

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГАОУ ВО НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук
Образовательная программа «Прикладная математика и информатика»

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

Исследовательский проект на тему:

Оптимизация размера сенсоров для физики элементарных частиц

Выполнил студент:

группы #БПМИ203, 4 курса

Зиманов Алихан

Принял руководитель ВКР:

Болдырев Алексей Сергеевич

Научный сотрудник

Факультет компьютерных наук НИУ ВШЭ

Москва 2024

Содержание

Аннотация	3
Abstract	4
1 Введение	5
2 Обзор литературы	7
2.1 Смежные работы	7
2.2 Используемые методы	9
3 Методология	10
3.1 Данные	10
3.2 Модели	11
3.3 Процедура обучения	13
3.4 Метрики качества	13
3.5 Аугментации	15
3.6 Постановка экспериментов	15
3.7 Ресурсы	16
3.8 Анализ результатов	16
4 Результаты	16
5 Заключение	20
6 Благодарности	21
7 Использование ИИ	21
Список литературы	23

Аннотация

Эксперименты по физике частиц часто связаны с точным измерением энергии и положения частиц, таких как фотоны, взаимодействующих с детекторами типа электромагнитных калориметров. Оптимизация размера сенсора этих детекторов имеет решающее значение для получения точных реконструкций свойств частиц с учетом экономической эффективности. В этом исследовании мы изучаем оптимизацию размера датчика для экспериментов по физике частиц, уделяя особое внимание реконструкции энергии фотонов и их положения в калориметре. Мы предлагаем использовать методы глубинного обучения для решения этой задачи и оценить различные разрешения датчиков от 10×10 до 40×40 при фиксированном физическом размере калориметра. Наш подход предполагает обучение моделей глубинного обучения, включая архитектуры ResNet18 и Vision Transformer, и сравнение их качества с базовыми моделями, такими как аналитическая модель и линейная регрессия. Мы используем такие метрики, как средняя квадратичная ошибка (MSE), средняя абсолютная ошибка (MAE), корень из средней квадратичной ошибки (RMSE), а также взвешенные версии этих метрик для оценки точности реконструкции для датчиков разного размера. Кроме того, мы изучили влияние функций потерь, стратегий обучения и параметров моделей на качество работы моделей глубинного обучения. Полученные нами результаты подчеркивают важность выбора оптимального разрешения сенсоров, обеспечивающего баланс между точностью реконструкции и экономической эффективностью при проведении экспериментов по физике частиц.

Ключевые слова

Физика элементарных частиц, оптимизация размера сенсоров, глубинное обучение, компьютерное зрение, реконструкция энергии фотонов, БАК, электромагнитные калориметры

Abstract

Particle physics experiments often involve the precise measurement of energy and position of particles, such as photons, interacting with detectors like electromagnetic calorimeters. Optimizing the sensor size of these detectors is crucial for achieving accurate reconstructions of particle properties while considering cost-effectiveness. In this research, we explore the optimization of

sensor size for particle physics experiments, focusing on the reconstruction of photon energy and position within calorimeter cells. We propose using deep learning methods to address this task and evaluate various sensor granularities ranging from 10×10 to 40×40 keeping the same physical size of the calorimeter. Our approach involves training deep learning models, including ResNet18 and Vision Transformer architectures, and comparing their performance against baseline models such as analytical formula-based model and Linear Regression. We employ metrics like Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and weighted versions of these metrics to assess the reconstruction accuracy across different sensor granularities. Additionally, we investigate the impact of loss functions, training strategies, and model parameters on the performance of the deep learning models. Our findings highlight the importance of selecting an optimal sensor size that balances reconstruction accuracy and cost-effectiveness for particle physics experiments.

Keywords

Particle Physics, Sensor Optimization, Deep Learning, Computer Vision, Photon Energy Reconstruction, LHC, ECAL

1 Введение

Добавить ссылки на всю используемую литературу

Эксперименты по физике элементарных частиц направлены на раскрытие фундаментальных структурных элементов Вселенной и сил, управляющих их взаимодействием. Эти эксперименты часто связаны со столкновением частиц высоких энергий и последующим измерением и анализом свойств частиц, образующихся при этих столкновениях. В таких экспериментах детекторы играют ключевую роль в улавливании признаков различных частиц, включая фотоны, электроны, мюоны, нейтрино и другие. Среди различных типов детекторов, используемых в экспериментах по физике частиц, электромагнитные калориметры предназначены для измерения энергии частиц, в частности фотонов и электронов, с высокой точностью.

Электромагнитные калориметры обычно состоят из плотных материалов, таких как свинец или вольфрам, расположенных слоями, перемежающимися с чувствительными элементами, обычно изготовленными из таких материалов, как кремний или сцинтилляционные кристаллы. Когда высокоэнергетические фотоны или электроны взаимодействуют с материалами калориметра, они создают электромагнитные ливни, являющиеся каскадом испускания вторичных и последующих частиц. Энергия, накопленная этими частицами в калориметре, измеряется чувствительными элементами, предоставляя ценную информацию об энергии и положении падающих частиц. Пример электромагнитного калориметра и электромагнитного ливня можно увидеть на Рисунке 1.1.

Одним из важнейших аспектов при разработке и оптимизации электромагнитных калориметров является выбор размера датчика. Под размером датчика понимаются размеры отдельных чувствительных элементов калориметра, которые напрямую влияют на зернистость и пространственное разрешение детектора. Более мелкая детализация, достигаемая за счет использования датчиков меньшего размера, позволяет лучше локализовать осаждение энергии в ячейках калориметра, что приводит к улучшению пространственного разрешения и реконструкции энергии. Однако уменьшение размера датчика сопряжено с определенными трудностями, включая повышение сложности считывания данных с датчика, увеличение стоимости изготовления и потенциальное ухудшение соотношения сигнал/шум.

Оптимизация размера датчика для электромагнитных калориметров предполагает поиск баланса между точностью реконструкции и экономической эффективностью. Хотя датчики меньшего размера обеспечивают более тонкое пространственное разрешение и потенциально более высокую точность реконструкции, они не всегда могут быть практичными или

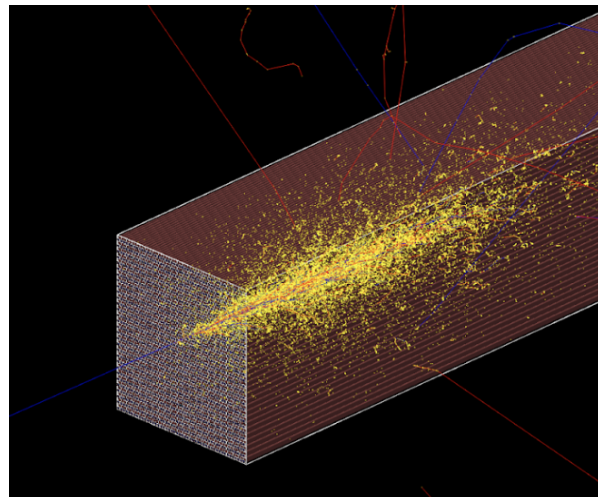
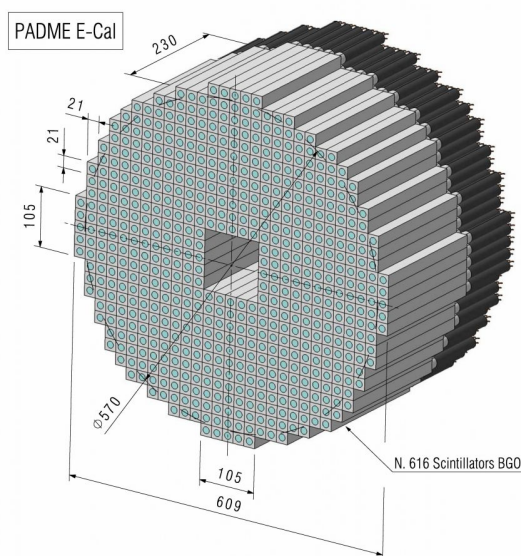


