НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
сентябрь–октябрь 2026 Том 26 № 5 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
September–October 2026 Vol. 26 No 5 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

doi: 10.17586/2226-1494-2026-26-5-963-970
УДК 004.032.26

Диагностика затухания межслойной связанности и минимальная коррекция остаточными соединениями в глубоких последовательных сетях

Максим Антонович Большим[1]✉, Александр Владимирович Кугаевских[2]

[1,2] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

[1] mabolshim@itmo.ru✉, https://orcid.org/0009-0009-3037-9852
[2] avkugaevskikh@itmo.ru, https://orcid.org/0000-0002-6676-0518

Аннотация

Введение. Проектирование последовательных нейронных сетей требует выбора числа и расположения остаточных соединений, тогда как методы автоматического поиска архитектур ресурсоемки и не локализуют структурные дефекты. В работе предложен способ диагностики локализованного затухания межслойной связанности до обучения и использовать его для минимальной коррекции остаточной топологии. **Метод.** Диагностика основана на внедиагональных блоках матрицы Гессе функции потерь. Введены глобальный индекс коллапса и максимальный локальный провал. По диагностическому профилю жадный алгоритм размещает не более заданного числа остаточных соединений без обучения сети и градиентного поиска по архитектуре. Метод проверен на CIFAR-10 с использованием многослойных перцептронов с узким местом. **Основные результаты.** Исследованы 16 конфигураций с 8, 12, 16 или 20 скрытыми блоками и базовой шириной 64 или 128 (по три инициализации, 48 запусков), а также архитектура с 16 блоками, базовой шириной 256, шириной узкого места 16 и позицией узкого места 8. Коэффициент Спирмена между максимальным локальным провалом и итоговой точностью равен –0,692; *p*-значение менее 0,001. После учета архитектурных параметров статистически значимая связь индексов с итоговой точностью не выявлена. При фиксированном бюджете в одно ребро направленное размещение обеспечивает более высокую точность, чем случайное; *p*-значение равно 0,019 (разведочный анализ). Средние значения мягкого коэффициента Жаккара составляют 0,9583–0,9722. Для соседних целевых блоков 90 % доверительный интервал парной разности точности лежит внутри границ эквивалентности; большее *p*-значение двух односторонних тестов равно 0,014. **Обсуждение.** Гессиановый профиль используется для локализации провала связанности и выбора минимального остаточного вмешательства, а не как самостоятельный предиктор качества обучения. Выводы относятся к сетям с искусственным узким местом; перенос метода на другие архитектуры требует проверки.

Ключевые слова

матрица Гессе, межслойная кривизна, остаточные соединения, диагностика архитектуры, нейронные сети

Благодарности

Работа выполнена в рамках государственного задания (проект FSER-2025-0004).

Ссылка для цитирования: Большим М.А., Кугаевских А.В. Диагностика затухания межслойной связанности и минимальная коррекция остаточными соединениями в глубоких последовательных сетях // Научно-технический вестник информационных технологий, механики и оптики. 2026. Т. 26, № 5. С. 963–970. doi: 10.17586/2226-1494-2026-26-5-963-970

Diagnostics of inter-layer connectivity decay and minimal correction by residual connections in deep sequential networks

Maksim A. Bolshim[1]✉, Alexander V. Kugaevskikh[2]

[1,2] ITMO University, Saint Petersburg, 197101, Russian Federation

[1] mabolshim@itmo.ru✉, https://orcid.org/0009-0009-3037-9852
[2] avkugaevskikh@itmo.ru, https://orcid.org/0000-0002-6676-0518

© Большим М.А., Кугаевских А.В., 2026

Abstract
Deep sequential network design requires choosing the number and placement of residual connections, while architecture-search methods are resource-intensive and do not localize structural defects. The objective is to diagnose pre-training inter-layer coupling decay and use it for minimal residual-topology correction. The procedure analyzes off-diagonal Hessian blocks and introduces a global collapse index and a maximum local drop. A greedy algorithm places residual connections based on the profile without training or gradient-based architecture search. The method is evaluated on multilayer perceptrons with an artificial bottleneck on CIFAR-10. Sixteen configurations with 8, 12, 16, or 20 blocks and base width 64 or 128 were evaluated with three initializations each (48 runs), along with a 16-block architecture of base width 256, bottleneck width 16, and bottleneck position 8. The Spearman correlation between the maximum local drop and final accuracy is –0.692; the *p*-value is less than 0.001. After architectural adjustment, no significant association with accuracy is detected. With a fixed one-edge budget, guided placement yields higher accuracy than random placement; the *p*-value is 0.019 (exploratory analysis). Soft Jaccard coefficients range from 0.9583 to 0.9722. For adjacent target blocks, the 90 % confidence interval for the paired accuracy difference lies within the equivalence bounds; the larger *p*-value of the two one-sided tests is 0.014. The Hessian profile localizes a coupling drop and guides residual intervention rather than serving as an independent predictor of training quality. Conclusions obtained are limited to networks with an artificial bottleneck.

Keywords
hessian matrix, inter-layer curvature, residual connections, architecture diagnostics, neural networks

Acknowledgements
The work was carried out as part of a government assignment (project FSER-2025-0004).

