

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ «НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЯДЕРНЫЙ
УНИВЕРСИТЕТ «МИФИ»

УДК 531.3, 539.1.05

ОТЧЕТ
О ПРЕДДИПЛОМНОЙ ПРАКТИКЕ
**РЕКОНСТРУКЦИЯ СИГНАЛОВ В
ДЕТЕКТОРЕ FARICH УСТАНОВКИ SPD
НА КОЛЛАЙДЕРЕ NICA**

Научный руководитель

к. ф-м. н.

_____ Е. Ю. Солдатов

Научный консультант

к. ф-м. н.

_____ А. С. Жемчугов

Студент

_____ К. Ю. Массалов

Москва 2024

СОДЕРЖАНИЕ

Введение	4
1 Экспериментальная установка Spin Physics Detector на строящемся в ОИЯИ коллайдере NICA	6
1.1 Конструкция детектора	6
1.2 FARICH	8
2 Алгоритмы распознавания колец	9
2.1 Разделение частиц по вторым моментам	9
2.2 Итерационный метод	9
2.3 Отношение функций максимального правдоподобия	10
2.4 Метод масок	10
2.5 Нечеткая кластеризация	10
2.6 Преобразование Хафа	11
2.7 СОР	11
2.8 Фитирование эллипса	12
3 Моделирование детектора	13
3.1 GEANT4	13
3.2 Моделирование детектора и событий	13
3.3 Вид обучающей выборки	14
4 Применение нейросетевых методов	16
4.1 Варианты обработки сигналов детектора с использованием информации о треках частиц	16
4.1.1 Варианты архитектур	16
4.1.2 Подходы к обработке множества треков	17
4.2 Позиционное кодирование	18
4.2.1 Архитектура	19

4.2.2	Результаты	19
4.3	Устойчивость к шуму	20
4.3.1	Без дополнительной очистки	21
4.3.2	С дополнительным этапом очистки	22
4.3.3	Оптимизированная архитектура с этапом очистки	22
4.3.4	Финальное сравнение	24
4.4	Сегментация исходного изображения	24
4.5	Исследование масштабируемости моделей	26
4.6	Исследование интерпретируемости моделей	27
4.7	Варианты обработки сигналов детектора без использования информации о треках частиц	29
	Заключение	32

Список использованных источников	33
---	-----------

ВВЕДЕНИЕ

Относительный вклад видимой барионной материи во Вселенную составляет около 5% от общей массы. Хотя барионная материя представляется достаточно изученной, вопрос о том, почему нуклоны имеют именно такую структуру и свойства, остается открытым. Одним из главных нерешенных вопросов в квантовой хромодинамике (КХД) остается понимание структуры и основных свойств нуклонов исходя из динамики кварков и глюонов внутри них.

Нуклон обладает спином, который отвечает за его магнитный момент и другие свойства. Изучение спиновой структуры нуклонов является важной задачей. Центральной задачей является понимание того, как спин нуклона формируется из спинов и орбитальных моментов кварков и глюонов.

За последние 25 лет эксперименты по глубоко неупругому рассеянию (CERN, DESY, JLab, SLAC) и высокоэнергетические столкновения протонов (RHIC) были основными источниками информации о спиновых структурных функциях нуклонов. Наши знания о внутренней структуре нуклонов все еще ограничены, особенно в отношении глюонного вклада.

Spin Physics Detector — универсальная установка для исследования спиновой структуры нуклонов и других спиновых явлений. Установка будет использовать поляризованные пучки протонов и дейтронов. Проект основан на большом опыте работы с поляризованными пучками в ОИЯИ. Основная цель эксперимента - комплексное исследование неполяризованного и поляризованного глюонной составляющей нуклона. Измерения спина в SPD на коллайдере NICA откроют новые возможности для понимания спиновой структуры нуклона.

На установке планируется частота соударений до 4 МГц и число выходных каналов с детекторов ~ 700000 . Таким образом, ожидается поток данных примерно 20 Гбайт/с или 200 Пбайт/год. Такой объем информа-

ции необходимо уменьшать, избавляясь от неинформативных событий и шумов. Для этого планируется использовать онлайн-фильтр, представляющий собой высокопроизводительную систему, которая будет включать в себя гетерогенные вычислительные платформы, аналогичные многим высокопроизводительным вычислительным кластерам. Для реконструкции и отбрасывания неинтересных событий можно использовать методы машинного обучения, поскольку проблема классических алгоритмов в том, что они имеют невысокую скорость и плохо параллелятся, в отличие от нейронных сетей. Проблема же нейронных сетей в сложности интерпретации их решений: никогда нельзя знать наверняка, почему сеть выдала именно такой результат, и насколько она права.

Таким образом, планируется использовать нейронные сети и классические алгоритмы одновременно: первые обеспечат скорость работы, вторые позволят проверять правильность результатов нейросетей.

Цель работы состоит в создании нейросетевых методов реконструкции сигналов с детектора FARICH, изучении их характеристик и сравнении их с классическими подходами.

1 ЭКСПЕРИМЕНТАЛЬНАЯ УСТАНОВКА SPIN PHYSICS ДЕТЕКТОР НА СТРОЯЩЕМСЯ В ОИЯИ КОЛЛАЙДЕРЕ NICA

Коллаборация Spin Physics Detector предлагает установить универсальный детектор во второй точке взаимодействия строящегося ускорителя NICA в Дубне. Этот детектор будет использоваться для изучения спиновой структуры протона и дейтрона и других спиновых эффектов с помощью поляризованных пучков протонов и дейтронов с энергией столкновения до 27 ГэВ и светимостью до $10^{32} \text{ см}^{-2} \text{ с}^{-1}$. Эксперимент SPD в NICA позволит заполнить кинематический пробел между измерениями при низких энергиях на установках ANKE-COSY и SATURNE, измерениями при высоких энергиях на ускорителе тяжелых ионов и будущими экспериментами на Большом адронном коллайдере (БАК). Уникальность NICA заключается в возможности работы с пучками поляризованных дейтронов при этих энергиях.

1.1 КОНСТРУКЦИЯ ДЕТЕКТОРА

Экспериментальная установка SPD (Рис. 1.1) проектируется как универсальный детектор 4π с расширенными возможностями для отслеживания и идентификации частиц на основе современных технологий. Вершинный детектор (vertex detector) на основе кремния обеспечит разрешение по позиции вершин на уровне менее 100 микromетров, необходимое для восстановления вторичных вершин от распадов D- мезонов. Система трекинга (straw tracker), помещенных в соленоидное магнитное поле с напряженностью до 1 Тесла на оси детектора, обеспечивает разрешение по попе-

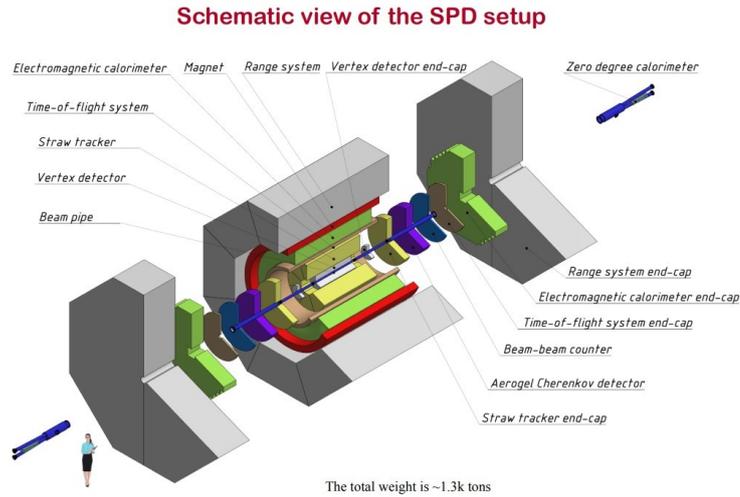


Рисунок 1.1 — Конструкция детектора

речной импульсу $\sigma_{p_T}/p_T \approx 2\%$ для частиц с импульсом 1 ГэВ/с. Время-пролетная система (time-of-flight system) с временным разрешением около 60 пикосекунд обеспечивает разделение 3σ для пионов/каонов и для каонов/протонов до энергий 1.2 и 2.2 ГэВ соответственно. Использование черенковского детектора на базе аэрогеля — FARICH — поможет расширить этот диапазон вплоть до 6 ГэВ. Обнаружение фотонов обеспечивается электромагнитным калориметром (electromagnetic calorimeter) с энергетическим разрешением $5\%/\sqrt{E}$. Для идентификации мюонов планируется использовать мюонную систему (range system). Она также может использоваться как грубый адронный калориметр. За локальную поляризацию и контроль светимости будет отвечать пара счетчиков пучков (beam-beam counter) и калориметров малого угла, или малоапертурных калориметров (zero-degree calorimeter).

