

УДК 004.51

**Никифорова Лариса Владимировна**

Амурский государственный университет

г. Благовещенск, Россия

*E-mail:* chepak.lv@amursu.ru**Добровольская Наталья Юрьевна**

Кубанский государственный университет

г. Краснодар, Россия

*E-mail:* dnu10@mail.ru**Nikiforova Larisa Vladimirovna**

Amur State University

Blagoveshchensk, Russia

*E-mail:* chepak.lv@amursu.ru**Dobrovolskaya Natalia Yurievna**

Kuban State University

Krasnodar, Russia

*E-mail:* dnu10@mail.ru**РАСПОЗНАВАНИЕ ЭЛЕМЕНТОВ ДАКТИЛЬНОЙ АЗБУКИ  
НА ОСНОВЕ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ****RECOGNITION OF DACTYLIC ALPHABET ELEMENTS  
BASED ON NEURAL NETWORK TECHNOLOGIES**

*Аннотация.* В статье рассматривается подход к распознаванию элементов дактильного языка, основанный на нейронной сети с долгой краткосрочной памятью. Выполнен краткий обзор существующих программных решений. Разработан собственный датасет и архитектура LSTM. Предложенная архитектура позволяет получить 92,52% категориальной точности распознавания.

*Abstract.* The article considers an approach to the recognition of elements of a classical language based on a neural network with a long short-term memory. A brief overview of existing software solutions has been performed. Our own dataset and LSTM architecture have been developed. The proposed architecture allows to obtain 92.52% categorical recognition accuracy.

*Ключевые слова:* компьютерное зрение, распознавание образов, рекуррентная нейронная сеть, дактильный язык, датасет.

*Key words:* computer vision, pattern recognition, recurrent neural network, dactyl language, dataset.

В условиях цифровизации большое внимание уделяется современным информационным технологиям, способствующим повышению качества жизни людей с ограниченными возможностями. Слабослышащие получили возможность общения по видеоканалам, а нейросетевые технологии, решая задачу распознавания образов, способны осуществить связь между тактильным языком и письменной речью. Подобные программы помогут слабослышащим без ограничений и дополнительного персонала участвовать в различных конференциях, совещаниях и т.д. Основным функционалом таких разработок должно быть распознавание жестов в реальном времени и трансформация их в текст.

Задаче распознавания жестов, как одной из центральных задач компьютерного зрения, посвящено много исследований. Так, Д.М. Булыгин и Т.Е. Мамонова анализируют различные технологии распознавания жестов в режиме реального времени, на основе нейронных сетей Segnet и Unet приводят собственную архитектуру сверточной нейронной сети, качественно работающей при дальности до 30 см [1]. Д.А. Рюмин использует для распознавания движения рук глубокую нейронную сеть модели Single Shot MultiBox Detector, причем точность распознавания составляет 79% [2]. И.А. Кагиров, А.А. Аксенов, А.А. Карпов, Д.А. Рюмин на основе расширенного аналитического обзора моделей реализации распознавания жестовых языков делают вывод, что наиболее эффективным подходом является использование архитектур двухпоточных сверточных и рекуррентных сетей [3].

В нашем исследовании для распознавания элементов тактильной азбуки выбрана архитектура рекуррентной нейронной сети. Для программной реализации используется язык Python с подключенной библиотекой компьютерного зрения и обработки изображения OpenCV. Конструирование нейронной сети и ее обучение выполнены с помощью библиотеки машинного обучения TensorFlow, мультимодальные конвейеры машинного обучения MediaPipe создаются посредством фреймворка, представленного Google.

Рекуррентные нейронные сети используются в компьютерной лингвистике, при анализе временных рядов и представляют собой процесс обработки последовательных элементов, связанных друг с другом. Архитектура этого типа нейронных сетей позволяет запоминать данные предыдущего элемента для использования их при обработке следующего элемента. Однако простая рекуррентная сеть имеет склонность к затухающему или взрывному градиенту при обучении на достаточно длинных последовательностях элементов. Решение этой проблемы – нейронная сеть с архитектурой долгой краткосрочной памяти (LSTM). Эта сеть включает блоки памяти, обладающие свойством избирательного запоминания данных.

Архитектура нейронной сети для распознавания тактильных жестов представлена тремя слоями с долгой краткосрочной памятью (LSTM) и тремя плотными слоями. Первый, второй и третий LSTM слои имеют 64, 128 и 64 нейрона соответственно. Первый, второй и третий плотные слои имеют 64, 32 и 3 нейрона соответственно. При необходимости количество слоев и нейронов в них можно изменить. Функция активации всех слоев – *relu*, кроме последнего плотного слоя, имеющего функцию активации *softmax*. Для обучения модели ис-

пользуется оптимизатор adam. В качестве функции потерь применена категориальная кросс-энтропия.

