На правах рукописи



Соколов Василий Александрович

Методы активного обучения детекторов объектов в задачах видеоаналитики на основе масштабируемых представлений

Специальность 1.2.1. –
Искусственный интеллект и машинное обучение

АВТОРЕФЕРАТ
диссертации на соискание учёной степени
кандидата технических наук

Москва – 2026

Работа выполнена в Федеральном государственном автономном образовательном учреждении высшего образования «Московский физико-технический институт (национальный исследовательский университет)».

Научный руководитель:
доктор технических наук
Матвеев Иван Алексеевич

Официальные оппоненты:
Орлов Алексей Александрович,
доктор технических наук, доцент,
Муромский институт (филиал) Владимирского государственного университета имени Александра Григорьевича и Николая Григорьевича Столетовых, заведующий кафедрой физики и прикладной математики

Ярушев Сергей Александрович,
кандидат технических наук,
Российский экономический университет имени Г. В. Плеханова, директор Центра перспективных исследований в искусственном интеллекте

Ведущая организация:
Федеральное государственное бюджетное образовательное учреждение высшего образования «Тульский государственный университет»

Защита состоится «___» _____ 2026 г. в _____ на заседании диссертационного совета 24.1.224.03 при Федеральном исследовательском центре «Информатика и управление» Российской академии наук по адресу: 119333, г. Москва, ул. Вавилова, д. 42.

С диссертацией можно ознакомиться в библиотеке Федерального исследовательского центра «Информатика и управление» Российской академии наук и на сайте https://www.frccsc.ru.

Автореферат разослан «___» _____ 2026 г.

Учёный секретарь
диссертационного совета 24.1.224.03,
кандидат технических наук И. А. Рейер

Общая характеристика работы

Актуальность темы исследования. Нейросетевые детекторы объектов применяются в промышленном контроле, мониторинге транспортной и городской инфраструктуры. Для их обучения требуется разметка классов и ограничивающих прямоугольников объектов; при переносе системы в новый домен её стоимость может стать главным ограничением.

Активное обучение позволяет итеративно выбирать полезные для модели изображения при ограниченном бюджете разметки. В задаче детекции необходимо учитывать новизну отдельных объектов, их масштаб и внутриклассовую вариативность, которые не описываются глобальным сходством кадров.

Локальные эмбеддинги позволяют описывать объекты по многомасштабным картам признаков детектора и оценивать их новизну относительно размеченного множества. Однако большие векторы требуют значительных затрат на хранение и сравнение, а короткие могут терять сведения о визуальном содержании объекта. Поэтому актуальна разработка методов отбора на основе локальных представлений изменяемой размерности с сохранением их информативности.

Степень разработанности темы. Критерии активного обучения объединяют методы по неопределённости, разнообразию данных и гибридные методы. Отбор по неопределённости для обучения классификаторов текстов исследовали Д. Льюис и У. Гейл, а отбор представительного подмножества данных в пространстве признаков нейронной сети – О. Сенер и С. Саварезе. Системный обзор семейств методов активного обучения проведён Б. Сеттлзом[1]. В задаче детекции Т. Юань и соавторы рассматривали неопределённость отдельных объектов, В. Юй и соавторы – согласованность предсказаний при преобразованиях изображений. Операцию RoI Align предложили К. Хэ и соавторы, основы обучения вложенных векторных представлений – А. Кусупати и соавторы, метод приближённого поиска HNSW – Ю. А. Малков и Д. А. Яшунин[2]. В настоящей работе совместно исследуются локальная новизна объектов, использование многомасштабных карт признаков и ускорение поиска за счёт вложенных представлений.

Целью диссертационной работы является исследование и разработка метода активного обучения, обеспечивающего быстрый и эффективный отбор данных для обучения нейросетевых детекторов с учётом признаков отдельных объектов, их внутриклассовой вариативности и различий в масштабе.

---

[1] *Lewis*, *D. D.* A Sequential Algorithm for Training Text Classifiers / D. D. Lewis, W. A. Gale // Proc. 17th Annual Intern. ACM SIGIR Conf. Research and Development in Information Retrieval. Dublin, Ireland, 1994. P. 3—12; *Sener*, *O.* Active Learning for Convolutional Neural Networks: A Core-Set Approach / O. Sener, S. Savarese // Proc. 6th Intern. Conf. Learning Representations. Vancouver, Canada, 2018; *Settles*, *B.* Active Learning Literature Survey : tech. rep. / B. Settles ; University of Wisconsin–Madison. 2009. P. 1—67. No. 1648.

[2] *Malkov*, *Y. A.* Efficient and Robust Approximate Nearest Neighbor Search Using Hierarchical Navigable Small World Graphs / Y. A. Malkov, D. A. Yashunin // IEEE Trans. Pattern Analysis and Machine Intelligence. 2020. Vol. 42, no. 4. P. 824—836; Mask R-CNN / K. He [et al.] // Proc. IEEE Intern. Conf. Computer Vision. Venice, Italy, 2017. P. 2961—2969; Matryoshka Representation Learning / A. Kusupati [et al.] // Advances in Neural Information Processing Systems. Vol. 35. 2022. P. 30233—30249.

Для достижения поставленной цели решены следующие **задачи**:

1. Сформулировать требования к критериям отбора данных при активном обучении нейросетевых детекторов объектов с учётом специфики задач видеоаналитики и установить, насколько существующие подходы удовлетворяют этим требованиям.
2. На основе сформулированных требований разработать и реализовать метод отбора изображений по локальным представлениям содержащихся в них объектов. Исследовать различные метрические оценки полезности изображения. Экспериментально оценить влияние предложенного метода на качество детекции при фиксированном объёме разметки.
3. Разработать и реализовать способ формирования вложенных векторных представлений объектов, позволяющий использовать начальные части вектора как самостоятельные представления разной размерности с сохранением информации с различных пространственных масштабов. Верифицировать метод на выбранных наборах данных, оценить влияние предложенного способа на качество детекции, качество и время поиска похожих объектов.
4. Разработать и реализовать метод активного обучения детекторов объектов, использующий масштабируемые представления и последовательное уточнение результатов поиска при увеличении размерности сравниваемых векторов. Определить его преимущества и ограничения относительно метода на основе представлений фиксированной размерности и существующих методов активного обучения по качеству детекции и полному времени подготовки данных для разметки.

