.24 | 166.4 | 98.49 | 99.13 | - |
| Мобильная сетьV3 большой + ССДЛайт | 10.1 | 2.46 | 51.1 | - | - | 71.49 |

Таблица 3. Моделирование результатов обучения на HaGRID. В качестве классификационной метрики был выбран балл F1. Процессор Intel(R) Core(TM) i9-9880H с тактовой частотой 2,30 ГГц используется для вычисления времени вывода.

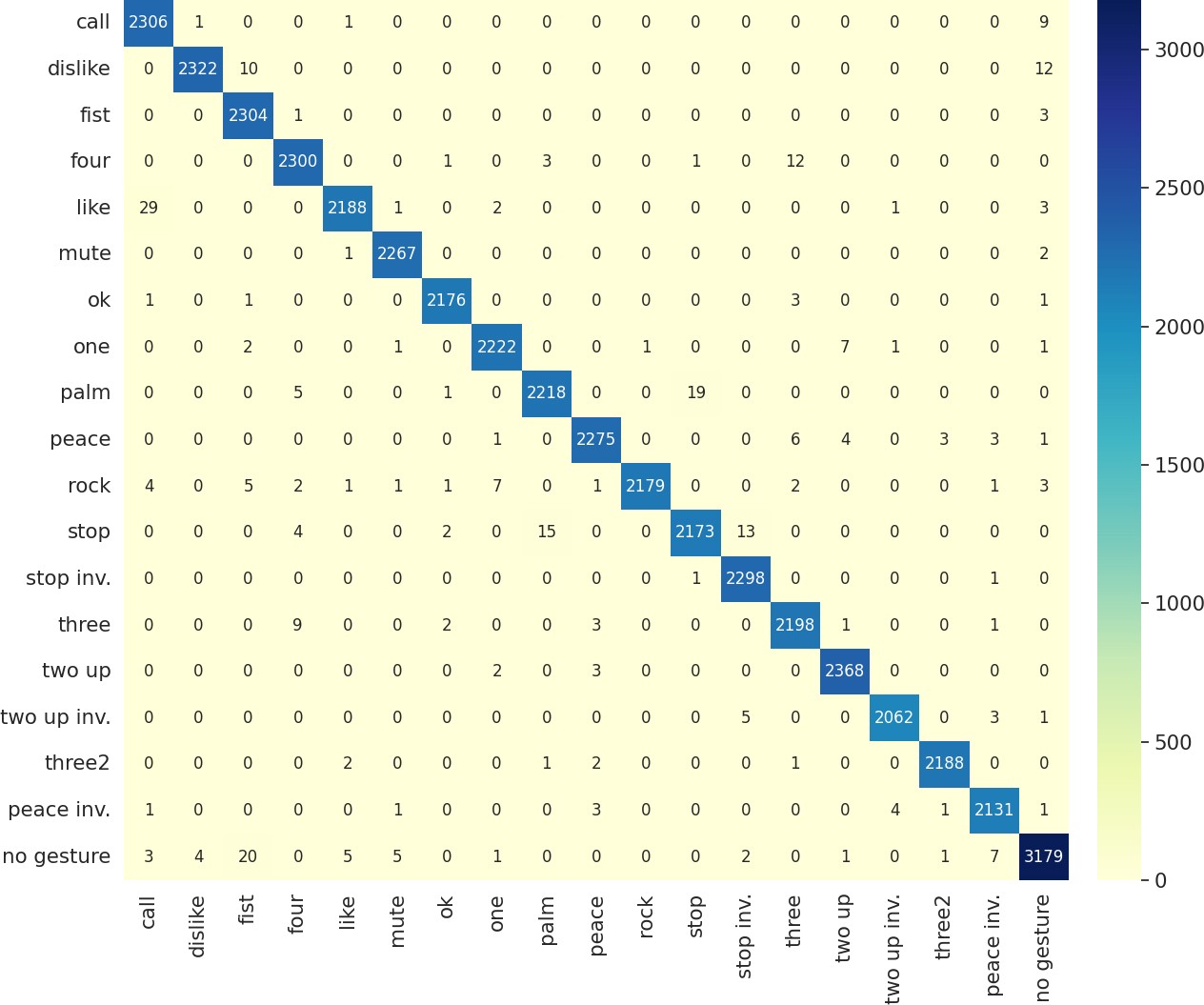


Рисунок 5. Путаница в матрице классификации жестов на Ha-GRID с использованием модели ResNeXt-101.

Система HGR. Другим направлением исследований является квантование моделей, поскольку важно запускать систему HGR на устройствах. Перспективными направлениями исследований также считаются квантование-ориентированное обучение (QAT) и нейронный архитектурный поиск (NAS ).

**Набор данных.** HaGRID предназначен для использования в ручных системах распознавания. Кроме того, набор данных можно использовать для поиска ведущей руки. Использование ведущей стрелки также может позволить увеличить количество жестов в 2 раза, чтобы соответствовать большому количеству компьютерных ответов. Наша работа по этой теме заключается в увеличении размера набора данных путем добавления новых статических жестов и образцов с natu-ral поведением рук пользователей, похожим на целевые жесты. Важно отметить, что мы планируем расширить разметку с помощью аннотаций additional, таких как пол, маски seg-mentation, ключевые моменты и так далее. Кроме того, мы скоро собираемся представить новые наборы данных для воспроизведения изображений и видео и некоторых популярных задач компьютерного зрения.

