

BAIKAL-GVD · ГЛУБОКОВОДНЫЙ НЕЙТРИННЫЙ ТЕЛЕСКОП

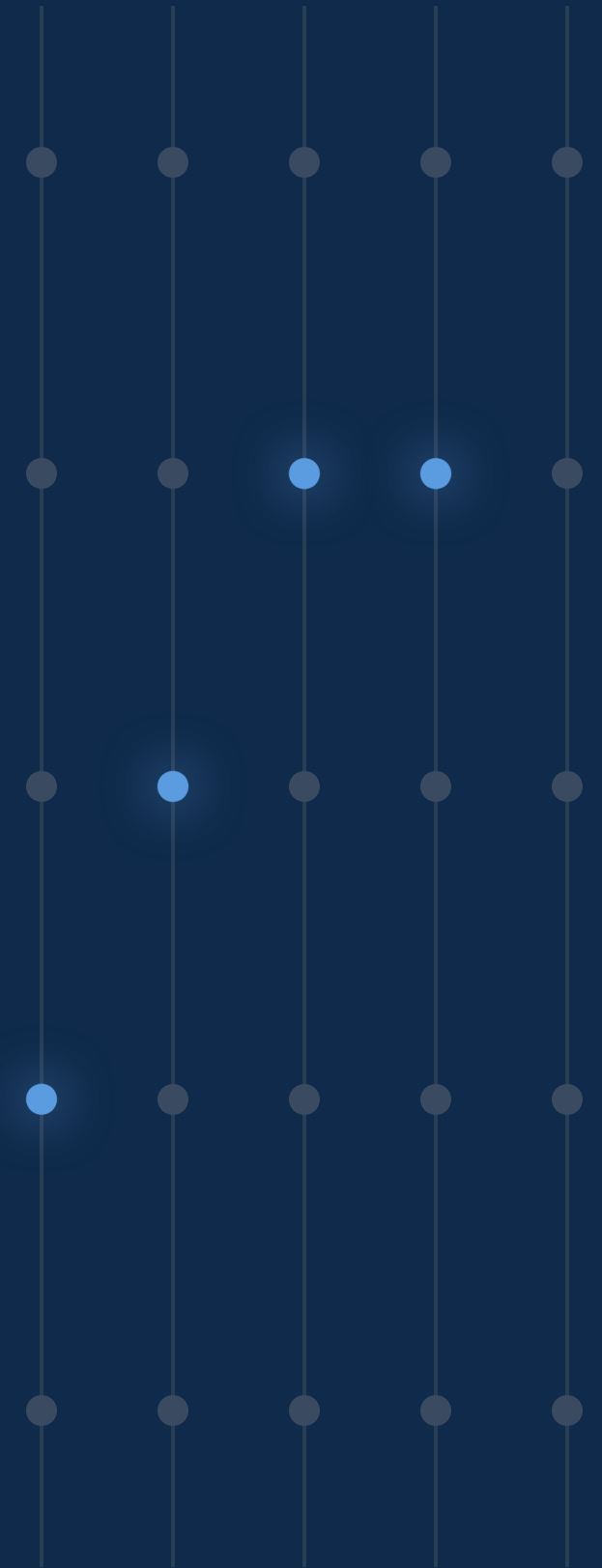
Классификация хитов графовыми нейросетями

Три поколения архитектур: свёртки → внимание → message passing

GCN

GAT

MPNN



Физическая задача и цель

Оптические модули на вертикальных гирляндах-«стрингах» регистрируют **черенковский свет** от трека релятивистской частицы в глубине озера.

Каждое **событие** — набор сработавших модулей, «хитов»: часть — сигнал, часть — фон.

