



Классификатор временных треков автомобильного радара с многомасштабной адаптацией к дисбалансу классов

Е. А. Лопухова^а, младший научный сотрудник, orcid.org/0000-0001-6802-6010

Г. С. Воронков^а, канд. техн. наук, доцент, orcid.org/0000-0002-8788-2696, voronkov.gs@ugatu.su

Е. П. Топольская^а, канд. техн. наук, доцент, orcid.org/0000-0003-2207-2702

^аУфимский университет науки и технологий, Заки Валиди ул., 32, Уфа, 450076, РФ

Введение: распознавание дорожных объектов по данным автомобильного радара осложняется малым числом отражений в кадре и сильной неравномерностью классов, из-за которой редкие объекты хуже представлены в обучении. **Цель:** разработать классификатор радарных треков, устойчивый к дисбалансу классов и применимый к открытым наборам дорожного движения. **Методы:** для проведения исследований разработан и применен метод классификации радарных треков, сочетающий кадровую обработку неупорядоченных множеств радарных отражений разделяемым многослойным персептроном с инвариантной агрегацией признаков, временную трансформерную агрегацию с обучаемым классификационным токеном и прототипную классификационную голову в нормированном косинусном пространстве. Дисбаланс классов учитывался с помощью отступа, зависящего от распределения меток, классово-зависимого температурного масштабирования, весов функции потерь, рассчитанных по эффективному числу образцов, и взвешенной выборки при формировании мини-батчей; дополнительно применялось вспомогательное предсказание статистических моментов радиальной скорости. **Результаты:** эксперименты выполнены на наборах RadarScenes и Boreas-Objects-V1. RadarScenes содержит 158 последовательностей с поточечной разметкой пяти классов, а Boreas-Objects-V1 — 4699 треков с трехмерной разметкой четырех классов, включая редчайший класс с долей 1,5 %. На Boreas-Objects-V1 сбалансированная точность выросла с 0,945 до 0,970, а полнота редчайшего класса — с 0,833 до 0,917. Абляционное исследование показало, что отключение временной агрегации снижает сбалансированную точность на RadarScenes с 0,998 до 0,327, что указывает на принадлежность последовательной информации к ключевым источникам признаков, а не к вспомогательным элементам модели. **Практическая значимость:** классификатор пригоден для обучения систем радарного распознавания дорожных объектов в условиях ограниченного числа примеров редких классов.

Ключевые слова — автомобильный радар, радарный трек, распознавание дорожных объектов, дисбаланс классов, временная агрегация, метрическое обучение, нейронные сети.

Для цитирования: Лопухова Е. А., Воронков Г. С., Топольская Е. П. Классификатор временных треков автомобильного радара с многомасштабной адаптацией к дисбалансу классов. *Информационно-управляющие системы*, 2026, № 4, с. 60–74. doi:10.31799/1684-8853-2026-4-60-74, EDN: GGKFJQ

For citation: Lopukhova E. A., Voronkov G. S., Topolskaya E. P. Temporal track classifier for automotive radar with multi-scale adaptation to class imbalance. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2026, no. 4, pp. 60–74 (In Russian). doi:10.31799/1684-8853-2026-4-60-74, EDN: GGKFJQ

Введение

Работоспособность функций активной безопасности транспортного средства (ТС) напрямую зависит от того, насколько надежно бортовые сенсоры распознают дорожную обстановку при переменных условиях видимости. Камеры и лидары дают высокое угловое и пространственное разрешение, но в тумане, при сильном дожде или в ночное время они деградируют ощутимо: рассеяние лидарного сигнала на аэрозоле растет с оптической глубиной трассы, а визуальное распознавание страдает от потери контраста и перекрытий [1]. В отечественных и зарубежных работах по интеллектуальным транспортным системам отмечается та же проблема: устойчивое восприятие дорожной сцены требует сенсоров с принципиально разной физической природой сигнала [2–4]. Автомобильный радар в подобных условиях держится заметно лучше — длина волны

миллиметрового диапазона существенно превышает характерный размер тумановых капель, что резко снижает рассеяние по сравнению с оптическими сенсорами [2, 5, 6]. Именно поэтому радар остается одним из ключевых компонентов систем помощи водителю при неблагоприятных погодных условиях [2, 3, 5].

У радарных данных есть своя принципиальная особенность, которую нельзя игнорировать. Один кадр от движущегося объекта, как правило, дает одну-три точки отражения, т. е. никакой пространственной информации, сопоставимой с плотным лидарным облаком, там нет [6]. Дискриминирующая сила сосредоточена в другом месте: динамика радиальной скорости, смещение отражений между кадрами, доплеровские сигнатуры, характерные для конкретного типа движения — все это проявляется именно во временном контексте [7, 8]. Распознавание по траекторным признакам последовательностей отражений давно

применяется в задачах классификации дорожных пользователей и в более широких задачах радарного распознавания [9]. Это означает, что агрегирование признаков по последовательности кадров — не дополнительная опция, а структурно необходимый элемент любого классификатора радарных треков.

Для обработки неупорядоченных наборов точек широко используется архитектура PointNet, обеспечивающая инвариантность к перестановке детекций за счет общего поточечного перцептрона и симметричного объединения признаков; на наборе ModelNet40 она достигла 89,2 % общей точности классификации [10]. Ее расширение PointNet++ извлекает локальную структуру на нескольких масштабах и повышает точность на том же наборе до 91,9 % [11]. В российских работах по нейросетевой классификации радарных объектов также активно разрабатываются подходы к формированию информативных представлений, инвариантных к порядку поступления отражений и плотности облака [12]. В рамках радарного восприятия близкие идеи были использованы в RadarGNN, где графовая нейронная сеть была применена к данным RadarScenes. В результате макроусредненный F1, т. е. среднее гармоническое точности и полноты, одинаково усредненное по всем классам независимо от их частоты, составил 77,1 % в задаче сегментации, а средняя прецизионность, усредненная по классам и порогам перекрытия с разметкой, составила 60,2 % в задаче обнаружения объектов [13].

Для моделирования последовательностей эффективным решением является трансформер с обучаемым токеном-агрегатором Classifier Token (classification token, CLS), который через механизм самовнимания аккумулирует информацию всей последовательности и формирует компактное представление для классификации [14, 15]. Важно, что трансформерная обработка последовательностей относится не к узкоспециализированным решениям для одной предметной области, а к междисциплинарному классу методов: в отечественных исследованиях она применяется, в частности, к задаче классификации пользовательских запросов при определении профильного врача, что подтверждает переносимость механизма самовнимания и классификационного токена на данные различной природы [16]. В задачах радарного распознавания близкую роль выполняют темпоральные модули, например RadarTCN, где последовательная обработка радарных кадров используется для онлайн-классификации автомобильных радарных целей [17].

Радарный сигнал от движущегося объекта накапливается приемником в виде последовательности наборов отражений — радарного трека. Именно на уровне всего трека и выполняется

классификация: пешеход, велосипедист, легковой автомобиль или тяжелое ТС. Присвоенная метка задает поведенческую модель объекта в прогнозирующем алгоритме и прямо влияет на формирование предупреждения о столкновении — ошибка здесь обходится дорого [18, 8]. Однако в реальном трафике распределение классов катастрофически неравномерно: стандартный классификатор тяготеет к доминирующей категории (прежде всего — легковым автомобилям), тогда как редкие и уязвимые классы: пешеходы, группы пешеходов, велосипедисты — получают слишком слабый градиентный сигнал [18, 19].

Отдельное направление исследований посвящено адаптации обучения к дисбалансу классов. Эта проблема имеет междисциплинарный характер: она возникает не только в задачах радарного распознавания, но и в других прикладных областях машинного обучения, где редкие, но практически значимые классы оказываются недостаточно представлены в обучающей выборке [19]. Поэтому методические выводы таких работ: необходимость контролировать качество на миноритарных классах, использовать специальные метрики и применять процедуры балансировки — могут быть перенесены на классификацию радарных треков, где редкие классы дорожных объектов также требуют отдельного контроля полноты [18, 19]. Взвешивание функции потерь через эффективное число образцов снижает ошибку на несбалансированном CIFAR-10 при коэффициенте дисбаланса 100 с 29,64 до 20,73 % [20], а метод Label-Distribution-Aware Margin (LDAM) с аддитивным отступом, зависящим от распределения меток, в схеме LDAM-DRW дает ошибку 14,82 % на том же варианте CIFAR-10 [21]. Современные обзоры по обучению на данных с длиннохвостым распределением (long-tailed learning) показывают, что основные способы работы с редкими классами включают ребалансировку выборки, модификацию функции потерь, обучение представлений и постобработку классификационной головы [22].

В метрическом обучении ArcFace применяет косинусный угловой отступ и достигает высокой точности верификации лиц за счет явного разделения классов в нормированном эмбединговом пространстве [23]. Близкая идея классификации по расстоянию до эталонов реализована в прототипических сетях, где каждому классу сопоставляется вектор-прототип, а решение принимается по близости объекта к этим прототипам [24]. Однако совместного применения покадровой обработки разреженных облаков точек, временной агрегации, метрической классификационной головы и адаптации к дисбалансу классов в единой модели для классификации радарных треков перечисленные подходы не рассматривали.