Объектом исследования являются нейросетевые методы детекции объектов и процедуры итеративного формирования обучающих множеств для таких моделей. Предметом исследования являются критерии активного обучения, основанные на локальных эмбеддингах объектов, а также архитектурные способы формирования вложенных представлений объектов в детекционной модели.

Научная новизна:

1. Предложен новый метод активного обучения детекторов объектов, в котором отбор изображений определяется за счёт расположения локальных представлений отдельных объектов относительно уже размеченного множества, а сами представления объединяют признаки из всех соответствующих объекту областей интереса на выходах модели, соответствующих разному пространственному разрешению.
2. Впервые предложен способ обучения детектора объектов, формирующего семейство двумерных вложенных согласованных представлений объектов, извлекаемых из карт активации. Предложенная структура эмбеддингов позволяет эффективно и быстро использовать их в поиске похожих изображений.
3. Предложен метод активного обучения на основе вложенных представлений объектов, сочетающий чередование признаков разных уровней

пирамиды и каскадный поиск ближайших соседей, что позволяет ускорить и повысить качество поиска кандидатов для последующей разметки.

Практическая значимость результатов работы состоит в расширении инструментария активного обучения для детекции объектов за счёт предложенных методов и повышении эффективности подготовки обучающих данных при ограниченном бюджете разметки. Разработанный критерий удовлетворяет требованиям, накладываемым спецификой обучения детекторов для видеоаналитики. Так, он совместим с наиболее быстрыми и точными архитектурами детекторов, т.е. детекторов, использующих карты активации на нескольких масштабах. Кроме того, он позволяет различать визуально разные группы объектов внутри одного и того же класса. Наконец, предлагаемый метод активного обучения опережает по скорости и качеству работы лучший из представленных на данный момент в литературе методов активного обучения детекторов. Результаты диссертационного исследования внедрены и используются в ООО «ВК ИТС» в специализированном программном обеспечении для видеомониторинга на основе глубоких нейросетевых детекторов. Внедрение подтверждено актом от 20 июля 2026 года.

Методология и методы исследования. Методы исследования включают в себя методы активного обучения, методы детекции объектов на основе глубоких нейронных сетей с помощью YOLOv8, методы извлечения локальных признаков с помощью RoI Align, методы метрического сравнения эмбеддингов, метод понижения размерности PCA (Principal Component Analysis), методы обучения вложенных представлений на основе Matryoshka Representation Learning, приближённый и каскадный поиск ближайших соседей, а также экспериментальную оценку качества детекции по мере качества mAP50–95 (mean Average Precision).

Основные положения, выносимые на защиту:

1. Метод активного обучения детекторов AREA, отличающийся способностью распознавать внутриклассовые визуальные различия объектов при отборе изображений по новизне отдельных объектов относительно размеченного множества за счёт объединения локальных эмбеддингов объектов, извлекаемых со всех уровней выходов модели без округления координат, за счет чего совместимый с наиболее распространенным, мультимасштабным классом детекторов.
2. Принцип модификации детекторов для формирования вложенных пространственных представлений объектов по схеме матрешки, основанный на совместном обучении выходного модуля детектора с общими весами на полных и сокращённых наборах каналов и чередовании уровней пирамиды признаков, что позволяет ускорить задачу нахождения похожих объектов за счет использования каскадного поиска и делает вложенные эмбеддинги пригодными для использования в различных приложениях.
3. Метод активного обучения детекторов MAREA, основанный на вложенных представлениях объектов из карт активации нейросети, каскадном

поиске и обеспечивающий преимущество перед AREA и стадией отбора по разнообразию PPAL (Plug and Play Active Learning) по совокупности качества детекции и скорости отбора данных в проведённых экспериментах.

Соответствие паспорту специальности. Содержание диссертации соответствует пунктам 4, 13, 16 и 17 паспорта научной специальности 1.2.1 «Искусственный интеллект и машинное обучение».

Достоверность результатов обеспечивается корректной постановкой задачи активного обучения, использованием общепринятых архитектур детекции объектов, сравнением с базовыми стратегиями отбора, экспериментальной проверкой на нескольких наборах данных, применением стандартных мер качества детекции и апробацией на практике. Статистическая значимость полученных результатов проверена с помощью многократных повторных запусков экспериментов, подсчёта средних значений и доверительного интервала.

Апробация работы. Основные результаты диссертационной работы докладывались и обсуждались на следующих конференциях: 15-й Международной конференции «Интеллектуализация обработки информации» (ИОИ-2024, Гродно, Беларусь, 2024); 67-й Всероссийской научной конференции Московского физико-технического института (Долгопрудный, Россия, 2025); 22-й Всероссийской конференции «Математические методы распознавания образов» (Муром, Россия, 22–26 сентября 2025 г.); 28-й Международной конференции по распознаванию образов (ICPR-2026) в рамках 10-го Международного семинара «Image Mining. Theory and Applications» (IMTA-X-2026, Лион, Франция, 21 августа 2026 г.).

Личный вклад. Все результаты, представленные в работах, включённых в диссертационное исследование, получены лично соискателем под научным руководством доктора технических наук Матвеева Ивана Алексеевича. В работах с соавторами личный вклад соискателя заключается в постановке задач, реализации методов, постановке основных численных экспериментов, анализе результатов и подготовке материалов к публикации.

