

**Московский государственный университет
имени М.В. Ломоносова
Химический факультет
Кафедра физической химии
Лаборатория строения конденсированных систем**

ХАЙБРАХМАНОВ АРТУР ИЛЬНУРОВИЧ

**ПРИМЕНЕНИЕ МЕТОДОВ ИИ ДЛЯ РЕШЕНИЯ ЗАДАЧ
РЕНТГЕНОДИФРАКЦИОННЫХ ИССЛЕДОВАНИЙ**

ДИПЛОМНАЯ РАБОТА

Научный руководитель:
к. х. н., с. н. с. Дмитриенко А. О.
д. х. н., доцент Лысенко К. С.

Москва, 2025

Оглавление

Введение	3
1 Обзор литературы	4
1.1 Рентгеновская кристаллография	4
1.2 Проблема фаз	4
2 Методология	7
2.1 Данные	7
2.2 Методы	9
3 Результаты и обсуждение	14
4 Выводы	21
Список литературы	22

Введение

Решение проблемы фаз является важной задачей рентгеноструктурного анализа, особенно актуальной для белковой кристаллографии ввиду отсутствия *ab initio* решений в этой области. Методы машинного обучения способны преодолеть данную задачу. Предлагается увеличивать разрешение дифракционной картины, предсказывая моделью машинного обучения дальние отражения по ближним, что позволит решить проблему фаз *ab initio* для биомолекул. В работе разработан генератор синтетических рентгенодифракционных данных, который может быть использован для решения прикладных задач методами ИИ, и пайплайн для проведения воспроизводимых экспериментов для решения задачи в рамках предложенной методологии. Также были разработаны модели FFT_UNet и XRD_Transformer, подходящие под специфику задачи; проведены их сравнительный анализ и интерпретация их работы с помощью GradCAM и связей внимания. Было показано и обосновано, что методы глубокого обучения способны считывать кристаллографические связи и законы в рентгенодифракционных данных, но их точности численного восстановления амплитуд структурных факторов не хватает для решения проблемы фаз в рамках предложенной методологии.

1 Обзор литературы

1.1 Рентгеновская кристаллография

Рентгеновская кристаллография является важнейшим инструментом определения трехмерных структур кристаллов. Этот метод позволяет получить бесценные сведения об атомных и молекулярных структурах кристаллов, что крайне важно для понимания свойств и функций материалов в различных областях, включая химию, биологию и материаловедение. Рентгеноструктурный анализ основан на упругом рассеянии монохроматического рентгеновского излучения на трехмерной регулярной решетке атомов твердого вещества, что приводит к интерференции рентгеновских лучей.

Дифракция в кристалле описывается как отражения от семейств параллельных кристаллографических плоскостей элементарной ячейки кристалла. Каждая точка дифракционной картины задается набором из индексов Миллера h, k, l . Любое отражение имеет математическое представление, известное как структурный фактор $F(h, k, l)$, зависящий от расположения и коэффициентов рассеяния всех атомов в элементарной ячейке кристалла. Структура кристалла получается при расчете электронной плотности с помощью Фурье-преобразования:

$$\rho(x, y, z) = V^* \sum_{h,k,l} F(h, k, l) \exp[-2\pi i(hx + ky + lz)],$$

где V^* – объем элементарной обратной ячейки [1].

1.2 Проблема фаз

Структурный фактор является комплексной величиной, и её амплитуда $|F(h, k, l)|$ легко определяется по результатам обычного рентгенодифракционного эксперимента как квадратный корень из зарегистрированной интенсивности дифрагированной волны, а информация о фазе теряется. Данная проблема получила название "проблема фаз". Без знания о фазе процесс восстановления распределения электронной плотности из дифракционных данных

затруднен.

Проблема фаз давно является одной из главных задач рентгеновской кристаллографии. Для ее решения были разработаны традиционные методы, такие как прямые [2] и charge-flipping [3], но они, как правило, ограничены атомным разрешением рентгенодифракционных данных. Эти методы требуют полного набора дифракционных пиков и часто требуют выращивания высококачественных образцов кристаллов, что может быть сложной задачей, особенно для слабо рассеивающих кристаллов или больших молекул, таких как белки [4]. Для определения структур макромолекулярных соединений разработаны методы молекулярного замещения и изоморфного замещения, требующие дополнительную информацию — знание о структуре белка с той же аминокислотной последовательностью или результаты рентгенодифракционных экспериментов той же структуры с добавлением тяжелых атомов, соответственно [5]. Таким образом, решение проблемы фаз является особенно актуальной задачей белковой кристаллографии из-за отсутствия *ab initio* решений.

Применение машинного обучения в области рентгеновской дифракции лишь недавно зародилось и сейчас бурно развивается. Чаще всего авторы пытаются обойти решение проблемы фаз, предсказывая кристаллическую структуру из доступных экспериментальных данных. Так, были разработаны модели глубокого обучения, позволяющие миновать проблему фаз, работая напрямую с функцией Паттерсона, которые получаются из дифракционных данных, не требуя информации о фазах отражений [6]. Эти модели представляют собой сверточные нейронные сети для предсказания распределения электронной плотности структуры по функции Паттерсона, демонстрируя многообещающие результаты на простых примерах дипептидов и подчеркивая потенциал для более широкого применения в белковой кристаллографии.

