

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 Используемые методы | 9 |
| 3 Методология | 9 |
| 3.1 Данные | 9 |
| 3.2 Модели | 10 |
| 3.3 Процедура обучения | 10 |
| 3.4 Метрики качества | 11 |
| 3.5 Аугментации | 12 |
| 3.6 Постановка экспериментов | 12 |
| 3.7 Ресурсы | 13 |
| 3.8 Анализ результатов | 13 |
| 4 Результаты | 13 |
| 5 Обсуждение | 17 |
| 6 Заключение | 19 |
| 7 Благодарности | 20 |
| 8 Использование ИИ | 20 |
| Список литературы | 21 |
| A Пример секции аппендикса | 22 |

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 калориметрических ячейках. Мы предлагаем использовать методы глубинного обучения для решения этой задачи и оценить различные размеры датчиков от 10 до 40×40 матриц. Наш подход предполагает обучение моделей глубокого обучения, включая архитектуры ResNet18 и Vision Transformer, и сравнение их производительности с базовыми моделями, такими как CatBoost и линейная регрессия. Мы используем такие метрики, как средняя квадратичная ошибка (MSE), средняя абсолютная ошибка (MAE), корневая средняя квадратичная ошибка (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 модели на производительность моделей глубинного обучения. Полученные нами результаты подчеркивают важность выбора оптимального размера датчика, 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 sizes ranging from 10×10 to 40×40 matrix shapes. Our approach involves training deep learning models, including ResNet18 and Vision Transformer architectures, and comparing their performance against baseline models such as CatBoost 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 sizes. 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 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 выделяются своей способностью измерять энергию 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 датчиками, обычно изготовленными из таких материалов, как кремний или сцинтилляционные кристаллы. Когда высокоэнергетические фотоны или электроны взаимодействуют с этими материалами, они создают электромагнитные ливни, характеризующиеся испусканием вторичных частиц. Энергия, накопленная этими частицами в калориметре, измеряется датчиками, 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 экономичными, особенно в крупномасштабных экспериментах. И наоборот, датчики больше-

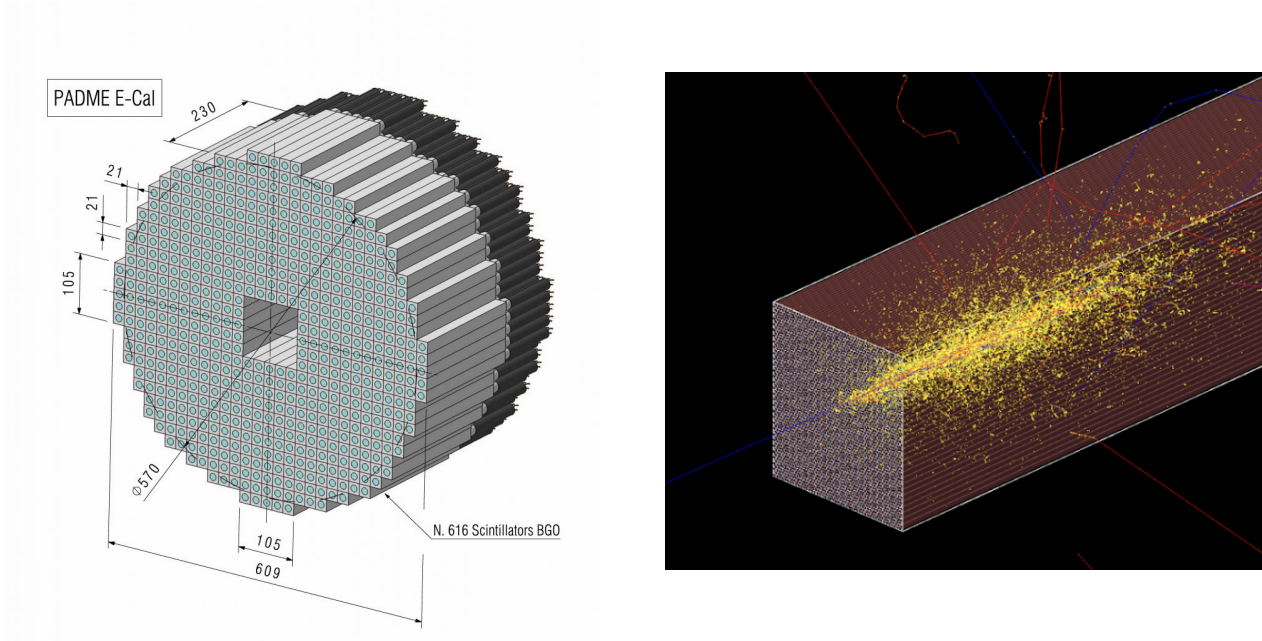


Рис. 1.1: Электромагнитный калориметр PADME и симуляция электромагнитного ливня с помощью GEANT4¹.

