УДК 004.81

СОБЫТИЙНО-ТОКЕННАЯ U-ОБРАЗНАЯ АРХИТЕКТУРА НА ОСНОВЕ СЕЛЕКТИВНЫХ STATE-SPACE БЛОКОВ ДЛЯ ДЕТЕКЦИИ И КЛАССИФИКАЦИИ СОБЫТИЙ В 1D-СИГНАЛАХ

К.А. АФАНАСЕНКО (магистрант УО «БГУИР»), Н.А. ЛАРЧЕНКО (студент УО «БГУИР»), Майорова Е.А. (студент УО «БГУИР»)

Введение. Задачи анализа одномерных сигналов с временной разметкой включают выделение событий (с указанием времени начала и конца), их классификацию по типам, а также получение агрегированной оценки состояния по окну наблюдения или по всей записи. Такие постановки возникают при мониторинге технических объектов, телеметрии, акустике, киберфизических системах и иных источниках данных, где информативные эпизоды локализованы во времени и часто разрежены относительно фона. Для практического применения важны две группы требований: точная локализация границ событий, поскольку ошибка в разметке по времени приводит к смещению интервалов и ухудшению интерпретации и устойчивое моделирование дальнего контекста, поскольку решение может зависеть от структуры сигналов на больших временных масштабах.

Классические сверточные подходы и их темпоральные варианты показали пригодность для широкого класса задач последовательностей, но качество и устойчивость на длинных зависимостях зависят от рецептивного поля и способов агрегации признаков. U-образные энкодер-декодер архитектуры широко используются для задач плотной разметки благодаря сочетанию многошкальных признаков и связей пропуска (skip-connections), позволяющих сохранять локальные детали при наличии контекстной информации [1]. В то же время модели на основе self-attention получили высокую гибкость и качество в задачах последовательностей, однако базовая операция внимания имеет квадратичную зависимость по длине последовательности, что затрудняет использование на длинных окнах без специальных модификаций.

Цель настоящей статьи состоит в формулировке архитектуры, ориентированной на совместное решение задач плотной временной сегментации, детекции интервалов событий и классификации состояния, с акцентом на длинные последовательности и разреженные события. Предлагается событийно-токенная U-образная архитектура (далее – EGSS-U-TokenNet), использующая селективные state-space блоки в качестве основного механизма дальнего контекста и контур токен-уточнения для объектного представления событий.

Основная часть. Пусть наблюдается многоканальный дискретный сигнал $x \in R^{T \times C}$, где T – длина окна в отсчетах, C – число каналов. Разметка по времени задается как множество интервалов событий $\{(k_i, t_s^{(i)}, t_e^{(i)})\}$, где $k_i \in \{1, \ldots, K\}$ – класс события, $t_s^{(i)}$, $t_e^{(i)}$ – начало и конец. Требуется получить плотную сегментацию $\hat{y}(t) \in \{0, \ldots, K\}$ для каждого t (0 – фон), предсказанные интервалы событий с оценкой уверенности $\hat{p}_j$, агрегированную метку или распределение по состояниям $\hat{c}$, для окна/записи (например, режим работы, наличие дефекта, класс контекста).

U-образные энкодер-декодер структуры являются базовым решением для плотной разметки, поскольку энкодер формирует многошкальные признаки, а декодер восстанавливает временное разрешение с использованием skip-соединений, что концептуально аналогично U-Net [2]. Для 1D-сегментации временных рядов ранее демонстрировались модификации U-подхода, обрабатывающие длинные записи и выдающие плотные карты классов с последующей агрегацией под доступную разметку, что подтверждает применимость U-принципа в 1D-сценариях [3].

Однако в задачах с длинным контекстом и разреженными событиями возникает противоречие: локальная точность границ требует сохранения высокочастотных деталей, а корректная классификация события и окна может зависеть от дальних зависимостей. Стандартный Transformer опирается на self-attention, который обеспечивает контент-зависимую агрегацию, но имеет неблагоприятную масштабируемость по длине последовательности в базовом виде. SSM-семейство предлагает альтернативу: моделирование последовательности через скрытое состояние, что позволяет представлять длинные зависимости; S4 формализует эффективный вычислительный путь для структурированных SSM. Селективные SSM (Mamba) вводят входозависимое управление переносом/забыванием информации, сохраняя линейную зависимость вычислений от длины [1].

