

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

ИНСТИТУТ ЯДЕРНОЙ ФИЗИКИ И ТЕХНОЛОГИЙ
КАФЕДРА №40 «ФИЗИКА ЭЛЕМЕНТАРНЫХ ЧАСТИЦ»

УДК 53.01

**ПОЯСНИТЕЛЬНАЯ ЗАПИСКА
К БАКАЛАВРСКОЙ ДИПЛОМНОЙ РАБОТЕ**

**ПРОСТРАНСТВЕННАЯ РЕКОНСТРУКЦИЯ X-Y
КООРДИНАТ СОБЫТИЙ В ДЕТЕКТОРЕ DARKSIDE-20K**

Студент _____ Н. А. Полякова

Научный руководитель,

к.ф.-м.н.

_____ А. В. Гробов

Москва 2021

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА БАКАЛАВРА

**ПРОСТРАНСТВЕННАЯ РЕКОНСТРУКЦИЯ X-Y
КООРДИНАТ СОБЫТИЙ В ДЕТЕКТОРЕ DARKSIDE-20K**

Студент	_____ Н. А. Полякова
Научный руководитель, к.ф.-м.н.	_____ А. В. Гробов
Рецензент, д. ест. н., проф.	_____ А. Киш
Секретарь ГЭК, к.ф.-м.н.	_____ А. А. Кириллов
Зав. каф. №40, д.ф.-м.н., проф.	_____ М. Д. Скорохватов

РЕФЕРАТ

Отчет 38 с., 42 рис., 12 источн.

КЛЮЧЕВЫЕ СЛОВА

Темная материя, WIMP - частица, эксперимент DarkSide-20k, машинное обучение, сверточные нейронные сети.

ЦЕЛЬ РАБОТЫ

Изучение принципов работы детектора DarkSide-20k, обработка данных с применением машинного обучения и сверточных нейронных сетей при помощи специальных библиотек языка программирования Python. Реконструкция положения событий в детекторе. Оценка максимального числа вышедших из строя ФЭУ детектора без потери точности реконструкции.

ИНСТРУМЕНТЫ, ИСПОЛЬЗУЕМЫЕ В РАБОТЕ

- 1) Python - язык программирования
- 2) Jupyter Notebook - интерактивная оболочка для Python
- 3) Google colab - веб-оболочка для Python
- 4) G4DS - модификация geant4 для Монте-Карло моделирования процессов для DarkSide

РЕЗУЛЬТАТЫ

- 1) Получены методом Монте-Карло данные, на основе конструкции детектора.
- 2) Разработана сверточная нейронная сеть, реконструирующая положения событий в плоскости X-Y.
- 3) Оптимизированы параметры нейронной сети для получения необходимой точности (≈ 2 см).
- 4) Оценено максимальное число SiPM, которые могут выйти из строя.

СОДЕРЖАНИЕ

Введение	4
1 Теоретическая часть	6
1.1 Благородные газы. Жидкий аргон	6
1.2 Эксперимент DarkSide	7
1.2.1 Детектор DarkSide-50	8
1.2.2 DarkSide-20k	11
1.3 Монте-Карло моделирование	13
2 Практическая часть	14
2.1 Реконструкция событий	14
2.1.1 Сверточные нейронные сети	15
2.1.2 Входные данные и алгоритм реконструкции	16
2.2 Определение оптимальных параметров для нейронной сети	18
2.3 Обработка данных	22
2.4 Реорганизация данных	28
2.5 Определение максимального числа SiPM, которые могут выйти из строя	31
2.6 Создание собственных Monte-Carlo данных, при помощи G4DS	34
2.7 Заключение	37
Список литературы	38

ВВЕДЕНИЕ

При расчете скорости движения звезд на границах галактики относительно ее центра, согласно законам Кеплера, она должна уменьшаться. Однако в действительности этого не происходит, скорость на достаточно больших расстояниях становится практически неизменной, что противоречит предположениям ньютоновской механики, поскольку плотность видимого вещества уменьшается при отдалении от центра галактики. Данный эффект может быть объяснен наличием «невидимой», или как ее позже назвали, темной материей, которая обеспечивает постоянство скорости вращения звезд в эллиптических галактиках при удалении от центра [1].

Таким образом, один из самых актуальных вопросов современной физики – это природа темной материи.

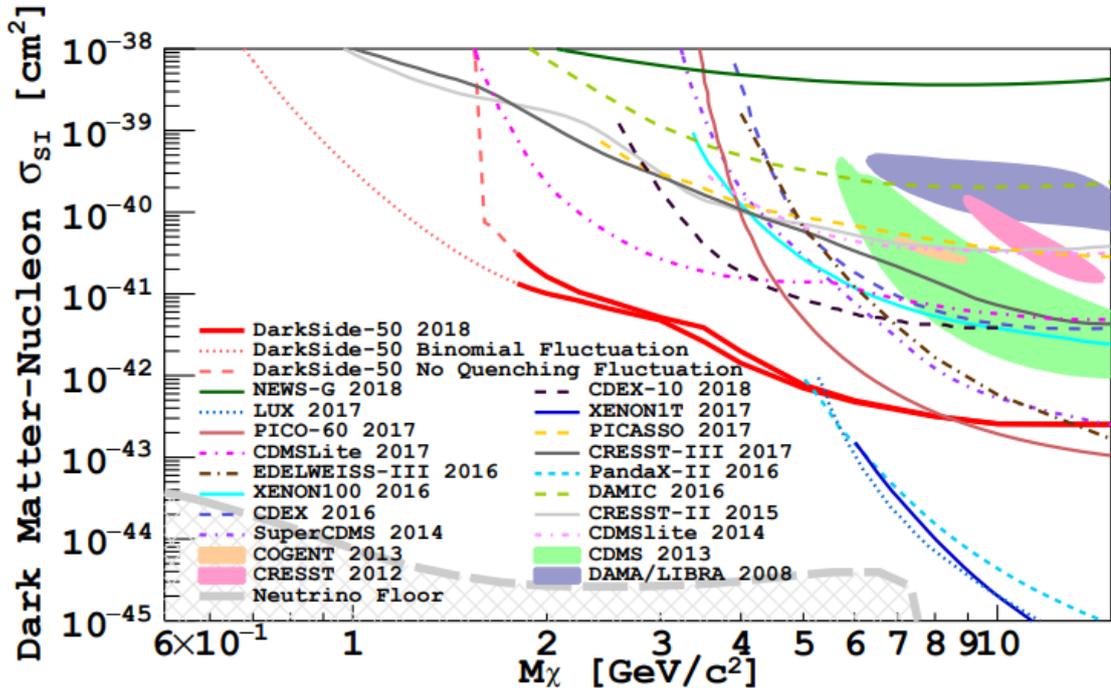


Рисунок 1 — Ограничение на зависимость сечения взаимодействия частиц темной материи от их массы. Текущие и предположительные результаты прямых экспериментов по поиску темной материи. [2].

Считается, что примерно 85% вещества во Вселенной находится в некоторой небарионной форме, которая не излучает и не поглощает элек-

тромагнитное излучение [3]. Существует множество частиц-кандидатов на роль темной материи, например, такими частицами являются так называемые вимпы – Weakly Interacting Massive Particles (WIMPs), – слабо взаимодействующие массивные частицы. Соответственно, эти частицы слабо взаимодействуют с обычным барионным веществом (Рисунок 1 – зависимость сечения взаимодействия от массы частицы темной материи), поэтому для их обнаружения используются детекторы, имеющие низкий фон и обладающие отличными свойствами для их регистрации. Например, детекторы на жидких благородных газах.

1. ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

1.1. БЛАГОРОДНЫЕ ГАЗЫ. ЖИДКИЙ АРГОН

Жидкие благородные газы, чаще всего используются аргон Ar и ксенон Xe, реже неон Ne, являются распространенными мишенями для детектирования вимпов благодаря ряду причин.

- 1) Простота эксплуатации — эти газы можно довести до сжиженного состояния азотным охлаждением.
- 2) Аргон и ксенон являются яркими сцинтилляторами и они прозрачны для собственного сцинтилляционного света.
- 3) В отличие от детекторов на основе кристаллов, детекторы на жидких благородных газах могут быть увеличены в масштабе в связи с доступностью аргона и ксенона.

Существует два типа аргона, отличающихся происхождением: атмосферный аргон AAr и подземный аргон UAr. Аргон, полученный из атмосферы, имеет в составе нестабильные изотопы - ^{37}Ar , ^{39}Ar и ^{42}Ar . Основной радиоактивный изотоп — это ^{39}Ar , он является самым активным из трех благодаря ряду причин, таких как время жизни и концентрация. Согласно исследованиям, его активность составляет 1 Бк/кг для AAr [4]. Для UAr активность составляет 0.73 мБк/кг. Поэтому в дальнейших экспериментах DarkSide планируется использовать именно подземный аргон, для минимизации фона от ^{39}Ar [5].

Принцип работы детекторов на жидком аргоне будет подробно описан в главе про детекторы 1.2.1.

1.2. ЭКСПЕРИМЕНТ DARKSIDE

Проект DarkSide - это обширная программа прямого поиска темной материи, которая ориентирована на серию экспериментов на основе жидкого аргона, но также включает в себя дополнительные проекты для разработки и оптимизации новых технологий, необходимых для повышения эффективности работы используемых детекторов [6].

Первым детектором, созданным в 2010 г. под руководством коллаборации стал DarkSide-10, объем активного вещества которого составлял 10 кг. Позднее в 2015 г. на его основе был создан детектор DarkSide-50, содержащий 50 кг сначала атмосферного аргона $A\text{Ar}$, а после его заполнили подземным аргоном $U\text{Ar}$. На данный момент в разработке находится детектор DarkSide-20k, который будет содержать 20 тонн жидкого аргона в чувствительном объеме.

Работая в двухфазном режиме, такие детекторы позволяют измерять как сцинтилляционный свет, так и ионизацию. В связи с тем, что взаимодействия с WIMP крайне редки, необходимо минимизировать всевозможные фоны.

В аргоновых детекторах важную роль играют следующие два детектируемых сигнала: $S1$ и $S2$, сцинтилляции, полученные, в первом случае, от возбужденных димеров аргона, а во втором от электронов.

В эксперименте DarkSide электрическое поле направлено вдоль оси Z детектора, и поэтому глубина события может быть с большой точностью определена путем измерения времени дрейфа — временного промежутка между сцинтилляционными сигналами $S1$ и $S2$ в жидкой и газообразной фазе. Что же касается координат X и Y , то это является уже более трудоемкой задачей.

Решением этой задачи может послужить алгоритм с использованием сверточных нейронных сетей. Для этого необходимо проанализировать данные, получаемые в ходе эксперимента, и понять как их можно использовать для реконструкции положения события в плоскости X и Y . Затем необходимо смоделировать данные методом Монте-Карло, однако с уже заранее известными координатами $x - y$ и на их основе подготовить наборы данных для обучения нейронной сети.

1.2.1. ДЕТЕКТОР DARKSIDE-50

Эксперимент DarkSide-50 (DS-50) находится в Национальной лаборатории Гран-Сассо(LNGS) в Италии, на глубине 3800 м [7]. Это последний работающий детектор коллаборации DarkSide, сейчас его работа приостановлена. На основе результатов работы DarkSide-50 [8] планируется запуск нового детектора DarkSide-20k.

DS-50 состоит из трех вложенных друг в друга детекторов (Рисунок 1.1). В направлении от центра к краям расположены три детектора: время-проекционная камера с жидким аргоном (TPC), которая является детектором темной материи, вето с жидким сцинтиллятором (LSV), служащее защитой от радиогенных и космогенных нейтронов, γ -лучей и космических мюонов и водный черенковский детектор (WCD), служащий защитой от космических мюонов.

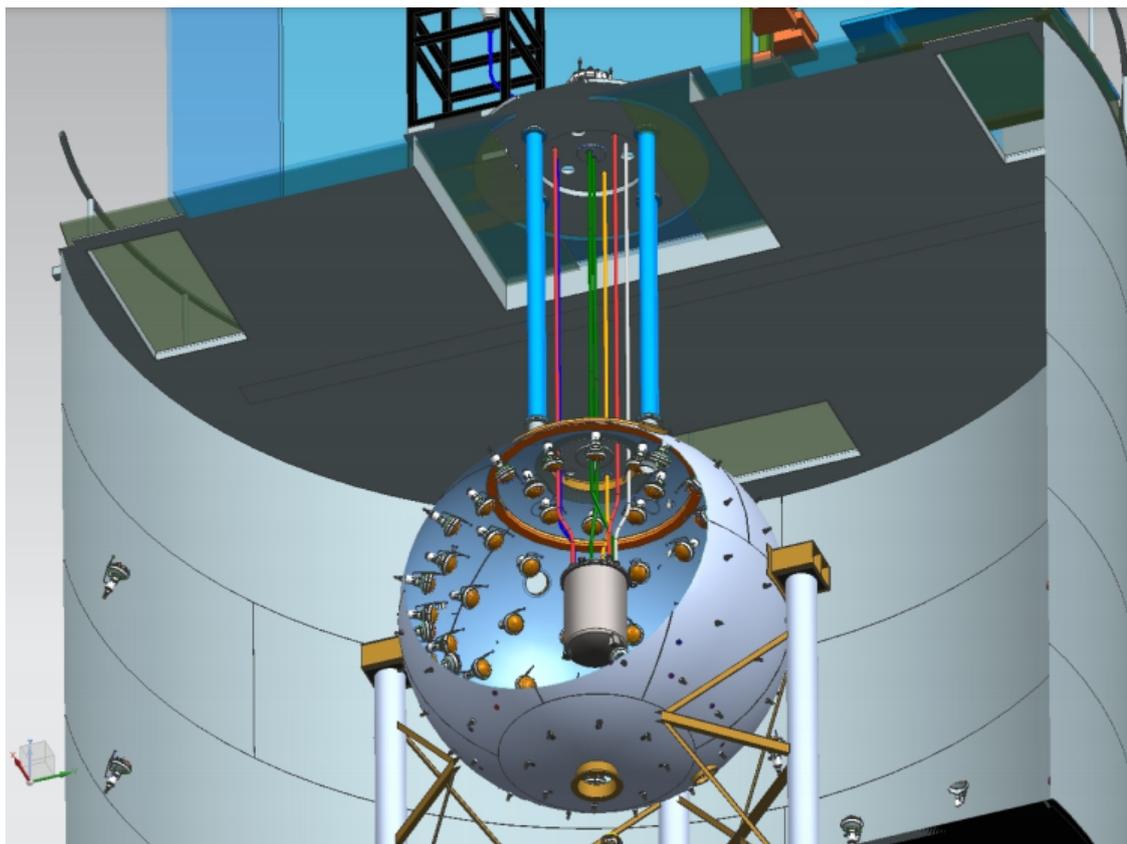
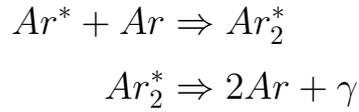


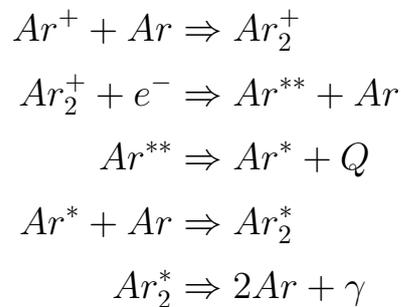
Рисунок 1.1 — Устройство DarkSide-50 [9].

