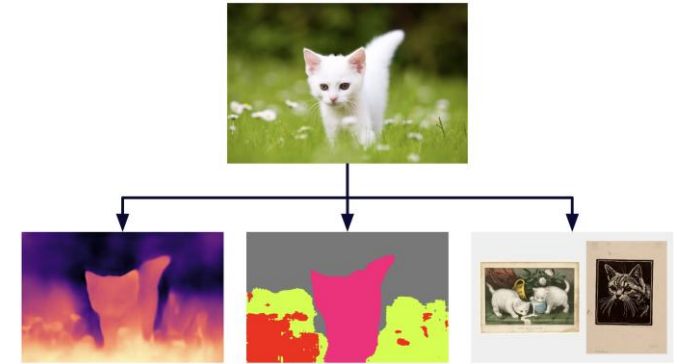


Методология создания базовых  
универсальных моделей трансформерных  
архитектур нейронных сетей для анализа  
процессов на коллайдерах

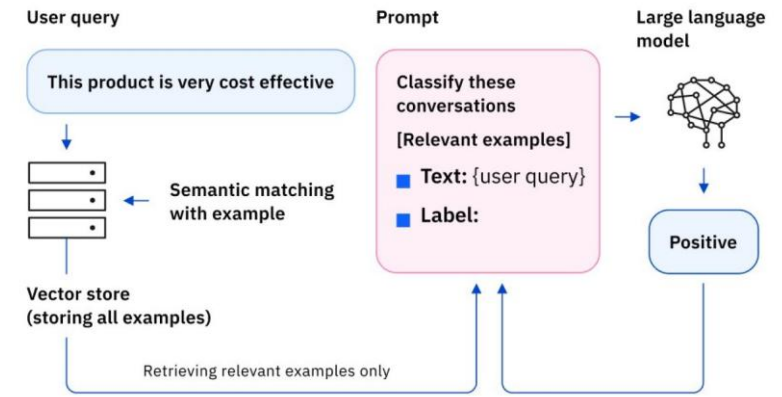
Абасов Э.Э., Волков П.В., Дудко Л.В., Заборенко А.Д., Иудин Е.С.,  
Маркина А.А., Перфилов М.А.

# Универсализация обработки данных

- В областях компьютерного зрения и обработки естественного языка часто используются фундаментальные модели, предобученные на большом и разнообразном объеме данных. Такая тренировка позволяет модели выявить закономерности в данных, полезные для многих последующих задач.
- В области Физики Высоких Энергий (НЕР) такой подход также может быть перспективным, например, для классификации событий. Фундаментальная модель, выделяющая общие физические закономерности, потенциально может улучшить точность узкоспециализированного анализа.
- Переходный шаг к таким моделям – «базовая» модель архитектуры Трансформер.



Извлечение визуальных признаков трансформером DinoV2, <https://arxiv.org/abs/2304.07193>

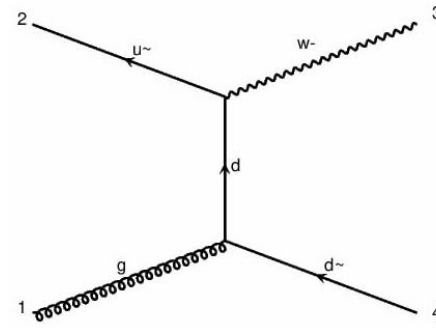


Использование предобученной языковой модели для few-shot классификации

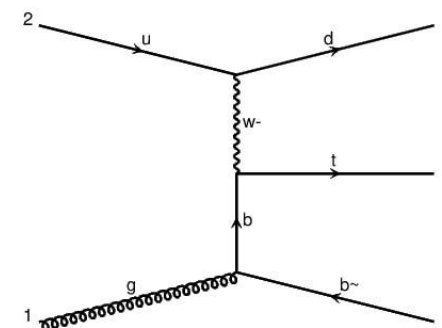
# Используемый датасет

- Использовался датасет из смоделированных коллайдерных событий, состоящий из пяти наборов процессов, сгруппированных по числу топ-кварков
- Генераторы: MadGraph5 и CompHEP
- Моделирование отклика детектора: DELPHES