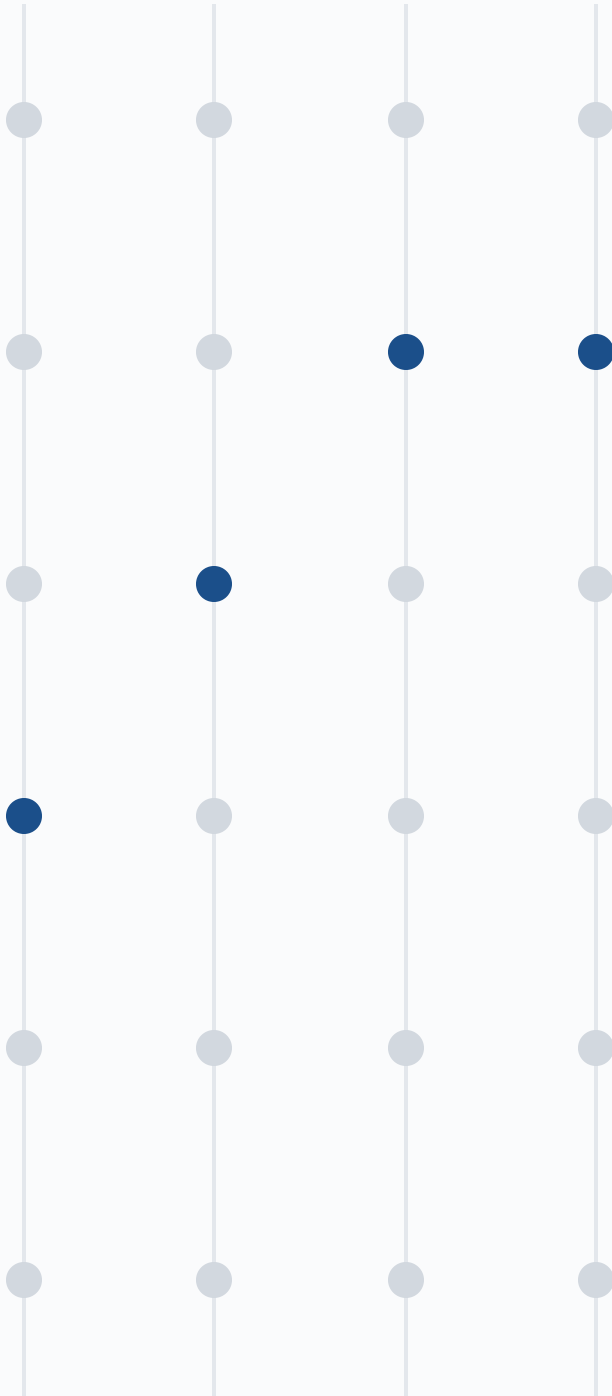
Задача — для каждого хита решить, **сигнал** он или **шум**.
Это первичная очистка и **фундамент** всей дальнейшей реконструкции.

● СИГНАЛ

Черенковский свет от трека частицы

● ФОН

Шум ФЭУ, случайные совпадения



Данные и детектор

18.7 млн

событий Монте-Карло

1.43 млрд

хитов всего

5

признаков на хит — Q, T, X, Y, Z

ГЕОМЕТРИЯ ТЕЛЕСКОПА

Стрингов в кластере	8
Модулей на стринге	36
Шаг по вертикали	~15 м
Рабочие глубины	750–1275 м

МАСШТАБ СОБЫТИЯ

Хитов в среднем	~76
Типичный разброс	десятки–сотни
Активных стрингов	7–8

У каждого хита есть истинная метка сигнал / шум — её и учимся предсказывать.

