

Автоматическая предварительная разметка судовых снимков в задаче обнаружения плавающего морского мусора

И. А. Герасимов¹, М. А. Борисов¹, М. П. Погожева^{2,3,4}, В. А. Спирина³, М. А. Криницкий^{1,4}

1 — Московский физико-технический институт (национальный исследовательский институт)

2 — МГУ им. М. В. Ломоносова, географический факультет

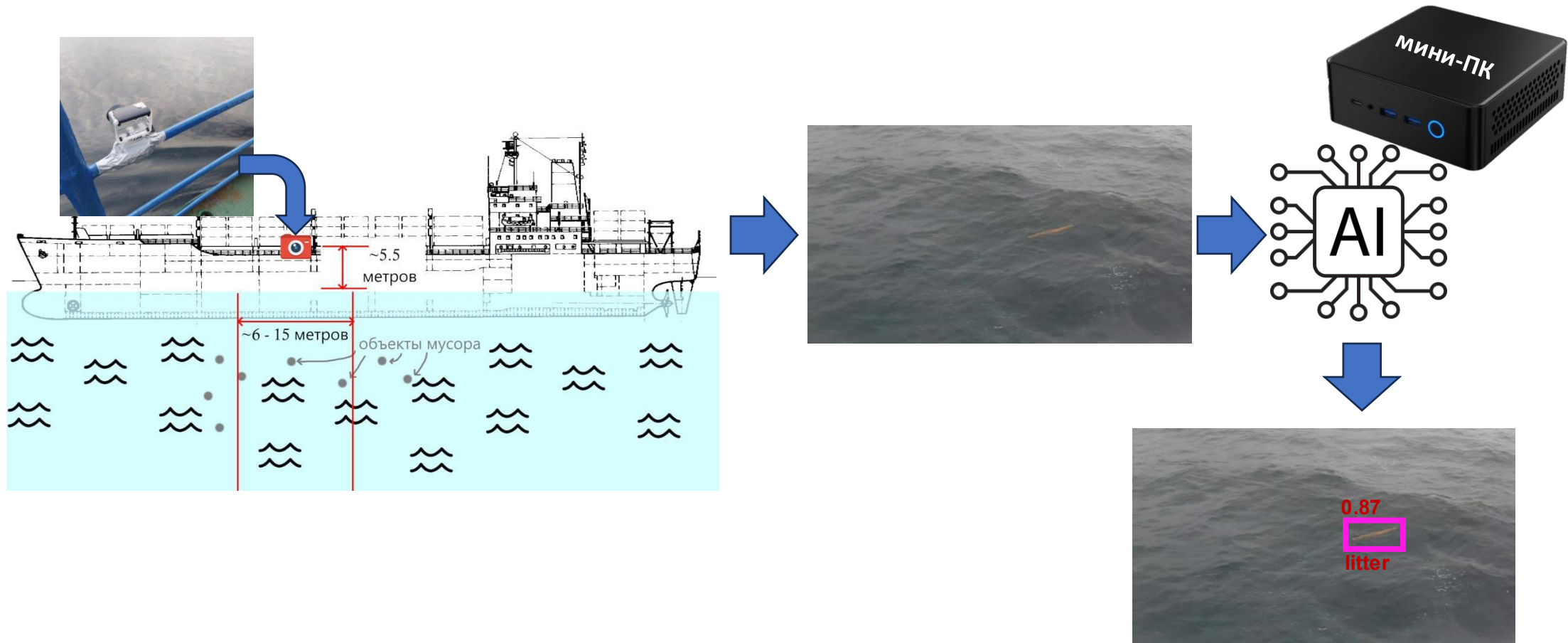
3 — Государственный океанографический институт им. Н. Н. Зубова

