

# ДОМЕННАЯ АДАПТАЦИЯ ДЕТЕКТОРОВ ОБЪЕКТОВ С ПРИМЕНЕНИЕМ МОДЕЛИ УЧИТЕЛЯ С ПОДАНСАМБЛЯМИ

## DOMAIN ADAPTATION OF OBJECT DETECTORS USING THE TEACHER MODEL WITH SUB ENSEMBLES

A. Medvedev

*Summary.* Domain adaptation for object detection remains a critical issue in computer vision, especially when models trained in the same domain are deployed to significantly different target domains. Traditional approaches, such as fine-tuning using labeled target domain data or using uniform teacher-student designs, often fail to cope with the complexity and variability of real-world scenarios. This paper presents a novel Sub-Ensemble Teacher (SET) approach that leverages the strengths of multiple teacher models to provide robust and comprehensive guidance to the student model for domain adaptation. The SET framework involves training multiple teacher models on different subsets of domain input data, each representing different aspects of the domain. These sub-ensembles generate a rich set of pseudo-labels for the unlabeled target domain data, which are then used to train the student model. This ensemble method improves the robustness and generalizability of the student model by efficiently approximating Bayesian inference, accounting for model uncertainty, and reducing label noise.

Experimental results on benchmark datasets show that the SET approach significantly outperforms traditional single-supervised models on cross-domain object detection tasks. The subensemble technique not only improves detection accuracy, but also provides a more robust estimate of uncertainty, making it a powerful tool for adapting object detectors to diverse and complex environments.

*Keywords:* object detection, domain adaptation, computer vision, deep learning, one-stage object detectors, sub-ensembles, object localization, uncertainty estimation.

Медведев Анатолий Андреевич

Аспирант, Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»  
medvedev.workspace@gmail.com

*Аннотация.* Адаптация доменов в задачах детекции объектов остается важной проблемой в области компьютерного зрения, особенно когда модели, обученные на одних наборах данных, применяются в существенно отличающихся целевых доменах. Традиционные методы, такие как тонкая настройка с использованием размеченных данных целевого домена или подходы «учитель-ученик», часто не справляются со сложностью и изменчивостью реальных условий. В данной статье предложен новый метод Sub-Ensemble Teacher (SET), который использует преимущества нескольких моделей учителей для обеспечения надежного и всестороннего руководства при адаптации модели ученика. Структура SET включает обучение нескольких моделей учителей на разных подмножествах входных данных, каждая из которых охватывает различные аспекты предметной области. Эти подансамбли создают богатый набор псевдометок для неразмеченных данных целевого домена, которые затем используются для обучения модели ученика. Этот подход увеличивает надежность и обобщаемость модели ученика за счет эффективной аппроксимации байесовского вывода, учета неопределенности модели и уменьшения шума меток. Экспериментальные результаты на эталонных наборах данных показывают, что метод SET превосходит традиционные модели с одним учителем в задачах кроссдоменной детекции объектов. Подход с подансамблями не только повышает точность обнаружения, но и обеспечивает более надежную оценку неопределенности, что делает его мощным инструментом для адаптации детекторов объектов к разнообразным и сложным условиям.

*Ключевые слова:* детекция объектов, доменная адаптация, компьютерное зрение, глубокое обучение, одноступенчатые детекторы объектов, подансамбли, локализация объектов, оценка неопределенности.

**Б**ыстрые успехи в области глубокого обучения значительно расширили возможности систем обнаружения объектов, позволив использовать их в самых разных сферах: от автономного вождения до медицинской визуализации. Обнаружение объектов, являясь фундаментальной задачей компьютерного зрения, включает в себя идентификацию и локализацию объектов на изображении. Несмотря на значительный прогресс, остается одна важная проблема: способность этих систем хорошо обобщать данные между доменами. Такие изменения происходят, когда распределение данных в целевом домене отличается от распределения в исходном домене, что часто приводит к значительному снижению эффективности обнаружения. Методы адаптации домена нацелены на преодоление этого разрыва, пере-

нося знания из размеченного исходного домена в неразмеченный или слабо размеченный целевой домен.