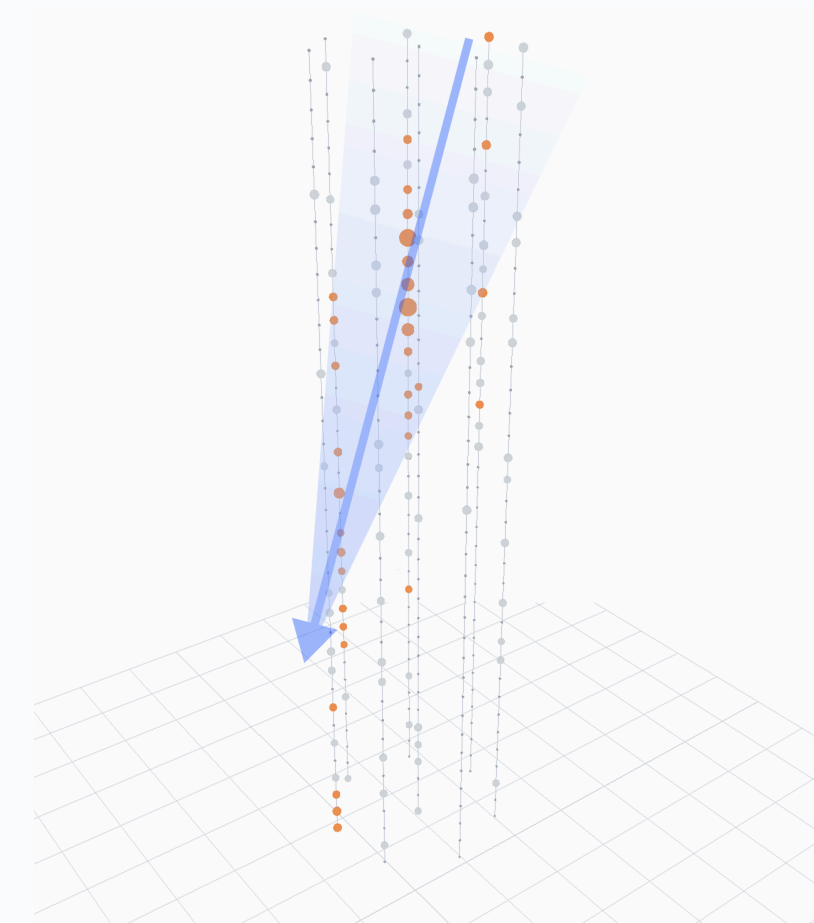
Событие как граф

Граф — **естественное и эффективное** представление для наших данных: событие само по себе устроено как набор связанных хитов. Событие — это граф:

узлы хиты со своим вектором признаков

рёбра соседи — к ближайшим по времени хитов

Ширина окна k управляет плотностью графа — и оказывается одним из рычагов качества.



● сигнал — трек частицы ● шум

Реальный event display: оранжевые модули — сигнал вдоль трека частицы (стрелка), бледные — шум. Это событие и представляем графом.

Как учится графовая сеть

На каждом слое узел «смотрит» на соседей, собирает их информацию и обновляет свой вектор. Три части доклада различаются тем, **как именно узел собирает информацию с соседей**:

ЧАСТЬ I · GCN

Свёртка

Просто **усредняет** векторы соседей.

ЧАСТЬ II · GAT

Внимание

Взвешивает соседей — учится, кто важнее.

ЧАСТЬ III · MPNN

Message passing

Для каждой пары считает **сообщение** отдельной нейросетью.

Метрика качества

P@R0.9

Чистота выборки (precision), если настроить порог так, чтобы поймать **90% сигнала** (recall = 0.9).

Иначе говоря: ловим 90% сигнальных хитов и смотрим, насколько чист отобранный набор — какая доля в нём действительно сигнал, а не шум.

Несколько первых экспериментов оценивались обычным recall при пороге 0.5 — пока мы не пришли к этой ключевой метрике.

ПОЧЕМУ НЕ ACCURACY

Сигнальных хитов в событии заметно **меньше**, чем шумовых — при таком дисбалансе классов ассигасу почти бесполезна.

Нам важно не пропустить сигнал и держать выборку чистой — именно это и измеряет **precision при фиксированном recall**.