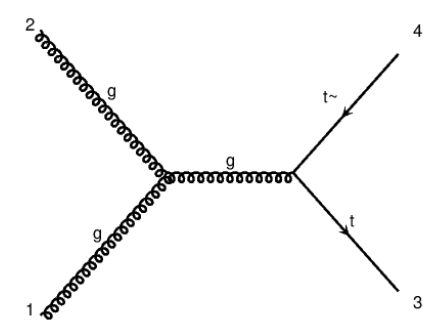
Примеры диаграмм Фейнмана для групп процессов:



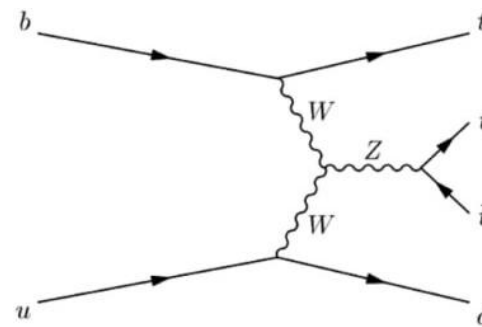
0 топ-кварков



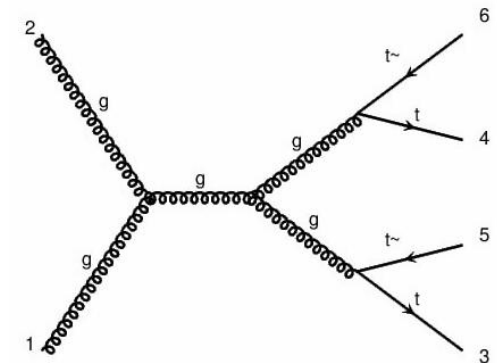
1 топ-кварк



2 топ-кварка



3 топ-кварка



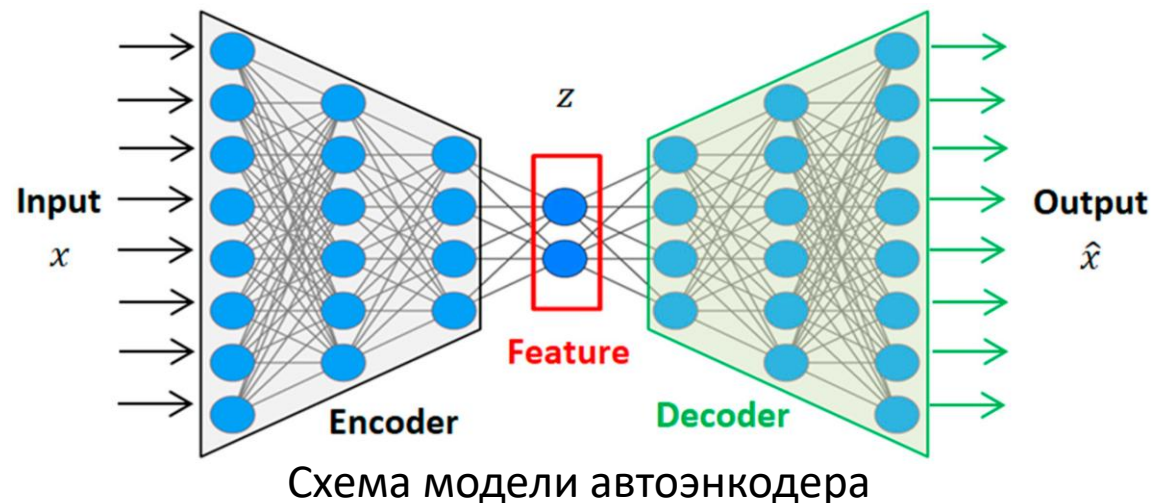
4 топ-кварка

# Универсальный набор входных переменных

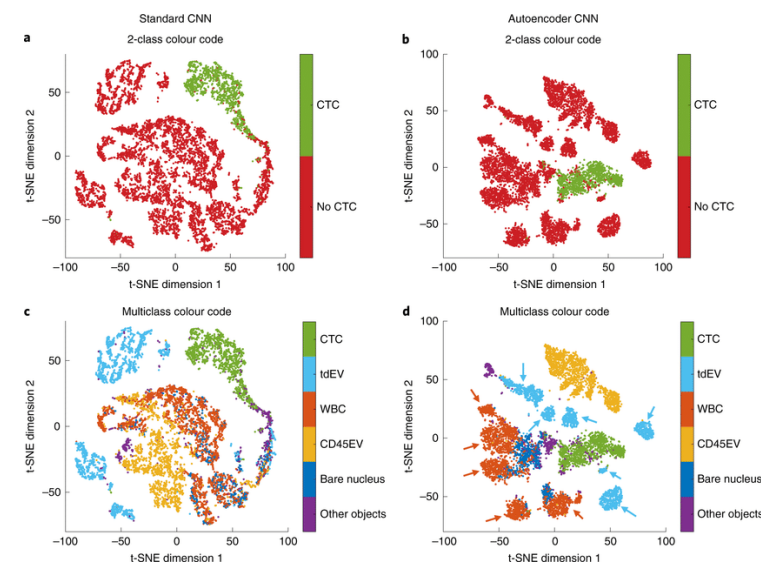
- В анализах НЕР изучаются коллайдерные события, продукты которых – **джеты и лептоны**, кинематические свойства которых измеряются современными детекторами.
- Тогда базовый набор переменных, характеризующий отдельное событие – это коллекция кинематических переменных всех возможных объектов. В нашем анализе для каждого объекта используются следующие переменные:  $p_T, \eta, \phi, \bar{P}_x, \bar{P}_y, \bar{P}_z$ .
- Также используются некоторые переменные, общие для всего события:  $N_{jets}, N_{b-jets}, N_\nu, N_l, MET$ .
