

ИССЛЕДОВАНИЕ ПРИМЕНИМОСТИ НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧАХ АНАЛИЗА ГИДРОАКУСТИЧЕСКИХ СИГНАЛОВ ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ОБНАРУЖЕНИЯ ПОДВОДНЫХ ЦЕЛЕЙ

Григоренко А. Г.¹, Ситдигов Д. С.²

DOI:10.21681/3034-4050-2025-6-20-27

Ключевые слова: сверточные сети, рекуррентные сети, трансформеры, автоэнкодеры, спектрограммы, классификация судов.

Аннотация

Целью работы является анализ применимости методов глубокого обучения к автоматизированной обработке и классификации гидроакустических сигналов для повышения эффективности обнаружения подводных целей, а также экспериментальная проверка одного из подходов на открытом наборе данных DeepShip с использованием архитектуры ResNet-50.

Метод исследования: применение систематического обзора публикаций 2020–2025 гг., охватывающий сверточные, рекуррентные, трансформерные модели и денойзинг-автоэнкодеры и проведение эксперимента, который включает преобразование аудиозаписей в трёхканальные спектрограммы (две низкочастотные полосы и одна высокочастотная), лог-нормализацию и масштабирование в $[0,1]$, случайное разбиение выборки (80/20) с сохранением пропорций классов, тонкую настройку предобученной ResNet-50 при функции потерь бинарной кросс-энтропии, оптимизаторе Adam, ранней остановке и выборе лучшей модели по точности.

Результаты исследования: обобщены современные подходы к анализу гидроакустических данных: сверточные сети над спектрограммами надёжно извлекают локальные признаки и обеспечивают высокую точность классификации; рекуррентные (LSTM/GRU) улучшают учёт временной динамики и трекинг целей; трансформеры с механизмом самовнимания повышают качество при длинных зависимостях и эффективны для детекции на сонарных изображениях; денойзинг-автоэнкодеры снижают влияние нестационарных помех и повышают отношение сигнал/шум. В проведённом эксперименте ResNet-50 на трёхканальных LF/HF-спектрограммах достигла 95 % точности на валидации и 93 % тесте, что подтверждает практическую применимость подхода. Отмечены ограничения реальной эксплуатации: вариативность шумовой среды, дефицит размеченных данных и риски недостаточной интерпретируемости решений; предложены направления совершенствования – расширенные аугментации, перенос обучения.

Научная новизна заключается в экспериментально подтверждённой схеме трёхканального спектрального представления для ResNet-50, задающей практические требования к точности и дальнейшей интеграции в системы подводного наблюдения.

Введение

Гидроакустический мониторинг – ключевой элемент обеспечения морской безопасности и подводного наблюдения: в отличие от радиолокационных и оптических средств, акустические системы работают на десятки километров под водой, что позволяет выявлять суда и подводные аппараты и контролировать состояние морской среды. При этом подводная обстановка чрезвычайно сложна для анализа: множественные источники шума (биологического и гидродинамического происхождения),

многолучевые отражения, переменные условия распространения звука и слабая сигнальная составляющая целей делают ручную обработку затруднительной, а традиционные алгоритмы – пороговые детекторы, шаблонные методы и линейные фильтры – оказываются недостаточно адаптивными и теряют точность при смене обстановки. Методы глубокого обучения (CNN, RNN, трансформеры, автоэнкодеры) способны автоматически извлекать нелинейные закономерности из спектрограмм и подстраиваться к динамике

¹ Григоренко Александр Сергеевич, младший научный сотрудник, Военная академия связи, Санкт-Петербург, Россия. E-mail: grigorenko.201@mail.ru

² Ситдигов Дмитрий Сергеевич, младший научный сотрудник, Военная академия связи, Санкт-Петербург, Россия. E-mail: dima.sitdikov.99@mail.ru

среды; недавние обзоры отмечают высокую эффективность CNN при классификации гидроакустических сигналов, а гибридные модели на базе CNN и трансформеров демонстрируют устойчивость к низкому отношению сигнал/шум [1]. Вместе с тем большинство работ ориентированы на узкие задачи и небольшие наборы данных, что подчеркивает необходимость систематического сравнения подходов и практических рекомендаций.