В недавней статье [7] с помощью методов глубокого обучения авторы предприняли попытку решить проблему фаз. В качестве объектов предсказания они выбрали центросимметричные структуры, для которых фазы отражений принимают два возможных значения — 0 и 1. Авторы презентовали нейронную сеть, представляющую собой

бинарный классификатор из блоков трехмерных свёрток и многослойных перцептронов. Обучение проводилось на синтетических кристаллических молекулярных структурах. Этот подход продемонстрировал способность решать фазовую задачу с разрешением всего 2\AA , используя всего 10–20% данных, требуемых традиционными методами. Также в работе реализована идея phase recycling — исходные данные прогоняются несколько раз через модель, что увеличивает точность классификации. Таким образом, впервые был продемонстрирован потенциал машинного обучения для решения проблемы фаз, но только для центросимметричных структур.

На основе методов глубокого обучения были также предложены решения аналогичной задачи в физике, особенно в области когерентной безлинзовой микроскопии. В обзорной статье [8] выделяются три подхода: DL-post-processing, где уточняются «плохие фазы», полученные из исходных интенсивностей; DL-in-processing, в котором нейронная сеть используется для непосредственного расчета фаз из интенсивностей; и DL-pre-processing, при котором обученная модель повышает разрешение микроскопического изображения, а фазы из полученного изображения определяются детерминированными методами. Однако, в кристаллографии дифракционные данные имеют гораздо меньшее разрешение, чем в когерентной микроскопии, что исключает возможность прямого применения существующих решений.

Таким образом, решение фазовой задачи белковой кристаллографии является актуальной задачей, нерешаемой рутинными методами. Инструменты на основе методов глубокого обучения способны преодолеть ограничения традиционных подходов, позволяя определять кристаллические структуры на основе ограниченных данных с низким разрешением. По мере развития исследований в этой области, вероятно, алгоритмы глубокого обучения станут незаменимым инструментом в области кристаллографии, облегчая решение сложных структур, которые ранее были неразрешимыми. Таким образом, разработка методов решения проблемы фаз с помощью ИИ является целью работы.

2 Методология

2.1 Данные

Было разработано программное обеспечение, позволяющие генерировать рентгенодифракционные данные случайных органических структур (github.com/blackwood168/xrd_simulator). С помощью библиотеки CCTBX (Computational Crystallography Toolbox [9]) создаются кристаллические решетки, в которых случайным образом расставлены атомы, и рассчитываются структурные факторы. Также генератор поддерживает последующий расчет данных рентгеновской порошковой дифракции. Так, можно получить достоверные дифрактограммы синтетических структур, в которых реализована профильная функция Псевдо-Войдта и аксиальная расходимость пучка. Данный генератор может быть полезен для генерации рентгенодифракционных синтетических данных для использования в различных прикладных задачах машинного обучения в данной области.

Начальное обучение моделей глубокого обучения было решено проводить на синтетических данных, которые могут быть рассчитаны с помощью разработанного генератора. Высокое разрешение выбрано 0.8 \AA , низкое — 1.5 \AA . Рассматриваются только симметрийно независимые отражения, так как данные о симметрии известны и без решения проблемы фаз. Так как общее количество симметрийно независимых отражений структур (для ячеек одного размера) зависит от сингонии (класса кристаллических решеток), все структуры в данных были моноклинными. Таков выбор неслучаен — моноклиновые группы симметрии являются одними из наиболее распространенными для белковых структур в базе данных белков (rcsb.org/stats/distribution-space-group). Соответственно группы симметрии — $P2_1$, $C2$. Типы атомов — C, N, O, Cl; число симметрийно независимых атомов 10–30. Для обучения были сгенерированы 400.000, 100.000 и 100.000 структур для тренировочной, валидационной и тестовой выборок, соответственно.

Также в работе использовались реальные моноклинные молекулярные структуры малых молекул из Кембриджского Банка Структурных Данных [10], для которых были расчитаны дифракционные отражения. 10.000 структур использовались для дообучения моделей на реальных структурах, 2000 — были отложены для тестирования.

Так как отражения являются точками обратного пространства, каждое из них можно однозначно описать индексами Миллера (h,k,l) . Тогда дифракционную картину можно описать трехмерным тензором, в котором записаны данные каждого отражения. Размер тензора был выбран $(26,18,23)$ — в матрицу такого размера помещаются все отражения для самой большой моноклинной структуры из синтетических данных. Соответственно для обучения данные об отражениях были помещены в тензор, в позициях (h,k,l) соответственно, где дифракционные максимумы отсутствуют — стоят нули.

Для выполнения работы валидным является предсказание как интенсивности, так и амплитуды структурных факторов ($|F| = \sqrt{I}$). Типичные распределения этих данных для сгенерированных структур представлены на рис. 2.1. Как можно заметить, распределение амплитуд больше похоже на стандартное, поэтому именно амплитуды были выбраны для решения задачи. Дифракционные данные также были отнормированы в диапазон 0-1.

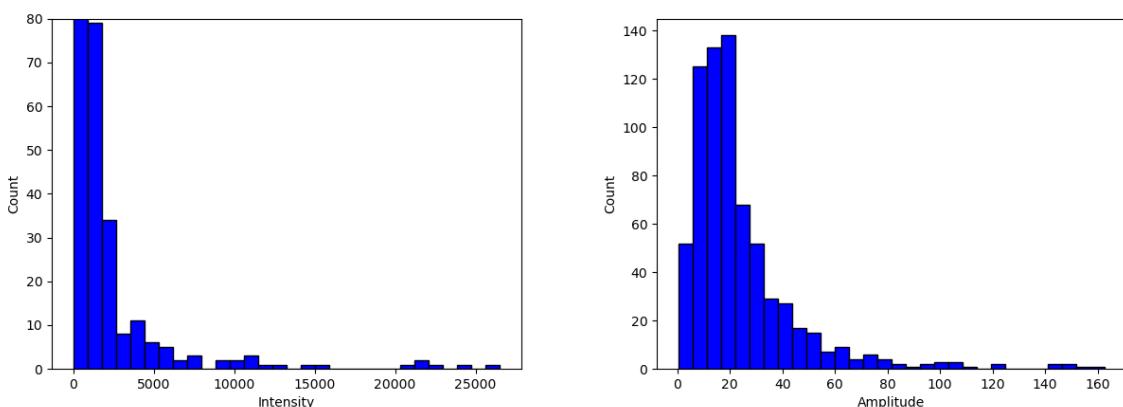


Рисунок 2.1 – Типичные распределения интенсивностей (слева) и амплитуд (справа) дифракционной картины