For citation: Bolshim M.A., Kugaevskikh A.V. Diagnostics of inter-layer connectivity decay and minimal correction by residual connections in deep sequential networks. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2026, vol. 26, no. 5, pp. 963–970 (in Russian). doi: 10.17586/2226-1494-2026-26-5-963-970

Введение

Глубокие нейронные сети достигли высокого качества в ряде задач машинного обучения [1], однако проектирование архитектур по-прежнему опирается на эвристики: число и расположение остаточных соединений [2], преодолевающих деградацию обучения глубоких моделей [3, 4], определяются эмпирически. Методы автоматического поиска архитектур [5, 6] требуют значительных вычислительных затрат, а суррогатные метрики нулевой стоимости [7] ранжируют кандидатов, но не локализуют конкретных структурных дефектов.

Анализ ландшафта функции потерь через матрицу Гессе [8, 9] показал, что спектральная структура гессиана отражает свойства обучаемости, однако работы данного направления оперируют преимущественно полным гессианом или его диагональю, не выделяя информацию о взаимодействии между группами параметров различных слоев.

Результаты, полученные авторами в работах [10, 11], показали, что внедиагональные блоки блочной матрицы Гессе — блоки $\mathbf{H}_{i,j}$ при различных индексах i и j — содержат количественную информацию о степени связанности блоков i и j и что эта связанность закономерно затухает с расстоянием $|i-j|$ в последовательных архитектурах.

Цель работы — разработать метод диагностики локализованного затухания межслойной связанности до обучения и использовать диагностический профиль для минимальной коррекции остаточной топологии.

Для достижения цели определены скалярные индексы Γ (глобальный индекс коллапса) и $\Delta_{\max}$ (максимальный локальный провал), вычислимые до обучения на одном калибровочном пакете данных; сформулирован жадный алгоритм размещения остаточных соединений с бюджетом B (максимальным числом добавляемых остаточных соединений); проведена экспериментальная проверка на многослойных перцептронах (MLP) с искусственно заданным узким местом.

Область применения метода ограничена диагностикой локального провала межблочной связанности и выбором позиции остаточного соединения; задача полного автоматического поиска архитектуры в работе не рассматривается.

Диагностическая гипотеза состоит в том, что индексы Γ и $\Delta_{\max}$ характеризуют архитектурно обусловленный режим затухания связанности второго порядка и позволяют локализовать область максимального провала межблочной связанности по координатам (a^*, j^*), где первый индекс соответствует якорному блоку, а второй — целевому блоку; эти координаты относятся к провалу связанности и не обязаны совпадать с позицией узкого блока. Практическая гипотеза состоит в том, что в исследованной конфигурации с узким местом направленное размещение остаточного соединения по диагностическому профилю дает более высокую точность, чем случайное размещение при одинаковом бюджете.

Анализ второго порядка применяется и в оптимизации нейронных сетей: метод кронекерово-факторизованного приближения кривизны (Kronecker-Factored Approximate Curvature, K-FAC) [12] использует блочно-диагональное представление информационной матрицы Фишера, однако отбрасывает межблочные члены $\mathbf{H}_{i,j}$, содержащие информацию о взаимодействии блоков.

В работах [8, 9] исследован спектр полного гессиана как индикатор обучаемости без установления связи между спектральными свойствами отдельных пар блоков и архитектурными дефектами. В [10] предложен послойный гессиан, выявляющий локальные свойства отдельных слоев, а в [11] введена блочная структура $\mathbf{H}_{i,j}$ для архитектур на основе направленных ациклических графов и показано закономерное затухание нормированной связанности $C_{i,j}$ с расстоянием $|i-j|$. Указанное

затухание используется для диагностики структурных дефектов и целенаправленного размещения остаточных соединений.

Модель межслойной связанности и алгоритм коррекции

Блочный гессиан и межслойная связанность. Рассмотрим сеть с L нумерованными скрытыми диагностическими блоками. Пусть $\pi(\mathbf{x})$ — отдельная входная проекция, $\mathbf{h}_0 = b_0(\pi(\mathbf{x}); \theta_0)$, $\mathbf{h}_i = b_i(\mathbf{h}_{i-1}; \theta_i)$, $i = 1, \ldots, L-1$, а c — выходной классификатор. Здесь $\mathbf{x}$ — входной вектор, $\mathbf{h}_i$ — активация на выходе блока i, θ_i — параметры соответствующего блока, y — метка класса. Функцию потерь обозначим $J = \ell(c(\mathbf{h}_{L-1}), y)$. Входной вектор $\mathbf{x}$, проекция π и классификатор c не входят в число L диагностических блоков. Следуя [11], представим входной гессиан в виде блочной матрицы $\mathbf{H} = (\mathbf{H}_{i,j})$, $i, j = 0, \ldots, L-1$, где $\mathbf{H}_{i,j} = \partial^2 J/(\partial \mathbf{h}_i\, \partial \mathbf{h}_j) \in \in \mathbb{R}^{d_i \times d_j}$, а $d_i = \dim \mathbf{h}_i$. Диагональные блоки $\mathbf{H}_{i,i}$ описывают локальную кривизну по активациям блока i; внедиагональные блоки $\mathbf{H}_{i,j}$ при различных индексах i и j отражают межблочную связанность кривизны. Здесь b_i обозначает преобразование i-го диагностического блока, а ℓ — функцию потерь.

Ненормированная связность определяется нормой Фробениуса блока: $R_{i,j} = \|\mathbf{H}_{i,j}\|_F$.

