

Сравнение современных методов глубокого обучения для автоматической детекции белух в естественной среде

Булкин Антон Павлович

2025

Аннотация

В работе исследуются современные архитектуры глубокого обучения для автоматического мониторинга белух — редкого и охраняемого вида, требующего внимательного и регулярного наблюдения. Цель исследования — повысить эффективность мониторинга животных, минимизируя затраты времени и ресурсов на ручную обработку данных, а также создания системы отслеживания взаимодействия животных между собой в рамках одного видеофайла. Предложен комплексный подход: классические решения CV, полуавтоматическая разметка видеоматериалов, обучение и сравнение современных детекторов и трекеров, а также анализ их точности и скорости.

1 Введение

Белуха — редкий и охраняемый вид морских млекопитающих, для которого необходимы регулярные и объективные наблюдения в естественной среде обитания. Традиционные подходы — полевые экспедиции, акустический мониторинг, визуальный подсчёт — трудоёмки, подвержены субъективности и ограничены по пространственному и временному охвату. Автоматизация на основе методов компьютерного зрения и глубокого обучения позволяет снизить затраты, повысить воспроизводимость и масштабировать мониторинг.

Задача детекции морских млекопитающих в изображениях и видео обычно решается локализацией объектов с ограничивающими рамками и последующим трекингом. Ранние работы использовали классические техники и первые версии одноэтапных и двухэтапных детекторов, показывая базовую применимость, но чувствительность к условиям съёмки и фоновым помехам. Для видео широко применялись связки «детектор + трекер», повышающие устойчивость за счёт межкадровой ассоциации. Современные одноэтапные архитектуры семейства YOLO существенно улучшили соотношение точности и скорости и пригодны для сценариев реального времени. Трансформерные детекторы упрощают инференс, снимая необходимость в постобработке наподобие подавления немаксимумов, и повышают согласованность предсказаний на сложном фоне. Отдельное направление посвящено снижению стоимости аннотирования за счёт полуавтоматической разметки и стратегий активного обучения.

Ключевые трудности домена — малый видимый размер животных в кадре, блики и рябь воды, частичные окклюзии и близость визуально похожих фоновых объектов (лодки, буи, скалы, водоросли). Существенны доменный сдвиг между локациями и условиями съёмки и дисбаланс классов в многоклассовых сценариях. Высокая стоимость

ручной аннотации приводит к ограниченным наборам и ошибкам разметки, что снижает переносимость моделей. Наконец, несогласованные протоколы оценки в литературе — разные наборы данных, пороги пересечения, препроцессинг — затрудняют прямое сравнение результатов и выбор решений для практики.

В работе предлагается воспроизводимый бенчмарк детекции белух в реальной морской среде с унифицированным протоколом обучения и оценки. Подготовлены и сопоставлены наборы изображений разного масштаба — 400, 800, 1200 и около 8 000 кадров, — где расширение корпуса достигается полуавтоматической разметкой: начальное обучение модели-помощника, автоматическая аннотация неразмеченных кадров и ручная валидация. Рассматриваются одноклассовая и многоклассовая разметки, что позволяет учитывать типичные фоновые объекты на воде.

2 Обзор литературы / Related works

Исследования, посвящённые автоматическому мониторингу белух и других морских млекопитающих, активно развиваются в последние годы и охватывают как аэровидео съёмку, так и анализ спутниковых данных. Одной из наиболее значимых современных работ является статья Alsaidi *et al.* [1], где предложена система детекции и трекинга белух на аэровидео с использованием YOLOv7 и алгоритма DeepSORT. Модель показала высокую точность и полноту, а также устойчивость трекинга после постобработки. Аналогичные идеи развиваются в работе Harasyn *et al.* [2], где YOLOv4 и DeepSORT применялись для детекции белух, каяков и лодок в видеопотоках с дронов; достигнута точность около 74% и полнота 72%. Lee *et al.* [3] исследовали RetinaNet и Faster R-CNN на данных залива Камберленд, отметив, что использование тайлинга крупных изображений существенно повышает точность и уменьшает количество ложных срабатываний.

Особое внимание уделяется построению и масштабированию наборов данных. Araújo *et al.* [4] представили датасет *Beluga-5k* с более чем 5500 фотографиями белух и предложили полуавтоматическую разметку. Лучшая модель (YOLOv3-tiny) достигла 97,05% mAP@0.5, корректно обнаруживая белух даже в условиях сложного фона. Boulent *et al.* [5] предложили интерактивную схему «человек в контуре», при которой нейросеть, обученная на 100 изображениях, достигла 91% совпадения с экспертами и позволила ускорить аннотацию более чем в пять раз. Cubaynes и Fretwell [6] опубликовали набор спутниковых изображений *Whales from Space*, включающий сотни размеченных примеров китов, что обеспечило возможность обучения моделей на данных высокого разрешения.