Основная же сложность работать с чистыми амплитудами

структурных факторов — то есть с точками в обратном пространстве — заключается в том, что данные не являются локально связанными, как изображения (рис. 2.2). Это вносит специфику в данную задачу и создает трудности, так как типичные архитектуры моделей для работы с трехмерными данными, например, видео или медицинскими данными [11] могут не сработать.

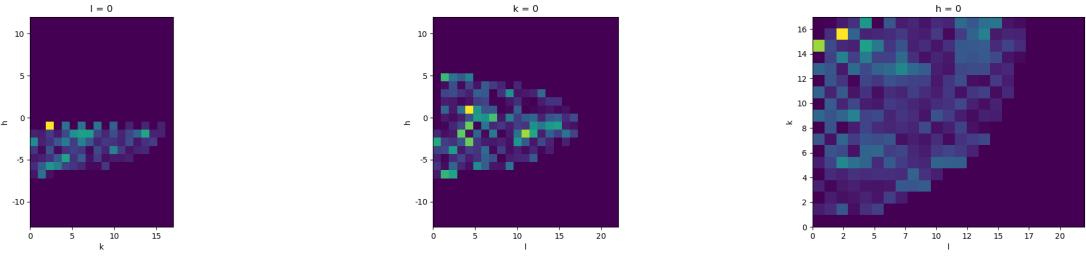


Рисунок 2.2 – Типичные сечения тензора отражений для реальных структур из CSD

2.2 Методы

работе было предложено предсказывать амплитуды дифракционных отражений белков, которые нельзя получить из эксперимента, по известным из того же эксперимента. После предсказания достаточного количества рефлексов, разрешения должно хватить для определения фаз и расчета электронной плотности одним из рутинных *ab initio* методов, в качестве которого был выбран SHELLX [12].

Таким образом, задача сводится к восстановлению трехмерного тензора. Inference моделей глубокого обучения должен выглядеть следующим образом: на вход подается тензор с рентгенодифракционными экспериментальными данными, на выходе должен быть тензор с дополнительными интенсивностями. В ходе обучения планируется научить модель восстанавливать тензор отражений по данным малых органических молекул. Для этого будут обнулены интенсивности дальних отражений так, чтобы длина разрешения полученной дифракционной картины соответствовала типичному разрешению белковых соединений (1.5 \AA).

В качестве baseline была обучена модель UNet, адаптированная для

трехмерных тензоров (рис. 2.3). Данная модель выбрана потому что она хорошо себя проявляет в простейших задачах Super Resolution, однако в данной задаче она не достигнет высокой точности из-за локальной несвязанности рентгенодифракционных данных.

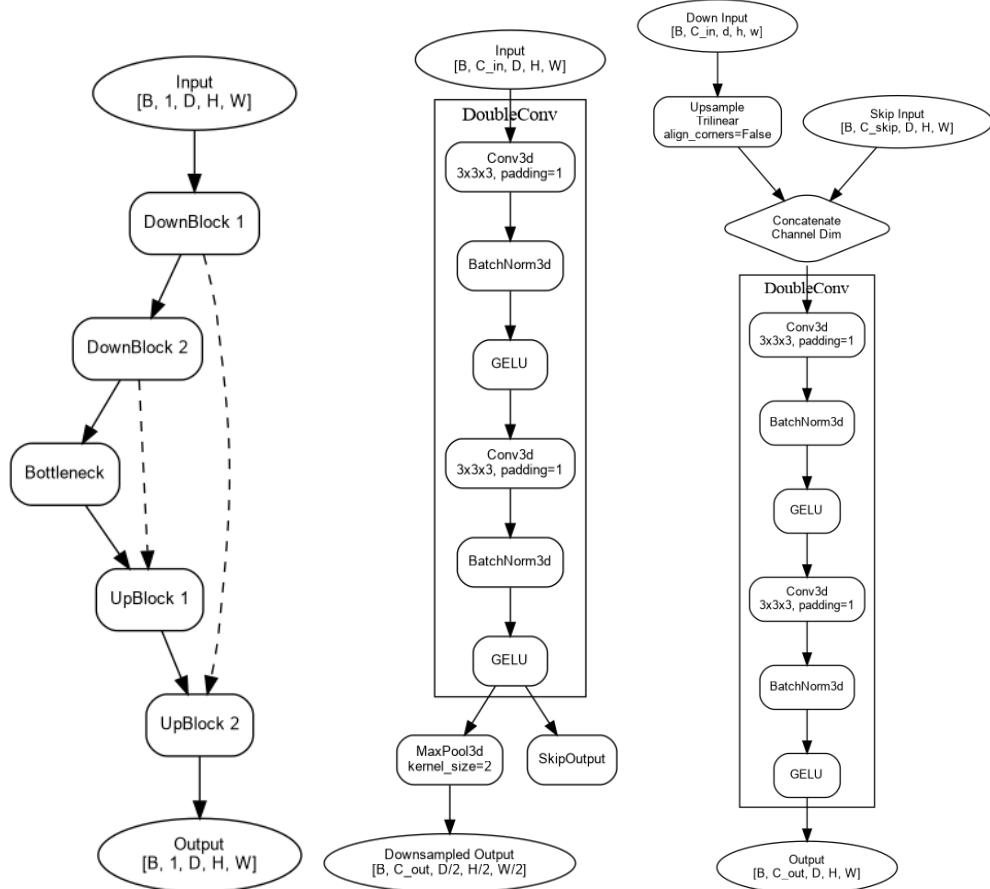
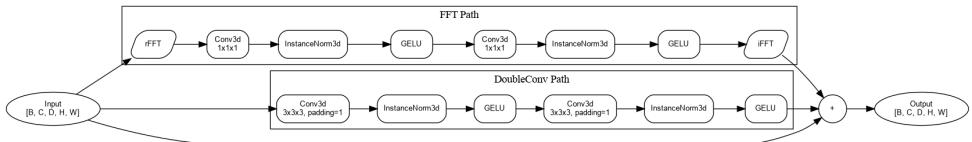
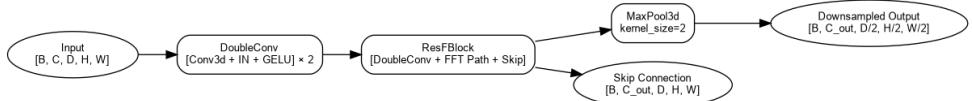