При взаимодействии частицы с ядрами жидкого аргона LAr, происходит выделение энергии несколькими путями: сцинтилляцией, ионизацией

и в виде тепловой энергии. Возбужденные атомы аргона способны создавать слабосвязанные димеры, которые при распаде испускают γ -кванты с длиной волны 128 нм и происходит сцинтилляция.



В свою очередь ионизированные атомы аргона рекомбинируют и передают часть своей энергии близлежащим атомам посредством теплового движения.



Отношение числа ионизированных атомов к числу сцинтилляций, зависит от многих факторов и в общем случае меньше единицы. Например, при наличии электрического поля направленного вдоль ТРС это отношение уменьшается с увеличением напряженности поля, т.к. свободные электроны будут дрейфовать от катода к аноду и тем самым вероятность рекомбинации будет уменьшаться.

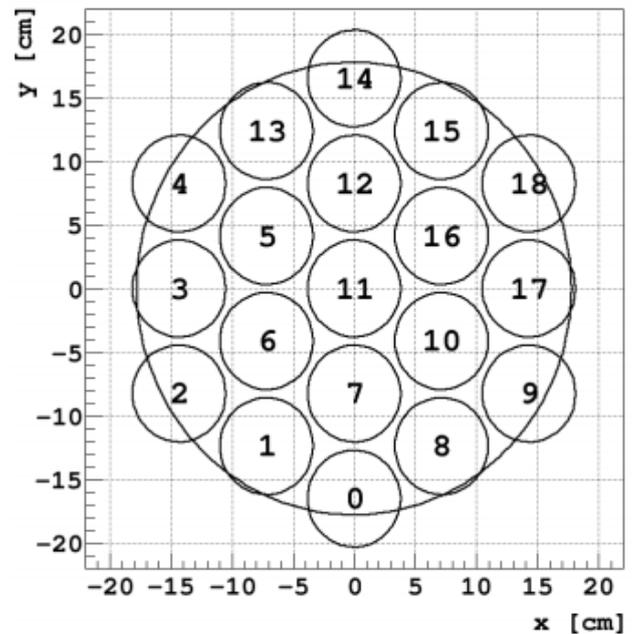


Рисунок 1.2 — Расположение фотомножительных трубок в массивах DS-50.

Схематичное изображение время-проекционной камеры представлено на рисунке 1.3.

ТРС состоит из двух массивов фотоумножительных трубок, расположенных сверху и снизу емкости камеры. Вдоль цилиндрической стенки располагаются медные кольца, которые находясь под напряжением создают электрическое поле, направленное по оси z детектора, которое и заставляет дрейфовать свободные электроны. Камера заполнена не только жидким, но и газообразным аргоном. Над фазовой границей на расстоянии 5 мм располагается экстракционная сетка, на которую подается напряжение -5.6 кВ.

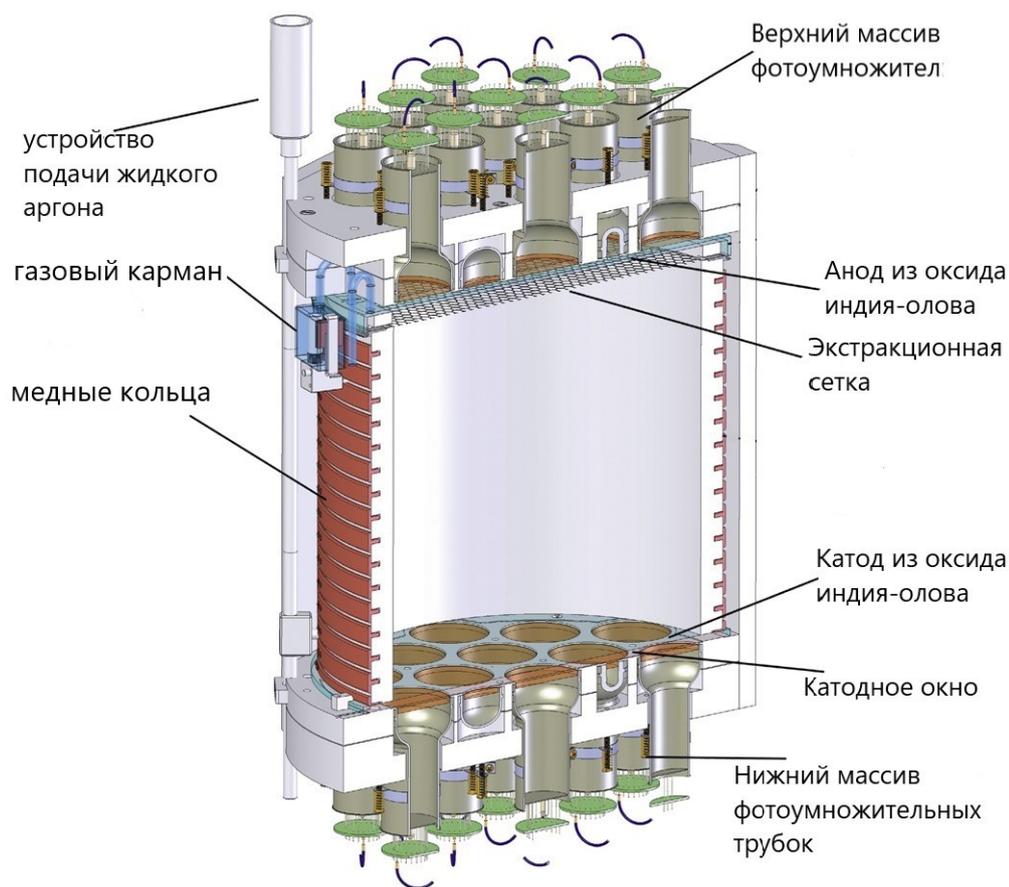


Рисунок 1.3 — Устройство двухфазной время-проекционной камера (ТРС) на примере эксперимента DarkSide-50. [9].

1.2.2. DARKSIDE-20K

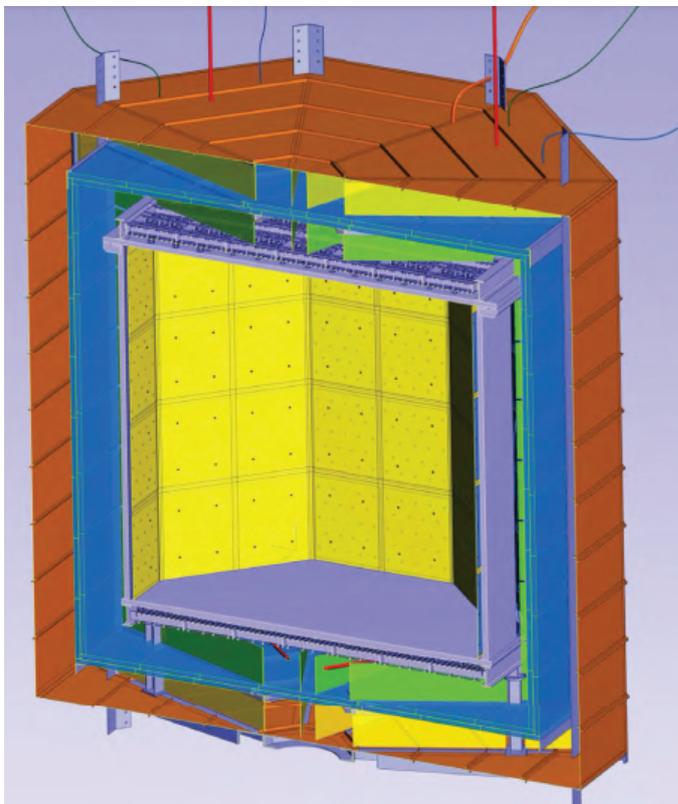


Рисунок 1.4 — Вид DS-20k вето детектора. Двухфазная время-проекционная камера (TPC) эксперимента DarkSide-20k изображена сиреневым цветом. Клетка Фарадея коричневым. [10]

более высокую эффективность регистрации фотонов (PDE) и лучшее разрешение по одному фотону, при этом они работают при более низком напряжении смещения.

Кремниевые фотоумножители будут располагаться сверху и снизу рабочего объема детектора (Рисунок 1.4). SiPM будут собраны в фотодетектирующие модули (PDU) по 24 штуки, а из модулей будут складываться так называемые пады, плитки, содержащие по 25 PDU, если это квадратная плита и по 15, если это треугольная плита (Рисунок 1.5 и 1.6). Всего детектор будет содержать 8240 SiPM, по 4120 соответственно на нижний и верхний модуль, состоящие из падов.

Детектор DarkSide-20k находится в стадии разработки. Планируется, что его первый запуск произойдет в 2022 г. В отличие от предыдущей версии на 50 кг, детектор будет содержать кремниевые фотоумножители (SiPMs), вместо ранее использовавшихся фотоумножительных трубок (PMTs) и объем рабочего вещества будет содержать уже 20 тонн жидкого аргона.

Кремниевые фотоумножители (SiPM) - одна из ключевых технологий, эффективно работающих при проведении крупномасштабных экспериментов с темной материей на основе LAr. Они имеют ряд преимуществ в эффективности по сравнению с традиционными PMT, включая

Принцип работы детектора претерпит небольшие изменения. Он так же будет содержать 3 детектора: время-проекционную камеру с жидким аргоном (TPC), но veto с жидким сцинтиллятором (LSV) и ранее водный черенковский детектор (WCD) будут заполнены ≈ 700 тонн атмосферного аргона. Изменится также его размер и форма TPC, с цилиндрической на восьмиугольную.

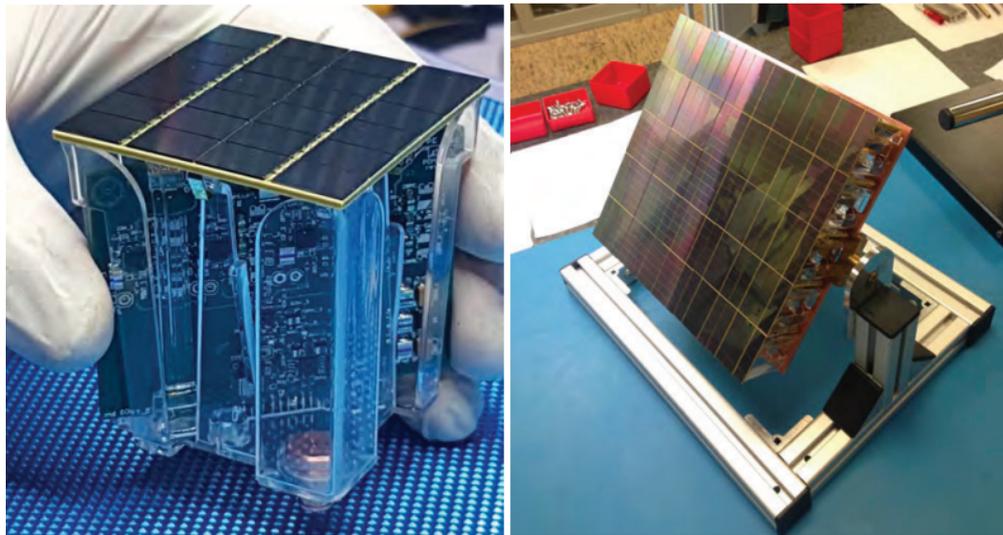


Рисунок 1.5 — Слева расположен фотодетектирующий модуль (PDM), состоящий из 24 SiPM. Справа расположен фотодетектирующий пад, состоящий из 25 PDM. [10]

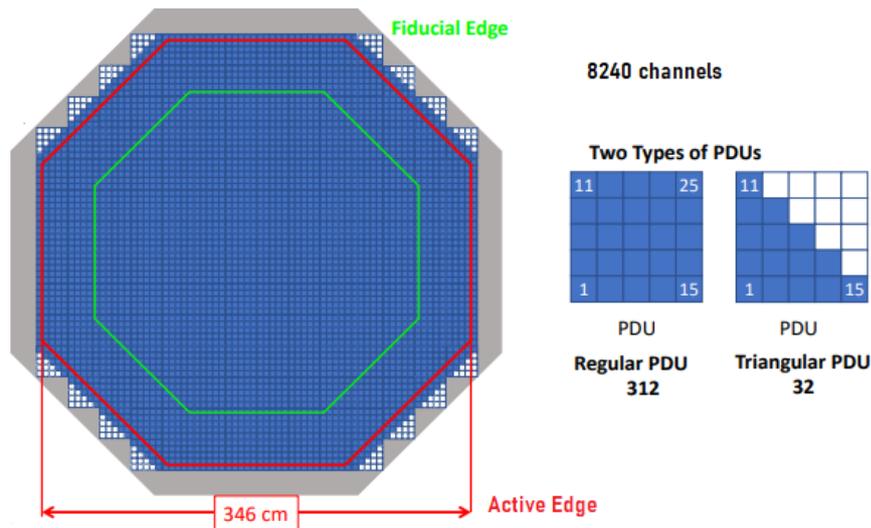


Рисунок 1.6 — Расположение кремниевых фотоумножителей

1.3. МОНТЕ-КАРЛО МОДЕЛИРОВАНИЕ

Моделирование событий происходит при помощи специального пакета программ G4DS.

