МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра теоретических основ компьютерной безопасности и криптографии

**Использование нейросетей для обработки сигналов и изображений**

РЕФЕРАТ

студента 5 курса 531 группы

специальности 10.05.01 Компьютерная безопасность

факультета компьютерных наук и информационных технологий

Кузнецова Егора Дмитриевича

|  |  |  |
| --- | --- | --- |
| Преподаватель  доцент | \_\_\_\_\_\_\_\_\_\_ | И. И. Слеповичев |
|  | подпись, дата |  |

Саратов 2024

**СОДЕРЖАНИЕ**

[ВВЕДЕНИЕ 3](#_Toc185383774)

[ИСПОЛЬЗОВАНИЕ НЕЙРОСЕТЕЙ ДЛЯ ОБРАБОТКИ ИЗОБРАЖЕНИЙ 4](#_Toc185383775)

[1 Общие сведения 4](#_Toc185383776)

[2 Метрики для операций обработки изображений 6](#_Toc185383777)

[2.1 Метрики для предварительной обработки изображений 6](#_Toc185383778)

[2.2 Показатели для сегментации изображений 8](#_Toc185383779)

[2.3 Метрики для извлечения и классификации признаков 9](#_Toc185383780)

[3 Восстановление изображений с помощью нейронных сетей 10](#_Toc185383781)

[4 Улучшение изображений с помощью нейронных сетей 13](#_Toc185383782)

[6 Извлечение признаков с помощью нейронных сетей 17](#_Toc185383783)

[7 Классификация изображений с помощью нейронных сетей 20](#_Toc185383784)

[ИСПОЛЬЗОВАНИЕ НЕЙРОСЕТЕЙ ДЛЯ ОБРАБОТКИ СИГНАЛОВ 22](#_Toc185383785)

[1 Общие сведения 22](#_Toc185383786)

[2 Оценка и улучшение качества речевых сигналов 25](#_Toc185383787)

[3 Распознавание и синтез речевых сигналов 28](#_Toc185383788)

[ЗАКЛЮЧЕНИЕ 33](#_Toc185383789)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 34](#_Toc185383790)

## **ВВЕДЕНИЕ**

Нейронные сети – это адаптивные системы для обработки и анализа данных, которые представляют собой математическую структуру, имитирующую некоторые аспекты работы человеческого мозга и демонстрирующие такие его возможности, как способность к неформальному обучению, способность к обобщению и кластеризации неклассифицированной информации, способность самостоятельно строить прогнозы на основе уже предъявленных временных рядов. Возможность обучения является одним из главных преимуществ нейронных сетей перед традиционными алгоритмами. Технически обучение заключается в нахождении коэффициентов связей между нейронами. В процессе обучения нейронная сеть способна выявлять сложные зависимости между входными и выходными данными, а также выполнять обобщения. Это означает, что в случае успешного обучения сеть сможет создать верный результат на основании данных, которые или отсутствовали в обучающей выборке, или были неполными, зашумленными и частично искаженными.

Главным их отличием от других методов является то, что нейросети в принципе не нуждаются в заранее известной модели, а строят ее сами только на основе предъявляемой информации. Именно поэтому нейронные сети и генетические алгоритмы вошли в практику всюду, где нужно решать задачи прогнозирования, классификации, управления, иными словами, в области человеческой деятельности, где есть плохо алгоритмизуемые задачи, для решения которых необходимы либо постоянная работа группы квалифицированных экспертов, либо адаптивные системы автоматизации, каковыми и являются нейронные сети.

## **ИСПОЛЬЗОВАНИЕ НЕЙРОСЕТЕЙ ДЛЯ ОБРАБОТКИ ИЗОБРАЖЕНИЙ**

### **Общие сведения**

В области обработки изображений существует множество методик, предназначенных для решения таких задач, как шумоподавление, улучшение качества, сегментация, извлечение признаков и классификация. Эти методы в совокупности решают проблемы и открывают возможности, связанные с различными аспектами анализа и обработки изображений, что позволяет применять их в различных областях. Каждая из этих методик способствует углублению нашего понимания изображений, извлечению важной информации и принятию обоснованных решений на основе визуальных данных.

Традиционные методы обработки изображений и DL-модели представляют собой два различных подхода к решению задач анализа изображений. Традиционные методы часто основаны на созданных вручную алгоритмах и эвристике, включающих серию предопределенных шагов для обработки изображений. Модели DL изучают представления объектов непосредственно из данных, что позволяет им автоматически извлекать сложные функции, которые традиционные методы могли бы пропустить.

В области шумоподавления выделяются такие методы, как Self2Self нейронные сети, шумоподавляющие CNNS (сверточные нейронные сети), DFT-сети (сети на основе глубокого преобразования объектов) и MPR-CNN (параллельная сверточная нейронная сеть с мультиразрешением), которые обеспечивают снижение уровня шума при одновременном решении задач увеличения объема данных и настройки параметров.

Улучшение изображений с помощью таких подходов, как R2R (преобразование изображений при слабом освещении в изображения при нормальном освещении) и LeNet, демонстрирует потенциал для повышения качества изображения, хотя сложности, связанные с изображениями естественной среды и аутентичностью, сохраняются.

Методы сегментации, в том числе PSPNet и Mask-RCNN (сверточная нейронная сеть на основе маски), демонстрируют точность при выделении объектов, а также справляются со сложностями, такими как перекрывающиеся объекты и проблемы с надёжностью.

Для извлечения признаков такие методы, как CNN и HLF-DIP (предварительный анализ изображений на основе функции потерь Хубера), способны автоматизировано работать при выявлении характеристик изображения, но при этом имеют недостатки в плане интерпретируемости и сложности.

Методы классификации варьируются от рекуррентных нейронных сетей до CNN-LSTM (свёрточные нейронные сети на основе долгой краткосрочной памяти), демонстрируя свой потенциал в точности классификации, несмотря на проблемы, связанные с вычислительными требованиями и интерпретируемостью.

По мере развития этой области решение таких проблем, как вычислительные ресурсы и надёжность, по-прежнему имеет ключевое значение для максимального раскрытия потенциала методов обработки изображений [1].

### **Метрики для операций обработки изображений**

Метрики служат ключевыми инструментами для оценки эффективности и влияния различных методов обработки изображений. Они предоставляют количественные измерения, которые позволяют исследователям и практикам проводить непредвзятый анализ и облегчают сравнение результатов, полученных с помощью различных методов. С помощью этих показателей сложная и зачастую субъективная область обработки изображений может стать более объективной, что приводит к принятию обоснованных решений и прогрессу в этой области. В этом разделе использовалась информация из источников [1], [2], [3].

### **Метрики для предварительной обработки изображений**

* + 1. **Среднеквадратичная ошибка (MSE)**

Среднеквадратичная ошибка применяется в случаях, когда требуется подчеркнуть большие ошибки и выбрать модель, которая дает меньше именно больших ошибок.

где M и N – размеры изображения. – это значения пикселей в позиции на исходном и обработанном изображениях соответственно.

* + 1. **Пиковое отношение сигнала к шуму (PSNR)**

PSNR обычно используется для оценки качества восстановленных изображений. Он сравнивает исходное и восстановленное изображения, учитывая среднеквадратичную ошибку между значениями пикселей.