Кроме того, мы надеемся, что HaGRID, с любой дополнительной разметкой, может быть полезным набором данных для многих компьютерного зрения.

Заключение

В этой статье мы представляем набор данных HAnd Gesture Recogni- tion под названием HaGRID, который является одним из крупнейших и самых разнообразных с точки зрения предметов и коллекционных кондиций. Набор данных жестов в основном предназначен для использования в устройствах управления системой, но потенциал его применения довольно огромен. HaGRID является самым сложным набором данных, сопоставленным с другими,поскольку он был собран примерно из 35 000 сцен с различным освещением и расстоянием до камеры. Кроме того, предложены исходные условия оценки задач ГГР. Весь набор данных, пробная версия со 100 изображениями на класс, предварительно обученные модели и demo находятся в открытом доступе в репозитории[12](#_bookmark25) или зеркале[13](#_bookmark26).

Ссылки

1. С. Клаф и М. К. Дафф. Роль жеста в коммуникации и познании: последствия для понимания и лечения нейрогенных расстройств коммуникации. *Границы в нейробиологии человека*, 14, 2020.
2. К. А. Пикеринг, К. Дж. Исследование технологий распознавания жестов рук и приложений для взаимодействия человека с транспортным средством. В *2007 году 3-я Конференция по инженерии и технологиям по автомобильной электронике*, страницы 1-15, 2007.
3. Ф. Парада-Луара, Э. Гонз алез-Агулла и Дж. Л. Альба-Кастро. Жесты рук для управления информационно-развлекательным оборудованием в автомобилях. В *2014 IEEE Intelligent Vehicles Symposium Proceedings*, страницы 1–6, 2014.
4. П. Н. Арати, С. Артика, С. Понмитра, К. Шринивасан, и

В. Руккумани. Система домашней автоматизации на основе жестов. В *2017 году Международная конференция nextgen Electronic Tech- nologies: Silicon to Software (ICNETS2),* стр. 198–201, 2017.

1. Х. С. Хасан и С. А. Карим. Взаимодействие человека с компьютером для распознавания жестов рук на основе зрения: опрос. В
2. Л. Чэнь, Ф. Ван, Х. Дэн и К. Цзи. Опрос по распознаванию жестов рук. В *2013 году Международная конференция по компьютерным наукам и приложениям*, страницы 313–316, 2013.
3. З. Халим и Г. Аббас. Система распознавания жестов на языке жестов на языке жестов на основе кинекта для лиц с нарушениями слуха и речи: пилотное исследование пакистанского языка жестов. *Ассистивные технологии*, 27:34–43, 2015.
4. Ашиш С. Никам и Аарти Г. Амбекар. Двуязычное распознавание знаков с использованием техники жестов рук на основе изображений для людей с нарушениями слуха и речи. В *2016 году Международная конференция по управлению вычислительной связью и ау-томации (ICCUBEA),* страницы 1–6, 2016.
5. Т. Старнер, Д. Уивер, Д. Ченг и А. Пентланд. Распознавание американского языка жестов в режиме реального времени с использованием настольного и носимого компьютерного видео. *IEEE Transactions on Pattern Anal- ysis and Machine Intelligence*, 20(12):1371–1375, 1998.
6. Т.К. Ким, С.Ф. Вонг и Р. Чиполла. Тензорный канонический корреляционный анализ для классификации действий. В *2007 ieee Con- ference on Computer Vision and Pattern Recognition*, стр. 1–8, 2007.
7. С. Эскалера, Дж. Гонц а'лез, X. Баро', М. Реда, О. Лопес
   1. Гийон, В. Атитсос и Х. Эскаланте. Мультимодальное распознавание жестов 2013: набор данных и результаты. *Труды 15-й АКМ по Международной конференции по мультимодальному взаимодействию*, стр. 445–452, 2013.
8. J. Wan, Y. Zhao, S. Zhou, I. Guyon, S. Escalera и S. Z. Литий. Chalearn смотрит на людей rgb-d изолированные и непрерывные наборы данных для распознавания жестов. В *2016 году IEEE Con- ference on Computer Vision and Pattern Recognition Work- shops (CVPRW),* 2016.
9. П. Молчанов, X. Ян, С. Гупта, Ким К., С. Тайри, и

Ю. Каутц. Онлайн-обнаружение и классификация динамических жестов рук с рекуррентной 3d сверточной нейронной сетью. В *2016 году IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* страницы 4207–4215, 2016.

1. Ю. Чжан, К. Цао, Д. Чэн и Х. Лу. Эгогестура: Новый набор данных и эталон для эгоцентрического распознавания жестов рук . *IEEE Транзакции на мультимедиа*, 20(5):1038–1038, 2018.
2. Г. Латиф, Н. Мохаммад, Дж. Альгазо, Р. Аль-Халаф, и

Р. Аль-Халаф. Arasl: набор данных языка жестов арабского алфавита.