Идея графовой свёртки

$$h_i^{(l+1)} = \sigma \left(W^{(l)} \cdot \text{mean}_{j \in \mathcal{N}(i)} h_j^{(l)} \right)$$

`mean` — усреднение по соседям `W` — обучаемая матрица слоя `σ` — нелинейность (GELU)

Ключевое свойство – и будущее ограничение: свёртка усредняет всех соседей одинаково.

Базовый GCN: где потолок

ОКНО К ПОМОГАЕТ (Recall)		ЁМКОСТЬ — НЕТ (P@R0.9)		ПРИЗНАКИ — ЧАСТИЧНО (P@R0.9)	
k = 12	0.441	128 кан.	0.353	5 базовых	~0.39
k = 4	0.632	512 кан.	0.353	геометрия — до 9 признаков	0.453
k = 2	0.689	2048 кан.	~0.38	физика — до 13 признаков	0.472

ГЕОМЕТРИЯ (ДО 9 ПРИЗНАКОВ)

Производные координаты хита — цилиндрические и сферические (r, φ, θ) , степень узла в графе. Где хит расположен в детекторе.

ФИЗИКА (ДО 13 ПРИЗНАКОВ)

Признаки взаимодействия — невязка времени пролёта света, заряд и время ближайших соседей, локальная плотность. Как хит связан с треком.

Ни ёмкость, ни признаки не пробивают потолок до конца. Узкое место — **сама операция усреднения** в свёртке. → [главная проблема глубоких GNN](#)

Остаточные связи: ResGCN

$$h^{(l+1)} = \text{GELU}\left(\text{GCN}(h^{(l)}) + \text{proj}(h^{(l)})\right)$$

ABLATION ПРИ РАВНОМ ЧИСЛЕ ПАРАМЕТРОВ

With Residual	0.598
Linear Only (параметры те же, без skip)	0.360
Without Residual (обычная GCN)	0.354

Дело **не в параметрах**, а в операции «прибавить вход». Остаточные связи — **необходимое условие** глубоких GNN.

Проблема глубины — **пересглаживание**: узлы становятся неразличимы. Лекарство — «обходной путь» для сигнала: к выходу слоя прибавляем вход.

Собрав выходы всех слоёв (**Jumping Knowledge**) и расширив вход до 21 признака, свёрточный подход выходит на свой потолок — **P@R0.9 ≈ 0.91**.

Среди добавленных до 21 признаков: dt (сдвиг времени), ToF Res (невязка времени пролёта), NeighDist / NeighQ (соседи), ρ, r_xy, cosθ (геометрия), EventNhits (размер события).

Идея внимания на графе

$$h_i^{(l+1)} = \sigma \left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} W^{(l)} h_j^{(l)} \right), \quad \sum_j \alpha_{ij} = 1$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_k \exp(e_{ik})}, \quad e_{ij} = a^{(l)\top} \text{LeakyReLU} \left(\Theta^{(l)} [h_i \| h_j] \right)$$

W — преобразование значений соседа, Θ — отдельная матрица для веса внимания (это **разные** матрицы), a — вектор внимания. Внутри слоя они общие для всех соседей; у каждого слоя (l) — свой независимый набор.