Для заполнения этого пробела предлагается классификатор временных радарных треков (КВРТ) с многомасштабной адаптацией к дисбалансу классов. Архитектура объединяет: по-кадровый кодировщик, в котором каждая точка радарного облака обрабатывается одним и тем же многослойным перцептроном с разделяемыми между точками весами, а поточечные признаки затем сводятся в единый вектор кадра симметричной функцией агрегации, инвариантной к порядку точек [10, 12]; временной трансформер с CLS [14–16]; классификационную голову с косинусной метрикой [23] и обучаемыми векторами-эталоном классов [24]. Эффективность и вклад компонентов исследуются на двух открытых наборах данных:

1) RadarScenes — 158 последовательностей и пять классов: car (легковой автомобиль), pedestrian (пешеход), pedgroup (группа пешеходов), twowheeler (двухколесное ТС) и largevehicle (крупногабаритное ТС), — при этом доля редчайшего класса twowheeler не превышает 5 % треков [18];

2) Boreas-Objects-V1 — 4699 треков четырех классов: car (45,7 %), pedestrian (48,5 %), cyclist (велосипедист, 4,3 %) и misc (прочие объекты, 1,5 %, или 70 треков) [25].

Абляционное исследование количественно проверяет, является ли временная агрегация структурно необходимой, достаточен ли мощный кодировщик без адаптации к дисбалансу и какие компоненты балансировки дают измеримый прирост точности на редких классах.

Методология

Основная концепция метода заключается в том, чтобы классифицировать не отдельные радарные точки или кадры, а весь радарный трек как временную последовательность разреженных множеств отражений. Для этого применяется трехступенчатая архитектура: 1) кодирование множества точек внутри каждого кадра; 2) агрегирование показанных признаков по времени; 3) сравнение трека с обучаемыми прототипами классов.

Адаптация к дисбалансу классов осуществляется не одним методом, а несколькими согласованными механизмами: весами функции потерь; повторной выборкой редких классов; классово-зависимым отступом; температурным масштабированием.

Наборы данных

Набор данных RadarScenes сформирован с использованием четырех серийных автомобильных радарных датчиков с рабочей частотой 77 ГГц, установленных на одном ТС; дальность обнаружения каждого датчика составляет до 100 м, угловое поле зрения — $\pm 60^\circ$ относительно оси, разрешение по дальности — 0,15 м, разрешение

по радиальной скорости — 0,1 км/ч, средний период обновления — 60 мс (≈ 17 Гц). Каждая детекция описывается шестью признаками: плоскими координатами цели в системе отсчета носителя (x, y), компенсированной радиальной скоростью v_r , эффективной площадью рассеяния (мощностью отражения), дальностью и азимутальным углом; эти же шесть признаков образуют входной вектор в предложенном методе. Набор данных Boreas-Objects-V1 построен на основе датасета Boreas, записанного с использованием сканирующего кругового радара Navtech CIR204-H (диапазон — до 200 м, разрешение по дальности — 0,060 м, угловое разрешение — $0,9^\circ$, частота вращения — 4 Гц); треки сформированы из трехмерных аннотаций объектов и содержат те же шесть признаков детекций, что и при обработке RadarScenes. Минимальный порог для включения трека в обучающую выборку составлял четыре детекции и два кадра; максимальная длина трека ограничена 24 кадрами, среднее число детекций в кадре — два.

Постановка задачи

Задача, рассматриваемая в настоящей работе, обусловлена двумя структурными особенностями автомобильных радарных данных. Первая: одиночный радарный кадр содержит от одной до трех точек отражения — информации для надежной классификации там недостаточно, и распознавание опирается на всю временную последовательность кадров, а не на отдельный «снимок». Вторая: реальный трафик дает сильно несбалансированное распределение классов. В наборе Boreas-Objects-V1 редчайший класс занимает лишь 1,5 % треков — при таком соотношении стандартный классификатор просто не успевает выучить граничные условия для этой категории.

Пусть задано множество классов $\mathcal{C} = \{1, \dots, K\}$, где K обозначает число распознаваемых категорий дорожных объектов. Радарный трек рассматривается как последовательность кадров, упорядоченных по времени:

$$\mathcal{T}_i = (\mathbf{X}_{i,1}, \mathbf{X}_{i,2}, \dots, \mathbf{X}_{i,T_i}, y_i, s_i),$$

$$\mathbf{X}_{i,t} \in \mathbb{R}^{N_{i,t} \times d}, y_i \in \mathcal{C}, \quad (1)$$

где $\mathcal{T}_i$ — i -й трек; $\mathbf{X}_{i,t}$ — матрица признаков t -го кадра этого трека; T_i — длина трека в кадрах; y_i — метка класса; s_i — идентификатор сцены записи; $N_{i,t}$ — число радарных отражений в кадре; d — размерность признакового описания одной детекции. Формула (1) задает структуру входных данных: каждый трек содержит не один вектор, а последовательность матриц переменного размера, поэтому дальнейшая модель должна уметь работать и с множеством точек внутри кадра, и с временной последовательностью кадров.

Чтобы исключить утечку данных, разбиение выполняется не по отдельным трекам, а по сценам записи [25]:

$$\forall T_i \in \mathcal{D}_{train}, \forall T_j \in \mathcal{D}_{test} : s_i \neq s_j,$$

где $\mathcal{D}_{train}$ и $\mathcal{D}_{test}$ — обучающая и тестовая подвыборки треков соответственно, а индексы i и j независимо пробегают треки из этих подвыборок. Это условие означает, что треки из одной и той же сцены не могут одновременно попасть в обучающую и тестовую выборки. Такое ограничение необходимо, поскольку соседние треки в пределах одной сцены могут иметь общий контекст движения, близкие условия наблюдения и схожие ошибки разметки, а значит, случайное разбиение по трекам могло бы завысить оценку качества модели.

Признаки одной детекции имеют разную физическую природу и разные численные масштабы, поэтому перед обучением они нормализуются по статистикам обучающей выборки:

$$\tilde{x}_{rj} = \frac{x_{rj} - \mu_j^{train}}{\sigma_j^{train} + \varepsilon},$$

где x_{rj} — значение j -го признака у r -й реальной детекции; μ_j^{train} и σ_j^{train} — среднее и стандартное отклонения этого признака, вычисленные только на обучающей части; ε — малая положительная константа для предотвращения деления на нуль. Нормализация делает признаки сопоставимыми по масштабу и предотвращает ситуацию, при которой признаки с большими численными значениями начинают доминировать в градиентном обучении только из-за своей размерности.

Кодировщик множества точек

Каждый радарный кадр представляет собой неупорядоченное множество отражений, поэтому его представление не должно зависеть от порядка строк во входной матрице. Для этого каждая детекция обрабатывается одним и тем же разделяемым многослойным персептроном $\phi(\cdot)$ — небольшой нейронной сетью с общими весами для всех точек, после чего точечные признаки объединяются симметричными операциями максимального и среднего агрегирования [10]:

$$\mathbf{h}_t = \psi \left(\left[\begin{array}{c} \max_{n: m_{t,n}=1} \phi(\tilde{\mathbf{x}}_{t,n}); \frac{\sum_{n=1}^{N_{\max}} m_{t,n} \phi(\tilde{\mathbf{x}}_{t,n})}{\sum_{n=1}^{N_{\max}} m_{t,n}} \end{array} \right] \right),$$

где $\mathbf{h}_t$ — итоговый вектор признаков кадра; $\psi(\cdot)$ — проекционный слой; $\phi(\cdot)$ — общий для всех

точек персептрон; $\tilde{\mathbf{x}}_{t,n}$ — нормализованный вектор признаков n -й детекции в t -м кадре; $m_{t,n} \in \{0, 1\}$ — маска, отделяющая реальные детекции от нулевого заполнения. Максимальное агрегирование выделяет наиболее выраженные признаки среди детекций, среднее агрегирование описывает распределенную статистику всего кадра, а их конкатенация позволяет сохранить оба типа информации.

Преимуществом многослойного персептрона является инвариантность к перестановке точек внутри кадра. Если изменить порядок строк в $\mathbf{X}_t$, значения максимума и среднего по множеству реальных детекций не изменятся, поэтому и итоговый вектор $\mathbf{h}_t$ останется тем же. Это свойство важно именно для радарных данных, где порядок записей в таблице детекций не несет физического смысла [10].

Временной преобразователь с агрегирующим служебным токеном

После кодирования отдельных кадров требуется получить одно представление всего трека, поскольку класс объекта часто определяется не по форме одного кадра, а по изменению признаков во времени [7–9, 17, 18]. Для этого используется временной преобразователь на основе самовнимания с обучаемым служебным токеном CLS [14–16]:

$$\mathbf{z} = \text{Transformer}([\mathbf{c}_{\text{CLS}}; \mathbf{h}_1 + \mathbf{a}_1; \dots; \mathbf{h}_{T_{\max}} + \mathbf{a}_{T_{\max}}], \mathbf{m}^{\text{frame}})_{\text{CLS}},$$

где $\mathbf{z}$ — итоговый вектор признаков всего трека; $\mathbf{c}_{\text{CLS}}$ — обучаемый токен агрегирования; $\mathbf{h}_t$ — представление t -го кадра; $\mathbf{a}_t$ — позиционный вектор, задающий номер кадра во временной последовательности; $\mathbf{m}^{\text{frame}}$ — маска реальных и дополненных кадров. Индекс CLS означает, что после обработки трансформером берется не среднее по всем выходам, а выход, соответствующий служебному токenu.