Рис. 1.1: Электромагнитный калориметр PADME и симуляция электромагнитного ливня с помощью GEANT4¹.

экономичными, особенно в крупномасштабных экспериментах. И наоборот, датчики большего размера могут обеспечить достаточную производительность реконструкции при меньших затратах, но при этом может пострадать общая точность детектора.

В этом исследовании мы изучаем оптимизацию размера датчика для экспериментов по физике элементарных частиц, сосредоточившись на реконструкции энергии и положения фотонов в ячейках электромагнитного калориметра. Мы предлагаем использовать методы глубинного обучения для решения этой задачи и оцениваем производительность калориметров, состоящих из чувствительных элементов в количестве от 10×10 до 40×40 . В частности, мы обучаем модели глубинного обучения, включая сверточные нейронные сети (CNN), такие как ResNet18, и архитектуры на основе трансформеров, такие как Vision Transformer (ViT), для восстановления энергии и положения фотонов из смоделированных данных калориметра. Кроме того, мы сравниваем производительность этих моделей глубинного обучения с базовыми методами, такими как аналитическая модель и линейная регрессия (LinReg), учитывая такие метрики, как средняя квадратичная ошибка (MSE), средняя абсолютная ошибка (MAE), корневая средняя квадратичная ошибка (RMSE) и их взвешенные варианты.

Кроме того, мы изучаем влияние различных функций потерь, стратегий обучения и гиперпараметров моделей глубинного обучения на точность восстановления для датчиков разного размера. Систематически анализируя эти факторы, мы стремимся понять компро-

¹Источники изображений: <https://padme.lnf.infn.it/ecal/> и <https://ep-news.web.cern.ch/content/geant4-modern-and-versatile-toolkit-detector-simulations>

миссы, связанные с выбором оптимального размера датчика для электромагнитных калориметров в экспериментах по физике частиц. Наши выводы будут способствовать разработке и оптимизации будущих детекторных систем, позволяющих проводить более точные измерения свойств частиц с учетом практических ограничений и стоимости.

Оставшееся часть работы организована следующим образом. Раздел 2 рассказывают о существующих продвижениях в применении глубинного обучения для задач физики высоких энергий, а также приводит краткое описание используемых в данной работе моделей. В Разделе 3 подробно расписывается постановка решаемой задачи, приводится описание данных, используемые модели и техники. В Разделе 4 приводятся и обсуждаются результаты проведенных экспериментов. В Разделе 5 мы подводим итоги наших выводов.

Оформить гитхаб с документацией и комментариями

Реализация всей процедуры обучения, а также проведения экспериментов доступны в интернете².

2 Обзор литературы

2.1 Смежные работы

В области физики частиц наблюдается всплеск исследований, изучающих применение методов глубинного обучения для задач моделирования и реконструкции в экспериментах по физике высоких энергий. В работе [2] были представлены алгоритмы глубинного обучения для моделирования и реконструкции калориметрии с использованием детального моделирования ливня в качестве обучающих данных. Их нейронная сеть для реконструкции, обученная на уровне калориметрических ячеек, продемонстрировала значительные улучшения по сравнению с современными алгоритмами, позволяя одновременно идентифицировать частицы и восстанавливать энергию. Эта работа очень актуальна для нашего исследования, поскольку мы также изучаем оптимизацию моделей глубокого обучения для реконструкции частиц в электромагнитных калориметрах, уделяя особое внимание оптимизации размера датчика и точности реконструкции.

Кроме того, в работе [3] предложен подход с использованием компьютерного зрения к алгоритмам Particle Flow (PFlow) в экспериментах по физике высоких энергий, использующий изображения калориметра для превосходной реконструкции энергетических распадов нейтральных частиц на фоне столкновения заряженных частиц. Их исследование продемон-

²<https://github.com/dxtvzw/ecal-optimization/tree/master/src>

стрировало эффективность методов глубинного обучения в улучшении реконструкции кинематических свойств частиц, что соответствует нашей цели оптимизации размера датчика для улучшения пространственного разрешения и реконструкции энергии в электромагнитных калориметрах.

В смежном исследовании [4] представил новую методику, основанную на глубинном обучении, для реконструкции распадов частиц, усиленных преобразованием Лоренца, минуя традиционные аналитические методы реконструкции. Их подход, проверенный на данных о столкновениях на БАК, продемонстрировал замечательную производительность в восстановлении свойств частиц непосредственно из минимально обработанных данных детектора. Хотя основное внимание было уделено реконструкции распадов частиц, их методология и стратегии обучения актуальны для наших исследований по оптимизации размера датчика и точности реконструкции в электромагнитных калориметрах.

Кроме того, в работе [1] сравнивалась производительность глубинных нейронных сетей, включая сверточные нейронные сети (CNN) и графовые нейронные сети (GNN), с современными методами восстановления энергии в сегментированных калориметрах. Сравнительный анализ показал значительное улучшение энергетического разрешения, достигнутое CNN, обученными на чистых образцах частиц, что подчеркивает важность скрытых представлений характеристик, используемых нейросетевыми методами. Это исследование дает ценное представление о потенциальных преимуществах подходов глубинного обучения для реконструкции энергии в калориметрах, что поможет нам в исследовании оптимизации размера датчика и точности реконструкции.

Наконец, в работе [5] представлено исследование реконструкции фотонов в электромагнитном калориметре Belle II с помощью графовых нейронных сетей (GNN), демонстрирующее значительное улучшение энергетического разрешения по сравнению с существующими алгоритмами реконструкции. Их работа, сосредоточенная на восстановлении как изолированных, так и перекрывающихся фотонов, продемонстрировала многообещающий потенциал GNN в повышении энергетического разрешения и уменьшении хвостов в восстановленных энергетических распределениях. Это исследование совпадает с нашей исследовательской задачей по оптимизации размера датчика для улучшения реконструкции фотонов в электромагнитных калориметрах, что подчеркивает актуальность методов глубинного обучения для совершенствования алгоритмов реконструкции частиц.

