

УДК 004.032.26

РАСПОЗНАВАНИЕ ОБЪЕКТОВ ИТК С ИСПОЛЬЗОВАНИЕМ ГЛУБИННЫХ И СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

Федотовских А.В.,

*к.э.н, профессор РАОИФН им. М.В. Ломоносова,
председатель Комитета по науке и инновациям РСПП-Заполярье,
член Российской ассоциации ИТК*

Аннотация: технологии машинного обучения всё активнее проникают в повседневную жизнь и все чаще используются для решения научных задач, могут применяться при работе с различными видами информации. Одной из нерешенных проблем инструментальной транскомуникации (ИТК) является трактовка полученной аудио- и визуальной информации, независимо от метода. В статье в краткой форме описываются сервисы для ИТК, созданные на основе искусственных нейронных сетей.

Ключевые слова: искусственные нейронные сети, инструментальная транскомуникация, машинное обучение.

Fedotovskikh A.V.,

*Candidate of Economic Sciences, Professor of the RACBS,
Chairman of the Science and Innovation Committee of the RUIE-Arctic,
member of the Russian ITC Association,*

RECOGNITION OF ITC OBJECTS USING DEEP AND CONVOLUTIONAL NEURAL NETWORKS

***Abstract:** Machine learning technologies are increasingly penetrating into everyday life and are increasingly used to solve scientific problems, can be used when working with various types of information. One of the unsolved problems of instrumental transcommunication (ITC) is the interpretation of the received audio and visual information, regardless of the method. The article briefly describes services for ITCs based on artificial neural networks.*

***Key words:** artificial neural networks, instrumental transcommunication, machine learning.*

Сторонники и противники ИТК придерживаются совершенно различных точек зрения по вопросу трактовки полученных сведений. Считающие, что ИТК - это не научный подход, приводят аргументы, что за тысячи лет человеческий мозг эволюционировал и может распознавать какие угодно шаблоны и даже в шуме мозг «обнаружит» для себя известные ему слова без их разумного и понятного источника. Важную роль играет и ожидание ответов в процессе «общения». В целом, несмотря на технические новации, ИТК зачастую присуще низкое качество распознавания звуков и образов, психологические аспекты, физическое состояние оператора, интерпретация полученных ответов человеческим мозгом. Скептики ФЭГ считают, что для чистоты эксперимента нужны иные более научные методы распознавания получаемых данных. Многие практики и сторонники ИТК считают, что не всегда дело в технической стороне вопроса, а голоса и изображения проявляются только в момент когда человек находится в некотором особенном, измененном состоянии сознания. В этом случае проявляется внешнее влияние мозга, сравнимое с эффектом наблюдателя, известным по школьным основам физики в эксперименте со щелями Томаса Юнга. Уверенных доказательств влияния оператора на ответы с той стороны или их отсутствие практически нет, поэтому, будем считать, что для чистоты эксперимента влияния внешнего

наблюдателя быть не должно. Для того, чтобы получить более точные результаты интерпретировать полученные аудио- и видеоматериалы может искусственный интеллект, нейросети, тем более, что даже современный уровень технологии может помочь справиться с этой задачей.

Искусственная нейронная сеть — математическая модель, а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей — сетей нервных клеток живого организма [1]. Нейронные сети можно классифицировать по многим параметрам. Например, по характеру обучения. Наиболее продвинутыми являются глубинные и сверточные нейросети. Глубокие или глубинные нейронные сети состоят из нескольких скрытых слоёв [2]. Сверточная нейросеть имеет специальную архитектуру, которая позволяет ей максимально эффективно распознавать образы. Идея такой нейросети основывается на чередовании сверточных и субдискретизирующих слоев, а структура является однонаправленной [3].

Распознавание голоса — бесконтактная, неинвазивная и простая в использовании технология. Нейросеть анализирует аудиопоток, затем разделяет его на отдельные фрагменты - фонемы. Каждая фонема анализируется, сверяется с базой эталонных звуков обученной нейросети и сопоставляется букве, слогу или целому слову. После многократного анализа фонем они расшифровываются в текст. Получившаяся текстовая запись снова сравнивается с базой слов нейросети. После выполнения всех действий нейросеть выдает готовый текст [4]. Аудиоанализ — область, включающая автоматическое распознавание речи (ASR), цифровую обработку сигналов, а также классификацию, тегирование и генерацию музыки — представляет собой развивающийся поддомен приложений глубокого обучения. Самые популярные системы машинного обучения, такие как Alexa, Siri и Google Home, созданы на основе моделей, извлекающих информацию из аудиосигналов [5].