Постановка задачи

Необходимо выбрать оптимальную архитектуру модели нейронной сети, провести адаптацию модели (доработать входной слой и разработать классификатор), провести обучение модели, выполнить анализ ограничений и направлений развития.

Цель работы – проанализировать эффективность и ограничения нейросетевых методов в задаче анализа гидроакустических сигналов, разработать и экспериментально обосновать трёхканальную нейросетевую модель классификации гидроакустических сигналов по спектрограммам и наметить направления дальнейшего развития для повышения надёжности систем подводного наблюдения.

Решение задачи.

Анализ архитектур нейронных сетей

Для успешного применения нейронных сетей в анализе подводных сигналов требуется тщательная предобработка данных. Сырые гидрофонные сигналы обладают широкой полосой частот и низким отношением сигнал/шум, поэтому их преобразование в информативное представление является важнейшим этапом. Обычно применяются методы кратковременного преобразования Фурье (STFT) для получения спектрограмм, в том числе специальных: LOFAR-спектрограммы (Low Frequency Analysis and Recording) и спектрограммы Meier. LOFAR-спектрограмма представляет собой двухмерное временно-частотное представление сигнала и широко используется для идентификации подводных целей [1]. Спектрограмма Meier имитирует частотную чувствительность человеческого слуха на низких частотах и сохраняет основные особенности акустического сигнала [2]. Кроме того, для признакового описания применяются мел-спектрограммы и кепстральные коэффициенты MFCC, а также методы временно-частотного анализа вроде преобразования

Гильберта–Хуан [2]. В процессе подготовки данных производится фильтрация фоновых шумов (например, биологических и гидродинамических), нормализация по амплитуде и логарифмическое масштабирование спектральных коэффициентов. Совмещение нескольких спектральных представлений и их укладка в многоканальный формат повышают распознавание подводных объектов [3], показательно, что авторы данного исследования [4] подавали в CNN два канала (R, G) – реальные и мнимые части ковариационной матрицы, оформленные как RGB-изображение, и получили точную оценку направления прихода звука. Для увеличения объёма данных и устойчивости моделей применяют аугментации [5]: случайные сдвиги по времени; растяжение/сжатие (time-stretching); изменение высоты тона (pitch-shifting); добавление искусственного белого шума; случайное обнуление частей спектрограммы (frequency masking, time masking); смешивание аудиофрагментов из разных классов (Mixup). В совокупности фильтрация, продуманный выбор (и комбинация) спектральных признаков, нормализация и аугментации образуют критически важный конвейер предобработки для гидроакустического анализа.

Сверточные нейронные сети (CNN)

Сверточные сети – класс моделей, использующих операции свёртки для извлечения локальных признаков из двумерных представлений; обучаемые фильтры скользят по входу (спектрограмме), формируя карты признаков, над которыми применяются нелинейности, нормализация и подвыборка, что обеспечивает иерархическое выделение как низкоуровневых деталей, так и более глобальных структур при частичной трансляционной инвариантности. Изначально разработанные для изображений, CNN эффективно адаптированы к спектрограммам гидроакустических сигналов: глубокие архитектуры демонстрируют высокую обобщающую способность и адаптивность при классификации «акустических отпечатков» судов по LOFAR- и мел-спектрограммам, причём сочетание мел-спектров с глубокими CNN даёт выигрыш относительно линейных спектрограмм [6]; в задачах сегментации гидролокационных изображений (кадров переднего/бокового обзора) свёрточные слои автоматически выделяют объекты (включая малоразмерные мины) на фоне шумного

дна; для оценки направления прихода звука (DOA) показан подход, где матрица ковариации кодируется как RGB-изображение (реальная и мнимая части по каналам), и обученная CNN достигает высокой точности определения направления [4]. Для повышения устойчивости к помехам и уменьшения переобучения применяются архитектуры ResNet, техники BatchNorm и Dropout, а также использование предобученных моделей на больших массивах, например ImageNet, AudioSet, с последующим тонким дообучением на профильных данных. Таким образом, CNN способны надёжно извлекать информативные пространственно-частотные паттерны и обеспечивать высокую точность распознавания объектов.

