УДК 004.93

ДООБУЧЕНИЕ КОДИРОВЩИКА НЕЙРОННОЙ СЕТИ WHISPER С ПОМОЩЬЮ ПРОМТ-ВЕКТОРОВ И БЛОКОВ АДАПТАЦИИ ДЛЯ ЗАДАЧИ ОБНАРУЖЕНИЯ ШУМОВ СЕРДЦА

Топников А.И., *к.т.н., доцент кафедры цифровых технологий и машинного обучения Ярославского государственного университета им. П.Г. Демидова, e-mail: a.topnikov@uniyar.ac.ru*
Смирнова Т.А., *студент 4 курса Ярославского государственного университета им. П.Г. Демидова, e-mail: t.smirnova9@stud.uniyar.ac.ru*

FINE-TUNING THE WHISPER NEURAL NETWORK ENCODER USING PROMT VECTORS AND ADAPTATION BLOCKS FOR HEART MURMUR DETECTION

Topnikov A.I., Smirnova T.A.

The problem of heart murmur detection from phonocardiograms using neural networks is considered. To solve the problem the WhisPAr neural network architecture is used. It's based on the Whisper network encoder and using promts and adapters to fine-tune on a small amount of data. The publicly available part of the CirCor DigiScope signal database is used. The use of prompts and adapters makes it possible to increase the weighted accuracy compared to the basic version of the classifier based on the Whisper network encoder with a trained classification layer at the output. The results obtained can be used to further improve approaches to the detection of heart pathologies based on the analysis of phonocardiograms.

Key words: phonocardiogram, heart murmurs, neural network, deep learning, transfer learning, audio signals classification.

Ключевые слова: фонокардиограмма, шумы сердца, нейронная сеть, глубокое обучение, трансферное обучение, классификация звуковых сигналов.

Рассматривается задача обнаружения шумов сердца на фонокардиограммах с использованием нейронных сетей. Для решения задачи применяется нейронная сеть WhisPAr, основанная на кодировщике сети Whisper и использующая для его дообучения на небольшом объеме данных промт-векторы и блоки адаптации. Для обучения и тестирования используется общедоступная часть размеченной базы фонокардиограмм CirCor DigiScope. Применение указанных методов дообучения позволяет повысить взвешенную точность классификации по сравнению с базовым вариантом алгоритма на основе кодировщика сети Whisper с обучаемым слоем классификации на выходе. Полученные результаты могут использоваться для дальнейшего совершенствованию подходов к выявлению патологий сердца на основе анализа фонокардиограмм.

Введение

Анализ биомедицинских сигналов является важным приложением цифровой обработки сигналов. Современные нейросетевые архитектуры активно применяются для создания систем поддержки принятия врачебных решений. Наибольший успех наблюдается в задачах, связанных с обработкой и анализом электрокардиограмм [1], электроэнцефалограмм [2], а также биомедицинских изображений [3-5]. Развитие методов классификации звуковых сигналов также позволяет решать ряд медицинских задач [6-10].

В кардиологии важным приложением анализа звуковых сигналов является обнаружение шумов сердца, которые наиболее характерны для пациентов детского возраста. Сложность решения этой задачи для врачей хорошо иллюстрируют следующие слова С. Манджони: «Аускультация сердца – это важнейшая часть физикальной диагностики, а распознавание шумов сердца – это настоящее испытание для врача. Оно требует распознавания нескольких звуков, сжатых во временном интервале меньше 0,8 секунды, часто перекрывающих друг друга и не так уж редко находящихся на пороге слышимости» [11].

Попытки применения нейронных сетей для обнаружения шумов сердца наблюдаются ни одно десятилетие [12], однако переходу к практическому внедрению пре пятствует не только исходная сложность задачи, но и отсутствие достаточного объема данных для обучения. Это можно объяснить следующими причинами:

– невозможностью достоверного обнаружения шумов сердца для разметки фонокардиографических записей без привлечения медицинских специалистов высокой квалификации;

– относительной редкостью (низкой частотой проявления) и высокой значимостью одного из классов (наличия шумов сердца);

– сложностью получения фонокардиограмм высокого качества, не искаженных внешними шумами, не относящимися к работе сердца.