G4DS - это набор инструментов моделирования на основе Geant4, специально разработанный для DarkSide. Она была разработана для описания энергетических и временных характеристик всех детекторов, принадлежащих программе DarkSide, а именно DarkSide-10, DarkSide-50 и DarkSide-20k. Для каждого из них G4DS предоставляет богатый набор генераторов частиц, детализированную геометрию, физические процессы и полное оптическое распространение фотонов, создаваемых сцинтилляцией в жидком аргоне и электролюминесценцией в газообразном аргоне.

Основными целями G4DS являются: точное описание энергетических сигналов S1 и S2 и формы импульса, выраженного переменной f_{90} .

G4DS отслеживает фотоны вплоть до их преобразования в фотоэлектроны, которое происходит, когда фотон достигает активной области ФЭУ. Электронное моделирование также обеспечивает возможность перекрывать моделируемые события и реальные исходные данные, чтобы обеспечить более реалистичное моделирование. На выходе программа создает формы сигналов для каждого канала с тем же форматом данных, что и реальные данные.

2. ПРАКТИЧЕСКАЯ ЧАСТЬ

2.1. РЕКОНСТРУКЦИЯ СОБЫТИЙ

Роль реконструкции состоит в том, чтобы получить из необработанных данных сокращенный набор физических параметров, необходимый для физического анализа.

Реконструкция объединяет информацию от ТРС и вето детекторов. Типичный алгоритм реконструкции принимает один или несколько наборов информации в качестве входных данных, вызывает набор модульных инструментов и выводит один или несколько наборов реконструированных данных. Удобно для реконструкции использовать сверточные нейронные сети.

В данной работе реконструкция проводится следующим образом. Сначала происходит обработка данных, в тот вид, на котором нейронной сети будет проще обучаться. В данном случае к рассмотрению берутся две величины: pe_pmt — число срабатываний конкретных фотоумножителей для каждого события, и npe — общее число фотоэлектронов, полученное за событие.

Из этих величин формируется таблица, где каждая строка соответствует одному событию, а столбец конкретному номеру ФЭУ, и на выходе получается таблица, где каждой ячейке соответствуют некоторые числа фотоэлектронов, отнормированные на число фотоэлектронов за событие.

У каждого события заведомо известна пара координат x и y . Нейросеть учится сопоставлять данные из таблицы и координаты, чтобы в дальнейшем предсказывать значения координат.

2.1.1. СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ

Сверточные нейронные сети (СНС) — Convolutional Neural Networks — это мощный класс искусственных нейронных сетей, которые очень хорошо справляются с проблемами восстановления изображений.

Принцип работы этих сетей основан на обработке входных данных несколькими слоями, в частности обработки изображений, каждый из которых состоит из линейного оператора и действует как свертка, за которым следует нелинейное преобразование с некоторой функцией активации.

Каждый фрагмент входного изображения умножается на матрицу (ядро) свертки поэлементно, а результат суммируется и записывается в аналогичную позицию выходного изображения [11].

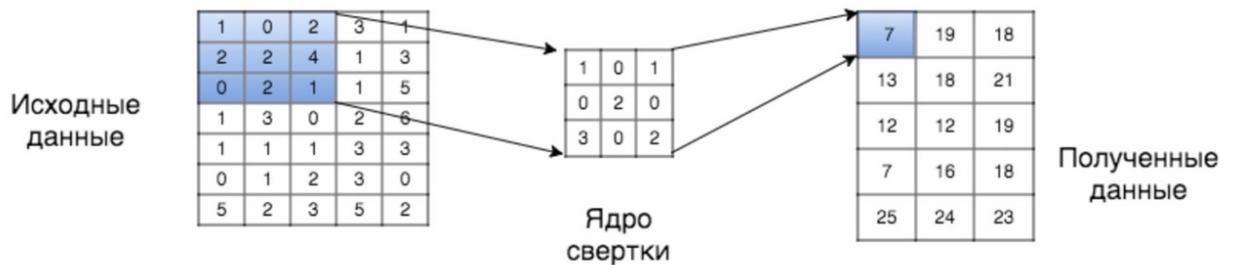


Рисунок 2.1 — Пример работы сверточного слоя.

Форма вывода у СНС выражается как:

$$y = F(\Theta, x) = f_n(W_n f_{n-1}(\dots(f_2(W_2 f_1(W_1 x + b_1) + b_2)\dots) + b_n), \quad (2.1)$$

где x — это входные данные, а y выходные, W_i — матрица свертки i -го слоя, b_i — смещение i -го слоя свертки, f_i — нелинейная функция, Θ — набор всех настраиваемых параметров.

Основная цель работы СНС состоит в том, чтобы найти оптимальный набор параметров Θ и количество эпох K и в последующем минимизировать потери, выражающиеся следующей формулой:

$$\sum_{k=1}^K L(y_k, F(\Theta, x_k)), \quad (2.2)$$

где $L(y_k, F(\Theta, x_k))$ — функция потерь для k -ой эпохи.

2.1.2. ВХОДНЫЕ ДАННЫЕ И АЛГОРИТМ РЕКОНСТРУКЦИИ



Рисунок 2.2 — Алгоритм реконструкции событий в виде блок-схемы.

Получить результат предсказания, максимально удовлетворяющий заданным критериям.

Пятый этап — использование обученной и настроенной нейронной сети для реконструкции координат.

В данной работе присутствовало три типа данных: первый — Монте-Карло данные предоставленные коллаборацией DarkSide в виде, экспериментальном или сыром, второй — Монте-Карло данные, также предоставленные коллаборацией, но прошедшие предварительный отбор и представленные в виде таблицы, и третий — самостоятельно смоделированные Монте-Карло данные (Рисунок 2.3). Все данные содержали информацию об верхнем массиве фотоумножителей, где события были случайным образом распределены по площади детектора.

Реконструкция представляет собой последовательность действий, представленных на Рисунке 2.2.

Первым и самым важным этапом является сбор данных, от их качества зависит весь дальнейший анализ. Это должна быть простая система, потому что нейронной сети легче обучаться на обработанной, например некоторым физическим анализом исключая различные, группы данных.

Следовательно, вторым этапом происходит организация данных в необходимый "упрощенный" вид.

Третий этап — это обучение нейронной сети.

Четвертый этап — подбор параметров таким образом, чтобы получить результат предсказания, максимально удовлетворяющий заданным критериям.

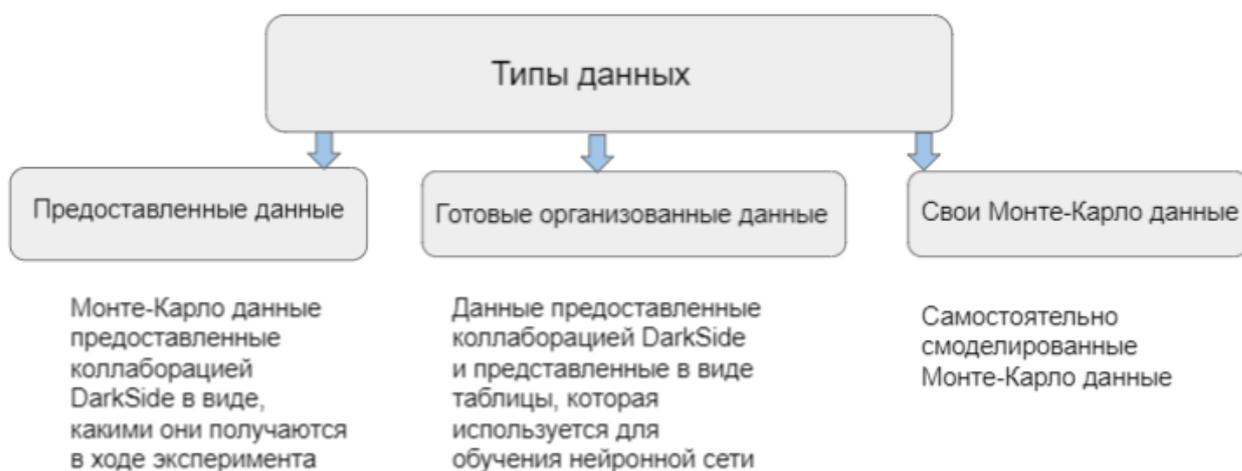


Рисунок 2.3 — Типы данных, использовавшихся в работе.

Для реконструкции были выбраны сверточные нейронные сети, в частности библиотеки *Keras* программного языка *Python* [12]. Схема работы нейронной сети, использовавшейся в исследовании, представлена на Рисунке 2.4. Всего, как упоминалось ранее (Рисунок 1.6), DS-20k содержит 4120 каналов кремниевых фотосенсоров, поэтому первый слой содержит такое же число нейронов, далее число уменьшается, пока не дойдет до 2 значений. В качестве выходных данных нейронная сеть предсказывает X-Y значения координат для одного события.



Рисунок 2.4 — Схема работы нейронной сети.

2.2. ОПРЕДЕЛЕНИЕ ОПТИМАЛЬНЫХ ПАРАМЕТРОВ ДЛЯ НЕЙРОННОЙ СЕТИ

Методом Монте-Карло было смоделировано более 200 тысяч событий для детектора DarkSide-20k, эти данные были предоставлены коллаборацией DarkSide и заранее подготовлены для обучения нейронной сети. Для каждого события заранее были известны каналы с максимальным значением фотоэлектронов в верхней части детектора, а также координаты X и Y . В предположении, что они будут использованы для обучения сверточной нейронной сети, чтобы потом применять ее к реальным данным.

Для реконструкции событий было принято решение разработать сверточную нейронную сеть, состоящую из нескольких слоев, которая должна обучаться на 90 процентах данных и тестироваться на оставшихся 10 процентах.

В связи с большим количеством данных, из-за физических ограничений на оперативную память, анализ проводился частями, по 20 тысяч событий за раз на онлайн-площадке Google Colab.

Точность предсказанных нейросетью значений в большой степени зависит от размера выборки и количества эпох обучения. Достаточно быстро был определен оптимальный размер выборки, он составил 10 для 20 тысяч событий, а с количеством эпох все было не так однозначно. Потому что график потерь в зависимости от количества эпох выходит на плато после 10-12 эпохи (Рисунок 2.6а и 2.6б), однако реконструкция для 12 и даже 20 эпох оказалась в разы менее точной, чем ожидалось в сравнении с 30 эпохами (Рисунок 2.7 и рисунок 2.8).

Для оценки правильности реконструкции положения событий в пространстве были построены графики для каждой переменной x и y зависимости предсказанного значения и реального (Рисунок 2.5). Для каждого из графиков был определен угловой коэффициент методом наименьших квадратов.

Таблица 2.1 — Полученные средние ошибки для координат

	x	y
Угловой коэфф-т, α	1.0024 ± 0.0003	1.0063 ± 0.0003

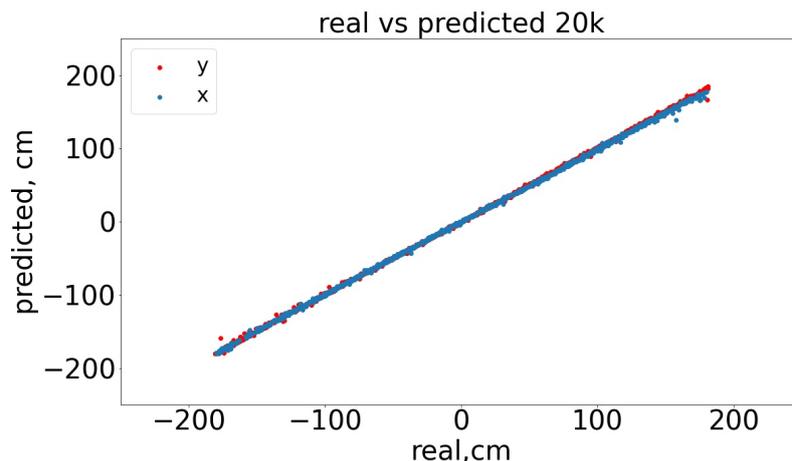
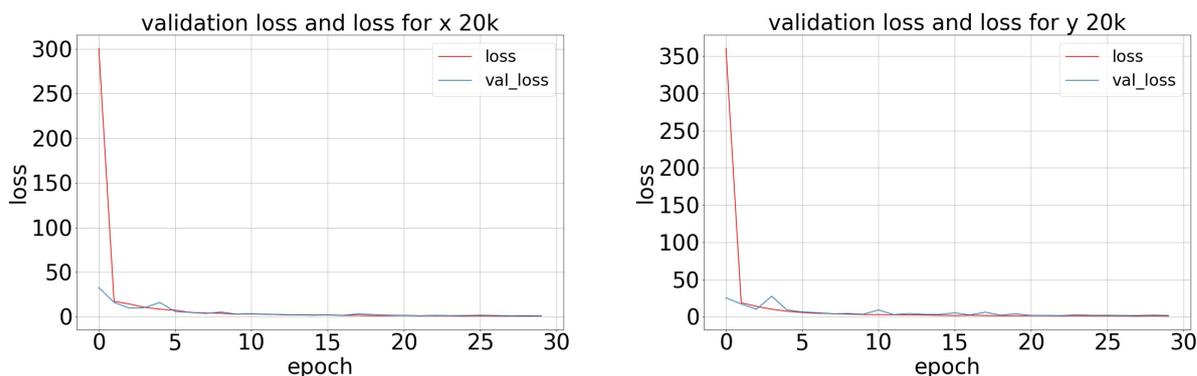


Рисунок 2.5 — График зависимости предсказанного значения от реального для x и y.

Несмотря на точность, которая возрастает с увеличением числа эпох, время на обучение тратится пропорционально их количеству, а значит с каждым разом все больше. Следовательно, необходимо было найти такое число эпох, чтобы обучение занимало приемлемое время, а точность при этом не была потеряна. Для 50 эпох разность между реальным и предсказанным значением координат минимальная из полученных, а на рисунке 2.76 для 20 эпох эта величина имеет большой разброс. Но если посмотреть на рисунок 2.8а для 30 эпох разность принимает относительно адекватное значение, и время для обучения также оптимально.

