DOI - 10.32743/UniTech.2021.92.11.12582

РАСПОЗНАВАНИЕ ЖЕСТОВ РУК С ПОМОЩЬЮ MobileNetV2

Махмудов Миршод Дилшод угли
магистр,
Ташкентский университет информационных технологий имени Мухаммада аль Хоразми,
Республика Узбекистан, г. Ташкент
E-mail: makhmudov0907 @gmail.com

Фазилова Дилбархон Шамурадовна
д-р. ф.-мат. наук,
Астрономический институт Академии наук Узбекистана,
Республика Узбекистан, г. Ташкент
E-mail: dil_faz@yahoo.com

HAND GESTURE RECOGNITION USING MobileNetV2

Mirshod Makhmudov
master,
Tashkent information technology university,
Republic of Uzbekistan, Tashkent

Dilbarkhon Fazilova
Doctor of Physical and Mathematical Sciences,
Astronomical Institute of the Uzbekistan Academy of Sciences,
Uzbekistan, Tashkent

АННОТАЦИЯ

В работе используется новейшая платформа Tensorflow с открытым исходным кодом от Google и модель сверточной сети SSD MobileNetV2 FPN Lite для построения мобильной системы распознавания жестов. Эксперимент проводился с использованием самостоятельно собранного набора данных. Результаты экспериментов показывают, что можно легко настраивать структуру сети, быстро находить оптимальную модель и лучше выполнять задачу распознавания жестов.

ABSTRACT

The paper uses Google newest open-source Tensorflow framework and convolution neural network SSD MobileNet V2 FPN Lite model to build the mobile system of hand gesture recognition. The experiment is designed with the self-collected datasets. The experimental results show that one can easily adjust the structure of the network, quickly find the optimal model, and better perform the gesture recognition task.

Ключевые слова: конволюционная нейронная сеть, трансферное обучение, SSD MobileNet V2 FPN Lite.
Keywords: convolution neural network, transfer learning, SSD MobileNetV2 FPN Lite.

В настоящее время прямой контакт через такие устройства, как мышь, клавиатура, пульт дистанционного управления, сенсорный экран, является доминирующей формой взаимодействия между пользователем и машиной. Взаимодействие же людей друг с другом осуществляется с помощью более естественных и интуитивно понятных бесконтактных методов, таких как звук и физические движения. Гибкость и эффективность бесконтактных методов делают актуальными проведение исследований по обеспечению возможности поддержки взаимодействия человека с компьютером. Среди них распознавание жестов — это область активных текущих исследований в области компьютерного зрения и машинного обучения. Она имеет множество потенциальных приложений в сфере здравоохранения, образования и индустрии развлечений, для поддержки пользователей с особыми потребностями и пожилых людей, управления роботами или создания музыки. Исторически использовались информационные перчатки для определения угла и положения каждого сустава при жесте пользователя. Но, неудобство и стоимость одеваемого на руку датчика, ограничивают широкое использование такого метода. Методы распознавания жестов, основанные на бесконтактном визуальном осмотре,

Библиографическое описание: Махмудов М.Д., Фазилова Д.Ш. РАСПОЗНАВАНИЕ ЖЕСТОВ РУК С ПОМОЩЬЮ MobileNetV2 // Universum: технические науки : электрон. научн. журн. 2021. 11(92). URL: *https://7universum.com/ru/tech/archive/item/12582*

в настоящее время популярны из-за их низкой стоимости и удобства для пользователя [1]. Работы, связанные с системой распознавания языка жестов на основе зрения, можно разделить на две категории: традиционные и подходы, связанные со сверточной нейронной сетью (CHH). Среди применяемых подходов алгоритмы k-ближайшего соседа и нейронной сети [2]. Они обычно имеют более низкую точность прогнозирования. Вместо создания CHH с нуля использование трансферного обучения на этих архитектурах может значительно сократить время, необходимое для создания рабочей модели для конкретной проблемы. Тестирование моделей VGG16, VGG19, ResNet, Xception, Inception V3, MobileNet и SqueezeNet выполнялось для распознавания жестов рук. Наивысшая точность трансферного обучения была достигнута с использованием модели MobileNet и составила 99,45% [3]. Целью данной работы является создание системы распознавания жестов рук с использованием новой архитектуры нейронной сети SSD MobileNetV2 FPN Lite, адаптированной для мобильных сред и сред с ограниченными ресурсами, такими как мобильные устройства, робототехника и беспилотные транспортные средства [4].