Целью работы является решение задачи обнаружения шумов сердца с использованием дообучаемого на небольшом объеме данных кодировщика нейронной сети Whisper, изначально созданной для распознавания речевых сигналов.

База фонокардиограмм CirCor DigiScope и научный конкурс 2022 года

Значительный рост интереса специалистов в области глубокого обучения к рассматриваемой тематике связан с появлением относительно крупной размеченной базы фонокардиограмм CirCor DigiScope [13, 14] и проведением на ее основе международного научного конкурса в 2022 году [15].

База фонокардиограмм CirCor DigiScope состоит из записей (рисунок 1), собранных в Бразилии в рамках двух кампаний в 2014 и 2015 годах с использованием электронного стетоскопа Littmann 3200. Длительность сигналов, представленных в базе, варьируется от нескольких секунд до нескольких десятков секунд. Частота дискретизации равна 4 кГц, разрядность представления отсчетов – 16 бит.

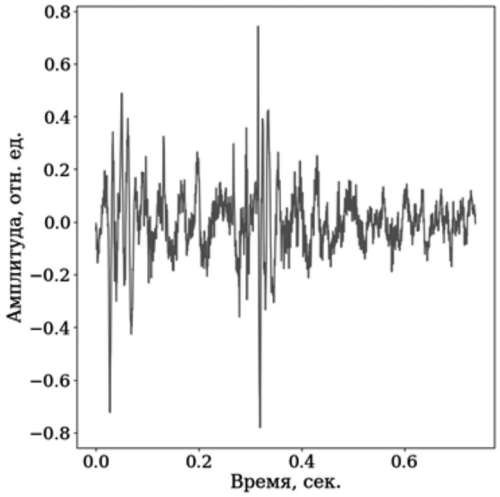


Рисунок 1. Осциллограмма фрагмента сигнала

Общий объем базы – 5282 записи, преимущественно записанные в четырех основных точках аускультации сердца у 1568 пациентов в возрасте от 0,1 до 356,1 месяцев. Средний возраст пациентов составляет 73,4 месяца. Общедоступной является только часть базы, предоставленная организаторами научного конкурса для обучения алгоритмов классификации. Анонсировалось, что остальные записи, использовавшиеся организаторами конкурса для тестирования алгоритмов участников, станут публичными после завершения конкурса, но этого не было сделано.

Конкурс 2022 года посвящен решению двух последовательных задач: обнаружению шумов и предсказанию клинического исхода. Общий файл разметки содержит информацию о наличии шумов и исходе для каждого пациента, а также другую полезную информацию, включая пол, возраст и т.д. Разметка проведена высококвалифицированными специалистами-медиками. Для пациентов с шумами сердца также указываются базовые характеристики шумов (высота, характер звука и т.д.), точки аускультации, в которых определяется наличие шумов, а также точки, в которых они наиболее интенсивны. Каждая запись также снабжается файлом с временными границами фаз сердечного цикла.

В общедоступной части базы представлены фонокардиограммы 695 пациентов с отсутствием шумов сердца (класс «Absent»), 179 – с их наличием (класс «Present») и 68 пациентов, отнесенных к классу «Unknown». Эта метка выставлялась в случае невозможности уверенной интерпретации фонокардиограмм медицинскими специалистами, преимущественно, в связи с сильной зашумленностью сигнала.

Оценка алгоритмов обнаружения шумов в конкурсе 2022 года осуществлялась с помощью взвешенной точности со следующими весами: 5 – для класса «Present», 3 – для «Unknown», 1 – для «Absent». Решения, представленные лидерами конкурса, позволили достичь на тестовой выборке значений взвешенной точности в интервале 0,76-0,78 [15]. Эти результаты преимущественно получены с использованием сверточных нейронных сетей, использующих в качестве входных признаков представление фонокардиограмм в виде мел-спектрограмм.