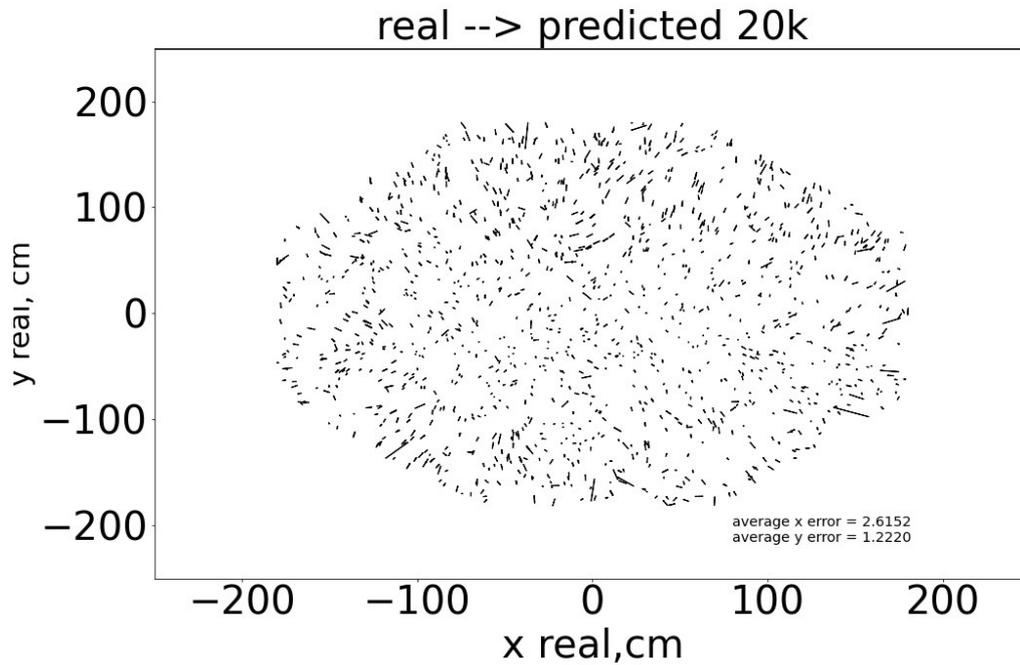


(а) График validation loss и loss для x координаты 30 эпох. (б) График validation loss и loss для y координаты 30 эпох.

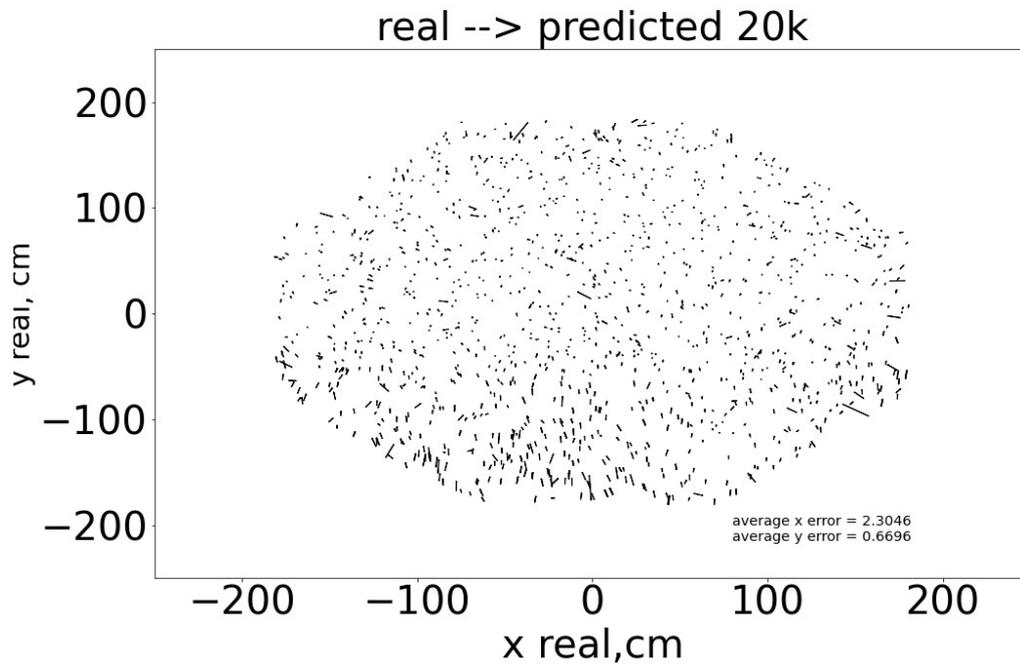
Рисунок 2.6 — Основные графики и гистограммы для 30 эпох.

Таким образом, оптимальное число эпох, при котором реконструкция нейронной сети является точной ($\lesssim 2$ см) и при этом не занимает много

времени – это 30 эпох.

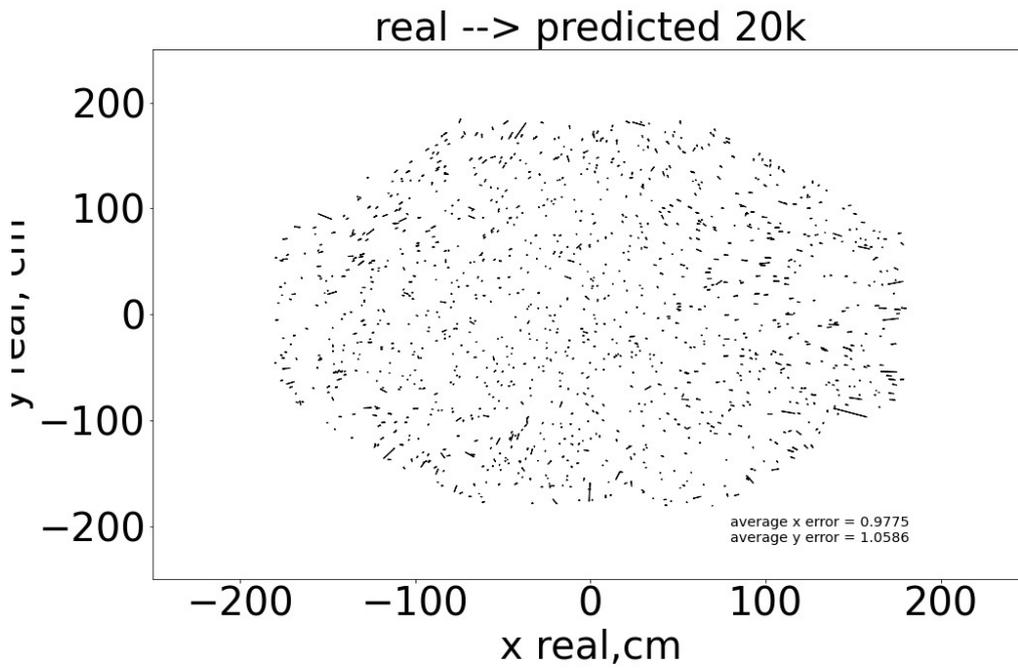


(а) Распределения разности предсказанного и реального значений координат. 12 эпох.

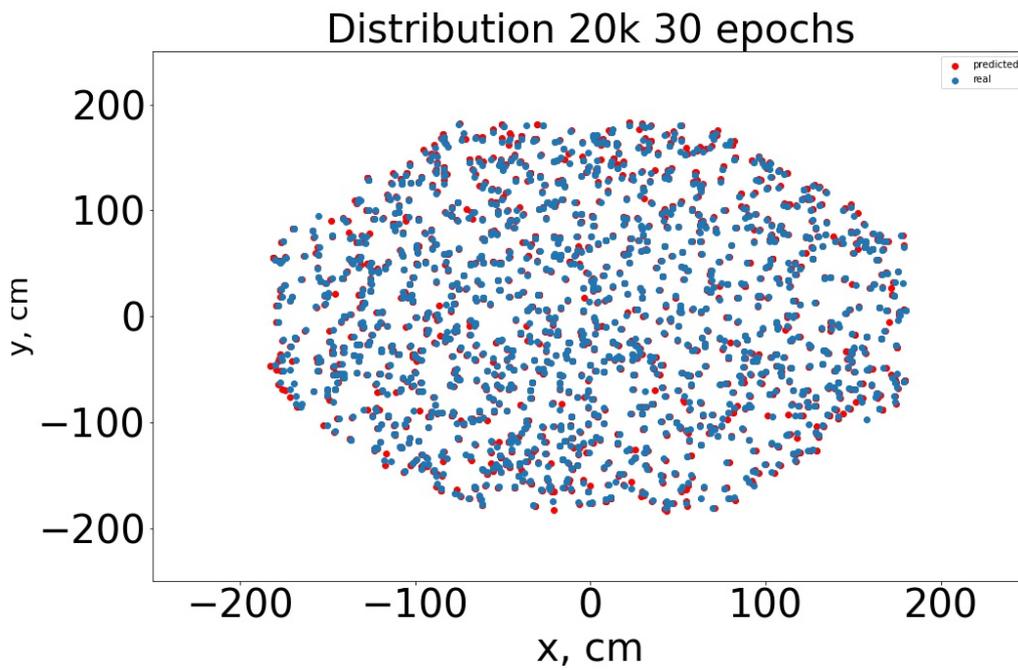


(б) Распределения разности предсказанного и реального значений координат. 20 эпох.

Рисунок 2.7 — Основные графики и гистограммы для 20 и 12 эпох.



(a) Распределение разности предсказанного и реального значений координат.



(б) Предсказанное распределение координат X и Y всех событий.

Рисунок 2.8 — Основные графики для 30 эпох.

2.3. ОБРАБОТКА ДАННЫХ

Коллаборацией DarkSide для сравнительного анализа были предоставлены данные двух типов, смоделированные методом Монте-Карло. Одни были заранее подготовлены для обучения нейронной сети (те, что использовались для подбора параметров сети в предыдущем разделе), а другие предстояло обработать и привести в вид, подходящий для обучения. Оба набора данных содержали X и Y координаты событий. Подготовленные данные представляли собой 3 переменные для каждого события: пара координат x , y и двумерный массив, содержащий номера сработавших фотоумножителей и число фотоэлектронов дошедших до него, отнормированное на общее число фотоэлектронов данного события. Второй набор содержал информацию в таком виде, в каком она формируется в ходе выполнения эксперимента.

Также было известно заранее истинное расположение реальных участвующих в работе фотоумножителей, которое изображено на рисунке 2.9.

Необходимо было оценить корректность и пригодность организованных "сырых" данных. Для этого на готовых "правильных" данных заранее обучалась нейросеть, чтобы после этого можно было сравнить результаты предсказания на обоих наборах данных и сделать выводы о качестве подготовленного материала. Графики и гistogramмы результатов обучения представлены ниже. Основными из них являются: графики потерь 2.14, распределение реальных и предсказанных нейросетью значений координат 2.12, гistogramмы ошибок для каждой из координат (Рисунок 2.11).

Средние ошибки Δx и Δy , приведенные в таблице 2.2 посчитаны следующим образом

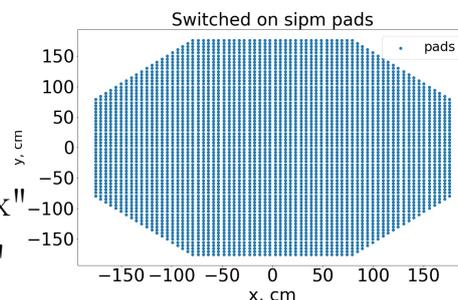


Рисунок 2.9 — Расположение фотоумножителей в детекторе DarkSide-20k

$$\Delta y = \frac{\sum_{i=1}^n |y_{pred} - y_{real}|}{n}, \Delta x = \frac{\sum_{i=1}^n |x_{pred} - x_{real}|}{n}, \quad (2.3)$$

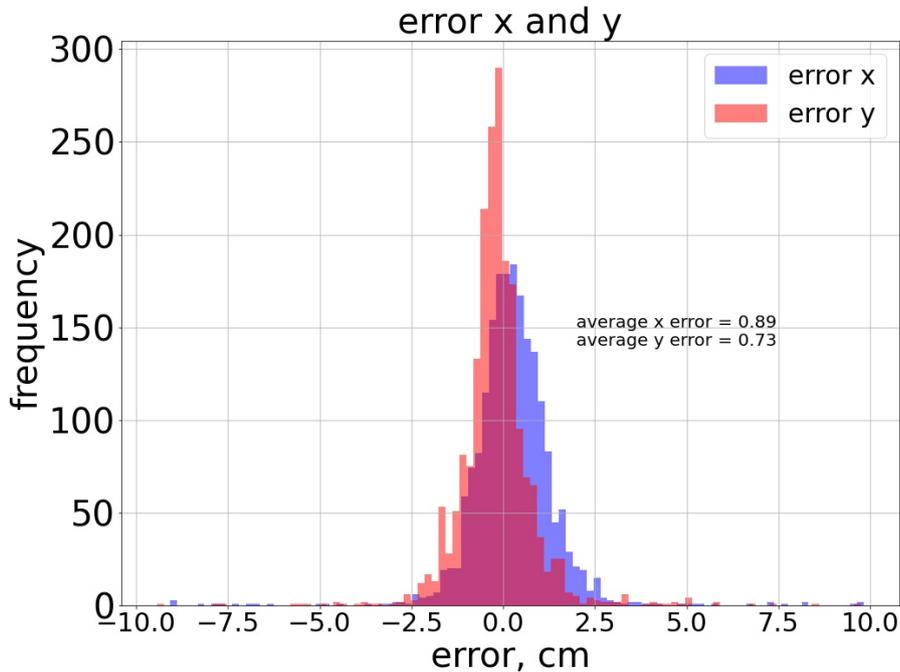


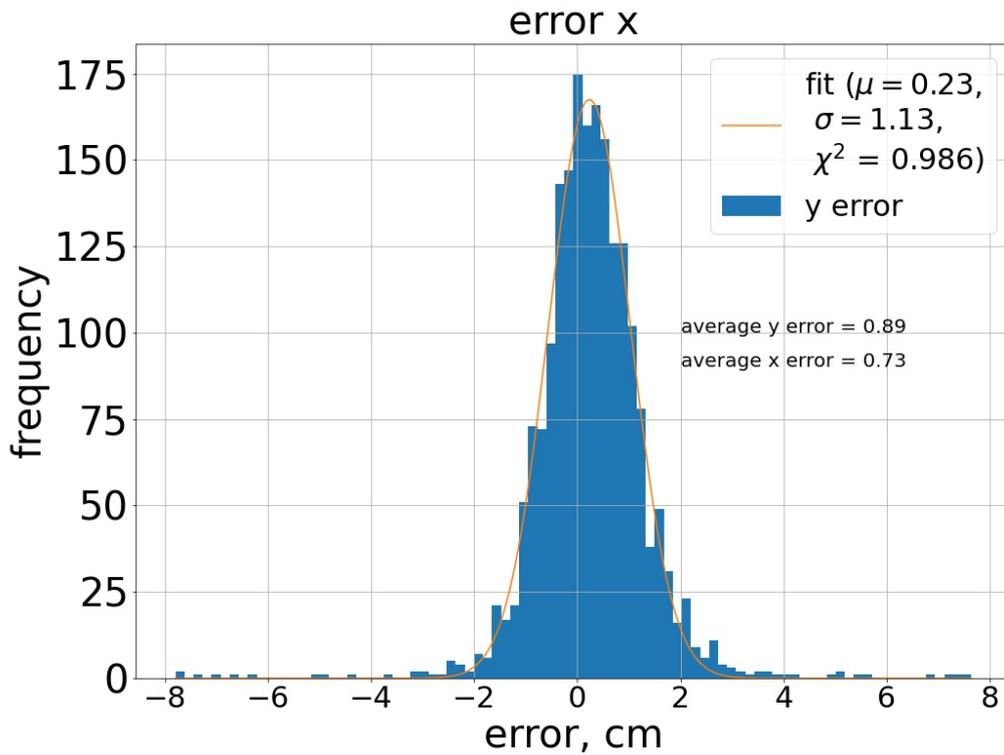
Рисунок 2.10 — Распределение ошибки — разности предсказанного и реального значений для обоих координат.