Указанные соображения приводят к выбору следующей комбинации: U-образная многошкальная схема для точной локализации; state-space блоки в качестве механизма дальнего контекста на каждом масштабе; событийные токены как объектное представление, позволяющее агрегировать информацию о событиях, связать сегментацию и классификацию через единый “объектный” слой, реализовать контур уточнения, корректирующий плотные карты на основе компактной событийной структуры.

Архитектура включает четыре логических подсистемы:

1) Стем и многошкальный энкодер. Вход xxx проходит через начальные 1D-свертки (или эквивалентный локальный миксер) для выделения короткозависимых паттернов. Далее последовательность последовательно downsample-ится по времени (stride/pooling), формируя уровни s = 0, …, S с длиной $T_s = T/2^s$ и размерностью признаков D_s. На каждом уровне применяются остаточные блоки вида “локальный миксер + EGSS”.

2) EGSS-блок (Event-Gated Selective State-Space). EGSS рассматривается как модуль дальнего контекста, реализующий рекуррентную динамику скрытого состояния в стиле SSM, но с входозависимым управлением вкладом текущего входа (селективность). Концептуально это соответствует идее селективных SSM, где параметры/гейты зависят от входа, что применяется для управления переносом информации вдоль времени. В отличие от attention-механизма, EGSS не требует явной матрицы попарных взаимодействий размера T×T\, что является ключевым ограничением базового Transformer-подхода на длинных последовательностях [4].

3) Декодер и связи пропуска. Декодер последовательно upsample-ит признаки до исходного временного разрешения, объединяя их со skip-признаками энкодера. Этот принцип соответствует U-подходу, применяемому в задачах сегментации, и переносится на 1D-сигналы аналогично существующим 1D U-архитектурам.

4) Событийные токены и контур уточнения. Из промежуточного представления на выбранном масштабе (например, среднем s=s*) вычисляется набор NNN токенов $z_n \in R^{Dz}$, где каждый токен должен соответствовать одному событию-кандидату. Сбор токена выполняется как взвешенное суммирование признаков по времени вокруг центра μ_n:

$$z_n = \sum_{t=1}^{T_{s^*}} w_n(t)h(t), \qquad w_n(t) \infty \exp\left(-\frac{(t-\mu_n)^2}{2\sigma^2}\right), \tag{1}$$

где h(t) – признаки на масштабе s*, σ – параметр ширины. Центры μ_n и/или маски w_n предсказываются отдельной токенной головой. Далее токены проходят через token-mixer (например, легкую attention- или SSM-подобную агрегацию) и возвращаются в декодер через механизм кросс-влияния (cross-gating или кросс-смешивание), формируя корректирующий сигнал Δh(t) для уточнения сегментации и границ. Данный контур предназначен для согласования “объектного” уровня (события как сущности) и “пиксельного” уровня (плотная разметка).