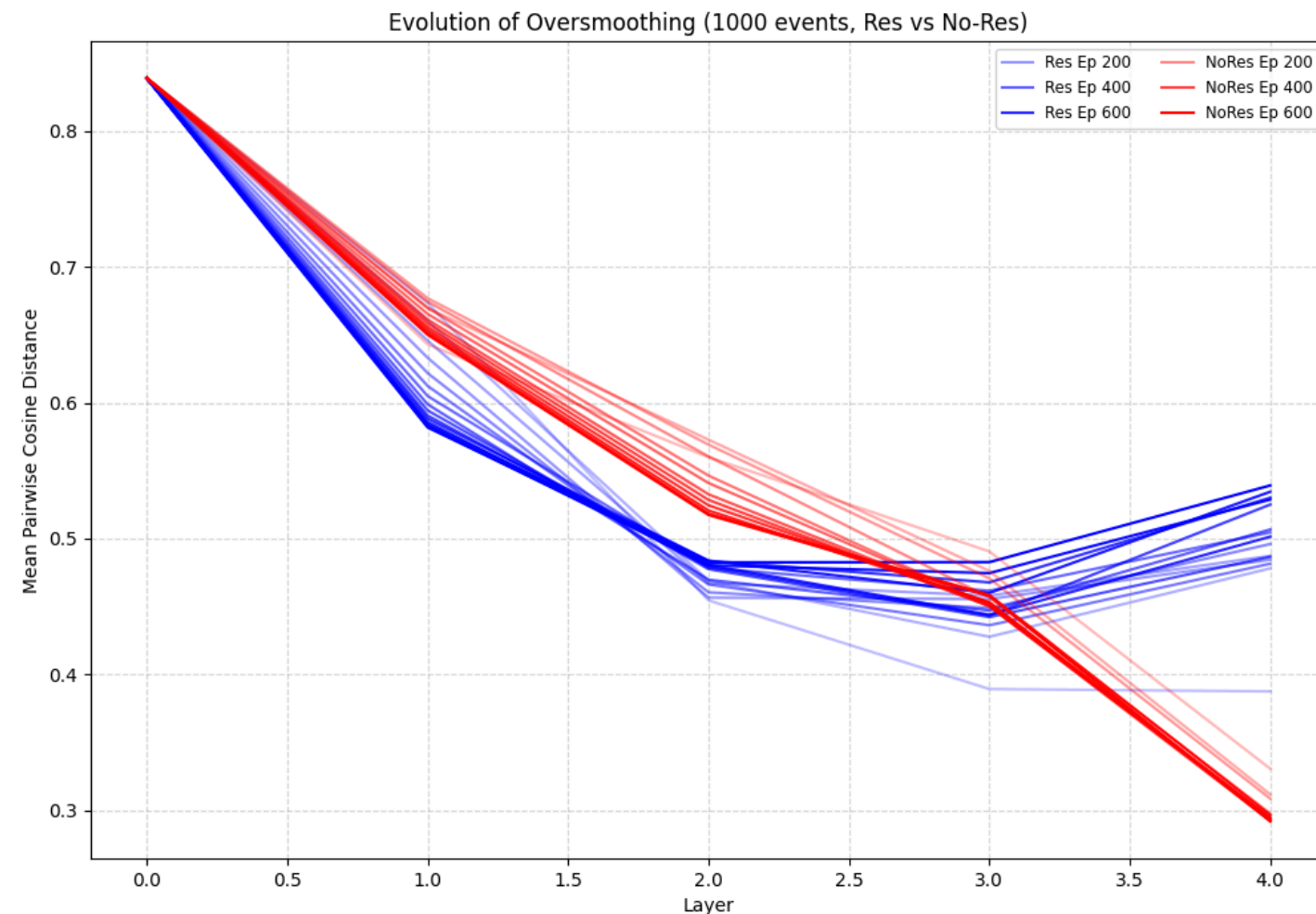
GATv1 · статическое

Ранжирование соседей почти не зависит от того, какой узел «спрашивает».

GATv2 · динамическое

Важность соседа по-настоящему зависит и от i , и от j — выразительнее.

Внимание — и отдельно пересглаживание



Само по себе локальное внимание дало **~0.92** — на уровне свёрточного потолка. **Скачка нет.**

Здесь я прямо **измерил пересглаживание** — среднее косинусное расстояние между узлами по слоям. Борьба с ним через residual улучшает модель **независимо** от механизма внимания.

0.30

без residual — узлы неразличимы

0.51

с residual — «якорь» держит

Динамика по эпохам: без residual (красные) обучение усиливает схлопывание, с residual (синие) — расстояние держится.