Адаптация домена имеет критическое значение для обнаружения объектов, поскольку реальные сценарии часто значительно отличаются от условий, в которых модели изначально обучались. Традиционные методы адаптации домена для обнаружения объектов обычно включают дообучение моделей с использованием небольшого количества размеченных данных целевого домена [1, 2] или использование единой структуры «учитель-ученик» [3]. Однако эти подходы часто испытывают трудности с обобщением данных в значительной степени отличающихся доменах из-за их ограниченной способности охватывать весь спектр вариаций домена

и являются непрактичными для многих реальных приложений, где размеченные данные в целевом домене недостаточны или отсутствуют. Последние достижения в области методов несупервизируемой адаптации домена (UDA) направлены на преодоление этого ограничения за счет использования размеченных данных из исходного домена и неразмеченных данных из целевого домена [4]. Эти методы обычно включают в себя сопоставление распределений признаков [5] между доменами для минимизации сдвига домена и улучшения производительности обнаружения в целевом домене.

Для решения этих проблем мы предлагаем новую структуру Sub-Ensemble Teacher (SET) для адаптации домена детекторов объектов. Подход SET использует несколько моделей-учителей, которые генерируют разнообразный набор псевдо-меток для неразмеченных данных целевого домена. Ключевое новшество структуры SET заключается в использовании подансамблевых моделей для аппроксимации байесовского вывода. Этот ансамблевый метод повышает надежность и обобщающие способности модели-ученика за счет снижения шума в метках и улучшения качества псевдо-меток.

Вкратце, основные вклады данной работы заключаются в следующем:

- Мы предлагаем новый подход Subensemble Teacher, который использует несколько моделей-учителей, каждая из которых обучена на разных подмножествах данных исходного домена, для генерации надежных и полных псевдо-меток для целевого домена. Эта структура устраняет ограничения традиционных моделей с одним учителем, повышая надежность и обобщаемость.
- Подход SET эффективно аппроксимирует байесовский вывод путем агрегирования предсказаний от ансамбля подмоделей. Этот метод учитывает неопределенность модели, предоставляя вероятностную меру уверенности в предсказаниях, что критически важно для надежной адаптации домена.
- Путем усреднения псевдо-меток, полученных от различных подмоделей, структура SET снижает влияние шумных меток. Это приводит к более чистым и точным псевдо-меткам, улучшая процесс обучения и повышая эффективность адаптации модели-ученика.

Адаптация домена для обнаружения объектов привлекла значительное внимание в последние годы, что привело к разработке различных инновационных подходов. В этом разделе представлен обзор значимых связанных работ, классифицированных по нескольким ключевым методологиям: выравнивание признаков, соревновательное обучение, самообучение, несупервизируемая адаптация домена, а также ансамблевые и подансамблевые модели для обнаружения объектов.

*Выравнивание признаков.* Методы выравнивания признаков стремятся уменьшить несоответствие доменов путем согласования распределения признаков в исходном и целевом доменах. Одним из инновационных подходов является Domain Adversarial Neural Network (DANN), предложенная Ганиным и др. [6], которая вводит слой инверсии градиента для стимулирования модели к изучению доменно-инвариантных признаков. Эта концепция была расширена для обнаружения объектов Ченом и др. [7] в их работе над Domain Adaptive Faster R-CNN, которая выравнивает признаки на уровне изображений и экземпляров с использованием соревновательного обучения. Более современные техники сосредотачиваются на согласовании более высокоуровневых признаков и включении жестких согласованностей. Например, работа Хэ и др. [8, с. 208] вводит подход многоуровневого выравнивания признаков (MLFA), который выравнивает признаки через несколько слоев системы обнаружения объектов, повышая надежность между доменами.

*Соревновательное обучение.* Соревновательное обучение стало краеугольным камнем в адаптации домена, используя соревновательные сети для минимизации разрыва между доменами. Сайто и др. [9] представили метод сильного и слабого выравнивания распределений (SWDA), который использует комбинацию стратегий сильного и слабого выравнивания для балансирования между выравниванием признаков и производительностью классификации.