Модель получает последовательность кадровых признаков и с помощью самовнимания оценивает, какие кадры важнее для классификации. Позиционные векторы необходимы, потому что без них трансформер воспринимал бы последовательность как набор элементов без временного порядка.

Классификационная голова на основе сравнения с обучаемыми прототипами классов

После получения трекового вектора $\mathbf{z}$ выполняется сравнение с обучаемыми прототипами классов. Пусть $\mathbf{P} \in \mathbb{R}^{K \times D}$ — матрица прототипов, где строка $\mathbf{p}_k$ соответствует k -му классу,

а D — размерность пространства эмбедингов (embedding dimension). Косинусный логит k -го класса вычисляется как скалярное произведение нормированных векторов:

$$c_k = \left\langle \frac{\mathbf{z}}{\|\mathbf{z}\|_2}, \frac{\mathbf{p}_k}{\|\mathbf{p}_k\|_2} \right\rangle, \quad k = 1, \dots, K,$$

где $\mathbf{z}$ — вектор представления трека; $\mathbf{p}_k$ — вектор-прототип k -го класса, оба нормированы по евклидовой норме; c_k показывает угловое сходство между представлением трека и прототипом k -го класса. Нормирование по евклидовой норме $\|\cdot\|_2$ устраняет влияние длины векторов, поэтому решение зависит прежде всего от направления эмбединга, а не от его масштаба.

После перехода к угловому пространству возникает задача учесть неравномерность числа обучающих треков разных классов. Для этого в работе используется метод LDAM — подход к обучению на несбалансированных выборках, в котором к логиту истинного класса вводится аддитивный отступ, зависящий от представленности класса. В отличие от весов функции потерь, которые изменяют вклад примеров разных классов в значение ошибки, LDAM воздействует на само условие правильной классификации: на этапе обучения логит истинного класса уменьшается на величину m_{y_i} , поэтому модель должна распознать объект с дополнительным запасом по угловому сходству. Для редких классов этот запас задается больше, так как m_k выбирается пропорционально $n_k^{-1/4}$. Такая зависимость следует из анализа обобщающей способности классификатора при дисбалансе меток и позволяет аналитически связать отступ с числом обучающих примеров класса. Таким образом, к логиту истинного класса применяется классово-зависимый аддитивный отступ:

$$\tilde{c}_{ik} = c_{ik} - m_{y_i} \mathbb{I}[k = y_i],$$

где c_{ik} — косинусный логит k -го класса для i -го объекта; m_{y_i} — отступ истинного класса y_i ; $\mathbb{I}[k = y_i]$ — индикатор, равный единице только для правильного класса. Отступ усложняет задачу классификации на обучении: чтобы объект был правильно распознан, логит истинного класса должен быть не просто максимальным, а достаточно большим после вычитания m_{y_i} .

После применения отступа логиты масштабируются классово-зависимой обратной температурой:

$$\ell_{ik} = \tau_k \tilde{c}_{ik}, \quad (2)$$

где ℓ_{ik} — итоговый логит, передаваемый в softmax, а τ_k — обратная температура k -го класса.

Чем больше τ_k , тем более резким становится распределение вероятностей для соответствующего класса, а меньшие значения τ_k делают решение мягче, что полезно для классов с ограниченным числом обучающих примеров.

Аналитическая настройка параметров дисбаланса

Параметры τ_k и m_k задаются через число обучающих образцов класса n_k , а не подбираются независимо для каждого класса. Сначала вычисляется эффективное число образцов k -го класса

$$E_k = \frac{1 - \beta^{n_k}}{1 - \beta},$$

где n_k — фактическое число обучающих треков этого класса; $\beta \in [0, 1)$ — параметр, задающий степень насыщения информации при добавлении похожих объектов. Если класс уже содержит много близких примеров, новый пример увеличивает E_k слабее, чем увеличивает простое количество n_k , поэтому эффективное число отражает не только объем, но и убывающий вклад дополнительных наблюдений.

На основе E_k вычисляется обратная температура:

$$\tilde{\tau}_k = E_k^\alpha, \quad \tau_k = \bar{\tau} \frac{K \tilde{\tau}_k}{\sum_{j=1}^K \tilde{\tau}_j},$$

где $\tilde{\tau}_k$ — ненормированная обратная температура; α — параметр степени зависимости от представленности класса; $\bar{\tau}$ — заданное среднее значение обратной температуры; K — число классов. Нормировка во второй части формулы сохраняет средний масштаб логитов, потому что среднее значение τ_k по классам становится равным $\bar{\tau}$.

Классово-зависимый отступ задается по принципу LDAM:

$$m_k = m_{\max} \frac{n_k^{-1/4}}{\max_j n_j^{-1/4}},$$

где m_k — отступ для k -го класса; $m_{\max}$ — максимальное допустимое значение отступа, а множитель $n_k^{-1/4}$ делает отступ больше для редких классов и меньше для частых. Нормировка на $\max_j n_j^{-1/4}$ гарантирует, что самый большой отступ среди всех классов точно равен $m_{\max}$.

Функция потерь и задача оптимизации

Обучение модели выполняется по сумме основной классификационной ошибки и вспомогательной регрессионной ошибки:

$$L = L_{cls} + \lambda \cdot L_{aux}, \quad (3)$$

где L_{cls} — основная функция потерь классификации; L_{aux} — вспомогательная ошибка предсказания статистических моментов радиальной скорости; λ — коэффициент, задающий вклад вспомогательной задачи. Такая структура связывает обучение классификатора с физическими характеристиками радарного сигнала, поскольку радиальная скорость и ее статистика несут информацию о типе движущегося объекта.

Основная классификационная ошибка записывается как взвешенная перекрестная энтропия (Cross-Entropy, CE):

$$L_{cls} = -\frac{1}{B} \sum_{i=1}^B w_{y_i} \sum_{k=1}^K q_{ik} \log \frac{\exp(\ell_{ik})}{\sum_{j=1}^K \exp(\ell_{ij})},$$

где B — размер мини-батча; w_{y_i} — вес истинного класса i -го объекта; q_{ik} — сглаженная целевая метка; ℓ_{ik} — логит после отступа и температурного масштабирования (2), а дробь внутри логарифма — вероятность класса после softmax. Веса w_{y_i} усиливают вклад редких классов, а сглаживание меток снижает переобучение за счет отказа от полностью жесткой целевой вероятности, равной единице только для истинного класса.

Вспомогательная ошибка имеет вид средне-квадратичной регрессии:

$$L_{aux} = \frac{1}{B} \sum_{i=1}^B \|\hat{\mathbf{s}}_i - \mathbf{s}_i\|_2^2,$$

где $\hat{\mathbf{s}}_i$ — предсказанный моделью вектор статистических моментов радиальной скорости, а $\mathbf{s}_i$ — соответствующий целевой вектор, вычисленный по данным трека. Минимизация этой ошибки заставляет скрытое представление $\mathbf{z}$ сохранять информацию о динамике объекта, что особенно важно для различения классов с похожей геометрией, но различным характером движения [18].

Классовые веса в основной функции потерь вычисляются через эффективное число образцов:

$$w_k = \frac{1/E_k}{\frac{1}{K} \sum_{j=1}^K 1/E_j}, \quad \frac{1}{K} \sum_{k=1}^K w_k = 1. \quad (4)$$

В выражении (4) редкие классы получают больший вес, потому что их эффективное число E_k меньше, а величина $1/E_k$ больше. Нормировка сохраняет средний вес равным единице, поэтому меняется относительный вклад классов, но общий масштаб функции потерь остается стабильным.

Задача оптимизации модели формулируется как минимизация средней функции потерь на обучающем множестве:

$$\theta^* = \arg \min_{\theta} \frac{1}{|\mathcal{D}_{train}|} \sum_{i: T_i \in \mathcal{D}_{train}} L(T_i, y_i; \theta),$$

где θ^* — значения параметров после обучения; θ — совокупность всех обучаемых параметров модели, включая параметры кодировщика, трансформера, прототипов классов и вспомогательной головы. Формула задает стандартную эмпирическую минимизацию риска, но с функцией потерь, специально адаптированной к дисбалансу классов и временной природе радарных треков.

Для выбора лучшей модели используется составная валидационная метрика:

$$Q_{val} = \frac{1}{2}(\text{BA} + F1_{macro}),$$

где BA — сбалансированная точность, т. е. средняя полнота по классам; $F1_{macro}$ — макроусредненная F1-мера, одинаково учитывающая все классы независимо от их частоты. Такая метрика подходит для несбалансированных данных, потому что стандартная точность может быть высокой даже при плохом распознавании редких классов.

Контрольная базовая постановка использует линейную классификационную голову:

$$\mathbf{l}^{base} = \mathbf{W}\mathbf{z} + \mathbf{b},$$

где $\mathbf{l}^{base}$ — вектор логитов по классам; $\mathbf{W}$ — матрица весов линейного классификатора; $\mathbf{b}$ — вектор смещений. Это выражение служит базовым сравнением: в нем отсутствуют прототипы, косинусная нормализация, классово-зависимые отступы и температурное масштабирование.