Глобальный слой события

АРХИТЕКТУРА	P@R0.9
FC Last Layer (глоб. внимание, k=4)	0.953
FC Last · k=8	0.962
FC Last · k=12	0.9623

FC Last Layer

Первые слои — на локальном графе (k=4), а последний слой **полносвязный**: каждый хит видит все хиты события сразу.

+0.04 к P@R0.9 при ~677k параметров

Один глобальный слой поверх лёгкого локального графа даёт больше, чем наращивание слоёв или ширины. Дальше локальный граф полезно уплотнять до $k \approx 12$ — потом насыщение.

Идея message passing

$$m_{ij} = f([h_i \| h_j]), \quad a_i = \text{mean}_j m_{ij}, \quad h'_i = \text{GELU}(u([h_i \| a_i]) + \text{proj}(h_i))$$

И свёртка, и внимание — **частные случаи** этой схемы.

f и u — двуслойные MLP: вклад соседа считается **нелинейной** функцией.

Метрика теперь — **P@R0.95**: по P@R0.9 модели хуже различимы.

General MPNN: рёбра и мастер-вершина

Заменяю готовые слои GATv2 на собственный message passing и добавляю две параллельные идеи:

ФИЗИКА РЕБРА

В сообщение явно подаётся геометрия пары хитов — разница времён, расстояние, интервал Минковского. Лучший результат — когда для каждого скаляра выделяется **обучаемое направление** в пространстве представлений хитов.

МАСТЕР-ВЕРШИНА

Одна виртуальная вершина, связанная со всеми хитами: быстрый канал **«событие ↔ хит»** для глобального контекста.

ПОЛНЫЙ ОБЪЁМ · 25.6М СОБЫТИЙ · P@R0.95

мастер + рёбра edge12	0.9558
мастер + рёбра edge3_learn8	0.9542
обучаемые рёбра (edge3_learn8)	0.9509
геометрия рёбер (edge3)	0.9502

В эпоху MPNN архитектуры сошлись в узкий коридор 0.950–0.956. Главный рычаг — **явная физика ребра**.

ИТОГИ

Сводная картина

ЧАСТЬ	АРХИТЕКТУРА / НЮАНС	МЕТРИКА СВОЕГО ВРЕМЕНИ
I · Свёртки	Базовый GCN (k=2, без LayerNorm)	$P@R0.9 \approx 0.35$
	+ инженерия признаков (5 → 13)	$P@R0.9 = 0.47$
	ResGCN — остаточные связи + 21 признак	$P@R0.9 = 0.77$
	JK-ResGCN + 21 признак	$P@R0.9 \approx 0.91$
II · Внимание	GATv2 (+ Res / JK)	$P@R0.9 = 0.92\text{--}0.93$
	FC Last Layer + плотный k=12	$P@R0.9 = 0.9623$
III · MPNN	General MPNN + физика ребра / мастер (25.6M)	$P@R0.95 = 0.950\text{--}0.956$
	лучший — мастер + рёбра edge12	$P@R0.95 = 0.9558$

Сквозные уроки

- 01 Способ сбора информации с соседей — **главный рычаг** (mean → attention → MLP).
- 02 Топология графа решает не меньше архитектуры.
- 03 Остаточные связи — **необходимое условие** глубоких GNN.
- 04 Глобальный контекст события важнее глубины иерархии.
- 05 Явная физика ребра — лучший текущий рычаг в эпоху MPNN.
- 06 Строгий протокол измерений — единая шкала метрики и одинаковый объём обучающих данных — обязателен для честного сравнения архитектур.

ЧТО ДАЛЬШЕ

Лидер (мастер + рёбра) не насытился к эпохе 2660 — есть запас. Расширять обучаемый базис признаков рёбер.

От очистки хитов — к **реконструкции треков**: те же графовые представления должны давать направление и энергию частицы.

Спасибо. Вопросы?