

**ВЛИЯНИЕ СТРУКТУРНО-ОРИЕНТИРОВАННОЙ
АУГМЕНТАЦИИ НА ДИНАМИКУ ИНФОРМАЦИОННОГО
СЖИМАЮЩЕГОСЯ ПОТОКА В ГИБРИДНЫХ
НЕЙРОСЕТЯХ**

Зверева Анна Константиновна

Аспирант

Московский физико-технический институт, Москва, Россия

E-mail: zvereva.ak@phystech.edu

Научный руководитель — Грабовой А.В.

Гибридные нейросетевые архитектуры, объединяющие пространственные кодировщики типа визуального трансформера [3] и рекуррентные модули типа сетей долгой краткосрочной памяти [5, 7] (ViT-LSTM), показывают высокую эффективность в анализе последовательностей функциональной магнитно-резонансной томографии (фМРТ). Однако при дефиците размеченных медицинских данных такие модели склонны к переобучению [1]. В работе исследуется влияние структурно-ориентированной аугментации (введение синтетических аномалий и масок) на динамику обработки сигнала с позиций теории информационного сжимающегося потока [6].

Оценка потоков информации на l -ом слое энкодера T_l производилась через вычисление прокси-метрик: энтропии активаций $\hat{I}(X; T_l)$ для оценки степени сжатия и диагональной аппроксимации матрицы Фишера $\hat{I}(T_l; Y)$ для оценки релевантности признаков:

$$\hat{I}(X; T_l) \approx - \sum_i \sigma(t_i) \log_2 \sigma(t_i), \quad \hat{I}(T_l; Y) \approx \sum_i \left(\frac{\partial \mathcal{L}}{\partial t_i} \right)^2$$

где $\sigma(t_i)$ — softmax-нормализованные активации слоя, \mathcal{L} — функция потерь кросс-энтропии.

Эксперименты на сбалансированной выборке фМРТ (202 объекта, включая 101 объект класса «норма» [4] и 101 объект класса «патология» [2]) показали, что базовая архитектура ViT-LSTM без аугментации характеризуется высокой информационной избыточностью. Оценка $\hat{I}(X; T_l)$ снижается незначительно, а градиент релевантности $\hat{I}(T_l; Y)$ остается низким на всей глубине сети. Это свидетельствует о низкой способности модели выделять специфические паттерны на фоне высокой вариативности сигналов здоровых областей.

Внедрение предложенной аугментации со структурными масками кардинально меняет информационную траекторию (рис. 1).

Иллюстрации

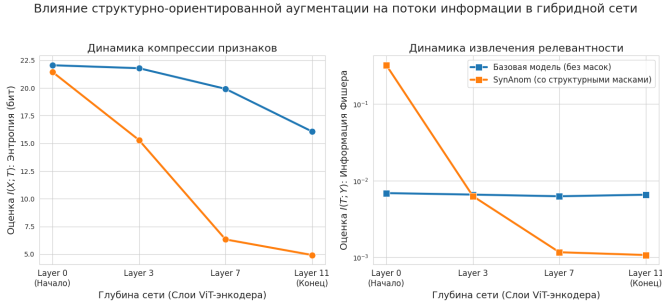


Рис. 1. Динамика показателей $\hat{I}(X; T_i)$ (сжатие) и $\hat{I}(T_i; Y)$ (релевантность) в слоях Vision Transformer.

Маски выступают в роли сильного индуктивного смещения, провоцируя резкий рост извлечения предиктивной информации уже на начальном слое трансформера (рост $\hat{I}(T_0; Y)$ на два порядка). После локализации релевантных признаков на старте, последующие слои внимания переходят в режим агрессивной компрессии: энтропия $\hat{I}(X; T_i)$ стремительно падает до 4.92 бит. Таким образом, теоретико-информационный анализ подтверждает, что структурная аугментация качественно трансформирует стратегию сети, формируя оптимальный информационный сжимающийся поток.

В заключение автор выражает глубокую признательность научному руководителю Грабовому А.В. и научному консультанту Каприеловой М.С. за помощь в постановке задач, ценные советы и обсуждение результатов.

Литература

1. Воронцов К.В. Математические методы обучения по прецедентам. М., 2020.
2. Damir V. et al. Pre-Post rehabilitation fMRI data. OpenNeuro, 2022.
3. Dosovitskiy A. et al. An image is worth 16x16 words. ICLR, 2021.
4. Hanke M. et al. Forrest Gump. OpenNeuro, 2018.
5. Hochreiter S., Schmidhuber J. Long short-term memory. 1997.

6. Shwartz-Ziv R., Tishby N. Opening the black box of deep neural networks. 2017.
7. Zvereva A.K., Grabovoy A.V., Kaprielova M.S. Dynamic division of labor in hybrid AI // Doklady RAN. Math. Inf. Proc. Upr. 2025. Vol. 527. P. 117-133.