где MAX – максимально возможное значение пикселя (255 для 8-битных изображений), MSE – среднеквадратичная ошибка между исходным и очищенным от шума изображениями.

* + 1. **Индекс структурного сходства (SSIM)**

SSIM применим и для восстановления изображений. Он оценивает сходство между исходным и восстановленным изображениями с точки зрения яркости, контрастности и структуры. Более высокие значения SSIM указывают на более высокое качество восстановления.

где – средние значения исходного изображения и изображения с шумоподавлением, – отличиям исходного изображения от изображения с шумоподавлением, – ковариация между исходным изображением и изображениями с шумоподавлением, – константы, позволяющие избежать деления на ноль.

* + 1. **Средний индекс структурного сходства (MSSIM)**

MSSIM расширяет SSIM на несколько участков изображения и вычисляет среднее значение SSIM для этих участков.

где – фрагменты исходного и улучшенного изображений.

* + 1. **Средняя абсолютная погрешность (MAE)**

Средняя абсолютная погрешность измеряет, насколько прогнозы близки к фактическим результатам. Это среднее значение всех ошибок модели, где ошибка модели – абсолютное расстояние между значением прогнозируемой метки и значением правильной метки.

где n – количество выборок.

* + 1. **NIQE (оценка качества изображения по критерию естественности)**

NIQE количественно оценивает естественность изображения, измеряя отклонение локальной статистики от естественных изображений. Он вычисляет среднее значение локальных различий в яркости и контрастности.

* + 1. **FID (дистанция начала Фреше)**

FID измеряет расстояние между двумя распределениями (реальными и сгенерированными изображениями) с помощью расстояния Фреше между их представлениями признаков, рассчитанными предварительно обученной нейронной сетью.

### **Показатели для сегментации изображений**

**2.2.1 Пересечение через объединение (IoU)**

**IoU** – это метрика, используемая для оценки производительности и точности алгоритмов обнаружения объектов. Она количественно измеряет степень перекрытия между двумя ограничивающими рамками: одна представляет «ground truth» (фактическое расположение объекта), а другая — «предсказание» модели для того же объекта.

**2.2.2 Средняя точность (AP)**

AP измеряет точность при различных уровнях полноты и вычисляет площадь под кривой «точность-полнота». Используется для оценки моделей обнаружения объектов и сегментации экземпляров.

**2.2.3 Коэффициент сходства игральных костей**

Коэффициент сходства Диза, также известный как коэффициент Соренсена-Диза, – распространённый показатель для оценки сходства между двумя наборами данных. В контексте сегментации изображений он количественно оценивает перекрытие между прогнозируемой сегментацией и эталонной сегментацией, принимая во внимание как истинные, так и ложные срабатывания. Значение DSC варьируется от 0 до 1, где более высокие значения указывают на лучшее перекрытие между прогнозируемой сегментацией и эталонной сегментацией. Значение DSC, равное 1, соответствует идеальному совпадению.

**2.2.4 Средняя точность (AA)**

Средняя точность измеряет общую точность сегментации путём вычисления процента правильно классифицированных пикселей по всем классам.

где N – количество классов, – истинные положительные и истинно отрицательные результаты для i-го класса, – общее количество пикселей в классе.

### **Метрики для извлечения и классификации признаков**

**2.3.1 Accuracy**

Отношение количества правильно предсказанных примеров к общему количеству примеров. Обычно используется для сбалансированных наборов данных, но может вводить в заблуждение при работе с несбалансированными наборами данных.

**2.3.2 Precision**

Отношение количества точных положительных прогнозов к общему количеству положительных прогнозов. Это показатель способности модели избегать ложных положительных результатов.

**2.3.3 Recall**

Отношение количества верно предсказанных положительных результатов к общему количеству реальных положительных результатов. Это показатель способности модели правильно определять положительные результаты.

**2.3.4 Показатель F1**

Показатель F1 также называется сбалансированной F-оценкой или F-мерой. Это среднее гармоническое значение точности и полноты. Показатель F1 полезен в том случае, если необходимо найти баланс между точностью и полнотой.

**2.3.5 Specificity**

Отношение истинных негативных прогнозов к общему числу фактических негативных случаев.

**2.3.6 Кривая ROC**

Графическое представление компромисса между долей истинно положительных результатов и долей ложноположительных результатов при изменении порога классификации. Эти показатели обычно используются при бинарной классификации.

### **Восстановление изображений с помощью нейронных сетей**

Восстановление изображений – важнейший процесс, направленный на восстановление целостности и визуального качества изображений, которые были повреждены или искажены.

Серьезным препятствием на пути к получению качественных изображений является шум – непреднамеренное изменение значений пикселей, которое приводит к появлению нежелательных артефактов и может привести к потере важной информации. Различные типы шума, такие как гауссов шум, характеризующийся случайным распределением, импульсный шум, вызывающий появление ярких и темных пикселей, и спекл-шум, возникающий из-за помех, могут ухудшать качество изображений. Исторически сложилось так, что традиционные методы восстановления изображений включали в себя ряд способов смягчения последствий деградации и зашумления. К этим методам относятся фильтры наименьших квадратов с ограничениями, методы слепой деконволюции, направленные на устранение эффекта размытия, фильтры Вайнера и обратные фильтры для повышения соотношения сигнал/шум. Методы шумоподавления такие как шумоподавление по общей вариации (TVD) и нелокальные средства (NLM), дополнительно способствуют эффективному снижению уровня случайных шумов при сохранении важных деталей изображения, что в совокупности повышает целостность и визуальную чёткость изображений.

Свёрточные нейронные сети хорошо обучаются и извлекают сложные признаки из изображений, что позволяет им распознавать закономерности и нюансы, которые могут быть трудноразличимы для традиционных методов. Благодаря обширному обучению на больших наборах данных эти сети могут значительно повысить качество восстановленных изображений, часто превосходя возможности традиционных методов.

В таблице 1 представлено краткое описание моделей глубокого обучения для восстановления изображений, включая их преимущества и недостатки [1].