*Краткие данные*, 23:103777, 2019.

1. Г. Бенитес-Гарсия, Х. Олivарес-Меркадо, Г. С а'нчез-П ерез и К. Янай. Ipn hand: набор видеоданных и эталон для непрерывного распознавания жестов рук в режиме реального времени. В *2020 году 25-я Международная конференция по экопознанию паттерна R(ICPR),* стр. 4340–4347, 2021.
2. К. Нуцци, С. Пасинетти, Р. Пагани, Г. Коффетти и Дж. Сансони. Руки: набор данных rgb-d статических жестов рук для взаимодействия человека и робота. *Вкратце*, 35:106791, 2021.
3. Деян Чандра Гопе. Взаимодействие жестов рук с человеком-компьютером. *Глобальный журнал компьютерных наук и технологий*, 2012.
4. С. Каррино, М. Каон, О. Абу Халед, Р. Ингольд, и

Э. Мугеллини. Функциональные жесты для взаимодействия человека и окружающей среды. том 8007, 2013.

1. М. Ремпель Дэвид, Д. Камиллери Мэтт и Л. Ли Дэвид. Дизайн жестов рук для человеко-компьютерного взаимодействия: Уроки от сурдопереводчиков. *Международная конференция человеко-компьютерных исследований*, 72 10-11:728–735, 2014.
2. М. Карам и М. К. Шрефель. Таксономия жестов в взаимодействиях человека с компьютером. 2005 год.
3. Руки медиапайпа. [https://solutions.mediapipe.](https://solutions.mediapipe.dev/hands)  [разработчик/руки](https://solutions.mediapipe.dev/hands), 2019.
4. С. Руффье, Д. Лаланн, Э. Мугеллини и О. Абу Халед. Обзор наборов данных для распознавания жестов человека. В *человеко-компьютерном взаимодействии.* *Расширенные модальные измерения взаимодействия и методы*, страницы 337–348, 2014.
5. Заунер К. Реализация и бенчмаркинг хэш-функций перцептивных изображений. 2010 год.
6. К. Хе, Х. Чжан, С. Жэнь и Д. Сунь. Глубокое остаточное обучение для распознавания изображений. В *2016 году IEEE Conference on Com-puter Vision and Pattern Recognition (CVPR),* страницы 770–778, 2016.
7. А. Досовицкий, Л. Бейер, А. Колесников, Д. Вайсенборн,

X. Чжай, Т. Унтертинер, М. Дехгани, М. Миндерер,

Г. Хейгольд, С. Гелли, Дж. Ушкорейт и Н. Хоулсби. Изображение стоит 16x16 слов: Трансформеры для распознавания изображений в масштабе. *ArXiv*, abs/2010.11929, 2021.

1. М. Сэндлер, А. Говард, М. Чжу, А. Жмогинов, Л. -К. Чен. Mobilenetv2: Перевернутые остатки и линейные узкие места. В *2018 году IEEE/ CVF Conference on Computer Vision and Pattern Recognition*, страницы 4510–4520, 2018.
2. С. Се, Р. В. Гиршик, П. Долл а'р, З. Ту и К. Он. Ag-грегированные остаточные преобразования для глубоких нейронных сетей. *Конференция IEEE по компьютерному зрению и распознаванию* образов *(CVPR)* 2017 года, страницы 5987–5995, 2017.
3. А. Говард, М. Сандлер, Б. Чэнь, В. Ван, Л.-К. Чен

М. Тан, Г. Чу, В. Васудеван, Ю. Чжу, Р. Панг, Х. Адам, К. Ле. Поиск mobilenetv3. В *2019 году IEEE/CVF International Conference on Computer Vision (ICCV),* стр. 1314–1324, 2019.