Дактильная азбука русского языка содержит тридцать два символа, причем девять из них динамические (рис.1). Для обучения нейронной сети разработан собственный датасет, содержащий для каждого символа дактильного алфавита тридцать вариантов жеста, соответствующих тридцати кадрам.

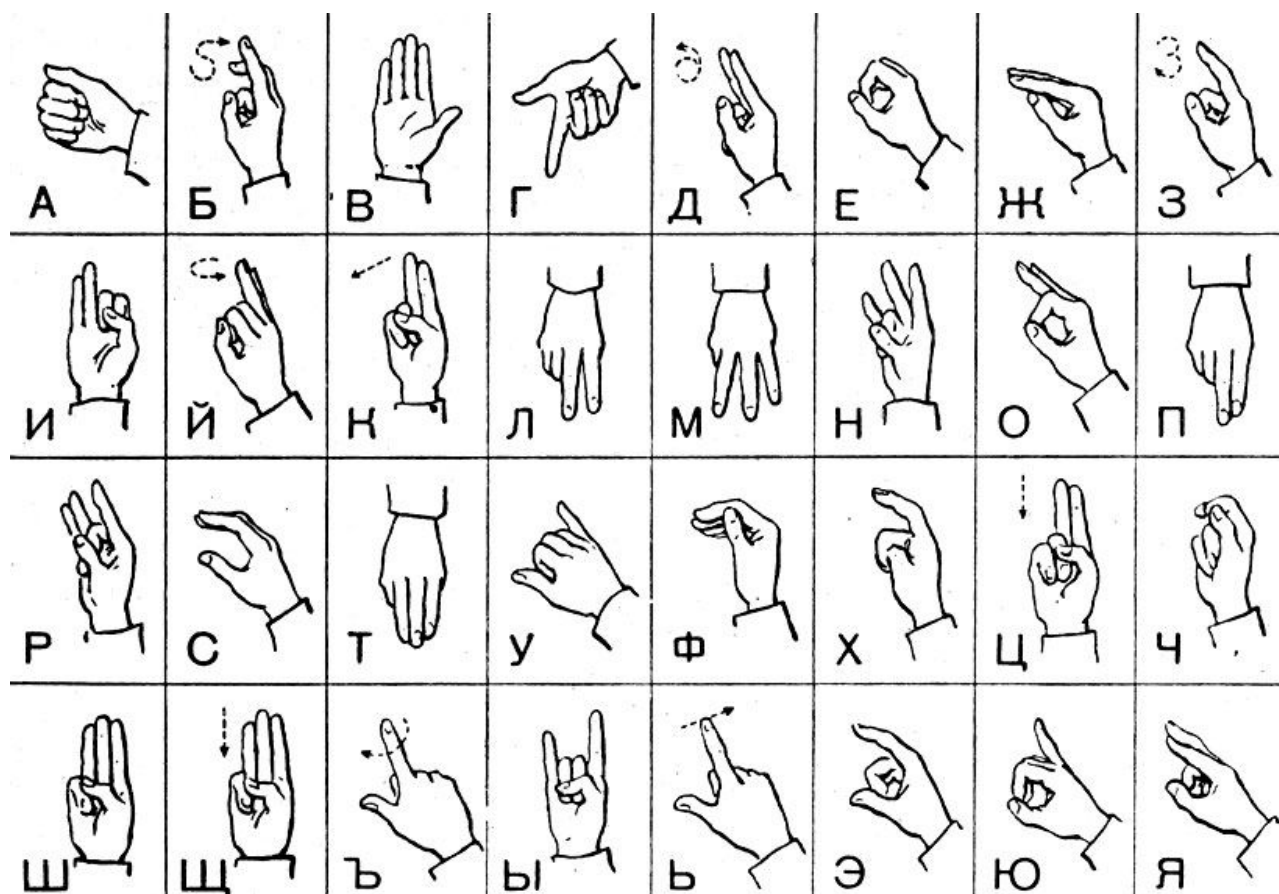


Рис. 1. Дактильный алфавит.

Стандартный контейнер библиотеки MediaPipe считывает с изображения двадцать одну ключевую точку кисти руки. Мы будем использовать расширенный функционал библиотеки и получать массив трехмерных координат сорока двух ключевых точек анатомических областей кисти. Таким образом, в датасете для одного жеста хранится 113400 координат, на основе которых нейронная сеть выполняет обучение по распознаванию этого жеста. Наполнение контента датасета автоматизировано и позволяет пользователю один раз перед работой программы распознавания собрать кадры жестов, продемонстрировав символы дактильной азбуки перед веб-камерой. Эта функция дает возможность индивидуализировать датасет, сформировав обучающую выборку под анатомические особенности конкретного пользователя.