Сеть Whisper в задаче классификации звуковых сигналов

Нейросетевая архитектура Трансформер лежит в основе многих современных нейронных сетей, применяемых для распознавания речи, а также других задач речевой и звуковой обработки [16-19]. Значительные успехи в распознавании речи связаны с появлением нейронной сети Whisper [20], построенной по схеме кодировщик-декодировщик (encoder-decoder), в которой обе части выполнены по архитектуре Трансформер [21]. В отличие от ряда других современных решений, использующих принцип самообучения, сеть Whisper обучалась на размеченных речевых сигналах. Высокие значения ключевых метрик, продемонстрированные сетью для задачи распознавания речевых сигналов, дают возможность предположить, что использование современных достижений в области машинного обучения и большого объема размеченных данных позволило ее создателям обучить кодировщик, эффективно извлекающий и компактно представляющий информацию, содержащуюся в звуковых сигналах. Это предположение нашло подтверждение в успешном применении кодировщика сети Whisper для решения ряда подзадач классификации звуковых сигналов [22, 23].

В большинстве случаев использование кодировщика Whisper для решения задачи классификации звуковых сигналов сводится к обучению классификатора, принимающего на вход либо выходное скрытое состояние, либо все скрытые состояния блоков кодировщика. При этом сам кодировщик Whisper и предшествующие ему два сверточных слоя чаще не подвергаются дообучению. Применимость такого подхода подтверждают результаты исследования, демонстрирующего способность кодировщика сети Whisper формировать подходящие для решения задачи классификации представления не только для речевых сигналов, но и для более широкого класса звуковых сигналов [23]. Однако в случаях, когда классифицируемые сигналы сильно отличаются от речевых сигналов или требуется выявлять тонкие различия между близкими классами звуковых сигналов, дообучение кодировщика Whisper совместно с обучением классификатора может быть полезно [24].

Подходы к дообучению кодировщика сети Whisper

В области классификации звуковых сигналов известно много задач, для которых не существует больших баз сигналов, что существенно осложняет обучение нейронных сетей с большим количеством параметров, к которым относится и сеть Whisper. Существование аналогичных проблем в ряде других задач и приложений, в которых показали свою эффективность сети на основе архитектуры Трансформер, привело к разработке ряда подходов, позволяющих производить дообучение, затрагивающее относительно небольшое количество коэффициентов [25, 26]. В задаче классификации звуковых сигналов исследовано применение подходов на основе промт-векторов (промтов) и блоков адаптации (адаптеров) для дообучения сети Whisper [24]. Созданный таким образом нейросетевой классификатор звуковых сигналов получил название WhisPAr.

Методика на основе промт-векторов подразумевает добавление обучаемых векторов к признакам, поступающим на вход блока Трансформера. В сети WhisPAr на вход блока вместе с признаками, содержащими также позиционный код, подаются классификационный и промт-векторы [24]:

$$EP_1^j = \left\{c^{(j-1)}, s_1^{(j-1)}, s_2^{(j-1)}, \cdots, s_n^{(j-1)}, p_1^{(j)}, \cdots, p_h^{(j)}\right\},$$

где $EP_1^j \in \mathbb{R}^{(1+n+h)\times d}$, $c^{(j-1)}$ – классификационный вектор на входе j-го блока Трансформера, $\left\{s_1^{(j-1)}, s_2^{(j-1)}, \cdots, s_n^{(j-1)}\right\}$ – последовательность векторов признаков длиной n на входе j-го блока Трансформера, $\left\{p_1^{(j)}, \cdots, p_h^{(j)}\right\}$ – последовательность промт-векторов длиной h на входе j-го блока Трансформера, d – размерность вектора признаков, для модели Whisper-base, равная 512. После прохождения EP_1^j j-го блока Трансформера промт-векторы $\left\{p_1^{(j)}, \cdots, p_h^{(j)}\right\}$ удаляются. Из выхода заключительного блока Трансформера с номером m извлекается вектор $c^{(m)}$, который поступает на обучаемый полносвязный классификационный слой.