Публикации. Результаты по теме диссертации опубликованы в 5 научных работах, включая 2 статьи в журналах из списка ВАК, индексируемых в Web of Science и Scopus [1; 2], и 3 публикации в сборниках тезисов докладов конференций [3—5]. Получены 3 свидетельства о государственной регистрации программ для ЭВМ [6—8].

Объём и структура работы. Диссертация состоит из введения, четырёх глав и заключения. Полный объём составляет 99 страниц, включая 28 рисунков и 13 таблиц; список литературы содержит 83 наименования.

Основное содержание работы

Во **введении** обоснована актуальность активного обучения детекторов объектов при ограниченном бюджете разметки, сформулированы цель и зада-

чи исследования, научная новизна и положения, выносимые на защиту. Работа последовательно решает задачи выбора локального представления объекта, формирования его вложенных вариантов и использования этих представлений для отбора изображений.

Первая глава посвящена анализу методов активного обучения и архитектур детекторов объектов в задачах видеоаналитики. Видеопоток содержит множество близких по содержанию кадров, однако редкие события и изменения условий съёмки образуют слабо представленные визуальные группы, для распознавания которых требуется дополнение обучающей выборки. При отборе таких кадров необходимо учитывать, какие объекты они добавляют к размеченному множеству и насколько эти объекты отличаются от уже известных детектору.

Пусть L_{t-1} – размеченное множество перед раундом t, U_{t-1} – неразмеченный пул, $f_{\theta_{t-1}}$ – обученный на L_{t-1} детектор. При бюджете b изображений требуется выбрать подмножество $S_t \subset U_{t-1}$ размера b и получить для него разметку. Множества обновляются:

$$\begin{aligned} L_t &= L_{t-1} \cup \{(x, y_x) : x \in S_t\}, \\ U_t &= U_{t-1} \setminus S_t, \end{aligned} \tag{1}$$

где y_x – разметка изображения x. После разметки выбранных изображений детектор обучается на расширенном множестве, и процедура повторяется, причём полезность очередного изображения требуется оценивать до получения его разметки.

Рассмотрены методы отбора по неопределённости предсказаний, разнообразию данных и сочетанию этих критериев, а также особенности их применения к изображениям, содержащим несколько объектов. Агрегация неопределённости и сходства по всему изображению не описывает визуальное разнообразие отдельных объектов, поэтому в работе обоснован переход к локальным эмбеддингам – векторам признаков отдельных объектов.

Сформулированы четыре требования к критерию отбора данных для промышленной видеоаналитики:

1. оценивать изображения через представления содержащихся в них объектов;
2. различать визуальные подгруппы внутри одного семантического класса;
3. использовать признаки нескольких пространственных масштабов;
4. оценивать новизну относительно уже размеченного множества.

В результате анализа 36 архитектур детекторов установлена необходимость работы с многомасштабными картами признаков: для мелких и крупных объектов значимы разные уровни пирамиды, поэтому выбор одного уровня ограничивает описание объекта. Рассмотренные критерии активного обучения не удовлетворяют всем четырём требованиям одновременно, что определяет задачу разработки метода AREA, представленную во второй главе.

Во второй главе предложен метод AREA (Active learning with Region Embedding Alignment), в котором изображения отбираются по новизне локальных эмбеддингов объектов относительно размеченного множества [4; 5; 2].

Для объекта o с ограничивающим прямоугольником $\mathbf{b}(o)$ его координаты проецируются на карты каждого уровня пирамиды. При обычном округлении координат часть области может смещаться или исчезать, особенно на картах низкого пространственного разрешения. В AREA применяется RoI Align: значения признаков в дробных координатах вычисляются билинейной интерполяцией, после чего область приводится к фиксированному размеру $r_h \times r_w$. Операция выполняется независимо на каждом уровне пирамиды, после чего полученные тензоры разворачиваются и объединяются в единый вектор, содержащий признаки разных пространственных масштабов:

$$\mathbf{e}(o) = \big[\mathbf{e}^{(1)}(o) \| \cdots \| \mathbf{e}^{(L)}(o)\big], \tag{2}$$

$$\mathbf{e}^{(l)}(o) = \mathrm{vec}\big(\mathrm{RoIAlign}(F_l, \mathbf{b}(o))\big), \tag{3}$$

где F_l – карта признаков уровня l, L – число уровней, $\|$ – конкатенация, vec – разворачивание тензора в вектор. Схема извлечения показана на рисунке 1; размерность полученного представления не зависит от исходных размеров объекта, поскольку на каждом уровне используется один и тот же целевой размер RoI Align.

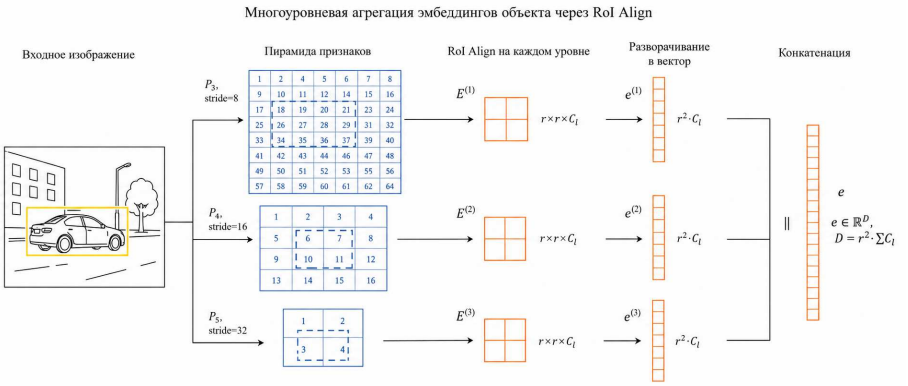


Рисунок 1 — Формирование локального эмбеддинга объекта с нескольких уровней пирамиды.