*Самообучение.* Методы самообучения используют псевдо-метки, созданные на основе прогнозов модели в целевом домене, для итеративного улучшения производительности. Хсу и др. [10] представили прогрессивную структуру адаптации домена (PDA), которая поэтапно улучшает псевдо-метки через многостадийный процесс обучения. Этот метод снижает шум в псевдо-метках и улучшает способность модели адаптироваться к целевому домену.

*Несупервизированная адаптация домена.* Техники несупервизированной адаптации домена особенно ценны, когда размеченные данные в целевом домене недостаточны или отсутствуют. Эти методы используют размеченные данные из исходного домена и неразмеченные данные из целевого домена. Недавние достижения включают ConfMix [11] и DACA [12], которые интегрируют стратегию дополнения данных домена для создания псевдо-меток, повышая надежность методов самообучения.

*Ансамблевые модели для обнаружения объектов.* Ансамблевые методы объединяют несколько моделей для улучшения надежности и производительности. Ансамбли особенно эффективны в задачах обнаружения

объектов, так как они могут усреднять ошибки отдельных моделей и предоставлять более надежные прогнозы. Ранним примером является работа Ли и др. [13], демонстрирующая эффективность ансамблевых методов в задачах обнаружения объектов. Подансамблевые модели, особый тип ансамблевого подхода, включают обучение нескольких моделей-учителей на разных подмножествах исходных данных домена. Это разнообразие улучшает способность модели обобщать данные между доменами. Структура Sub-Ensemble Teacher (SET), предложенная в данной работе, использует эту концепцию для создания богатого набора псевдо-меток для целевого домена, улучшая производительность адаптации домена.

Концепция использования моделей учитель-ученик становится все более популярной в адаптации предметной области благодаря их способности передавать знания от предварительно обученной модели-учителя к модели-ученику [14]. Структура SET состоит из трех основных компонентов: модели-учителя с подансамблями, модели-ученика и механизма адаптации домена. В этом разделе мы опишем принцип работы и преимущества модели SET, включая соответствующие формулы.

*Модель учителя с подансамблями.* Несколько подансамблевых моделей обучаются на различных подмножествах исходных данных домена. Эти подмножества спроектированы таким образом, чтобы охватывать широкий спектр вариаций в исходном домене, обеспечивая специализацию каждой модели-учителя на разных аспектах домена. Это разнообразие имеет критическое значение для предоставления всесторонних псевдо-меток, которые отражают различные характеристики целевого домена и помогают модели более эффективно адаптироваться, обучаясь на более широком спектре сценариев и функций.

*Модель ученика.* Модель ученика первоначально обучается с использованием размеченных данных из исходного домена. Во время фазы адаптации домена она дообучается с использованием псевдо-меток, созданных подансамблевыми моделями-учителями на неразмеченных данных целевого домена. Псевдо-метки комбинируются из ансамбля для предоставления более сбалансированного и точного сигнального эталона.

Формулы, описывающие взаимодействие моделей в структуре SET, можно представить следующим образом:

1. Пусть  $D_s$  и  $D_t$  обозначают исходный и целевой домены соответственно.
2. Модели-учителя  $T_i$  обучаются на различных подмножествах данных исходного домена  $D_s$ :  $T_i = f(D_s^i)$  для  $i = 1, \dots, N$ , где  $D_s^i$  является подмножеством  $D_s$ , а  $N$  — количество подансамблевых моделей.

3. Модель ученика  $S$  первоначально обучается на полном наборе данных исходного домена  $D_s$ :  $S = g(D_s)$ .
4. Псевдо-метки для данных целевого домена  $D_t$  генерируются путем усреднения прогнозов подансамблевых моделей-учителей:  $\hat{y}_t = (1/N) * \sum_{i=1}^N T_i(D_t)$ .
5. Модель ученика дообучается с использованием этих псевдо-меток:  $S' = h(D_t, \hat{y}_t)$ .

Где  $f$ ,  $g$  и  $h$  обозначают функции обучения для моделей учителя и ученика соответственно.

Таким образом, структура SET обеспечивает более сбалансированное и точное обучение модели ученика, улучшая адаптацию к целевому домену за счет использования разнообразных и обширных псевдо-меток, полученных от ансамбля моделей-учителей.