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Авторы | Использованные методы | Наборы данных | Метрики | Данные метрик | Преимущества | Недостатки |
| Ворку Дж. Сори и др. (2021) | DFT-сеть для шумоподавления и обнаружения | Снимки компьютерной томографии из KDSB и LUNA 16 | Accuracy, recall, specificity | A: 0.878  R: 0.874  S: 0.891 | Эффективное устранение дисбаланса меток изображений | Потенциальная потеря деталей при шумоподавлении |
| Сяобэнь Цзян и др. (2021) | MPR-CNN для параллельного подавления остаточных шумов | Рентгеновские снимки грудной клетки с COVID-19 | PSNR,  SSIM | PSNR: 36.368  SSIM: 0.895 | Надежность и временная сложность | Настройка гиперпараметров |
| Тонъяо Панг и др. (2021) | Метод R2R для снижения уровня шума | Использованный набор данных SIDD Benchmark | PSNR,  SSIM | Шум 50  PSNR: 26.13  SSIM: 0.709 | Результаты, эквивалентные контролируемому обучению | Вычислительные требования и обработка шума |
| Раджу Хасти, Дмитрий Шин (2022) | Использовались стандартные и модифицированные архитектуры CNN | Изображения, полученные в результате рассеяния Ми | MSE,  PSNR | MSE: 0.0053  PSNR: 22.757 | Предотвращение переобучения | Потребляемая память и временная сложность |
| Кейван Фаги Нирези, Чонг-Юнг Чи (2022) | HLF-DIP алгоритм для обработки гиперспектральных изображений | Гиперспектральные изображения из HYDICE | MPSNR, MSSIM,  MSAM,  MFSIM | Шум 40  MPSNR: 49.49  MSSIM: 0.998  MSAM: 0.024  MFSIM: 0.999 | Отсутствие регуляризаторов, удобство использования | Настройка по одному параметру |
| Мирал С. Тауфик и др. (2022) | Оцененные модели (Noise2Noise) для шумоподавления | Анализ изображений микрокомпьютерной томографии | PSNR, SSIM | PSNR: 20.607  SSIM: 0.546 | Временная сложность | Предельное обобщение |
| Ичжэнь Мэн, Цзюнь Чжан (2022) | Сеть ConvNet для шумоподавления на серых изображениях | BSD86 | PSNR, SSIM, FOM | PSNR: 26.44  SSIM: 0.6797  FOM: 1 | Улучшенное рецептивное поле | Сложности интерпретации |

Таблица 1 – Модели глубокого обучения для восстановления изображений

### **Улучшение изображений с помощью нейронных сетей**

Улучшение изображения – процесс обработки изображения с целью повышения его визуального качества и удобства восприятия человеком.

Стандартные методы улучшения изображений включают в себя ряд приёмов, в том числе сопоставление гистограмм для настройки распределения интенсивности пикселей, адаптивное выравнивание гистограммы с ограничением контрастности (CLAHE) для повышения локального контраста и такие фильтры, как фильтр Винера и медианный фильтр, для уменьшения шума. Для повышения чёткости и резкости изображения также часто используются линейная регулировка контрастности и фильтрация размытия.

В последние годы методы глубокого обучения стали эффективным подходом к улучшению изображений. Эти методы используют большие массивы данных и сложные архитектуры нейронных сетей для изучения закономерностей и особенностей изображений, что позволяет восстанавливать и улучшать изображения с впечатляющими результатами.

Исследования включают в себя ряд инновационных методов, в том числе интеграцию теории Retinex и априорных данных о глубоких изображениях в новом методе RetinexDIP, повышение надёжности с помощью нечёткой логики для снижения риска переобучения, а также объединение таких известных методов, как размытие по Гауссу, высокочастотная акцентирующая фильтрация и CLAHE с архитектурами EfficientNet-B4, ResNet-50 и ResNet-18 для повышения обобщаемости и надёжности. Кроме того, универсальный фреймворк pix2pixHD и LeNet вносят свой вклад в исследования. Глубокие свёрточные нейронные сети демонстрируют надёжность, но требуют тщательной настройки гиперпараметров. Наконец, MSSNet-WS (многомасштабная многоэтапная сеть) эффективно сходится и устраняет переобучение.

В таблице 2 представлено краткое описание моделей глубокого обучения для восстановления изображений, включая их преимущества и недостатки [1].

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Авторы | Использованные методы | Наборы данных | Метрики | Данные метрик | Преимущества | Недостатки |
| Юхуэй Цюань и др. (2020) | CV-CNN для удаления размытости изображения | Лодка, Пара, Мужчина, Зебра, Лена | PSNR, SSIM | PSNR: 49.2  SSIM: 0.990 | Эффективность модели, предотвращение переобучения | Настройка фильтров, обобщение признаков |
| Цзин-Та Лу и др. (2021) | Полносвязные нейронные сети с фильтрами для подавления шума | Изображение Лены, изображение лодки из Google | PSNR/  MSSIM | Шум 10  Лена: 49.2/0.999  Лодка: 41.7/0.998 | Обработка шума, улучшение пикселей | Потенциальная потеря деталей |
| Дэн Джин и др. (2021) | pix2pixHD для улучшения изобра-жений мультидете-кторной компью-терной томографии | Изображения микрокомпьютерной и мультидетекторной комп. томографии | SSIM,  FID | SSIM:  0.804 ± 0.037  FID:  43.598 ± 9.108 | Повышение качества, стоимость | Требования к данным, возможность переобучения |
| Гофа Ли (2021) | LeNet для восстановления изображений в условиях низкой освещенности | BDD100K | MSE,  PSNR,  SSIM | MSE: 401.26  PSNR: 24.38  SSIM: 0.91 | Обобщение, надежность | Ограничения естественной среды (Real-world limitations) |
| Сяньцзе Гао (2022) | RetinexDIP для улучшения изображения с помощью сравнений | DICM, Fusion, LIME, MEF, NPE, VV | NIQE, NIQMC, CPCQI | NIQE: 3.5294  NIQMC: 5.0398  CPCQI: 1.0437 | Более быстрая конвергенция, эффективность по времени | Проблемы в сложных сценах |
| Киен Ким и др. (2022) | MSSNet-WS для удаления размытос-ти с изображения | Изображения из реального мира, полученные с GoPro | PSNR, MSSIM, | PSNR: 31.83  MSSIM: 0.950 | Вычислительная эффективность | Ограничения естественной среды, размы-тие границ |

Таблица 2 – Модели глубокого обучения для улучшения изображений.

1. **Сегментация изображений с помощью нейронных сетей**

Сегментация изображения – важнейший процесс, который заключается в разделении изображения на отдельные сегменты на основе определённых различимых характеристик, таких как интенсивность, цвет, текстура или пространственная близость. Этот метод делится на две основные категории: семантическая сегментация и сегментация экземпляров. Семантическая сегментация относит каждый пиксель к определённому классу на входном изображении, что позволяет идентифицировать отдельные области объектов. Сегментация экземпляров делает ещё одно действие, не только распределяя пиксели по классам, но и различая отдельные экземпляры этих классов на изображении. Традиционные методы сегментации предполагают разделение данных, например изображений, на чётко определённые сегменты в соответствии с заранее заданными критериями. К распространённым методам относятся пороговая сегментация, при которой пиксели классифицируются на области объекта и фона с использованием определённых пороговых значений интенсивности, сегментация на основе областей, при которой пиксели со схожими характеристиками группируются в однородные области, и обнаружение границ для выявления значительных переходов интенсивности, которые могут указывать на потенциальные границы. Тем не менее, традиционные методы сталкиваются со сложностями, когда дело доходит до обработки сложных форм, динамического фона и шума в данных. Более того, ручное создание признаков для различных сценариев может быть трудоёмким и не всегда подходит для разных контекстов. Глубокие нейронные сети в свою очередь обладают удивительной способностью извлекать сложные признаки непосредственно из необработанных данных, устраняя необходимость в ручной обработке признаков. Это позволяет им улавливать тонкие пространственные взаимосвязи и адаптироваться к изменениям, эффективно устраняя ограничения, присущие традиционным методам.

В таблице 3 перечислены сильные и слабые стороны различных изученных моделей глубокого обучения [1].