4 — Институт океанологии им. П. П. Ширшова РАН

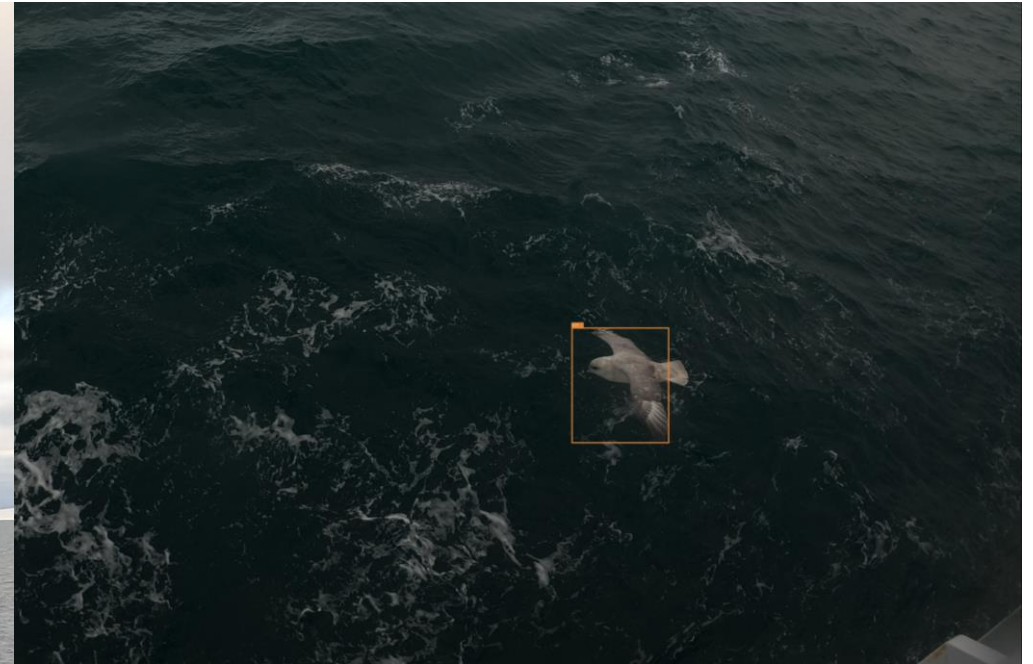
Актуальность

- Судовой фотомониторинг морского мусора даёт большие объёмы кадров высокого разрешения, которые сейчас просматриваются вручную.
- Целевые объекты — фрагменты пластика, плавающая древесина — занимают малую долю кадра; подавляющее большинство кадров не содержит целевых объектов.
- Обучение детекторов с учителем требует размеченных выборок, формирование которых затруднено из-за вариативности условий съёмки и редкости объектов.
- Задача: автоматически отсеять пустые кадры и снабдить непустые предварительной разметкой, снизив нагрузку на экспертов.

Целевой дизайн системы



Актуальность



Актуальность



Цель: Разработать автоматическую систему предварительной разметки судовых снимков, снижающую нагрузку ручного просмотра и формирующую основу для обучающей выборки.

Задачи:

- Выбор базовой модели детекции по текстовым описаниям (zero-shot), без дообучения на домене.
- Тайловый инференс для чувствительности к малоразмерным объектам на 4К-кадрах.
- Настройка фильтров для подавления ложных срабатываний.
- Автоматическое разделение кадров на «пустые» и «с объектами»;
- Анализ результатов и подготовка к формированию обучающей выборки.

Данные и постановка задачи

- Судовая фотосъёмка морской поверхности; кадры высокого разрешения $\approx 4K$ (3840×2160).
- Съёмка по бортам (ЛБ / ПБ) и датам; целевые классы: пластик, древесина, морские птицы.
- Выборка настоящего отчёта: 169 380 кадров, обработанных системой.
- **Особенности задачи:**
 - малый объект в большом кадре — медиана площади бокса ≈ 0.004 % кадра;
 - редкость: целевой объект присутствует лишь в ≈ 21 % кадров выборки;
 - вариативность фона: блики, пена, гребни волн, элементы судна $\rightarrow$ ложные срабатывания.