В реализации AREA источником служат выходы детектора, содержащие оценки классов и координаты прямоугольников. При трёх уровнях пирамиды и целевом размере 12×6 размерность для COCO с 80 классами составляет $12 \cdot 6 \cdot 84 \cdot 3 = 18\,144$. Для ускорения сравнения выполняется снижение размерности методом главных компонент до 512 координат. Проекция Φ строится

по совокупности эмбеддингов размеченного множества и неразмеченного пула; к обоим множествам применяется одна проекция.

Пусть E_L – множество эмбеддингов объектов размеченных изображений, $\mathcal{O}(x)$ – объекты, обнаруженные текущим детектором на изображении x. Мера новизны отдельного объекта определяется расстоянием до ближайшего объекта, представленного в размеченном множестве, а оценка изображения – максимумом этих расстояний по всем обнаруженным на нём объектам:

$$a_{\mathrm{AREA}}(x) = \max_{o \in \mathcal{O}(x)} \min_{z \in \Phi(E_L)} d_{\cos}\big(\Phi(\mathbf{e}(o)), z\big), \tag{4}$$

$$d_{\cos}(u,v) = 1 - \frac{u^{\mathsf{T}} v}{\|u\|_2 \|v\|_2}. \tag{5}$$

Изображение получает высокую оценку, если содержит хотя бы один объект, визуально непохожий на уже размеченные объекты, поэтому вклад редкого объекта сохраняется даже на кадре с большим количеством типичных объектов. Из полученных оценок формируется ранжирование неразмеченного пула, по которому для последующей разметки выбираются b изображений с наибольшими значениями критерия.

Дополнительно исследован критерий по плотности, определяемой суммой обратных квадратов расстояний до k ближайших размеченных объектов; изображения отбираются по возрастанию минимальной плотности среди обнаруженных объектов. По результатам сравнения основным вариантом AREA выбран критерий по расстоянию, который непосредственно выражает отсутствие похожего объекта в размеченном множестве и далее используется в MAREA.

Экспериментальная проверка проведена на Pascal VOC, COCO 2017, CrowdHuman и ind_safety, охватывающих многоклассовую детекцию, плотные одноклассовые сцены и задачи промышленной видеоаналитики. В цепочках активного обучения начальная доля размеченных изображений равна 0.2; далее на каждом раунде добавляется 0.1 обучающей выборки до достижения доли 0.7. На каждом раунде сохраняются ранее выбранные изображения, а объекты неразмеченного пула оцениваются по предсказаниям детектора, обученного на текущем размеченном множестве.

Основная модель – YOLOv8s; размер входного изображения составляет 640×640, размер пакета – 48, продолжительность обучения – 65 эпох. Качество измеряется по mAP50–95 на валидационном множестве; AREA сопоставляется со случайным отбором при одинаковых долях размеченных изображений и общих параметрах обучения.

На четырёх наборах данных максимальные выигрыши AREA над случайным отбором по mAP50–95 составляют от 0.25 до 0.93 п.п. Метод обеспечивает отбор с учётом отдельных объектов, их внутриклассовых визуальных различий и нескольких пространственных масштабов. Для дальнейшего ускорения требуется отказаться от отдельного обучения проекции и обеспечить информативность

представлений небольшой размерности, чему посвящено совместное обучение вложенных представлений в третьей главе.

Третья глава посвящена модификации выходного модуля YOLOv8 по принципу Matryoshka Representation Learning [1]. Вложенное представление устроено так, что начальные части полного вектора заранее выбранной длины могут использоваться как самостоятельные эмбеддинги, благодаря чему затраты на их хранение и сравнение изменяются после обучения одной модели. Это позволяет адаптировать модель к требованиям по скорости и памяти без переобучения.

Карта признаков l-го уровня на входе выходного модуля имеет вид $X_l \in \mathbb{R}^{B\times C_l\times H_l\times W_l}$, где B – размер пакета, C_l – число каналов, H_l и W_l – пространственные размеры. Задаётся набор делителей $\{d_1,\dots,d_K\}$, последний из которых равен единице. Режиму k соответствует число входных каналов

$$g_{l,k}=\max\left(1,\left\lfloor\frac{C_l}{d_k}\right\rfloor\right),\qquad X_{l,k}=X_l[:,:g_{l,k},:,:]. \tag{6}$$

В рабочей конфигурации используются делители 8,4,1, соответствующие одной восьмой, одной четверти и всем входным каналам, причём пространственные размеры карт во всех режимах сохраняются.

Регрессионная и классификационная ветви выходного модуля обрабатывают каждый режим с общими весами. Срез по входным каналам выполняется только для первой свёртки каждой ветви; её выходная размерность и последующие слои не изменяются. При таком устройстве дополнительные обучаемые параметры не вводятся, а в режиме полной ширины архитектура совпадает со стандартной YOLOv8 и различается только обученными значениями весов. Для всех режимов применяется стандартная детекционная функция потерь, и общая потеря имеет вид

$$\mathcal{L}_{\text{mat}}=\sum_{k=1}^{K}\alpha_k\mathcal{L}_{\text{det}}(H_k), \tag{7}$$

где H_k – совокупность выходов всех уровней пирамиды в режиме k. Для делителей 8,4,1 коэффициенты α_k равны 0.2,0.4,1.0. Схема модификации приведена на рисунке 2.

Для устойчивого обучения назначение предсказаний целевым объектам вычисляется один раз по выходам полной ширины и используется во всех вспомогательных проходах, устраняя различия положительных позиций и конкурирующие обучающие сигналы разных режимов.