Методика дообучения на основе блоков адаптации подразумевает добавление в структуры сети блоков с относительно небольшим количеством обучаемых параметров. Блок Трансформера с адаптерами, используемый в сети WhisPAr [24], представлен на рисунке 2. Схема адаптера представлена на рисунке 3. Серой заливкой обозначены блоки, подвергающиеся обучению.

Работу блока адаптации можно описать следующим выражением:

$$Adapt(x) = \mathrm{Re}LU\left(xW_1 + b_1\right)W_2 + b_2,$$

где x – вектор признаков на входе адаптера, $\mathrm{Re}LU(\,)$ – функция активации $\mathrm{Re}LU$, $W_1 \in \mathbb{R}^{d\times b}$, $b_1 \in \mathbb{R}^{1\times b}$ – параметры линейного слоя понижения размерности, $W_2 \in \mathbb{R}^{b\times d}$, $b_2 \in \mathbb{R}^{1\times d}$ – параметры линейного слоя повышения размерности, b (b<d) – размерность пространства так называемого «бутылочного горлышка» (bottleneck).

Исследования показали, что совместное применение промт-векторов и блоков адаптации позволяет существенно повысить точность классификации звуковых сигналов в ряде задач, подразумевающих наличие тонких отличий между разными классами [24]. Этот результат позволяет сделать предположение о применимости этих наработок для решения задачи обнаружения шумов сердца на фонокардиограммах.

Обучение и тестирование нейронной сети

Ряд исследований, использующих для обнаружения шумов сердца на фонокардиограммах нейронные сети на основе архитектуры Трансформер, позволяют сделать вывод о преимуществах применения этой архитектуры на текущем уровне развития данной предметной области [27-29]. Результаты, демонстрируемые в задаче распознавания речи нейронной сетью Whisper, в основе кодировщика и декодировщика которой также ле жит архитектура Трансформер [20], и примеры ее удачного дообучения для решения ряда смежных задач [22, 23], позволяют сделать предположение о перспективности

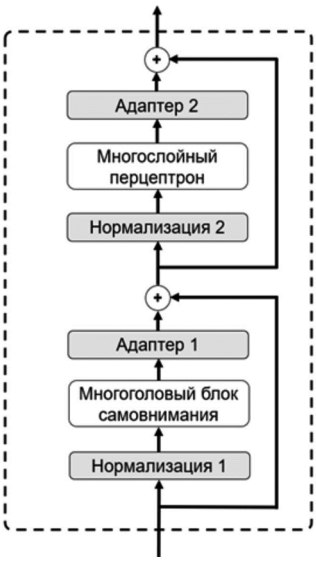


Рисунок 2. Схема расположения адаптеров в блоке Трансформера

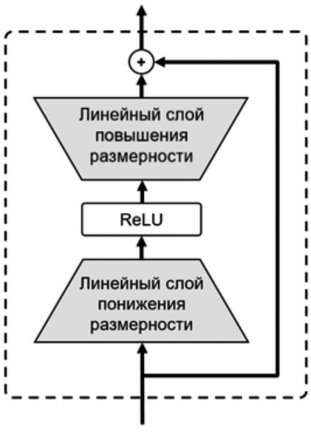


Рисунок 3. Схема блока адаптации (адаптера)

применения кодировщика сети Whisper в задаче обнаружения шумов сердца. Для проверки этого проведено исследование, затрагивающее также вопросы дообучения кодировщика с помощью промт-векторов и блоков адаптации, как это предложено создателями сети WhisPAr [24].

Для проведения исследований общедоступная часть базы фонокардиограмм CirCor DigiScope разделена на три подмножества: обучающее, валидационное и тестовое. Для этого использовалась реализация алгоритма, предложенная коллективом-участником научного конкурса 2022 года [30]. В публикации этих авторов [31] указывается, что оценки метрик на подмножествах, полученных из неполной базы с использованием предложенного алгоритма разбиения, для их алгоритма хорошо согласуются с итоговыми результатами, измеренными на оригинальном тестовом подмножестве базы CirCor DigiScope, которым располагали организаторы научного конкурса 2022 года.