Рисунок 2.3 – Схемы архитектуры модели UNet (слева) и её DownBlock(посередине) и UpBlock(справа)

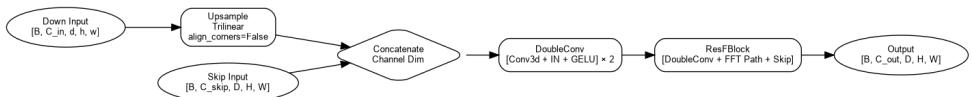
Также была разработана и обучена UNet-like модель с кастомными слоями, содержащими Фурье-преобразование (рис. 2.4). Данный подход был продемонстрирован в работе [13] при работе с дифракционными данными и он является многообещающим и для нашей задачи, поскольку при переходе в прямое пространство наши данные являются локально связанными.



(a) Кастомный слой с Фурье-преобразованием



(b) DownBlock



(c) UpBlock

Рисунок 2.4 – Кастомная FFT_UNet модель и её основные блоки

Поскольку тензоры дифракционных картин не являются локально связанными, актуально использование механизма внимания. Он позволит модели находить связи между не близколежащими отражениями. Так, был разработан трансформер XRD_Transformer для нашей задачи (рис. 2.5).

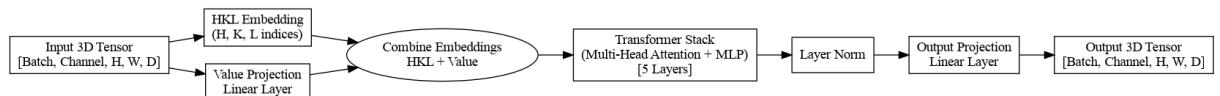


Рисунок 2.5 – Архитектура модели XRD_Transformer

В модели формируется единый эмбеддинг из эмбеддинга индексов Миллера (h, k, l) и проекции значений амплитуды отражений. Затем эмбеддинг подается в 5 слоев трансформера, которые состоят из Multi-Head Attention и полносвязного слоев. Затем после нормализации и обратного проецирования в исходное пространство получается восстановленный тензор дифракционной картины.

Так как нам известны все возможные значения индексов Миллера (h, k, l) эмбеддинг (рис. 2.6) формируется следующим образом: каждый индекс кодируется one-hot энкодингом. Затем каждый вектор с помощью обучаемого линейного слоя проецируется к размерности

`embed_dim/3`; после конкатенации получается эмбеддинг позиции отражения размерности `embed_dim`. Также в модели реализована возможность эмбеддинга через полносвязный слой, который проецирует позицию (h,k,l) сразу в вектор эмбеддинга, однако она не использовалась при обучении модели, так как первый способ является более физическим для нашей задачи, поскольку мы используем только симметрийно независимые отражения.

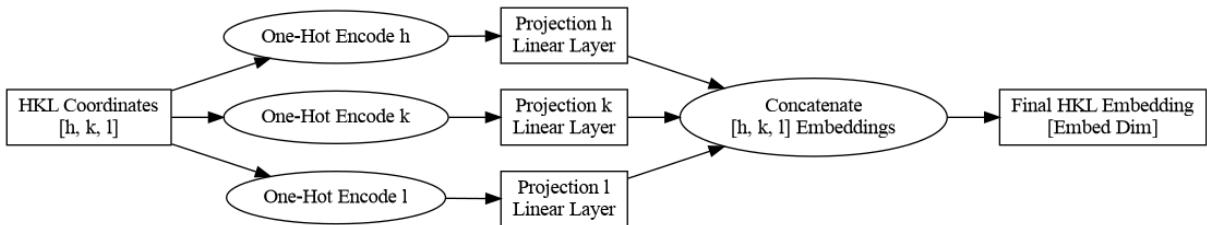


Рисунок 2.6 – Схема формирования эмбеддинга HKL

В ходе выполнения работы также предложен постпроцессинг, включающий в себя учёт систематических погасаний — "зануления" некоторых значений интенсивностей, что определяется симметрией структуры; в ходе обучения происходит явное восстановления части тензора, которую не нужно предсказывать.

Для обучения моделей UNet и FFT_UNet использовался оптимизатор Adam с планировщиком plateau scheduler (обучение: learning rate = 0.01, factor = 0.1; дообучение: learning rate = 0.0001, factor = 0.5); для XRD_Transformer — оптимизатор AdamW с планировщиком cosine annealing (обучение: learning rate = 0.01; дообучение: learning rate = 0.0001).