го 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 датчиков различных размеров от 10×10 до 40×40 матриц. 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 CatBoost 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 компромиссы, связанные с выбором оптимального размера датчика для электромагнитных калори-

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

метров 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.

Добавить описание структуры последующего текста работы

Хорошо оформить гитхаб с документацией и комментариями

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

Сказать, что в идеале модель должна давать одинаковое качество на всех размерах.

2 Обзор литературы

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 продемонстрировало 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 с сильным усилением Лоренца, минуя традиционные аналитические методы реконструкции. Их подход, проверенный на данных о столк-

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

новениях 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 глубоких 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 глубокого 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 глубокого обучения для усовершенствования алгоритмов реконструкции частиц.

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

2.1 Используемые методы

Добавить описание моделей ResNet и ViT

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

Добавить картинки с распределением данных

Для данного исследования была использована следующая структура электромагнитного калориметра: результаты считывания замеров датчиков 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. 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 положение входной точки фотона и его соответствующая энергия в диапазоне от 1 до 100 ГэВ. 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. 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 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.

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

3.2 Модели

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

Добавить больше деталей про размеры моделей и patch size = 5 для ViT

- **Аналитическая модель (AnaModel)**. Эта модель используют простые аналитические формулы для оценки целевых значений.
- **Линейная регрессия (LinReg)**. Модель линейной регрессии обучается предсказывать целевые значения на основе входных признаков.
- **ResNet18**. Архитектура остаточной нейронной сети используется для выявления сложных закономерностей в данных.
- **Простые сверточные сети (CNN)**. Базовые архитектуры конволюционных нейронных сетей изучаются на предмет их способности извлекать пространственные признаки.
- **Vision Transformer (ViT)**. Рассматриваются архитектуры на основе трансформеров, адаптированные для задач зрения, чтобы использовать механизмы внимания для извлечения признаков.

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.

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 learning rate. 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 задачи⁴. 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.

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

3.5 Аугментации

Добавить описание аугментаций и картинки

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.

⁴Увеличение α дает больший приоритет задаче восстановления энергии, в то время как уменьшение α способствует лучшему решению задачи реконструкции позиции.

3.7 Ресурсы

Для проведения экспериментов использовался кластер с AMD EPYC 9354 CPU и семью RTX4090 GPU. Суммарное время исполнения всех экспериментов оценивается в 60 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

Подробнее проанализировать результаты

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

Реализовать нахождение барицентра для модели AnaModel и добавить в таблицу

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. 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.

Показать, что RMSLE также устойчива при домножении на константу

Далее было проведено сравнение качества обучения модели ViT на различные функции потерь для реконструкции энергии (функция потерь для восстановления позиции оста-

| Модель | Метрика | Размер матрицы | | | | | |
|----------|------------------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| | | 10×10 | 15×15 | 20×20 | 25×25 | 30×30 | 40×40 |
| AnaModel | $\mathcal{L}_{\text{total}}$ | 0.1549 | 0.1555 | 0.1563 | 0.1575 | 0.1560 | 0.1561 |
| | \mathcal{L}_{eng} | 0.0213 | 0.0221 | 0.0226 | 0.0229 | 0.0220 | 0.0222 |
| | \mathcal{L}_{pos} | 0.2887 | 0.2891 | 0.2901 | 0.2924 | 0.2901 | 0.2901 |
| 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.

валась изначальной $\mathcal{L}_{\text{pos}} = \text{RMSE}$). Из Таблицы 4.2 можно видеть, что все нормализованные⁵ 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 .

| 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.