Определение 1 (нормированная связанность).

$$C_{i,j} = \frac{R_{i,j}}{\sqrt{R_{i,i} R_{j,j}} + \varepsilon}, \tag{1}$$

где $R_{i,i} = \|\mathbf{H}_{i,i}\|_F$; ε — малая положительная стабилизирующая константа; диагональную нормированную связанность $C_{i,i}$ полагаем равной 1 при положительном $R_{i,i}$. В расчетах использовано $\varepsilon = 10^{-12}$. Нормированная межслойная связанность определяется формулой (1).

Для кусочно-линейных активаций значения $C_{i,j}$ принадлежат интервалу от 0 до 1 [11]. В используемых архитектурах LayerNorm не является кусочно-линейным преобразованием, поэтому указанная теоретическая гарантия формально не распространяется на рассматриваемый случай; в проведенных экспериментах значений выше 1 не наблюдалось. Блок j исключается из вычисления профиля как мертвый, если $R_{j,j}$ меньше порога δ; использовано $\delta = 10^{-6}$.

Профиль затухания и диагностические индексы

Определение 2 (профиль затухания). Для якорного блока a относительный профиль затухания определяется формулой:

$$S_a(j) = \frac{C_{a,j}}{C_{a,a+1} + \varepsilon}, \quad j > a. \tag{2}$$

По определению $S_a(a+1)$ близко к 1; существенно меньшие значения $S_a(j)$ свидетельствуют о затухании связанности.

На основе профиля определим два диагностических индекса.

Определение 3 (диагностические индексы). Глобальный индекс коллапса при якоре a, равном 0, т. е. первом нумерованном скрытом блоке:

$$\Gamma = \frac{1}{L-1} \sum_{j=1}^{L-1} [-\ln(S_0(j) + \varepsilon)]. \tag{3}$$

Максимальный локальный провал (оптимизация по всем якорям):

$$\Delta_{\max} = \max_a \max_{j>a} [-\ln(S_a(j) + \varepsilon)]. \tag{4}$$

Индекс Γ, определенный формулой (3), измеряет среднюю скорость затухания связанности от первого диагностического блока. Индекс $\Delta_{\max}$, определенный формулой (4), фиксирует пару (a^*, j^*) с максимальным локальным провалом; a^* — индекс якорного блока, j^* — индекс целевого блока. Оба индекса вычисляются до обучения на одном калибровочном пакете данных. Вычислительная стоимость составляет $O(L^2 K)$ произведений матрицы Гессе на вектор, где K — число проб Хатчинсона; в экспериментах использовано $K = 30$.

Стохастическое оценивание. Норма $R_{i,j} = \|\mathbf{H}_{i,j}\|_F$ оценивается через тождество Хатчинсона [13] с радемахеровыми зондами $\mathbf{z} \in \{-1, +1\}^{d_j}$. Произведение $\mathbf{H}_{i,j}\mathbf{z} = \nabla_{\mathbf{h}_i}[(\nabla_{\mathbf{h}_j} J)^\top \mathbf{z}]$ вычисляется без явного хранения блока по методу [14]. Оценка $\hat{R}^2_{i,j} = K^{-1} \sum_{k=1}^{K} \|\mathbf{H}_{i,j}\mathbf{z}_k\|_2^2$ имеет математическое ожидание, равное $\|\mathbf{H}_{i,j}\|_F^2$. В отличие от блочно-диагонального приближения K-FAC [12], здесь используются внедиагональные блоки гессиана функции потерь.

Жадный алгоритм размещения

Алгоритм 1. Жадное размещение остаточных соединений.

Вход: профиль $\{C_{i,j}\}$; бюджет B — максимально допустимое число добавляемых ребер; порог τ; ширины диагностических блоков $\{w_i\}$.

Выход: множество остаточных ребер E, где s и d — индексы исходного и целевого блоков соответственно.

1: $a \leftarrow 0$; $E \leftarrow \varnothing$
2: для $t = 1, \ldots, B$:
3: вычислить $S_a(j)$ для $j > a$ по формуле (2)
4: $j^* \leftarrow \min\{j > a : S_a(j) < \tau\}$
5: если j^* не найден:
6: выход из цикла ▷ провала нет
7: конец если
8: $(s, d) \leftarrow (a, \max(j^* - 1, a + 1))$ ▷ блок перед провалом
9: если $w(s) \neq w(d)$:
10: если существует совместимый целевой блок:
11: $d \leftarrow \max\{r \in [s+1, d] : w(r) = w(s)\}$ ▷ ближайший совместимый блок
12: иначе, если существует совместимый исходный блок:
13: $s \leftarrow \min\{r \in [s+1, d) : w(r) = w(d)\}$
14: иначе: выход из цикла ▷ совместимой пары нет
15: конец если
16: $E \leftarrow E \cup \{(s, d)\}$; $a \leftarrow d$
17: конец цикла.

Алгоритм 1 реализует принцип минимального вмешательства: остаточное соединение $\mathbf{h}_d \leftarrow \mathbf{h}_d + \lambda \mathbf{h}_s$ с коэффициентом λ, равным 1, добавляется только при обнаружении провала $S_a(j)$ ниже порога τ и направляется в ближайший предшествующий разрыву блок со-

вместимой ширины. В алгоритме t — номер итерации, r — индекс проверяемого блока. Порог τ, равный 0,3, соответствует падению нормированной связанности до 30 % от базового уровня связанности между якорным и соседним блоками, т. е. относительному затуханию не менее 70 %, и используется как эмпирический критерий выраженного разрыва.