В целом, вышеупомянутые исследования подчеркивают растущий интерес к использованию методов глубинного обучения для задач моделирования и реконструкции в экспериментах по физике частиц, особенно в контексте калориметрии. Опираясь на идеи и ме-

тодологии, представленные в этих работах, наше исследование призвано внести вклад в оптимизацию размеров датчиков для электромагнитных калориметров, что в конечном итоге повысит точность и эффективность реконструкции частиц в экспериментах по физике высоких энергий.

2.2 Используемые методы

ResNet (Residual Neural Network) - это глубокая нейронная сеть, предназначенная для решения задач компьютерного зрения. Основная идея ResNet состоит в использовании блоков с остаточным соединением (residual connections), которые позволяют эффективно обучать глубокие сети, предотвращая проблему затухающего градиента.

Архитектура ResNet состоит из стека блоков, каждый из которых содержит несколько сверточных слоев. Важной особенностью ResNet является наличие skip-connections, которые позволяют пропускать через блок исходные признаки. Это позволяет избежать потери информации и упрощает обучение сети.

Принцип работы ResNet заключается в передаче информации через блоки с остаточным соединением, что помогает сети эффективно изучать более глубокие представления изображений. В процессе обучения, градиенты легче распространяются благодаря skip-connections, что способствует более стабильному и быстрому обучению сети.

Модель Vision Transformer (ViT) представляет собой архитектуру нейронной сети, разработанную для решения задач компьютерного зрения. Она основана на идее применения трансформеров, которые изначально были разработаны для обработки последовательностей, к обработке изображений.

Архитектура ViT состоит из двух основных компонентов: энкодера и классификатора. Энкодер состоит из нескольких слоев трансформера, включая механизм внимания (attention mechanism). Изображение разбивается на патчи, которые затем подаются на вход энкодера. После обработки энкодером, полученные представления патчей объединяются и подаются на вход классификатору для выполнения задачи классификации. Можно заменить классификатор на регрессор и тогда модель будет решать задачи регрессии.

Одно из преимуществ ViT по сравнению с ResNet заключается в его способности эффективно обрабатывать изображения любого размера. В ResNet изображения обычно масштабируются до фиксированного размера, что может привести к потере информации или искажению изображения. В ViT изображение разбивается на патчи, поэтому размер изображения не ограничен, и модель может работать с изображениями различных размеров без

необходимости их масштабирования.

Кроме того, ViT показывает хорошие результаты на ряде стандартных задач компьютерного зрения, демонстрируя сопоставимое или даже лучшее качество по сравнению с классическими архитектурами, такими как ResNet, при правильной настройке гиперпараметров и аугментации данных.

3 Методология

В этом Разделе мы описываем методику оптимизации размеров датчиков в электромагнитных калориметрах с помощью различных моделей глубокого обучения, включая аналитические модели, линейную регрессию, ResNet, простые сверточные сети и модели, основанные на архитектуре трансформеров.

3.1 Данные

Для данного исследования была использована следующая структура электромагнитного калориметра: результаты считывания замеров чувствительных элементов сенсора можно представить в виде матрицы $N \times N$ неотрицательных вещественных чисел, при этом сам калориметр состоит из 5×5 модулей. То есть если итоговая размерность калориметра это 40×40 , это на самом деле значит, что он состоит из 5×5 модулей, каждый из которых разбит на 8×8 единичных ячеек. Физический размер всего сенсора калориметра это $60.6 \text{ см} \times 60.6 \text{ см}$. Увеличение количества ячеек означает уменьшение размера каждой отдельной ячейки. Набор данных состоит из смоделированных симуляций, представляющих взаимодействие одного фотона с электромагнитным калориметром. Всего для каждого размера сенсора имеется по 64'000 примеров. Каждый элемент данных содержит такую информацию, как матрица неотрицательных значений (входные данные), а также положение входной точки фотона относительно центра сенсора и его соответствующая энергия в диапазоне от 1 до 100 ГэВ (целевые данные). Используя аналитические методы можно сделать сдвиг всей матрицы сенсора на необходимую позицию, что позволяет сделать предположение о том, что все фотоны прилетают в калориметр в его центральной клетке³. На Рисунке 3.1 показан пример входных данных и целевой позиции фотона. Симулирование данных было сделано с помощью инструмента GEANT4. В силу сложной структуры в которой хранятся результаты симуляций GEANT4, набор данных предварительно обрабатывался путем преобразования данных и

³В случае четной размерности сенсора и отсутствия центральной клетки, фотон сдвигался в правую нижнюю клетку, соседнюю с центром всего сенсора.