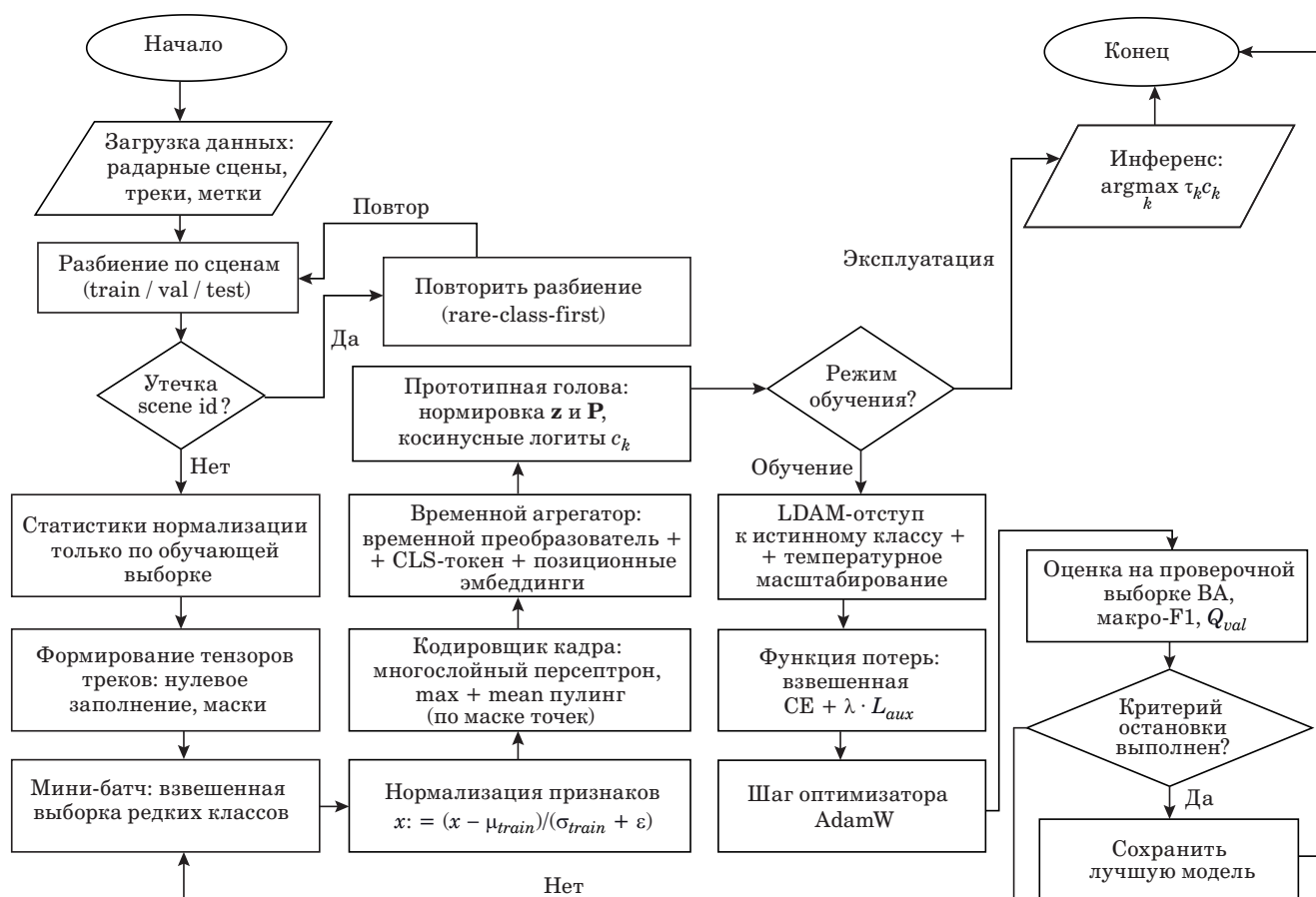
Протокол обучения

Исходные данные делятся по сценам записи в пропорции примерно 60/20/20: 20 % сцен направляются в валидационную выборку, 20 % — в тестовую, остальная часть — в обучающую. После разбиения программно проверяется отсутствие пересечений по идентификаторам сцен и треков между подмножествами: треки из одной записи не должны одновременно оказаться, например, в обучающей и тестовой выборках. Затем по обучающей части вычисляются статистики нормализации, формируются тензоры треков с нулевым заполнением и масками точек, а также рассчитываются классовые параметры n_k , E_k , τ_k , m_k и w_k , которые используются в классификационной голове и функции потерь.

Для уменьшения влияния дисбаланса классов мини-батчи формируются с помощью взвешенной случайной выборки. При таком способе треки редких классов получают большую вероятность попадания в батч, поэтому их вклад чаще учитывается при вычислении градиента. Это дополняет взвешивание функции потерь: выборка управляет составом батча, а веса w_k изменяют вклад классов уже внутри значения ошибки.

В ходе обучения для каждого мини-батча последовательно выполняются нормализация входных признаков, кодирование точек каждого кадра разделяемым многослойным персептроном, временная агрегация показровых признаков, вычисление косинусных логитов, применение LDAM-отступа к логиту истинного класса, температурное масштабирование и расчет полной функции потерь. Параметры модели θ обновляются оптимизатором AdamW. На этапе инференса LDAM-отступ не используется, поскольку истинная метка объекта неизвестна; предсказание выполняется по температурно масштабированным косинусным логитам.

Общая структура предлагаемого алгоритма показана на рис. 1, где наглядно отображены этапы подготовки данных, их разбиение на сцены, формирование мини-батчей, нормализация признаков, покадровое кодирование, временная агрегация, прототипическая классификация, а также разделение режимов обучения и инференса. Схема отражает полный путь обработки данных: от подготовки сцен и формирования батчей до получения трекового представления, классификации и контроля качества на валидационной выборке. Отдельное ветвление подчеркивает различие между режимами обучения и инференса: при обучении используются отступ, функция потерь и обновление параметров, тогда как при инференсе выполняется только прямой проход модели с выбором класса по итоговым логитам. Возврат от критерия ранней остановки к формированию батча показывает итерационный характер обучения и связь процедуры оптимизации с валидационным контролем.



■ Рис. 1. Структурная схема алгоритма КВРТ

■ Fig. 1. Structural scheme of the temporal radar track classifier algorithm

Результаты экспериментов

Все эксперименты проводились на двух открытых наборах данных: RadarScenes (пять классов, 31 619 треков, доля редчайшего класса — *twowheeler* — менее 5 %) и Boreas-Objects-V1 (четыре класса, 4699 треков, экстремальный дисбаланс 66:1 по классу *misc*). На RadarScenes применялась скользящая проверка по записям на трех наборах (*scene-level 3-fold CV*) с тремя значениями инициализации генератора случайных чисел — итого девять независимых запусков. На Boreas-Objects-V1 использовалось единственное стратифицированное разбиение 70/30 по сценам (одно разбиение, два независимых запуска).

Во всех основных экспериментах модель обучалась 30 эпох с ранней остановкой по сбалансированной точности (BA) на валидационном разбиении. Оптимизатор — AdamW, скорость обучения — 3×10^{-4} , размер батча — 64, размерность итогового эмбединга трека (TRACK_DIM) = 256. Базовый метод отличался линейной головой классификатора, равномерными весами функции потерь, случайным разбиением треков, остальная же нейросетевая архитектура была идентична KBPT.

Главным критерием качества является сбалансированная точность, дополненная макро-F1 и поклассовыми показателями полноты. Целевой порог полноты — $\text{recall} \geq 0,80$ по каждому классу.

Общее число обучаемых параметров KBPT составляет около 709 тыс., из которых 84 % сосредоточено в блоке временной агрегации — трехслойном преобразователе с четырьмя головами самовнимания. Кодировщик множества точек содержит 59 тыс. параметров, прототипная классификационная голова добавляет лишь 1280 параметров — по 256 на каждый из пяти классов. Вычислительная стоимость одного прямого прохода, оцененная аналитически для трека из 20 кадров при среднем числе отражений в кадре, равном двум, составляет около 29 млн операций с плавающей точкой. Поскольку классификация выполняется на уровне трека, а не отдельного кадра, задержка вывода не является узким местом: при частоте обновления радара 17 Гц трек из 20 кадров формируется примерно за 1,2 с, тогда как классификация занимает единицы миллисекунд, т. е. менее 1 % времени накопления трека. Это означает, что вычислительная нагрузка модели не является ограничивающим фактором при ее интеграции в систему помощи водителю.

Эксперименты на наборе данных RadarScenes

Модель обучалась и тестировалась с использованием трехкратной перекрестной проверки с тремя случайными инициализациями параметров, что дало девять независимых запусков.

В каждом запуске вычислялись сбалансированная точность (BA), макро-F1 и поклассовые значения полноты (*recall*).

По результатам девяти запусков средние значения составили: $BA = 0,9967 \pm 0,0048$, макро-F1 = $0,9961 \pm 0,0058$. Целевой порог полноты $\text{recall} \geq 0,80$ был достигнут во всех классах во всех девяти запусках. В семи из девяти случаев значение BA равнялось 1,0; в двух оставшихся зафиксированы единичные ошибки классификации: два трека из 95 в разбиении 2 (инициализация 1) были перепутаны между классами *largevehicle* и *twowheeler*.