- Максимальное число объектов: **12** струй (из них 4 b-струи) + **4** лептона. В рамках каждой группы объекты отсортированы по энергии.
- Если в событии отсутствует определенный объект, то в его переменных записывается ноль, такие значения обрабатываются отдельно.

# Формирование пространства представлений

- Один из возможных шагов в анализе процессов с различными кинематическими особенностями – создание объединенного пространства представлений. Обычно это выход промежуточного слоя нейронной сети, обученной выполнять какую-то задачу на объединенном датасете. Ожидается, что в этом пространстве более явно выделены закономерности в данных, лучше определены релевантные кластеры.
- Частым выбором для такой задачи является либо понижение размерности (автоэнкодер), либо моделирование замаскированных переменных (masked modelling).

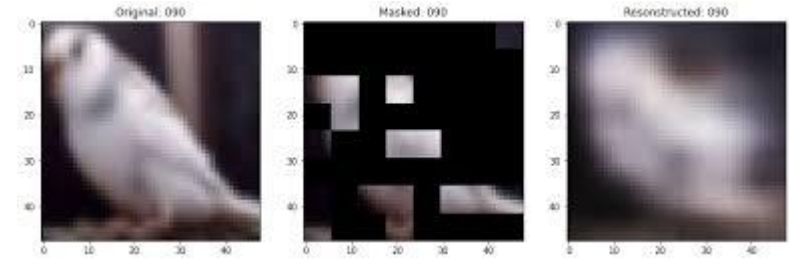


Визуализация пространства представлений для обычной сети и автоэнкодера (домен изображений):

