

МОДЕЛЮВАННЯ ЧАСОВИХ ПОСЛІДОВНОСТЕЙ З ВИКОРИСТАННЯМ РЕВЕРСИВНИХ ТРАНСФОРМЕРІВ

¹ Вінницький національний технічний університет

Анотація

Показано актуальність проблеми оптимізації споживання пам'яті під час навчання моделей глибокого навчання для обробки часових послідовностей. Запропоновано архітектуру реверсивного часового трансформера, яка дозволяє відв'язати об'єм споживаної пам'яті від довжини часової послідовності завдяки механізму часово-агностичного зворотного поширення помилки.

Ключові слова: штучний інтелект, машинне навчання, глибоке навчання, часові послідовності, трансформери, реверсивні архітектури, ефективність пам'яті.

Abstract

There has been shown the relevance of the problem of optimizing memory consumption during the training of deep learning models for temporal sequence processing. A reversible temporal transformer architecture is proposed, which decouples memory footprint from temporal depth using a time-agnostic backpropagation strategy.

Keywords: artificial intelligence, machine learning, deep learning, temporal sequences, transformers, reversible architectures, memory efficiency.

Сфера генеративного моделювання демонструє стрімкий прогрес, проте значним викликом залишається застосування цих моделей до часових послідовностей та даних, що безперервно еволюціонують (наприклад, генерація відео, робототехніка, довгострокове прогнозування). Головним "вузьким місцем" у цих доменах є величезне споживання пам'яті глибокими нейромережевими архітектурами, особливо під час навчання на довгих послідовностях [1].

Традиційні підходи, такі як рекурентні нейронні мережі (RNN, LSTM), страждають від проблеми зникаючого градієнта [2]. Архітектури на базі Трансформерів краще моделюють довгострокові залежності, але вартість їхнього механізму "самоуваги" (self-attention) квадратично зростає залежно від довжини послідовності [3]. Стандартний метод зворотного поширення помилки у часі (Backpropagation Through Time) вимагає збереження всіх проміжних активацій під час прямого проходу, що призводить до швидкого вичерпання фізичної пам'яті при збільшенні горизонту прогнозування.

Для вирішення цієї проблеми було розроблено реверсивний часовий трансформер TempVerseFormer. Натхненням для підходу став біологічний механізм "зворотного відтворення" (reverse replay) у гіпокампі, коли мозок ретроспективно переоцінює минулі стани в контексті нових даних [4]. Архітектура TempVerseFormer складається з трьох ключових модулів:

- Варіаційний автоенкодер (VAE) [5], який стискає вхідні кадри (зображення) у менш вимірний латентний простір.
- Реверсивний часовий трансформер, який працює з послідовностями латентних векторів.
- Механізм часового зв'язування, який дозволяє ітеративно подавати вихідні дані моделі з попереднього кроку як вхідні дані для наступного, формуючи цикл зворотного зв'язку.

Загальна архітектура запропонованого підходу наведена на рисунку 1. Ключовою інновацією підходу є використання часово-агностичної стратегії зворотного поширення помилки (Time-Agnostic Backpropagation). На відміну від традиційних мереж, реверсивні блоки TempVerseFormer математично сконструйовані таким чином, що вони є строго інвертованими. Це означає, що під час прямого проходу мережі не потрібно зберігати проміжні активації у пам'яті. Натомість, під час зворотного проходу при поширенні помилки, необхідні активації реконструюються на основі вихідних даних з попереднього блока. Схема роботи цього механізму зображена на рисунку 2.