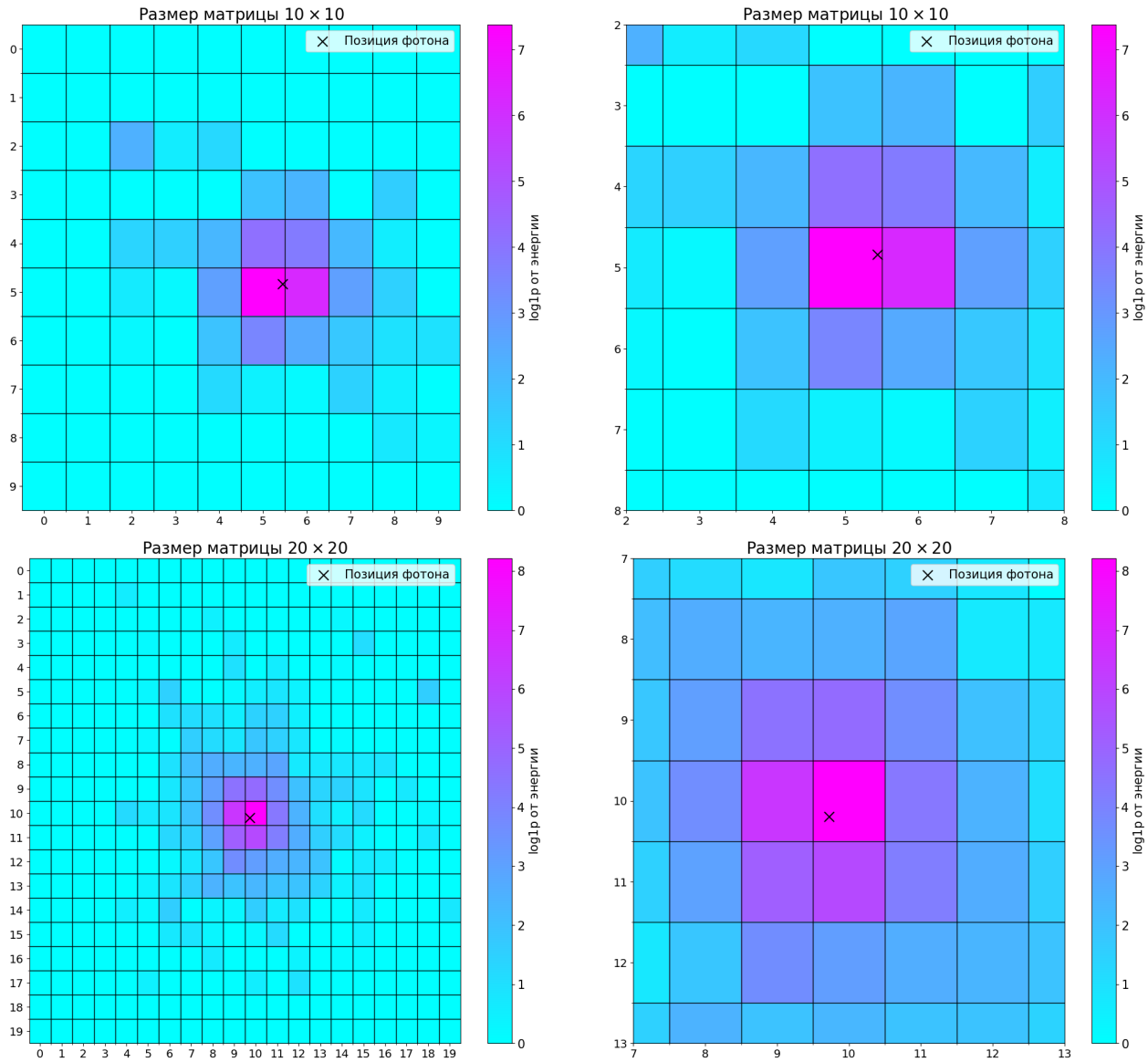


Рис. 3.1: Примеры данных, считанных калориметром для разрешений 10×10 и 20×20 . Слева показаны полноразмерные изображения, справа увеличенные центральные клетки. Настоящая позиция фотона обозначена черным крестиком. Для большей наглядности значения матриц показаны после применения функции $\log_{1p} = \log(1 + x)$.

их сохранения в виде NumPy массивов для ускорения загрузки во время обучения. Позиции входа фотонов для удобства нормализованы к диапазону $[0, 1]$. Значения энергии не нормируются и сохраняются в исходных единицах. Отношение обучающей и тестовой выборки выбрано равное 4 к 1.

3.2 Модели

Несколько моделей глубинного обучения оцениваются на предмет их эффективности при реконструкции энергии и позиции для разных размеров датчиков:

- **Аналитическая модель (AnaModel).** Эта модель используют простые аналитиче-

ские формулы для оценки целевых значений. Для задачи восстановления энергии она берет сумму значений всех ячеек матрицы и пропускает через один обучаемый линейный слой⁴. Для задачи реконструкции позиции модель создает в центре каждой ячейки калориметра точку с массой, равной значению считанной энергии данной ячейкой и вычисляет барицентр (центр масс) полученного множества взвешенных точек. Далее координаты полученной точки нормируются в нужный отрезок значений.

- **Линейная регрессия (LinReg).** Модель линейной регрессии обучается предсказывать целевые значения на основе входных признаков. Данная модель состоит из одного единственного обучаемого слоя, переводящего все элементы матрицы в три вещественных числа (одно для энергии и два для позиции фотона).
- **ResNet18.** Описание данной модели можно найти в Разделе 2.2. У базовой архитектуры ResNet был удален первый слой, так как он был рассчитан на изображения размера 224×224 , а в нашей задаче разрешение матриц сильно меньше и данный слой бы практически полностью терял информацию об исходных данных.
- **Простые сверточные сети (CNN).** Базовые архитектуры сверточных нейронных сетей изучаются на предмет их способности извлекать пространственные признаки.
- **Vision Transformer (ViT).** Рассматривается архитектура на основе трансформеров, адаптированная для задач зрения, чтобы использовать механизмы внимания для извлечения признаков (см. Раздел 2.2). Основным гиперпараметром модели является размер патча (patch size) на которые разбивается изображение, подающееся на вход модели. Было принято решение зафиксировать этот параметр равным 5 (то есть каждая матрица входных данных разбивается на подматрицы размера 5×5), потому что модель требует целочисленной делимости размера изображения на размер патча, а в нашей задаче по построению калориметр состоит из 5×5 модулей, а значит разрешение калориметра всегда делится на 5.

Современные модели глубинного обучения типично состоят из двух частей: основное тело (backbone) и голова (prediction head). Тело модели отвечает за извлечение сложных и показательных признаков из данных, а голова необходима для, собственно, решения задачи основываясь на этих полученных признаках. В данной работе предлагается одновременное решение поставленных задач (реконструкция энергии и реконструкция позиции фотона) с помощью добавления дополнительной головы к вышеуказанным моделям.

⁴Данный слой играет роль аффинного преобразования для подходящего выравнивания распределений.

3.3 Процедура обучения

Выбранные модели обучаются на подготовленном наборе данных. Процесс обучения включает в себя оптимизацию параметров модели для минимизации комбинированной функции потерь, которая представляет собой взвешенную сумму двух отдельных потерь: одной для восстановления энергии и другой для восстановления положения фотонов. Коэффициенты взвешенной суммы варьируются в разных экспериментах для изучения их влияния на производительность. Обучение проводилось с использованием оптимизатора AdamW, базовой длиной шага (learning rate) 0.007 и нулевым сокращением веса (weight decay). Использовалось косинусное расписание с прогревом (linear warmup with cosine annealing). Модели обучались в течение 100 эпох включая 15 эпох на прогрев расписания. Размер подвыборки (batch size) для шага оптимизации во всех экспериментах был 512.

3.4 Метрики качества

В данном исследовании рассматривались следующие функции потерь: Предположим, что дана выборка данных $(x_i, y_i)_{i=1}^n$, где x_i — входные данные и y_i — целевые значения. Далее пусть $\{a_i\}_{i=1}^n$ предсказания модели для этих данных. Тогда

- Корень из среднеквадратичной ошибки (RMSE):

$$\text{RMSE}(a, y) = \sqrt{\frac{1}{n} \sum_{i=1}^n (a_i - y_i)^2}.$$

- Средняя абсолютная ошибка (MAE):

$$\text{MAE}(a, y) = \frac{1}{n} \sum_{i=1}^n |a_i - y_i|.$$

- Корень из среднеквадратичной логарифмической ошибки (RMSLE):

$$\text{RMSLE}(a, y) = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(a_i + 1) - \log(y_i + 1))^2}.$$

- Взвешенный корень из среднеквадратичной ошибки (RMSE/E):

$$\text{RMSE/E}(a, y) = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{a_i - y_i}{y_i} \right)^2}.$$

- Взвешенная средняя абсолютная ошибка (MAE/E):

$$\text{MAE/E}(a, y) = \frac{1}{n} \sum_{i=1}^n \frac{|a_i - y_i|}{y_i}.$$

В качестве метрик итогового качества для реконструкции энергии и позиции были выбраны $\mathcal{L}_{\text{eng}} = \text{RMSE/E}$ и $\mathcal{L}_{\text{pos}} = \text{RMSE}$ соответственно. Эти же две функции используются как функции потерь для соответствующих задач если не сказано иное, а именно итоговой функцией обучения будет

$$\mathcal{L}_{\text{total}} = \alpha \cdot \mathcal{L}_{\text{eng}} + (1 - \alpha) \cdot \mathcal{L}_{\text{pos}} \quad (1)$$

где $\alpha \in [0, 1]$ — гиперпараметр важности той или иной задачи⁵. Дополнительно, через $\mathcal{L}_{\text{total}}^{\text{train}}, \mathcal{L}_{\text{eng}}^{\text{train}}, \mathcal{L}_{\text{pos}}^{\text{train}}$ и $\mathcal{L}_{\text{total}}^{\text{val}}, \mathcal{L}_{\text{eng}}^{\text{val}}, \mathcal{L}_{\text{pos}}^{\text{val}}$ будем обозначать значение данных метрик на обучающей и валидационной (тестовой) выборках соответственно. По умолчанию везде будет считаться, что $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{total}}^{\text{val}}, \mathcal{L}_{\text{eng}} = \mathcal{L}_{\text{eng}}^{\text{val}}$ и $\mathcal{L}_{\text{pos}} = \mathcal{L}_{\text{pos}}^{\text{val}}$.