Общий алгоритм работы процедуры распознавания дактильных жестов представлен на рис. 2.

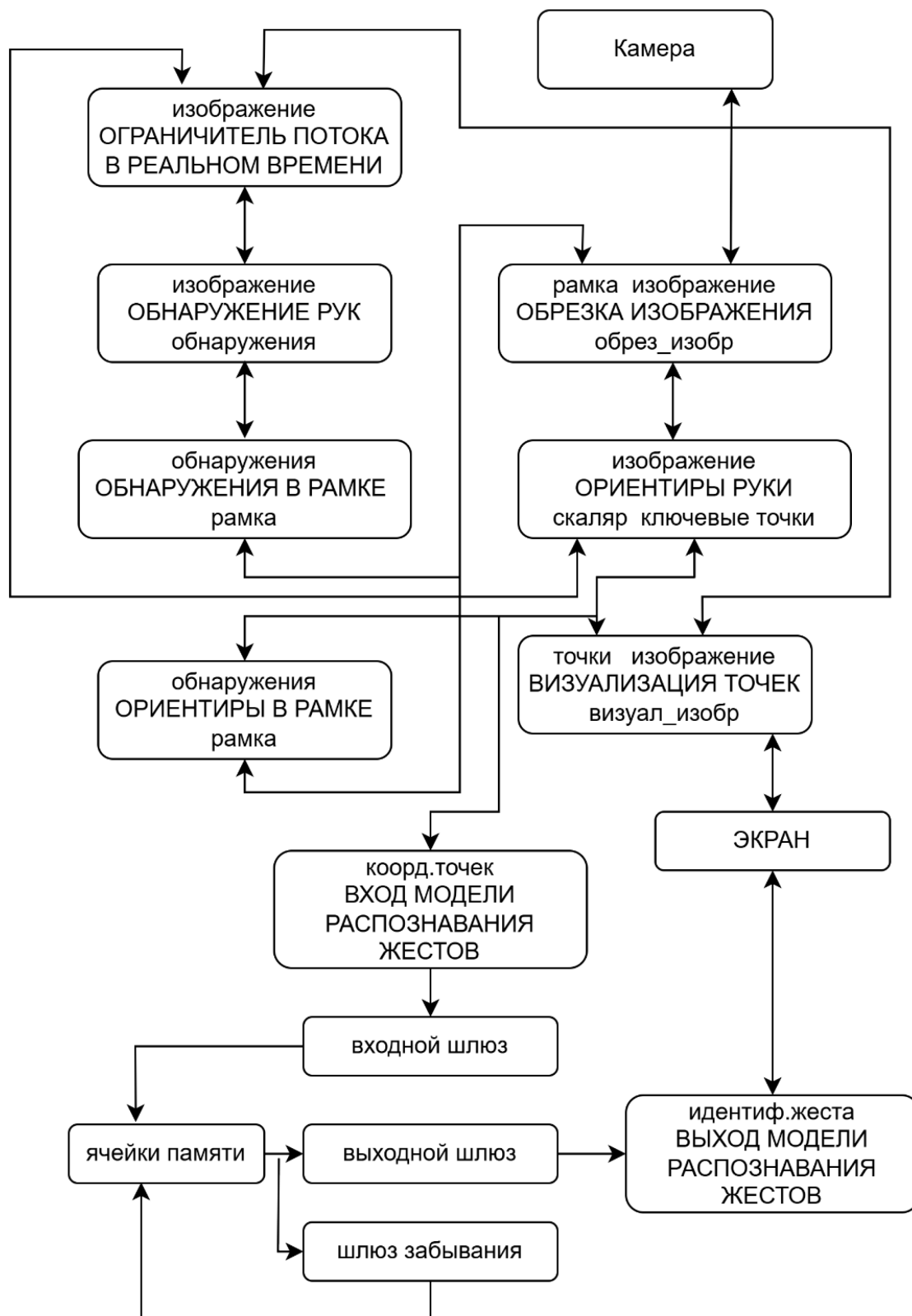


Рис. 2. Алгоритм распознавания дактильных жестов.

При обучении нейронной сети подбиралось и анализировалось количество эпох. Графики потерь и категориальной точности для обучающейся модели (160 эпох) показали переобучение после 130-й эпохи (рис. 3, 4).