1.2 FARICH

FARICH (Focusing Aerogel Ring-Imaging Cherenkov detector) — черенковский детектор с аэрогелем, используемым в качестве рабочего тела. Черенковский детектор работает на основе эффекта Черенкова, который заключается в излучении света заряженными частицами, движущимися в среде со скоростью, превышающей фазовую скорость света в этой среде: $\beta n > c$. Это излучение имеет вид конуса, ось которого совпадает с направлением движения частицы, а угол раствора зависит от скорости частицы и показателя преломления среды. Излучение регистрируется массивом фотодетекторов, в результате чего на выходе получается срез конуса плоскостью. Если частица летела перпендикулярно детектору, получается набор точек, формирующих окружность, которую в дальнейшем на онлайн фильтре надо будет реконструировать при помощи методов машинного обучения.

В качестве среды планируется использовать четыре слоя аэрогеля каждый со своим показателем преломления, что позволит улучшить разрешающую способность детектора [1].

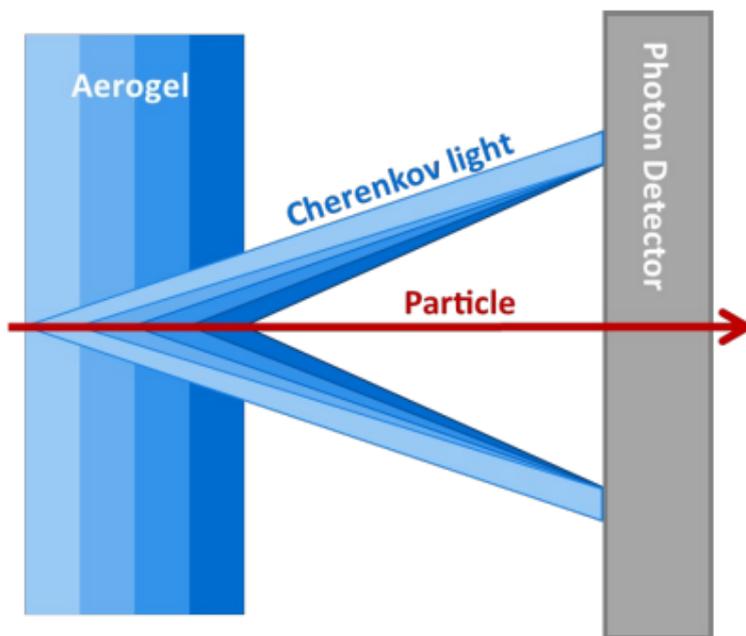


Рисунок 1.2 — Принципиальная схема детектора FARICH

2 АЛГОРИТМЫ РАСПОЗНАВАНИЯ КОЛЕЦ

2.1 РАЗДЕЛЕНИЕ ЧАСТИЦ ПО ВТОРЫМ МОМЕНТАМ

Простейшим методом идентификации является разделение по вторым моментам. Для каждого значения импульса второй центральный момент для кольца будет меняться в зависимости от типа частицы. Зная зависимость вторых моментов от импульса для разных частиц, можно определить их тип.

2.2 ИТЕРАЦИОННЫЙ МЕТОД

Для данного метода необходима дополнительная информация о предположительных центрах колец, которую можно получить независимо, например, в результате экстраполяции трека из трекового детектора.

Принцип работы следующий: при инициализации рассчитывается расстояние от каждого хита до каждого трека. Полученными значениями с весом 1 заполняют гистограммы (по числу треков). Затем, на каждой итерации обновляют веса хитов по формуле:

$$w_k = \frac{y_k}{\sum_i y_i}, \quad (2.1)$$

где w_k — вес хита в k -той гистограмме, y_k — содержание ячейки гистограммы, в которую попал данный хит. В результате получаются пики, которые соответствуют центрам колец [2].

2.3 ОТНОШЕНИЕ ФУНКЦИЙ МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Данный метод также требует дополнительной информации о центрах колец. На основании информации о конструкции детектора и параметрах трека делается предсказание положения черенковского фотона при гипотезе о типе частицы. На основе предсказания и наблюдения строится функция правдоподобия, а затем выбирается функция с максимальным правдоподобием [2].

2.4 МЕТОД МАСОК

На предварительно центрированное изображение накладывается трафарет с известными параметрами. Наложение заключается в сложении пикселей маски и изображения по модулю два. Чем больше нулей в результате, тем больше данная маска соответствует изображению [2].

2.5 НЕЧЕТКАЯ КЛАСТЕРИЗАЦИЯ

Метод заключается в минимизации следующего функционала:

$$J = \sum_{i=1}^C \sum_{j=1}^N u_{ij} d_{ij}^2, \quad (2.2)$$

где u_{ij} — весовой коэффициент, показывающий вероятность принадлежности хита i к кластеру j ($\sum_j u_{ij} = 1$), d_{ij} — расстояние между i хитом и j кластером, C — количество колец, N — количество хитов.

В данном методе необходимо также знать центры колец и их количество [2].

2.6 ПРЕОБРАЗОВАНИЕ ХАФА

Изначально преобразование Хафа предназначено для распознавания прямых и заключается в идее перейти из пространства координат в пространство параметров и нахождения там точки с наибольшей «интенсивностью», т.е. точки с наибольшим числом пересечений прямых. Полученная точка в пространстве параметров и есть параметры, описывающие прямую, проходящую через исходные точки.

Затем данное преобразование можно обобщить и на случай окружностей, перейдя в трехмерное пространство параметров [3].

2.7 СОР

Алгоритм СОР (Chernov-Ososkov-Pratt) позволяет по точкам оценить параметры окружности [4].

Для подгонки окружности методом наименьших квадратов необходимо минимизировать функционал:

$$L(a, b, R) = \sum_{i=1}^N (\sqrt{(x_i - a)^2 + (y_i - b)^2} - R)^2, \quad (2.3)$$

где a, b — x и y координаты центра окружности, R — ее радиус, x_i и y_i — координаты i -й точки, N — число точек.

Однако, такой метод требует достаточно много времени для вычисления. Для ускорения работы алгоритма можно использовать следующий функционал:

$$M(a, b, c) = \sum_{i=1}^N ((x_i - a)^2 + (y_i - b)^2 - R^2)^2 \quad (2.4)$$

Минус такого подхода в том, что при измерениях, не покрывающих всю окружность, результат может оказаться довольно далеко от истины.

В алгоритме СОР предлагается использовать функционал:

$$M(a, b, c) = \sum_{i=1}^N [((x_i - a)^2 + (y_i - b)^2 - R^2)^2 / R^2] \quad (2.5)$$

Нахождение минимума такого функционала эквивалентно минимизации (2.3), однако вычислительно менее затратно.

2.8 ФИТИРОВАНИЕ ЭЛЛИПСА

Для фитирования эллипса предлагается использовать функционал вида [5]:

$$L(x_{F_1}, y_{F_1}, x_{F_2}, y_{F_2}, a) = \sum_{i=1}^N \left(\sqrt{(x_i - x_{F_1})^2 + (y_i - y_{F_1})^2} + \sqrt{(x_i - x_{F_2})^2 + (y_i - y_{F_2})^2} - 2a \right)^2, \quad (2.6)$$

где $x_{F_1}, y_{F_1}, x_{F_2}, y_{F_2}$ — координаты x и y первого и второго фокуса эллипса соответственно, a — длина большой полуоси.

Минимизация осуществляется при помощи градиентного спуска.

3 МОДЕЛИРОВАНИЕ ДЕТЕКТОРА

3.1 GEANT4

GEANT4 — это программный инструментарий на C++ для моделирования экспериментов в области физики элементарных частиц. Он разрабатывается и поддерживается Европейской организацией по ядерным исследованиям (CERN) [6–8].