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Авторы | Использованные методы | Наборы данных | Метрики | Данные метрик | Тип | Преимущества | Недостатки |
| Крити Махаджан и др. (2021) | Модель глубокого итеративного обучения на основе кластеризации с использованием CNN | гиперспектральные наборы данных Indian Pines, University of Pavia | OA, AA, время вычислений | OA: 91.21  AA: 78.14  Time: 0.525 | Экземпляр | Последовательность, точность и скорость | Сложность кривой обучения |
| Лина Лю (2021) | CNN, CCL, MSFA | ISBI2017 | Мера Жаккара, коэффицент Диза,  ACC, SE, SP | Мера Жаккара 79.46  (ACC): 94.32 SEN: 88.76 | Семантический | Надежность | Противоречивые результаты в определенных случаях |
| Артзай Пикон и др. (2022) | PSPNet | Изображения различных видов сорняков | BAC, DSC | DSC: 47.97 | Семантический | Снижение затрат и усилий на аннотацию изображений растительности | Ограничения и точность синтетических данных для визуально по-хожих видов сорняков |
| Хасан Ашраф и др. (2022) | UNet, ResUNet, ResUNet ++ | ISIC-2016, ISIC-2017 | Коэффициент Диза, мера Жаккара, Precision, Recall | 80.73% и 90.02% | Семантический | Высокая масштабируемость и надежность | Влияние условий приобретениния заболевания |

Таблица 3 – Модели глубокого обучения для сегментации изображений.

### **Извлечение признаков с помощью нейронных сетей**

Извлечение признаков – фундаментальный процесс в обработке изображений и компьютерном зрении, который заключается в преобразовании необработанных пиксельных данных в более компактное и информативное представление, часто называемое признаками. Эти признаки отражают важные характеристики изображения, что упрощает для алгоритмов понимание и анализ изображений для решения различных задач, таких как распознавание объектов, классификация изображений и сегментация. Традиционные методы извлечения признаков были распространены до появления глубокого обучения и включали в себя методы, которые анализировали информацию на уровне пикселей. Анализ основных компонентов (PCA) – статистический метод, который уменьшает размерность данных при сохранении максимально возможной исходной дисперсии. Он определяет ортогональные оси (главные компоненты), вдоль которых данные варьируются в наибольшей степени. Независимый компонентный анализ (ICA) направлен на поиск линейного преобразования данных в статистически независимые компоненты. Он часто используется для разделения смешанных источников на изображениях, например, для отделения разных источников изображения от одного смешанного изображения. Локально-линейное вложение (LLE) – нелинейный метод уменьшения размерности, который направлен на сохранение локальной структуры точек данных. Он находит низкоразмерное представление данных, сохраняя при этом связи между соседними элементами. Эти традиционные методы извлечения признаков широко используются и позволяют получать ценную информацию и представления для различных задач анализа изображений. Однако они часто основаны на ручных методах, разработанных экспертами или специалистами в предметной области, что может быть трудоёмким и не всегда подходит для разных типов изображений или задач.

Несмотря на эффективность традиционного извлечения признаков для решения конкретных задач и его способность предоставлять аналитическую информацию о данных, у него есть свои ограничения. Традиционные методы часто требуют участия экспертов для создания признаков, что может быть трудоёмким процессом и приводить к упущению из виду сложных взаимосвязей или закономерностей в данных. Кроме того, традиционные методы могут сталкиваться с трудностями при работе с данными высокой размерности или в сценариях, где признаки сложно определить.

Глубокие нейронные сети в свою очередь самостоятельно учатся извлекать значимые признаки непосредственно из необработанных данных, устраняя необходимость в ручном извлечении признаков. Это упрощает выявление сложных взаимосвязей, закономерностей и многогранных взаимодействий, которые могут быть упущены из виду при использовании традиционных методов.

В таблице 4 кратко описаны показатели, сильные стороны и ограничения различных моделей глубокого обучения, используемых для улучшения характеристик [1].

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Авторы | Использованные методы | Наборы данных | Метрики | Данные метрик | Тип | Преимущества | Недостатки |
| Сонгита и Шарма (2021) | CNN | Изображения сетчатки | Class score, accuracy, precision, specificity, recall | 97% | Паттерн | Гибкость и адаптивность | Отсутствие кастомизации модели |
| К. Шанкар (2022) | FM-ANN, GLCM, GLRM, LBP | Рентгеновские снимки грудной клетки | Accuracy, Sensitivity, Specificity, F-показатель | 95.1% и 95.7% | Текстура | Эффективное извлечение признаков и настройка параметров | Интерпретация модели |
| Шахаб Ахмад и др. (2022) | AlexNet-GRU | PCam от Kaggle | Accuracy,  Precision, Sensitivity, Specificity | 99, 5% | Цвет | Высокая точность и улучшенные показатели производительности | Высокая вре-менная слож-ность и требо-вания к специ-ализированному оборудованию |
| Р. Арти, Г. Ришма и др. (2023) | MRCNN | Изображения отходов в режи-ме реального времени | Accuracy | 97% | Геометрия | Надежность | Эффективность систем сбора и разделения отхо-дов в режиме ре  ального времени |

Таблица 4 – Модели глубокого обучения для извлечения признаков изображений.

### **Классификация изображений с помощью нейронных сетей**

Классификация изображений – фундаментальная задача компьютерного зрения, которая заключается в распределении изображений по заранее заданным классам или меткам.

Традиционная классификация – метод анализа данных, который предполагает распределение точек данных по определённым классам или категориям на основе заранее установленных правил и признаков. До появления глубокого обучения для этой цели широко использовались несколько традиционных методов, в том числе деревья решений, метод опорных векторов (SVM), наивный байесовский классификатор и метод k-ближайших соседей (k-NN). В традиционной классификации эксперты тщательно разрабатывают и выбирают признаки, которые содержат важную информацию из данных. Эти признаки обычно выбираются на основе знаний и идей из предметной области с целью выявления отличительных характеристик. Традиционные методы классификации, эффективные в различных сценариях, часто требуют ручной разработки признаков, что может занимать много времени и не позволяет в полной мере выявить сложные закономерности, присутствующие в сложных наборах данных. Появление глубокого обучения произвело революцию в классификации, автоматизировав процесс извлечения признаков. Глубокие нейронные сети напрямую обучаются иерархическим представлениям на основе необработанных данных, устраняя необходимость в создании признаков вручную. Примечательно, что свёрточные нейронные сети (CNN) отлично справляются с задачами классификации изображений, а рекуррентные нейронные сети (RNN) демонстрируют эффективность при работе с последовательными данными. Эти модели глубокого обучения часто превосходят традиционные методы при решении сложных задач в различных областях.

Таблица 5 представляет собой краткий обзор преимуществ и ограничений в области классификации изображений, в котором рассматриваются различные модели глубокого обучения [1].