Рекуррентные нейронные сети (RNN)

Подводные акустические сигналы являются по природе временными рядами, поэтому модели, учитывающие последовательность данных, нашли широкое применение. Классические рекуррентные сети и их современные варианты – долгая краткосрочная память (LSTM) и сети на базе блоков GRU – способны хранить информацию о предыдущих отсчётах и тем самым учитывать динамику изменения акустического спектра. LSTM/GRU решают проблему затухания градиента, что позволяет обучать их на длительных последовательностях без потери памяти о ранних сигналах [7]. В гидроакустических задачах рекуррентные слои часто комбинируются со сверточными (архитектуры типа CRNN) для одновременного учёта пространственных (спектральных) и временных особенностей. Подобные гибридные сети успешно применены для трекинга подводных целей. В частности, алгоритм пассивного слежения был реализован с помощью связи: на первом этапе кубатурный фильтр Калмана (вариант нелинейного фильтра) оценивал движение цели, а на втором этапе рекуррентная сеть GRU обучалась по выходам фильтра предсказывать отклонения и корректировать траекторию. Симуляционные испытания показали, что такой подход (GRU и СКФ) обеспечивает лучшую точность и стабильность оценки состояния автономного подводного аппарата (UUV) по сравнению с традиционными методами фильтрации [7]. Кроме того, RNN широко используются для прогнозирования параметров акустических сигналов во времени (например, доплеровских сдвигов

или изменений тональной структуры шума корабля), что полезно при длительном наблюдении объектов. Однако у RNN есть ограничения в использовании, они требовательны к памяти, хуже параллелизуются и испытывают трудности на очень длинных последовательностях, поэтому в задачах с большими объёмами данных всё чаще используют модели с самовниманием, однако для умеренных длительностей и трекинга RNN остаются эффективным и практичным выбором.

Трансформеры и механизмы внимания

В последние годы в области анализа последовательностей (речи, текста) наибольших успехов достигли архитектуры типа Transformer, основанные на механизме самовнимания. Их ключевое преимущество – способность моделировать долгосрочные зависимости во входных данных, взвешивая вклад различных фрагментов последовательности. Для гидроакустических сигналов адаптация трансформеров стала следующим шагом развития. В 2022 году была предложена модель Spectrogram Transformer Model (STM) для распознавания подводных целей [8]. В этой модели исходный шумовой сигнал преобразуется в последовательность фрагментов спектрограммы, которая затем обрабатывается энкодером трансформера. Данный подход впервые внедрил самовнимание в область подводной акустики, продемонстрировав превосходство над классическими CNN: так, точность классификации STM-модели достигла 97 % (против 84 % у базового CNN) на тестовых данных [8]. Трансформеры также интегрируются в гибридные архитектуры для обработки сонарных изображений. Например, для повышения качества обнаружения объектов на гидролокационных снимках бокового обзора разработаны улучшения к алгоритму YOLO: добавлены блоки Swin-Transformer, реализующие механизм динамического внимания и глобального обзора сцены [9]. Это позволяет сети фокусироваться на релевантных областях сонограммы, учитывая при этом контекст всего изображения, что особенно важно при поиске небольших или малоразличимых целей на сложном фоне. Таким образом, интеграция трансформер-модулей повышает точность детекции на несколько процентов, сохраняя при этом допустимую скорость работы модели в реальном времени.