Сложным с технической точки зрения является процесс отделения речевой информации от шума. Речевое шумоподавление нейросети удалит шум из речевых сигналов при улучшении качества и разборчивости речи. Нейронные сети для глубокого обучения уже способны выделять человеческую речь на фоне звуков громкой техники и может идентифицировать голоса одного или нескольких человек. Компания Alibaba создала нейросеть, способную распознать голос человека в шумных местах, точность достигает 94–95% даже несмотря на сильный акцент оратора. В шанхайском метро на основе этой нейросети введена система оплаты проезда. Компания решила проблему с использованием аппаратной и программной составляющей при помощи системы микрофонов и алгоритма нейросети, отсекающей лишние голоса. Для обработки естественного языка в реальном времени используют облачные сервисы. В 2018 г. разработчики Google научились использовать нейросети для выделения конкретного голоса в толпе. Они обучили две отдельные нейросети: одна распознаёт говорящего, другая сравнивает звуковые спектрограммы. Тесты новой разработки, названной VoiceFilter, показали снижение частоты ошибок в распознавании слов с 55,9% до 23,4% [6]. Ученые из Массачусетского технологического института на основе полученного голоса создают портрет носителя речи, что для ИТК крайне актуально. Специалисты создали и обучили алгоритм Speech2Face, способный проанализировать тон, интонации и иные характеристики голоса с аудиозаписи и создать портрет человека на этой основе. Нейросети пришлось анализировать миллионы записей с говорящими людьми и учиться соотносить тембр голоса с внешними характеристиками [7]. В 2019 г. разработчики в Google AI опубликовали нейросетевую модель, которая распознает спикеров на аудиозаписи. Нейросеть была протестирована на задаче распознавания аудиозаписей медицинских обследований. В сравнении с state-of-the-art моделью предложенный

подход сокращает процент ошибок с 15.8% до 2.2% [8].

В России также работают сервисы распознавания речи. Так, программа Voice2Med экономит время врача и производит голосовое заполнение документов на основе специализированных словарей, позволяющих верно распознавать сложные медицинские термины [9].

Поднаправление искусственного интеллекта, которое занимается работой с изображениями и видеопотоком, называется компьютерное зрение. Наиболее широкое распространение нейросети получили именно в области работы с изображениями. К примеру, на практике они активно используются в медицинской диагностике и скрининге патологий. Машинное зрение стало работать настолько хорошо, что мы можем перенести эту технологию на другие области, в т.ч. и на ИТК. Нейросеть с высокой долей вероятности сможет распознавать не только изображения человека, но и окружающей обстановки. Для достижения высокой точности распознавания нейросеть предобучается на большом массиве изображений, например, таком, как в базе данных MegaFace. Это основной метод обучения для распознавания лиц [10]. В 2015 г. в Google заставили нейросеть видеть несуществующие образы. В качестве примера можно привести узнавание человеком разных картинок в облаках, однако, уровень нейросети в разы выше. Эксперимент был проведён над системой распознавания картинок. Кроме того, нейросеть может генерировать образы из белого шума [11].

Специалисты из Лаборатории информатики и искусственного интеллекта (CSAIL) Массачусетского технологического института в 2016 г. презентовали нейросеть SoundNet (<http://projects.csail.mit.edu/soundnet/>), которую обучили распознавать объекты по звуку [12].

Однако нейросеть для ИТК должна иметь ряд особенностей:

1. Уметь выстраивать слова в предложения, т.к. ответы чаще всего одиночные.

2. Носить характер мультилингвальной.

3. Учитывать, что полученный голос может отличаться от того, который человек имел при жизни, ведь после смерти он бестелесный, а в теле голос - продукт деятельности мышц, которые управляются от полушарий головного мозга до спинного мозга. Изображение также может полностью не совпадать с прижизненным.

4. Сравнивать с образцами и определять носителя, для этого иметь массивный dataset – набор данных, звуков и фотографий из десятков тысяч образцов. Такой большой dataset нужен для повышения качества распознавания при обучении нейронной сети.