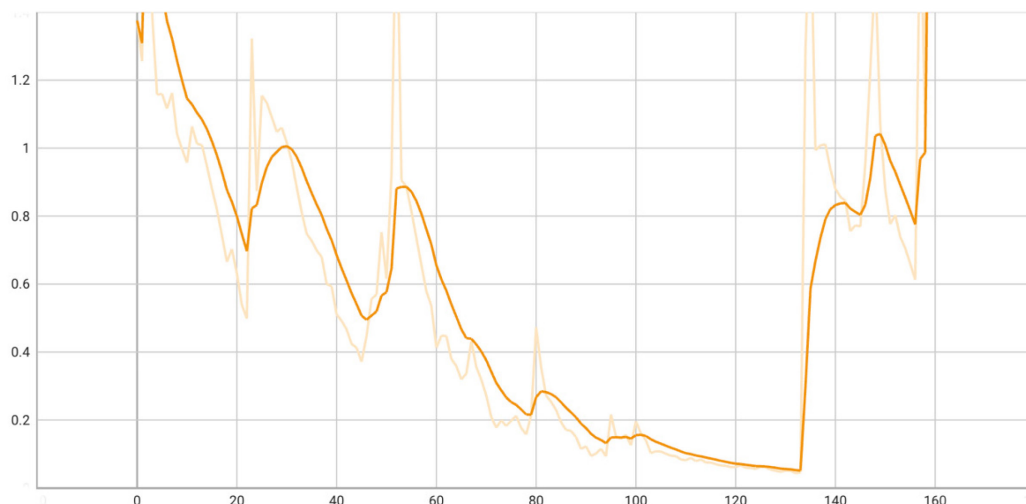


Рис. 3. График потерь для 160 эпох.

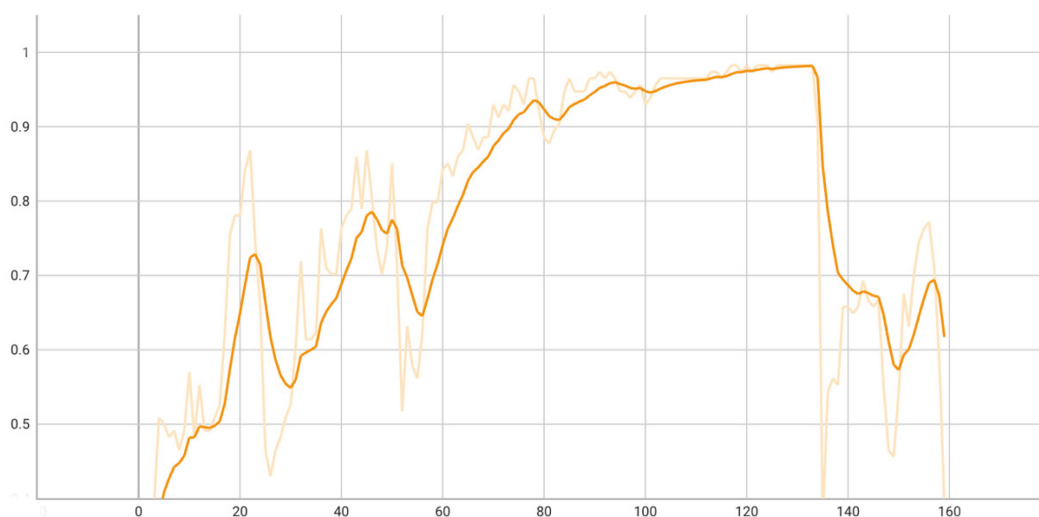


Рис. 4. График категориальной точности для 160 эпох.

Для повышения качества распознавания принято решение снизить количество эпох до 120 (рис. 5, 6).