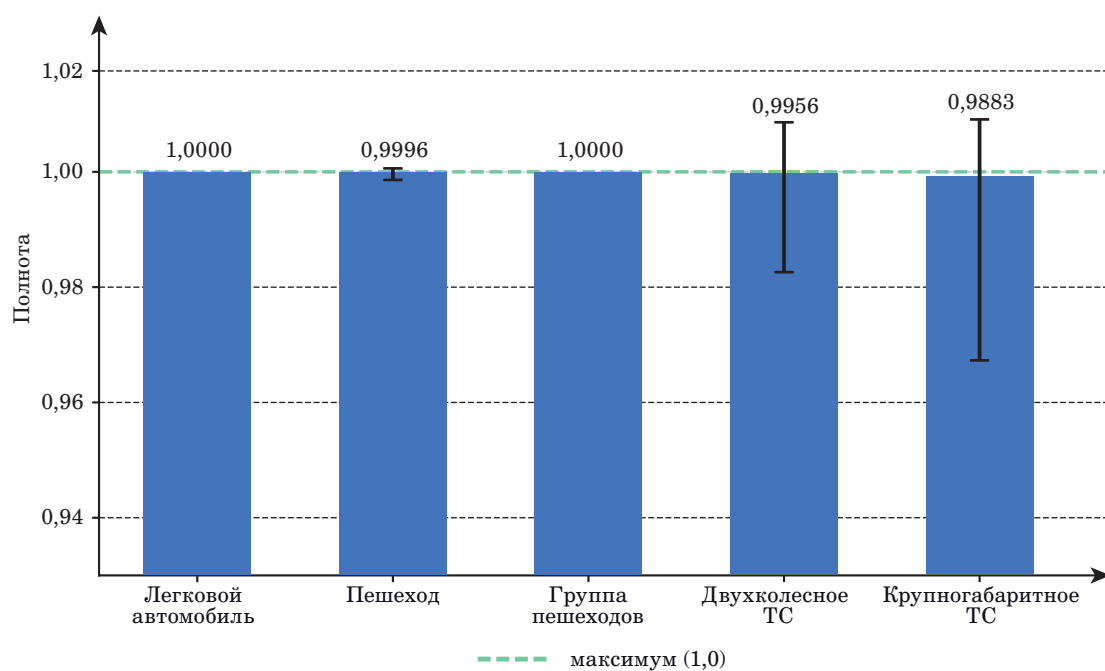
Поклассовые значения полноты представлены на рис. 2. Наиболее стабильные результаты показали классы *car* (легковой автомобиль) и *pedgroup* (группа пешеходов) ($\text{recall} = 1,000 \pm 0,000$ в обоих случаях). Класс *pedestrian* (пешеход) также демонстрирует высокую точность классификации: $\text{recall} = 0,9996 \pm 0,001$. Наибольший разброс наблюдается у классов *largevehicle* (крупногабаритное ТС) ($\text{recall} = 0,9883 \pm 0,021$) и *twowheeler* (двухколесное ТС) ($\text{recall} = 0,9956 \pm 0,013$), что соответствует их меньшей представленности в обучающей выборке.

Высокие показатели при низкой вариативности свидетельствуют о насыщении задачи при 30 эпохах обучения. Значение этого явления для интерпретации абляционного исследования подробно рассматривается ниже в подразделе «Абляционное исследование».

Сравнение с базовым методом на наборе данных Boreas-Objects-V1

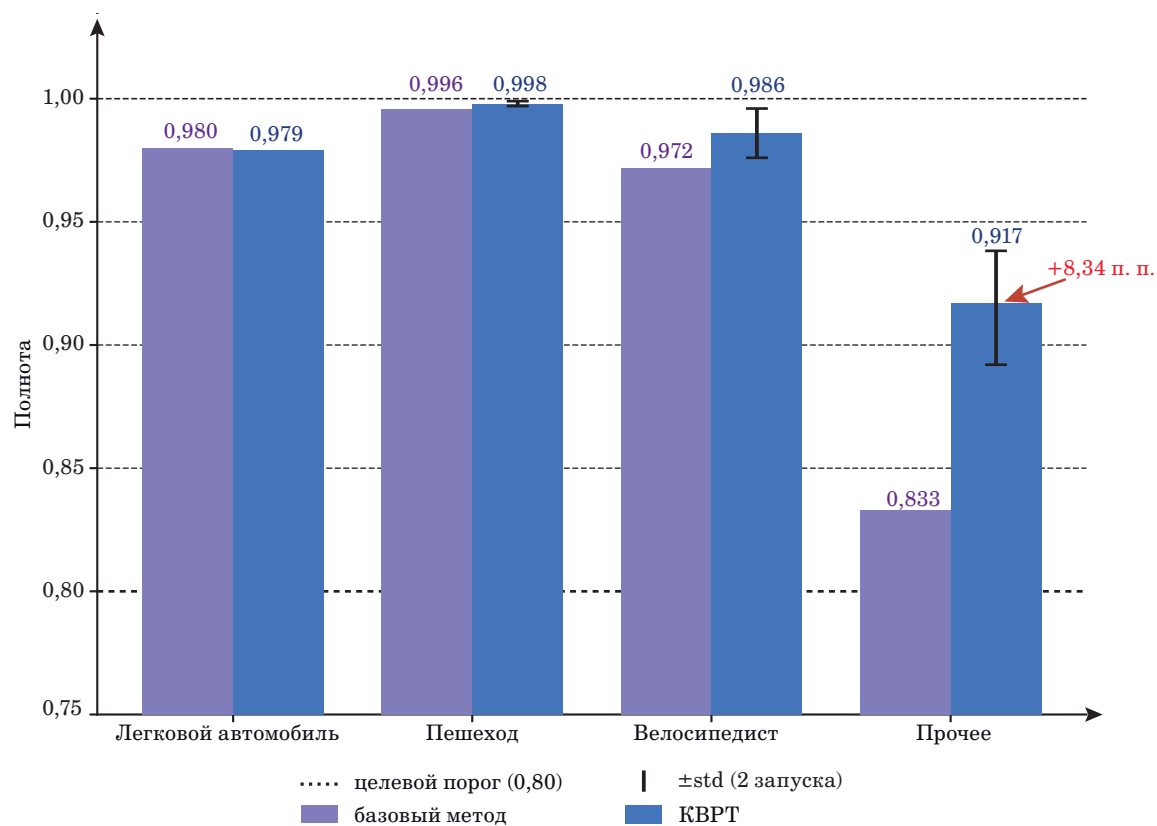
Набор Boreas-Objects-V1 содержит 4699 треков, распределенных по четырем классам, при этом класс прочее представлен лишь 70 треками (1,5 % выборки). При стратифицированном разбиении 70/30 по сценам тестовая часть включает не более 20–22 треков этого класса, что делает данный набор подходящим для оценки вклада механизмов адаптации к дисбалансу классов. Каждый из двух методов — предложенный (KBPT) и базовый — запускался дважды; для KBPT приведены среднее значение и СКО по двум запускам.

Классификатор временных радарных треков превосходит базовый метод по всем ключевым метрикам. Сбалансированная точность составила $0,970 \pm 0,003$ против 0,945 у базового метода (прирост 0,025), макро-F1 — $0,938 \pm 0,010$ против 0,927 (прирост 0,011 п. п.). Наиболее выраженный эффект наблюдается для редчайшего класса *misc*: полнота возросла с 0,833 до $0,917 \pm 0,025$ (прирост 0,083 п. п.). При этом адаптация к дисбалансу не привела к деградации на доминирующих классах: полнота *car* снизилась лишь на 0,001, тогда как *pedestrian* и *cyclist* показали прирост на 0,002 и 0,014 п. п. соответственно. Целевой порог полноты не ниже 0,80 выполнен обоими методами для всех классов.



■ **Рис. 2.** Поклассовые значения полноты модели КВРТ на наборе данных RadarScenes

■ **Fig. 2.** Per-class recall of the proposed method on RadarScenes



■ **Рис. 3.** Поклассовые значения полноты для КВРТ и базового метода на наборе данных Boreas-Objects-V1

■ **Fig. 3.** Per-class recall for the temporal radar track classifier and the baseline on Boreas-Objects-V1

Поклассовые значения полноты с планками погрешностей для КВРТ приведены на рис. 3.

Результат по классу *misc* (70 тестовых треков) свидетельствует о том, что базовый метод без балансировки при столь малом числе образцов стабильно недообучается на миноритарном классе, КВРТ же способно выделять паттерны данных в этих условиях.

Абляционное исследование

Для изолированной оценки вклада каждого компонента КВРТ было проведено абляционное исследование: восемь конфигураций на RadarScenes и пять конфигураций на Boreas-Objects-V1, по 30 эпох обучения, одна инициализация, одно разбиение на каждый набор данных. Конфигурации конструировались путем поочередного удаления или замены отдельных компонентов при прочих равных условиях. Исследование отвечает на три исследовательских вопроса, поставленных во введении.