GEANT4 используется для моделирования прохождения частиц через материю, включая взаимодействия, которые происходят при столкновении частиц с атомами в материале. Эта информация может быть использована для проектирования экспериментов, понимания результатов этих экспериментов и создания симуляций будущих экспериментов.

В данной работе пакет GEANT4 использовался для моделирования детектора FARICH.

3.2 МОДЕЛИРОВАНИЕ ДЕТЕКТОРА И СОБЫТИЙ

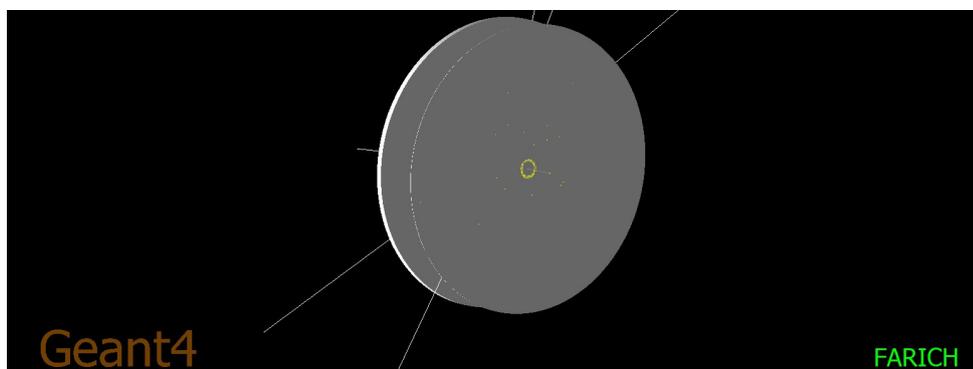


Рисунок 3.1 — Моделирование детектора FARICH

Модель детектора (Рис. 3.1)— два диска радиусом 847 мм. Первый диск состоит из четырех слоев аэрогеля на основе диоксида кремния с из-

меняющимся показателем преломления: 1.037, 1.041, 1.043, 1.047. Толщины слоев 7 мм, 10 мм, 9 мм, 10 мм соответственно. Второй — чувствительный детектор, который регистрирует черенковские фотоны в диапазоне примерно от 400 нм до 450 нм — в области чувствительности кремниевых фотодетекторов. Расстояние между аэрогелем и фотоприемником — 164 мм.

Для обучения и проверки моделей было смоделировано 210000 одночастичных событий — по 70000 событий с пионом, каоном и протоном. Затем отобраны события, в которых число хитов больше 5, а также χ^2 которых при фитировании эллипсом меньше 3. В результате получилось 70000 одночастичных событий.

3.3 ВИД ОБУЧАЮЩЕЙ ВЫБОРКИ

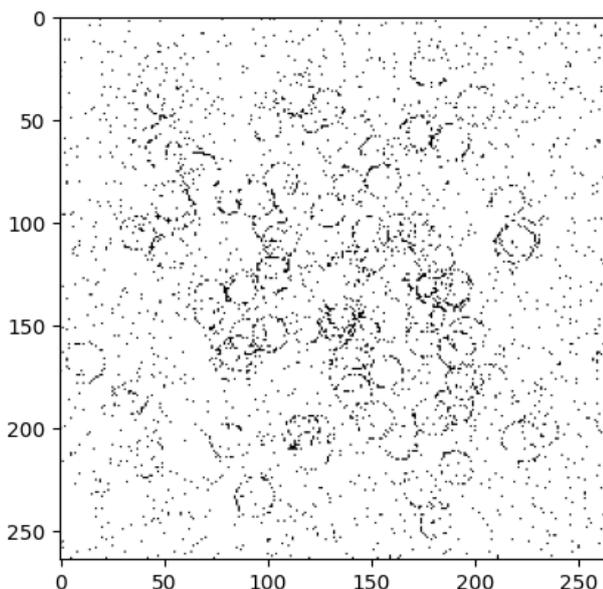


Рисунок 3.2 — Вид одного события из 100 частиц с детектора. Дополнительно каждый пиксель с вероятностью 0.025 — шумовой

Во время работы установки детектор будет за раз регистрировать порядка 100 частиц, поэтому одно событие с детектора создается путем сэмплирования из набора одночастичных событий. Кроме того, моделируется шум для каждого пикселя из распределения Бернулли. Также случайно добавляется случайный сдвиг на один пиксель в разные стороны, имитирующее ошибку экстраполяции трека.

Также, кроме сигналов с непосредственно детектора, есть возможность получить экстраполированные треки частиц и их импульсы. Таким образом, можно локализовать интересующие области.

Пример одного события представлен на рис. 3.2.

4 ПРИМЕНЕНИЕ НЕЙРОСЕТЕВЫХ МЕТОДОВ

В связи с быстрым развитием методов машинного обучения и, в частности, нейросетевых подходов, возникает возможность применения этих методов для анализа данных, полученных с FARICH детектора. Нейросетевые методы обладают уникальной способностью выявлять сложные зависимости в данных, что делает их идеальным инструментом для распознавания колец и анализа параметров частиц, регистрируемых в эксперименте. Такой подход не только повышает точность распознавания, но и позволяет создавать более эффективные модели, способствуя глубокому пониманию физических процессов.

4.1 ВАРИАНТЫ ОБРАБОТКИ СИГНАЛОВ ДЕТЕКТОРА С ИСПОЛЬЗОВАНИЕМ ИНФОРМАЦИИ О ТРЕКАХ ЧАСТИЦ

4.1.1 ВАРИАНТЫ АРХИТЕКТУР

Предлагаются два варианта архитектуры нейронных сетей для решения поставленной задачи:

- 1) Обработка неупорядоченных наборов точек, представленных координатами x и y .

- 1.1 Для получения векторного представления набора точек можно усреднить или максимизировать значения вдоль каждого канала (координаты x и y), затем выполнить конкатенацию полученного вектора с каждым вектором из набора точек и применить два полносвязных слоя с нелинейностью. Последовательное приме-

нение нескольких таких блоков с операцией глобального пулинга позволит получить векторное описание события, пригодное для дальнейшего анализа. Этот подход основан на архитектуре PointNet [9].

1.2 Набор точек также можно рассматривать как неупорядоченную последовательность, что позволяет использовать трансформеры и одномерные свертки.

2) Использование сверточных нейронных сетей (CNN) [10].

Основное преимущество первого подхода — меньшая вычислительная сложность, поскольку обработке подвергаются только ненулевые пиксели, число которых значительно меньше общего количества (например, на рис. 3.2 из 264x264 пикселей ненулевыми являются всего 3717). Однако в экспериментах данный подход показал значительно худшие результаты по сравнению с использованием двумерных сверток, поэтому в дальнейшем будут рассматриваться только они.

4.1.2 ПОДХОДЫ К ОБРАБОТКЕ МНОЖЕСТВА ТРЕКОВ

Для обработки множества треков предлагаются следующие три подхода:

- 1) Создание сжатого векторного представления (эмбединга) исходного изображения с использованием сверток и последующая обработка этого эмбединга совместно с вектором, описывающим трек каждой частицы (координаты и импульс). Этот подход может быть вычислительно эффективным при большом количестве частиц на событие (>100).
- 2) Обработка каждого трека независимо путем выделения интересующей области вокруг точки, обозначающей пересечение экстраполированного трека с плоскостью детектора.
- 3) Сегментация изображения: для каждого пикселя предсказание набора интересующих значений, а затем выделение только тех значений, которые лежат на точке пересечения экстраполированного трека с плоскостью детектора.

На практике эффективное сжатие пространственной информации в

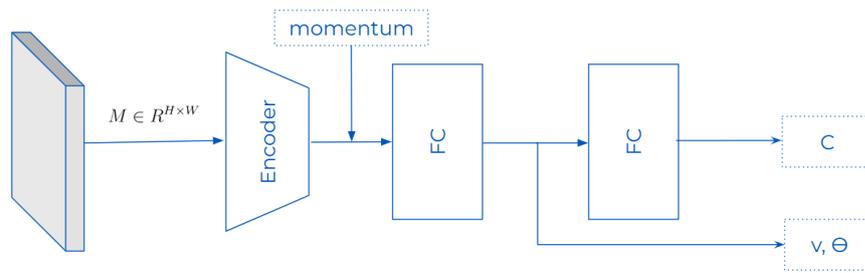


Рисунок 4.1 — Схема модели (вариант 1)

один вектор оказалось невозможным. В результате, при использовании первого подхода нейронная сеть предсказывала только константное значение, минимизирующее соответствующую функцию потерь.