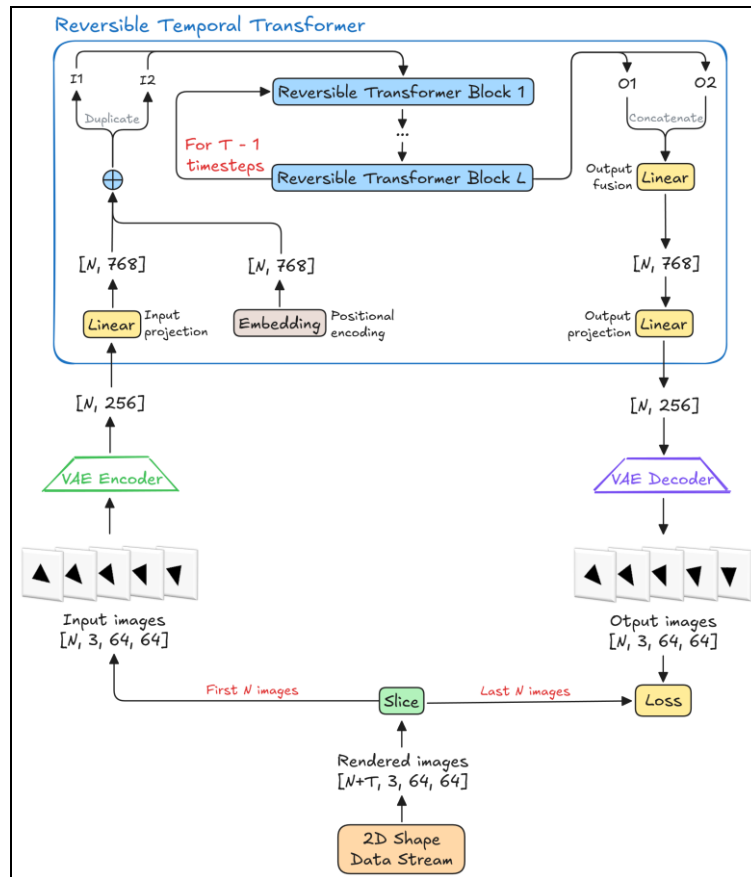


Рисунок 1 – Архітектура TempVerseFormer

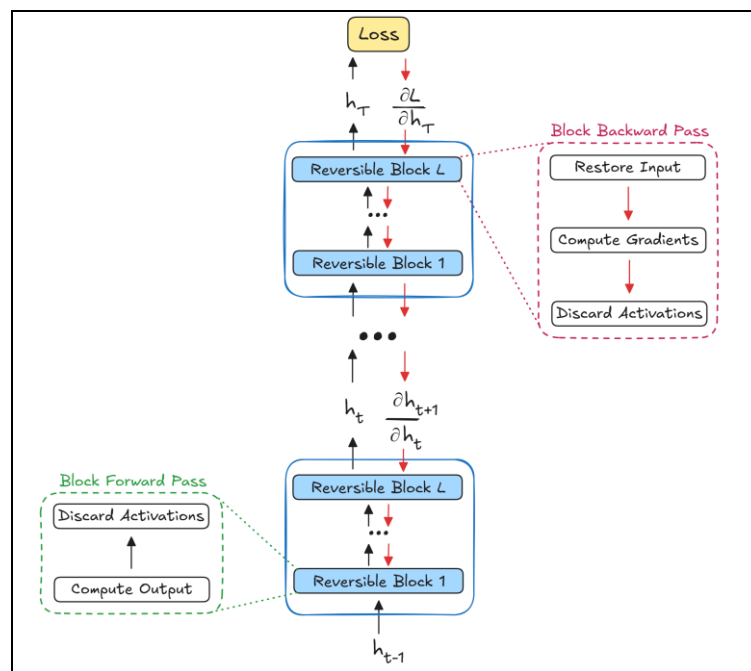


Рисунок 2 – Стратегія зворотного поширення помилки

Завдяки комбінації реверсивних блоків та ітеративного часового зв'язування, споживання пам'яті стає практично незалежним від того, на скільки кроків уперед робиться прогноз. Пам'ять витрачається лише на активації одного блоку в один момент часу, а не на всю історію часової послідовності.

Висновки

Показано, що проблема експоненційного зростання пам'яті при обробці довгих часових послідовностей може бути ефективно вирішена за допомогою реверсивних трансформерів. Запропонований підхід, TempVerseFormer, дозволяє генерувати довгострокові прогнози з константним споживанням пам'яті відносно довжини послідовності. Детальний опис архітектури, математичного підґрунтя реверсивного методу, а також результати вичерпних експериментів на синтетичних та реальних просторово-часових датасетах, які підтверджують високу ефективність підходу, наведено у роботі [6].

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Aidan N. Gomez, Mengye Ren, et al “The Reversible Residual Network: Backpropagation Without Storing Activations” Advances in Neural Information Processing Systems, 30, NIPS 2017.
2. Razvan Pascanu, Tomas Mikolov, Yoshua Bengio “On the difficulty of training recurrent neural networks” Proceedings of the 30th International Conference on Machine Learning, 2013.
3. Ashish Vaswani, Noam Shazeer, et al “Attention is all you need” Advances in Neural Information Processing Systems, 30, NIPS 2017.
4. David J. Foster, Matthew A. Wilson “Reverse replay of behavioural sequences in hippocampal place cells during the awake state” Nature, 440(7084), 2006.
5. Diederik P. Kingma, Max Welling “Auto-Encoding Variational Bayes” International Conference on Learning Representations (ICLR), 2014.
6. Leonid Kulyk “Temporal modeling with reversible transformers” Neural Computing and Applications, 2026.

Кулик Леонід Русланович – аспірант факультету інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: leonidkulik2707@gmail.com

Мокін Олександр Борисович – доктор технічних наук, професор, професор кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: abmokin@vntu.edu.ua

Kulyk Leonid – a graduate student, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, email: leonidkulik2707@gmail.com

Mokin Oleksandr – Dr. Sc. (Eng.), Professor of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: abmokin@vntu.edu.ua