

# РЕШЕНИЕ ЗАДАЧИ РАСПОЗНАВАНИЯ ЖЕСТОВ РУК НА ОСНОВЕ НЕЙРОННОЙ СЕТИ MASK-RCNN

*Т.Е. Мамонова, к.т.н., доц. ОАР ИШИТР,  
Д.А. Бульгин, студент гр. 8ЕМ91  
Томский политехнический университет  
E-mail: [dab28@tpu.ru](mailto:dab28@tpu.ru)*

## Введение

В настоящее время всё больше исследований направлено на решение задач с применением компьютерного зрения и искусственного интеллекта. Наиболее частыми являются решения и подходы с использованием распознавания жестов на основании инфракрасных сенсоров или нейронных сетей.

Целью данной работы является анализ работы нейронной сети Mask-RCNN, специально обученной для распознавания и выделения жестов в кадре.

## Существующие методы распознавания жестов

Индийскими учеными Adithya V. и Rajesh R. было предложено использовать сверточную нейронную сеть с тремя сверточными слоями для распознавания жестов рук в статике [1]. Данная нейронная сеть способна распознавать 24 жеста, с общей точностью около 96%.

Другой способ распознавания жестов был рассмотрен китайскими учеными Yao Huang, Jianyu Yan. При помощи устройства Microsoft Kinect, они смогли выделить область в кадре, в которой находится человеческая кисть. Затем, алгоритм исследует эту область и по наличию определенных признаков определяет жест [2].

В Тунисе рассмотрели способ распознавания жестов в динамике. Safa Ameer, Anouar Ben Khalifa, Med Salim Bouhleb, на основе устройства Leap Motion и нейронной сети типа LSTM смогли создать алгоритм, распознающий жесты по движению рук [3]. Их алгоритм способен успешно распознавать 12 динамических жестов с точностью до 90%.

## Теоретическое описание работы нейронной сети Mask-RCNN

Архитектура сети R-CNN (Regions With CNNs) была разработана командой из UC Berkeley для применения Convolution Neural Networks к задаче обнаружения объектов. Процедуру детектирования объектов сетью R-CNN можно разделить на следующие шаги:

- Выделение регионов-кандидатов.
- Преобразование региона в размер, принимаемый CNN.
- Получение при помощи CNN 4096-размерного вектора признаков.
- Проведение N бинарных классификаций каждого вектора признаков при помощи N линейных SVM.
- Линейная регрессия параметров рамки региона для более точного охвата объекта.

Mask R-CNN развивает архитектуру Fast R-CNN путём добавления ещё одной ветки, которая предсказывает положение маски, покрывающей найденный объект, и решает уже задачу сегментации объектов.

Маска представляет собой просто прямоугольную матрицу, в которой 1 на некоторой позиции означает принадлежность соответствующего пикселя объекту заданного класса, 0 — что пиксель объекту не принадлежит [4].

## Применение Mask-RCNN для распознавания и классификации жестов

Данная нейронная сеть обучена на наборе данных, включающем в себя пять классов жестов рук:

1. Большой палец (Thumb Finger);
2. Кулак (Fist);
3. Открытая ладонь (Palm);
4. Телефон (Phone);
5. Победа (Victory).

Все существующие типы изображений приведены на рисунке 1.



Рис. 1. Пять классов жестов

Обучающий набор данных содержит 300 изображений для обучения и 50 для тестирования нейронной сети.

Нейронная сеть Mask-RCNN обучалась на обучающей выборке десять эпох, в результате итоговые ошибки равны отражены в Таблице 1. Общая точность распознавания жестов на тестовой выборке данных равна 94%.

Таблица 1. Результаты обучения нейронной сети Mask-RCNN

Ошибка предсказания класса жеста	Ошибка наложения маски	Ошибка выделения контура	Ошибка предсказания класса
0,207	0,0998	0,0464	0,209

Работа нейронной сети Mask-RCNN была проверена на реальных изображениях. Результат работы Mask-RCNN можно увидеть на рисунке 2.



Рис. 2. Результаты работы нейронной сети

Как видно по рисунку 2, нейронная сеть верно распознает класс жеста, но наложение маски происходит некорректно. Можно заметить, что контур и сама маска не покрывают полностью жест. Это происходит вследствие недостаточного обучения последней части Mask-RCNN, а именно в слое, обратном слою свертки изображения.

### Заключение

В результате работы, было проведено обучение нейронной сети Mask-RCNN для распознавания жестов рук. Была подготовлена обучающая выборка, содержащая в сумме 350 изображений. Также была оценена работа нейронной сети на реальных данных. В ходе работы было выявлено, что классификация и обнаружение происходят корректно, но наложение маски на найденные жесты требует дополнительного обучения.

### Список использованных источников

1. Adithya V., Rajesh R.A. Deep Convolutional Neural Network Approach for Static Hand Gesture Recognition. Procedia Computer Science. – 2020. – vol. 171. – P. 2353–2361.
2. Huang Y., Yang J. A multi-scale descriptor for real time RGB-D hand gesture recognition. Pattern Recognition Letters. – 2021. – vol. 144. – P. 97–104.
3. Ameer S., Khalifa B., Bouhlel S. A novel hybrid bidirectional unidirectional LSTM network for dynamic hand gesture recognition with Leap Motion. Entertainment Computing. – 2020. – vol. 35.
4. Архитектура современной нейронной сети для сегментации объектов на изображениях [Электронный ресурс]. – URL: [https://libeloc.bsuir.by/bitstream/123456789/33348/1/Goncharov\\_Arkhitektura.PDF](https://libeloc.bsuir.by/bitstream/123456789/33348/1/Goncharov_Arkhitektura.PDF) (Дата обращения: 16.02.2021).