В следующих секциях будут подробнее рассмотрены варианты два и три.

4.2 ПОЗИЦИОННОЕ КОДИРОВАНИЕ

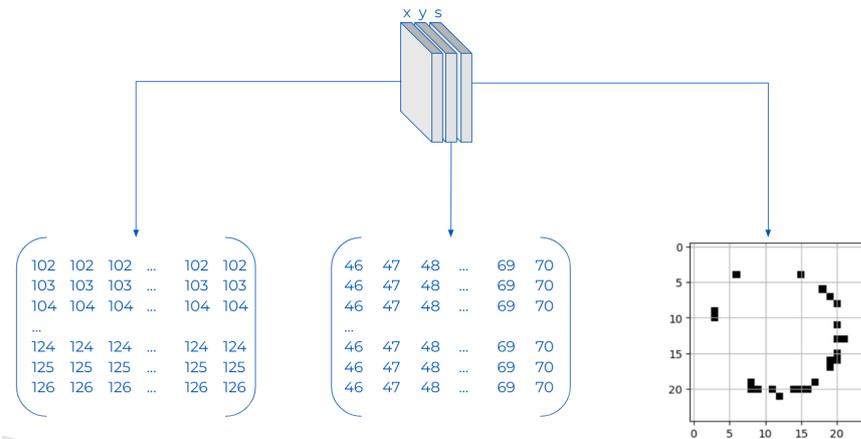


Рисунок 4.2 — Пример входа модели с координатами

Поскольку нейронная сеть обрабатывает область, вырезанную из полного изображения, она лишается информации об абсолютном положении трека. Для компенсации этой потери используется позиционное кодирование.

В данной секции будет рассмотрена важность позиционного кодирования. Для этого используется модель с независимой обработкой треков.

4.2.1 АРХИТЕКТУРА

Как показано на рис. 4.1, модель состоит из энкодера (последовательность сверточных и пулинговых слоев), который служит для извлечения признаков из входного изображения, и набора полносвязных слоев. Входной тензор имеет размерность $1 \times 25 \times 25$ без позиционного кодирования и $3 \times 25 \times 25$ с позиционным кодированием. Выходные значения соответствуют уверенности модели в идентификации зарегистрированной частицы. Для повышения интерпретируемости и обобщающей способности модели, за несколько слоев до классификационного выхода добавлены дополнительные выходы регрессии, предсказывающие скорость и угол черенковского конуса.

Размер входного тензора выбран, исходя из следующих соображений: размер пикселя составляет $5.8 \times 5.8 \text{ мм}^2$, а максимальный радиус окружности, образуемой сечением черенковского конуса плоскостью детектора, составляет приблизительно 6 см. Следовательно, размер области 25×25 пикселей обеспечивает достаточное покрытие для обработки данных без потери информации.

При использовании позиционного кодирования в два дополнительных канала записываются номера бинов вдоль осей x и y (см. рис. 4.2). Эти значения затем нормируются на интервал $(-1, 1)$ для улучшения сходимости модели. Это позволяет модели: 1) получить информацию о масштабе колец, важную для определения скорости и угла, и 2) понимать положение точки входа частицы, что, в свою очередь, позволяет оценить угол ее вылета и угол наклона оси эллипса относительно плоскости детектора.

4.2.2 РЕЗУЛЬТАТЫ

На рисунке 4.3 показаны двумерные гистограммы, отображающие распределение угла черенковского конуса в зависимости от импульса, а также ROC-кривые. Микроусредненное значение ROC AUC составляет 0.987 ± 0.002 .

На рисунке 4.4 представлены аналогичные результаты: двумерные гистограммы для распределения угла черенковского конуса в зависимости

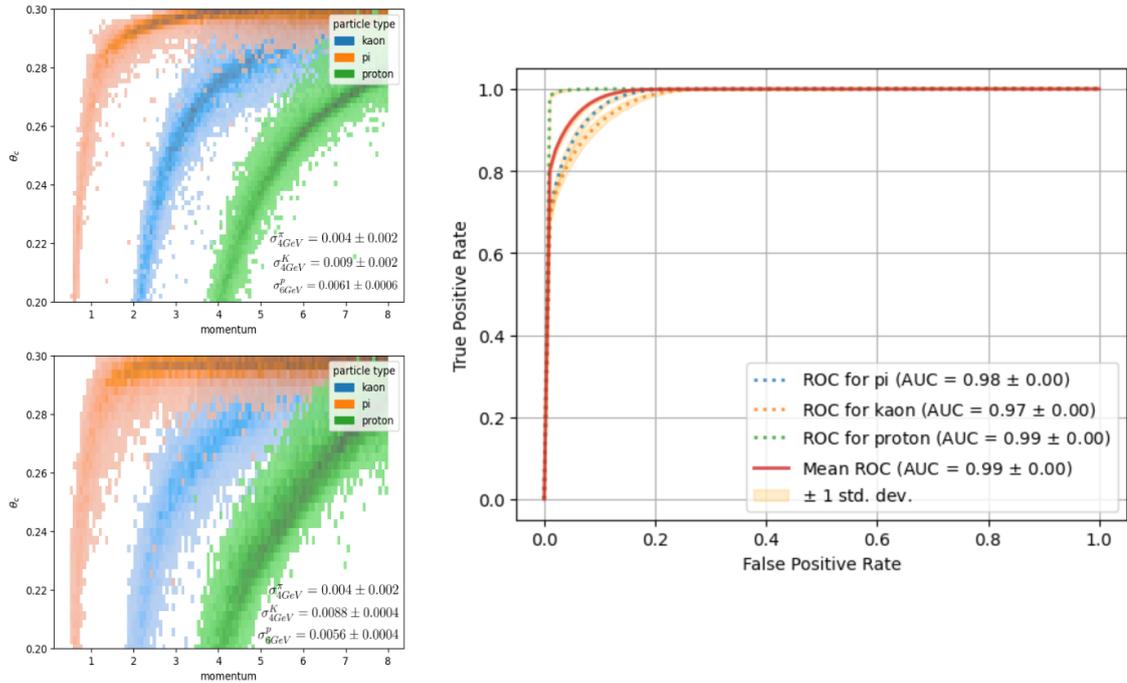


Рисунок 4.3 — Слева – двумерные гистограммы в координатах импульс-угол черенковского конуса; сверху угол восстановлен из предсказанной скорости, снизу – угол берется напрямую из выхода нейросети. Справа – ROC для классификации частицы на каждый из трех классов по отдельности, а также в общем

от импульса и ROC-кривые. Микроусредненное значение ROC AUC в этом случае достигает 0.9994 ± 0.0001 , что свидетельствует о значительно более высокой точности модели. В дальнейшем эта модель будет использоваться в качестве основы для будущих исследований.

4.3 УСТОЙЧИВОСТЬ К ШУМУ

Следующим важным шагом является исследование устойчивости модели к шуму. Детектор в процессе работы будет генерировать ложные сигналы, поэтому важно обучить модель игнорировать их.

Для этого к входному изображению добавлялся шум, моделируемый распределением Бернулли: каждый пиксель с вероятностью $*p*$ инвертировался. Было исследовано влияние различных уровней шума ($*p*$ принимало разные значения) на качество работы модели.