В качестве функции потерь была выбрана среднеквадратичная ошибка, минимизация которой должна приводить к восстановлению тензора рентгеновских отражений. В качестве метрики для анализа используется R-фактор: $R = \frac{\sum |F_{obs} - F_{calc}|}{\sum |F_{obs}|}$, где $|F_{obs}|$ — экспериментальные структурные факторы, $|F_{calc}|$ — рассчитанные структурные факторы. R-фактор является общепринятым стандартом в кристаллографическом сообществе для оценки качества структурных моделей. Нулевое значение R-фактора отвечает идеальному соответствуанию между данными модельной структуры и экспериментальными данными.

Также в качестве метрики рассматривался SSIM, однако эксперименты с ним в качестве добавки к функции потери привели к более низкому качеству восстановления модели с точки зрения R-фактора. Данный результат можно объяснить отсутствием локальности наших данных.

Эффективность предсказания обученных моделей глубокого обучения проверялась на тестовой части синтетического датасета, а также тестовой части рентгенодифракционных данных моноклинных структур из CSD.

3 Результаты и обсуждение

В рамках данной работы был разработан пайплайн, позволяющий проводить воспроизводимые эксперименты по решению проблемы фаз с помощью предложенной методологии. (github.com/blackwood168/xrd_phase_ml). В нем реализовано обучение и тестирование моделей, а также inference на реальных массивах данных и структурах кристаллических соединений. В репозитории присутствуют маленькие датасеты из сгенерированных и реальных структур малых органических молекул. Также там представлены веса обученных в работе моделей. Воспроизводимость обучения обеспечивает фиксирование начальных значений генераторов случайных состояний.

Результаты обучения

Таблица 3.1 – Эффективность моделей до и после дообучения на синтетическом и реальном датасете

Model	Metric	Synth	CSD
UNet	Before, R	0.477	0.590
	After, R	0.632	0.393
	$\Delta, \%$	-32.6	33.4
FFT_UNet	Before, R	0.726	0.646
	After, R	0.619	0.336
	$\Delta, \%$	14.7	48.0
XRD_Transformer	Before, R	0.346	0.581
	After, R	0.615	0.358
	$\Delta, \%$	-77.9	38.4

Было проведено обучение на синтетических данных и последующее дообучение на реальных. Сравнение R-фактора для моделей до и после дообучения представлено в таблице 3.1. После дообучение на реальных данных точность предсказания моделей увеличивается минимум в 1.5 раза, однако теряют в точности на синтетических данных, кроме кастомного UNet с Фурье-преобразованием. Лучшую точность имеет модель UNet_FFT, от которой немного отстает трансформер. В сводной таблице 3.2 представлены значения метрик на реальных данных финальных моделей. Таким образом, UNet_FFT занимает меньше видеопамяти, работает быстрее и достигает лучшей метрики на тестовых реальных данных.

Таблица 3.2 – Значения метрик на тестовом реальном наборе данных моделей после дообучения

Metric \ Model	UNet	FFT_UNet	XRD_Transformer
MSE·10 ⁻³	1,39	1,20	1,31
R	0,393	0,336	0,358

Анализ моделей

Также был проведен анализ полученных моделей. Для трансформера были построены карты внимания (рис. 3.1) при обработке одной из тестовых реальных структур, из которых видно, что модель в первых блоках имеет более рассеяное и равномерное внимание, что говорит о том, что сначала производится сбор общего контекста. В последующих блоках внимание становится более сфокусированным и локализованным, показывая, что модель учится выделять важные взаимосвязи. Также наблюдается способность модели связывать дальние отражения, что важно в нашей задаче. Более сильные связи наблюдается вдоль оси l; также можно заметить диагональные паттерны внимание, что соответствует систематическим погасаниям. Таким образом, можно сделать вывод, что модель научилась распознавать кристаллографические закономерности.