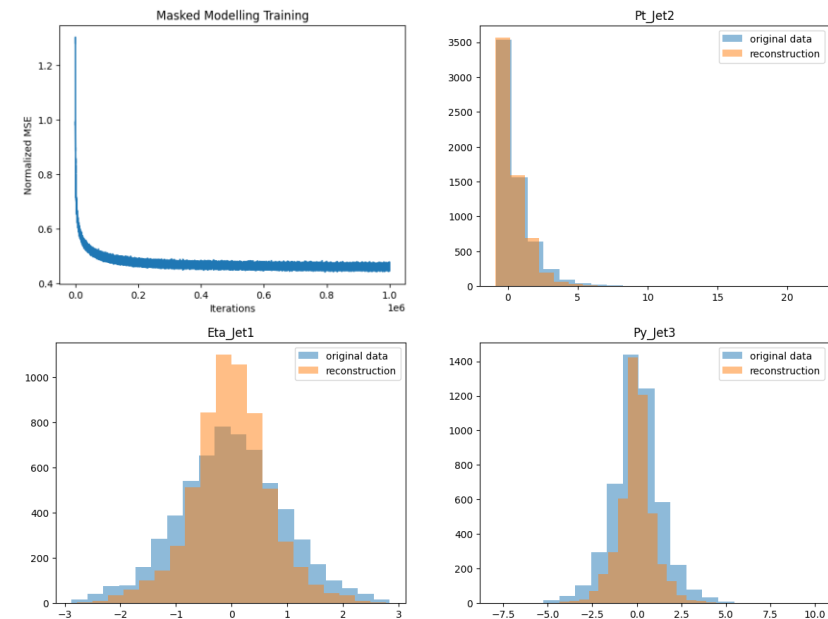


# Формирование пространства представлений

- Для нашей задачи был выбран подход реконструкции замаскированных переменных. В сравнении с автоэнкодером такой метод может быть более точным, т.к. в процессе обучения маска применяется ко всем переменным, и модель выучивает все связи между ними, а не руководствуется только средним качеством реконструкции.
- Потенциальное направление для предобучения – использование генераторной информации: регрессия компонент нейтрино, комбинаторные задачи.



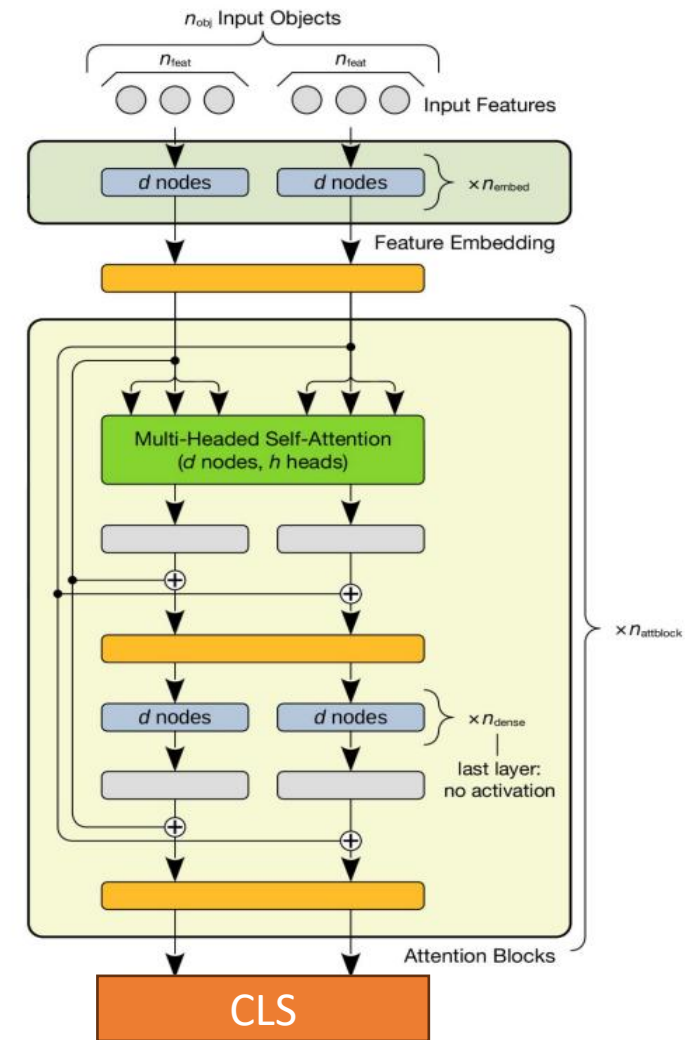
Пример реконструкции для изображений



Средняя ошибка реконструкции и восстановленные переменные.  
Вероятность маскирования переменной – 30%, переменные стандартизированы