Автоэнкодеры и удаление шумов

Одной из наиболее острых проблем подводной акустики является нестабильный шумовой фон. Для улучшения качества сигналов применяются нейросетевые модели, обучающиеся восстанавливать исходный сигнал из зашумлённого. Денойзинг-автоэнкодеры (DAE) обучаются на парах «шум и чистый сигнал», позволяя без ручного выбора признаков выделить информативную составляющую. В гидроакустике разрабатываются, как правило, сверточные автоэнкодеры, учитывающие пространственные структуры спектрограмм. Исследования показывают, что такие сети способны увеличивать отношение сигнал/шум на выходе, существенно улучшая различимость признаков цели. В одной из работ предложен полносверточный автоэнкодер для шумоподавления записей из базы ShipsEar; в результате уровень SNR повысился примерно на 10 дБ, а метрика качества (SI-SNR) – на 9,5 дБ [10], что прямо улучшает различимость тональных «отпечатков» винтов и двигателей на LOFAR/мел-спектрограммах. Преимущество автоэнкодеров в том, что они учатся непрерывно: при обновлении поступающих данных модель можно дообучить на новых типах шумов или сигналов. Кроме того, их можно объединять с другими сетями – например, использовать кодировщик автоэнкодера как слой признаков перед классификатором (так называемые гибридные модели). Практическую эффективность комбинированного подхода демонстрирует фреймворк AquaSignal: модуль U-Net для удаления шума и ResNet-18 для классификации и автоэнкодер для обнаружения аномалий, что на объединенном наборе DeepShip и ONC дало 71 % точности классификации и 91 % точности детекции неизвестных сигналов [11]. Следует отметить, что при крайне низком исходном SNR ни одна модель не гарантирует полного восстановления – часть информации теряется безвозвратно. Тем не менее, комбинация передовых методов шумоподавления (включая нейросетевые) с последующей классификацией представляет наиболее перспективное направление для работы в реальных условиях морского среды.

Эксперимент: классификация судов с помощью ResNet-50

Для оценки эффективности методов глубокого обучения в задаче распознавания морских

объектов проведён эксперимент с использованием сверточной нейросети ResNet-50. В качестве данных использован открытый датасет DeepShip, содержащий аудиозаписи (формат WAV) различных типов судов, полученные с гидрофонов с частотой дискретизации 32 кГц [12]. Каждая запись была преобразована в набор спектрограмм, формирующих трёхканальное изображение (RGB): два канала (R и G) кодировали спектрограммы низкочастотного диапазона, разделённого на две полосы – R: 0–449,22 Гц и G: 449,22–898,44 Гц, а третий канал (B) – спектрограмму высокочастотной составляющей сигнала 900–14 718,75 Гц.

Перед подачей в сеть спектрограммы были логарифмически нормированы и масштабированы в диапазон [0, 1] для устранения влияния абсолютного уровня громкости, дополнительно применяли легкие аугментации (временные сдвиги, добавление шума, маскирование по времени/частоте), повышающие устойчивость к вариациям среды.

Разметка данных включала несколько классов целей (различные типы судов) плюс класс фонового шума. Для обучения нейросети выборка была случайным образом разделена в соотношении 80/20 (обучающая и валидационная), при этом валидационная часть сохраняла пропорции по каждому классу, а для финальной проверки использовали независимый тестовый поднабор, не участвовавший в обучении.

Описание архитектуры и процедуры обучения

В качестве базовой модели взята архитектура ResNet-50, предобученная на датасете ImageNet (классические объекты с изображений) – такой подход позволяет воспользоваться уже усвоенными низкоуровневыми признаками. Входной слой ResNet был модифицирован под размерность 3×H×W (трёхканальное изображение спектрограмм), дописан оригинальный классификатор, выходной слой которого настроен на число классов, соответствующее количеству типов судов в наборе DeepShip. В процессе обучения использовалась функция потерь бинарной кросс-энтропии (БКЭ) с учетом многоклассовой классификации, оптимизатор Adam, регуляризация (L2, Dropout) и критерий ранней остановки по метрике ассурасу на валидационных данных. Обучение проводилось в течение 200 эпох.

На рисунке 1 представлен график функции потерь на обучающей и валидационной выборках по эпохам, отражающий сходимость процесса обучения. В начале обучения наблюдалось быстрое снижение ошибки, стабилизировавшееся после 50 эпох. Лучшее значение точности на валидации составило 95,1 %, что говорит о высокой способности модели правильно классифицировать типы судов по их спектрограммам. График точности модели на валидационной выборке (ВВ) представлен на рисунке 2. На независимом тестовом наборе (сигналы судов, не встречавшиеся при обучении) модель достигла аналогичной точности в 93 %. Полученные результаты подтверждают эффективность применения современных сверточных сетей для классификации «акустических отпечатков» судов по спектральным признакам.