RadarScenes. На наборе данных RadarScenes при 30 эпохах обучения задача классификации достигает насыщения: большинство конфигураций показывают $BA = 1,000$, т. е. неотличимы от полной модели. Это явление принято называть потолочным эффектом. Его причина состоит в том, что при достаточном числе эпох любая разумная конфигурация с корректной агрегацией по временной оси успешно решает данную задачу. В силу этого вклад отдельных компонентов: прототипной классификационной головы КВРТ, функции LDAM, классово-условного масштаби-

рования температуры — не поддается изолированному измерению при 30 эпохах. Все результаты по RadarScenes сведены в табл. 1.

Приведенные данные позволяют сформулировать два вывода с различным уровнем надежности. Первый вывод, подкрепленный наиболее убедительными данными: агрегация по временной оси является структурно необходимой. При замене временного трансформера покадровой классификацией сбалансированная точность снижается с 0,998 до 0,327 ($\Delta = -0,671$). Класс *pedgroup* полностью исчезает из предсказаний: все 223 трека этого класса отнесены к *car* или *largevehicle*, что объясняется недостаточностью одной-трех точек одиночного радарного кадра для разграничения близких классов без учета временного контекста. При этом замена трансформерной агрегации простым усреднением по кадрам не приводит к деградации ($BA = 1,000$): принципиальную роль играет сам факт агрегации по временной оси, а не конкретная ее форма.

Измеримый вклад при 30 эпохах установлен для двух компонентов: вспомогательной головы микроплеровских моментов и взвешивания функции потерь по классовым весам. Оба при исключении снижают сбалансированную точность на 0,007 (до 0,993) сугубо за счет ухудшения полноты класса *twowheeler* (с 1,000 до 0,965, т. е. $-3,5$ п. п.). Совпадение характера деградации в обоих случаях указывает на то, что данный редкий класс ($\approx 5\%$ обучающей выборки) чувствителен к ослаблению градиентного сигнала, а не к конкретному механизму его усиления.

■ **Таблица 1.** Абляционное исследование (RadarScenes)
■ **Table 1.** Component exclusion study (RadarScenes)

Конфигурация	Исключаемый компонент	BA	Полнота		Примечание
			twowheeler	pedgroup	
Полная модель (КВРТ)	—	0,998	1,000	1,000	Контрольная конфигурация
Усреднение вместо трансформера	CLS-трансформер → простое усреднение	1,000	1,000	1,000	Потолочный эффект
Без прототипной головы	Голова КВРТ → линейная голова	1,000	1,000	1,000	То же
Без отступа LDAM	Аддитивный отступ m_k	1,000	1,000	1,000	— “ —
Без температурного масштабирования	Классово-условная температура τ_k	1,000	1,000	1,000	— “ —
Без вспомогательной головы	Вспомогательная голова микро-Доплера	0,993	0,965	1,000	$-0,5$ п. п. по BA
Без взвешенной функции потерь	Классово-условные веса функции потерь	0,993	0,965	1,000	$-0,5$ п. п. по BA
Покадровая классификация	Вся агрегация по временной оси	0,327	0,070	0,000	Критическое снижение

■ **Таблица 2.** Абляционное исследование (Boreas-Objects-V1)

■ **Table 2.** Component exclusion study (Boreas-Objects-V1)

Конфигурация	Исключаемый компонент	BA	Макро-F1	Полнота: cyclist
Полная модель (КВРТ)	—	0,973	0,922	0,976
Без выравнивающей функции потерь	Функция потерь на выравнивание вложений	0,973	0,922	0,976
Без проекционного адаптера входа	Адаптер размерности признакового пространства	0,973	0,922	0,976
Без временной агрегации	Трансформерная агрегация по кадрам	0,973	0,918	0,976
Без прототипной головы (межнаборная)	Голова КВРТ (межнаборная конфигурация)	0,978	0,918	1,000

Boreas-Objects-V1. На наборе данных Boreas-Objects-V1 абляционное исследование охватывало шесть конфигураций (одно случайное состояние). Ни одна из проверенных конфигураций не показала статистически значимого отклонения от полной модели: все значения сбалансированной точности находятся в диапазоне 0,972–0,978, что не превышает одного стандартного отклонения полной модели по двум запускам. Результаты приведены в табл. 2.

Отдельного пояснения требует конфигурация «без выравнивающей функции потерь»: ее исключение не изменяет ни одну из метрик. Это объясняется тем, что функция выравнивания вложений разработана для согласования признаковых пространств двух различных наборов данных при переносе модели; при обучении на одном наборе она не выполняет никакой значимой регуляризирующей функции. Конфигурация «без прототипной головы» соответствует замене прототипной головы линейным классификатором при использовании проекционного адаптера входа; показатель BA при этом не снижается, а полнота класса cyclist достигает 1,000 при незначительном снижении макро-F1.

Обсуждение

Эксперименты показали, что архитектура действительно работает там, где мы ожидали — на уровне трека, а не отдельного кадра. Покадровый блок устроен по принципу PointNet: детекции обрабатываются общим преобразованием и объединяются симметричной операцией, что дает инвариантность к порядку строк. Но, в отличие от классической задачи с плотными облаками точек,

в одном радарном кадре их буквально 1–3 штуки, и именно поэтому основная различающая сила перемещается на уровень временной последовательности кадров.

Наборы данных RadarScenes и Boreas-Objects-V1 охватывают основные категории дорожных объектов: car, pedestrian, twowheeler и largevehicle. Однако они не отражают в полной мере реальный состав участников дорожного движения; ряд практически значимых категорий, таких как мотоциклисты (объединенные с велосипедистами в один класс) и дорожные рабочие, в обоих датасетах отдельно не представлены. Расширение классового состава при появлении соответствующих размеченных данных не требует переобучения модели целиком: прототипная классификационная голова допускает введение новых классов путем добавления прототипных векторов, что рассматривается как направление дальнейших исследований.

На наборе RadarScenes по трем разбиениям и трем случайным состояниям (девять запусков) была получена сбалансированная точность $0,997 \pm 0,005$ и макро-F1 $0,996 \pm 0,006$, при этом целевой уровень полноты не ниже 0,80 был достигнут для всех пяти классов в каждом запуске.

Непосредственное сопоставление результатов с опубликованными результатами по другим методам затруднено различием входных представлений и использованных наборов данных. Метод RadarGNN [13], также примененный на RadarScenes, работает с отдельными кадрами облака точек в задачах сегментации и обнаружения объектов и достигает макроусредненной F1-меры 77,1 % при сегментации; предложенный же метод решает иную по постановке задачу — классификацию треков с явным учетом дисбаланса

классов, что не позволяет сравнивать эти числа напрямую. Метод RadarTCN [17] использует двумерные спектральные карты дальности и скорости в качестве входного представления и оценивается на наборе данных CARRADA, содержащем принципиально иной тип радарного сигнала по сравнению с разреженным облаком точек, составляющим основу RadarScenes и Boreas-Objects-V1; количественное сопоставление с RadarTCN в данных условиях методически некорректно. Таким образом, предложенный подход занимает самостоятельную нишу: в отличие от RadarGNN он ориентирован на временную последовательность кадров и адаптацию к дисбалансу, а в отличие от RadarTCN — оперирует разреженными облаками точек без предварительного формирования двумерных проекций.

Наиболее четкая разница проявилась на Boreas-Objects-V1, где класс прочее — всего 70 треков из 4699, т. е. 1,5 %. Именно здесь адаптация к дисбалансу показала себя: сбалансированная точность выросла с 0,945 до 0,970, а полнота на редчайшем классе — с 0,833 до 0,917 (+8,34 п. п.) при том, что на доминирующих классах деградации нет. Улучшение оказалось адресным — только там, где оно было нужно.

Абляционное исследование подтвердило структурную роль временной агрегации: при ее отключении и переходе к классификации на уровне отдельных кадров сбалансированная точность на RadarScenes снижается с 0,998 до 0,327, а полнота класса *pedgroup* обращается в ноль. Это согласуется с известной ролью CLS-агрегации в трансформерных архитектурах, но впервые количественно показано для классификации радарных треков. На RadarScenes часть упрощенных конфигураций также достигает BA = 1,000 за 30 эпох, что отражает эффект насыщения метрики и ограничивает возможность разделить вклад отдельных компонентов метрической головы на этом наборе данных.

Ограничения работы связаны с объемом валидации и областью применимости. На Boreas-Objects-V1 использовано одно разбиение и два случайных состояния, что не позволяет строго оценить статистическую значимость различий. Не проверена работа метода на треках, формируемых реальным детектором и трекером, а не на заранее размеченных треках наборов данных. Дальнейшие исследования включают воспроизведение экспериментов на Boreas-Objects-V1 в режиме «3 разбиения × 3 случайных состояния», сравнение с базовой моделью при одинаковой размерности трекового вектора, анализ

чувствительности по параметрам τ_0 , m_0 и β и унификацию постановки задачи для прямого сопоставления с методами семантической сегментации радарных данных, такими как RadarGNN.

Заключение

Разработан классификатор временных радарных треков с многомасштабной адаптацией к дисбалансу классов — KBPT. Архитектура включает три блока: инвариантный к перестановке кодировщик детекций, временной трансформер с агрегирующим служебным токеном и классификационную голову на основе сравнения с обучаемыми эталонами классов. Параметры классово-условной температуры и аддитивного отступа выводятся аналитически из статистики обучающей выборки — подбирать гиперпараметры отдельно под каждый класс не нужно. Данные разбиваются на уровне сцен записи, а программная проверка исключает пересечения между подвыборками.

