

УДК 004.8

**РАЗРАБОТКА СИСТЕМЫ РАСПОЗНАВАНИЯ ЖЕСТОВ СУРДОПЕРЕВОДА
В РЕАЛЬНОМ ВРЕМЕНИ ДЛЯ СИСТЕМ ВИДЕОСВЯЗИ С ИСПОЛЬЗОВАНИЕМ
СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ И МЕТОДОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ**

**Шевчук Марат Александрович, студент
Просвирнина Ирина Борисовна, доцент кафедры СБиКБ
Гродненский государственный университет имени Янки Купалы**

**DEVELOPMENT OF A REAL-TIME SIGN LANGUAGE GESTURE RECOGNITION SYS-
TEM FOR VIDEO COMMUNICATION SYSTEMS USING CONVOLUTIONAL NEURAL
NETWORKS AND COMPUTER VISION METHODS**

**Shevchuk Marat, student, marat2656671@gmail.com
Prosvirnina Irina, Associate Professor of the Department of SPCS, PhD, i.prosvirnina@grsu.by
Yanka Kupala State University of Grodno**

Аннотация. В данной работе рассматривается разработка системы распознавания жестов сурдоперевода в реальном времени для интеграции в системы видеосвязи. Основное внимание уделено применению сверточных нейронных сетей и методов компьютерного зрения для обработки видеопотока и классификации жестов. Предложенный подход позволяет повысить доступность коммуникации для людей с нарушениями слуха и речи. Рассматриваются архитектуры нейронных сетей, методы предварительной обработки изображений и особенности работы системы в реальном времени.

Ключевые слова: компьютерное зрение, сверточные нейронные сети, распознавание жестов, сурдоперевод, видеосвязь, машинное обучение.

Abstract. This paper presents the development of a real-time sign language gesture recognition system for integration into video communication platforms. The study focuses on the application of convolutional neural networks and computer vision techniques for video stream processing and gesture classification. The proposed approach improves communication accessibility for people with hearing and speech impairments. Neural network architectures, image preprocessing methods, and real-time system constraints are discussed.

Keywords: computer vision, convolutional neural networks, gesture recognition, sign language, video communication, machine learning.

В последнее время специалисты активно развивают технологии видеосвязи. Эти инструменты стали необходимы, так как общество переходит на цифровые системы и люди чаще взаимодействуют на расстоянии. Но хотя платформы Zoom, Microsoft Teams и другие сервисы стали популярными, люди с нарушениями слуха и речи все еще сталкиваются с трудностями при общении. Для решения этого вопроса инженеры создают интеллектуальные системы, которые автоматически распознают жесты сурдоперевода.

Сурдоперевод - это система общения через визуальные сигналы, которая состоит из движений рук, выражений лица и поз человека. Если сравнивать этот процесс с обычным текстом или речью, то жестовый язык часто меняется и зависит от ситуации. По этой причине компьютерным программам трудно распознавать его автоматически. Чтобы такие системы работали успешно, разработчики используют методы компьютерного зрения и глубокого обучения.

Одним из основных инструментов в этой сфере являются сверточные нейронные сети (Convolutional Neural Networks, CNN). Эти алгоритмы работают результативно, когда компьютер обрабатывает изображения и видео. Главное свойство CNN - это способность программы самостоятельно находить важные характеристики в данных, такие как границы, узоры и очертания предметов. И это имеет значение, когда система изучает жесты, потому что для понимания смысла важны местоположение рук, их форма и то, как они перемещаются в пространстве.

В системе, которую создают специалисты, процесс распознавания жестов в реальном времени разделен на функциональные шаги. На первом шаге система получает видеопоток, когда камера пользователя передает данные. Этот поток является последовательностью кадров, и программа обрабатывает каждый из них отдельно. Для того чтобы анализ движения был точным, важно, чтобы видео содержало большое количество кадров в секунду.

На втором этапе программа предварительно обрабатывает изображения. В ходе этого процесса система приводит данные к единым параметрам, меняет размер картинок до нужных значений и удаляет лишние помехи. Дополнительно алгоритмы могут разделять изображение на части, чтобы найти область, где находятся руки пользователя. С этой целью применяются методы, которые анализируют цвета, выделяют границы объектов или используют современные модели, которые находят предметы.