# Приложения

**A. Информация о сборе данных**

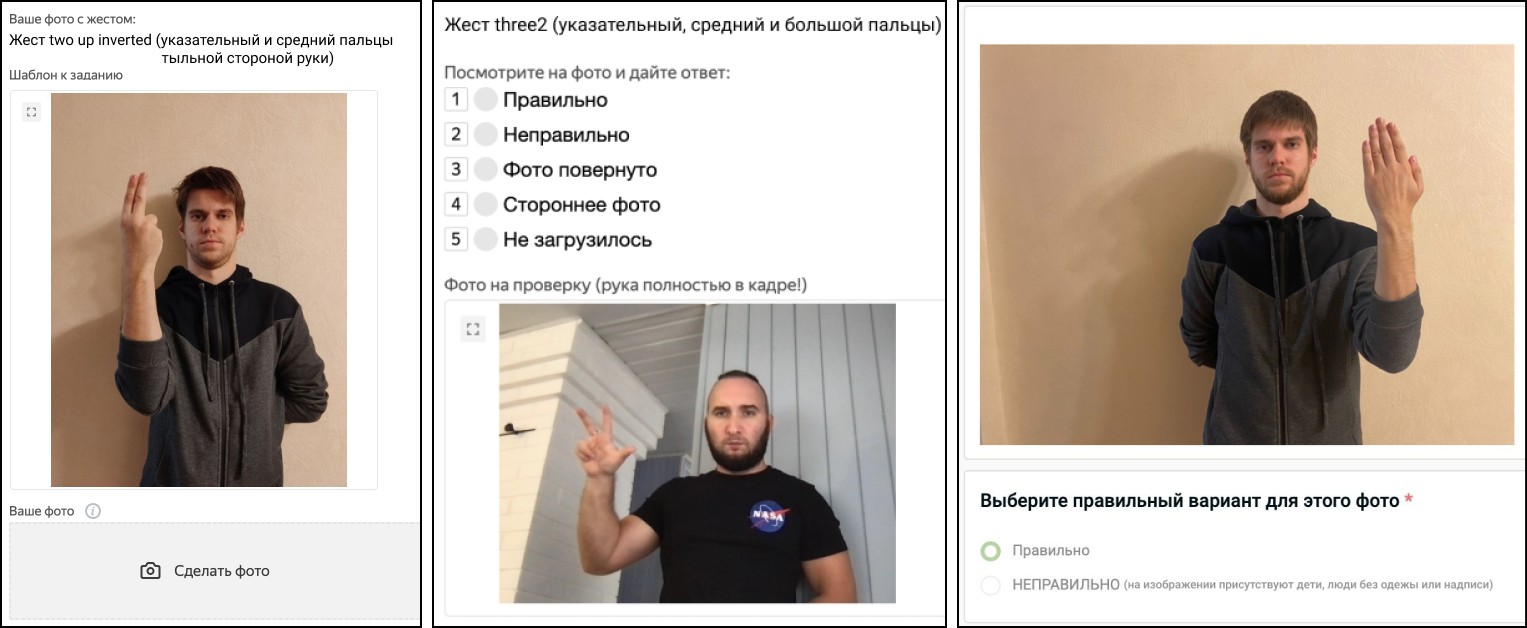


Рисунок 6. Веб-интерфейс для задач майнинга (слева), задач проверки (посередине) и задач фильтрации (справа).