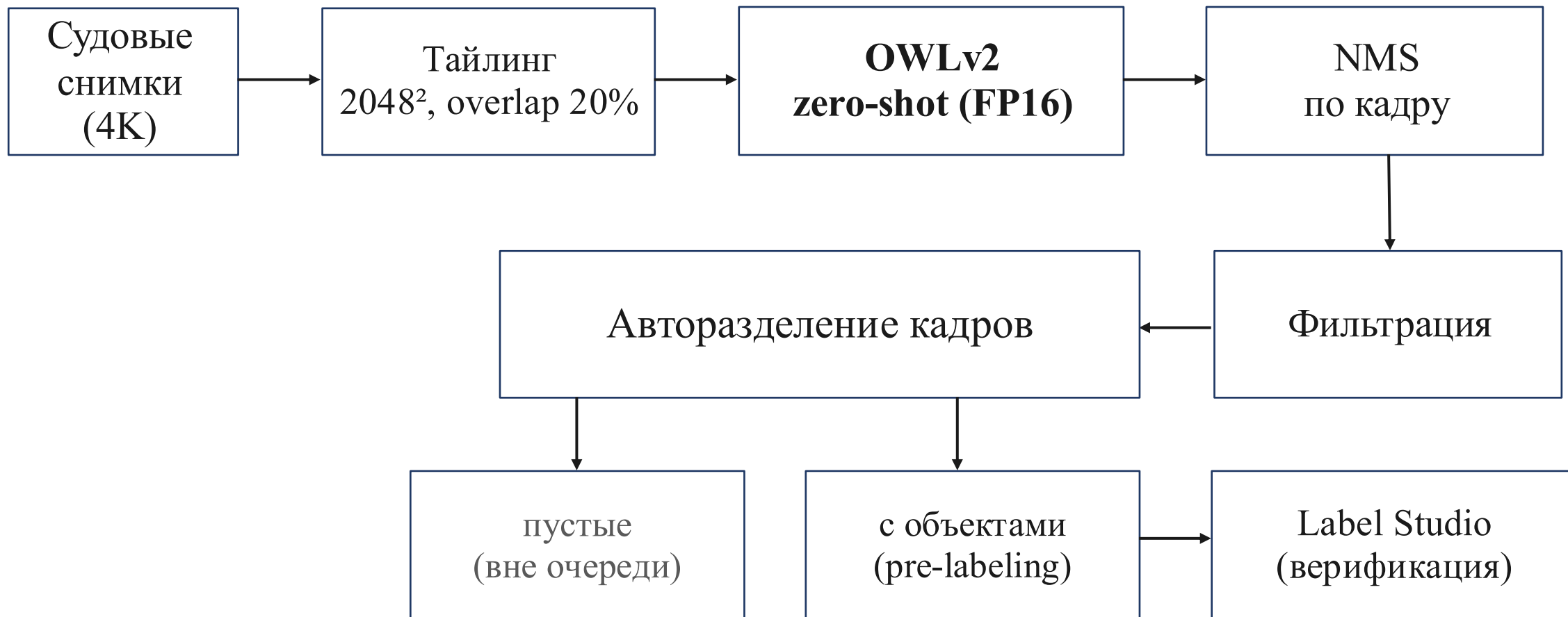
Выбор базовой модели

Пилотное сравнение zero-shot детекторов на валидационной выборке (30 «пустых» + 8 «с объектом» кадров, инференс по целому кадру, без тайлинга и фильтров):

Модель	Промпты / классы	Полнота (объекты, n=8)	Ложные (пустые, n=30)	Итог
Florence-2 large	текстовый промпт	0 / 8	0 / 30	не обнаруживает
YOLO-World (X)	4 класса	1 / 8	0 / 30	низкая полнота
OWLv2 large-ensemble	5 дескрипторов	5 / 8	13 / 30	выбрана

Вывод. OWLv2 даёт наибольшую полноту (5 из 8) — критично при редких объектах; её ложные срабатывания далее подавляются per-class порогами, тайлингом и каскадом фильтров. Florence-2 не обнаружил ни одного объекта, YOLO-World — лишь один.

Схема системы



Zero-shot детекция OWLv2

- Open-vocabulary детектор: находит объекты по текстовым промптам без дообучения на целевом домене.
- Модель: *google/owlv2-large-patch14-ensemble*; инференс в FP16.
- **Целевые классы (промпты):**
 - brightly colored plastic object
 - wooden log floating on water
 - seabird sitting on water · seabird flying over sea

Фильтры и авторазделение

- **Фильтры (последовательно):**
 - порог уверенности по классу;
 - макс. площадь бокса ($> 10\%$ кадра — галлюцинация);
 - макс. размер бокса по классу.
- Результат → формат Label Studio (координаты в %).
- **Авторазделение:**
 - пустые → empty/;
 - непустые → pre-labeling/;
 - пустые исключаются из очереди просмотра.
 - Импорт непустых в Label Studio (SDK, батчи).