Использование спутниковых снимков для мониторинга млекопитающих подтверждается рядом работ. Green *et al.* [7] применили YOLOv5 для обнаружения серых китов на спутниковых снимках и достигли точности 80–94% при полноте 84–89%. Guirado *et al.* [8] предложили каскадный подход, объединяющий классификацию наличия китов и подсчёт особей, что увеличило общую точность на 36% по сравнению с одноэтапными методами. Bogowicz *et al.* [9] показали, что CNN, обученные на аэрофотоснимках, могут быть перенесены на спутниковые данные без значительной потери качества.

В области общих архитектур объектной детекции значительную роль сыграли двухэтапные модели (Faster R-CNN [10]), обеспечившие высокую точность за счёт сети предложений регионов. Одноэтапные детекторы, начиная с YOLO [11] и YOLO9000 [12], позволили объединить локализацию и классификацию в одном прогоне сети, достигнув 45–60 FPS. RetinaNet [13] с функцией Focal Loss улучшил устойчивость к дисбалансу классов и повысил точность при сохранении скорости. Последующие версии YOLO (в частности, YOLOv7 [14]) установили новый стандарт по соотношению точности и производительности.

Трансформерные подходы (DETR [15], RT-DETR [16]) позволили отказаться от постобработки (NMS) и выполнять end-to-end детекцию. RT-DETR показал сопоставимую с YOLO точность при сохранении real-time скорости, что делает его перспективным для видеоаналитики. Модели открытого словаря (ViLD [17], GLIP [18], YOLO-World [19]) расширили возможности детекции на неизвестные классы, включая морские объекты, что особенно важно для практических систем, где структура сцены заранее не фиксирована.

Обзор Tuia *et al.* [20] подчёркивает важность машинного обучения для экологического мониторинга. Авторы указывают, что интеграция глубоких моделей с полевыми данными и дронами системами позволяет существенно сократить стоимость и повысить масштабируемость наблюдений. Совокупно, эти исследования формируют основу для разработки универсального бенчмарка детекции белух и показывают актуальность сопоставления современных архитектур на единых данных и метриках.

Список литературы

- [1] M. Alsaïdi et al. Localization and tracking of beluga whales in aerial video using deep learning. *Frontiers in Marine Science*, 2024.
- [2] Madison L. Harasyn, Wayne S. Chan, Emma L. Ausen, and David G. Barber. Detection and tracking of belugas, kayaks and motorized boats in drone video using deep learning. *Journal of Unmanned Vehicle Systems*, 2022.
- [3] P. Q. Lee et al. Beluga whale detection in the cumberland sound bay using convolutional neural networks. *Canadian Journal of Remote Sensing*, 2021.
- [4] Voncarlos M. Araújo, Ankita Shukla, Clément Chion, Sébastien Gambs, and Robert Michaud. Machine-learning approach for automatic detection of wild beluga whales from hand-held camera pictures. *Sensors*, 22(11):4107, 2022.
- [5] J. Boulent et al. Scaling whale monitoring using deep learning: A human-in-the-loop solution for analyzing aerial datasets. *Frontiers in Marine Science*, 2023.
- [6] H. C. Cubaynes and P. T. Fretwell. Whales from space dataset: an annotated satellite image dataset of whales for training machine learning models. *Scientific Data*, 2022.
- [7] K. M. Green et al. Gray whale detection in satellite imagery using deep learning. *Remote Sensing in Ecology and Conservation*, 2023.
- [8] E. Guirado et al. Whale counting in satellite and aerial images with deep learning. *Scientific Reports*, 2019.
- [9] A. Borowicz et al. Aerial-trained deep learning networks for surveying cetaceans from satellite imagery. *PLOS One*, 2019.
- [10] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *NeurIPS*, 2015.
- [11] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *CVPR*, 2016.
- [12] Joseph Redmon and Ali Farhadi. Yolo9000: Better, faster, stronger. In *CVPR*, 2017.

- [13] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *ICCV*, 2017.
- [14] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*, 2022.
- [15] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *ECCV*, 2020.
- [16] Yian Zhao et al. Detrs beat yolos on real-time object detection (rt-detr). *arXiv preprint arXiv:2304.08069*, 2023.
- [17] X. Gu et al. Open-vocabulary object detection via vision-and-language knowledge distillation (vild). In *CVPR*, 2021.
- [18] L. H. Li et al. Grounded language-image pre-training (glip). In *CVPR*, 2022.
- [19] T. Cheng et al. YOLO-world: Real-time open-vocabulary object detection. In *CVPR*, 2024.
- [20] D. Tuia et al. Perspectives in machine learning for wildlife conservation. *Nature Communications*, 2022.