Во вспомогательных проходах BatchNorm использует статистики текущего пакета без обновления накопленных оценок; полный режим сохраняет стандартную нормализацию. Вспомогательные потери включаются с задержкой и постепенно растущими коэффициентами, а на каждом шаге используются два режима, один из которых обязательно полный. Совместное применение этих приёмов сохраняет обучение вложенных представлений и снижает замедление относительно стандартной YOLOv8 до 6 %.

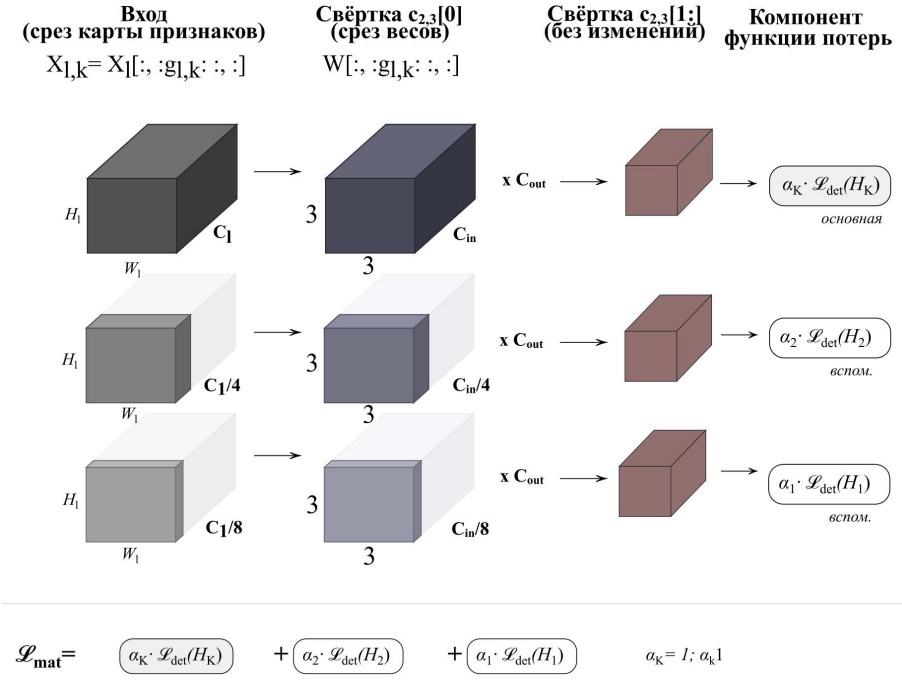


Рисунок 2 — Совместное обучение выходного модуля YOLOv8 на полных и сокращённых наборах входных каналов.

После обучения локальные признаки извлекаются с входов выходного модуля, то есть из промежуточных карт переходного модуля, с помощью RoI Align размера 4×4 на трёх уровнях пирамиды. При простой конкатенации уровней короткий начальный подвектор содержал бы признаки только первого уровня, поэтому для сохранения всех пространственных масштабов предложено переплетение признаков. Каждый уровень делится по каналам на блоки $B_{l,j}$, а блоки с одинаковым номером объединяются раньше перехода к следующему номеру:

$$e(o) = \left[B_{1,1} \| \cdots \| B_{L,1} \| B_{1,2} \| \cdots \| B_{L,2} \| \cdots \right]. \tag{8}$$

Начальная доля координат такого вектора включает соответствующую долю признаков каждого уровня пирамиды. Схема показана на рисунке 3.

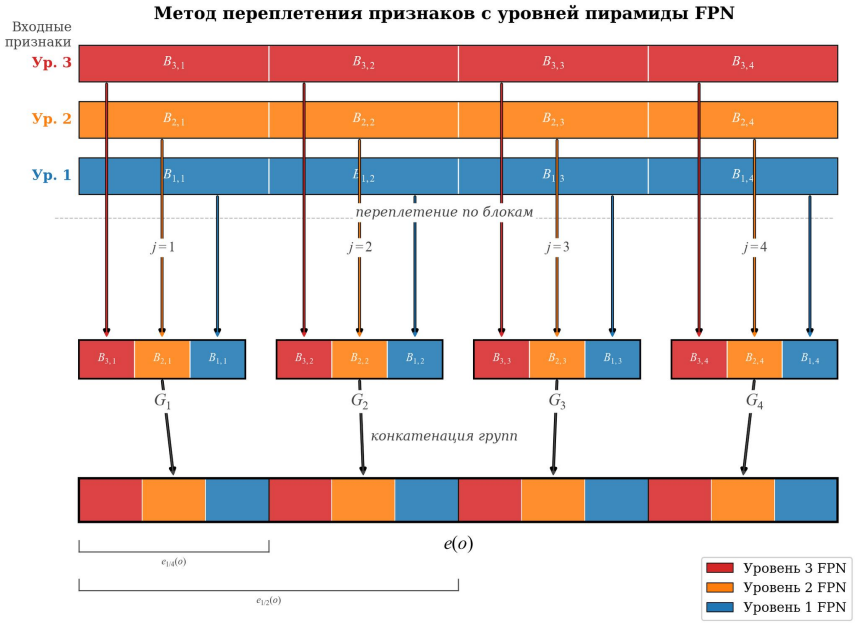


Рисунок 3 — Переплетение блоков признаков разных уровней пирамиды; схема показана для долей $1/4$, $1/2$ и 1.

В основной конфигурации объект описывается семейством

$$\mathcal{E}(o) = \{e_{1/8}(o), e_{1/4}(o), e_1(o)\}. \tag{9}$$

Масштабируемость обеспечивается выбором длины подвектора при сохранении в нём информации разных пространственных масштабов; перед сравнением представлений выполняется L2-нормализация, а отдельное обучение PCA-проекции не требуется.