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Авторы | Использованные методы | Наборы данных | Метрики | Данные метрик | Тип | Преимущества | Недостатки |
| Молуд Абдар и др. (2021) | TWDBDL (трехстороннее байесовское глубокое обучение на основе принятия решений) | Наборы данных по раку кожи | Area under the curve (AUC) | 88.95% и 90.96% | Многоклассовый | Гибкий гибридный подход, эффективная количественная оценка неопределенности | Ограничения вычислительной мощности и временных ресурсов |
| Абдуллахи Умар Ибрагим и др. (2021) | Модель AlexNet | Заболевания легких | Sensitivity, Specificity, Accuracy (%) | 94% | Многоклассовый | Экономия времени, сокращение трудозатрат | Относительно небольшой набор данных о случаях пневмонии при COVID-19 |
| Юнчен Конг и др. (2022) | CNN и SVM | Caltech256 | Accuracy | 93,4% | Многоклассовый | Улучшенная генерализация, предотвращение переобучения | Помеченные выборки, измен  чивость набора данных, интер  претируемость |
| Хармандип Сингх Джилл и др. (2022) | Гибрид CNN-RNN | Фрукты | Precision, Recall, F-показатель, Accuracy (%) | 99, 5% | Многоклассовый | Сравнение с другими мето-дами, после  довательная маркировка | Зависимость от данных, интер-претируемость, вычислительная трудоемкость |
| Абу-Джейми и др. (2023) | VGG16 | Фрукты | Precision, Recall, F-показатель, Accuracy (%) | 100% | Многоклассовый | Высокий уровень точности, эффективное использование CNN | Потенциальное переобучение, смещение набора данных |

Таблица 5 – Модели глубокого обучения для классификации изображений.

## **ИСПОЛЬЗОВАНИЕ НЕЙРОСЕТЕЙ ДЛЯ ОБРАБОТКИ СИГНАЛОВ**

### **Общие сведения**

В области речевой обработки, как и одно-два десятилетия назад, большинство алгоритмов работает с сигналами, представленными в частотной области или иной области трансформант. Зачастую вопрос выбора формы представления сигнала для решения конкретной задачи является не менее важным, чем вопросы, связанные с последующей обработкой. Этап формирования признаков (представления) сигнала необходим для создания компактного и информативного представления данных, с которыми работает

конкретный алгоритм. В рамках концепции глубокого обучения роль выбора признаков может показаться менее значительной, однако знание и понимание базовых основ в этой области все же необходимы [7].

В последние десятилетия в обработке речевых и других звуковых сигналов вместо традиционных спектрограмм, полученных на основе оконного преобразования Фурье (в соответствии с рисунками 1, 2), часто прибегают к мелчастотным кепстральным коэффициентам (MFCC – Mel Frequency Cepstral Coefficients) и спектрограммам на их основе.

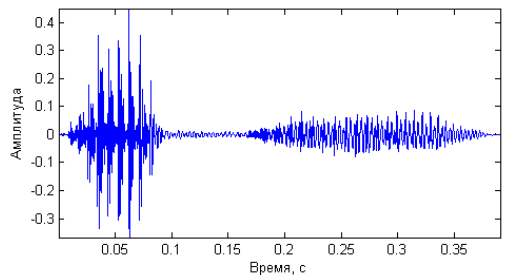


Рисунок 1 – Представление речевого сигнала во временной области

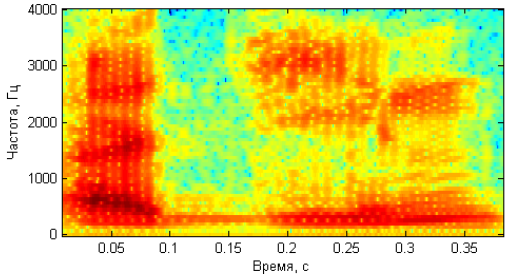


Рисунок 2 – Представление речевого сигнала в частотной области

Действительный кепстр сигнала x, иногда называемого просто кепстр, вычисляется путем определения натурального логарифма величины преобразования Фурье x, затем получения обратного преобразования Фурье получившейся последовательности:

где – спектр входного сигнала.

В настоящее время кепстральный анализ рассматривается в качестве вида гомоморфной обработки и находит применение в самых разных областях [5].

Мел – психофизическая единица высоты звука, применяется главным образом в музыкальной акустике. Было установлено, что высота отдельных звуков нелинейно связана с их частотой. Отдельные научные группы изучали эту взаимосвязь, поэтому существует несколько единиц высоты звука. В настоящее время наиболее востребованными и широко используемыми являются мел-единицы. По определению звуковые колебания частотой 1000 Гц при уровне громкости 40 фон, воздействующие на человека с нормальным слухом, вызывают у него восприятие высоты звука, приравненной к 1000 мел.

Формулы, преобразующие мелы в Герцы и в обратном направлении:

где – частота, выраженная в Герцах, – высота звука в мелах [4].