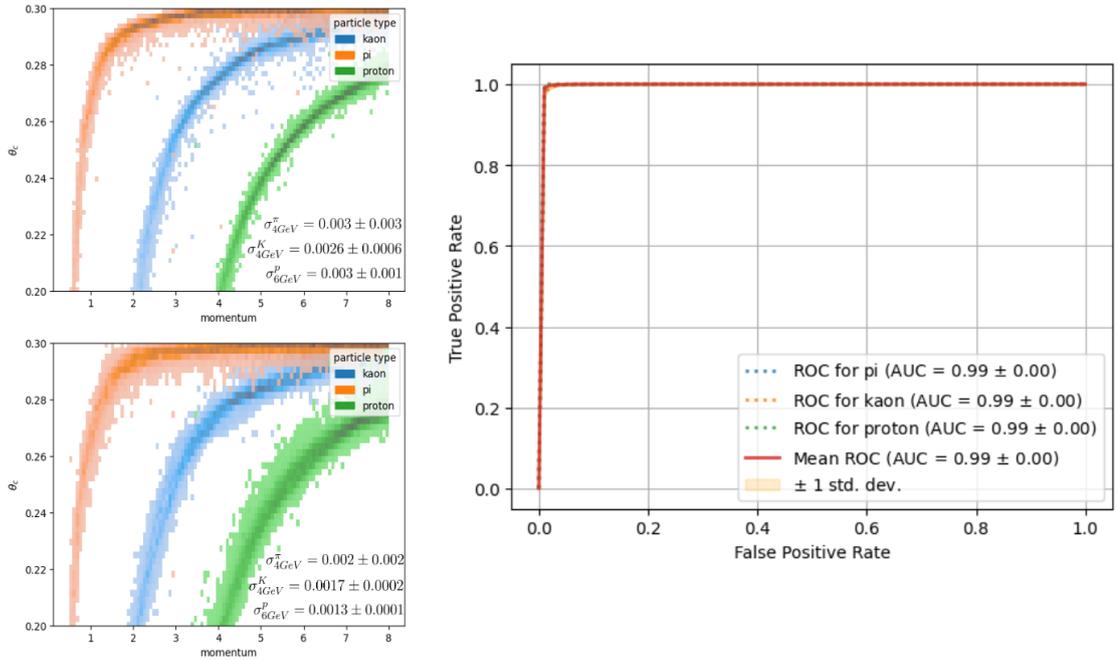


Рисунок 4.4 — Слева – двумерные гистограммы в координатах импульс-угол черенковского конуса; сверху угол восстановлен из предсказанной скорости, снизу – угол берется напрямую из выхода нейросети. Справа – ROC для классификации частицы на каждый из трех классов по отдельности, а также в общем

4.3.1 БЕЗ ДОПОЛНИТЕЛЬНОЙ ОЧИСТКИ

В этой секции представлены результаты, полученные для базовой модели, описанной в секции 4.2.1. Дополнительная очистка не применялась: зашумленное изображение подавалось непосредственно на вход модели, а на выходе оценивались скорость, угол и тип частицы.

Noise Level	ROC AUC	MAE
0.01-0.02	0.9972 ± 0.0005	0.0026 ± 0.0003
0.03-0.04	0.995 ± 0.001	0.0029 ± 0.0002
0.05-0.06	0.991 ± 0.002	0.0034 ± 0.0003
0.07-0.08	0.989 ± 0.002	0.0038 ± 0.0004
0.09-0.10	0.985 ± 0.003	0.0041 ± 0.0004

Таблица 4.1 — Качество модели без дополнительной очистки при разных уровнях шума

В таблице 4.1 представлены значения ROC AUC и MAE для угла черенковского конуса при различных уровнях шума. Как и ожидалось, качество предсказания ухудшается с увеличением уровня шума.

4.3.2 С ДОПОЛНИТЕЛЬНЫМ ЭТАПОМ ОЧИСТКИ

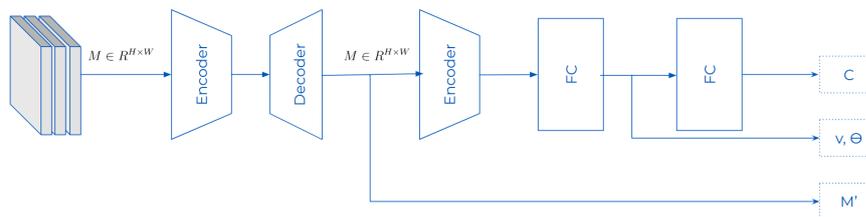


Рисунок 4.5 — Модель с дополнительным этапом очистки от шума

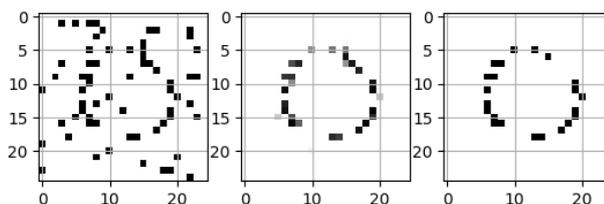


Рисунок 4.6 — Результат очистки от шума. Слева исходное изображение, по центру – результат работы нейросети, справа – искомое изображение

Для повышения устойчивости к шуму был применен подход, включающий предварительную очистку изображения. Для этого к базовой модели (рис. 4.1) перед энкодером была добавлена небольшая UNet-подобная [11] сеть, состоящая из пары энкодер-декодер (рис. 4.5). Пример результата очистки зашумленного изображения представлен на рисунке 4.6.

В таблице 4.2 представлены значения ROC AUC и MAE для угла черенковского конуса при различных уровнях шума после применения этапа очистки. Качество предсказания по-прежнему ухудшается с увеличением уровня шума, хотя и в меньшей степени. Как видно из таблицы, применение этапа очистки позволяет несколько улучшить метрики качества по сравнению с моделью без очистки.

4.3.3 ОПТИМИЗИРОВАННАЯ АРХИТЕКТУРА С ЭТАПОМ ОЧИСТКИ

Предыдущая архитектура является вычислительно затратной из-за наличия дополнительного энкодера и декодера. Для снижения вычислительной нагрузки в новой архитектуре (рис. 4.7) второй энкодер был уда-

Noise level	ROC AUC	MAE
0.01-0.02	0.9988 ± 0.0002	0.0021 ± 0.0004
0.03-0.04	0.9979 ± 0.0004	0.0022 ± 0.0003
0.05-0.06	0.9967 ± 0.0007	0.0025 ± 0.0001
0.07-0.08	0.994 ± 0.001	0.0027 ± 0.0002
0.09-0.10	0.986 ± 0.002	0.0039 ± 0.0002

Таблица 4.2 — Качество модели с дополнительной очисткой при разных уровнях шума

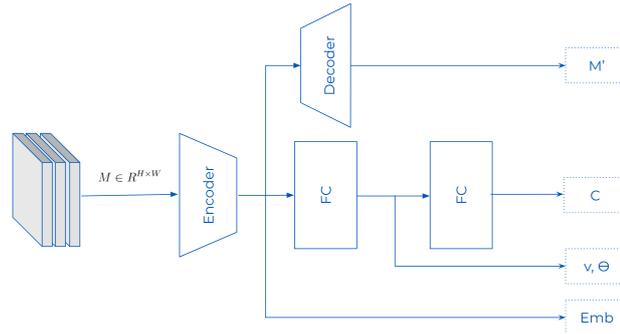


Рисунок 4.7 — Модель с вынесенным в сторону декодером

лен, а блок декодера вынесен за пределы основной вычислительной линии, что позволяет отключать его, когда очистка изображения не требуется.

В процессе обучения в функцию потерь добавляется слагаемое, представляющее собой косинусное расстояние между выходами энкодера для зашумленного и очищенного изображений. Целью является максимизация этого расстояния, что способствует сохранению полезной информации при очистке.

Результаты, полученные для данной архитектуры, представлены в таблице 4.3.

Noise level	ROC AUC	MAE
0.01-0.02	0.997 ± 0.001	0.0025 ± 0.0004
0.03-0.04	0.995 ± 0.001	0.0027 ± 0.0003
0.05-0.06	0.992 ± 0.001	0.0032 ± 0.0005
0.07-0.08	0.989 ± 0.002	0.0034 ± 0.0004
0.09-0.10	0.985 ± 0.004	0.0039 ± 0.0005

Таблица 4.3 — Качество модели с дополнительной очисткой при разных уровнях шума

4.3.4 ФИНАЛЬНОЕ СРАВНЕНИЕ

В таблице 4.4 представлены итоговые значения метрик и время работы для трех рассмотренных моделей. Модель с отдельным этапом очистки (вторая модель) демонстрирует наивысшую точность, но и наибольшее время выполнения. Базовая модель (первая модель) и оптимизированная модель с очисткой (третья модель) показывают сравнимые результаты, однако первая не позволяет оценивать качество очистки изображения.