Следующим объектом исследования был гиперпараметр α из выражения (1), отве-

⁵устойчивые к изменению магнитуды целевых значений

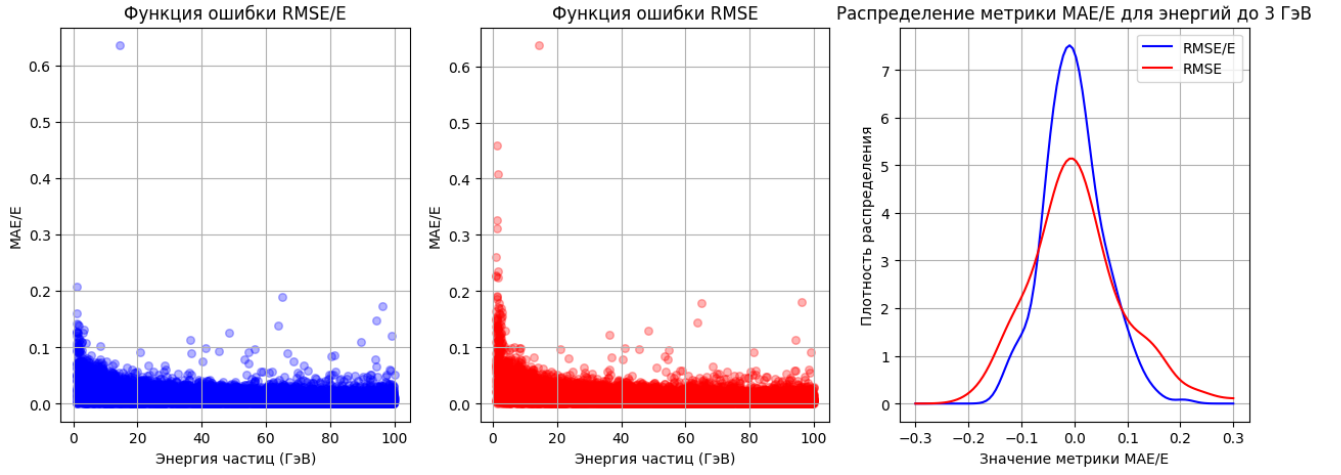


Рис. 4.1: Сравнение распределения метрики MAE/E на тестовых данных для модели ViT, обученной с помощью одной из двух функций потерь RMSE/E и RMSE в задаче реконструкции энергии.

чающий за отношение функций потерь двух задач в итоговой функции потерь. Исходя из 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 ⁶ видно, что обучение модели на обе задачи одновременно практически никак не сказывается на результатах в каждой задаче отдельно.

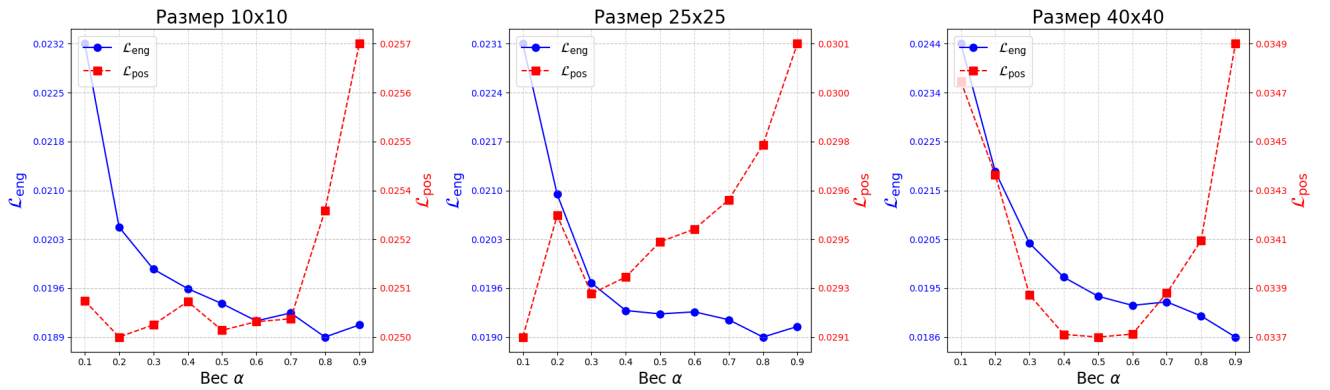


Рис. 4.2: Сравнение метрик качества модели ViT при подборе гиперпараметра α из выражения функции потерь (1).