Вложенные представления применены к поиску похожих объектов. Для объекта-запроса требуется найти K ближайших объектов в галерее $\mathcal{G}$. На первом этапе HNSW-индекс, построенный по коротким подвекторам, формирует расширенный список из N_1 кандидатов, где $N_1 \gg K$ и $N_1 \ll |\mathcal{G}|$. На втором этапе расстояния пересчитываются по полным представлениям только для этих кандидатов, после чего выбираются K ближайших. Таким образом, дорогие сравнения полной размерности выполняются для небольшой части галереи.

На Pascal VOC каскад $e_{1/8} \to e_1$ сокращает среднее время запроса с 20.5 до 6.7 мс, то есть в 3.1 раза, при Recall@10, равном 96.95 % против 96.92 % у поиска по полным векторам. В данной постановке каждому запросу соответствует один релевантный объект; Recall@10 показывает долю запросов, для которых он найден среди первых десяти результатов. При использовании одной восьмой входных каналов выходного модуля на ind_safety получено mAP50, равное 92.4 %

против 15.3 % при наивном усечении стандартной модели, что соответствует выигрышу 77.1 п.п. Полученные результаты подтверждают пригодность коротких представлений как для детекции, так и для ускоренного поиска.

В четвёртой главе предложен метод MAREA (Matryoshka Active learning with Region Embedding Alignment), использующий вложенные локальные представления для активного обучения детекторов объектов. Источником эмбеддингов служит детектор YOLOv8s, обученный совместно на полных и сокращённых наборах каналов; отбор сохраняет смысл AREA, но выполняется в пространстве признаков переходного модуля с каскадным поиском и без PCA.

Пусть $E(x)$ – вложенные эмбеддинги объектов изображения x, E_L – эмбеддинги объектов размеченного множества. Оценка новизны в режиме полного поиска определяется выражением

$$a_{\mathrm{MAREA}}(x) = \max_{e \in E(x)} \min_{z \in E_L} d_{\cos}(e_1, z_1). \tag{10}$$

Для ускорения применяется трёхэтапный каскад $e_{1/8} \to e_{1/4} \to e_1$. Грубый отбор по одной восьмой вектора формирует $4b$ кандидатов; уточнение по одной четверти оставляет $2b$; финальное ранжирование по полным представлениям даёт пакет из b изображений. Пакетный критерий записывается как

$$S_t^* = \operatorname*{arg\,max}_{S \subset U_{t-1},\ |S|=b} \sum_{x \in S} a_{\mathrm{MAREA}}(x). \tag{11}$$

Отобранные изображения размечаются и добавляются к обучающему множеству по схеме (1), после чего детектор переобучается, а представления следующего раунда вычисляются с учётом его нового состояния.

Для проверки используется общая схема экспериментов второй главы. MAREA сопоставляется со случайным отбором, AREA и стадией отбора по разнообразию PPAL (Plug and Play Active Learning). Последняя применяется ко всему неразмеченному пулу; сравнение с ней проведено на Pascal VOC и CrowdHuman. Контрольный вариант RANDOM-MAT использует ту же модель с вложенными представлениями при случайном выборе изображений; близость RANDOM и RANDOM-MAT показывает, что прирост MAREA связан с критерием отбора данных.

На четырёх наборах данных максимальные выигрыши MAREA по mAP50–95 составляют от 0.41 до 1.19 п.п. над случайным отбором и от 0.30 до 0.65 п.п. над AREA. При пяти долях разметки от 0.3 до 0.7 MAREA превосходит случайный отбор во всех 20 сравнениях, а AREA – в 18 из 20. Цепочки активного обучения для четырёх наборов данных представлены на рисунках 4–7, где качество детекции сопоставлено при последовательном увеличении доли размеченных изображений.

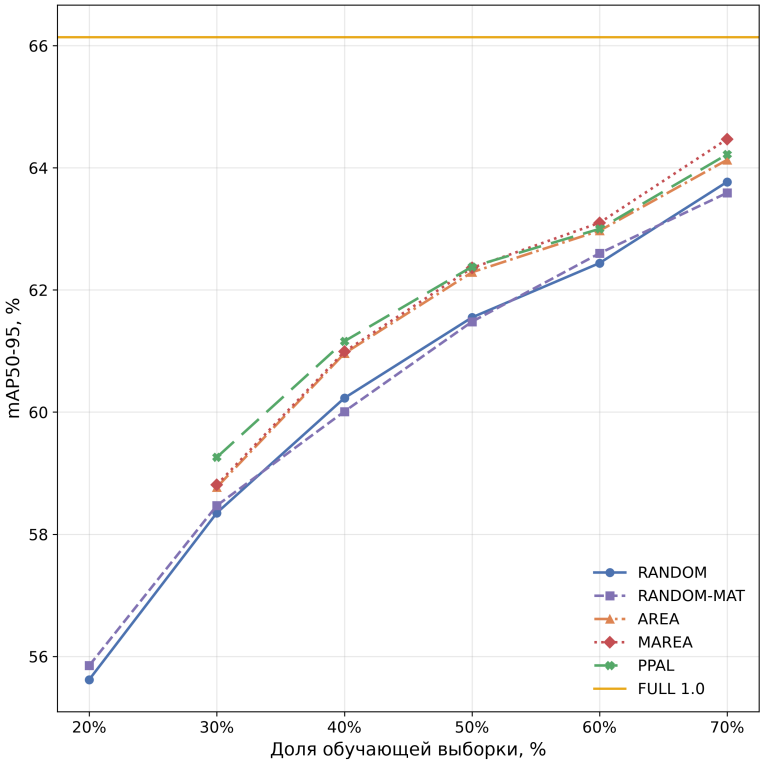


Рисунок 4 — Качество детекции на Pascal VOC при увеличении доли размеченных изображений; горизонтальная линия соответствует обучению на полном наборе данных.

На CrowdHuman MAREA превосходит стадию отбора по разнообразию PPAL при всех пяти долях разметки, достигая максимального выигрыша 1.06 п.п. при доле 0.4: 54.29 % против 53.23 %.