Эксперименты и результаты

Общий протокол. Все эксперименты проведены на вариантах многослойного перцептрона, выбранных для упрощения поблочного анализа: PlainMLP(L, W) содержит L скрытых диагностических блоков Linear → LayerNorm → ReLU базовой ширины W; BottleneckMLP(L, W, W_b, q) отличается тем, что блок с номером q имеет выходную ширину W_b; SparseSkipMLP дополняет одну из этих архитектур множеством остаточных соединений $\mathbf{h}_d \leftarrow \mathbf{h}_d + \mathbf{h}_s$ при совпадении ширин исходного и целевого блоков. Параметр q обозначает номер блока узкого места. Буква p в статистических результатах обозначает p-значение соответствующего критерия.

Набор данных и обучение. Использован CIFAR-10 [15]: цветные изображения размером 32 × 32 пикселя с тремя цветовыми каналами (красным, зеленым и синим) преобразовывались в векторы размерности 3072. Обучение выполнялось в течение 15 эпох методом стохастического градиентного спуска (SGD) с импульсом 0,9, начальной скоростью обучения 0,01 и косинусным планировщиком [16]. Размер обучающего пакета составлял 128 изображений; применялось ограничение евклидовой нормы градиента значением 1. Для диагностики использовались первые 32 изображения калибровочного пакета. Порог обучаемости T_α — номер эпохи, на которой точность впервые достигает уровня α: в эксперименте A уровень α составлял 0,4, в экспериментах B–F — 0,5. Если не указано иное, использовались три случайные инициализации 42, 43 и 44; эксперимент F выполнен для десяти инициализаций. После знака ± приведено описательное стандартное отклонение с делителем, равным числу наблюдений в соответствующей группе.

В табл. 1 и 3 p означает p-значение соответствующего статистического критерия. Для коэффициента Спирмена нулевая гипотеза — отсутствие ранговой корреляции; для критерия Манна–Уитни — равенство распределений; для критерия Уилкоксона — нулевая медиана парных разностей. В эксперименте F эквивалентность проверяется относительно границ –0,0094 и +0,0094 двумя односторонними парными t-тестами.

Эксперимент A (корреляция индексов с обучаемостью). Исследованы 16 конфигураций: число диагностических блоков L принимало значения 8, 12, 16 и 20; базовая ширина W — 64 или 128; узкое место либо отсутствовало, либо имело ширину W_b, равную 16, и находилось в середине сети, в позиции q, равной $L/2$. Для каждой конфигурации использованы три случайные инициализации, всего 48 запусков; в каждом запуске глобальный индекс коллапса и максимальный локальный провал вычислялись на одном калибровочном пакете до начала обучения.

Для максимального локального провала коэффициент Спирмена с итоговой точностью равен $\rho = -0{,}692$, а со временем достижения порога — $\rho = +0{,}729$. Для глобального индекса коллапса соответствующие коэффициенты равны –0,431 и +0,408 (табл. 1, рисунок).

После учета числа блоков L, базовой ширины W и наличия узкого места коэффициенты при Г и $\Delta_{\max}$ статистически незначимы: p-значения равны 0,38 и 0,94 соответственно; коэффициент детерминации полной модели R^2 равен 0,882. Следовательно, в данной модели статистически значимая связь индексов с итоговой точностью после учета указанных архитектурных параметров не выявлена. При этом профиль $S_a(j)$ сохраняет информацию о координатах локального провала (a^*, j^*).

Эксперимент B (направленное размещение). Использована архитектура BottleneckMLP(16, 256, 16, 8): 16 диагностических блоков, базовая ширина 256, ширина узкого места 16, позиция узкого места 8. Для этой архитектуры отдельная входная проекция переводит 3072-мерный входной вектор в представление ширины 256. Бюджет B равен 2. Сравнивались пять групп: plain без остаточных соединений; random и periodic с двумя соединениями; guided, использующая алгоритм 1 и одно найденное соединение; dense_ref — стратегия сравнения с пятью соединениями, не ограниченная бюджетом B, равным 2.

Группа guided достигает наибольшей средней точности и минимального T_α, используя одно соединение — меньше заданного бюджета (табл. 2). Алгоритм выбирает ребро (0, 13) для инициализации 42 и ребро (0, 14) для инициализаций 43 и 44. Эти значения являются результатом применения жадного алгоритма к диагностическому профилю. Индекс 0 обозначает выход блока 0 после отдельной входной проекции, а не исходный вектор изображения. Блоки 0, 13 и 14 имеют ширину 256, поэтому дополнительное преобразование размерности на остаточной ветви не требуется.