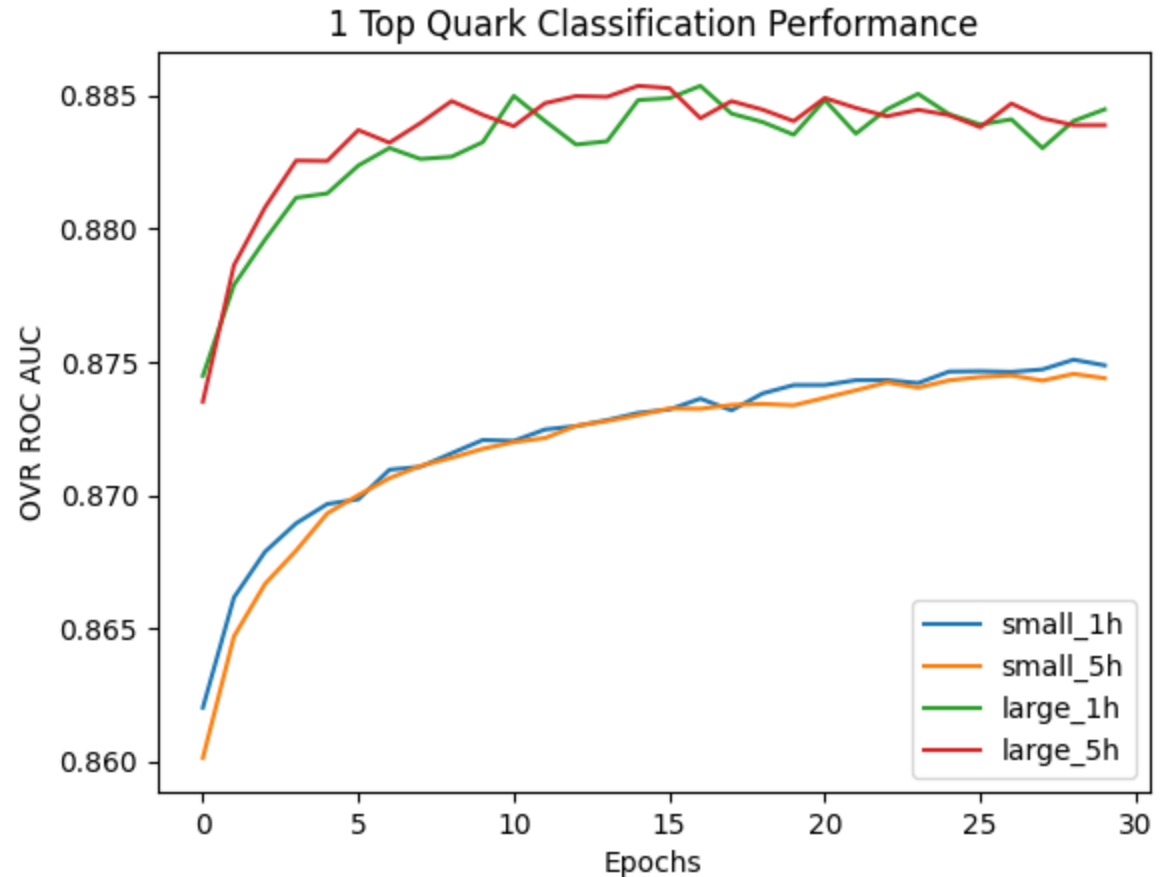
# Модель Transformer

- Для задачи классификации была выбрана архитектура Трансформер. Эту архитектуру можно эффективно параллелизовать при обучении на больших датасетах, а модуль внимания может моделировать сложные комбинаторные зависимости между частицами.
- Результат классификации – разделение событий с разным числом топ-кварков (относительно простая тестовая задача).