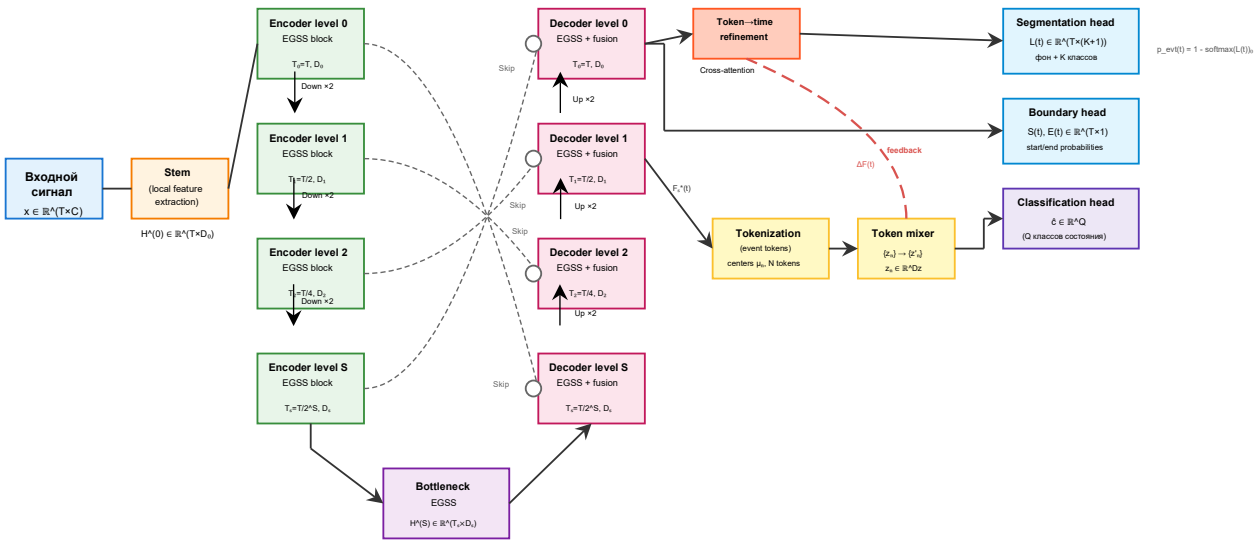


Рисунок 1 – Общая схема EGSS-U-TokenNet

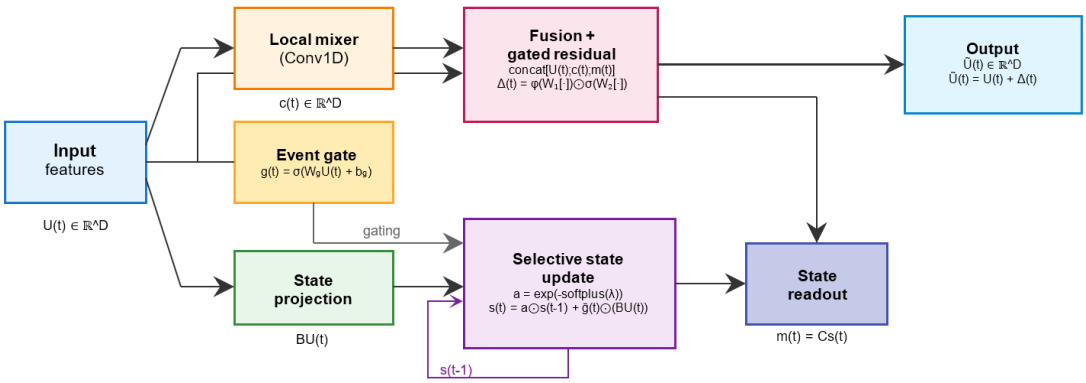
Архитектура формирует три основных выхода:

Первый выход – плотная сегментация $\hat{y}(t)$ или распределение $p(y_t|x)$, обучаемое по кросс-энтропии (с учетом дисбаланса классов). Второй выход – карты границ $p_s(t)$, $p_e(t)$ для начала и конца события, что упрощает извлечение интервалов и повышает чувствительность к локальным переходам. Третий выход – классификация окна/записи $\hat{c}$, вычисляемая либо по pooled-признакам, либо по агрегированному набору токенов (например, через attention-пулинг по $\{z_n\}$.