Важно было не допустить переобучения нейронной сети, когда результаты становятся слишком идеальными, т.е. параметры нейронной сети отлично подбираются только для одного набора данных, поэтому на любых других или на реальных данных с эксперимента обученная нейросеть может неверно предсказывать значения.

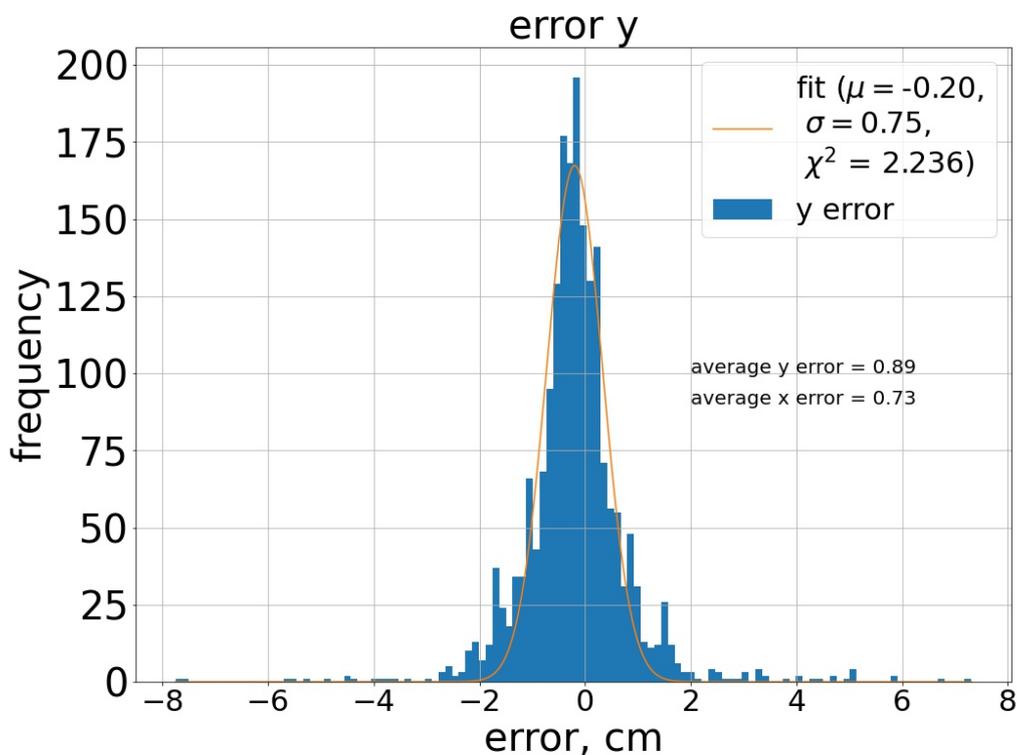
Таблица 2.2 — Полученные средние ошибки для координат

	x , см	y , см
Ошибка, Δ	0.89	0.73

Ранее было установлено, что оптимальное число эпох для существующей конфигурации нейронной сети составляет 30 эпох. Это также подтверждает график потерь для координаты y для 20000 событий на рисунке 2.14. Нейросеть еще не успевает переобучиться, но при этом точность реконструкции так же остается приемлемой (≈ 1 см).

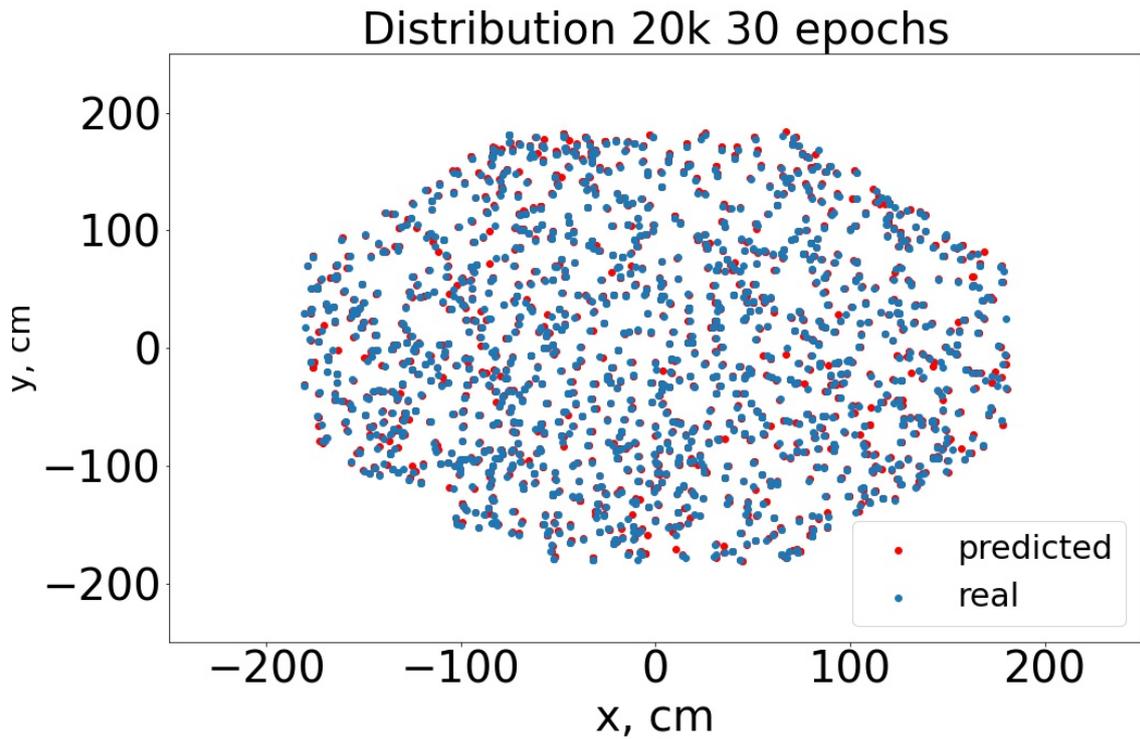


(a) Распределение ошибки (разности предсказанного и реального значений) координаты x.

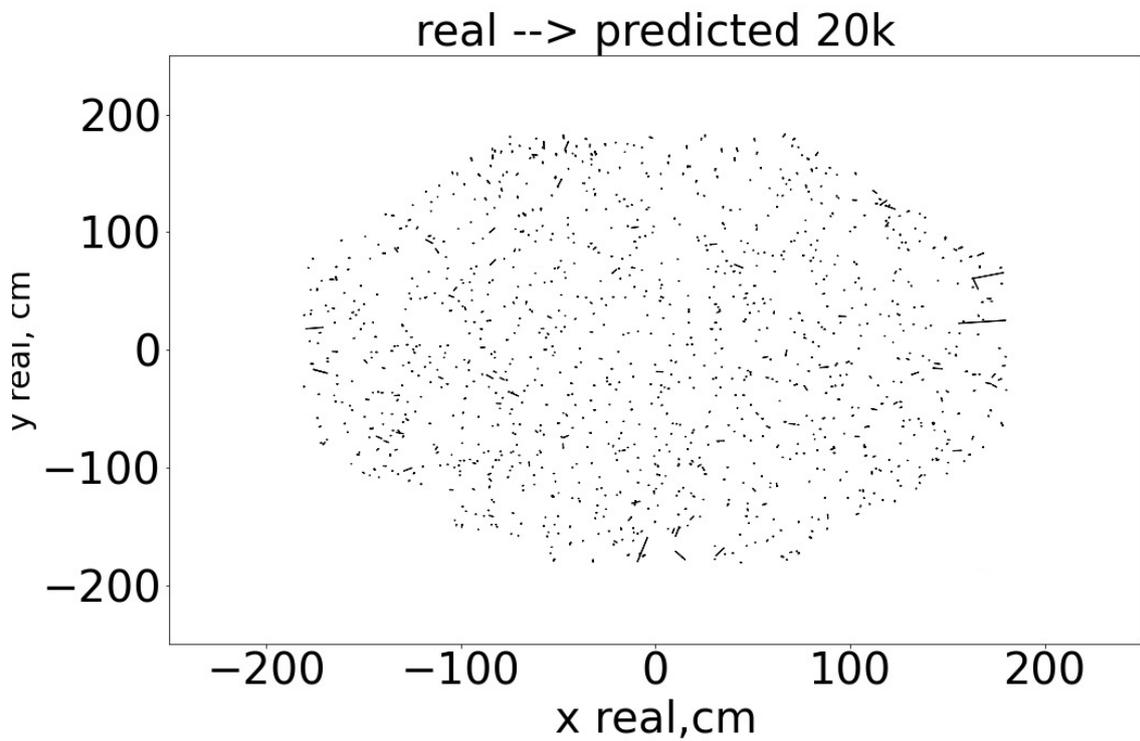


(б) Распределение ошибки (разности предсказанного и реального значений) координаты y.

Рисунок 2.11 — Распределение ошибки — разности предсказанного и реального значений координат для 30 эпох. Подготовленные заранее данные.

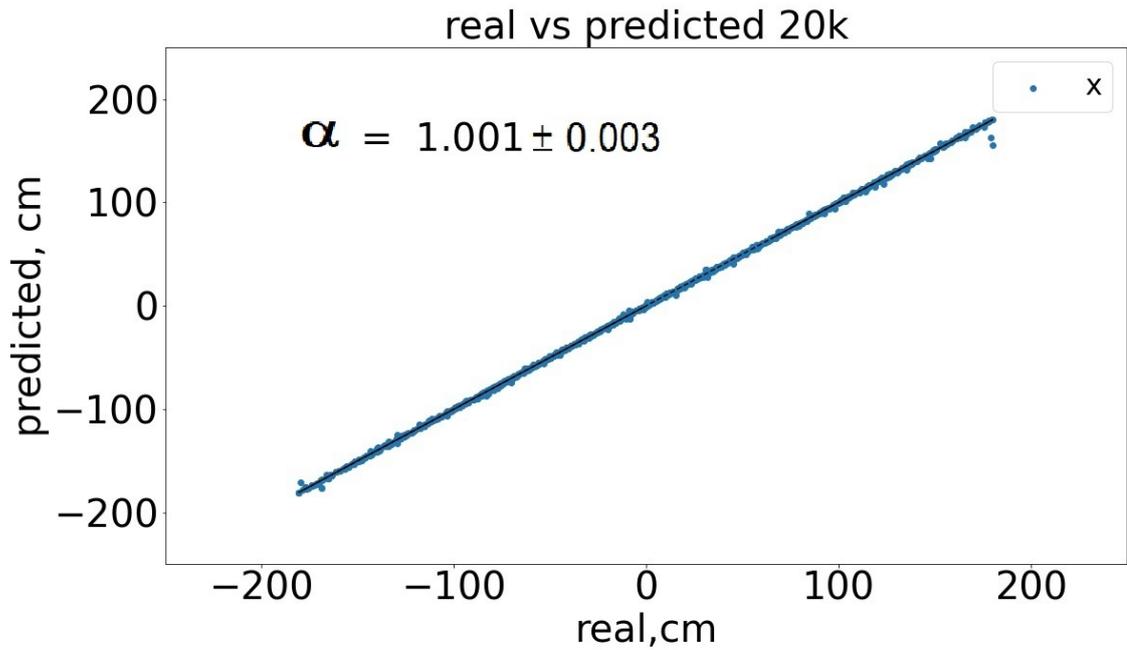


(a) Распределение предсказанных и реальных координат.

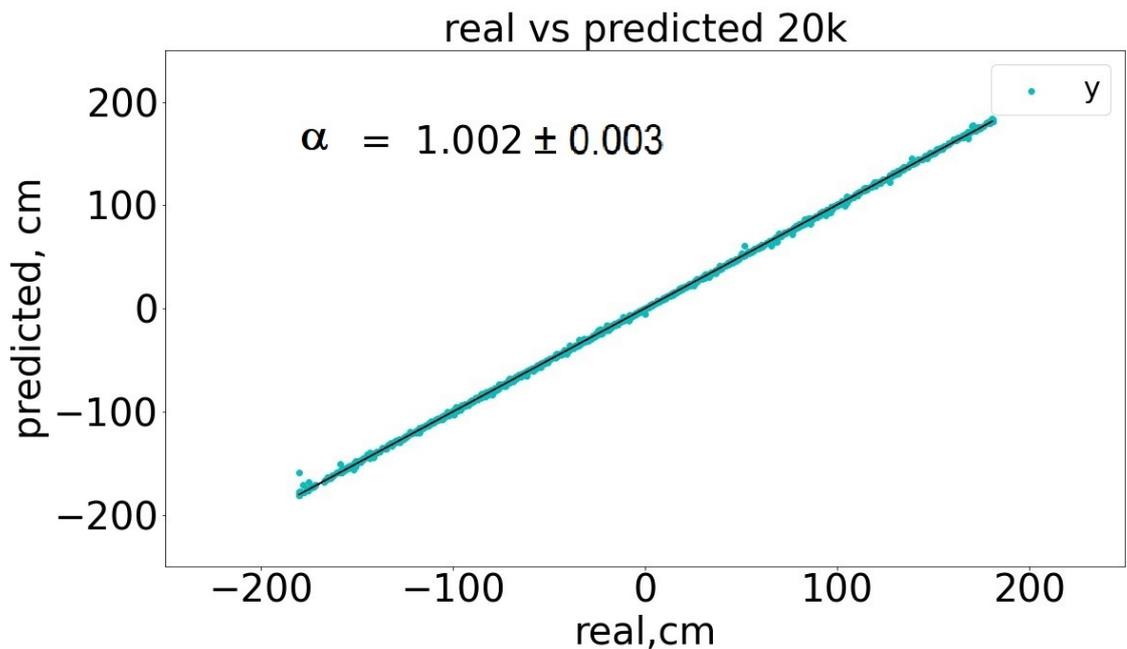


(б) Распределение разности между предсказанными и реальными координатами.