SSD MobileNetV2 FPN Lite — это архитектура сети с самой глубокой сверткой. Эти прикладные платформы не так мощны, как компьютер общего назначения, но все же требуют сопоставимой точности распознавания при своевременных вычислениях. SSD MobileNetV2 FPN Lite использует функцию активации выпрямленного линейного модуля (Rectified Linear Unit – ReLU), определяющую архитектуру построения свертки с разделением по глубине узких мест из базовых блоков с остатками [5]. Можно отметить два основных момента в сети: инвертированные остатки (инвертированная остаточная структура) и линейные узкие места. Полная стандартная свертка заменяется факторизованной версией и реализуется в два этапа. Первый шаг называется глубинной сверткой, он выполняет легкую фильтрацию, применяя единственный сверточный фильтр для каждого входного канала. Второй шаг — это свертка 1×1, называемая точечной сверткой, которая отвечает за создание новых функций путем вычисления линейных комбинаций между входными каналами. Архитектура MobileNetV2 содержит начальный полностью сверточный слой с 32 фильтрами, за которым следуют 19 уровней остаточных узких мест. Функция активации ReLU6 использована в качестве нелинейности из-за ее устойчивости при использовании с вычислениями с низкой точностью. Также использованы размер ядра 3 × 3, что является стандартом для современных сетей, прерывание и пакетная нормализация во время обучения (Рис.1). Реализация сети нами выполнена с помощью библиотеки моделей Tensorflow-Slim [6].

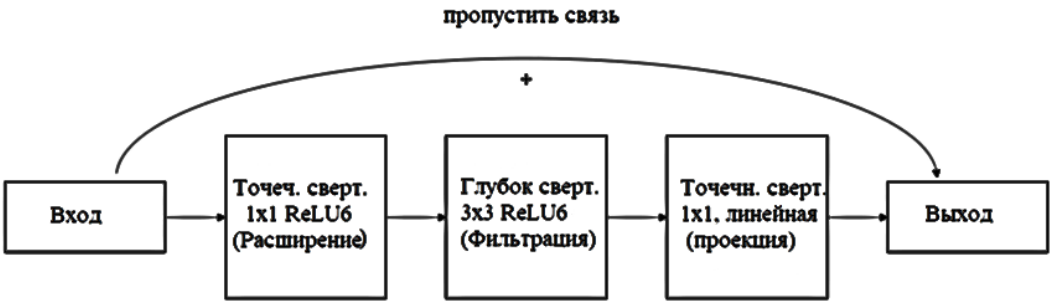

Рисунок 1. Инвертированный остаточный блок

Набор данных для данной работы создавался самостоятельно с помощью снимков камеры ноутбука. Все изображения были получены с разного расстояния камеры с небольшими изменениями освещения. В реальных приложениях исполнитель языка жестов может стоять на разном расстоянии от камеры нотебука. Обеспечено увеличение данных, которое широко применяется при обучении CNN с ограниченным набором данных, которое вводит больше вариаций с масштабированием, вращением, переворачиванием, сдвигом, сдвигом и т. д. У каждого знака есть вариации в захвате расстояния и освещении.

Для обучения было собрано 58 изображений жестов. Размер изображений 1280 × 720 пикселей. Входные данные были перемешаны случайным образом. Набор для проверки получали в режиме реального времени с камеры. Инструмент LabelImg использовался для графической маркировки изображений. Все эксперименты проводились на компьютере с процессором Intel Core i7, 32 ГБ оперативной памяти и NVIDIA GeForce RTX 720. Скорость модели SSD MobileNetV2 FPN Lite 39 мс. Примеры работы программы и точность распознавания жестов «Согласие», «Отказ» , «Спасибо» и «Солидарность» показаны на рисунке 2.

Преимуществом этой модели является, что даже при использовании сети глубокого обучения она може быть использована для работы на мобильных устройствах и использовать камеру смартфона для удобства и улучшния качества связи между глухонемыми и широкой публикой.