Проверка на двух открытых наборах данных дала устойчивые результаты в обоих режимах. На RadarScenes — сбалансированная точность $0,997 \pm 0,005$, полнота не ниже 0,80 для каждого из пяти классов в каждом из девяти запусков. На Boreas-Objects-V1 при выраженном дисбалансе — прирост полноты редчайшего класса +8,340 п. п. без потерь на доминирующих категориях. Абляционное исследование зафиксировало главное: без временной агрегации задача классификации треков не решается — точность на RadarScenes падает с 0,998 до 0,327.

К направлениям дальнейших исследований относятся: статистическая валидация на Boreas-Objects-V1 в схеме «3 разбиения × 3 случайных состояния»; проверка метода на треках, формируемых реальным детектором; анализ чувствительности по ключевым гиперпараметрам классификационной головы — базовой температуре косинусных логитов τ_0 , базовой величине аддитивного отступа LDAM m_0 и параметру эффективного числа образцов β , управляющему силой классово-условного перевзвешивания; унификация постановки задачи для прямого сопоставления с методами поточечной семантической сегментации радарных данных.

Финансовая поддержка

Исследование выполнено при поддержке Российского научного фонда (соглашение № 25-29-00868).

Литература

1. **Bijelic M., Gruber T., Mannan F., Kraus F., Ritter W., Dietmayer K., Heide F.** Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11679–11689. doi:10.1109/CVPR42600.2020.01170, EDN: PADCFG
2. **Srivastav A., Mandal S.** Radars for autonomous driving: A review of deep learning methods and challenges. *IEEE Access*, 2023, vol. 11, pp. 97147–97168. doi:10.1109/ACCESS.2023.3312382, EDN: RQXILU
3. **Соколик Н. В.** Определение скорости движения и дальности быстро движущихся объектов в РЛС с непрерывным линейно-частотно-модулированным излучением с использованием автокорреляционной схемы. *Известия высших учебных заведений России. Радиоэлектроника*, 2020, т. 23, № 2, с. 63–72. doi:10.32603/1993-8985-2020-23-2-63-72, EDN: JDERCM
4. **Сунь Х., Чжуан С., Костров А. А.** Обзор современных систем технического зрения, применяемых в транспортной отрасли. *Современные наукоемкие технологии*, 2024, № 9, с. 69–73. doi:10.17513/snt.40150, EDN: BPHTGT
5. **Zhou Y., Liu L., Zhao H., López-Benítez M., Yu L., Yue Y.** Towards deep radar perception for autonomous driving: Datasets, methods, and challenges. *Sensors*, 2022, vol. 22, no. 11, Art. 4208. doi:10.3390/s22114208, EDN: IUZPZK
6. **Артюхин И. В., Аверин И. М., Флакман А. Г., Рубцов А. Е.** Алгоритм оценки углов прихода сигналов в системе распределенных некогерентных автомобильных радаров. *Журнал радиоэлектроники [электронный журнал]*, 2023, № 4. doi:10.30898/1684-1719.2023.4.2, EDN: XSEZVF. http://jre.cplire.ru/jre/about_e.html (дата обращения: 02.05.2025).
7. **Coppola R., Ahmed S., Alouini M. S.** Road users classification based on bi-frame micro-Doppler with 24-GHz FMCW radar. *Frontiers in Signal Processing*, 2022, vol. 2, Art. 864538. doi:10.3389/frsip.2022.864538, EDN: AUYPYP
8. **Басов О. О., Толстой И. М., Ле Ань Ту.** Классификация движущихся наземных объектов по их доплеровским портретам с помощью методов машинного обучения. *Современные наукоемкие технологии*, 2021, № 10, с. 17–22. doi:10.17513/snt.38848, EDN: WPJJSE
9. **Лук Д. В., Коновалов А. А., Хоанг Л. М.** Алгоритм распознавания малоразмерных воздушных целей по траекторным признакам в полуактивной РЛС. *Известия высших учебных заведений России. Радиоэлектроника*, 2023, т. 26, № 5, с. 76–88. doi:10.32603/1993-8985-2023-26-5-76-88, EDN: CZDSXO
10. **Qi C. R., Su H., Mo K., Guibas L. J.** PointNet: Deep learning on point sets for 3D classification and segmentation. *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 652–660. doi:10.1109/CVPR.2017.16
11. **Qi C. R., Yi L., Su H., Guibas L. J.** PointNet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, vol. 30, pp. 5099–5108. arXiv:1706.02413
12. **Шпрехер Д. М., Минаков Е. И., Грачев А. Н., Башеров М. С.** Классификация воздушных целей в РЛС на основе нейросетевой модели с расширенным признаковым пространством. *Программные продукты и системы*, 2025, т. 38, № 4, с. 682–693. doi:10.15827/0236-235X.152.682-693, EDN: ZNAJAZ
13. **Fent F., Bauerschmidt P., Lienkamp M.** RadarGNN: Transformation invariant graph neural network for radar-based perception. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023, pp. 182–191. doi:10.1109/CVPRW59228.2023.00023
14. **Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I.** Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, vol. 30, pp. 5998–6008. arXiv:1706.03762
15. **Devlin J., Chang M.-W., Lee K., Toutanova K.** BERT: Pre-training of deep bidirectional transformers for language understanding. *Proc. 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 2019, pp. 4171–4186. doi:10.18653/v1/N19-1423
16. **Люткин Д. А., Поздняков Д. В., Соловьев А. А., Жуков Д. В., Малик М. Ш. И., Игнатов Д. И.** Применение трансформеров для определения профиля врача на основе запросов пользователей. *Автоматика и телемеханика*, 2024, № 3, с. 86–100. doi:10.31857/S0005231024030076, EDN: TQAELK
17. **Li Y., Zhang M., Jing H., Liu Z.** RadarTCN: Light-weight online classification network for automotive radar targets based on TCN. *Sensors*, 2024, vol. 24, no. 9, Art. 2813. doi:10.3390/s24092813, EDN: UXADLX
18. **Schumann O., Hahn M., Scheiner N., Weishaupt F., Tilly J. F., Dickmann J., Wöhler C.** RadarScenes: A real-world radar point cloud data set for automotive applications. *Proc. 24th International Conference on Information Fusion (FUSION)*, 2021, pp. 1–8. doi:10.23919/FUSION49465.2021.9627037
19. **Константинов А. Ф., Дьяконова Л. П.** Сравнительный анализ методов снижения дисбаланса классов при построении моделей машинного обучения в финансовом секторе. *Известия Кабардино-Балкарского научного центра РАН*, 2025, т. 27, № 1, с. 143–151. doi:10.35330/1991-6639-2025-27-1-143-151, EDN: XRYMDH
20. **Cui Y., Jia M., Lin T.-Y., Song Y., Belongie S.** Class-balanced loss based on effective number of samples. *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 9260–9269. doi:10.1109/CVPR.2019.00949

21. Cao K., Wei C., Gaidon A., Arechiga N., Ma T. Learning imbalanced datasets with label-distribution-aware margin loss. *Advances in Neural Information Processing Systems (NeurIPS)*, 2019, vol. 32, pp. 1567–1578. arXiv:1906.07413
22. Zhang Y., Kang B., Hooi B., Yan S., Feng J. Deep long-tailed learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, vol. 45, no. 9, pp. 10795–10816. doi:10.1109/TPAMI.2023.3268118, EDN: LSGVDA
23. Deng J., Guo J., Yang J., Xue N., Liu Z., Gong Y., Zafeiriou S. ArcFace: Additive angular margin loss for deep face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, vol. 44, no. 10, pp. 5962–5979. doi:10.1109/TPAMI.2021.3087709, EDN: XKIJAV
24. Snell J., Swersky K., Zemel R. Prototypical networks for few-shot learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, vol. 30, pp. 4077–4087. arXiv:1703.05175
25. Burnett K., Yoon D. J., Wu Y., Li A. Z., Zhang H., Lu S., Qian J., Tseng W.-K., Lambert A., Leung K. Y. K., Schoellig A. P., Barfoot T. D. Boreas: A multi-season autonomous driving dataset. *The International Journal of Robotics Research*, 2023, vol. 42, no. 1–2, pp. 33–42. doi:10.1177/02783649231160195, EDN: HIRFYI

UDC 004.93'1:621.396.96

doi:10.31799/1684-8853-2026-4-60-74

EDN: G GK F J Q

Temporal track classifier for automotive radar with multi-scale adaptation to class imbalance

E. A. Lopukhova^a, Junior Researcher, orcid.org/0000-0001-6802-6010

G. S. Voronkov^a, PhD, Tech., Associate Professor, orcid.org/0000-0002-8788-2696, voronkov.gs@ugatu.su

E. P. Topolskaya^a, PhD, Tech., Associate Professor, orcid.org/0000-0003-2207-2702

^aUfa University of Science and Technology, 32, Z. Validi St., 450076, Ufa, Russian Federation

Introduction: The recognition of road objects from automotive radar data is complicated by the small number of reflections in individual frames and by class imbalance, which leaves rare objects underrepresented during training. **Purpose:** To develop a radar-track classifier which is robust to class imbalance and applicable to open road-traffic radar datasets. **Methods:** We have developed and applied a radar-track classification method that combines frame-wise processing of unordered sets of radar reflections by a shared multilayer perceptron with invariant feature aggregation, temporal transformer-based aggregation with a learnable classification token, and a prototype-based classification head in a normalized cosine space. Class imbalance is considered through the use of a margin dependent on the label distribution, class-dependent temperature scaling, loss-function weights calculated from the effective number of samples, and weighted sampling during mini-batch formation; auxiliary prediction of statistical moments of radial velocity has been additionally applied. **Results:** The experiments have used RadarScenes and Boreas-Objects-V1. RadarScenes contains 158 sequences with point-wise labels for five classes, whereas Boreas-Objects-V1 contains 4,699 tracks with three-dimensional labels for four classes, including a rarest class with a 1.5% share. On Boreas-Objects-V1, balanced accuracy has increased from 0.945 to 0.970, and the recall of the rarest class has increased from 0.833 to 0.917. The ablation study has shown that removing temporal aggregation reduces balanced accuracy on RadarScenes from 0.998 to 0.327, which confirms the structural necessity of sequential information. **Practical relevance:** The classifier can be applied for training radar-based road-object recognition systems under limited rare-class examples.