Позже нейронная сеть извлекает признаки и определяет вид жеста. В качестве основы для этой сети могут выступать различные модели глубокого обучения, такие как ResNet, MobileNet или EfficientNet. Если разработчикам нужно достичь определенных показателей точности или быстродействия, они выбирают конкретную структуру сети. К примеру, модель MobileNet является подходящей для мобильных телефонов, так как она потребляет мало вычислительной мощности. Но при этом модель EfficientNet показывает высокую точность, когда ресурсы компьютера используются наиболее эффективно. В этой задаче специалисты изучают, как алгоритмы распознают динамические жесты. Под такими жестами люди понимают движения, которые следуют друг за другом в определенные промежутки времени. Для того чтобы обрабатывать информацию, инженеры применяют разные архитектуры. В их состав входят сверточные сети и рекуррентные нейронные сети, такие как LSTM. Также существуют способы, с помощью которых программы анализируют последовательности кадров. Дополнительно в системе работают технологии, которые находят ключевые точки (landmarks) на руках и теле человека. Современные инструменты для компьютерного зрения следят за тем, где находятся суставы и пальцы. Данный процесс происходит без задержек. Когда система использует эти методы, точность распознавания становится выше. При этом фон и свет меньше влияют на результат, а количество входных данных становится меньше.

Одной из задач является создание условий, при которых система работает в реальном времени. Это условие требует, чтобы алгоритмы обрабатывали каждый кадр очень быстро. Для того чтобы достичь такого результата, разработчики оптимизируют нейронные сети. Они применяют квантование и делают модели меньше по размеру. И, наконец, вычисления происходят быстрее, если задействовать GPU или специализированные ускорители. Если инженеры внедрят созданную систему в платформы для видеосвязи, это сделает общение более доступным. В частности, система автоматически переводит движения рук пользователя в текстовые символы или искусственный голос. С помощью этой функции люди, которые не знают жестовый язык, понимают собеседника. Для обратной связи программа превращает звуковую речь в текст и показывает его на экране.

Но специалисты продолжают активно работать в этой области, так как процесс еще не завершен. К числу текущих трудностей относятся малое количество точных данных с метками и ошибки при определении жестов, когда свет в помещении слабый или фон неоднородный. Также системе нужна настройка под разные национальные варианты жестового языка.

Специалисты считают, что создавать системы, которые распознают жесты сурдоперевода через сверточные нейронные сети и методы компьютерного зрения, важно и полезно. С помощью таких технологий государственные и частные цифровые сервисы станут удобнее. И это упростит повседневные задачи для людей, которые имеют физические ограничения.

Список использованных источников

1. Goodfellow, I. Deep Learning / I. Goodfellow, Y. Bengio, A. Courville. – Cambridge: MIT Press, 2016. – URL: <https://www.deeplearningbook.org/> – Дата доступа: 18.04.2026.
2. Krizhevsky, A. ImageNet Classification with Deep Convolutional Neural Networks / A. Krizhevsky, I. Sutskever, G. Hinton // Advances in Neural Information Processing Systems. – 2012. – URL: <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf> – Дата доступа: 18.04.2026.
3. Tan, M. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks / M. Tan, Q. Le. – 2019. – URL: <https://arxiv.org/abs/1905.11946> – Дата доступа: 18.04.2026.
4. Howard, A. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications / A. Howard [и др.]. – 2017. – URL: <https://arxiv.org/abs/1704.04861> – Дата доступа: 18.04.2026.
5. Molchanov, P. Hand Gesture Recognition with 3D Convolutional Neural Networks / P. Molchanov [и др.]. – 2015. – URL: <https://arxiv.org/abs/1506.00615> – Дата доступа: 18.04.2026.
6. Cao, Z. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields / Z. Cao [и др.]. – 2019. – URL: <https://arxiv.org/abs/1812.08008> – Дата доступа: 18.04.2026.
7. Lugaresi, C. MediaPipe: A Framework for Building Perception Pipelines / C. Lugaresi [и др.]. – 2019. – URL: <https://arxiv.org/abs/1906.08172> – Дата доступа: 18.04.2026.