	AUC	MAE(v)	MAE(Θ)	Inference time (batch, 1000 samples)
v1	0.9966 ± 0.0009	0.0010 ± 0.0001	0.0029 ± 0.0003	(12.50 \pm 0.02) ms
v2	0.9978 ± 0.0004	$0.00087 \pm$ 0.00006	0.0021 ± 0.0002	(73.80 \pm 0.02) ms
v3	0.996 ± 0.001	0.009 ± 0.0001	0.0026 ± 0.0003	(53.50 \pm 0.08) ms

Таблица 4.4 — Финальное сравнение устойчивости к шуму для разных моделей

В дальнейших главах шум создается с вероятностью, равномерно распределенной в интервале от 0 до 0.025.

4.4 СЕГМЕНТАЦИЯ ИСХОДНОГО ИЗОБРАЖЕНИЯ

Сегментация решает задачу классификации пикселей, присваивая каждому пикселю изображения метку, соответствующую определенному классу или области. В дополнение к этому, для распознавания сигналов с детектора FARICH необходимо для каждого пикселя решать задачу регрессии: предсказывать значение скорости (или угла черенковского конуса) частицы, проходящей через данный пиксель, основываясь на форме эллипса, создаваемого этой частицей (если таковой эллипс присутствует). В процессе работы (инференса) нейросети для повышения вычислительной эффективности достаточно рассматривать только те пиксели, через которые

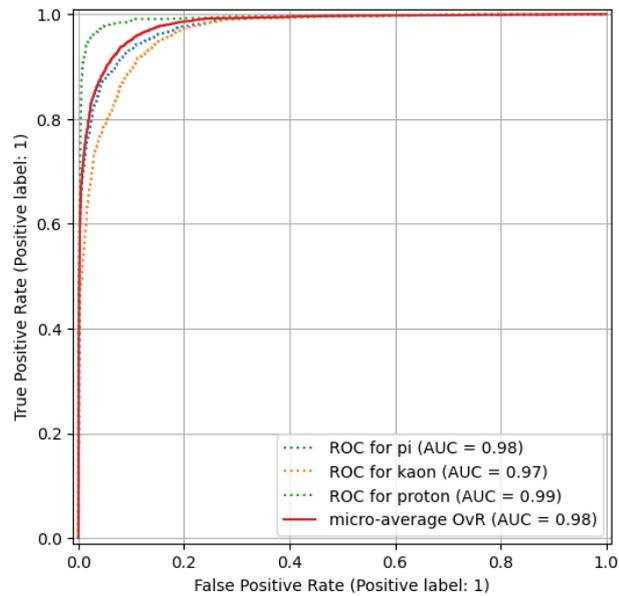


Рисунок 4.8 — ROC AUC для классификации частиц при модели сегментации

проходит реконструированный трек частицы.

Для решения этой задачи была реализована Unet-подобная [11] архитектура, адаптированная для задачи регрессии. Unet, известная своей эффективностью в задачах сегментации, обеспечивает сохранение пространственной информации, что критически важно для точного предсказания скорости.

На вход модели подается 4 канала: сигнал с детектора, два канала для позиционного кодирования, а также канал, в котором в соответствующих пикселях находятся значения импульсов, взятых из трекового детектора.

Результаты работы представлены на рис. 4.8. Видно, что такая модель по качеству на уровне модели с независимой обработкой треков.

Помимо задачи классификации частиц для оценки работы модели или для дополнительного анализа может потребоваться определение принадлежности хитов соответствующим трекам. Для этого для каждого пикселя решается задача бинарной классификации по числу частиц. Бинарная классификация была выбрана по той причине, что фотоны от нескольких частиц могут попасть в один хит. Пример работы нейросети показан на рис. 4.9.

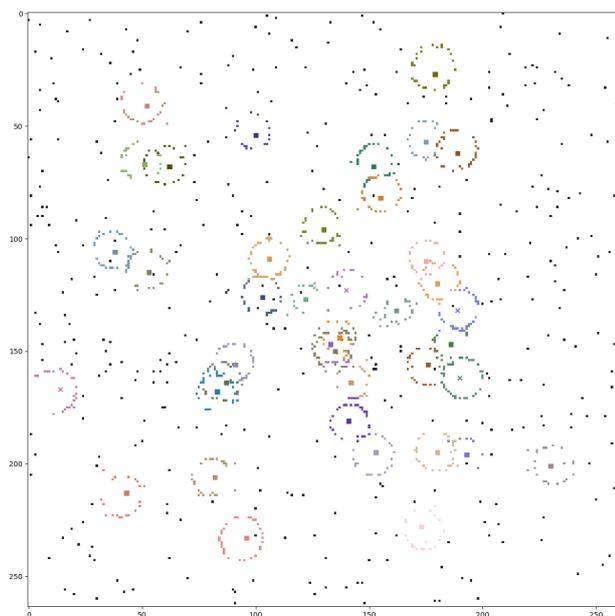


Рисунок 4.9 — Работа нейросети для определения, какому треку принадлежит каждый хит. Черным выделены шумовые хиты, соответствие обозначено одинаковыми цветами у хитов и центров треков. Крестиками обозначены хиты, для которых модель неверно определила частицу, квадратами – хиты, для которых модель верно определила частицу

4.5 ИССЛЕДОВАНИЕ МАСШТАБИРУЕМОСТИ МОДЕЛЕЙ

В этой секции исследуется влияние вычислительной сложности модели на ее эффективность. Для этого варьируются такие параметры, как глубина и ширина модели, а также функция активации.

Варьируемые параметры следующие:

- 1) Глубина: увеличение глубины на единицу соответствует добавлению 10 сверточных слоев.
- 2) Ширина: показывает количество каналов в скрытых слоях.
- 3) Функция активации: функция, обеспечивающая нелинейность в нейронной сети. Оценка вычислительной сложности (от наименьшей к наибольшей): ReLU, Tanh, Sigmoid.

Во время обучения к входным данным добавлялся шум, моделируемый распределением Бернулли, с вероятностью $*p*$ в диапазоне от 0 до 0.025.

На рисунке 4.10 представлены графики, показывающие зависимость

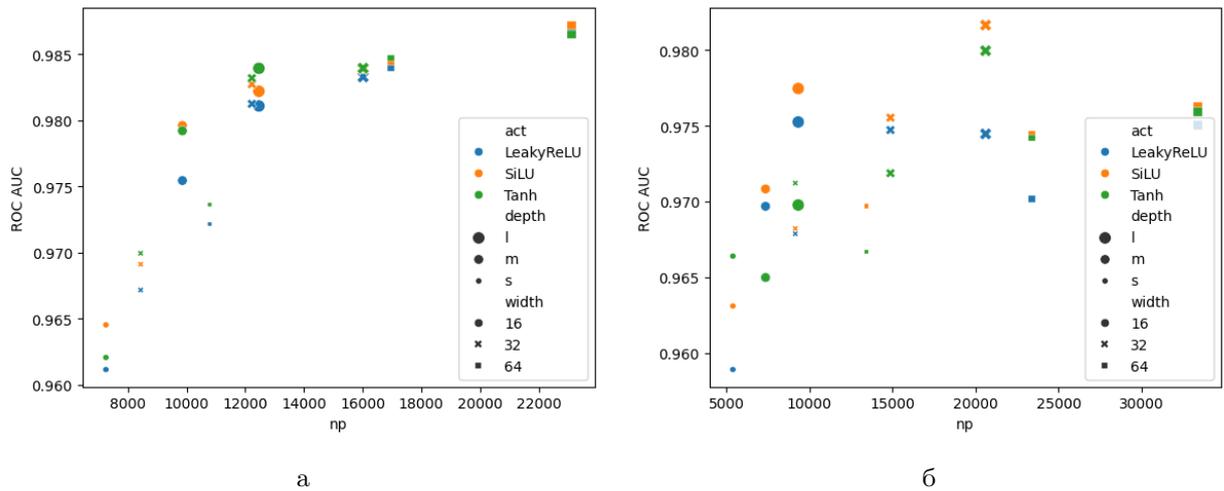


Рисунок 4.10 — Влияние масштабируемости на качество моделей: (а) независимая обработка треков; (б) сегментация.