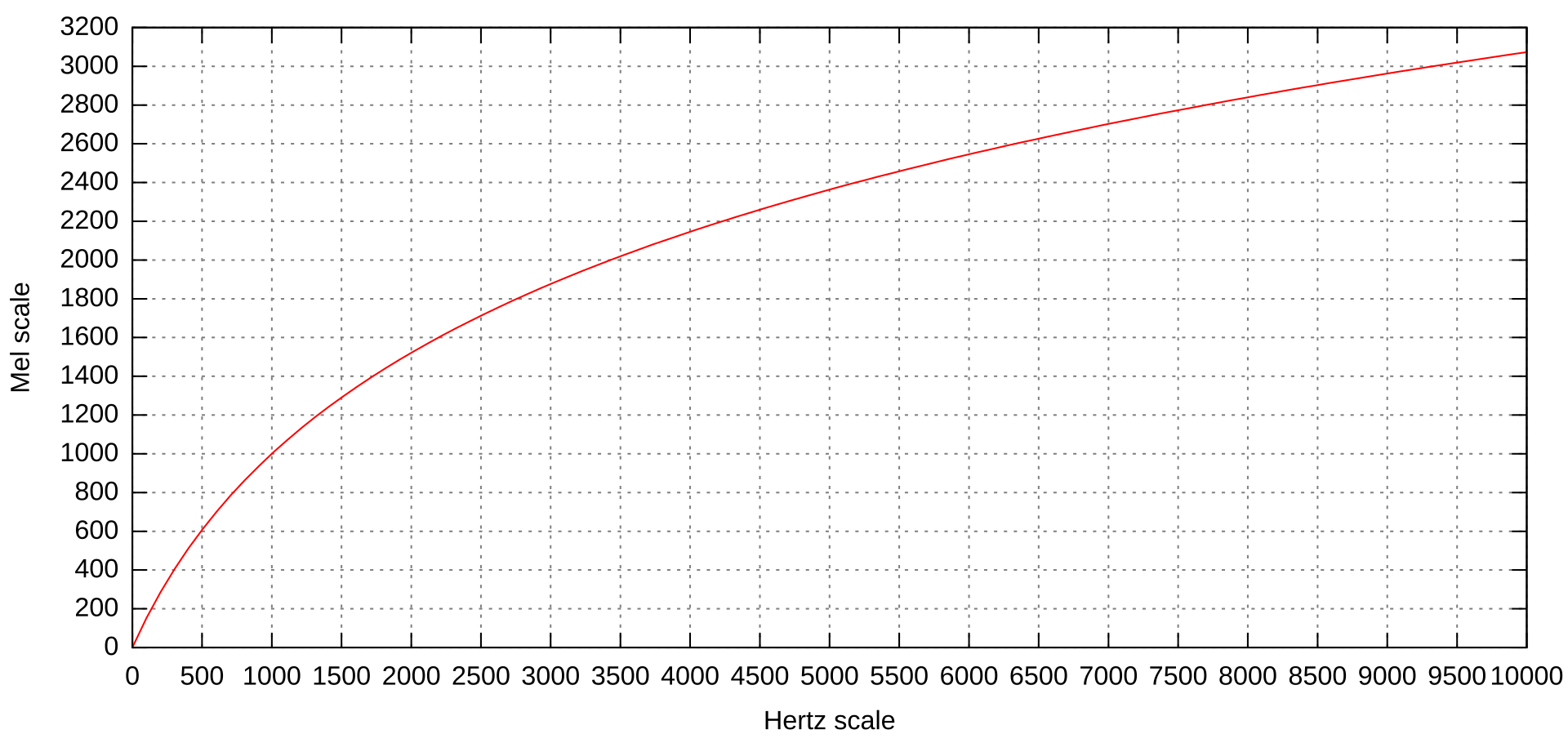


Рисунок 3 – Взаимосвязь высоты звука, выраженной в мелах, и его частоты в Герцах

Алгоритм вычисления мел-частотных кепстральных коэффициентов для произвольного звукового сигнала выглядит следующим образом: необходимо вычислить амплитудный спектр фрагмента звукового сигнала. Для этого используется оконное преобразование Фурье. Длина и шаг окна, а также тип используемой оконной функции зависят от особенностей решаемой задачи. Очень часто в речевой обработке используют окна длиной от 128 до 2048 отсчетов. Длину временного промежутка в секундах, которому соответствует конкретная длина окна, определяет частота дискретизации. Выбранная длина окна определяет количество спектральных отсчетов в найденном с помощью оконного преобразования Фурье спектре. В большинстве практических случаев их число измеряется сотнями, а иногда и тысячами. Слуховой системе человека, как и многим современным алгоритмам речевой обработки, не требуется столь детальное представление звуковых сигналов. Если в случае с человеком форма представления акустических сигналов определяется физиологией слуховой системы, то в нашем случае переход к более обобщенному, агрегированному спектральному представлению осуществляется за счет применения мел-фильтров. Банк таких фильтров можно построить и во временной области, но на практике чаще идут по пути применения треугольных функций (в соответствии с рисунком 4), равномерно распределённых на мел-шкале к спектрограмме, полученной на предыдущем шаге. Каждый столбец исходной спектрограммы умножается на соответствующий мел-фильтр (спроецированный на частотную шкалу), после чего получается вектор чисел, по размеру равный количеству мел-фильтров. Таким образом значительно сокращается количество составляющих, а разрешение по частоте становится неравномерным – с большим акцентом на низкие частоты.

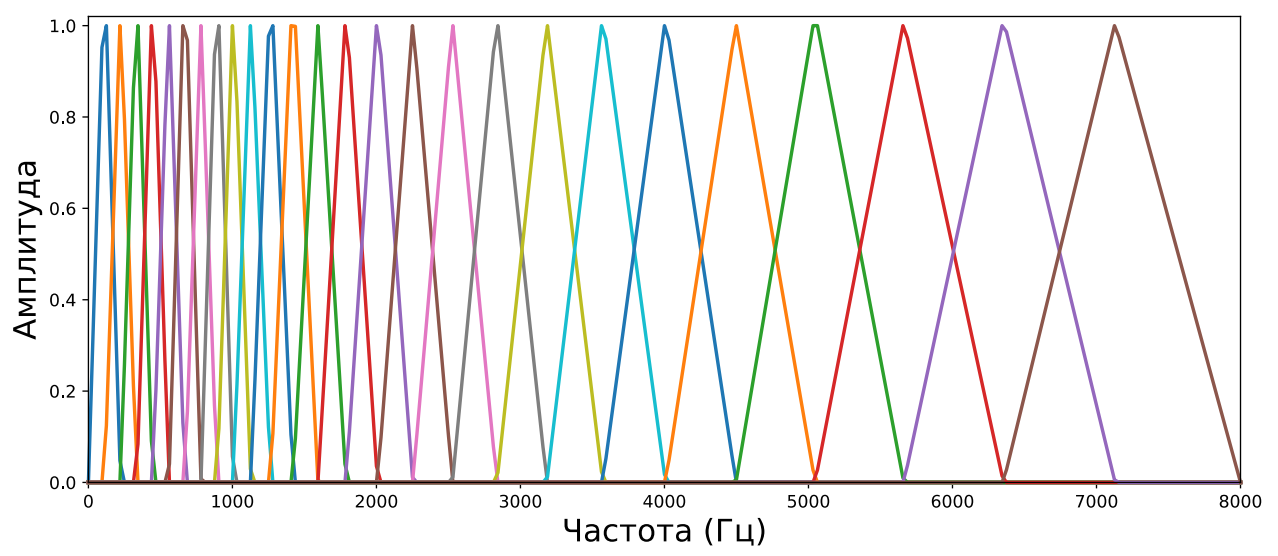


Рисунок 4 – АЧХ банка треугольных мел-фильтров

Многие современные алгоритмы, в том числе и нейросетевые, используют мелспектрограммы в качестве входных признаков звуковых сигналов.

### **Оценка и улучшение качества речевых сигналов**

Очень часто в реальных условиях речевой сигнал подвергается зашумлению и иным искажениям, ухудшающим качество и разборчивость речи. Для улучшения качества таких сигналов существуют специальные методы и алгоритмы. Однако возможное количество разных типов искажений велико и их характер столь отличен друг от друга, что, как правило, приходится идти по пути создания отдельных методов. Поэтому, кроме методов шумоподавления, существуют также методы дереверберации, эхоподавления, восстановления клиппированных сигналов и т.д. Сравнение существующих методов улучшения качества, а также их разработка невозможна без объективных показателей качества. В отличие от субъективных методов, в которых непосредственная оценка качества осуществляется людьми, объективные можно полностью реализовать программно или аппаратно, что существенно ускоряет и удешевляет процедуру.

Объективные показатели качества принято разделять на эталонные и неэталонные. Для работы последних, в отличие от эталонных, не требуется чистого (незашумленного, неискаженного) сигнала. Долгое время практически все показатели были эталонными. По сути, эти методы измеряют расстояние между чистым и зашумленным сигналами в некотором выбранном пространстве, однако выбор признаков и метрики – сложная научная задача. Самые простые подходы, основанные на измерении отношения сигнал/шум или вычислении среднеквадратической ошибки между сигналами, мало подходят для оценки качества речи. Результаты таких измерений будут слабо коррелировать с субъективными оценками качества, проведенными экспертами.