Таблица 1. Ранговые корреляции Спирмена [17] для 48 запусков

Table 1. Spearman rank correlations [17] for 48 runs

Пара переменных	Коэффициент Спирмена ρ	p-значение
Г и итоговая точность	–0,431	0,002
Г и T_α	+0,408	0,004
$\Delta_{\max}$ и итоговая точность	–0,692	менее 0,001
$\Delta_{\max}$ и T_α	+0,729	менее 0,001

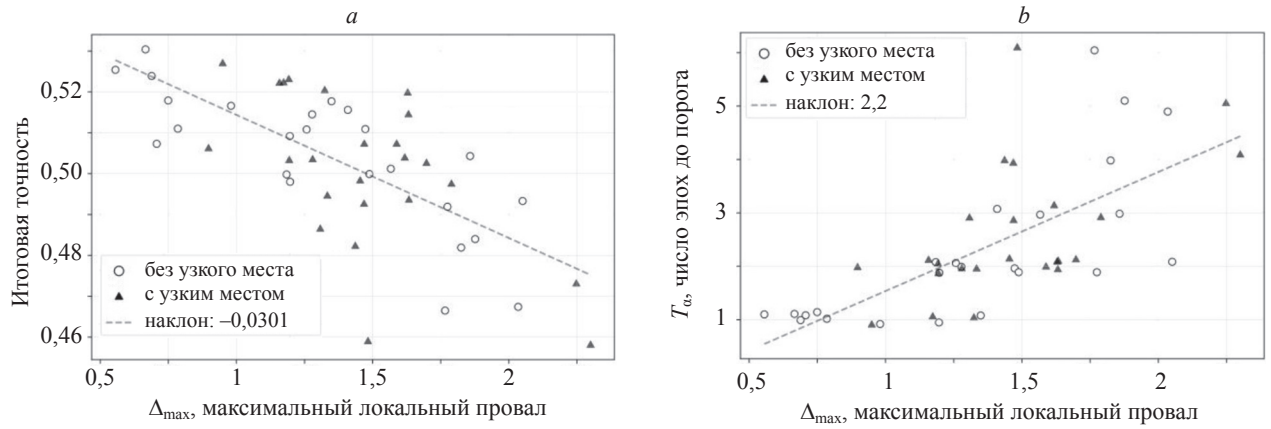


Рисунок. Зависимости итоговой точности (*a*) и времени достижения порога T_α (*b*) от Δ_{max} в эксперименте A для 48 запусков. Линейные регрессии (штриховые линии)

Figure. Final accuracy (*a*) and epochs to threshold T_α (*b*) versus Δ_{max} in Experiment A for 48 runs. Linear regressions (dashed lines)

Таблица 2. Итоговая точность и время достижения порога по группам (среднее ± стандартное отклонение для трех случайных инициализаций; эксперимент B)

Table 2. Final accuracy and time to threshold by group (mean ± standard deviation for three random initializations; Experiment B)

Группа	Число остаточных соединений	Точность (среднее ± стандартное отклонение)	T_α, эпох (среднее ± стандартное отклонение)
plain	0	0,5176 ± 0,0044	7,00 ± 0,82
random	2	0,5238 ± 0,0063	5,33 ± 0,47
periodic	2	0,5279 ± 0,0026	4,67 ± 0,47
guided	1	0,5357 ± 0,0028	3,00 ± 0,00
dense_ref	5	0,5238 ± 0,0010	5,33 ± 0,47

Сравнение guided с random и periodic в эксперименте B смешивает различия стратегии и числа ребер; влияние этого фактора отдельно проверяется в эксперименте E.

Эксперименты C и D (локализация и стабильность)

Эксперимент C (точность локализации). На архитектуре, использованной в эксперименте B, обучены группы guided (3 наблюдения), random (30 наблюдений) и shifted (смещение целевого блока на два блока в обе стороны, 30 наблюдений). Для guided и random критерий Манна–Уитни [18] дает p-значение 0,058; это значение превышает уровень значимости 0,05, а малая численность группы guided ограничивает мощность сравнения. Для guided и shifted p-значение равно 0,66. Статистически значимое различие guided и shifted не выявлено; поэтому эксперимент C не подтверждает единственность выбранной позиции. Поскольку варианты random и shifted вложены в три базовые инициализации, p-значения Манна–Уитни здесь интерпретируются как разведочные. Сравнение с random при фиксированном бюджете приведено в эксперименте E.

Эксперимент D (стабильность диагностики). Для трех случайных инициализаций 42, 43 и 44 и трех 128-элементных тестовых пакетов рассчитаны девять рекомендаций; одни и те же три пакета использовались для всех инициализаций, в каждом пакете анализировались первые 32 изображения. Каждая рекомендация содержала одно ребро — (0, 13) или (0, 14). Согласованность оценивалась коэффициентом Жаккара [19] и мягким коэффициентом Жаккара, для которого близость двух ребер задается функцией $\sigma((s_1, d_1), (s_2, d_2)) = \max\{0,1 - (|s_1 - s_2| + |d_1 - d_2|)/L\}$. Первая и вторая координаты каждого ребра в этой функции соответствуют индексам исходного и целевого блоков. Средний коэффициент Жаккара между инициализациями при фиксированном пакете равен 0,3333, между пакетами при фиксированной инициализации — 0,5556; среднее по всем 36 попарным сравнениям девяти рекомендаций равно 0,4444. Соответствующие средние значения мягкого коэффициента Жаккара равны 0,9583, 0,9722 и 0,9653. Диагностика устойчиво определяет область размещения с точностью до одного блока.

Эксперимент E (сравнение при фиксированном бюджете). Для отделения эффекта стратегии от числа ребер все сравниваемые стратегии используют одно остаточное соединение: guided_k1 и periodic_k1 имеют по три наблюдения, random_k1 — десять случайных вариантов для каждой из трех базовых инициализаций, всего 30 наблюдений; группа plain не содержит остаточных соединений и имеет три наблюдения. Варианты random_k1, полученные при одной базовой инициализации, не рассматриваются как независимые повторения. В табл. 3 n обозначает число наблюдений в группе, k — число остаточных соединений.