Рисунок 2.12 — Распределение координат.



(а) Зависимость реального значения от предсказанного для координаты x .



(б) Зависимость реального значения от предсказанного для координаты y .

Рисунок 2.13 — Зависимость реального значения от предсказанного.

На рисунке 2.12 представлены распределения предсказанных и реальных координат 2.12а и разности этих двух значений 2.12б полученные после обучения нейросети. На втором распределении видно насколько точным является предсказание координат, т.к. практически отсутствуют "длинные" фрагменты, показывающие большое различие между предсказанными и реальными координатами.

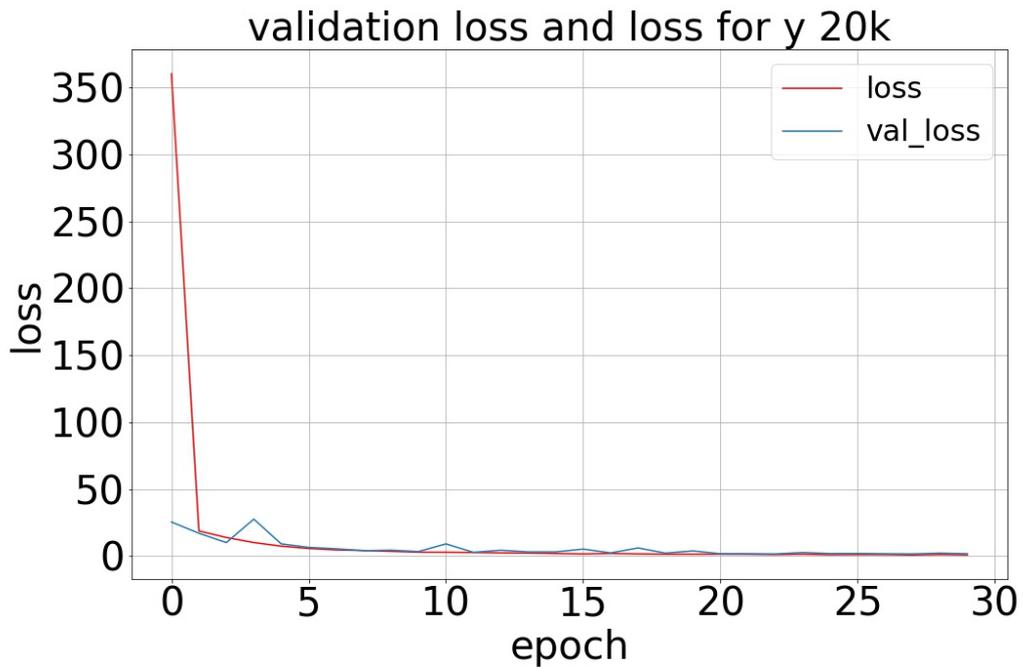


Рисунок 2.14 — График потерь — validation loss и loss для 30 эпох обучения

Эти результаты в ходе дальнейшей работы будут считаться эталонными, именно сравнивая с ними будет определяться корректность обработанных данных.

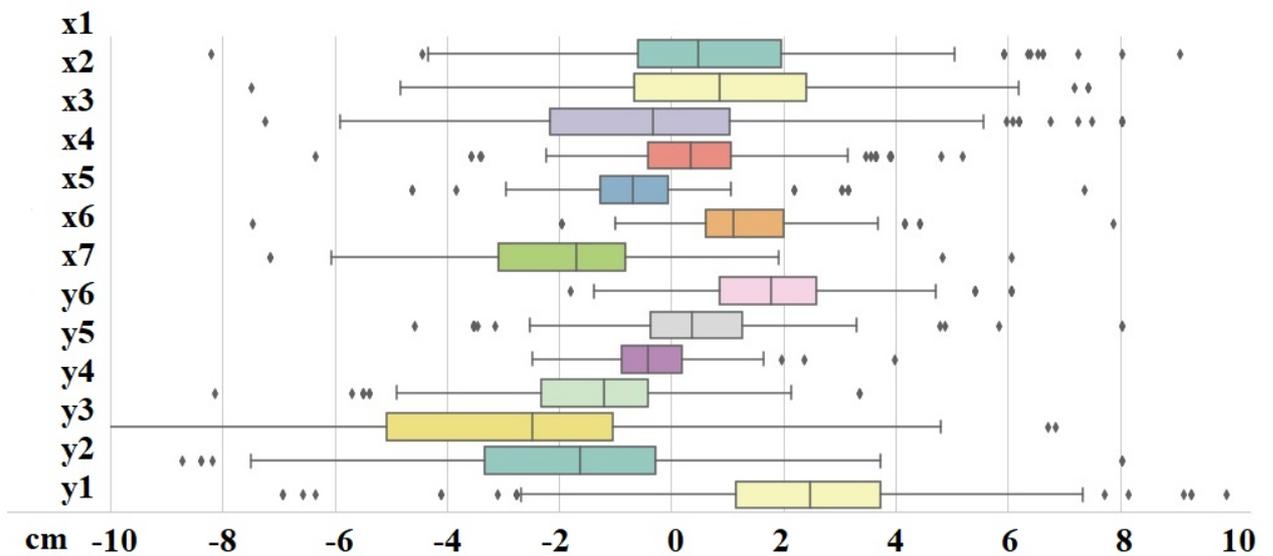


Рисунок 2.15 — Распределение ошибки — разности предсказанного и реального значений для координат для различных выборок.

Для дополнительного анализа была также построена зависимость реального значения координат от предсказанного. Линейный коэффициент,

определенный по МНК, для каждой из координат практически равен единице (рисунок 2.13). Также можно заметить, что вблизи концов прямых практически отсутствуют различного рода нелинейности, что в свою очередь говорит об корректной работе нейронной сети.

Данное исследование было проведено на различных выборках по 20 тысяч событий из 200 тысяч событий. На рисунке 2.15 представлены ошибки для 7 выборок для каждой из координат, каждая выборка обозначается соответствующим индексом на графике.

Средние значения ошибок расположены приблизительно в одном диапазоне. Те участки, которые обозначены точками, считаются выбросами и из дальнейшего рассмотрения исключаются.

Таким образом, можно сделать вывод, что алгоритм реконструкции работает корректно и при этом обеспечивает требуемую точность.

2.4. РЕОРГАНИЗАЦИЯ ДАННЫХ

Из "сырых" данных была сформирована таблица пригодная для обучения нейронной сети по подобию "готовых" данных. На этих данных обучалась нейросеть с теми же параметрами. Наименьшая ошибка, которую удалось получить по обоим координатам, ≈ 2 см.

Основные распределения, графики и гистограммы представлены на рисунках 2.17 и 2.18

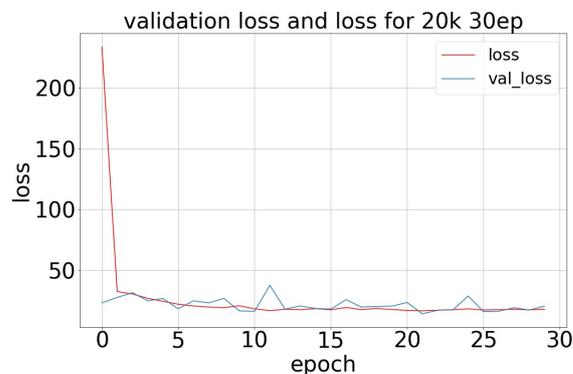
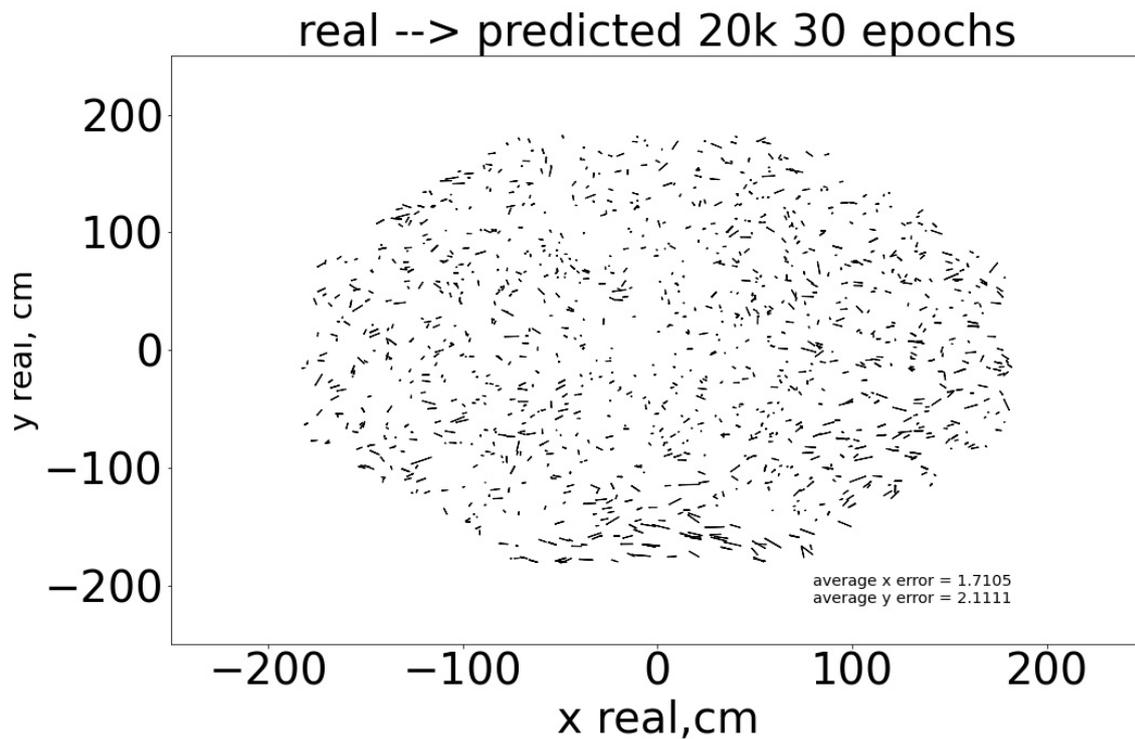


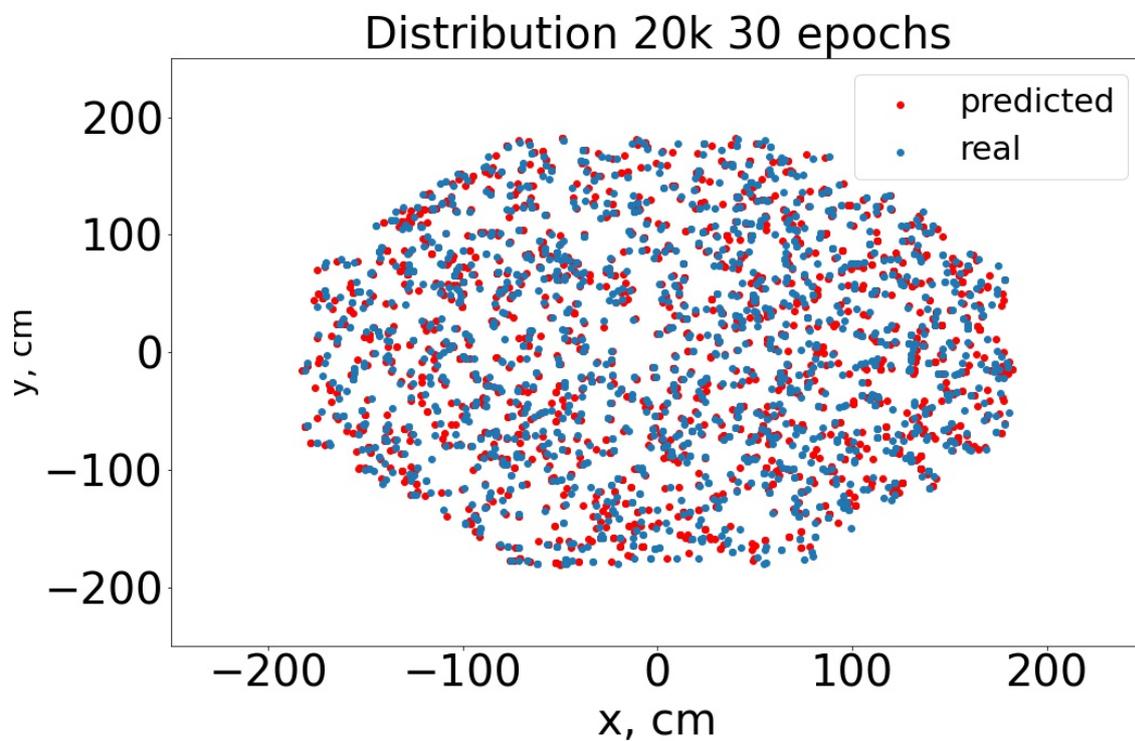
Рисунок 2.16 — График потерь — validation loss и loss

Таблица 2.3 — Полученные средние ошибки для координат

	x , см	y , см
Ошибка, Δ	1.71	2.11

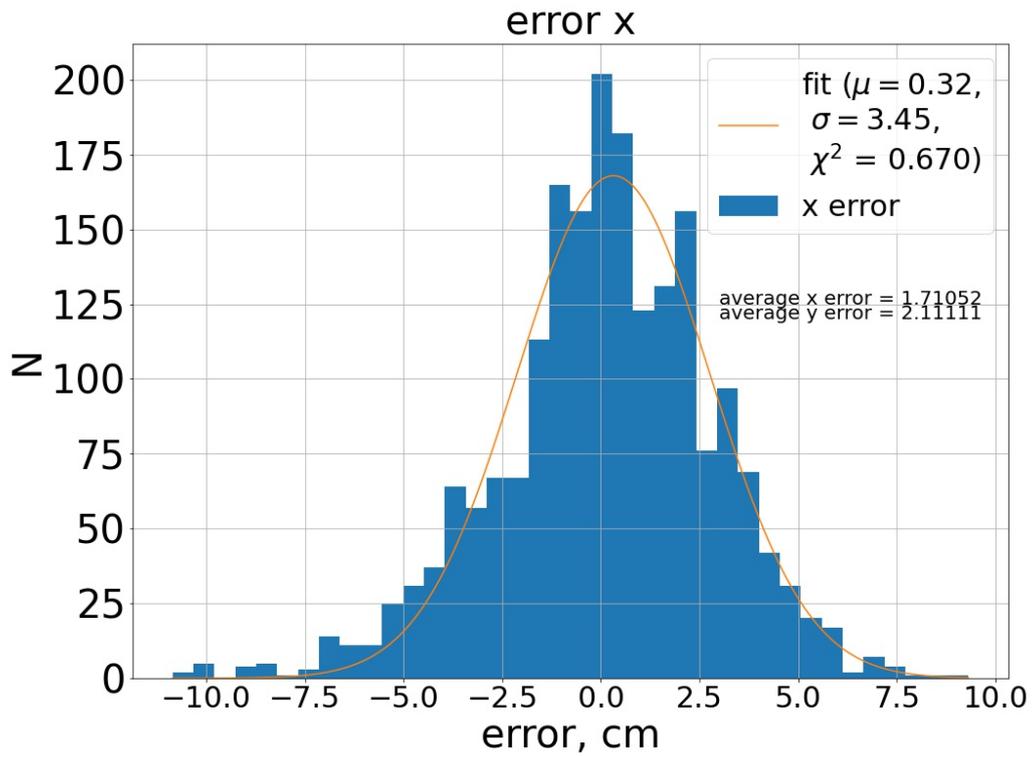


(а) Разность реального и предсказанного значения координат.