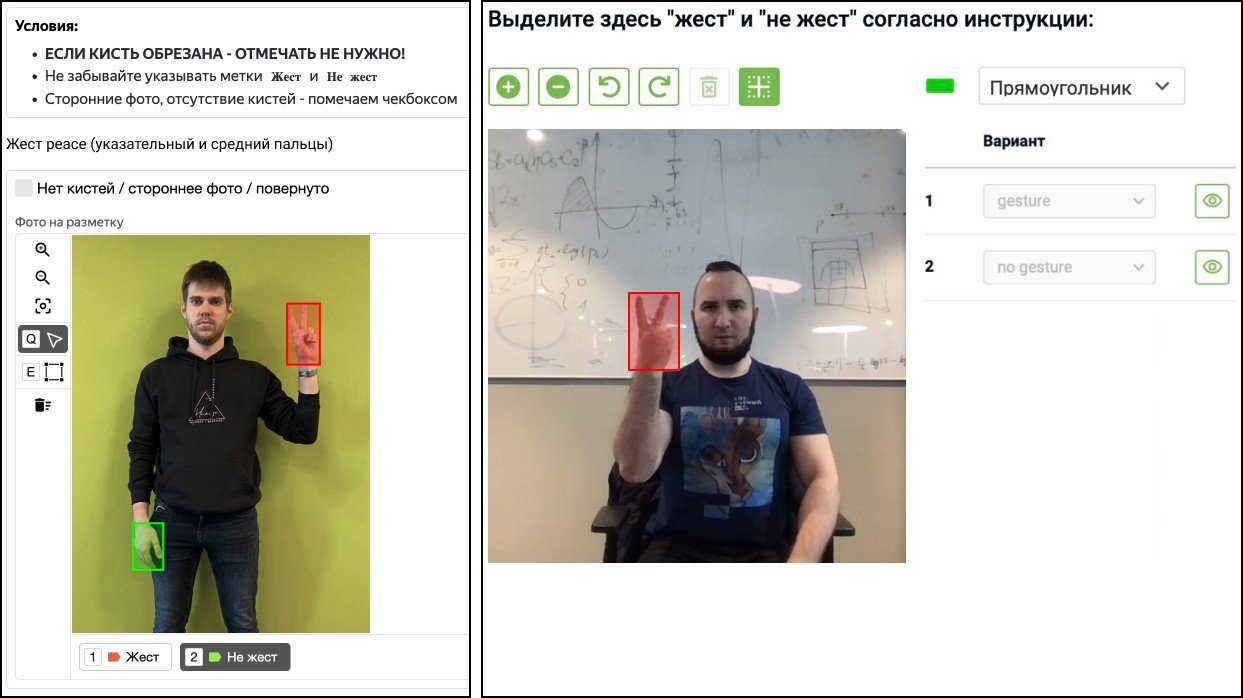


Рисунок 7. Веб-интерфейс для задач аннотации. Интерфейс Яндекс.Толока (слева). Abc Элементарный интерфейс (справа).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Жест | Добыча | Ратификация | Фильтрация | Аннотация |
| звать | 55,502 | 36,516 | 32,325 | 30,629 |
| неприязнь | 53,889 | 39,752 | 33,252 | 31,088 |
| кулак | 48,157 | 39,724 | 32,215 | 30,270 |
| Четыре | 50,590 | 39,376 | 33,408 | 31,372 |
| любить | 49,674 | 37,964 | 32,457 | 30,136 |
| немой | 49,056 | 38,782 | 33,725 | 31,409 |
| Хорошо | 63,017 | 37,305 | 32,460 | 30,390 |
| Один | 47,566 | 37,666 | 32,819 | 30,906 |
| ладонь | 47,507 | 38,299 | 32,954 | 30,753 |
| мир | 47,870 | 38,228 | 32,874 | 30,780 |
| мир перевернутый | 52,800 | 36,711 | 31,676 | 30,142 |
| скала | 48,374 | 37,175 | 32,157 | 30,195 |
| остановка | 50,564 | 38,120 | 32,888 | 30,363 |
| стоп инвертированный | 54,468 | 36,811 | 32,647 | 31,326 |
| Три | 47,616 | 37,373 | 32,423 | 30,423 |
| три2 | 54,479 | 36,319 | 31,012 | 30,121 |
| два вверх | 55,433 | 38,276 | 33,591 | 32,224 |
| два перевернутых вверх | 48,183 | 36,287 | 31,892 | 30,465 |
| **итог** | **924,745** | **680,684** | **586,775** | **552,992** |

Таблица 4. Количество образцов, которые остаются после каждого этапа .