Таблица 3. Точность и время достижения порога при одном остаточном соединении. Разведочное сравнение Манна–Уитни: guided_k1 против random_k1; *p*-значение для точности 0,019, для времени достижения порога 0,001 (эксперимент E)

Table 3. Accuracy and time to threshold with one residual connection. Exploratory Mann–Whitney comparison: guided_k1 vs random_k1; *p*-value 0.019 for accuracy and *p*-value 0.001 for time to threshold (Experiment E)

Группа	*n*	*k*	Точность (среднее ± стандартное отклонение)	T_α, эпох (медиана)	*p*-значение по точности относительно guided_k1
plain	3	0	0,5137 ± 0,0031	7	отсутствует
guided_k1	3	1	0,5356 ± 0,0041	3	отсутствует
periodic_k1	3	1	0,5323 ± 0,0039	4	0,200
random_k1	30	1	0,5262 ± 0,0067	6	0,019

При фиксированном бюджете в одно ребро средняя точность в группе guided_k1 выше, чем в группе random_k1; критерий Манна–Уитни дает *p*-значение 0,019 (табл. 3). Для времени достижения порога *p*-значение равно 0,001. Из-за вложенности вариантов random_k1 в три базовые инициализации оба значения следует трактовать как результаты разведочного анализа, а не как подтверждение на уровне независимых повторений. Различие guided_k1 и periodic_k1 по точности статистически незначимо (*p*-значение 0,20): периодическое ребро (0, 9) также пересекает узкое место. Индекс 0 в этих ребрах обозначает выход блока 0 после отдельной входной проекции.

Эксперимент F (эквивалентность близких ребер). Для десяти случайных инициализаций модель обучена попарно с ребрами (0, 13) и (0, 14) (табл. 4). Анализ выполнен по десяти парным разностям точности. Граница эквивалентности 0,0094 выбрана на основе независимого эксперимента E как разность средней точности 0,5356 для guided_k1 и 0,5262 для random_k1. Документального подтверждения фиксации этой границы до анализа результатов эксперимента F нет. Описательные стандартные отклонения в табл. 4 вычислены с делителем n; стандартная ошибка и доверительные интервалы для среднего рассчитаны по выборочному стандартному отклонению с делителем $n - 1$. 95 % доверительный интервал среднего построен по распределению Стьюдента при девяти степенях свободы; 90 %-й интервал использован в проверке эквивалентности. Процедура эквивалентности состоит из двух односторонних парных t-тестов относительно границ –0,0094 и +0,0094; в таблице приведено большее из двух односторонних *p*-значений.

Критерий знаковых рангов Уилкоксона [20] не выявил статистически значимого различия. В процедуре двух односторонних тестов [21] большее одностороннее *p*-значение равно 0,014; 90 % доверительный интервал средней парной разности [–0,0008; +0,0076] целиком находится внутри границ эквивалентности [–0,0094; +0,0094].

Обсуждение

Предсказательная роль диагностических индексов ограничена: после учета числа блоков, базовой ширины и наличия узкого места статистически значимая связь Γ и $\Delta_{\max}$ с итоговой точностью в используемой модели не выявлена. Профиль $S_a(j)$ содержит информацию о координатах локального провала (a^*, j^*); эти координаты относятся к области провала связанности и не обязаны совпадать с номером узкого блока. Профиль используется для выбора позиции остаточного соединения. Остаточная ветвь создает дополнительный путь распространения сигнала и градиента в обход узкого участка и тем самым может увеличивать межблочную связанность $C_{i,j}$. Проверка причинного характера этого механизма выходит за рамки настоящего исследования.

В исследованном семействе моделей BottleneckMLP направленное размещение одного остаточного соединения дает более высокую среднюю точность, чем случайное размещение при одинаковом бюджете; *p*-значение в разведочном анализе эксперимента E равно

Таблица 4. Парное сравнение ребер для десяти случайных инициализаций и процедура двух односторонних тестов с границей эквивалентности 0,0094 (эксперимент F)

Table 4. Paired comparison of edges for ten random initializations and two one-sided equivalence tests with equivalence margin 0.0094 (Experiment F)

Ребро	Точность	T_α, эпох
(0, 13)	0,5387 ± 0,0040	3,10 ± 0,30
(0, 14)	0,5353 ± 0,0053	3,00 ± 0,00
Средняя парная разность	+0,0034 ± 0,0069	+0,10 ± 0,30
95 % доверительный интервал средней парной разности	[–0,0018; +0,0086]	отсутствует
Критерий Уилкоксона, статистика / *p*-значение	14,0/0,19	0,0/1,00
Два односторонних теста, *p*-значение	0,014	отсутствует
90 % доверительный интервал средней парной разности	[–0,0008; +0,0076]	отсутствует

0,019. Согласованность диагностических рекомендаций при изменении инициализации и калибровочного пакета высока: средние значения мягкого коэффициента Жаккара составляют 0,9583–0,9722 (эксперимент D).

Предложенная диагностика использует блочную структуру гессиана, индексируемую по скрытым блокам, и правило совместимости соединений по ширине. Для сверточных сетей ширине диагностического блока может соответствовать число каналов; для трансформеров диагностическими блоками могут служить выходы последовательных модулей. Экспериментальная проверка такого переноса за пределами исследованных многослойных перцептронов не проводилась.