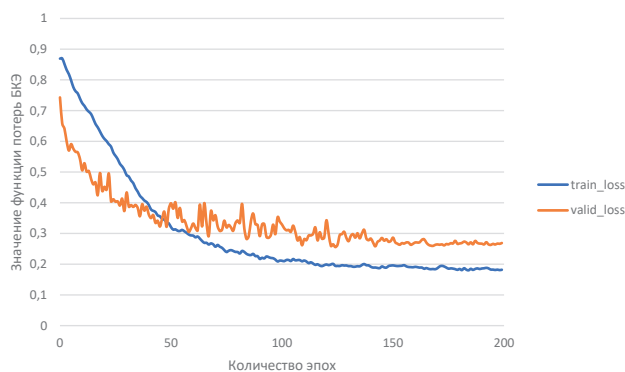


Рис. 1. График функции потерь на обучающей и валидационной выборках

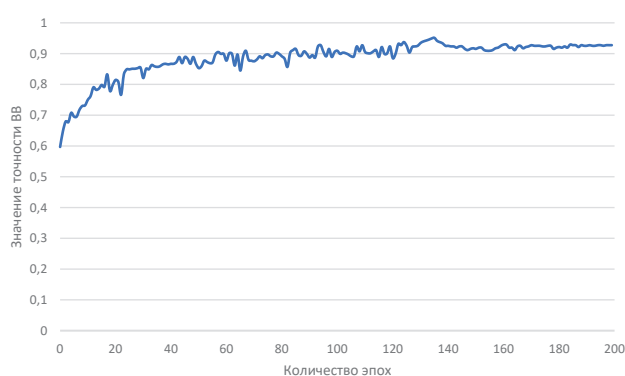


Рис. 2. График точности модели нейронной сети на валидационной выборке

Однако, при практическом использовании подобных моделей в реальных условиях необходимо учитывать ряд ограничений.

Во-первых, акустические помехи и вариативность среды: хотя нейросеть успешно обучилась

на имеющихся данных, в открытом океане возможны существенные отличия из-за шумового фона (шторм, биологический шум, реверберация от береговой черты). Если реальное отношение сигнал/шум окажется значительно ниже, чем в обучающем наборе, качество классификации может ухудшиться, поскольку модель будет получать на вход более искаженные спектрограммы.

Во-вторых, ограниченность обучающих данных: набор DeepShip представляет лишь определенный сценарий (записи ряда судов в конкретных условиях). На практике часто сталкиваются с необходимостью распознавания новых типов объектов или работы в новых локациях, для которых данных нет. Модели глубокого обучения чувствительны к распределению данных, поэтому при переносе их в новые условия требуется либо сбор дополнительных размеченных выборок, либо применение специальных техник (аугментация, синтез данных, transfer learning). В нашем эксперименте использовался перенос обучения и генерация дополнительных спектрограмм с шумами, что отчасти решает проблему, однако полностью избежать разрыва распределений (dataset bias) не удастся.

В-третьих, прозрачность алгоритмов: глубокие нейросети действуют как «черный ящик», и в критических приложениях (например, военные системы принятия решений) необходимо понимать причины, по которым модель отнесла сигнал к тому или иному классу. Для этого используют подходы XAI: тепловые карты важности на спектрограммах (например, Grad-CAM, Integrated Gradients), анализ attention-весов в трансформерах, а также post-hoc методы уровня признаков (LIME/SHAP), которые помогают верифицировать причины решения и повысить доверие к системе, при этом данная область активно развивается и остается предметом текущих исследований [13, 14].