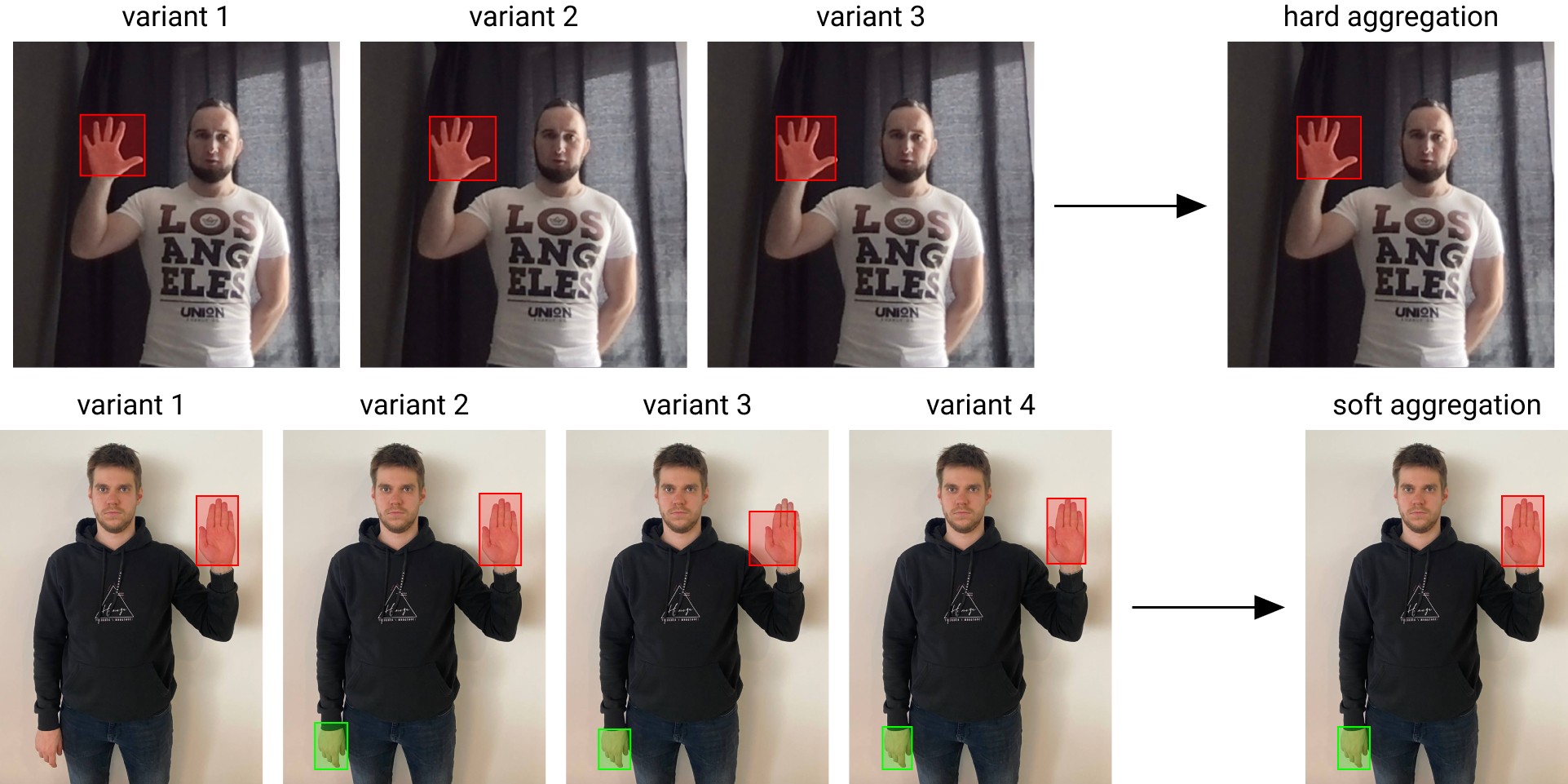


Рисунок 8. Примеры жесткой и мягкой агрегации. (вверху) Жесткая агрегация: изображение имеет три варианта разметки с одинаковым количеством коробок, которые принадлежат к одной группе; все поля помечены одним class и их группа *IoU >* 0*.* 7. (внизу) Мягкая агрегация: первые три варианта изображения не были усреднены с использованием жесткой агрегации, поэтому был добавлен 4-й вариант. Ограничительная рамка «Выбросы» с меткой жеста из третьего варианта была удалена. После этого варианты 1 и 3 имеют по одному отсутствующему блоку. В варианты были добавлены недостающие поля. Жесткая агрегация была применена снова, и результат был положительным.

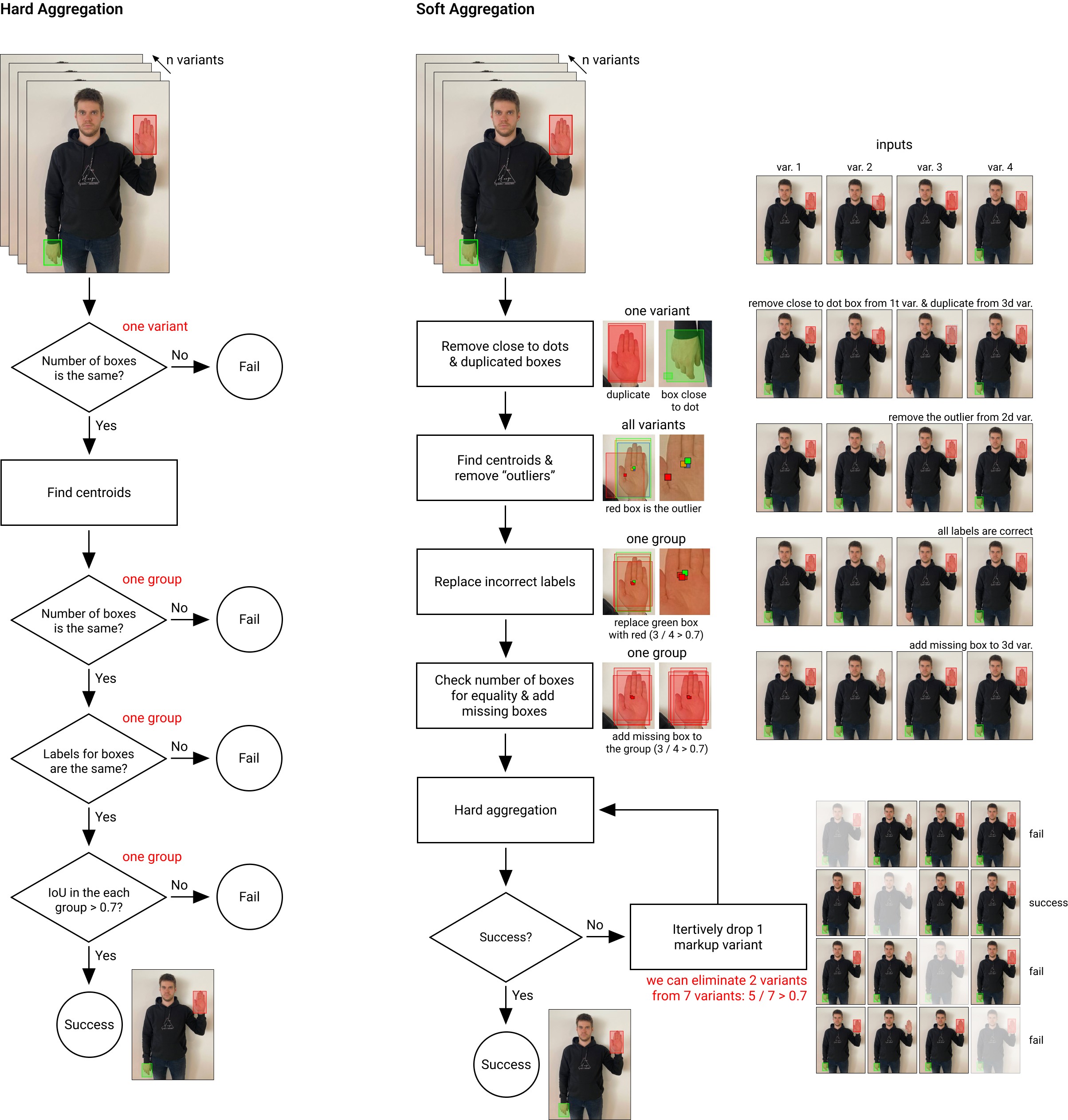
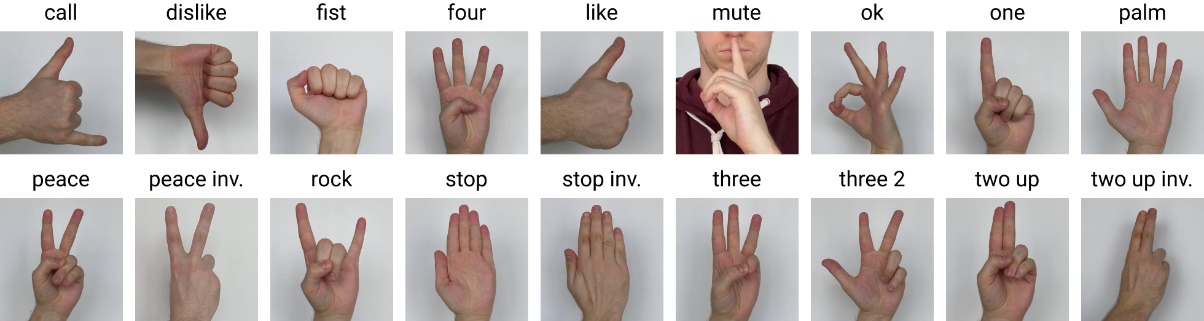