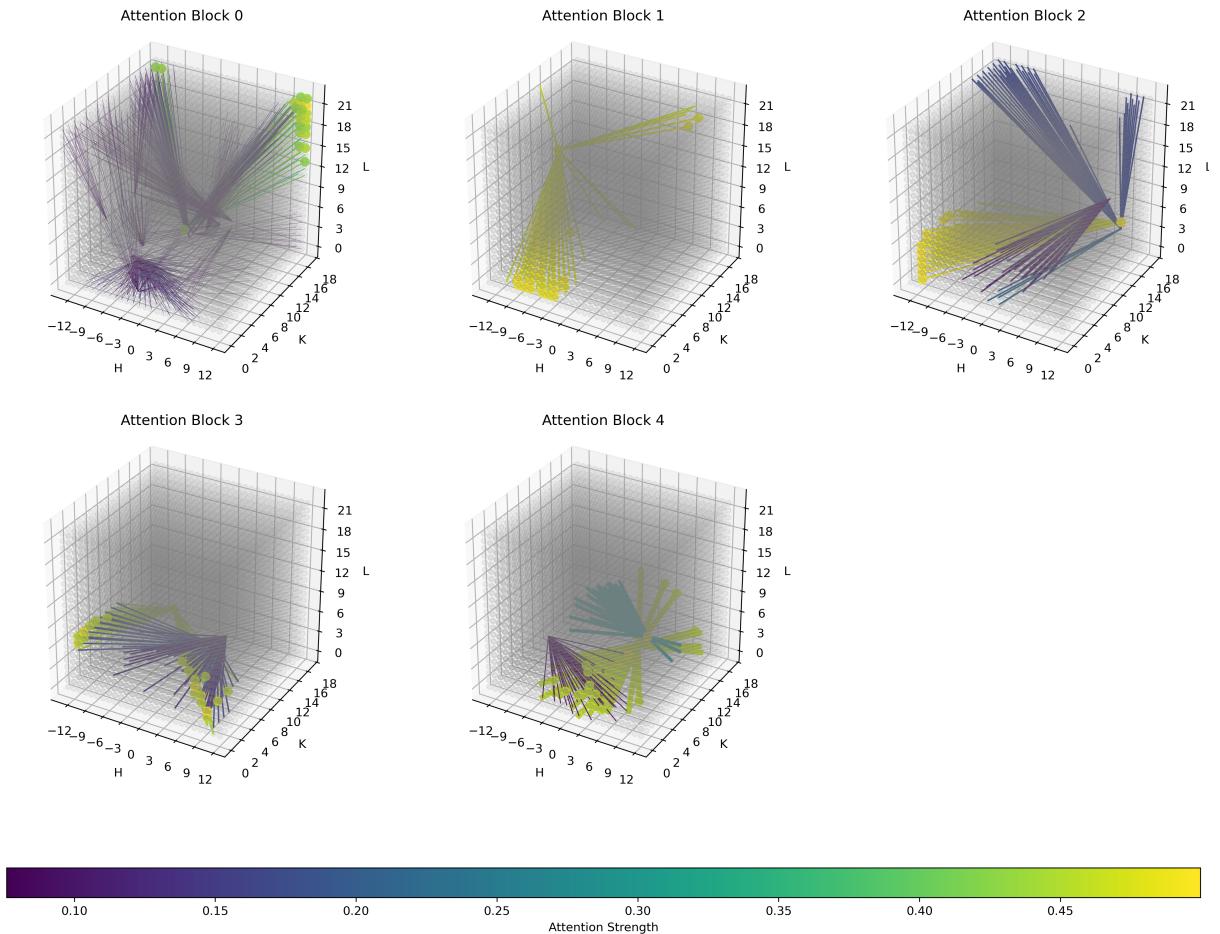


Рисунок 3.1 – Связи внимания в блоках трансформера

Также был проведен анализ первого блока моделей UNet и FFT_UNet с помощью GradCAM (рис. 3.2, 3.3). По паттерну для UNet можно заметить, что модель обращает исключительное внимание на границы дифракционной картины низкого разрешения. Смущает, что высокие значения Attribution и Overlay достигаются во всем диапазоне вне дифракционной картины на входе. Паттерн же для FFT_UNet кажется более разумным: Attribution и Overlay распространяются на входную картину и немного за неё, что показывает, что модель ищет информацию около границы дифракционной картины. Наиболее высокие значения внимания достигаются в высокоинтенсивной зоне, дальше от неё значения монотонно спадают.

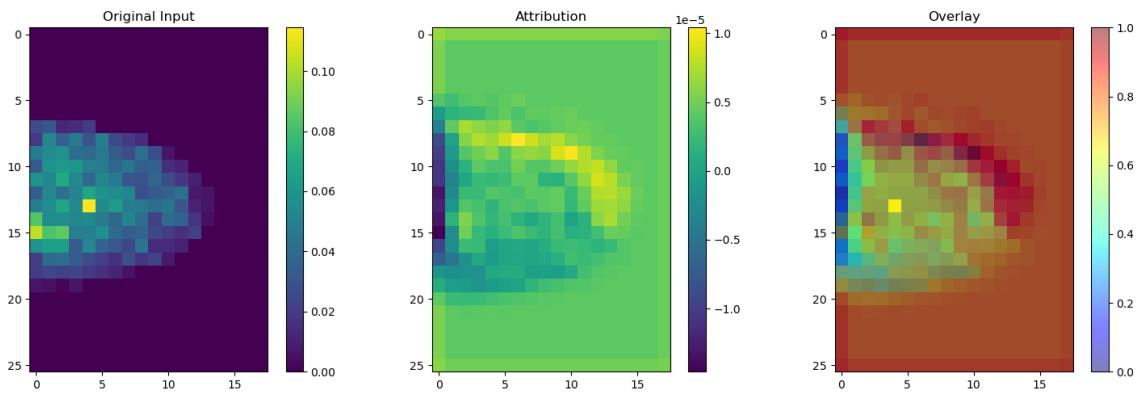


Рисунок 3.2 – Анализ первого блока UNet с помощью GradCAM

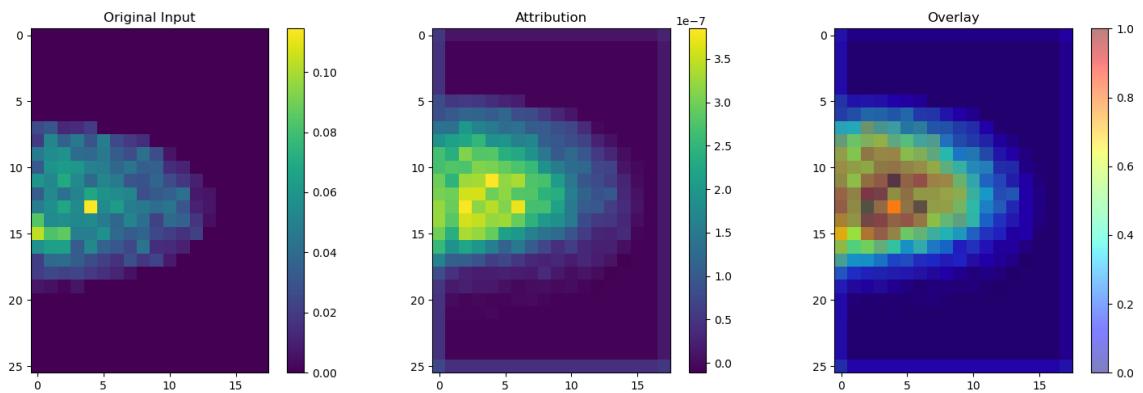


Рисунок 3.3 – Анализ первого блока FFT_UNet с помощью GradCAM

Также были получены карты чувствительности для данных моделей (рис. 3.4). По ним можно сделать вывод, что модели в целом робастные и стабильно работают при добавлении шума практически во всем обратном пространстве. Однако существуют единичные воксели, которые претерпевают большие возмущения из-за шума.