Исследование имеет следующие ограничения. Все эксперименты проведены на многослойных перцептронах с CIFAR-10 и искусственно заданным узким местом. Порог τ, равный 0,3, выбран эмпирически; систематический анализ чувствительности к этому порогу не проводился. Диагностические индексы не показали самостоятельной статистически значимой связи с итоговой точностью после учета включенных архитектурных параметров. В экспериментах C и E несколько случайных вариантов получены в рамках каждой из трех базовых инициализаций, поэтому стандартные *p*-значения Манна–Уитни не учитывают эту зависимость и трактуются как разведочные; подтверждающее сравнение требует большего числа независимых инициализаций. В эксперименте C сравнение guided с random имеет *p*-значение 0,058 и не достигает уровня статистической значимости 0,05; малый размер группы guided ограничивает мощность этого сравнения. Основной интервенционный результат получен для одного остаточного ребра; режим с несколькими остаточными ребрами требует отдельной проверки.

Заключение

Предложен метод локализации областей выраженного провала межблочной связанности в глубоких последовательных сетях с гетерогенной шириной блоков, основанный на анализе внедиагональных блоков матрицы Гессе. Введены глобальный индекс коллапса и максимальный локальный провал, а также жадный алгоритм размещения остаточных соединений с заданным бюджетом. Экспериментальная проверка на многослойных перцептронах с искусственным узким местом показала возможность применения диагностических индексов для локализации провала связанности; статистически значимая связь индексов с итоговой точностью после учета архитектурных параметров не установлена.

При фиксированном бюджете в одно ребро средняя точность при направленном размещении выше, чем при случайном размещении; *p*-значение в разведочном анализе эксперимента E равно 0,019. Диагностические рекомендации устойчивы к изменению инициализации и калибровочного пакета: средние значения мягкого коэффициента Жаккара составляют 0,9583–0,9722 в эксперименте D. Для соседних целевых блоков 13 и 14 90 % доверительный интервал средней парной разности точности целиком лежит внутри выбранных границ эквивалентности; большее *p*-значение двух односторонних тестов равно 0,014 в эксперименте F.

Перенос метода на сверточные и трансформерные архитектуры и анализ чувствительности к порогу размещения остаточных соединений требуют отдельной экспериментальной проверки.

Литература

1. LeCun Y., Bengio Y., Hinton G. Deep learning // Nature. 2015. V. 521. N 7553. P. 436–444. doi: 10.1038/nature14539
2. He K., Zhang X., Ren S., Sun J. Deep Residual learning for image recognition // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 770–778. doi: 10.1109/CVPR.2016.90
3. Glorot X., Bengio Y. Understanding the difficulty of training deep feedforward neural networks // Proc. of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS). 2010. V. 9. P. 249–256.
4. Schoenholz S.S., Gilmer J., Ganguli S., Sohl-Dickstein J. Deep information propagation // Proc. of the International Conference on Learning Representations. 2017.
5. Zoph B., Le Q.V. Neural architecture search with reinforcement learning // Proc. of the International Conference on Learning Representations. 2017.
6. Liu H., Simonyan K., Yang Y. DARTS: Differentiable architecture search // Proc. of the International Conference on Learning Representations. 2019.
7. Mellor J., Turner J., Storkey A., Crowley E.J. Neural architecture search without training // Proc. of the 38th International Conference on Machine Learning, PMLR. 2021. V. 139. P. 7588–7598.
8. Sagun L., Evci U., Güney V.U., Dauphin Y., Bottou L. Empirical analysis of the hessian of over-parametrized neural networks // arXiv. 2017. arXiv:1706.04454. doi: 10.48550/arXiv.1706.04454
9. Yao Z., Gholami A., Keutzer K., Mahoney M.W. PyHessian: neural networks through the lens of the hessian // Proc. of the IEEE International Conference on Big Data. 2020. P. 581–590. doi: 10.1109/BigData50022.2020.9378171

References

1. LeCun Y., Bengio Y., Hinton G. Deep learning. *Nature*, 2015, vol. 521, no. 7553, pp. 436–444. doi: 10.1038/nature14539
2. He K., Zhang X., Ren S., Sun J. Deep Residual learning for image recognition. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90
3. Glorot X., Bengio Y. Understanding the difficulty of training deep feedforward neural networks. *Proc. of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2010, vol. 9, pp. 249–256.
4. Schoenholz S.S., Gilmer J., Ganguli S., Sohl-Dickstein J. Deep information propagation. *Proc. of the International Conference on Learning Representations*, 2017.
5. Zoph B., Le Q.V. Neural architecture search with reinforcement learning. *Proc. of the International Conference on Learning Representations*, 2017.
6. Liu H., Simonyan K., Yang Y. DARTS: Differentiable architecture search. *Proc. of the International Conference on Learning Representations*, 2019.
7. Mellor J., Turner J., Storkey A., Crowley E.J. Neural architecture search without training. *Proc. of the 38th International Conference on Machine Learning, PMLR*, 2021, vol. 139, pp. 7588–7598.
8. Sagun L., Evci U., Güney V.U., Dauphin Y., Bottou L. Empirical analysis of the hessian of over-parametrized neural networks. *arXiv*, 2017. arXiv:1706.04454. doi: 10.48550/arXiv.1706.04454
9. Yao Z., Gholami A., Keutzer K., Mahoney M.W. PyHessian: neural networks through the lens of the hessian. *Proc. of the IEEE International Conference on Big Data*, 2020, pp. 581–590. doi: 10.1109/BigData50022.2020.9378171