Результаты и выводы

В ходе исследования проведен всесторонний анализ современных методов глубокого обучения применимых к обработке гидроакустических сигналов. Различные классы нейросетей продемонстрировали следующие ключевые особенности:

1. Сверточные сети эффективно справляются с выделением признаков на спектрограммах, достигая высокой точности

- классификации подводных объектов. Они особенно полезны для распознавания по акустическому «отпечатку» цели, а также для обработки акустических изображений (сонограмм), где позволяют обнаруживать объекты на сложном фоне.
2. Рекуррентные сети и их производные (LSTM, GRU) незаменимы при анализе временной структуры сигналов. Их использование существенно повышает точность трекинга и прогнозирования. RNN также улучшают классификацию долговременных сигналов, таких как протяжённые шумы судов.
 3. Трансформеры с механизмом самовнимания открыли новые возможности учёта глобальных связей в данных. Данные модели обещают быть особенно полезными в сценариях, где требуется учесть долгие зависимости (например, анализ акустического ландшафта за длительный период) или совместно обрабатывать множественные сенсорные данные.
 4. Автоэнкодеры зарекомендовали себя как эффективный инструмент для автоматического подавления шумов. Их применение позволяет повысить надёжность последующей классификации, поскольку на вход классификатору подаётся очищенный сигнал. Удалось достичь улучшения соотношения сигнал/шум на 8–10 дБ на реальных данных, что подтверждается исследованиями.

Практический эксперимент с моделью ResNet-50 подтвердил применимость глубокого обучения для классификации объектов по их акустическим шумам: даже относительно компактная сеть при правильной подготовке входных данных способна различать несколько типов кораблей с точностью 95 % на тестовых данных. Это результат на уровне современных мировых достижений и значительно выше, чем у традиционных алгоритмов. Тем не менее, эксперимент выявил и нерешённые проблемы. Главные из них – адаптация к нестационарным помехам и дефицит размеченных данных для обучения. Решение первой проблемы видится в комбинировании нейросетевых методов с физическими моделями распространения звука (для учёта среды) и в развитии более сложных архитектур

фильтрации (например, объединение нейросетей с фильтрами Калмана, вейвлет-преобразованиями). Вторая проблема – нехватка данных – может быть сглажена с помощью методов аугментации и генерации данных. В частности, показано, что применение GAN-генераторов для синтеза искусственных спектрограмм улучшает устойчивость моделей [15]. Активно используется перенос знаний: сети обучают на больших смежных датасетах перед тонкой настройкой под гидроакустические данные, что повышает обобщающую способность алгоритмов. Дефицит данных может помочь решить использование мультисенсорных схем – совместная обработка гидроакустических данных (спектрограммы, LOFAR), изображений бокового обзора (SSS), навигации/AIS. Такой подход помогает в некоторых задачах, например снижает неопределённость трекинга и повышает робастность классификации [16].

Ещё одним направлением исследований является повышение интерпретируемости глубоких моделей. В военных приложениях, где цена ошибки крайне высока, необходимо доверие к системе. Разработка методов XAI [14] для подводных нейросетевых систем – важная задача ближайших лет. Возможные подходы включают интеграцию механизмов внимания (подсвечивающих важные участки спектрограммы, повлиявшие на решение), оценку неопределённости (выход сети с доверительным интервалом) и создание гибридных моделей, где критические решения перепроверяются классическими алгоритмами.

Для работы в режиме реального времени на борту AUV/USV/буёв необходима аппаратная оптимизация: квантование (FP16/INT8), структурное прореживание, дистилляция [17].

Заключение

Глубокое обучение становится ключевым инструментом в анализе гидроакустических сигналов. Современные модели уже сейчас обеспечивают высокий уровень автоматизации и точности при мониторинге подводной обстановки. Их дальнейшая оптимизация – залог создания интеллектуальных гидроакустических комплексов нового поколения, способных надёжно работать в реальном времени даже в самых сложных условиях Мирового океана.