Некоторым промежуточным, но важным шагом стало создание показателя качества PESQ (Perceptual Evaluation of Speech Quality). Процесс оценки PESQ включает в себя сравнение эталонного (исходного) речевого сигнала с ухудшенным речевым сигналом после передачи или обработки. Алгоритм использует модель восприятия, которая имитирует слуховую систему человека, принимая во внимание характеристики человеческого слуха, такие как частотная характеристика, эффекты маскировки и временная маскировка. Анализируя влияние искажений на восприятие речи, PESQ генерирует оценку качества, которая отражает воспринимаемое качество ухудшенного сигнала [6].

Развитие нейронных сетей, а также упрощение интерфейсов средств разработки привели к появлению в последние годы значительного количества нейросетевых показателей качества, в том числе и неэталонных. В качестве признаков речевых сигналов на входе нейросети чаще всего используют спектрограммы, мел-спектрограммы, а также мел-частотные кепстральные коэффициенты. Существенной проблемой является то, что для обучения таких алгоритмов требуются большие базы речевых сигналов с указанными оценками качества, полученными субъективным методом. Однако и эта проблема может быть частично решена. Можно обучить нейронную сеть предсказывать значения PESQ (или другого существующего показателя качества) в неэталонном режиме. В этом случае речевую базу, искаженную различными способами, автоматически размечают с использованием эталонного метода, обладающего высокой достоверностью. Далее эта база используется для обучения неэталонного нейросетевого алгоритма. Примером реализации подобного подхода является алгоритм Quality-Net, построенный на основе двунаправленной сети долгой краткосрочной памяти (BLSTM). Другим примером использования нейронных сетей в задаче неэталонной оценки качества может служить алгоритм DNSMOS, разработанный сотрудниками компании Microsoft. Для его обучения использовались данные, размеченные в соответствии со стандартом оценки качества ITU-T P.808.

При решении задачи шумоподавления многие традиционные алгоритмы строятся на применении функции коррекции спектра (gain function) к спектрограмме зашумленного сигнала, а затем преобразованием результата обработки во временную область посредством обратного преобразования Фурье. Другая группа алгоритмов строится на применении в спектральной области бинарных масок. Основное отличие алгоритмов состоит в методиках оценки функции коррекции спектра или бинарных масок.

Анализ современных подходов к подавлению шума в речевых сигналах позволяет сделать вывод, что для решения этой задачи возможно применять самые разные виды нейросетей. Простейший алгоритм шумоподавления может быть реализован даже на основе простой полносвязной сети. Наибольшее распространение получили алгоритмы на основе сверточных и рекуррентных нейросетей, однако существуют и менее распространенные подходы, например на основе специально разработанной генеративной сети SEGAN и даже сети WaveNet, изначально разработанной для синтеза речи и других звуковых сигналов. Но даже при фиксированном типе нейросети, например сверточной, существует большее количество настраиваемых гиперпараметров, процедур подготовки данных, методик обучения и так далее, что способно породить большое количество исследований и получаемых в ходе них алгоритмов.

### **Распознавание и синтез речевых сигналов**

Под распознаванием речевых сигналов чаще всего подразумевают преобразование речи в текст. Именно такой вариант распознавания имеет наибольшую коммерческую востребованность, а вместе с тем является наиболее сложным. Исходя из особенностей структуры речи, долгое время основные надежды в решении этой задачи были связаны с непосредственным применением аппарата скрытых марковских моделей, однако на современном этапе доминирующую роль играют именно нейросетевые алгоритмы.

Для решения задачи распознавания обучаются две вероятностные модели: акустическая и языковая. Первая необходима для того, чтобы для каждого небольшого фрагмента сигнала на основе вычисленных из него признаков определить, какой фонеме он соответствует. Вторая на основе последовательности фонем (с соответствующими им вероятностями) формирует последовательности слов. На этом этапе, благодаря накопленной при обучении статистике, исправляется значительная часть ошибок, возникших при распознавании отдельных фонем.

В конце 1980-х коллектив авторов, включающий в себя Александра Вайбеля и Джеффри Хинтона, предложил архитектуру нейронной сети с временной задержкой (Time Delay Neural Network, TDNN) и продемонстрировал ее применение в задаче распознавания фонем. Эта структура похожа на полносвязную сеть, но нейрон осуществляет взвешенное суммирование не только отсчетов входного сигнала, но и их задержанных копий (в соответствии с рисунком 5). Учитывая математическое описание работы такой сети, ее можно отнести к одномерным сверточным сетям.

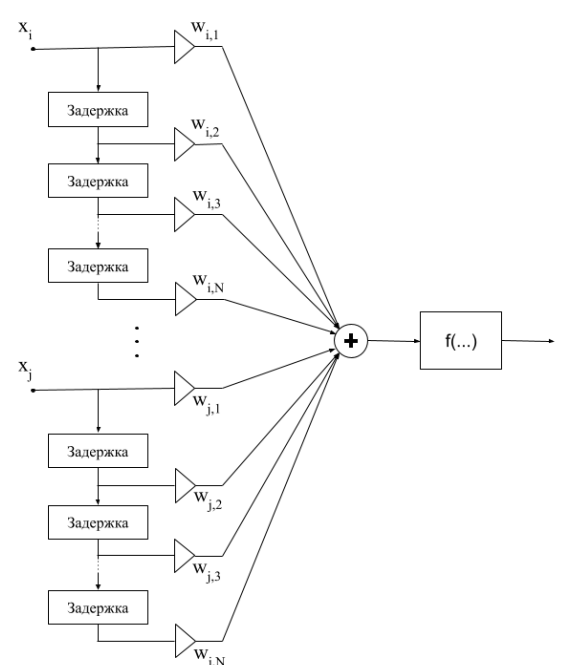


Рисунок 5 – Схема структурного блока нейронной сети с временной задержкой

Чтобы оценить современный уровень технологий в задаче распознавания речи, можно обратиться к решениям ИТ-гигантов: Google Cloud Speech-to-Text и Yandex SpeechKit. Оба варианта демонстрируют высокий уровень точности распознавания в благоприятных условиях, однако при работе с живыми записями, сделанными в условии воздействия значительных шумов, помех и реверберации, точность распознавания значительно снижается. Именно неудовлетворительная работа в таких сложных условиях является как основным фактором, сужающим потенциальные области применения алгоритмов распознавания речи, так и основным научным вызовом в данной области исследований.

К группе распознавания дикторов относятся прежде всего задачи идентификации и верификации говорящего (диктора), а также задача диаризации. Задача идентификации состоит в определении личности по голосу из множества зарегистрированных в системе дикторов. Регистрация подразумевает запись некоторого, обычно короткого, фрагмента голоса конкретного человека. Задача верификации направлена на подтверждение личности с помощью сравнения некоторой эталонной записи в базе с текущим произнесением диктора. То есть система верификации отвечает на вопрос, является ли диктор тем, за кого себя выдает. Как задача идентификации, так и задача верификации могут производиться текстозависимо или текстонезависимо. В первом случае диктору требуется произнести определенное слово или фразу, во втором – что-то произвольное. Текстозависимые системы идентификации более устойчивы к спуфингу – попыткам злоумышленника воспользоваться записью чужого голоса для обмана системы. Другим распространённым подходом противодействия спуфингу является использование многофакторной (мультимодальной) биометрии, когда идентификация производится не только по голосу, но и по другим признакам личности, например по изображению лица или отпечаткам пальцев. Задача диаризации состоит в разделении звукового потока на фрагменты, принадлежащие разным дикторам.