Выбор взвешенной метрики RMSE/E для реконструкции энергии обоснован важностью относительной погрешности по сравнению с абсолютной погрешностью в экспериментах по физике высоких энергий. Относительная ошибка учитывает пропорциональное расхождение между предсказанными и фактическими значениями энергии, что имеет решающее значение для поддержания точности в широком диапазоне уровней энергии. В экспериментах, связанных со столкновениями частиц, где измерения энергии имеют фундаментальное значение для понимания свойств и взаимодействия частиц, точная количественная оценка относительных отклонений необходима для точного анализа и интерпретации. Поэтому, взвешивая метрику RMSE/E, мы отдаем приоритет относительной точности реконструкции энергии, обеспечивая надлежащий масштаб расхождений по отношению к величине энергии, тем самым повышая общую надежность и интерпретируемость результатов.

Выбор метрики RMSE для реконструкции позиции обоснован тем, что эта метрика совпадает с евклидовым расстоянием на плоскости (с точностью до мультипликативной константы) и является естественным показателем качества.

Во всех метриках чем меньше получаемое значение, тем лучше, при этом минимальное возможное значение ограничено 0.

⁵Увеличение α дает больший приоритет задаче восстановления энергии, в то время как уменьшение α способствует лучшему решению задачи реконструкции позиции.

3.5 Аугментации

В ходе исследования было принято решение использовать аугментации данных для улучшения качества моделей глубинного обучения. Аугментация данных — это процесс модификации имеющихся данных путем применения различных трансформаций, таких как повороты и отражения, с целью увеличения разнообразия обучающей выборки и улучшения обобщающей способности моделей.

Одной из примененных аугментаций было случайное горизонтальное или вертикальное отражение матрицы калориметра. Это позволяет модели обучаться на различных вариантах изображений фотона, которые могут встречаться в реальных условиях эксперимента. При этом точка, представляющая позицию фотона в центральной клетке калориметра, также подвергается отражению в соответствии с производимой трансформацией, чтобы сохранить соответствие между позицией фотона и его отражением.

Другой примененной аугментацией был случайный поворот матрицы калориметра на углы вида 0° , 90° , 180° и 270° . Эта трансформация позволяет модели обучаться на различных ориентациях фотона в калориметре и учитывать различные варианты расположения позиции фотона. Важно отметить, что при повороте матрицы точка в центральной клетке также подвергается повороту на соответствующий угол, чтобы корректно отобразить новое положение фотона после поворота. Поворот применялся только на углы, кратные 90° для того, чтобы физический поворот сенсора переводил его в себя без лишних перекрытий и выходов за границы.

Использование указанных аугментаций позволяет обогатить обучающий набор данных и сделать модели более устойчивыми к различным вариантам входных данных, что в свою очередь способствует улучшению их обобщающей способности и точности реконструкции энергии и позиции фотона в электромагнитном калориметре.

3.6 Постановка экспериментов

Были проведены множественные эксперименты, включающие в себя сравнение различных архитектур моделей на всех имеющихся размерах сенсоров, сравнение эффективности функций потерь, влияние сложности модели и аугментации данных на итоговое качество.

Для более качественного сравнения во всех экспериментах были зафиксированы все источники случайности, кроме инициализации весов-параметров моделей, а также случайности компьютерных компонент. Замеры для каждого эксперимента проводились по 8 раз и в качестве итогового результата бралось медианное значение.

3.7 Ресурсы

Для проведения экспериментов использовался кластер с AMD EPYC 9354 CPU и семью RTX4090 GPU. Суммарное время исполнения всех экспериментов оценивается в 90 часов.

3.8 Анализ результатов

Результаты, полученные в ходе экспериментов, анализируются с целью выявления наиболее эффективной модели и конфигурации для оптимизации размеров датчиков в электромагнитных калориметрах. Также исследуется влияние различных размеров датчиков на энергию и точность восстановления положения.

4 Результаты

В этом Разделе мы представляем результаты экспериментов по оптимизации размеров датчиков в электромагнитных калориметрах с использованием различных моделей глубокого обучения. Результаты анализируются на основе метрик качества, определенных в Разделе 3.4.

Сперва было проведено сравнение качества различных моделей на всех имеющихся размерах сенсоров. Как показывает Таблица 4.1, аналитическая модель, а также линейная регрессия уже дают достаточно хорошее качество реконструкции энергии фотонов, но при этом сильно отстают в восстановлении позиции. Стоит заметить, что аналитическая модель получает лучшее качество в реконструкции позиции, чем линейная регрессия, потому что она по построению учитывает физическое расположение ячеек сенсора. Модель ResNet18 оказывается слишком сложной для маленьких размеров сенсоров и сильно проигрывает даже аналитической модели, но при этом показывает весьма хорошие результаты на сенсорах более высокого разрешения. Намного более простой сверточной нейронной сети CNN удастся получить приемлемые результаты на сенсорах меньшего разрешения, но она не справляется с задачей реконструкции позиции на некоторых из размеров сенсоров. Единственная архитектура, которая показывает отличные результаты на обеих задачах и всех размерах сенсоров оказывается ViT, поэтому во всех последующих экспериментах анализировалась именно данная модель.

Улучшение производительности, наблюдаемое при использовании более сложных моделей, таких как ResNet18 и ViT, подчеркивает важность использования сложности модели

Модель	Метрика	Размер матрицы					
		10×10	15×15	20×20	25×25	30×30	40×40
AnaModel	$\mathcal{L}_{\text{total}}$	0.0626	0.0518	0.0469	0.0448	0.0440	0.0467
	\mathcal{L}_{eng}	0.0213	0.0221	0.0226	0.0229	0.0220	0.0222
	\mathcal{L}_{pos}	0.1040	0.0817	0.0714	0.0670	0.0662	0.0716
LinReg	$\mathcal{L}_{\text{total}}$	0.0728	0.0738	0.0741	0.0745	0.0748	0.0751
	\mathcal{L}_{eng}	0.0213	0.0221	0.0226	0.0231	0.0221	0.0223
	\mathcal{L}_{pos}	0.1244	0.1258	0.1257	0.1264	0.1278	0.1282
Resnet18	$\mathcal{L}_{\text{total}}$	0.0926	0.0823	0.0248	0.0254	0.0707	0.0271
	\mathcal{L}_{eng}	0.0814	0.0705	0.0215	0.0215	0.0565	0.0207
	\mathcal{L}_{pos}	0.1042	0.0941	0.0284	0.0297	0.0851	0.0336
CNN	$\mathcal{L}_{\text{total}}$	0.0229	0.0240	0.0682	0.5922	0.0260	0.1326
	\mathcal{L}_{eng}	0.0197	0.0200	0.0201	0.8934	0.0197	0.0196
	\mathcal{L}_{pos}	0.0262	0.0281	0.1166	0.2912	0.0324	0.2462
ViT	$\mathcal{L}_{\text{total}}$	0.0221	0.0232	0.0243	0.0242	0.0253	0.0264
	\mathcal{L}_{eng}	0.0194	0.0198	0.0203	0.0194	0.0196	0.0193
	\mathcal{L}_{pos}	0.0250	0.0270	0.0285	0.0295	0.0311	0.0336