Указанные достижения и возможности появились благодаря обучению на массивных наборах данных. Единственная проблема в том, что использование нейросетей для ИТК требует большого количества машинного времени и ресурсов. Система, которая может генерировать внятную человеческую речь, должна быть очень мощной. Тем не менее, технологии развиваются, и специалисты утверждают, что через 5-10 лет ситуация может кардинально измениться.

В связи с отсутствием специальных нейросетей для ИТК можно использовать существующие открытые бесплатные онлайн-сервисы и софт.

Для обработки аудиоинформации, особенно голосов относящихся к классам В (временами затухает и вновь появляется через некоторое время) и С (очень слабый, едва слышимый сигнал):

- Сайт <https://mvsep.com/ru/> помогает разделить музыку и голос. Сайт выполняет разделение музыкального трека на составляющие: голос, музыка, барабаны, гитара и т.д. Процедура выделения вокала (голоса) из песни с музыкальным сопровождением и прочими шумами основана на сверточных нейросетях и дисциплине Feature Engineering [13].

- Программ Krisp (<https://krisp.ai/>). Разработчики прогнали через

нейросеть свыше 2,5 тысячи часов аудио из 20 тыс. различных источников шума. Программа обучена вычленять из аудиодорожки все посторонние звуки, что позволяет получить чистый звук. Krisp в режиме реального времени устраняет посторонние звуки во время звонков через Skype, Slack и другие VoIP-сервисы. Шумоподавление сервиса работает благодаря технологии машинного обучения. Команда проекта собрала два датасета: в первом было несколько десятков тысяч аудиозаписей с различными звуками-шумами, во втором — примеры чистой речи, записанной в студии [14].

- Нейросеть <https://speech2face.github.io/>. Помогает «восстановить» внешность человека по короткой аудиозаписи. Модель пока в точности не может воспроизводить изображения конкретных людей, но позволяет создавать изображения, отражающие различные физические характеристики говорящих, такие как возраст, пол и этническая принадлежность. Чем длиннее входная аудиозапись, тем и ближе сгенерированное изображение к истинному [15].

Для полученных изображений и спектрографии:

- Сайт <https://www.remove.bg/ru> работает с изображениями. Нейросеть Remove разделяет картинку и помогает за несколько секунд удалить фон с фотографии, оставить основной объект на переднем плане. Отлично справляется с фотографиями человека на фоне шума.

- Бесплатный онлайн сервис <https://letsenhance.io/> улучшает качество фотографий, увеличивая их разрешение в 4 раза. Текущая версия Let's Enhance использует нейронные сети, наибольшим образом ориентированные на пейзажные и портретные изображения, чтобы повысить качество снимков с низким разрешением.

- Сервис <https://colorize.cc/> создан российской компанией G-Core Labs на основе проекта DeOldify с открытым исходным кодом. Нейросеть умеет раскрашивать черно-белые фотографии в реалистичные цвета, что

очень актуально для изображений, полученных при ИТК.

- Сервис <https://cvl-demos.cs.nott.ac.uk/vrn/> создан учеными из Ноттингемского университета. В качестве основы использована сверточная нейросеть для распознавания объектов. Сервис умеет делать 3D-модели лица на примере одного фото. Обработка занимает 3-4 секунды, а сам результат можно «потрогать» мышкой или скачать в формате .obj.

Таким образом применение нейросетей позволит в меньшей степени использовать фото- и аудиоредакторы в ручном режиме.

Противники ФЭГ и ИТК утверждают, что сам феномен общения с иным миром является подделкой, несмотря на значительное количество фактов, подтверждающих это явление. Активное применение нейросетей улучшит техническую составляющую ИТК, опровергнет или подтвердит интерпретацию человека по поводу содержимого ответов, поможет в идентификации собеседника. Кроме того, нейросети помогут в борьбе с фейками и подделками, количество которых растет, а их авторы дискредитируют метод ИТК. Нейросети смогут определять подлинность представленных материалов и вычлнять из шума иные, не прослушиваемые явно ответы или фоновые изображения и использоваться как дополнение к уже существующим практическим технологиям ИТК.

Использованные источники:

1. А.А. Нейландс. Распознавание объектов с помощью нейронных сетей [Электронный ресурс] // Творческие проекты и работы учащихся. - URL: <https://tvorcheskie-proekty.ru/node/2557> (дата обращения: 23.11.2021).