Для предотвращения многократного повторения операции вычисления мел-спектрограмм звуковых сигналов, проведено однократное вычисление мел-спектрограмм всех фонокардиограмм общедоступной части базы CirCor DigiScope с последующим сохранениям в файлы, соответствующие обучающему, валидационному и тестовому подмножествам. Алгоритм и параметры вычисления мел-спектрограмм полностью соответствуют тем, что применяются в сети Whisper. В связи с этим частота дискретизации сигналов из базы повышалась до 16 кГц.

Для аугментации обучающих данных к мел-спектрограммам применялся метод mixup с параметром $\alpha = 0{,}2$ [32]. Предварительные исследования показали, что это значение параметра дает более высокие результаты, чем $\alpha = 0{,}4$ и $\alpha = 1$. Какой-либо адаптации метода аугментации, учитывающей особенности решаемой задачи и применяемой нейронной сети, не проводилось.

Исследуемый классификатор фонокардиограмм построен на основе кодировщика нейронной сети Whisper в конфигурации «base» с двумя сверточными слоями на входе, модифицированного аналогично [24]. В рамках дообучения оригинальные коэффициенты кодировщика сети Whisper, за исключением коэффициентов слоев нормализации, фиксировались. Обучение затрагивало слои нормализации, классификационный и промт-векторы, блоки адаптации и заключительный полносвязный слой, отвечающий за классификацию.

Наряду с общим небольшим объемом обучающих данных в исследовании решалась проблема их сильной несбалансированности по классам. Одним из базовых решений этой проблемы является повторное использование реализаций малопредставленных классов и/или неполное использование реализаций преобладающих классов для компенсации несбалансированности (в англоязычной терминологии oversampling и undersampling, соответственно). Эти подходы часто реализуются с помощью взвешенной случайной выборки, при этом обеспечивается не только общая сбалансированность, но и более строгая – в рамках отдельных батчей. Однако такой подход, реализованный с помощью класса Weighted Random Sampler фреймворка PyTorch, по результатам предварительных исследований оказался менее эффективен, нежели использование весов классов в функции потерь. Возможная причина этого может заключаться в очень сильной (многократной) несбалансированности данных. В связи с этим в качестве функции потерь применяется категориальная кросс-энтропия с весами классов, используемыми для вычисления взвешенной точности в рамках научного конкурса 2022 года [15]. Эти значения весов учитывают, прежде всего, практическую неравнозначность классов, а не их частоту встречаемости в базе. В дальнейшем возможно проведение исследования, направленного на подбор весовых коэффициентов, учитывающих как несбалансированность классов, так и их исходную (практическую) неравнозначность и максимизирующих значение ключевой метрики.

Для обучения нейронной сети использовался алгоритм оптимизации AdamW с линейным увеличением до 0,001 и последующим линейным уменьшением коэффициента, отвечающего за скорость обучения. Размер батчей равнялся 16, длительность обучения составляла 100 эпох. В качестве основной метрики, отслеживаемой для обучающего и валидационного множеств, использовалась взвешенная точность. Проводилось обучение нескольких конфигураций нейронной сети, в том числе без применения промт-векторов и/или блоков адаптации (адаптеров). Исследовались варианты с разным количеством промт-векторов h и разной внутренней размерностью блоков адаптации b. Значения метрик для валидационного подмножества представлены в таблице 1. Для решаемой задачи именно значение взвешенной точности является ключевой метрикой [15], значения точности приводятся в качестве дополнительной информации. В рамках обучения и валидации оценивалась только способность нейронной сети классифицировать отдельные записи, без последующего вынесения решения по пациенту.