*Принципы Байеса.* Использование подансамблевых моделей в адаптации домена, особенно в контексте обнаружения объектов, предлагает несколько преимуществ на основе байесовских принципов. Ансамблевые методы известны своей способностью уменьшать шум в метках за счет усреднения ошибок отдельных моделей. Подход SET использует это свойство для генерации более чистых и точных псевдо-меток, что критично для эффективной адаптации домена. Байесовский вывод предоставляет вероятностный подход к учету неопределенности модели, а ансамбли можно рассматривать как практическую аппроксимацию байесовских методов. Здесь мы обсудим, как подансамблевые модели соотносятся с байесовскими принципами и какие конкретные преимущества они предлагают в аппроксимации данных:

1. Байесовский вывод включает обновление вероятностной оценки гипотезы по мере появления новых доказательств или информации. Естественным образом он включает неопределенность, рассматривая параметры модели как случайные величины с априорными распределениями. Такой подход обеспечивает всестороннее понимание уверенности модели и неопределенности прогнозов. Подансамблевые модели аппроксимируют байесовский вывод, объединяя предсказания нескольких моделей, каждая из которых представляет отдельную гипотезу о данных. Разнообразие ансамбля позволяет ему захватывать более широкий спектр неопределенностей, аналогично байесовским апостериорным распределениям.
2. Байесовские методы интегрируют все возможные модели, взвешенные по их апостериорным вероятностям, предоставляя определенную степень неопределенности. Усредняя предсказания разных моделей, обученных на различных подмножествах данных, подансамблевые модели имитируют эту интеграцию, предлагая практический

способ оценки неопределенности без вычислительной сложности полного байесовского метода.

3. Подансамблевые модели снижают переобучение, сглаживая смещения и предвзятости отдельных моделей, подобно регуляризирующему эффекту байесовских априорных подходов. Это приводит к более высокой точности и надежности в задачах обнаружения объектов, особенно в различных целевых областях.

В этом разделе мы опишем экспериментальную установку, использованную для оценки подхода Sub-Ensemble Teacher (SET), включая подробности о моделях, наборах данных и метриках оценки.

Для моделей учителя и ученика мы использовали архитектуру YOLOv5 [15], известную своей эффективностью и точностью. YOLOv5, разработанная Ultralytics, является частью семейства моделей You Only Look Once (YOLO), предназначенных для обнаружения объектов в реальном времени. YOLOv5 особенно подходит для наших экспериментов благодаря балансу между скоростью и производительностью, что делает ее идеальной для сценариев с ограниченными вычислительными ресурсами.

Мы обучали и оценивали подход SET, используя набор данных Cityscapes (Рисунок 1), крупномасштабный набор данных, предназначенный для понимания городской сцены. Cityscapes широко используется в исследованиях компьютерного зрения, особенно для таких задач, как семантическая сегментация, сегментация экземпляров и обнаружение объектов. Изображения сделаны при различных погодных условиях (например, солнечно, дождливо, пасмурно) и в разное время суток, что делает этот набор данных сложным для адаптации домена.

На ранних этапах модель ученика, обученная с подансамблевыми учителями, показала более высокую точность в обнаружении объектов, что отражается в более высоких показателях precision и recall. Агрегированные псевдо-метки от нескольких моделей учителей уменьшили влияние шума в метках и повысили устойчивость модели ученика, позволяя ей более эффективно справляться с различными изменениями домена с самого начала. Модель ученика, обученная по методике SET, демонстрировала более быструю сходимость, достигая более высоких показателей производительности за меньшее количество эпох по сравнению с моделью с одним учителем (Рисунок 2).

Экспериментальные результаты на эталонных наборах данных демонстрируют эффективность подхода SET. По сравнению с традиционными моделями с одним учителем, структура SET достигает более высокой точности

и устойчивости в задачах обнаружения объектов в целевом домене (Таблица 1).

Таблица 1.