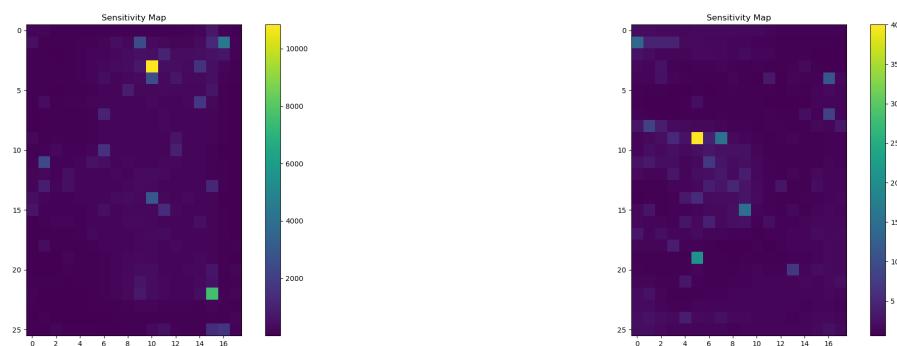


Рисунок 3.4 – Карты чувствительности для UNet (слева) и FFT_UNet (справа)

Проверка на реальных данных

Несмотря на неплохие значения метрик на реальных моноклинных структурах из Кембриджского Банка Структурных данных, качества восстановления тензора отражений недостаточно для решения структуры с помощью SHELXT. На рис. 3.5 представлены сечения тензора отражений для реальной структуры, R-фактор восстановленного тензора с помощью FFT_UNet равен 0.346. Нейросеть достаточно точно предсказывает границы дифракционной картины, средние значения по voxelам (если их брать по дуге) тоже достаточно близки. Но восстановленные значения в тензоре более смазаны вдоль этих дуг, картине не хватает более точного распределения амплитуд по соседним voxelам в каждой локальной области.

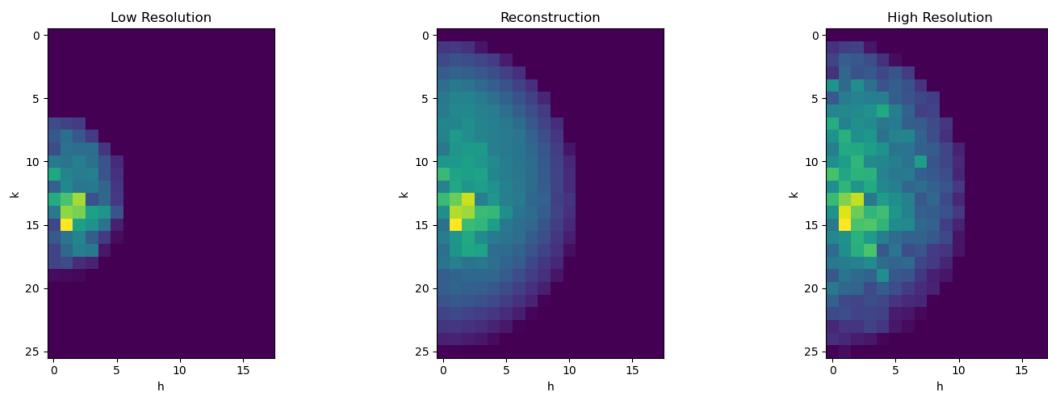


Рисунок 3.5 – Типичная реконструкция дифракции реальной структуры, $R = 0.346$

Как уже было оговорено, дифракционные отражения не обладают локальной связью, и для успешного решения задачи фаз требуется очень точно определить амплитуду в каждом voxelе. Несмотря на то, что модели машинного обучения успешно схватывают кристаллографические закономерности, им не хватает более точного численного определения в каждом voxelе данных. Таким образом, методы глубокого обучения успешно определяют кристаллографические паттерны и закономерности в обратном пространстве, однако им не хватает точности для решения проблемы фаз при восстановлении численных значений амплитуд отражений.

Дальнейшие планы и предложения

В работе было показано, что методы ИИ успешно схватывают кристаллографические закономерности, однако из-за специфики обратного пространства для решения проблемы фаз им не хватает точности. Совершив переход в локально связанное пространство, можно предположить, что нейронные сети справятся с поставленной задачей.

Для решения проблемы фаз часто используется так называемая функция Паттерсона:

$$P(u, v, w) = \sum_{h,k,l \in Z} |F(h, k, l)|^2 e^{-2\pi i(hu + kv + lw)}$$

Таким образом, функция Паттерсона является Фурье-преобразованием интенсивностей дифракционных максимумов, где за их фазы приняты нули. Таким образом, перейдя от тензора амплитуд/интенсивностей к трехмерной картине функции Паттерсона — получается классическая задача Super Resolution, поскольку для дифракционной картины низкого разрешения получится Фурье-образ низкого разрешения, разрешение которого требуется повысить так, чтобы восстановленная картина соответствовала Фурье-образу дифракционной картины высокого разрешения (рис. 3.6). После увеличения разрешения трехмерной картины функции Паттерсона, мы можем вернуться в обратное пространство, рассчитав интенсивности (и из них амплитуды) рентгенодифракционных отражений, и решить проблему фаз с помощью уже описанных *ab initio* методов.