Таблица 1. Выбор гиперпараметров нейронной сети

Вариант сети	Параметры	Точность (acc)	Взвешенная точность (Wacc)
Без промт-векторов и адаптеров	–	0,864	0,800
С промт-векторами	h=1	0,850	0,738
	h=2	0,824	0,736
	h=3	0,850	0,824
	h=4	0,814	0,735
С адаптерами	b=8	0,876	0,820
	b=16	0,858	0,806
	b=32	0,870	0,831
	b=64	0,842	0,792
С промт-векторами и адаптерами	h=2, b=8	0,866	0,815
	h=2, b=16	0,846	0,826
	h=2, b=32	0,856	0,827
	h=2, b=64	0,864	0,800
	h=3, b=8	0,816	0,791
	h=3, b=16	0,870	0,808
	h=3, b=32	0,864	**0,832**
	h=3, b=64	0,806	0,767

Таблица 2. Значение метрик на тестовом подмножестве

Классификация	Точность (acc)	Взвешенная точность (Wacc)	Полнота (recall) для класса «Present»	F-мера для класса «Present»
Отдельных записей	0,874	0,833	0,889	0,756
Пациентов	0,820	0,822	0,972	0,707

Видно, что одновременное применение промт-векторов и блоков адаптации для дообучения сети позволяет повысить значение взвешенной точности на валидационном подмножестве по сравнению с базовым вариантом (без их применения).

Далее производилось тестирование варианта сети, показавшего наибольшее значение взвешенной точности на валидационном подмножестве. Значение взвешенной точности классификации записей на тестовом подмножестве составило 0,833. Затем производилась оценка точности классификации пациентов. Для вынесения решения по пациенту использовалось простое правило: если хотя бы в q фонокардиограммах пациента алгоритм детектировал наличие шумов сердца, то аналогичное решение принималось для пациента в целом. Если это условие не выполнялось, но хотя бы одна из записей пациента относилась сетью к классу «Unknown», то это решение принималось для пациента в целом. В остальных случаях считалось, что шумов сердца у пациента не обнаружено. Значение порога q очевидным образом влияет на соотношение между точностью (precision) и полнотой (recall) для класса «Present». Для минимизации вероятности пропуска пациентов с шумами сердца, как и в более ранних исследованиях [29, 33], выбрано значение параметра $q = 1$. Значение взвешенной точности классификации пациентов на тестовом подмножестве составило 0,822. Значения других важных метрик приводится в табл. 2.

Достигнутое значение взвешенной точности превосходит результаты, продемонстрированные участниками научного конкурса 2022 года [15, 31] и близко к текущим достижениям исследователей, работающих с базой сигналов CirCor DigiScope [27]. Дальнейшее улучшение результатов возможно за счет совершенствования алгоритма принятия решения по пациентам и разработки механизмов, направленных на выявление записей, отнесенных к наименее представленному в базе классу «Unknown». Также перспективным является исследование применения других методов аугментации данных, использования для классификации всех скрытых состояний кодировщика Whisper, а также поиск новых нейросетевых архитектур.

Заключение

Рассмотрен вопрос дообучения кодировщика нейронной сети Whisper в конфигурации «base» для обнаружения шумов сердца на фонокардиограммах. Подход на основе применения промт-векторов и блоков адаптации для дообучения кодировщика показал преимущество перед обучением только слоев нормализации и заключительного классификационного слоя сети. Итоговое значение взвешенной точности классификации пациентов на тестовом множестве составило 0,822.

Дальнейшие исследования будут направлены на применение других методов аугментации данных, использование для классификации всех скрытых состояний кодировщика Whisper, а также на анализ возможности применения новых нейросетевых архитектур.

Литература

1. Avanzato R., Beritelli F. Automatic ECG diagnosis using convolutional neural network. Electronics, vol. 9, no. 6. 2020. 14 p.

2. Rajwal S., Aggarwal S. Convolutional neural network-based EEG signal analysis: A systematic review. Archives of Computational Methods in Engineering, vol. 30, no. 1. 2023, pp. 3585-3615.

3. Хрящев В.В., Завьялов Д.В., Андержанова А.С. Классификация эндоскопических изображений устья червеобразного отростка на основе методов глубокого машинного обучения. Цифровая обработка сигналов. 2023. № 1. С. 35-38.

4. Хрящев В.В., Приоров А.Л., Лебедев А.А., Котов Н.А. Обнаружение полипов на колоноскопических изображениях при помощи алгоритмов на базе нейросетевой архитектуры YOLO. Цифровая обработка сигналов. 2023. № 3. С. 27-31.