качества модели от количества параметров. Для модели с независимой обработкой треков наблюдается рост качества с увеличением числа параметров, однако темп роста постепенно снижается. В большинстве случаев наилучшие результаты достигаются при использовании гиперболического тангенса (Tanh) в качестве функции активации.

Качество модели с сегментацией также демонстрирует тенденцию к росту с увеличением числа параметров, особенно в группах моделей с фиксированной шириной. Для этой архитектуры наилучшей функцией активации в большинстве случаев оказалась SiLU.

4.6 ИССЛЕДОВАНИЕ ИНТЕРПРЕТИРУЕМОСТИ МОДЕЛЕЙ

В данной секции исследуется, насколько связаны ее предсказания между собой и насколько им можно доверять. Рассматриваться будут две модели (с независимой обработкой треков и с сегментацией) со следующими параметрами: глубина «m» (в терминах из прошлой главы), ширина 64, функция активации - гиперболический тангенс.

Так как модели (рис. 4.1, рис. 4.7) на последнем слое выдают класс, а из слоев, находящихся ранее, выдают физические параметры, можно посмотреть, насколько эти предсказания связаны. На рис. 4.11 показано, как распределение по предсказанным углам черенковского конуса и скоростям

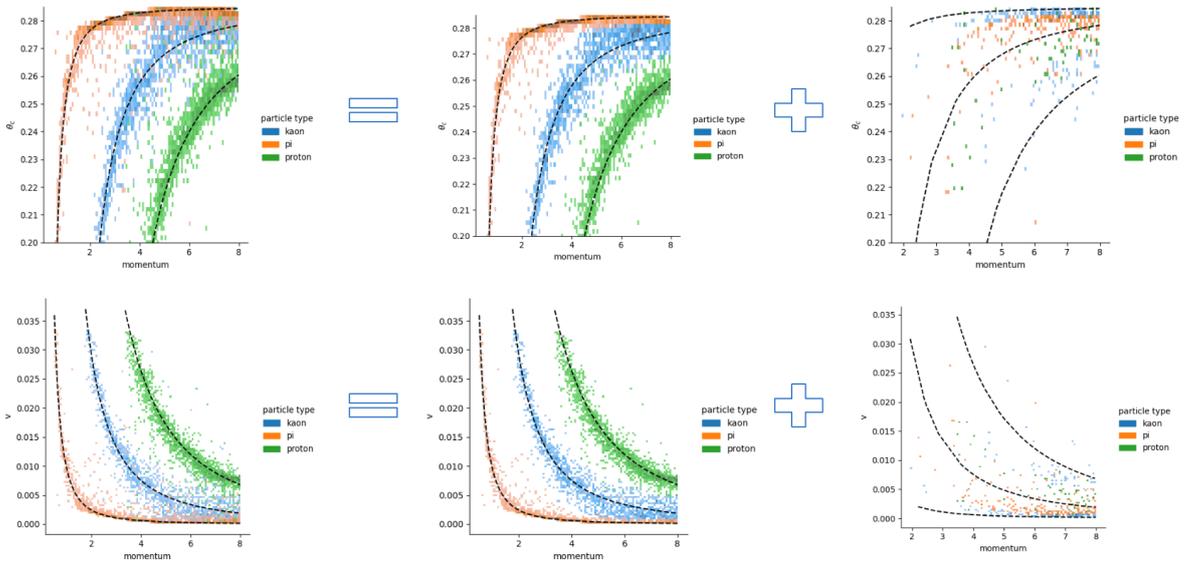


Рисунок 4.11 — Представление распределения по предсказанным углам черенковского конуса и скоростям в виде суммы распределений для верно классифицированных частиц и неверно классифицированных частиц

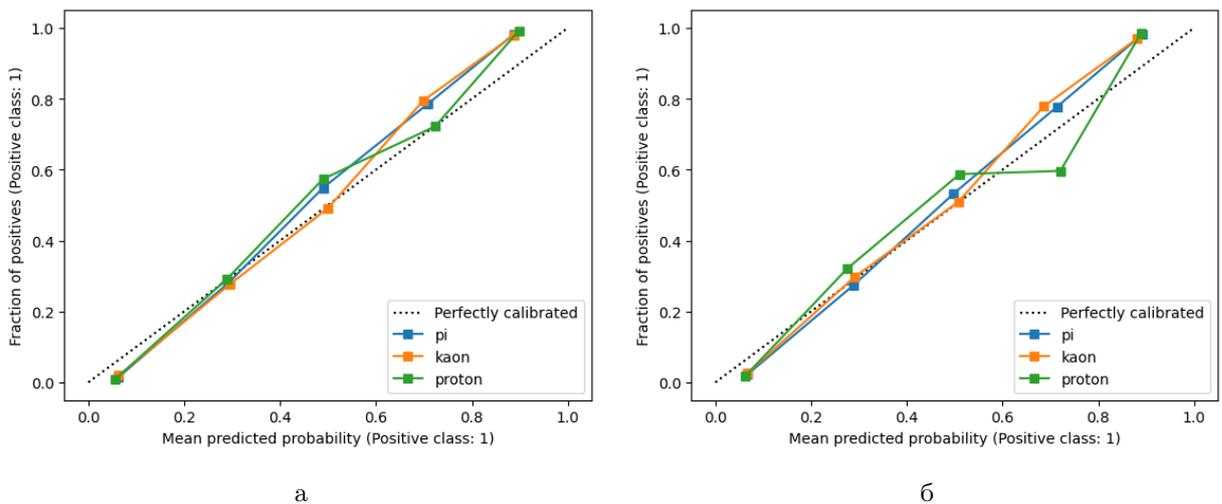


Рисунок 4.12 — Калибровочные кривые для моделей с: (а) независимой обработкой треков; (б) сегментацией.

раскладывается на сумму распределений для верно классифицированных частиц и неверно классифицированных частиц. Видна корреляция между неверной классификацией и неверной регрессией – если модель сильно ошиблась в предсказании скорости и угла черенковского конуса, то, скорее всего, она ошибется и в классификации.

Также, для классификации можно построить калибровочную кривую. Калибровочные кривые, также называемые диаграммами надежности ([12]), сравнивают, насколько хорошо откалиброваны вероятностные пред-

сказания бинарного классификатора. Чем ближе кривая к прямой $y = x$, тем лучше откалиброван классификатор и тем больше вероятности, которые он выдает, соответствуют действительности. На рис. 4.12 представлены калибровочные кривые для модели с независимой обработкой треков (4.12а) и с сегментацией (4.12б).

4.7 ВАРИАНТЫ ОБРАБОТКИ СИГНАЛОВ ДЕТЕКТОРА БЕЗ ИСПОЛЬЗОВАНИЯ ИНФОРМАЦИИ О ТРЕКАХ ЧАСТИЦ

В условиях необходимости быстрой обработки сигналов может быть гораздо эффективнее не ждать информации о треках из другого детектора, а начинать обработку сразу, как только данные пришли из установки. Для этого есть два возможных подхода: сегментация и детекция. Сегментация рассматривалась выше. Отличие будет в том, что на выходе не будет канала с импульсами, а на выходе не будет классификации.

Детекция отличается тем, что модель не предсказывает классы для каждого пикселя, а предсказывает ограничивающие рамки, в которых лежит объект, уверенность в том, что в этих рамках объект действительно есть, а также непосредственно класс объекта в рамке. Из-за того, что размер выхода модели фиксированный, некоторая часть рамок будет пустыми. Для того, чтобы их отсеять, и используется уверенность.

В случае детектора FARICH нет необходимости предсказывать рамку – гораздо важнее предсказать точку предполагаемого пересечения трека с детектором. Таким образом, модель предсказывает для каждого объекта четыре числа: координаты x и y , вероятность, что через точку с такими координатами проходит трек, а также интересующую нас скорость (или угол черенковского конуса).

В процессе обучения для каждого целевого значения берется ближайшее предсказание и для него считается функция ошибки для регрессии. На предсказание уверенности обучаются все предсказания. Для этого введена

эмпирическая формула:

$$conf = \begin{cases} \frac{0.6}{(r-0.75)^2+0.6} & \text{if } r > 0.75 \\ 1 & \text{else} \end{cases} \quad (4.1)$$

Здесь r – расстояние от предсказанной точки до ближайшего трека.