Количественные результаты (mAP) для эталона Cityscapes

Метод	Детектор	Бекбоун	mAP
SIGMA [16]	FCOS	VGG-16	53.7
SCAN [17]	FCOS	VGG-16	52.6
EPM [18]	FCOS	ResNet-101	51.2
ConfMix	YOLOv5	CSP-Darknet53	56.2
DACA	YOLOv5	CSP-Darknet53	60.6
SET (Наш)	YOLOv5	CSP-Darknet53	67.3

Улучшения особенно значимы в сценариях с существенными доменными сдвигами, где разнообразие и комплементарные сильные стороны ансамбля играют критическую роль в преодолении доменного разрыва. Основные выводы нашей работы заключаются в следующем:

1. Использование нескольких подмоделей в структуре SET улучшает устойчивость детектора объектов за счет захвата широкого диапазона изменений в исходном домене. Это разнообразие приводит к лучшей обобщаемости в различных целевых доменах, так как каждая подмодель привносит уникальную информацию, которая коллективно улучшает общую производительность.
2. Подход SET аппроксимирует байесовский вывод путем агрегации прогнозов от нескольких подмоделей. Эта аппроксимация эффективно захватывает неопределенность модели и предоставляет вероятностную меру уверенности в прогнозах, что критично для надежной адаптации домена.
3. Подход SET является осуществимым и масштабируемым. Обучение нескольких подмоделей на подмножествах исходных данных домена может быть параллелизовано, а агрегация псевдо-меток может выполняться эффективно. Это делает подход SET подходящим для задач адаптации домена в крупномасштабных и сложных средах.

Несмотря на значительные улучшения в адаптации домена для обнаружения объектов, достигнутые подходом SET, некоторые области требуют дальнейшего изучения. Будущие исследования могут изучить динамические стратегии взвешивания для агрегации псевдо-меток, где веса адаптируются на основе уверенности или производительности каждой подмодели в целевом домене. Оптимизация разнообразия подмоделей путем изучения различных способов разбиения данных исходного домена или путем включения различных архитектур моделей может еще больше улучшить устойчивость и производительность структуры SET.