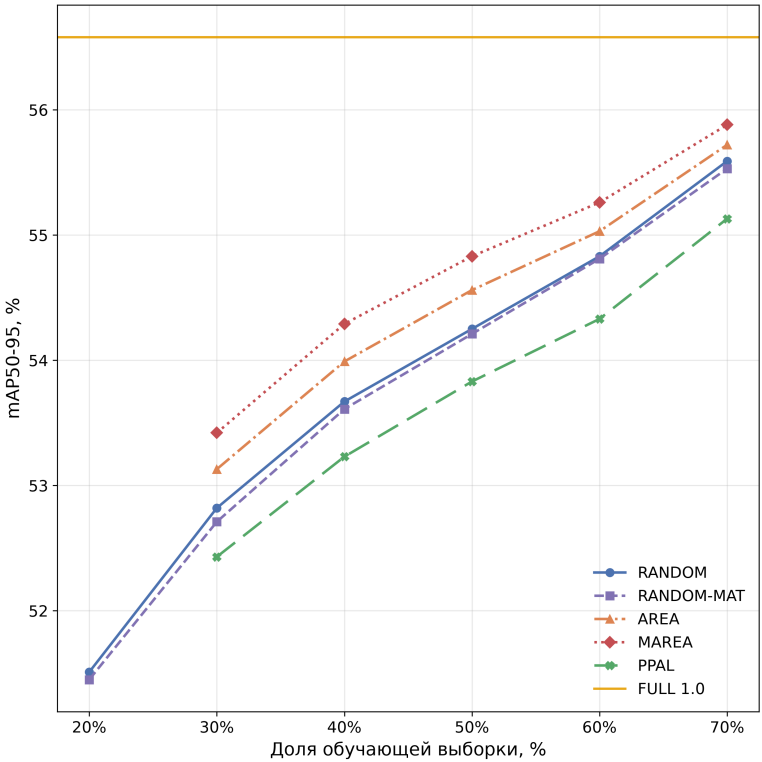


Рисунок 5 — Качество детекции на CrowdHuman при увеличении доли размеченных изображений; горизонтальная линия соответствует обучению на полном наборе данных.

На COCO 2017 MAREA превосходит AREA при долях разметки 0.3–0.6, а при доле 0.7 уступает на 0.15 п.п.; средний выигрыш над AREA составляет 0.17 п.п.

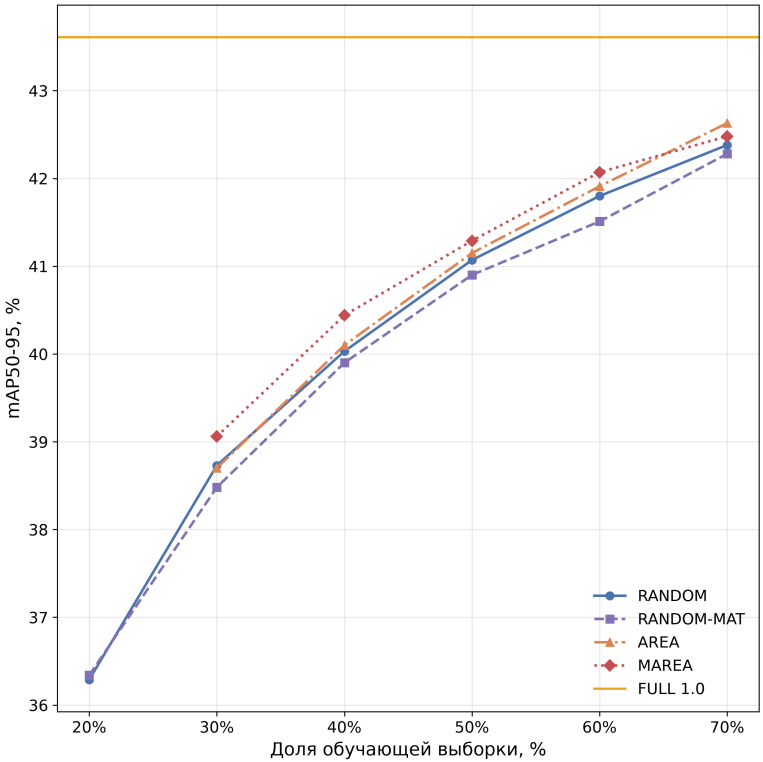


Рисунок 6 — Качество детекции на COCO 2017; горизонтальная линия соответствует обучению на полном наборе данных.

На ind_safety при использовании 60 % обучающих изображений получено mAP50–95, равное 67.18 %, против 66.85 % при обучении на полном наборе. Результаты для промышленного набора данных на рисунке 7 показывают повышение качества на 0.33 п.п. при сокращении числа размеченных изображений на 40 %.