Литература

1. Lin X., Dong R., Lv Z. Deep learning-based classification of raw hydroacoustic signal: A review // Journal of Marine Science and Engineering. – 2022. – Т. 11. – №. 1. – С. 3.
2. Lin X. et al. Efficient ship noise classification with positive incentive noise and fused features using a simple convolutional network // Scientific Reports. – 2023. – Т. 13. – №. 1. – С. 17905.
3. Guo T. et al. Underwater target detection and localization with feature map and CNN-based classification // 2022 4th International Conference on Advances in Computer Technology, Information Science and Communications (CTISC). – IEEE, 2022. – С. 1–8.
4. Liu Y., Chen H., Wang B. DOA estimation based on CNN for underwater acoustic array // Applied Acoustics. – 2021. – Т. 172. – С. 107594.
5. Abayomi-Alli O. O. et al. Data augmentation and deep learning methods in sound classification: A systematic review // Electronics. – 2022. – Т. 11. – №. 22. – С. 3795.
6. Mishachandar B., Vairamuthu S. Diverse ocean noise classification using deep learning // Applied Acoustics. – 2021. – Т. 181. – С. 108141.
7. Wang Y. et al. Passive sonar target tracking based on deep learning // Journal of Marine Science and Engineering. – 2022. – Т. 10. – №. 2. – С. 181.
8. Li P. et al. STM: Spectrogram transformer model for underwater acoustic target recognition // Journal of Marine Science and Engineering. – 2022. – Т. 10. – №. 10. – С. 1428.
9. Wen X. et al. Underwater side-scan sonar target detection: YOLOv7 model combined with attention mechanism and scaling factor // Remote Sensing. – 2024. – Т. 16. – №. 13. – С. 2492.
10. Song Y. et al. Underwater acoustic signal noise reduction based on a fully convolutional encoder-decoder neural network // Journal of Ocean University of China. – 2023. – Т. 22. – №. 6. – С. 1487–1496.
11. Panteli E., Santos P. E., Humphrey N. AquaSignal: An Integrated Framework for Robust Underwater Acoustic Analysis // arXiv preprint arXiv:2505.14285. – 2025.
12. Hummel H. I., van der Mei R., Bhulai S. A survey on machine learning in ship radiated noise // Ocean Engineering. – 2024. – Т. 298. – С. 117252.
13. Akman A., Schuller B. W. Audio explainable artificial intelligence: A review // Intelligent Computing. – 2024. – Т. 2. – С. 0074.
14. Cheng Z. et al. A Comprehensive Review of Explainable Artificial Intelligence (XAI) in Computer Vision // Sensors. – 2025. – Т. 25. – №. 13. – С. 4166.
15. Tian S. et al. Deep convolution stack for waveform in underwater acoustic target recognition // Scientific reports. – 2021. – Т. 11. – №. 1. – С. 9614.
16. Zhang Z. et al. ESL-YOLO: Edge-Aware Side-Scan Sonar Object Detection with Adaptive Quality Assessment // Journal of Marine Science and Engineering. – 2025. – Т. 13. – №. 8. – С. 1477.
17. Tang Y. et al. AUV-based side-scan sonar real-time method for underwater-target detection // Journal of Marine Science and Engineering. – 2023. – Т. 11. – №. 4. – С. 690.

STUDY OF THE APPLICABILITY OF NEURAL NETWORKS IN THE ANALYSIS OF HYDROACOUSTIC SIGNALS TO IMPROVE THE EFFICIENCY OF DETECTING UNDERWATER TARGETS

Grigorenko A. G.³, Sitdikov D. S.⁴

Keywords: convolutional networks, recurrent networks, transformers, autoencoders, spectrograms, ship classification.

Abstract

The aim of the work is to analyze the applicability of deep learning methods to automated processing and classification of hydroacoustic signals to improve the efficiency of detecting underwater targets, as well as to experimentally test one of the approaches on an open DeepShip dataset using the ResNet-50 architecture.

³ Alexander S. Grigorenko, Junior Researcher, Military Academy of Communications, St. Petersburg, Russia. E-mail: grigorenko.201@mail.ru

⁴ Dmitry G. Sitdikov, Junior Researcher, Military Academy of Communications, St. Petersburg, Russia. E-mail: dima.sitdikov.99@mail.ru

Research method: application of a systematic review of publications in 2020–2025, covering convolution, recurrent, transformer models and denoised autoencoders, and conducting an experiment that includes the transformation of audio recordings into three-channel spectrograms (two low-frequency bands and one high-frequency), log-normalization and scaling to [0,1], random sample splitting (80/20) while maintaining class proportions, fine-tuning of the pre-trained ResNet-50 at the binary cross-entropy loss, Adam optimizer, early stop and selection of the best model for accuracy.