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 восстано-

⁶На 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 .

ния энергии лишь немного ухудшается с увеличением размера модели. С другой стороны, слишком маленькие модели не способны выучиться решать задачу восстановления энергии.

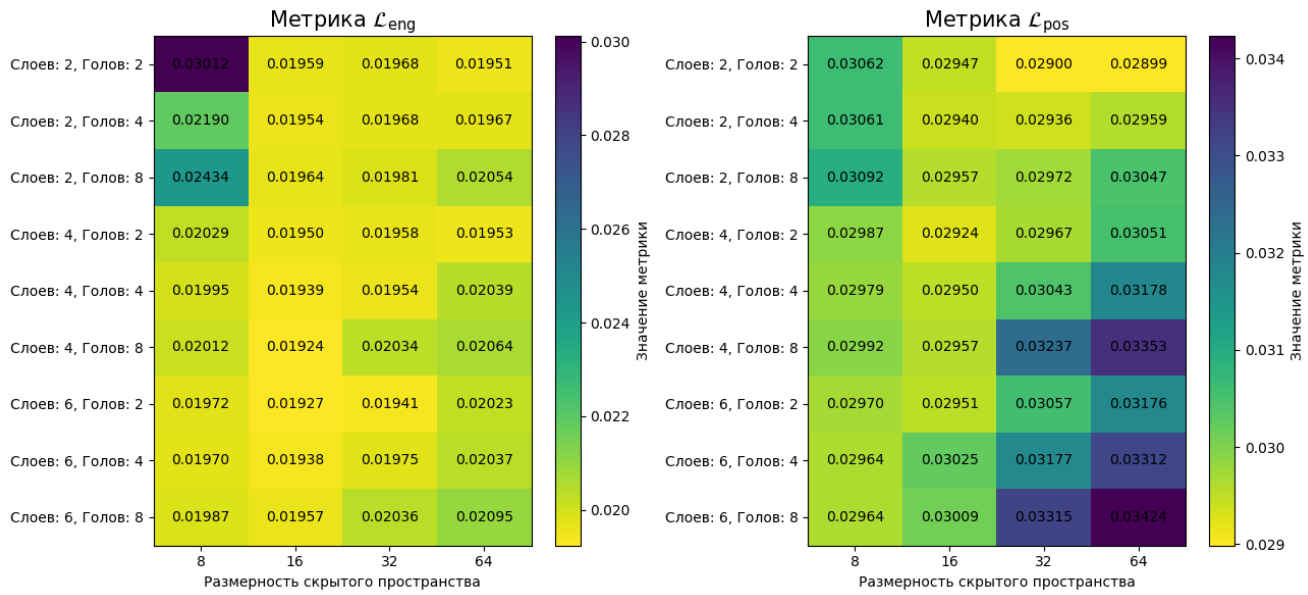


Рис. 4.3: Сравнение качества модели ViT на сенсоре размера 25×25 при изменении её параметров, таких как количество слоев, голов и размерность скрытого пространства (размерность MLP слоев была ровно в 4 раза больше размерности скрытого пространства).

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.

| | | Размер матрицы | | | | | | | |
|------------|----------|-----------------------------|-----------------------------|---------------------------|---------------------------|-----------------------------|-----------------------------|---------------------------|---------------------------|
| | | 15 × 15 | | | | 25 × 25 | | | |
| Перевороты | Повороты | $\mathcal{L}_{eng}^{train}$ | $\mathcal{L}_{pos}^{train}$ | \mathcal{L}_{eng}^{val} | \mathcal{L}_{pos}^{val} | $\mathcal{L}_{eng}^{train}$ | $\mathcal{L}_{pos}^{train}$ | \mathcal{L}_{eng}^{val} | \mathcal{L}_{pos}^{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).

Показать результат лучшей модели для каждого размера со всеми метриками

Сравнить свои результаты с результатами каких-то из статей

Несмотря на многообещающие результаты, полученные в данном исследовании, необходимо признать его ограниченность. 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 Обсуждение

Результаты, полученные в ходе экспериментов, проливают свет на эффективность различных моделей глубинного обучения в оптимизации размеров датчиков для электромагнитных калориметров. В этом разделе мы подробно обсудим последствия наших результатов и выясним основные факторы, влияющие на производительность каждого типа моделей.