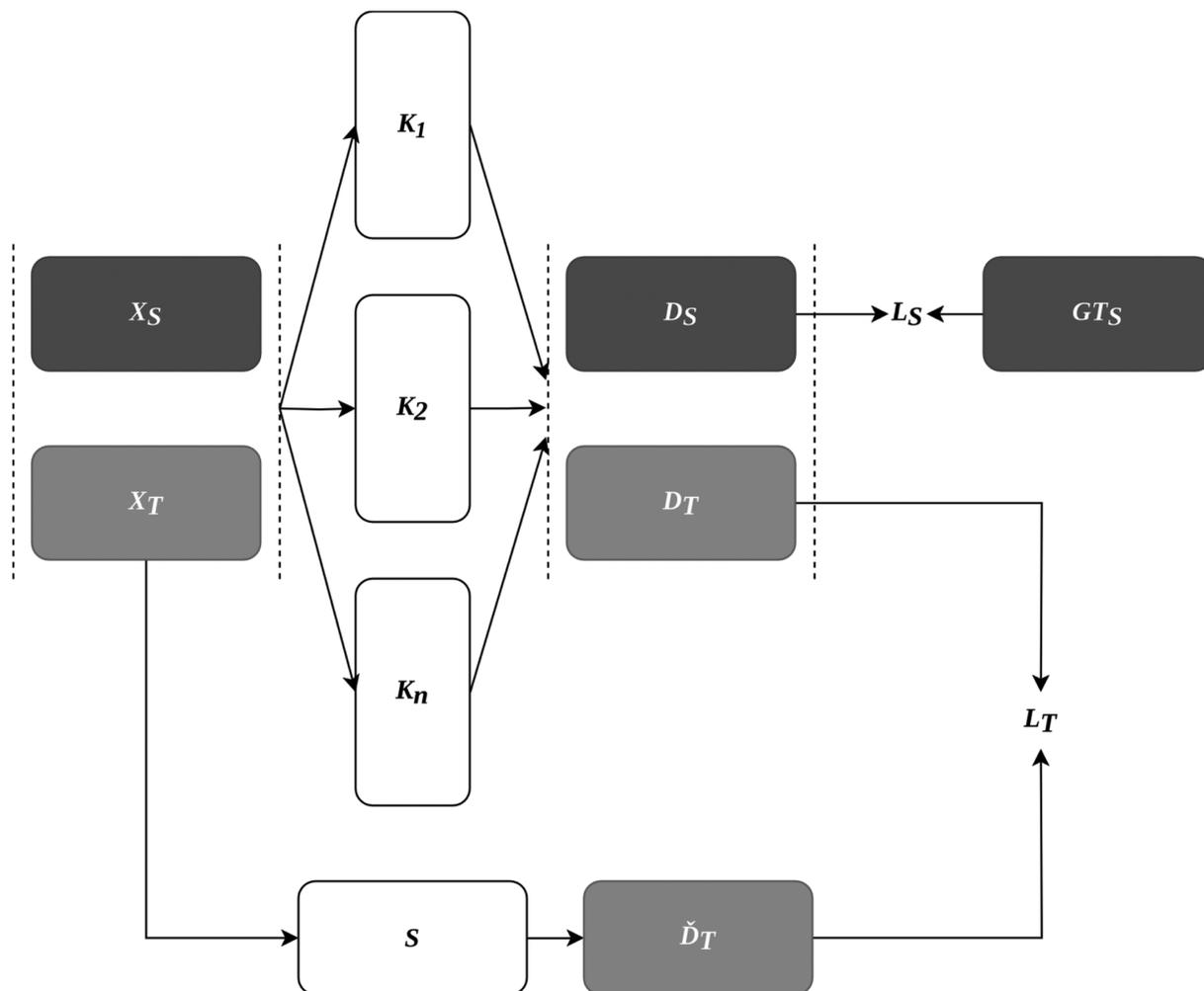


Рис. 1. Обзор предлагаемого подхода. Модель ученика  $S$  обучается с использованием псевдо-меток, созданных подансамблевыми моделями учителя  $K_n$  на неразмеченных данных целевого домена  $X_T$

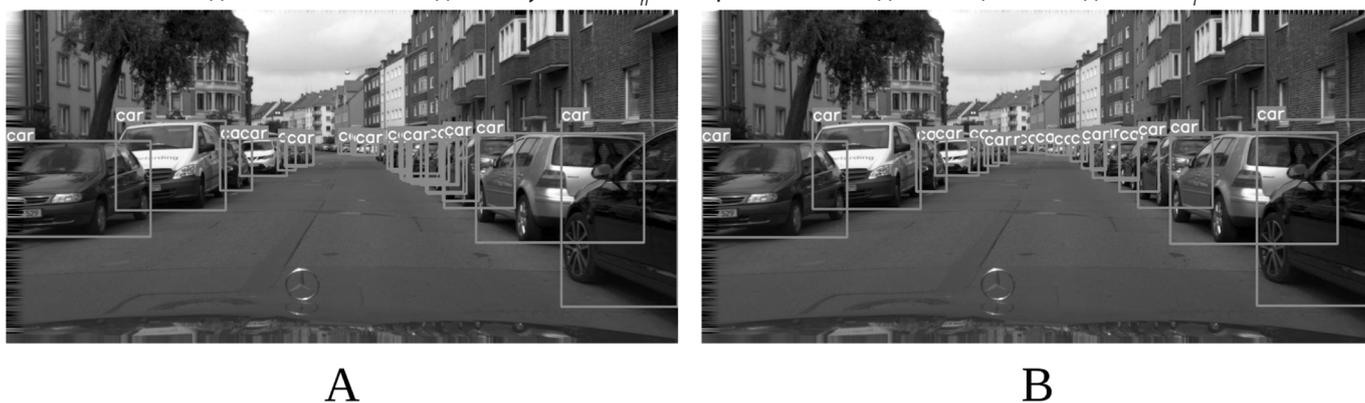


Рис. 2. Сравнение производительности генерации псевдо-меток с одним учителем (А) и учителем с подансамблями (В)

### Заключение

Область адаптации домена для обнаружения объектов изобилует разнообразными и инновационными подходами, которые способствуют созданию надежных и адаптируемых моделей. Несмотря на значительный прогресс, остаются нерешенные задачи, такие как не-

хватка размеченных данных целевого домена и сложность реальных условий. Подход Sub-Ensemble Teacher (SET) представляет собой перспективное достижение в этой области. Подансамблевые модели предлагают мощную и практическую аппроксимацию байесовских методов в контексте адаптации домена для обнаружения объектов. Используя сильные стороны нескольких

моделей, они охватывают широкий спектр неопределенностей, снижают переобучение и улучшают обобщаемость. Эти преимущества, основанные на байесовских принципах, делают подансамблевые модели надежным и эффективным инструментом для аппроксимации данных, особенно в сложных и разнообразных доменах.