На стадии работы модели точки с низким значением уверенности удаляются, а для близко лежащих точек применяется алгоритм «подавление не максимумов» (non maximum suppression – nms): из всех точек, лежащих внутри некоторого радиуса, выбирается точка с наибольшей уверенностью, а остальные удаляются.

Результат работы модели, а затем алгоритма nms представлены на рис. 4.13

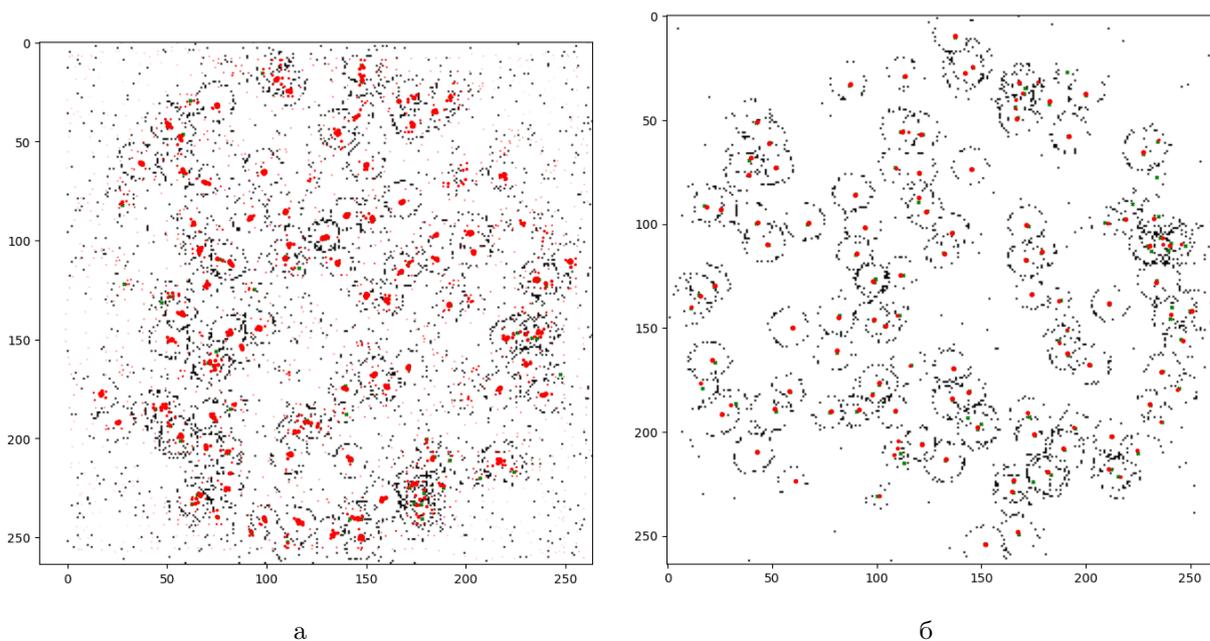
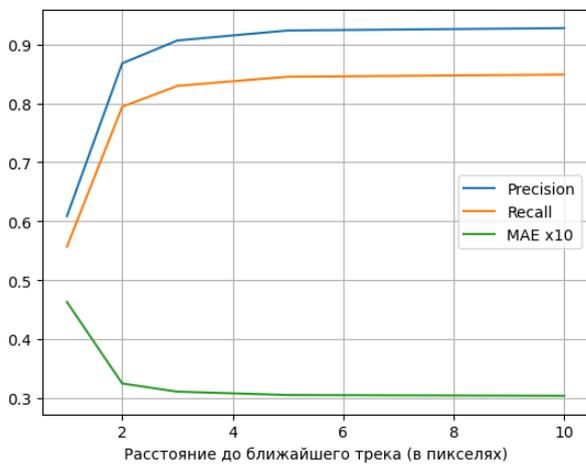
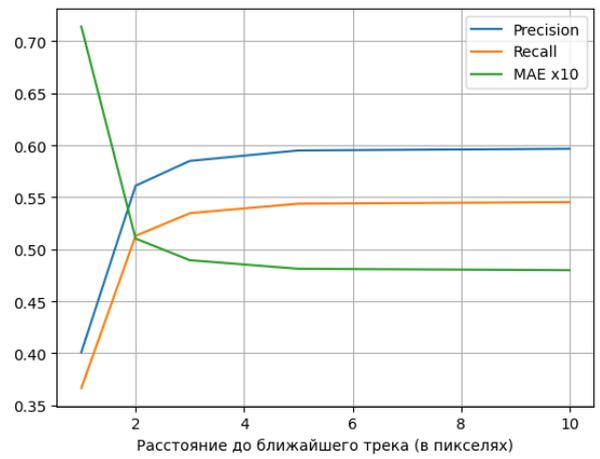


Рисунок 4.13 — Результат работы модели детекции (а) и алгоритма non maximum suppression (б)

Качество детекции и качество детекции при условии, что по найденному углу можно предсказать тип частицы, показано на рис. 4.14. Видно, что такая модель сильно уступает всем предыдущим моделям по качеству определения параметров частицы.



а



б

Рисунок 4.14 — Результат работы модели детекции (а) и алгоритма по maximum suppression (б)

ЗАКЛЮЧЕНИЕ

В ходе исследования была разработана нейросетевая модель для предсказания скорости и типа регистрируемых частиц. Предложен и экспериментально подтвержден метод улучшения характеристик модели, основанный на позиционном кодировании. Кроме того, была исследована устойчивость разработанных моделей к шумовым искажениям, имитирующим реальные условия работы детектора. Для повышения устойчивости к шуму предложена и протестирована архитектура, включающая этап предварительной очистки изображений от шума. Была исследована масштабируемость моделей, а также их интерпретируемость.

Дальнейшие исследования будут направлены на исследование повышения интерпретируемости модели, в том числе с использованием байесовских нейронных сетей, а также сетей Колмогорова-Арнольда (KAN).

Кроме того, планируется разработать нейросеть, которая будет работать без дополнительной экстраполяции трека частицы. Такой подход может быть важен в режиме онлайн-анализа, когда нет времени ждать результатов из другого детектора.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. RICH with multiple aerogel layers of different refractive index / S. Korpar [и др.] // Nucl. Instrum. Meth. A / под ред. J. Engelfried, G. Paic. — 2005. — Т. 553.
2. *Лебедев С. А.* Математическое обеспечение для реконструкции колец черенковского излучения и идентификации электронов в RICH детекторе эксперимента CBM : дис. ... канд. / Лебедев С. А. — Объединённый институт ядерных исследований, 2011.
3. *Hough V P. C.* Method and means for recognizing complex patterns. — 1962.
4. *Chernov N. I., Ososkov G. A.* Effective Algorithms of Circle Fitting // Comput. Phys. Commun. — 1984. — Т. 33. — С. 329—333.
5. *Lebedev S. A., Ososkov G. A.* Fast algorithms for ring recognition and electron identification in the CBM RICH detector // Phys. Part. Nucl. Lett. — 2009. — Т. 6. — С. 161—176.
6. Recent developments in Geant4 / J. Allison [и др.] // Nucl. Instrum. Meth. A. — 2016. — Т. 835. — С. 186—225.
7. Geant4 developments and applications / J. Allison [и др.] // IEEE Transactions on Nuclear Science. — 2006. — Т. 53, № 1. — С. 270—278.
8. GEANT4—a simulation toolkit / S. Agostinelli [и др.] // Nucl. Instrum. Meth. A. — 2003. — Т. 506. — С. 250—303.
9. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation / C. R. Qi [и др.]. — 2017. — arXiv: 1612.00593 [cs.CV].

10. *LeCun Y., Bengio Y.* Convolutional networks for images, speech, and time series // The Handbook of Brain Theory and Neural Networks. — Cambridge, MA, USA : MIT Press, 1998. — C. 255—258. — ISBN 0262511029.
11. *Ronneberger O., Fischer P., Brox T.* U-Net: Convolutional Networks for Biomedical Image Segmentation. — 2015. — arXiv: 1505.04597 [cs.CV].
12. *Wilks D., Wolfe D.* Optimal use and economic value of weather forecasts for lettuce irrigation in a humid climate // Agricultural and Forest Meteorology. — 1998. — T. 89, № 2. — C. 115—129. — ISSN 0168-1923.