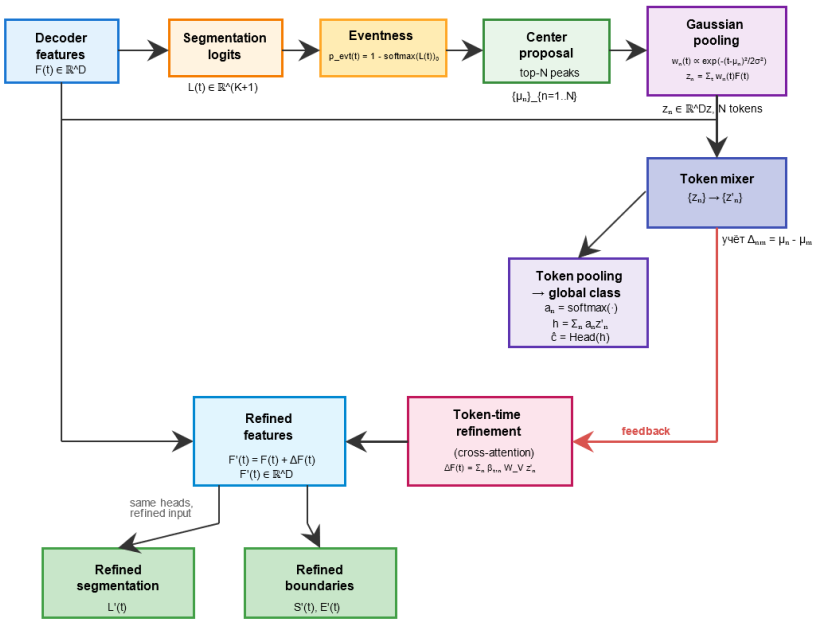
Суммарная функция потерь может быть записана в виде:

$$L=\alpha L_{seg} + \beta L_{bd} + \gamma L_{tok} + \delta L_{cls}, \qquad (2)$$

где L_{seg} – потери сегментации, L_{bd} – потери по границам (например, BCE для p_s, p_e или регрессия на расстояние до границы), L_{tok} – согласование токенов с истинными событиями (например, matching по IoU интервалов или по близости центров), L_{cls} – потери классификации. Подбор коэффициентов α, β, γ, δ в экспериментальной части может выполняться через нормировку вкладов по масштабам градиентов или через сопоставление средних значений компонент на начальных эпохах.



а)



б)

Рисунок 2 – EGSS-блок и контур token→time уточнения:
а) EGSS-блок; б) Контур token→time уточнений

Заключение. В данной работе сформулирована концепция архитектуры EGSS-U-TokenNet для анализа одномерных сигналов с временной разметкой, ориентированная на совместное получение плотной разметки по времени, оценок границ начала и конца событий и итоговой классификации состояния по окну наблюдения. Архитектура объединяет многошкальную U-образную схему, обеспечивающую сохранение локальной точности на исходном временном разрешении, с механизмом дальнего контекста на базе селективного state-space обновления, а также вводит событийно-токенное представление.

Ключевым элементом предложенного подхода является контур token→time уточнения, в котором токены событий, извлечённые из промежуточных признаков и согласованные с плотным предсказанием, используются для корректировки временных признаков и, как следствие, уточнения сегментации и границ. Такая постановка позволяет описывать выход модели в виде структурированного набора событий-кандидатов, что делает возможной дальнейшую агрегацию по окну и формирование интерпретируемых характеристик.

В рамках дальнейшего исследования целесообразно выполнить экспериментальную валидацию архитектуры в сравнении с базовыми темпоральными сверточными моделями и attention-ориентированными подходами, а также провести абляционный анализ влияния селективного state-space блока и токенного контура на качество локализации и классификации. Дополнительно требуется зафиксировать протокол извлечения интервалов из плотных карт и правила сопоставления токенов с истинными событиями, чтобы обеспечить воспроизводимость результатов и однозначную интерпретацию архитектурного вклада.

Список литературы

1. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2015. P. 234–241. DOI: 10.1007/978-3-319-24574-4_28. – Дата доступа: 14.12.2025.

2. Perslev M., Jensen M. H., Darkner S., Jennum P. J., Igel C. U-Time: A Fully Convolutional Network for Time Series Segmentation Applied to Sleep Staging // Advances in Neural Information Processing Systems (NeurIPS 2019). 2019. – Дата доступа: 14.12.2025.

3. Vaswani A., Shazeer N., Parmar N., et al. Attention Is All You Need. arXiv:1706.03762. 2017. – Дата доступа: 15.12.2025.

4. Gu A., Goel K., Ré C. Efficiently Modeling Long Sequences with Structured State Spaces. arXiv:2111.00396. 2021. – Дата доступа: 15.12.2025.