Применение подхода SET к сценариям реальной адаптации, таким как автономное вождение или медицинская визуализация, может предоставить ценные

данные о его практической эффективности и ограничениях. Сочетание подхода SET с другими передовыми методами адаптации домена, такими как самообучение или метаобучение, может привести к созданию еще более надежных и обобщаемых моделей для обнаружения объектов. В будущих работах можно было бы дальше изучать оптимизацию конфигураций подгрупп и интеграцию более сложных байесовских методов для повышения их эффективности.

#### ЛИТЕРАТУРА

1. Li Wanyi, Li Fuyu, Luo Yongkang, Wang Peng Deep Domain Adaptive Object Detection // Conference on Computer Vision, and Pattern Recognition. — 2020.
2. Xingxu Yao, Sicheng Zhao, Pengfei Xu, Jufeng Yang Multi-Source Domain Adaptation for Object Detection // Conference on Computer Vision and Pattern Recognition. — 2021.
3. Lawrence G. Phillips, David B. Grimes, Yihan Jessie Li Teacher-Student Domain Adaptation for Biosensor Models // 2020 International Conference on Learning Representations. — 2020.
4. Poojan Oza, Vishwanath A. Sindagi, Vibashan VS, Vishal M. Patel Unsupervised Domain Adaptation of Object Detectors: A Survey // ArXiv abs/2105.13502. — 2021.
5. Debjeet Majumdar, Vinay Namboodiri Unsupervised domain adaptation of deep object detectors // The European Symposium on Artificial Neural Networks. — 2018.
6. Yaroslav Ganin, E. Ustinova, Hana Ajakan, Pascal Germain, H. Larochelle, François Laviolette, M. Marchand, V. Lempitsky Domain-Adversarial Training of Neural Networks // Journal of machine learning research. — 2015.
7. Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, Luc Van Gool Domain Adaptive Faster R-CNN for Object Detection in the Wild // Computer Vision and Pattern Recognition Conference 2018. — 2018.
8. Jianfeng Dong, Zhongzi Long, Xiaofeng Mao, Changting Lin, Yuan He, Shouling Ji Multi-level Alignment Network for Domain Adaptive Cross-modal Retrieval // Neurocomputing. — 2021. — №440. — P. 207–219.
9. Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, Kate Saenko Strong-Weak Distribution Alignment for Adaptive Object Detection // Computer Vision, and Pattern Recognition Conference. — 2019.
10. Han-Kai Hsu, Chun-Han Yao, Yi-Hsuan Tsai, Wei-Chih Hung, Hung-Yu Tseng, Maneesh Singh, Ming-Hsuan Yang Progressive Domain Adaptation for Object Detection // 2020 IEEE Winter Conference on Applications of Computer Vision. — 2020
11. Giulio Mattolin, Luca Zanella, Elisa Ricci, Yiming Wang ConfMix: Unsupervised Domain Adaptation for Object Detection via Confidence-based Mixing // IEEE Workshop/Winter Conference on Applications of Computer Vision. — 2022.
12. Mohamed L. Mekhalfi, Davide Boscaini, Fabio Poiesi Detect, Augment, Compose, and Adapt: Four Steps for Unsupervised Domain Adaptation in Object Detection // British Machine Vision Conference. — 2023.
13. Hui Li, X. Wang, Shifei Ding Research, and development of neural network ensembles: a survey // Artificial Intelligence. — 2018. — №49 (2).
14. Matias Valdenegro Sub-Ensembles for Fast Uncertainty Estimation in Neural Networks // International Conference on Computer Vision Workshops. — 2023.
15. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi You Only Look Once: Unified, Real-Time Object Detection // IEEE Conference on Computer Vision and Pattern Recognition. — 2016.

© Медведев Анатолий Андреевич (medvedev.workspace@gmail.com)  
Журнал «Современная наука: актуальные проблемы теории и практики»