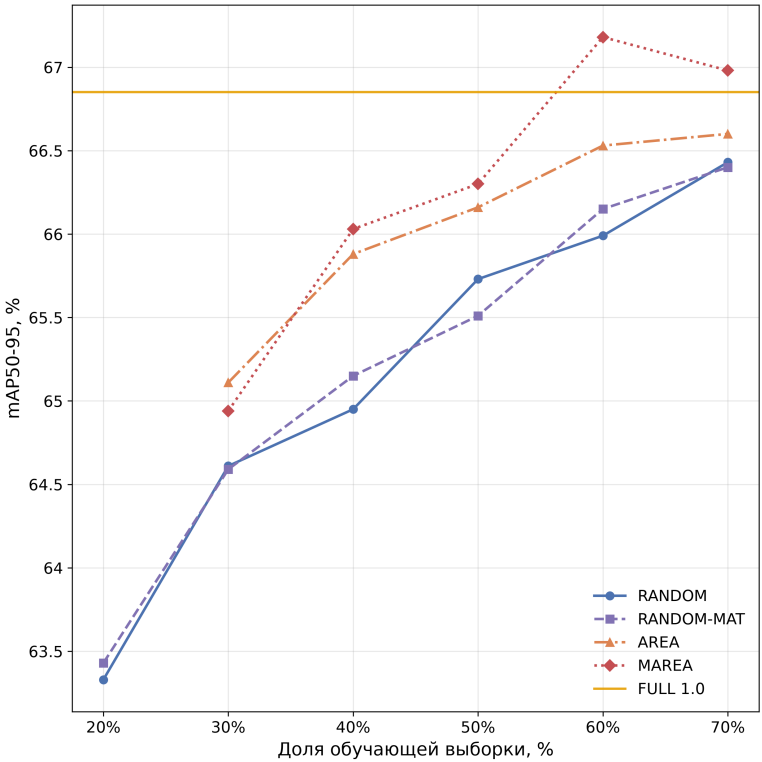


Рисунок 7 — Качество детекции на ind_safety; горизонтальная линия соответствует mAP50–95, равному 66.85 %, при обучении на полном наборе данных.

На CrowdHuman подготовка разбиения занимает 516 секунд для MAREA, 730 секунд для AREA и 1029 секунд для стадии отбора по разнообразию PPAL, что соответствует ускорению в 1.41 и 1.99 раза. Замена PCA-предобработки вложенными представлениями и последовательное уточнение поиска обеспечивают преимущество по совокупности качества детекции и скорости отбора данных.

Дальнейшие исследования связаны с объединением новизны и несогласованности вложенных представлений в гибридном критерии отбора, а также с переносом MAREA на другие детекторы с многомасштабными картами признаков.

Публикации автора по теме диссертации

1. *Матвеев, И. А.* Модификация нейросетевого детектора для получения вложенных эмбеддингов объектов / И. А. Матвеев, В. А. Соколов // Известия Российской академии наук. Теория и системы управления. — 2026. — № 4. — С. 78—87. — URL: https://doi.org/10.7868/S3034644426040076.
2. *Sokolov, V. A.* Extracting Embeddings for Active Learning Corresponding to Regions of Interest in an Image / V. A. Sokolov, I. A. Matveev, L. M. Teplyakov // Pattern Recognition and Image Analysis. — 2025. — Vol. 35. — P. 417—426.
3. *Соколов, В. А.* Использование эмбеддингов-матрёшек в критерии отбора для активного обучения свёрточных сетей / В. А. Соколов, И. А. Матвеев // Математические методы распознавания образов: тезисы докладов 22-й Всероссийской конференции. — Муром, Россия, 2025. — С. 83—85. — 22–26 сентября 2025 г.
4. *Соколов, В. А.* Выделение эмбеддингов для активного обучения, соответствующих регионам интереса на изображении / В. А. Соколов, И. А. Матвеев, Л. М. Тепляков // 15-я Международная конференция «Интеллектуализация обработки информации» (ИОИ-2024): тезисы докладов. — Гродно, Беларусь, 2024. — С. 45. — URL: https://mmro.ru/wp-content/uploads/2025/01/idp-2024.pdf.
5. *Соколов, В. А.* Критерий активного обучения на основе анализа локальных эмбеддингов / В. А. Соколов, И. А. Матвеев, Л. М. Тепляков // 67-я Всероссийская научная конференция Московского физико-технического института: тезисы докладов. Прикладная математика и информатика. — Долгопрудный, Россия, 2025. — С. 127—129. — URL: https://old.mipt.ru/upload/medialibrary/1e0/konferentsiya_5_fpmi.pdf.
6. owl.Base : свидетельство о государственной регистрации программы для ЭВМ № 2023689395 / М. В. Лазуков, Г. Е. Макаров, С. А. Марков, А. Д. Мершавка, Ю. А. Прохоров, В. А. Соколов, Л. М. Тепляков, А. А. Федоров, У. С. Цинпаев, И. А. Чайкин, Е. А. Швец, А. В. Шошин ; Общество с ограниченной ответственностью «НОРД ВИЖЕН ИНТЕЛЛЕДЖЕНС РЕСЕРЧ». — № 2023689496 ; заявл. 29.12.2023 ; зарегистрировано в реестре программ для ЭВМ 29.12.2023.
7. owl.Drill : свидетельство о государственной регистрации программы для ЭВМ № 2023689397 / М. В. Лазуков, С. А. Марков, А. Д. Мершавка, Ю. А. Прохоров, В. А. Соколов, Л. М. Тепляков, А. А. Федоров, А. В. Фомин, И. А. Чайкин, Е. А. Швец, А. В. Шошин, Д. И. Юрманов ; Общество с ограниченной ответственностью «НОРД ВИЖЕН ИНТЕЛЛЕДЖЕНС РЕСЕРЧ». — № 2023689498 ; заявл. 29.12.2023 ; зарегистрировано в реестре программ для ЭВМ 29.12.2023.

8. owl.Guard: версия 6.0 : свидетельство о государственной регистрации программы для ЭВМ № 2024617256 / М. В. Вербицкий, П. Г. Гуляев, Я. С. Гурьев, К. А. Ионкин, М. В. Лазуков, С. И. Литвинов, Г. Е. Макаров, С. А. Марков, А. Д. Мершавка, О. В. Пинарин, Ю. А. Прохоров, В. А. Соколов, Л. М. Тепляков, А. А. Федоров, А. В. Фомин, У. С. Цинпаев, Е. А. Швец, А. В. Шошин, Д. И. Юрманов ; Общество с ограниченной ответственностью «НОРД ВИЖЕН ИНТЕЛЛЕДЖЕНС РЕСЕРЧ». — № 2024616277 ; заявл. 29.03.2024 ; зарегистрировано в реестре программ для ЭВМ 29.03.2024.