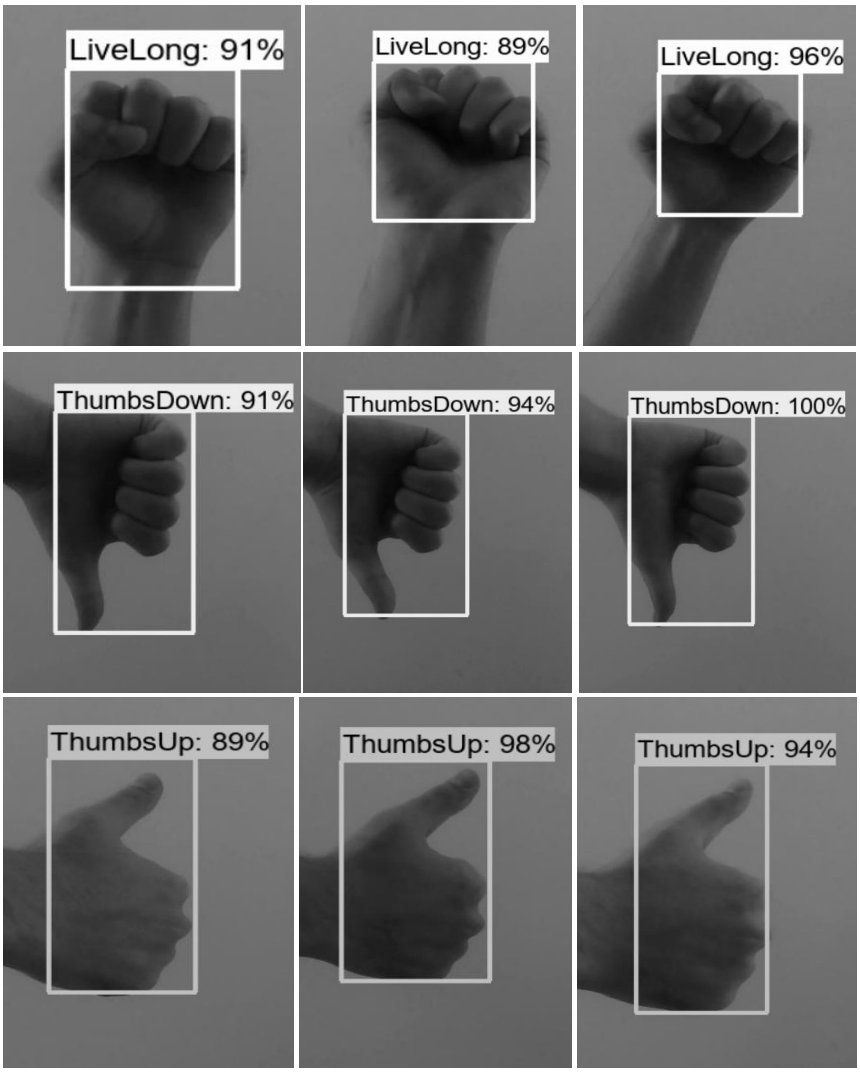

Рисунок 2. Примеры работы программы распознавания жестов в реальном режиме

Список литературы:

1. Chaudhary A., Raheja J.L., Das K., Raheja S. Intelligent approaches to interact with machines using hand gesture recognition in natural way: a survey // International Journal of Computer Science and Engineering Survey. – 2011. Vol. 2. №1. – P.122–133.
2. Hartanto R., Kartikasari A. Android based real-time static Indonesian sign language recognition system prototype // 8th International Conference on Information Technology and Electrical Engineering (ICITEE). (Yogyakarta, Indonesia, 5–6 October, 2016). – Indonesia, 2016. – P.1–6.
3. Alashhab S., Gallego A.J., Lozano M.Á., Hand gesture detection with convolutional neural networks // International Symposium on Distributed Computing and Artificial Intelligence: proc. (Toledo, Spain, 20–22 June, 2018). – Spain, 2018. – P. 45–52.
4. Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.C. Mobilenetv2: Inverted residuals and linear bottlenecks // IEEE conference on computer vision and pattern recognition: proc. (Salt Lake City, USA, 18–23 June, 2018). – USA, 2018. – P. 4520.
5. Chollet F. Xception: deep learning with depthwise separable convolutions / [Электронный ресурс]. – Режим доступа: URL: https://arxiv.org/abs/1610.02357 (дата обращения: 17.10.21).
6. Tensorflow / [Электронный ресурс]. – Режим доступа: URL: https://github.com/tensorflow/models/tree/master/research/slim/nets/mobilenet (дата обращения: 17.10.21).