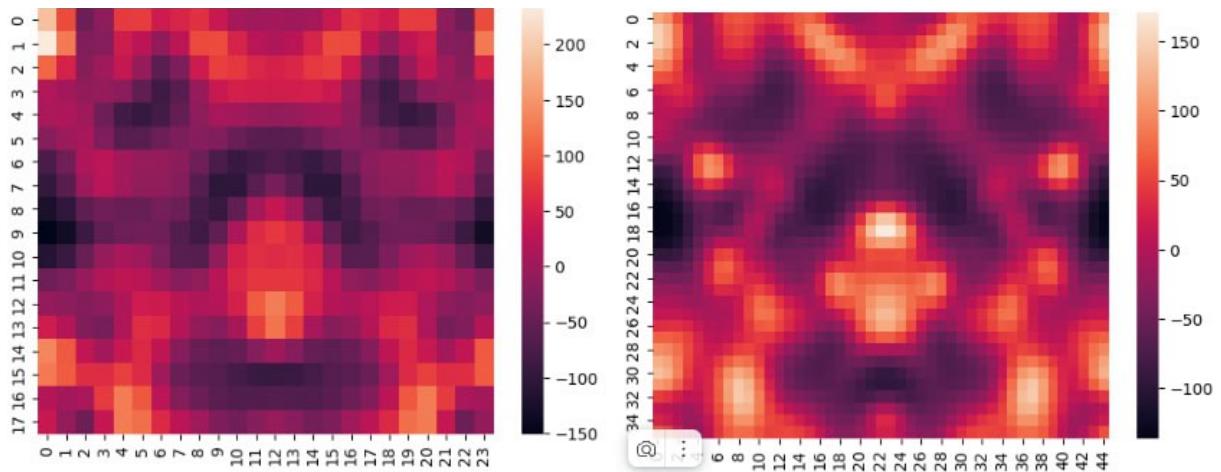


Рисунок 3.6 – Типичное сечение функции Паттерсона по данным низкого (слева) и высокого (справа) разрешения

4 Выводы

- Разработано программное обеспечение по генерации синтетических рентгенодифракционных данных, которое может быть использовано для решения прикладных задач рентгеновской дифракции с помощью ИИ (github.com/blackwood168/xrd_simulator)
- Разработан единый пайплайн (github.com/blackwood168/xrd_phase_ml), позволяющий проводить воспроизводимые эксперименты по обучению, тестированию и inference задачи предсказывания дальних отражений по ближним для решения проблемы фаз
- Разработаны и обучены на синтетических и реальных структурах UNet в качестве baseline, FFT_UNet — кастомный UNet с Фурье-преобразованием в слоях, XRD_Transformer — модель на основе трансформер со специфичным эмбеддингом, вписывающимся в физику задачи
- Проведен анализ моделей, подтверждающий, что модели глубокого обучения выявляют кристаллографические закономерности и имеют потенциал решения прикладных задач рентгеновской дифракции
- У моделей не достает качества численного восстановления рентгеновских отражений для решения проблемы фаз, несмотря на улавливание рентгенодифракционных паттернов; приведено обоснование и предложен план изменения методологии для решения задачи

Список литературы

- [1] Girolami G. S. X-ray Crystallography. — University Science Books, 2016.
- [2] Hauptman H. and Karle J. Solution of the Phase Problem. I. The Centrosymmetric Crystal.
- [3] Oszlányi G. and Sütő A. Ab initiostructure solution by charge flipping // Acta Crystallographica Section A Foundations of Crystallography. — 2004. — Vol. 60, no. 2. — P. 134–141.
- [4] Lattman E. and Loll P. Protein Crystallography. — Johns Hopkins University Press, 2008.
- [5] Margiolaki I. and Wright J. P. Powder crystallography on macromolecules // Acta Crystallographica Section A. — 2008. — Vol. 64, no. 1. — P. 169–180.
- [6] Pan T., Dun C., and Jin S. CrysFormer: Protein structure determination via Patterson maps, deep learning, and partial structure attention // Structural dynamics. — 2024.
- [7] Larsen A. S., Rekis T., and Madsen A. PhAI: A deep-learning approach to solve the crystallographic phase problem // Science. — 2024. — Vol. 385, no. 6708. — P. 522–528.
- [8] Wang K., Song L., and Wang C. On the use of deep learning for phase recovery // Light Sci Appl. — 2024. — Vol. 13, no. 4.
- [9] Grosse-Kunstleve R. W., Sauter N. K., Moriarty N. W., and Adams P. D. The *Computational Crystallography Toolbox*: crystallographic algorithms in a reusable software framework // Journal of Applied Crystallography. — 2002. — Vol. 35, no. 1. — P. 126–136.
- [10] Groom C. R. and Allen F. H. The Cambridge Structural Database in Retrospect and Prospect // Angewandte Chemie International Edition. — 2014. — Vol. 3. — P. 662–671.

- [11] Forigua C., Escobar M., and Arbelaez P. SuperFormer: Volumetric Transformer Architectures for MRI Super-Resolution. — 2022. — P. 132—141.
- [12] Sheldrick G. M. SHELXT - Integrated space-group and crystal-structure determination. — 2015. — P. 3–8.
- [13] Yang Y., Lian Q., Zhang X., Zhang D., and Zhang H. HIONet: Deep priors based deep unfolded network for phase retrieval // Digital Signal Processing. — 2023. — Vol. 132. — P. 103797.