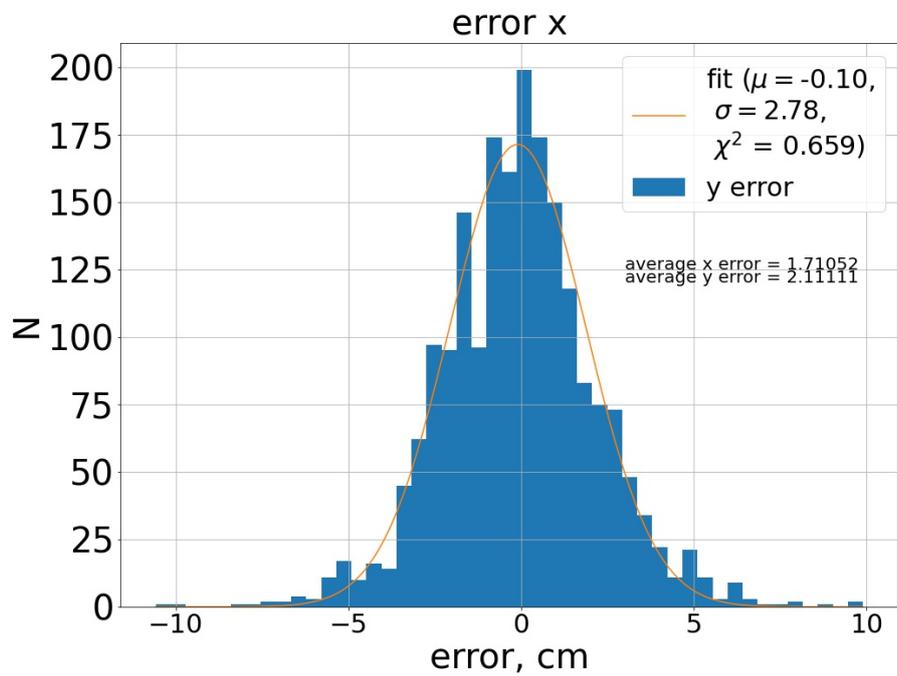


(б) Распределение реальных и предсказанных координат.

Рисунок 2.17 — Распределение координат для переорганизованных данных.



(а) Гистограмма ошибки x .



(б) Гистограмма ошибки y .

Рисунок 2.18 — Распределение координат для переорганизованных данных.

2.5. ОПРЕДЕЛЕНИЕ МАКСИМАЛЬНОГО ЧИСЛА SiPM, КОТОРЫЕ МОГУТ ВЫЙТИ ИЗ СТРОЯ

Как показывает практика, в ходе реального эксперимента кремниевые ФЭУ могут выйти из строя. Их замена является трудоёмкой для реализации задач, в связи с особенностями технической конструкции детекторов DarkSide.

Поэтому одной из задач данной исследовательской работы было определить максимальное количество SiPM, которые могут выйти из строя, при котором реконструкция остается точной. В качестве критерия оценки точности было принято, что точной является реконструкция, ошибка которой остается ≈ 2 см для каждой из координат.

Таблица 2.4 — Средние ошибки для разного количества выключенных фотоумножителей для координат

выключенные ФЭУ	0	250	500	1000
Δx , см	0.86	0.93	2.08	4.49
Δy , см	0.51	0.68	1.75	4.69

Для подготовленных данных достаточное число эпох для обучения равняется 30. Нагляднее в этом можно убедиться на рисунке 2.19, где для 20 тысяч событий изображена разность предсказанных и реальных координат при всех работающих SiPM. На последнем изображении 2.22 разброс значительно больше, чем при всех включенных ФЭУ на рисунке 2.19.

В таблице 2.4 представлены основные полученные значения средних ошибок для x и y при разном числе SiPM, вышедших из строя. SiPM выключались случайным образом в разных частях массива. При всех работающих фотоумножителях средняя ошибка < 1 см (рисунок 2.19). Для 250 нерабочих SiPM ошибка все так же остается меньше сантиметра. Оптимальным числом выключенных SiPM, при котором точность реконструкции удовлетворяет критериям, оказалось 500 из 4120 общего числа всех ФЭУ (рисунок 2.21). Для 1000 неработающих ФЭУ разброс становится очень большим > 4 см (рисунок 2.22).

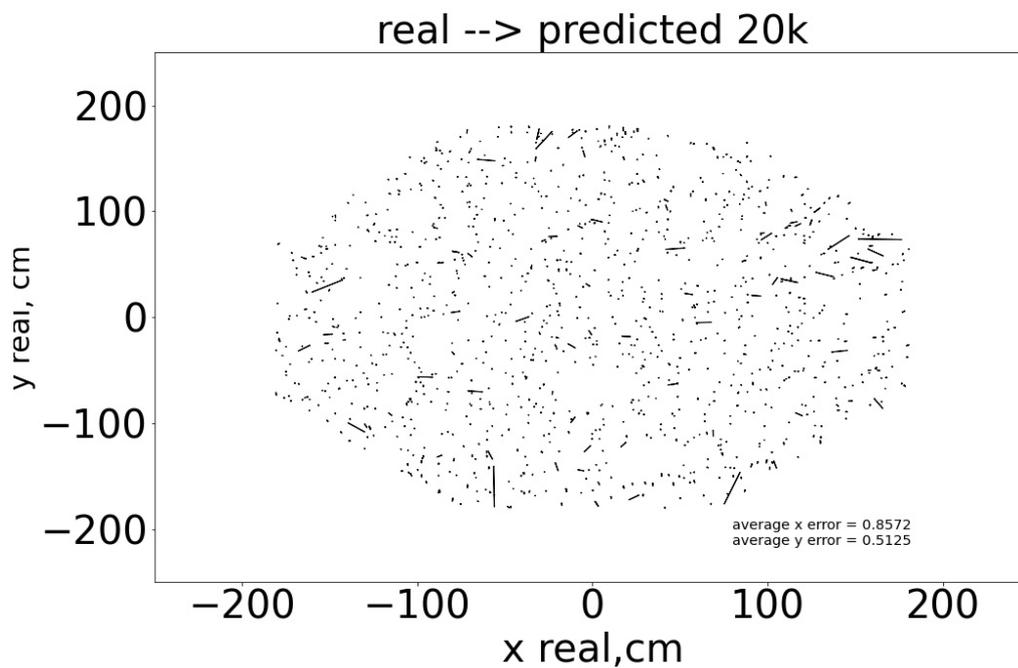


Рисунок 2.19 — Распределение ошибок для координат. Случай, когда все SiPM работают

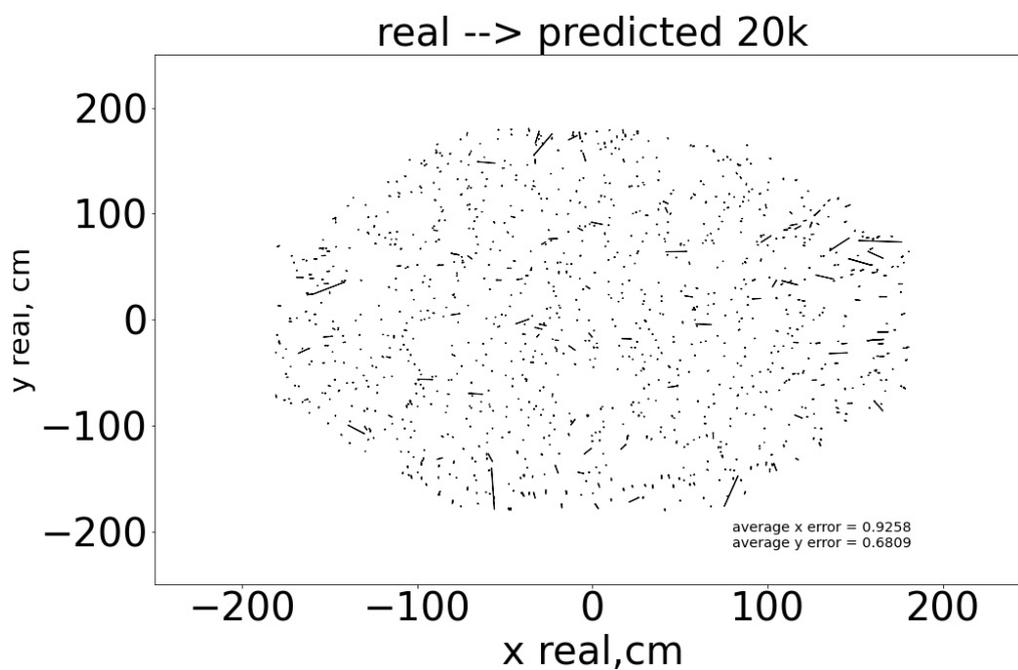


Рисунок 2.20 — Распределение ошибок для координат. 250 SiPM выключены.

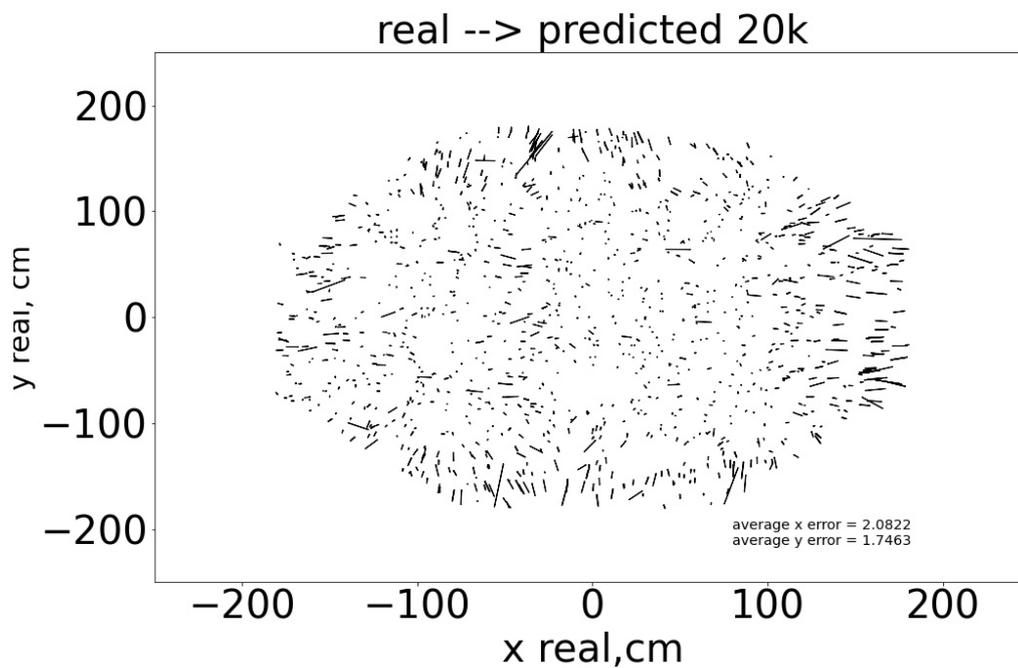


Рисунок 2.21 — Распределение ошибок для координат. 500 SiPM выключены

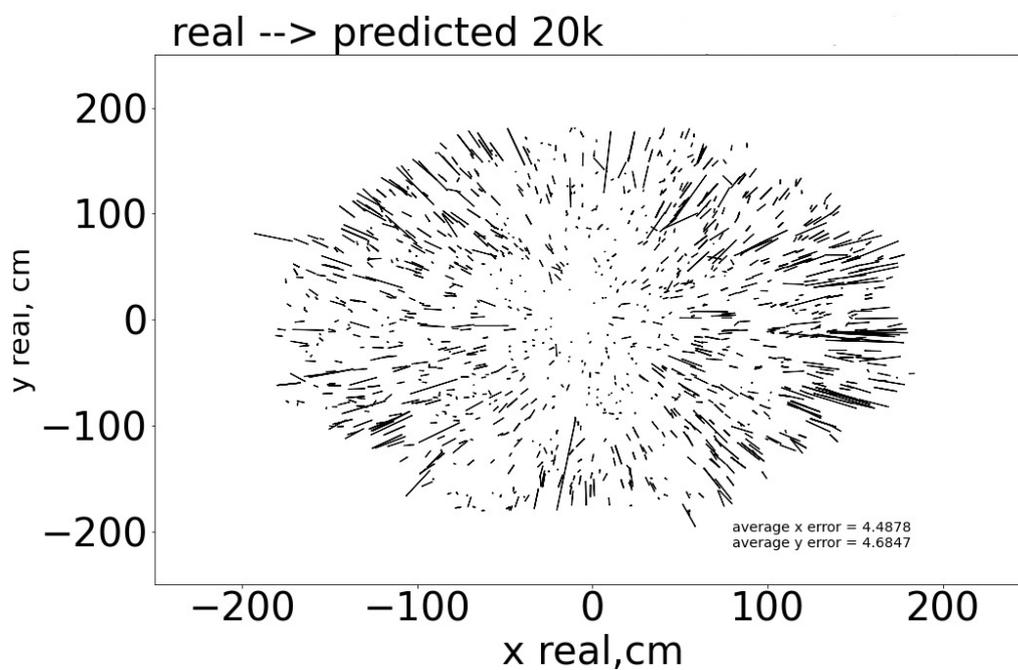
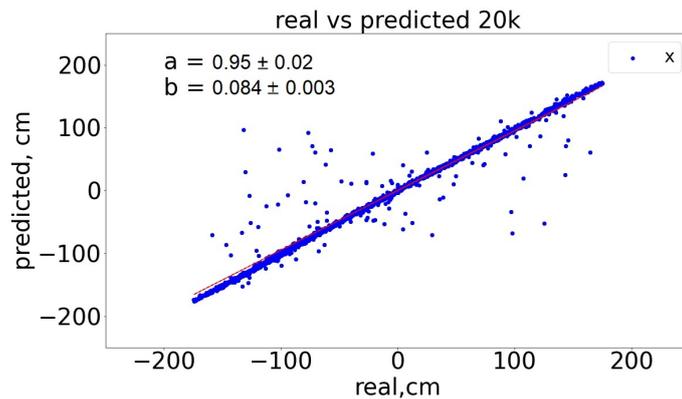