2. Нейросеть для распознавания лиц. Часть 2 [Электронный ресурс] // NtechLab (дата публикации: 17.04.2017). - URL: <https://findface.pro/blog/nejronnye-seti-i-raspoznavanie-obrazov-chast-2/> (дата обращения: 23.11.2021).

3. Нейронные сети: распознавание образов и изображений с

помощью ИИ [Электронный ресурс] // Центр 2М. - URL: <https://center2m.ru/ai-recognition> (дата обращения: 23.11.2021).

4. И.Калинин. Распознавание и синтез речи: как ИИ анализирует наши разговоры [Электронный ресурс] // Spark (дата публикации: 11.08.2020). - URL: <https://spark.ru/startup/redaktsiya-spark-ru/blog/64389/raspoznavanie-i-sintez-rechi-kak-ii-analiziruet-nashi-razgovori> (дата обращения: 23.11.2021).

5. Анализ аудиоданных с помощью глубокого обучения и Python (часть 1) [Электронный ресурс] // Nuances of programming (дата публикации: 28.02.2020). - URL: <https://nuancesprog.ru/p/6713/> (дата обращения: 23.11.2021).

6. Alibaba создала нейросеть, способную распознать голос человека в толпе [Электронный ресурс] // Klevo.Net (дата публикации: 03.12.2018). - URL: <https://klevo.net/alibaba-sozdala-nejroset-sposobnuju-raspoznat-golos-cheloveka-a-v-tolpe/> (дата обращения: 23.11.2021).

7. Ученые создали нейросеть, которая определяет внешность человека по голосу [Электронный ресурс] // Screenlifer Media (дата публикации: 02.07.2019). - URL: <https://zen.yandex.com/media/screenlifer/uchenye-sozdali-neiroset-kotoraia-opredeliaet-vneshnost-cheloveka-po-golosu-cto-da-5d1b859624e56600ad2b4781> (дата обращения: 23.11.2021).

8. Нейросеть от Google AI различает спикеров на аудиозаписи [Электронный ресурс] // neurohive.io (дата публикации: 19.08.2019).- URL: <https://neurohive.io/ru/papers/nejroset-ot-google-ai-razlichaet-spikerov-na-audiozapisi> (дата обращения: 23.11.2021).

9. Нейросети в медицине: что это и как работает? [Электронный ресурс] // Медицинские скрининг системы (дата публикации: 22.09.2020). - URL: <https://celsus.ai/blog/kak-rabotaet-mediczinskaya-nejroset/> (дата

обращения: 23.11.2021).

10. Нейросеть для распознавания лиц. Часть 2 [Электронный ресурс] // NtechLab (дата публикации: 17.04.2017). URL: <https://findface.pro/blog/nejronnye-seti-i-raspoznavanie-obrazov-chast-2> (дата обращения: 23.11.2021).

11. Нейроная сеть видит галлюцинации [Электронный ресурс] // LINUX.ORG.RU (дата публикации: 19.06.2015). - URL: <https://www.linux.org.ru/forum/talks/11711194> (дата обращения: 23.11.2021).

12. Машинный слух. Нейросеть SoundNet обучили распознавать объекты по звуку [Электронный ресурс] // Хабр (дата публикации: 04.12.2016). - URL: <https://habr.com/ru/post/399659/> (дата обращения: 23.11.2021).

13. Audio AI: выделяем вокал из музыки с помощью свёрточных нейросетей [Электронный ресурс] // Хабр (дата публикации: 20.09.2019). - URL: <https://habr.com/ru/post/441090/> (дата обращения: 23.11.2021).

14. Математики из Армении создали сервис, который убирает посторонние звуки во время звонков [Электронный ресурс] // vc.ru (дата публикации: 29.01.2019). - URL: <https://vc.ru/services/56580-matematiki-iz-armenii-sozdali-servis-kotoryy-ubiraet-postoronnie-zvuki-vo-vremya-zvonkov> (дата обращения: 23.11.2021).

15. Рассказываем, как работает нейросеть Speech2Face. Она создаёт лица по голосу [Электронный ресурс] // iphones.ru (дата публикации: 02.12.2019). URL: <https://www.iphones.ru/iNotes/hz-poka-kakoy-syuda-zag-post-pro-golos-11-16-2019> (дата обращения: 23.11.2021).