10. Bolshim M.A., Kugaevskikh A.V. Local properties of neural networks through the lens of layer-wise Hessians // arXiv. 2025. arXiv:2510.17486. doi: 10.48550/arXiv.2510.17486
11. Bolshim M.A., Kugaevskikh A.V. Inter-layer Hessian analysis of neural networks with DAG architectures // arXiv. 2026. arXiv:2604.11639. doi: 10.48550/arXiv.2604.11639
12. Martens J., Grosse R. Optimizing neural networks with Kronecker-factored approximate curvature // Proc. of the 32nd International Conference on Machine Learning. 2015. V. 37. P. 2408–2417.
13. Hutchinson M.F. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines // Communications in Statistics — Simulation and Computation. 1990. V. 19. N 2. P. 433–450. doi: 10.1080/03610919008812866
14. Pearlmutter B.A. Fast Exact multiplication by the Hessian // Neural Computation. 1994. V. 6. N 1. P. 147–160. doi: 10.1162/neco.1994.6.1.147
15. Krizhevsky A. Learning Multiple Layers of Features from Tiny Images. Technical Report. University of Toronto, 2009. 60 p.
16. Loshchilov I., Hutter F. SGDR: Stochastic gradient descent with warm restarts // Proc. of the International Conference on Learning Representations (ICLR). 2017.
17. Spearman C. The Proof and measurement of association between two things // The American Journal of Psychology. 1904. V. 15. N 1. P. 72–101. doi: 10.2307/1412159
18. Mann H.B., Whitney D.R. On a test of whether one of two random variables is stochastically larger than the other // The Annals of Mathematical Statistics. 1947. V. 18. N 1. P. 50–60. doi: 10.1214/aoms/1177730491
19. Jaccard P. The Distribution of the flora in the Alpine zone // New Phytologist. 1912. V. 11. N 2. P. 37–50. doi: 10.1111/j.1469-8137.1912.tb05611.x
20. Wilcoxon F. Individual comparisons by ranking methods // Biometrics Bulletin. 1945. V. 1. N 6. P. 80–83. doi: 10.2307/3001968
21. Schuirmann D.J. A comparison of the Two One-Sided Tests Procedure and the Power Approach for assessing the equivalence of average bioavailability // Journal of Pharmacokinetics and Biopharmaceutics. 1987. V. 15. N 6. P. 657–680. doi: 10.1007/BF01068419

10. Bolshim M.A., Kugaevskikh A.V. Local properties of neural networks through the lens of layer-wise Hessians. *arXiv*, 2025. arXiv:2510.17486. doi: 10.48550/arXiv.2510.17486
11. Bolshim M.A., Kugaevskikh A.V. Inter-layer Hessian analysis of neural networks with DAG architectures. *arXiv*, 2026. arXiv:2604.11639. doi: 10.48550/arXiv.2604.11639
12. Martens J., Grosse R. Optimizing neural networks with Kronecker-factored approximate curvature. *Proc. of the 32nd International Conference on Machine Learning*, 2015, vol. 37, pp. 2408–2417.
13. Hutchinson M.F. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics – Simulation and Computation*, 1990, vol. 19, no. 2, pp. 433–450. doi: 10.1080/03610919008812866
14. Pearlmutter B.A. Fast Exact multiplication by the Hessian. *Neural Computation*, 1994, vol. 6, no. 1, pp. 147–160. doi: 10.1162/neco.1994.6.1.147
15. Krizhevsky A. *Learning Multiple Layers of Features from Tiny Images*. Technical Report. University of Toronto, 2009, 60 p.
16. Loshchilov I., Hutter F. SGDR: Stochastic gradient descent with warm restarts. *Proc. of the International Conference on Learning Representations (ICLR)*, 2017.
17. Spearman C. The Proof and measurement of association between two things. *The American Journal of Psychology*, 1904, vol. 15, no. 1, pp. 72–101. doi: 10.2307/1412159
18. Mann H.B., Whitney D.R. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 1947, vol. 18, no. 1, pp. 50–60. doi: 10.1214/aoms/1177730491
19. Jaccard P. The Distribution of the flora in the Alpine zone. *New Phytologist*, 1912, vol. 11, no. 2, pp. 37–50. doi: 10.1111/j.1469-8137.1912.tb05611.x
20. Wilcoxon F. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1945, vol. 1, no. 6, pp. 80–83. doi: 10.2307/3001968
21. Schuirmann D.J. A comparison of the Two One-Sided Tests Procedure and the Power Approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, 1987, vol. 15, no. 6, pp. 657–680. doi: 10.1007/BF01068419

Авторы

Большим Максим Антонович — младший инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0009-0009-3037-9852, mabolshim@itmo.ru
Кугаевских Александр Владимирович — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, SC 56442745400, https://orcid.org/0000-0002-6676-0518, avkugaevskikh@itmo.ru

Authors

Maksim A. Bolshim — Junior Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0009-0009-3037-9852, mabolshim@itmo.ru
Alexander V. Kugaevskikh — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 56442745400, https://orcid.org/0000-0002-6676-0518, avkugaevskikh@itmo.ru

Статья поступила в редакцию 22.04.2026
Одобрена после рецензирования 19.08.2026
Принята к печати 19.09.2026

Received 22.04.2026
Approved after reviewing 19.08.2026
Accepted 19.09.2026

Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»