Рисунок 2.22 — Распределение ошибок для координат. 1000 SiPM выключены

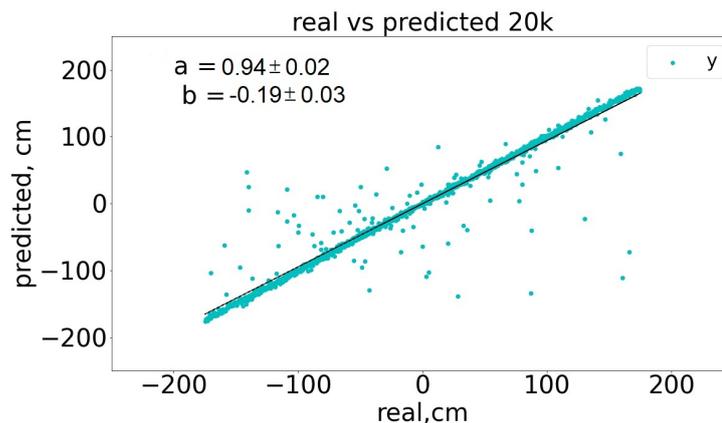
2.6. СОЗДАНИЕ СОБСТВЕННЫХ MONTE-CARLO ДАННЫХ, ПРИ ПОМОЩИ G4DS

В качестве завершающего этапа необходимо было создать собственные данные методом Монте-Карло и преобразовать их в необходимый для нейронной сети вид, чтобы в дальнейшем предсказать координаты.

Ниже на рисунках 2.23а для x и 2.23б для y представлены зависимости реального значения координат от предсказанного нейронной сетью. Большинство точек располагается на одной прямой, однако имеются и точки, лежащие в области рядом с прямой. Это свидетельствует о том, что реконструкция не является точной. Это подтверждает и график распределения координат.



(а) Зависимость реального значения от предсказанного для координаты x .

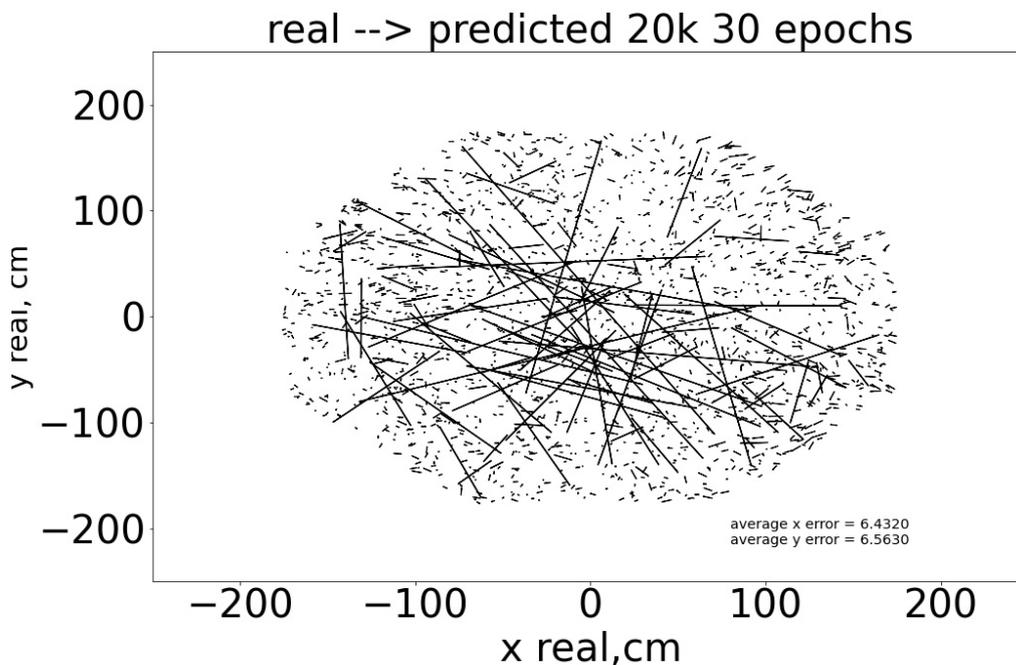


(б) Зависимость реального значения от предсказанного для координаты y .

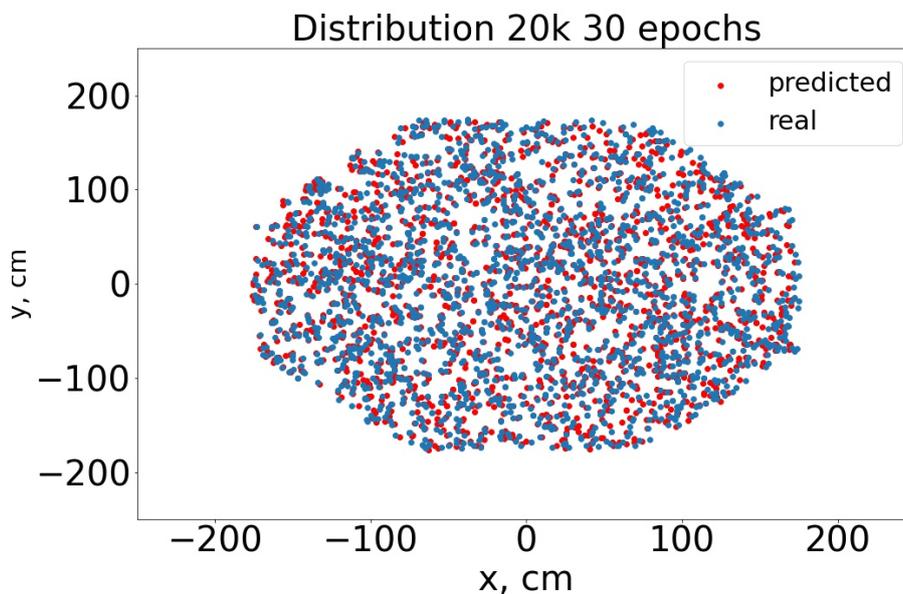
Рисунок 2.23 — Зависимость реального значения от предсказанного для собственных данных.

На рисунке 2.24а представлен график ошибки в виде линий, где на-

чальная точка – это реальные координаты x и y , а конечная точка – предсказанные координаты, присутствует разброс в предсказании координат, из-за чего общая точность предсказаний падает несмотря на то, что из рисунка видно преобладание "коротких" фрагментов, сигнализирующих о хорошей реконструкции.



(a) Распределение разности реального значения координат и предсказанного.



(б) Распределение реальных и предсказанных координат.

Рисунок 2.24 — Зависимость реального значения от предсказанного для собственных данных.

В отличие от выше полученных результатов, видно, что присутствует большой разброс. Наилучшие результаты, которые удалось получить представлены в таблице 2.5. Гистограмма ошибок для y координаты представлены на рисунке 2.25.

Таблица 2.5 — Полученные средние ошибки для координат

	x , см	y , см
Ошибка, Δ	6.43	6.56

Такого рода расходимости в точности, могут быть следствием двух причин:

- 1) Неправильно подобраны параметры для Монте-Карло моделирования.
- 2) Нейронная сеть требует усложнения в своей конфигурации, например добавление дополнительных сверточных слоев и т.п.
- 3) Требуется провести некоторый физический отбор по событиям, которые как раз и искажают результаты.

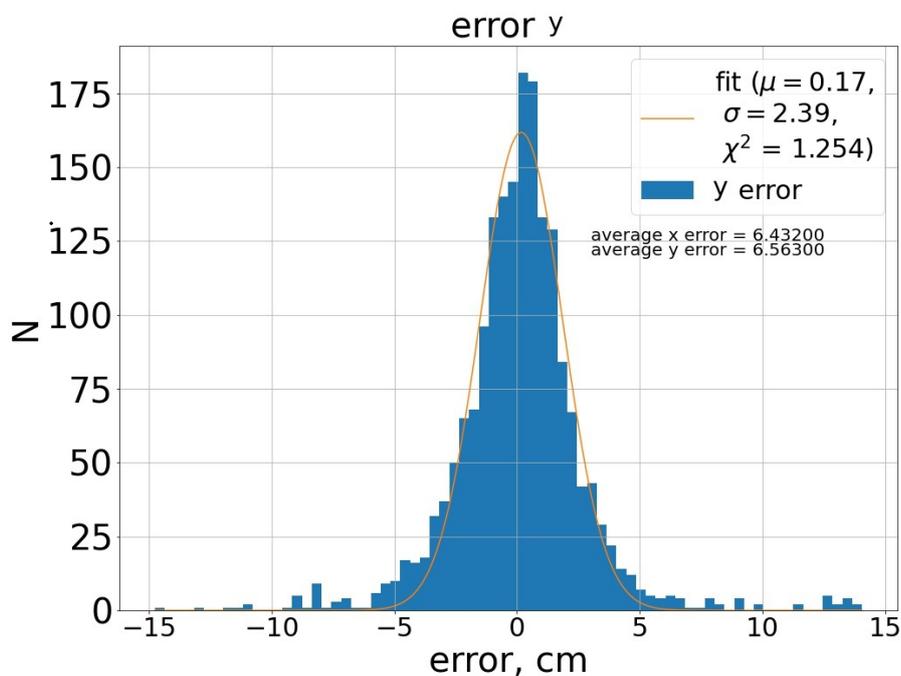


Рисунок 2.25 — Распределение ошибки для координаты y .

2.7. ЗАКЛЮЧЕНИЕ

При выполнении данной научно-исследовательской работы были изучены основные принципы работы эксперимента DarkSide-20k и сверточных нейронных сетей.

Разработана нейронная сеть, которая с необходимой точностью реконструирует положение события в пространстве в плоскости x и y .

Обработаны данные, смоделированные методом Монте-Карло, отобрана
нужная для реконструкции информация о событиях, с заведомо известными значениями координат x и y , на основе которой была обучена нейросеть.

Подобраны оптимальные параметры обучения, такие чтобы реконструкция была точной.

Определено число выключенных ФЭУ при работе детектора, при котором реконструкция событий остается возможной.

СПИСОК ЛИТЕРАТУРЫ

1. Kinematic studies of early-type stars. I. photometric survey, space motions, and comparison with radio observations. / V. C. Rubin [et al.] // The Astronomical Journal 67. — 1967. — DOI: [10.1086/108758](https://doi.org/10.1086/108758).
2. Low-Mass Dark Matter Search with the DarkSide-50 Experiment / P. Agnes [et al.]. — 2018 ; — arxiv: <https://arxiv.org/pdf/1802.06994.pdf>.
3. DarkSide-20k: A 20 tonne two-phase LAr TPC for direct dark matter detection at LNGS / C. E. Aalseth, F. Acerbi, P. Agnes, [et al.] // The European Physical Journal Plus. — 2018. — DOI: [10.1140/epjp/i2018-11973-4](https://doi.org/10.1140/epjp/i2018-11973-4).
4. Separating ^{39}Ar from ^{40}Ar by cryogenic distillation with Aria for dark matter searches / P. Agnes [et al.]. — 2021 ; — arxiv: <https://arxiv.org/pdf/2101.08686.pdf>.
5. *Garcia E. S.* DArT, a detector for measuring the ^{39}Ar depletion factor // Journal of Instrumentation. — 2020. — DOI: [10.1088/1748-0221/15/02/c02044](https://doi.org/10.1088/1748-0221/15/02/c02044).
6. *Bottino B.* Dark matter search with liquid argon in DarkSide: results with scientific and technological prototypes. — 2019 ; — http://www.infn.it/thesis/thesis_dettaglio.php?tid=13764.
7. Cosmogenic Backgrounds in Borexino at 3800 m water-equivalent depth / G. Bellini [et al.] // Journal of Cosmology and Astroparticle Physics. — 2013. — arxiv: <https://arxiv.org/abs/1304.7381>.
8. DarkSide-50 532-day dark matter search with low-radioactivity argon / P. Agnes [et al.] // Physical Review D. — 2018. — arxiv: <https://arxiv.org/abs/1802.07198>.
9. First results from the DarkSide-50 dark matter experiment at Laboratori Nazionali del Gran Sasso / P. Agnes [et al.] // Physics Letters B. — 2015. — arxiv: <https://arxiv.org/abs/1410.0653>.

10. *Collaboration T. G. A. D. M.* DarkSide-20k Design Report. — 2021.
11. *Cullo A.* Using Convolutional Neural Networks for DarkSide-20k Analysis. — 2018.
12. *Ketkar N.* Introduction to keras // Deep learning with Python. — Springer, 2017. — P. 97–111.