5. Хрящев В.В. Разработка системы поддержки принятия врачебного решения для эндоскопических исследований желудка. Доклады XXV Международной конференции «Цифровая обработка сигналов и ее применение» (DSPA). Москва, 2023. С. 249-253.

6. Aykanat M., Kılıç Ö., Kurt B., Saryal S. Classification of lung sounds using convolutional neural networks. EURASIP Journal on Image and Video Processing. 2017. 9 p.

7. Bardou D., Zhang K., Ahmad S.M. Lung sounds classification using convolutional neural networks. Artificial intelligence in medicine. 2018, pp. 58-69.

8. Wall C., Zhang L., Yu Y., Kumar A., Gao R. A deep ensemble neural network with attention mechanisms for lung abnormality classification using audio inputs. Sensors, vol. 22(15): 5556. 2022. 25 p.

9. Bansal V., Pahwa G., Kannan N. Cough classification for COVID-19 based on audio MFCC features using convolutional neural networks. 2020 IEEE international conference on computing, power and communication technologies (GUCON). 2020, pp. 604-608.

10. Pahar M., Klopper M., Warren R., Niesler T. COVID-19 cough classification using machine learning and global smartphone recordings. Computers in Biology and Medicine. 2021. 13 p.

11. Маджони С. Секреты клинической диагностики. М.: Бином. 2004. 604 с.

12. Cathers I. Neural network assisted cardiac auscultation. Artificial Intelligence in Medicine, vol. 7, no. 1. 1995. pp. 53-66.

13. Oliveira J., Renna F., Costa P., Nogueira M., Oliveira A.C., Elola A., Ferreira C., Jorge A., Bahrami Rad A., Reyna M., Sameni R., Clifford G., Coimbra M. The CirCor DigiScope phonocardiogram dataset. PhisioNet [Электронный

ресурс]. Режим доступа: https://physionet.org/content/circor-heart-sound/1.0.3/

14. Oliveira J.H., Renna F., Costa P.D., Nogueira M., Oliveira C., Ferreira C., Jorge A., Mattos S., Hatem T., Tavares T., Elola A., Rad A.-B., Sameni R., Clifford G., Coimbra M. The CirCor DigiScope dataset: from murmur detection to murmur classification. IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 6. 2021, pp. 2524-2535.

15. Reyna M., Kiarashi Y., Elola A., Oliveira J., Renna F., Gu A., Perez Alday E.A., Sadr N., Mattos S., Coimbra M., Sameni R., Bahrami Rad A., Koscova Z., Clifford G. Heart murmur detection from phonocardiogram recordings: The George B. Moody PhysioNet Challenge 2022. PLOS Digital Health, vol. 2, no. 9. 2023. 22 p.

16. Baevski A., Zhou Y., Mohamed A., Auli M. Wav2vec 2.0: A framework for self-supervised learning of speech representations. Advances in neural information processing systems, vol. 33. 2020, pp. 12449-12460.

17. Gulati A., Qin J., Chiu C.C., Parmar N., Zhang Y., Yu J., Pang R. Conformer: convolution-augmented transformer for speech recognition. arXiv preprint arXiv:2005.08100. 2020. 5 p.

18. Hsu W.N., Bolte B., Tsai Y.H.H., Lakhotia K., Salakhutdinov R., Mohamed A. HuBERT: self-supervised speech representation learning by masked prediction of hidden units. IEEE/ACM transactions on audio, speech, and language processing, vol. 29. 2021, pp. 3451-3460.

19. Jeffries N., King E., Kudlur M., Nicholson G., Wang J., Warden P. Moonshine: speech recognition for live transcription and voice commands. arXiv preprint arXiv:2410.15608. 2024. 7 p.

20. Radford A., Kim J.W., Xu T., Brockman G., McLeavey C., Sutskever I. Robust speech recognition via large-scale weak supervision. International conference on machine learning. 2023, pp. 28492-28518.

21. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser L., Polosukhin I. Attention is all you need. Advances in neural information processing systems. 2017, pp. 5998-6008.

22. Zhang L., Jiang N., Wang Q., Li Y., Lu Q., Xie L. Whisper-SV: adapting Whisper for low-data-resource speaker verification. Speech Communication, vol. 163. 2024.

23. Gong Y., Khurana S., Karlinsky L., Glass J. Whisper-AT: noise-robust automatic speech recognizers are also strong general audio event taggers. arXiv preprint arXiv:2307.03183. 2023. 5 p.

24. Shi B., Wang H., Lu C., Zhao M. WhisPAr: transferring pre-trained audio models to fine-grained classification via prompt and adapter. Knowledge-Based Systems, vol. 300, 112242. 2024. 10 p.

25. Houlsby N., Giurgiu A., Jastrzebski S., Morrone B., Laroussilhe Q.D., Gesmundo A., Attariyan M., Gelly S. Parameter-efficient transfer learning for NLP. Proceedings of the International Conference on Machine Learning. 2019, pp. 2790-2799.

26. Radford A., Wu J., Child R., Luan D., Amodei D., Sutskever I. Language models are unsupervised multitask learners. OpenAI blog, vol. 1, no. 8. 2019. 24 p.

27. Panah D.S., Hines A., McKeever S. Exploring wav2-vec 2.0 model for heart murmur detection. 31st European Signal Processing Conference (EUSIPCO), pp. 1010-1014.

28. Zhou G., Chien C., Chen J., Luan L., Chen Y., Carroll S., Dayton J., Thanjan M., Bayle K., Flynn P. Identifying pediatric heart murmurs and distinguishing innocent from pathologic using deep learning. Artificial Intelligence in Medicine, vol. 153. 2024. 13 p.

29. Топников А.И., Смирнова Т.А. Применение предобученных нейронных сетей в задаче детектирования шумов сердца на фонокардиограммах. Проблемы информатики в образовании, управлении, экономике и технике: материалы 24-й междунар. научно-техн. конф. Пенза, 2024. С. 239-244.

30. GitHub – Benjamin-Walker/heart-murmur-detection [Электронный ресурс]. Режим доступа: https://github.com/Benjamin-Walker/heart-murmur-detection

31. Walker B., Krones F, Kiskin I., Parsons G., Lyons T., Mahdi A. Dual Bayesian ResNet: a deep learning approach to heart murmur detection. Computing in Cardiology Conference (CinC). 2022. 5 p.

32. Zhang H., Cisse M., Dauphin Y.N., Lopez-Paz D. Mixup: beyond empirical risk minimization. International Conference on Learning Representations. 2018. 13 p.

33. Смирнова Т.А. Применение предобученной нейронной сети YAMNet в задаче детектирования шумов сердца на фонокардиограммах. Биотехнические, медицинские и экологические системы, измерительные устройства и робототехнические комплексы (БИОМЕДСИСТЕМЫ-2024): материалы 37-й всерос. научно-техн. конф. Рязань, 2024. С. 277-282.

НОВЫЕ КНИГИ



Бабаян П.В., Муравьев В.С., Смирнов С.А., Стротов В.В.

Обработка изображений в системах обнаружения и сопровождения объектов. Классические методы: Учебное пособие для вузов: Учебное издание, М.: Изд-во «Горячая линия-Телеком», 2024 г. 168 с.: ил.

Освещены вопросы пространственно-временной обработки и анализа последовательностей изображений применительно к задачам обнаружения и сопровождения объектов. Значительное внимание уделено алгоритмам восстановления изображений и определению параметров их геометрических преобразований.

Для студентов вузов, обучающихся по направлениям подготовки: 27.03.04 – «Управление в технических системах», 01.03.02 – «Прикладная математика и информатика», 15.03.06 – «Мехатроника и робототехника», 27.04.04 – «Управление в технических системах», а также по специальности 12.05.01 – «Электронные и оптико-электронные приборы и системы специального назначения». Будет полезна аспирантам и специалистам соответствующего профиля.