1. Использование сложности модели

Улучшение производительности, наблюдаемое при использовании более сложных моделей, таких как ResNet и трансформаторы зрения, подчеркивает важность использования сложности модели для улавливания сложных закономерностей в данных. Эти модели обладают большей репрезентативной способностью, что позволяет им изучать иерархические особенности и дальние зависимости, которые имеют решающее значение для точного восстановления энергии и положения. Благодаря использованию остаточных связей и механизмов самовнушения архитектуры ResNet и трансформаторов зрения демонстрируют более высокую производительность по сравнению с более простыми моделями, что подчеркивает важность глубины и выразительности модели при оптимизации размеров датчиков.

2. Извлечение пространственных признаков

Эффективность сверточных нейронных сетей (CNN) в захвате пространственных характеристик подчеркивает важность пространственной локальности при оптимизации размеров датчиков для электромагнитных калориметров. CNN отлично справляются с извлечением локальных паттернов и пространственных зависимостей из данных, что позволяет точно восстановить энергию фотонов и положение входа. Использование конволюционных слоев облегчает изучение особенностей различных областей калориметра, повышая способность модели различать тонкие изменения в осаждении энергии. Таким образом, извлечение пространственных признаков играет решающую роль в достижении точной оптимизации размера датчика.

3. Роль механизмов самозахвата

Превосходная работа трансформаторов зрения подчеркивает важность механизмов самовнимания для улавливания глобальных закономерностей и дальних зависимостей в дан-

ных. В отличие от CNN, которые полагаются исключительно на локальные рецептивные поля, трансформаторы зрения используют механизмы самовнимания 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 по оптимизации размера датчика. Способность трансформаторов зрения обращать внимание на соответствующие области калориметра повышает чувствительность модели к тонким закономерностям осаждения энергии, что приводит к повышению точности реконструкции.

4. Компромисс между сложностью и эффективностью модели

Хотя сложные модели, такие как ResNet и трансформаторы зрения, обеспечивают превосходную производительность, они часто обходятся повышенной вычислительной сложностью и требованиями к ресурсам. Таким образом, существует компромисс между сложностью модели и вычислительной эффективностью в контексте оптимизации размера датчика. Исследователи должны тщательно учитывать вычислительные ограничения и практическую целесообразность применения сложных моделей в реальных приложениях. Такие стратегии, как обрезка моделей, квантование и оптимизация архитектуры, могут быть использованы для уменьшения вычислительных затрат при сохранении производительности.

5. Обобщение и устойчивость

Важным аспектом оптимизации размера сенсора является обобщение и устойчивость моделей к различным экспериментальным условиям и наборам данных. Хотя эксперименты демонстрируют многообещающие результаты в контролируемых условиях, крайне важно оценить эффективность обобщения моделей в различных сценариях. Будущие исследования должны быть направлены на оценку устойчивости предложенных подходов в различных диапазонах энергий, геометрии детекторов и экспериментальных условиях. Для улучшения обобщения моделей и облегчения переноса знаний между областями могут быть использованы такие методы, как адаптация к области и трансферное обучение.

6. Интеграция со специфическими знаниями домена

Интеграция знаний о конкретной области и физических ограничениях в процесс разработки модели может дополнительно повысить точность и интерпретируемость алгоритмов оптимизации размеров датчиков. Интегрируя специфические для данной области знания в архитектуру модели и функцию потерь, исследователи могут использовать предыдущие знания о взаимодействии частиц и характеристиках датчиков, чтобы направлять процесс оптимизации. Включение ограничений, специфичных для данной области, гарантирует, что оптимизированные размеры датчиков будут соответствовать физическим принципам и экс-

периментальным требованиям, что приведет к более значимым и практичным результатам.

6 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 мы намерены 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 глубокого обучения и физики высоких энергий, мы сможем еще больше повысить точность и эффективность оптимизации размеров 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.

7 Благодарности

8 Использование ИИ

Написать где использовал генеративные модели.

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>.

А Пример секции аппендикса