Работу алгоритмов диаризации часто иллюстрируют осциллограммой звуковой записи, участки которой раскрашены разными цветами, при этом каждый цвет соответствует определенному диктору. Долгое время основным инструментом решения задач данной группы являлось применение моделей гауссовой смеси (Gaussian Mixture Models или GMM) и других методов, являющихся развитием этого подхода. Попытки применить нейронные сети начали предприниматься достаточно давно, однако существенный прогресс связан с появлением крупных баз сигналов, необходимых для обучения и тестирования глубоких нейронных сетей. Наиболее известным в этом направлении является проект VoxCeleb, в рамках которого созданы базы VoxCeleb1 и VoxCeleb2.

Если же говорить об архитектурах сетей и о специфических подходах, то в первую очередь стоит выделить системы распознавания дикторов на основе x-векторов. Этот подход базируется на применении нейронных сетей с временной задержкой, которые изначально были созданы для распознавания фонем и основаны на одномерной свертке. Альтернативой является применение сверточных сетей, использующих двумерную свертку. Такой тип сетей чаще используется для распознавания изображений, однако двумерность спектрограмм, в том числе и тех, что составлены из МЧКК-векторов, позволяет без существенных изменений использовать известные архитектуры из области машинного зрения для работы с речевыми сигналами. Другим интересным подходом в этой области является замена входного слоя сверточной сети на специально разработанный слой SincNet, представляющий собой банк обучаемых фильтров, осуществляющих обработку сигналов, представленных во временной области.

Попытки синтезировать отдельные звуки, слоги и даже слова осуществлялись еще во времена доминирования аналоговых технологий. С появлением цифровых технологий ряд подходов были адаптированы и продолжали применяться. Существенная же роль нейросетей в решении этой группы задач начала наблюдаться менее десяти лет назад. В 2016 году компанией Google была представлена нейросетевая архитектура WaveNet, которая позволила существенно улучшить качество синтезируемой речи по сравнению с подходами, используемыми ранее. Позже исследователи из разных стран предложили альтернативные подходы и архитектуры нейронных сетей, особенности которых позволяют говорить об отсутствии монопольного положения WaveNet в задачах, связанных с синтезом речи. В качестве примера можно назвать генеративные нейросети WaveGlow, MelGAN, HiFi-GAN. Принято выделять конкатенативный (concatenative) и параметрический речевой синтез. Первый строится на сопоставлении коротких фрагментов аудиозаписей, соответствующих отдельным звукам или слогам. Такой подход требует большого количества записанных фрагментов и является крайне негибким. Параметрический подход, как правило, подразумевает синтез речевых сигналов путем последовательного использования акустической модели, которая отвечает также за длительности и интонации, и вокодера, генерирующего звуковые сигналы. Акустическая модель может строиться на основе рекуррентной нейронной сети, однако простые решения на их основе, неспособные учитывать мультимасштабный контекст, демонстрируют низкую реалистичность синтезируемой речи. В последние годы наблюдается существенный интерес к применению нейронных сетей в рамках параметрического подхода. При этом нейронные сети используются как для построения акустической модели, так и в качестве вокодера. Например, архитектура Tacotron2, использует вокодер на основе сети WaveNet. Подобные решения в сочетании с рядом вспомогательных алгоритмов позволяют уже сейчас синтезировать речь, обладающую высоким уровнем реалистичности [5].

## **ЗАКЛЮЧЕНИЕ**

Анализ использования нейросетей для обработки сигналов и изображений показал, что нейросети предоставляют широкий спектр возможностей для решения сложных задач, таких как шумоподавление, восстановление и улучшение изображений, сегментация, классификация и синтез сигналов. Их гибкость и адаптивность позволяют эффективно справляться с проблемами, плохо поддающимися алгоритмизации, а также превосходить традиционные методы в точности и качестве обработки.

Наибольшее влияние нейронные сети оказывают на области, требующие высокой вычислительной мощности и способности работать с большими объемами данных. В то же время остаются вызовы, связанные с интерпретируемостью моделей, их вычислительной сложностью и потребностью в значительных объемах обучающих данных.

Результаты исследований подтверждают, что развитие технологий машинного обучения и их интеграция в области обработки сигналов и изображений продолжит формировать будущее автоматизации, анализа данных и искусственного интеллекта.

## **СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1. R. Archana & P. S. Eliahim Jeevaraj «Deep learning models for digital image processing: a review» [Электронный ресурс] // SpringerNatureLink [Электронный ресурс] : [сайт] – URL: <https://link.springer.com/article/10.1007/s10462-023-10631-z#Sec2> (дата обращения 16.12.2024) – Загл. с экрана. – Яз. англ.
2. Luis Quintanilla et al. «Оценка модели ML.NET с помощью метрик» [Электронный ресурс] // Microsoft Learn [Электронный ресурс] : [сайт] – URL: [https://learn.microsoft.com/ru-ru/dotnet/machine-learning/resources /metrics](https://learn.microsoft.com/ru-ru/dotnet/machine-learning/resources%20/metrics) (дата обращения 16.12.2024) – Загл. с экрана. – Яз. рус.
3. egor\_labintcev «Метрики в задачах машинного обучения» [Электронный ресурс] // Хабр [Электронный ресурс] : [сайт] – URL: https://habr.com/ru/companies/ods/articles/328372/ (дата обращения 17.12.2024) – Загл. с экрана. – Яз. рус.
4. Музафаров Р.Р., Соловьев Н.А. «Кепстральный анализ звука для идентификации пользователя по голосу» [Электронный ресурс] : [сайт] – URL: <https://s.science-engineering.ru/pdf/2021/3/1355.pdf> (дата обращения 16.12.2024) – Загл. с экрана. – Яз. рус.
5. А. И. Топников «Применение нейронных сетей в задачах обработки речевых сигналов» [Электронный ресурс] // Применение нейронных сетей в задачах обработки речевых сигналов: учебно-методическое пособие / А. И. Топников; Яросл. гос. ун-т им. П. Г. Демидова. – Ярославль : ЯрГУ, 2022. – 36 с. [Электронный ресурс] – URL: <http://www.lib.uniyar.ac.ru/edocs/iuni/20220705.pdf> (дата обращения 17.12.2024)
6. Статья «PESQ Perceptual Evaluation of Speech» [электронный ресурс] // Telecom Trainer [Электронный ресурс] : [сайт] – URL: <https://www.telecomtrainer.com/pesq-perceptual-evaluation-of-speech/> (дата обращения 17.12.2024) – Загл. с экрана. – Яз. англ.
7. П. Е. Овчинников «Применение искусственных нейронных сетей для обработки сигналов» [Электронный ресурс] // ПРИМЕНЕНИЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ОБРАБОТКИ СИГНАЛОВ. Автор: Овчинников П.Е.. Учебно-методическое пособие. – Нижний Новгород: Нижегородский госуниверситет, 2012. – 32 с. [Электронный ресурс] – URL: http://www.unn.ru/books/met\_files/Ovch\_ANN.pdf (дата обращения 17.12.2024)