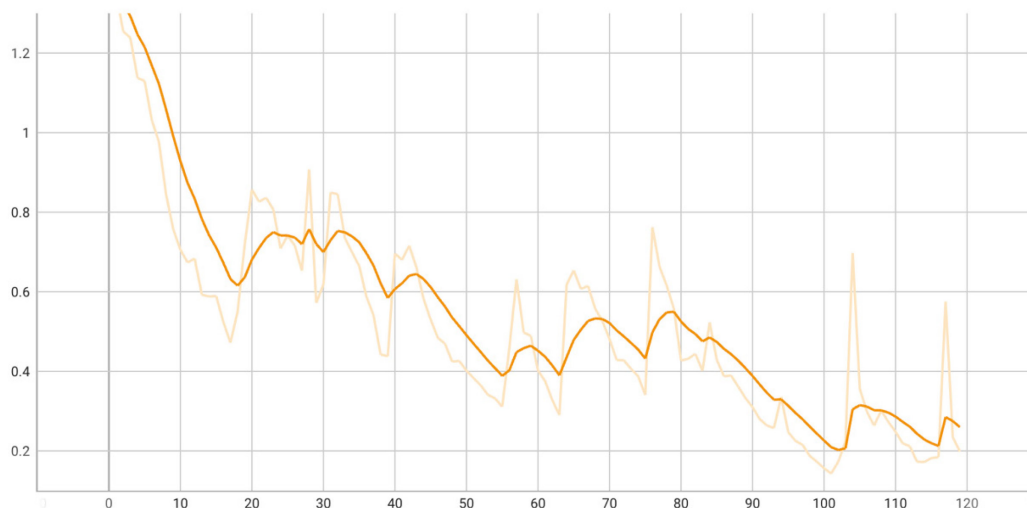


Рис. 5. График потерь для 120 эпох.

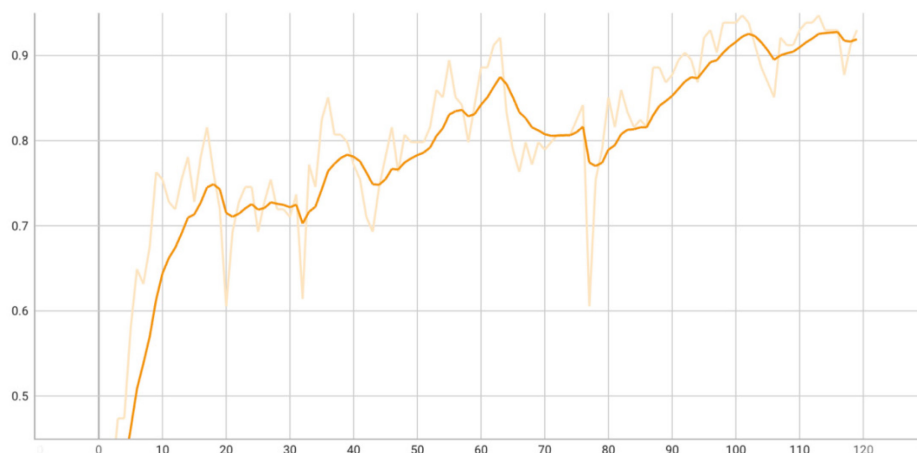


Рис.6. График категориальной точности для 120 эпох.

Результаты показали, что проблема переобучения решилась. При 120 эпохах категориальная точность составляет 92%, а при 160 эпохах – всего 63%, что явно недостаточно для качественного распознавания.

Далее проводился эксперимент по распознаванию жестов в условиях плохой освещенности, различной удаленности руки от камеры и разных углов поворота кисти (рис.7).

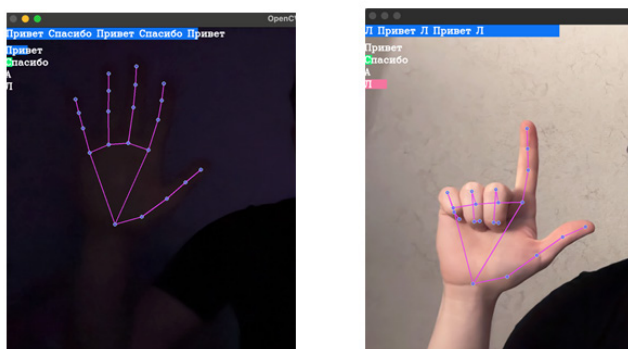


Рис.7. Эксперимент.

Эксперимент показал, что при наличии в датасете обучающих кадров с высокой удаленностью руки от камеры распознавание жестов будет приемлемым. Низкая освещенность не требует дополнительных примеров в датасете. Предложенная архитектура нейронной сети способна распознавать элементы дактильного русского языка с категориальной точностью 92,54% и может служить основой системы автоматического сурдоперевода и других человеко-машинных интерфейсов, позволяющих взаимодействовать с программами посредством жестов.

1. Булыгин, Д.А. Распознавание жестов рук в режиме реального времени / Д.А. Булыгин, Т.Е. Мамонова // Системы анализа и обработки данных. – 2020. – №1 (78). – С.25-40.

2. Рюмин, Д.А. Метод автоматического видеоанализа движений рук и распознавания жестов в человеко-машинных интерфейсах // Научно-технический вестник информационных технологий, механики и оптики. – 2020. – №4. – С. 525-531.

3. Рюмин, Д.А. Аналитический обзор моделей и методов автоматического распознавания жестов и жестовых языков / Д.А. Рюмин, И.А. Кагиров, А.А. Аксенов, А.А. Карпов // Информационно-управляющие системы. – 2021. – №6 (115). – С.10-20.