Рисунок 9. Трубопроводы жесткой и мягкой агрегации. (слева) Жесткая агрегация: результат агрегируется только после успешного прохождения всех проверок. (справа) Мягкая агрегация: после каждой неудачной проверки предпринимаются корректирующие действия.

Приложение (код)

**Мы** представляем большой набор данных изображений **HaGRID (набор** данных **изображений HA** и распознавания жестов) для систем распознавания жестов рук (HGR). Вы можете использовать его для классификации изображений или задач обнаружения изображений. Предлагаемый набор данных позволяет создавать системы HGR, которые можно использовать в сервисах видеоконференцсвязи (Zoom, Skype, Discord, Jazz и т. д.), системах домашней автоматизации, автомобильном секторе и т. д.

Размер HaGRID составляет **716 ГБ**, а набор данных содержит **552 992** изображения RGB FullHD (1920 × 1080), разделенных на **18** классов жестов. Также некоторые изображения имеют no\_gestureкласс, если в кадре есть вторая свободная рука. Этот дополнительный класс содержит **123 589** образцов. Данные были разделены на обучающие 92% и тестовые 8% наборов по предметам user\_id, с 509 323 изображениями для поезда и 43 669 изображениями для теста.

[](https://github.com/hukenovs/hagrid/blob/master/images/gestures.jpg)

Набор данных содержит **34 730** уникальных лиц и как минимум такое же количество уникальных сцен. Субъектами являются люди от 18 до 65 лет. Набор данных был собран в основном в помещении со значительными различиями в освещении, включая искусственное и естественное освещение. Кроме того, в набор данных включены изображения, сделанные в экстремальных условиях, например, лицом к окну и спиной к нему. Также испытуемые должны были показывать жесты на расстоянии от 0,5 до 4 метров от камеры.

Установка

Клонируйте и установите необходимые пакеты Python:

git clone

https://github.com/hukenovs/hagrid.git

# or mirror link:

cd hagrid

# Create virtual env by conda or venv

conda create -n gestures python=3.9 -y

conda activate gestures

# Install requirements

pip install -r requirements.txt

Установка докера

docker build -t gestures .

docker run gestures

Загрузки

Мы разбили набор данных поезда на 18 архивов по жестам из-за большого размера данных. Загрузите и разархивируйте их по следующим ссылкам:

| Жест | Размер | Жест | Размер |
| --- | --- | --- | --- |
| [call](https://sc.link/ykEn) | 39,1 ГБ | [peace](https://sc.link/l6nM) | 38,6 ГБ |
| [dislike](https://sc.link/xjDB) | 38,7 ГБ | [peace\_inverted](https://sc.link/mXoG) | 38,6 ГБ |
| [fist](https://sc.link/wgB8) | 38,0 ГБ | [rock](https://sc.link/kMm6) | 38,9 ГБ |
| [four](https://sc.link/vJA5) | 40,5 ГБ | [stop](https://sc.link/gXgk) | 38,3 ГБ |
| [like](https://sc.link/r7wp) | 38,3 ГБ | [stop\_inverted](https://sc.link/jJlv) | 40,2 ГБ |
| [mute](https://sc.link/q8vp) | 39,5 ГБ | [three](https://sc.link/wgBr) | 39,4 ГБ |
| [ok](https://sc.link/pV0V) | 39,0 ГБ | [three2](https://sc.link/vJA8) | 38,5 ГБ |
| [one](https://sc.link/oJqX) | 39,9 ГБ | [two\_up](https://sc.link/q8v7) | 41,2 ГБ |
| [palm](https://sc.link/nJp7) | 39,3 ГБ | [two\_up\_inverted](https://sc.link/r7w2) | 39,2 ГБ |

train\_val **аннотации** :[ann\_train\_val](https://sc.link/BE5Y)