Таблица 4.1: Сравнение качества моделей на различных размерах сенсоров. Каждая модель независимо обучалась для каждого размера сенсоров на обе задачи реконструкции одновременно.

для улавливания сложных закономерностей в данных. Эти модели обладают большей репрезентативной способностью, что позволяет им изучать иерархические особенности и глубокие зависимости, которые имеют решающее значение для точной реконструкции энергии и положения. В отличие от исключительно сверточных нейронных сетей, которые полагаются только на локальные рецептивные поля, Vision Transformers используют механизмы внимания для улавливания взаимодействий между удаленными областями калориметра. Это позволяет модели учитывать контекстную информацию из всего пространства входных данных, что приводит к принятию более обоснованных решений.

Далее было проведено сравнение качества обучения модели ViT на различные функции потерь для реконструкции энергии (функция потерь для восстановления позиции оставалась изначальной $\mathcal{L}_{\text{pos}} = \text{RMSE}$). Из Таблицы 4.2 можно видеть, что все нормализованные⁶ функции потерь RMSE/E, MAE/E и RMSLE приводят к заметно лучшим результатам, чем ненормализованная функция потерь RMSE. Более того, Рисунок 4.1 показывает, что наибольшие значения метрики качества (на графике показана метрика MAE/E) получаются на данных с наименьшими значениями энергии, при этом функция потерь RMSE/E приводит к более равномерному распределению, чем RMSE.

Следующим объектом исследования был гиперпараметр α из выражения (1), отвечающий за отношение функций потерь двух задач в итоговой функции потерь. Исходя из

⁶устойчивые к изменению магнитуды целевых значений

Функция потерь	Метрика	Размер матрицы					
		10×10	15×15	20×20	25×25	30×30	40×40
RMSE/E	\mathcal{L}_{eng}	0.0194	0.0198	0.0201	0.0192	0.0193	0.0195
	\mathcal{L}_{pos}	0.0250	0.0269	0.0284	0.0294	0.0310	0.0338
MAE/E	\mathcal{L}_{eng}	0.0198	0.0200	0.0204	0.0199	0.0199	0.0200
	\mathcal{L}_{pos}	0.0251	0.0269	0.0284	0.0294	0.0312	0.0336
RMSLE	\mathcal{L}_{eng}	0.0193	0.0196	0.0204	0.0191	0.0194	0.0194
	\mathcal{L}_{pos}	0.0249	0.0269	0.0285	0.0294	0.0310	0.0338
RMSE	\mathcal{L}_{eng}	0.0222	0.0222	0.0238	0.0214	0.0216	0.0219
	\mathcal{L}_{pos}	0.0290	0.0294	0.0332	0.0363	0.0356	0.0385

Таблица 4.2: Сравнение обучения модели ViT при различных функциях потерь для задачи реконструкции энергии фотона. Метрики качества остаются такими же, как было определено в Разделе 3.4.

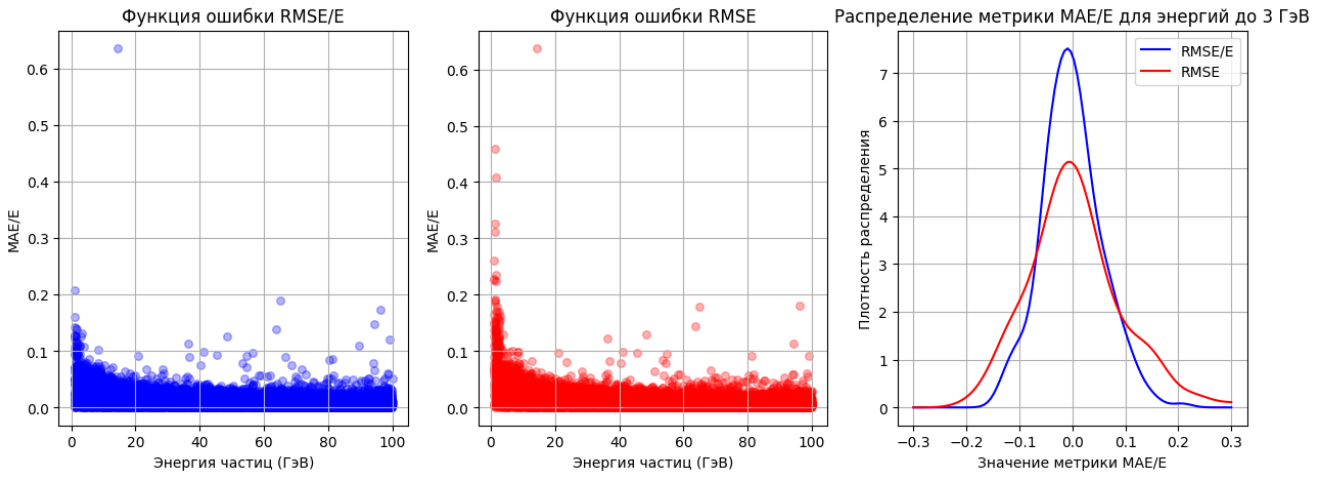


Рис. 4.1: Сравнение распределения метрики MAE/E на тестовых данных для модели ViT, обученной с помощью одной из двух функций потерь RMSE/E и RMSE в задаче реконструкции энергии.

Рисунка 4.2 можно заметить, что качество на задачах реконструкции энергии и реконструкции позиции и вправду ведет себя так же, как и соответствующее изменение гиперпараметра α (большой вес в функции потерь приводит к лучшему качеству на данной задаче). Также получается, что взятие $\alpha = 0.5$ приводит к лучшим результатам в совокупности двух задач. Также посмотрев на крайние значения α ⁷ видно, что обучение модели на обе задачи одновременно практически никак не сказывается на результатах в каждой задаче отдельно.

Очередным экспериментом было сравнение влияния размера модели на качество решения задач. На Рисунке 4.3 можно видеть, что чрезмерное увеличение размера модели плохо влияет на качество реконструкции позиции фотона, при этом качество восстановления энергии лишь немного ухудшается с увеличением размера модели. С другой стороны,

⁷На графиках не были указаны граничные случаи $\alpha = 0$ и $\alpha = 1$ потому что при данных значениях одна из функций потерь не обучалась вовсе и достигала слишком больших значений, что мешало визуальному сравнению более содержательных не граничных значений α .