Results of the study: modern approaches to the analysis of hydroacoustic data are summarized: convolutional networks over spectrograms reliably extract local features and provide high classification accuracy; recurrent networks (LSTM/GRU) improve time dynamics accounting and target tracking; transformers with a self-attention mechanism improve quality with long dependencies and are effective for detection on sonar images; denoising autoencoders reduce the influence of non-stationary interference and increase the signal-to-noise ratio. In the ResNet-50 experiment on three-channel LF/HF spectrograms, it achieved 95 % accuracy in validation and 93 % in the test, which confirms the practical applicability of the approach. The limitations of real operation are noted: variability of the noise environment, lack of labeled data and risks of insufficient interpretability of decisions; Directions for improvement are proposed, such as extended augmentations, transfer of training.

The scientific novelty lies in the experimentally confirmed scheme of three-channel spectral representation for ResNet-50, which sets practical requirements for accuracy and further integration into underwater surveillance systems.

References

1. Lin X., Dong R., Lv Z. Deep learning-based classification of raw hydroacoustic signal: A review // Journal of Marine Science and Engineering. – 2022. – Т. 11. – №. 1. – С. 3.
2. Lin X. et al. Efficient ship noise classification with positive incentive noise and fused features using a simple convolutional network // Scientific Reports. – 2023. – Т. 13. – №. 1. – С. 17905.
3. Guo T. et al. Underwater target detection and localization with feature map and CNN-based classification // 2022 4th International Conference on Advances in Computer Technology, Information Science and Communications (CTISC). – IEEE, 2022. – С. 1–8.
4. Liu Y., Chen H., Wang B. DOA estimation based on CNN for underwater acoustic array // Applied Acoustics. – 2021. – Т. 172. – С. 107594.
5. Abayomi-Alli O. O. et al. Data augmentation and deep learning methods in sound classification: A systematic review // Electronics. – 2022. – Т. 11. – №. 22. – С. 3795.
6. Mishachandar B., Vairamuthu S. Diverse ocean noise classification using deep learning // Applied Acoustics. – 2021. – Т. 181. – С. 108141.
7. Wang Y. et al. Passive sonar target tracking based on deep learning // Journal of Marine Science and Engineering. – 2022. – Т. 10. – №. 2. – С. 181.
8. Li P. et al. STM: Spectrogram transformer model for underwater acoustic target recognition // Journal of Marine Science and Engineering. – 2022. – Т. 10. – №. 10. – С. 1428.
9. Wen X. et al. Underwater side-scan sonar target detection: YOLOv7 model combined with attention mechanism and scaling factor // Remote Sensing. – 2024. – Т. 16. – №. 13. – С. 2492.
10. Song Y. et al. Underwater acoustic signal noise reduction based on a fully convolutional encoder-decoder neural network // Journal of Ocean University of China. – 2023. – Т. 22. – №. 6. – С. 1487–1496.
11. Panteli E., Santos P. E., Humphrey N. AquaSignal: An Integrated Framework for Robust Underwater Acoustic Analysis // arXiv preprint arXiv:2505.14285. – 2025.
12. Hummel H. I., van der Mei R., Bhulai S. A survey on machine learning in ship radiated noise // Ocean Engineering. – 2024. – Т. 298. – С. 117252.
13. Akman A., Schuller B. W. Audio explainable artificial intelligence: A review // Intelligent Computing. – 2024. – Т. 2. – С. 0074.
14. Cheng Z. et al. A Comprehensive Review of Explainable Artificial Intelligence (XAI) in Computer Vision // Sensors. – 2025. – Т. 25. – №. 13. – С. 4166.
15. Tian S. et al. Deep convolution stack for waveform in underwater acoustic target recognition // Scientific reports. – 2021. – Т. 11. – №. 1. – С. 9614.
16. Zhang Z. et al. ESL-YOLO: Edge-Aware Side-Scan Sonar Object Detection with Adaptive Quality Assessment // Journal of Marine Science and Engineering. – 2025. – Т. 13. – №. 8. – С. 1477.
17. Tang Y. et al. AUV-based side-scan sonar real-time method for underwater-target detection // Journal of Marine Science and Engineering. – 2023. – Т. 11. – №. 4. – С. 690.