Тест

| Тест | Архивы | Размер |
| --- | --- | --- |
| картинки | [test](https://sc.link/zlGy) | 60,4 ГБ |
| аннотации | [ann\_test](https://sc.link/DE5K) | 3,4 МБ |

Подвыборка

Подвыборка содержит 100 элементов на жест.

| Подвыборка | Архивы | Размер |
| --- | --- | --- |
| картинки | [subsample](https://sc.link/AO5l) | 2,5 ГБ |
| аннотации | [ann\_subsample](https://sc.link/EQ5g) | 153,8 КБ |

или с помощью скрипта Python

python download.py --save\_path <PATH\_TO\_SAVE> \

--train \

--test \

--subset \

--annotations \

--dataset

Выполните следующую команду с ключом --subset, чтобы загрузить небольшое подмножество (100 изображений на класс). Вы можете загрузить подмножество поездов с помощью --trainvalили тестовое подмножество с помощью --test. Скачать аннотации для выбранного этапа по --annotationsключу. Загрузите набор данных с изображениями по --dataset.

usage: download.py [-h] [--train] [--test] [--subset] [-a] [-d] [-t TARGETS [TARGETS ...]] [-p SAVE\_PATH]

Download dataset...

optional arguments:

-h, --help show this help message and exit

--train Download trainval set

--test Download test set

--subset Download subset with 100 items of each gesture

-a, --annotations Download annotations

-d, --dataset Download dataset

-t TARGETS [TARGETS ...], --targets TARGETS [TARGETS ...]

Target(s) for downloading train set

-p SAVE\_PATH, --save\_path SAVE\_PATH

Save path

Модели

Мы предоставляем несколько предварительно обученных моделей в качестве основы с классической архитектурой магистрали и двумя выходными головками — для классификации жестов и классификации ведущей руки.

| Классификаторы | Жесты F1 | F1 Ведущая рука |
| --- | --- | --- |
| [Реснет18](https://sc.link/KEnx) | 98,72 | 99,27 |
| [Реснет152](https://sc.link/O9rr) | 99,11 | **99,45** |
| [ResNeXt50](https://sc.link/GKjJ) | 98,99 | 99,39 |
| [ResNeXt101](https://sc.link/JXmg) | **99,28** | 99,28 |
| [MobileNetV3\_small](https://sc.link/XVEg) | 96,78 | 98,28 |
| [MobileNetV3\_large](https://sc.link/YXG2) | 97,88 | 98,58 |
| [витб32](https://sc.link/XV4g) | 98,49 | 99,13 |

Также мы предоставляем модель SSDLite с большой магистралью MobileNetV3 для решения проблемы обнаружения рук.

| Детектор | карта |
| --- | --- |
| [SSDLite](https://sc.link/YXg2) | 71,49 |

Аннотации

Аннотации состоят из ограничивающих прямоугольников рук в формате COCO [top left X position, top left Y position, width, height]с метками жестов. Также аннотации имеют пометки leading hands( leftили rightдля жеста руки) и leading\_confкак достоверность для leading\_handаннотации. Мы предоставляем user\_idполе, которое позволит вам самостоятельно разделить набор данных train/val.

"03487280-224f-490d-8e36-6c5f48e3d7a0": {

"bboxes": [

[0.0283366, 0.8686061, 0.0757000, 0.1149820],

[0.6824319, 0.2661254, 0.1086447, 0.1481245]

],

"labels": [

"no\_gesture",

"one"

],

"leading\_hand": "left",

"leading\_conf": 1.0,

"user\_id": "bb138d5db200f29385f..."

}

* Ключ - имя изображения без расширения
* Bboxes - список нормализованных bboxes[top left X pos, top left Y pos, width, height]
* Метки — список меток классов, например like, stop, ,no\_gesture
* Ведущая рука - rightили leftза руку, показывающую жест
* Leading conf - ведущее доверие дляleading\_hand
* Идентификатор пользователя — идентификатор субъекта (полезно для разделения данных для обучения/оценки подмножеств).

Ограничивающие рамки

| Объект | Поезд + Вал | Тест | Общий |
| --- | --- | --- | --- |
| жест | ~ 28 300 | ~ 2 400 | 30 629 |
| нет жеста | 112 740 | 10 849 | 123 589 |
| Всего коробок | 622 063 | 54 518 | 676 581 |

Тренировка модели

Вы можете использовать загруженные обученные модели, в противном случае выберите классификатор и параметры для обучения в формате default.yaml. Чтобы обучить модель, выполните следующую команду:

python -m classifier/run.py --command 'train' --path\_to\_config <PATH>

Каждый шаг, текущая потеря, скорость обучения и другие значения регистрируются в **Tensorboard** . Просмотрите все сохраненные метрики и параметры, открыв командную строку (при этом откроется веб-страница по адресу localhost:6006):

tensorboard --logdir=experiments

Тест

Протестируйте свою модель, выполнив следующую команду:

python -m classifier/run.py --command 'test' --path\_to\_config <PATH>

Демо

python demo.py -p <PATH\_TO\_DETECTOR>