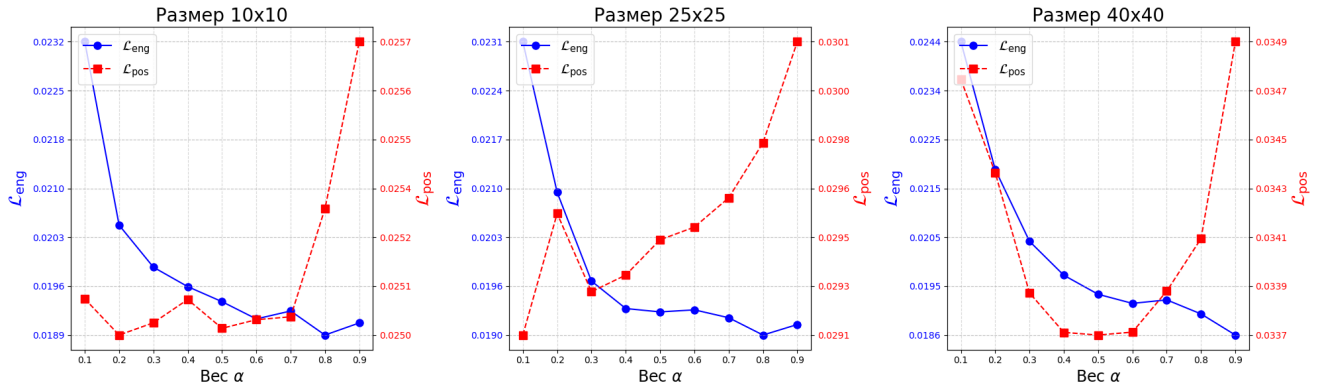


Рис. 4.2: Сравнение метрик качества модели ViT при подборе гиперпараметра α из выражения функции потерь (1).

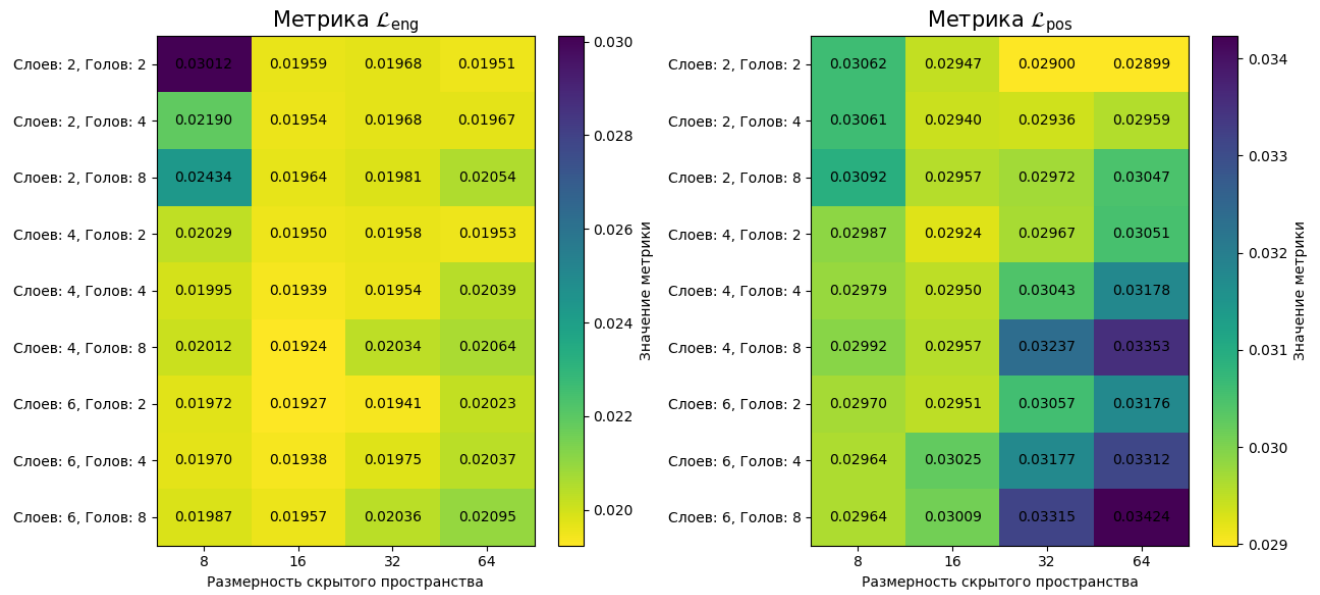


Рис. 4.3: Сравнение качества модели ViT на сенсоре размера 25×25 при изменении её параметров, таких как количество слоев, голов и размерность скрытого пространства (размерность MLP слоев была ровно в 4 раза больше размерности скрытого пространства).

слишком маленькие модели не способны выучиться решать задачу восстановления энергии.

Последним предметом экспериментов стало сравнение влияния аугментаций на итоговое качество модели. Судя по Таблице 4.3, применение аугментаций заметно улучшает качество для сенсоров размера 15×15 и немного для размера 25×25 . Более того, для обоих размеров аугментации приводят к сужению разрыва между качеством модели на обучающей и валидационной выборке, что показывает эффективность применяемых трансформаций.

В Таблице 4.4 указаны лучшие метрики качества, полученные после применения вышеописанных идей. Как можно видеть, модель ViT способна решать задачу реконструкции позиции с точностью, в 20 раз меньшую чем длина стороны центральной ячейки. Более того, данная модель решает задачу восстановления энергии с относительной ошибкой в 1.3%.

		Размер матрицы							
		15 × 15				25 × 25			
Отражения	Повороты	$\mathcal{L}_{\text{eng}}^{\text{train}}$	$\mathcal{L}_{\text{pos}}^{\text{train}}$	$\mathcal{L}_{\text{eng}}^{\text{val}}$	$\mathcal{L}_{\text{pos}}^{\text{val}}$	$\mathcal{L}_{\text{eng}}^{\text{train}}$	$\mathcal{L}_{\text{pos}}^{\text{train}}$	$\mathcal{L}_{\text{eng}}^{\text{val}}$	$\mathcal{L}_{\text{pos}}^{\text{val}}$
✗	✗	0.0182	0.0255	0.0217	0.0269	0.0181	0.0291	0.0192	0.0294
✓	✗	0.0186	0.0259	0.0194	0.0269	0.0186	0.0295	0.0191	0.0293
✗	✓	0.0185	0.0258	0.0193	0.0267	0.0186	0.0296	0.0189	0.0293
✓	✓	0.0185	0.0258	0.0192	0.0267	0.0186	0.0295	0.0189	0.0292

Таблица 4.3: Сравнение качества модели ViT при применении различных аугментаций (см. Раздел 3.5).