# Архитектурные вариации

- Было обучено несколько вариаций трансформера:
  - Варианты small и large отличаются размером эмбединга и полносвязной сети
  - Варианты 1h и 5h отличаются количеством голов внимания
- Выводы:
  - Увеличение размера модели влияет на метрики классификации, но увеличивает ресурсы для тренировки
  - Увеличение числа голов внимания не улучшает работу модели, а в случаях перетренировки деградирует ее сильнее

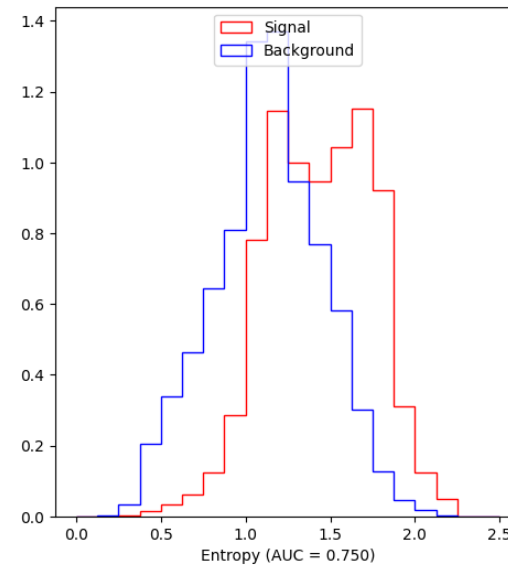




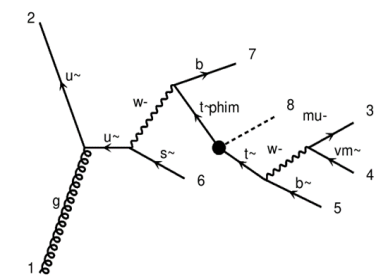
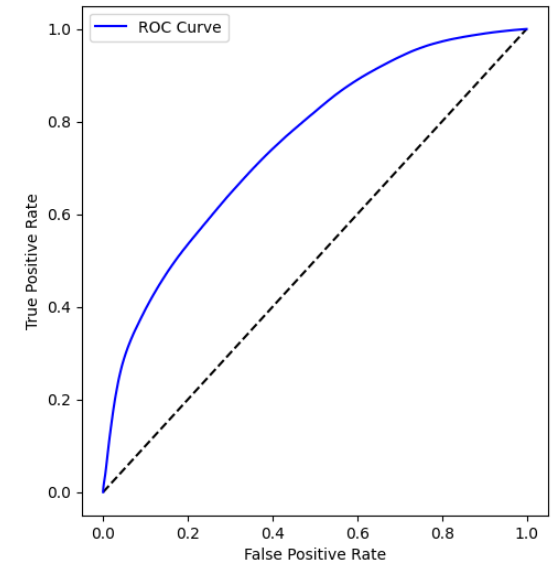
# Применение обученного трансформера для поиска Темной Материи

- В качестве примера применения обученного трансформера приведено использование энтропии распределения вероятностей трансформера (без его дообучения) для идентификации событий, смоделированных с использованием медиатора Темной материи.
- $H(X) := -\sum_{x \in X} p(x) \log_2 p(x)$  – функция энтропии, применяющаяся к предсказанным вероятностям для каждого события.
- Рассматриваются процессы т-канального рождения топ-кварка в Стандартной Модели и в рамках Упрощенной Модели ТМ со скалярным медиатором.

Entropy of Transformer Probability Predictions



ROC Curve



Диаграммы Фейнмана для фоновых и сигнальных процессов

# Заключение

- В докладе представлены методы создания универсальных больших нейронных сетей для унификации анализа данных НЕР в разных сигнатурах:
  - организация набора переменных и создание представления в некотором абстрактном оптимизированном пространстве
  - предобучение модели для создания пространства представлений
  - обучение большой модели трансформера для тестовой задачи классификации , на примере всех основных процессов рождения топ-кварка во всех возможных сигнатурах
- Следующие шаги будут включать последовательные оптимизации промежуточных шагов и усложнение задач, ставящихся перед базовой моделью, для применения в исследованиях топ-кварка на коллайдерах.

Backup

# Дополнительные метрики классификации