Keywords – automotive radar, radar track, road-object recognition, class imbalance, temporal aggregation, metric learning, neural networks.

For citation: Lopukhova E. A., Voronkov G. S., Topolskaya E. P. Temporal track classifier for automotive radar with multi-scale adaptation to class imbalance. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2026, no. 4, pp. 60–74 (In Russian). doi:10.31799/1684-8853-2026-4-60-74, EDN: G GK F J Q

Financial support

The research is supported by Russian Science Foundation (agreement No. 25-29-00868).

References

1. Bijelic M., Gruber T., Mannan F., Kraus F., Ritter W., Dietmayer K., Heide F. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11679–11689. doi:10.1109/CVPR42600.2020.01170, EDN: PADCFG
2. Srivastav A., Mandal S. Radars for autonomous driving: A review of deep learning methods and challenges. *IEEE Access*, 2023, vol. 11, pp. 97147–97168. doi:10.1109/ACCESS.2023.3312382, EDN: RQXILU
3. Sokolik N. V. Velocity and range measurement of fast-moving objects in a continuous linear frequency-modulated radar using an autocorrelation scheme. *Journal of the Russian Universities. Radioelectronics*, 2020, vol. 23, no. 2, pp. 63–72 (In Russian). doi:10.32603/1993-8985-2020-23-2-63-72, EDN: JDERCM
4. Sun H., Zhuang S., Kostrov A. A. An overview of modern vision systems used in the transport industry. *Modern High Technologies*, 2024, no. 9, pp. 69–73 (In Russian). doi:10.17513/snt.40150, EDN: BPHTGT
5. Zhou Y., Liu L., Zhao H., López-Benítez M., Yu L., Yue Y. Towards deep radar perception for autonomous driving: Datasets, methods, and challenges. *Sensors*, 2022, vol. 22, no. 11, Art. 4208. doi:10.3390/s22114208, EDN: IUZPZK
6. Artyukhin I. V., Averin I. M., Flaksman A. G., Rubtsov A. E. Direction-of-Arrival Estimation Algorithm for a Distributed Non-Coherent Automotive Radar System. *Journal of Radio Electronics* [online journal], 2023, no. 4. doi:10.30898/1684-1719.2023.4.2, EDN: XSEZVF. Available at: http://jre.cplire.ru/jre/about_e.html (accessed 2 May 2025) (In Russian).
7. Coppola R., Ahmed S., Alouini M. S. Road users classification based on bi-frame micro-Doppler with 24-GHz FMCW radar. *Frontiers in Signal Processing*, 2022, vol. 2, Art. 864538. doi:10.3389/frsip.2022.864538, EDN: AUYPYP
8. Basov O. O., Tolstoy I. M., Le Anh Tu. Classification of moving ground objects by their doppler signatures using machine learning methods. *Modern High Technologies*, 2021, no. 10, pp. 17–22 (In Russian). doi:10.17513/snt.38848, EDN: WPJJSE
9. Luk D. V., Konovalov A. A., Hoang L. M. Algorithm for recognition of small-sized aerial targets by trajectory features

- in a semi-active radar. *Journal of the Russian Universities. Radioelectronics*, 2023, vol. 26, no. 5, pp. 76–88 (In Russian). doi:10.32603/1993-8985-2023-26-5-76-88, EDN: CZDSXO
10. Qi C. R., Su H., Mo K., Guibas L. J. PointNet: Deep learning on point sets for 3D classification and segmentation. *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 652–660. doi:10.1109/CVPR.2017.16
 11. Qi C. R., Yi L., Su H., Guibas L. J. PointNet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, vol. 30, pp. 5099–5108. arXiv:1706.02413
 12. Shprekher D. M., Minakov E. I., Grachev A. N., Basherov M. S. Radar target classification based on a neural network model with an extended feature space. *Software & Systems*, 2025, vol. 38, no. 4, pp. 682–693 (In Russian). doi:10.15827/0236-235X.152.682-693, EDN: ZNAJAZ
 13. Fent F., Bauerschmidt P., Lienkamp M. RadarGNN: Transformation invariant graph neural network for radar-based perception. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023, pp. 182–191. doi:10.1109/CVPRW59228.2023.00023
 14. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I. Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, vol. 30, pp. 5998–6008. arXiv:1706.03762
 15. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proc. 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 2019, pp. 4171–4186. doi:10.18653/v1/N19-1423
 16. Lyutkin D. A., Pozdnyakov D. V., Solovov A. A., Zhukov D. V., Malik M. Sh. I., Ignatov D. I. Application of transformers to specialist physician identification based on user queries. *Automation and Remote Control*, 2024, no. 3, pp. 86–100 (In Russian). doi:10.31857/S0005231024030076, EDN: TQAEK
 17. Li Y., Zhang M., Jing H., Liu Z. RadarTCN: Lightweight online classification network for automotive radar targets based on TCN. *Sensors*, 2024, vol. 24, no. 9, Art. 2813. doi:10.3390/s24092813, EDN: UXADLX
 18. Schumann O., Hahn M., Scheiner N., Weishaupt F., Tilly J. F., Dickmann J., Wöhler C. RadarScenes: A real-world radar point cloud data set for automotive applications. *Proc. 24th International Conference on Information Fusion (FUSION)*, 2021, pp. 1–8. doi:10.23919/FUSION49465.2021.9627037
 19. Konstantinov A. F., Dyakonova L. P. Comparative analysis of class imbalance reduction methods for machine learning models in the financial sector. *News of the Kabardino-Balkarian Scientific Center of RAS*, 2025, vol. 27, no. 1, pp. 143–151 (In Russian). doi:10.35330/1991-6639-2025-27-1-143-151, EDN: XRYMDH
 20. Cui Y., Jia M., Lin T.-Y., Song Y., Belongie S. Class-balanced loss based on effective number of samples. *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 9260–9269. doi:10.1109/CVPR.2019.00949
 21. Cao K., Wei C., Gaidon A., Arechiga N., Ma T. Learning imbalanced datasets with label-distribution-aware margin loss. *Advances in Neural Information Processing Systems (NeurIPS)*, 2019, vol. 32, pp. 1567–1578. arXiv:1906.07413
 22. Zhang Y., Kang B., Hooi B., Yan S., Feng J. Deep long-tailed learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, vol. 45, no. 9, pp. 10795–10816. doi:10.1109/TPAMI.2023.3268118, EDN: LSGVDA
 23. Deng J., Guo J., Yang J., Xue N., Liu Z., Gong Y., Zafeiriou S. ArcFace: Additive angular margin loss for deep face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, vol. 44, no. 10, pp. 5962–5979. doi:10.1109/TPAMI.2021.3087709, EDN: XKIJAV
 24. Snell J., Swersky K., Zemel R. Prototypical networks for few-shot learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, vol. 30, pp. 4077–4087. arXiv:1703.05175
 25. Burnett K., Yoon D. J., Wu Y., Li A. Z., Zhang H., Lu S., Qian J., Tseng W.-K., Lambert A., Leung K. Y. K., Schoellig A. P., Barfoot T. D. Boreas: A multi-season autonomous driving dataset. *The International Journal of Robotics Research*, 2023, vol. 42, no. 1–2, pp. 33–42. doi:10.1177/02783649231160195, EDN: HIRFYI

ПАМЯТКА ДЛЯ АВТОРОВ

Поступающие в редакцию статьи проходят обязательное рецензирование.

При наличии положительной рецензии статья рассматривается редакционной коллегией. Принятая в печать статья направляется автору для согласования редакторских правок. После согласования автор представляет в редакцию окончательный вариант текста статьи.

Процедуры согласования текста статьи могут осуществляться как непосредственно в редакции, так и по e-mail (ius.spb@gmail.com).

При отклонении статьи редакция представляет автору мотивированное заключение и рецензию, при необходимости доработать статью — рецензию.

Редакция журнала напоминает, что ответственность за достоверность и точность рекламных материалов несут рекламодатели.