Метрика	Размер матрицы					
	10 × 10	15 × 15	20 × 20	25 × 25	30 × 30	40 × 40
$\mathcal{L}_{\text{total}}$	0.0219	0.0227	0.0239	0.0237	0.0249	0.0262
\mathcal{L}_{eng}	0.0189	0.0189	0.0194	0.0187	0.0189	0.0190
\mathcal{L}_{pos}	0.0247	0.0265	0.0283	0.0289	0.0308	0.0334
$\text{MAE}/E_{\text{eng}}$	0.0131	0.0127	0.0132	0.0124	0.0130	0.0129
$\bar{\mathcal{L}}_{\text{pos}}$	0.2117	0.1515	0.1211	0.0989	0.0881	0.0715
Размер одной ячейки	6.0600	4.0400	3.0300	2.4240	2.0200	1.5150

Таблица 4.4: Лучшее качество модели ViT на всех размерах сенсоров. Размер одной ячейки калориметра измеряется в сантиметрах. $\bar{\mathcal{L}}_{\text{pos}}$ — величина \mathcal{L}_{pos} , переведенная в сантиметры для сравнения с общей величиной центральной ячейки (эквивалентно евклидовому расстоянию между настоящей и предсказанной точками). Метрика $\text{MAE}/E_{\text{eng}}$ прямо соответствует модулю относительной ошибки.

Несмотря на многообещающие результаты, полученные в данном исследовании, необходимо признать его ограничения. Необходимы дальнейшие исследования для изучения устойчивости предложенных подходов в различных экспериментальных условиях и наборах данных. Кроме того, будущие направления исследований могут быть направлены на использование специфических знаний и методов адаптации к конкретной области для дальнейшего повышения эффективности моделей глубинного обучения при оптимизации размеров датчиков для электромагнитных калориметров.

5 Заключение

В заключение следует отметить, что данное исследование демонстрирует эффективность моделей глубинного обучения в оптимизации размеров датчиков для электромагнитных калориметров в экспериментах по физике высоких энергий. Благодаря всестороннему анализу различных архитектур моделей, включая ResNet18, сверточные нейронные сети (CNN) и модели, основанные на архитектуре трансформеров (ViT), мы показали, что использование передовых методов глубинного обучения может значительно повысить точность

восстановления энергии и положения.

Эксперименты, проведенные в этом исследовании, подчеркивают важность сложности модели, извлечения пространственных признаков и механизмов внимания при анализе размеров датчиков для электромагнитных калориметров. Сложные модели, такие как Vision Transformer, демонстрируют превосходную производительность, улавливая сложные паттерны и обширные зависимости в данных, что приводит к повышению точности реконструкции.

Кроме того, данное исследование подчеркивает компромисс между сложностью модели и вычислительной эффективностью, важность обобщения и робастности модели, а также важность интеграции знаний о конкретной области в процесс оптимизации.

В будущих исследованиях предполагается изучить новые архитектуры моделей, учесть специфические для данной области ограничения и оценить эффективность обобщения предложенных подходов в различных экспериментальных условиях. Решая эти задачи и используя последние достижения в области глубинного обучения и физики высоких энергий, можно еще больше повысить точность и эффективность моделей для решаемых задач на всех размерах датчиков для электромагнитных калориметров, что в конечном итоге улучшит наше понимание фундаментальных частиц и их взаимодействий.

В целом, данная работа вносит вклад в растущий объем исследований по оптимизации размеров датчиков в экспериментах по физике высоких энергий, предоставляя ценные идеи и прокладывая путь для будущих достижений в этой области.

6 Благодарности

7 Использование ИИ

В ходе проведения данного исследования были задействованы инструменты искусственного интеллекта, такие как ChatGPT и GitHub Copilot, для различных задач. ChatGPT был применен для автоматизации процессов создания кода для визуализации данных в виде графиков, форматирования таблиц в отчете с использованием языка разметки LaTeX, а также для получения консультаций по вопросам, связанным с библиотеками PyTorch и Scikit-learn. С другой стороны, GitHub Copilot использовался в процессе написания программного кода для проекта с целью создания документации и оптимизации повторяющихся участков кода. Основная часть программного кода была разработана без вмешательства генеративных моделей. Несмотря на то, что ChatGPT облегчал процесс создания графиков, было необходимо вручную корректировать результаты для достижения более точного отображе-

ния имеющихся данных.

Список литературы

- [1] N. Akchurin, C. Cowden, J. Damgov, A. Hussain и S. Kunori. “On the use of neural networks for energy reconstruction in high-granularity calorimeters”. в: *Journal of Instrumentation* 16.12 (дек. 2021), P12036. ISSN: 1748-0221. DOI: [10.1088/1748-0221/16/12/p12036](https://doi.org/10.1088/1748-0221/16/12/p12036). URL: <http://dx.doi.org/10.1088/1748-0221/16/12/P12036>.
- [2] Dawit Belayneh, Federico Carminati, Amir Farbin, Benjamin Hooberman, Gulrukh Khattak, Miaoyuan Liu, Junze Liu, Dominick Olivito, Vitória Barin Pacela, Maurizio Pierini, Alexander Schwing, Maria Spiropulu, Sofia Vallecorsa, Jean-Roch Vlimant, Wei Wei и Matt Zhang. “Calorimetry with deep learning: particle simulation and reconstruction for collider physics”. в: *The European Physical Journal C* 80.7 (июль 2020). ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-020-8251-9](https://doi.org/10.1140/epjc/s10052-020-8251-9). URL: <http://dx.doi.org/10.1140/epjc/s10052-020-8251-9>.
- [3] Francesco Armando Di Bello, Sanmay Ganguly, Eilam Gross, Marumi Kado, Michael Pitt, Lorenzo Santi и Jonathan Shlomi. “Towards a computer vision particle flow”. в: *The European Physical Journal C* 81.2 (февр. 2021). ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-021-08897-0](https://doi.org/10.1140/epjc/s10052-021-08897-0). URL: <http://dx.doi.org/10.1140/epjc/s10052-021-08897-0>.
- [4] A. Tumasyan и др. “Reconstruction of decays to merged photons using end-to-end deep learning with domain continuation in the CMS detector”. в: *Phys. Rev. D* 108 (5 сент. 2023), с. 052002. DOI: [10.1103/PhysRevD.108.052002](https://doi.org/10.1103/PhysRevD.108.052002). URL: <https://link.aps.org/doi/10.1103/PhysRevD.108.052002>.
- [5] F. Wemmer, I. Haide, J. Eppelt, T. Ferber, A. Beaubien, P. Branchini, M. Campajola, C. Cecchi, P. Cheema, G. De Nardo, C. Hearty, A. Kuzmin, S. Longo, E. Manoni, F. Meier, M. Merola, K. Miyabayashi, S. Moneta, M. Remnev, J. M. Roney, J.-G. Shiu, B. Shwartz, Y. Unno, R. van Tonder и R. Volpe. “Photon Reconstruction in the Belle II Calorimeter Using Graph Neural Networks”. в: *Computing and Software for Big Science* 7.1 (дек. 2023). ISSN: 2510-2044. DOI: [10.1007/s41781-023-00105-w](https://doi.org/10.1007/s41781-023-00105-w). URL: <http://dx.doi.org/10.1007/s41781-023-00105-w>.