Пороги уверенности по классам

класс	порог
plastic object	0.10
wooden log	0.20
seabird sitting	0.20
seabird flying	0.20

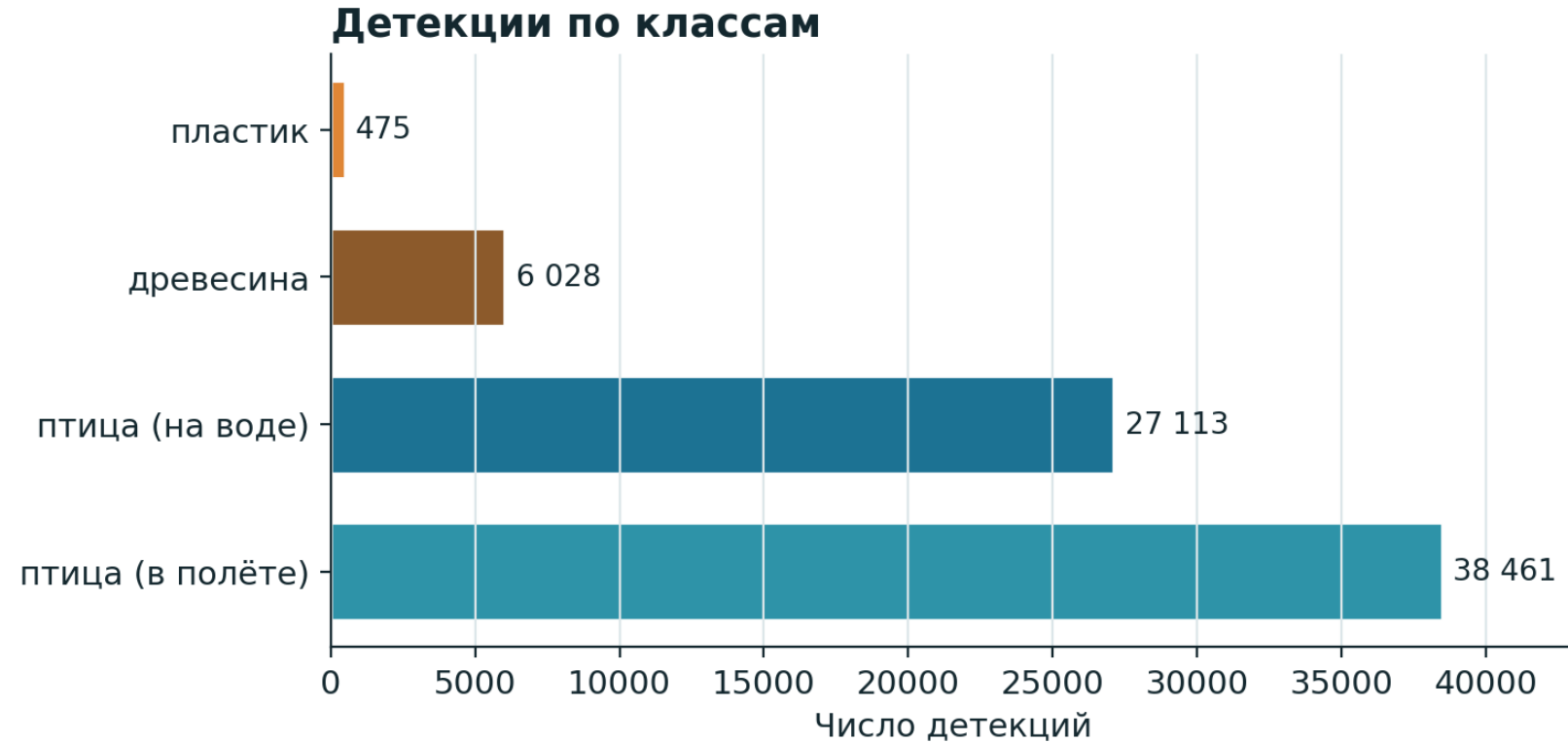
Результаты: разделение выборки



Выборка 169 380 кадров. «Пустые» включают кадры без целевого объекта.

- Из 169 380 кадров выборки только $\approx 20.7\%$ содержат целевой объект — остальные пусты.
- Авторазделение снимает эти $\approx 79\%$ с очереди ручного просмотра, фокусируя экспертов на непустых кадрах.

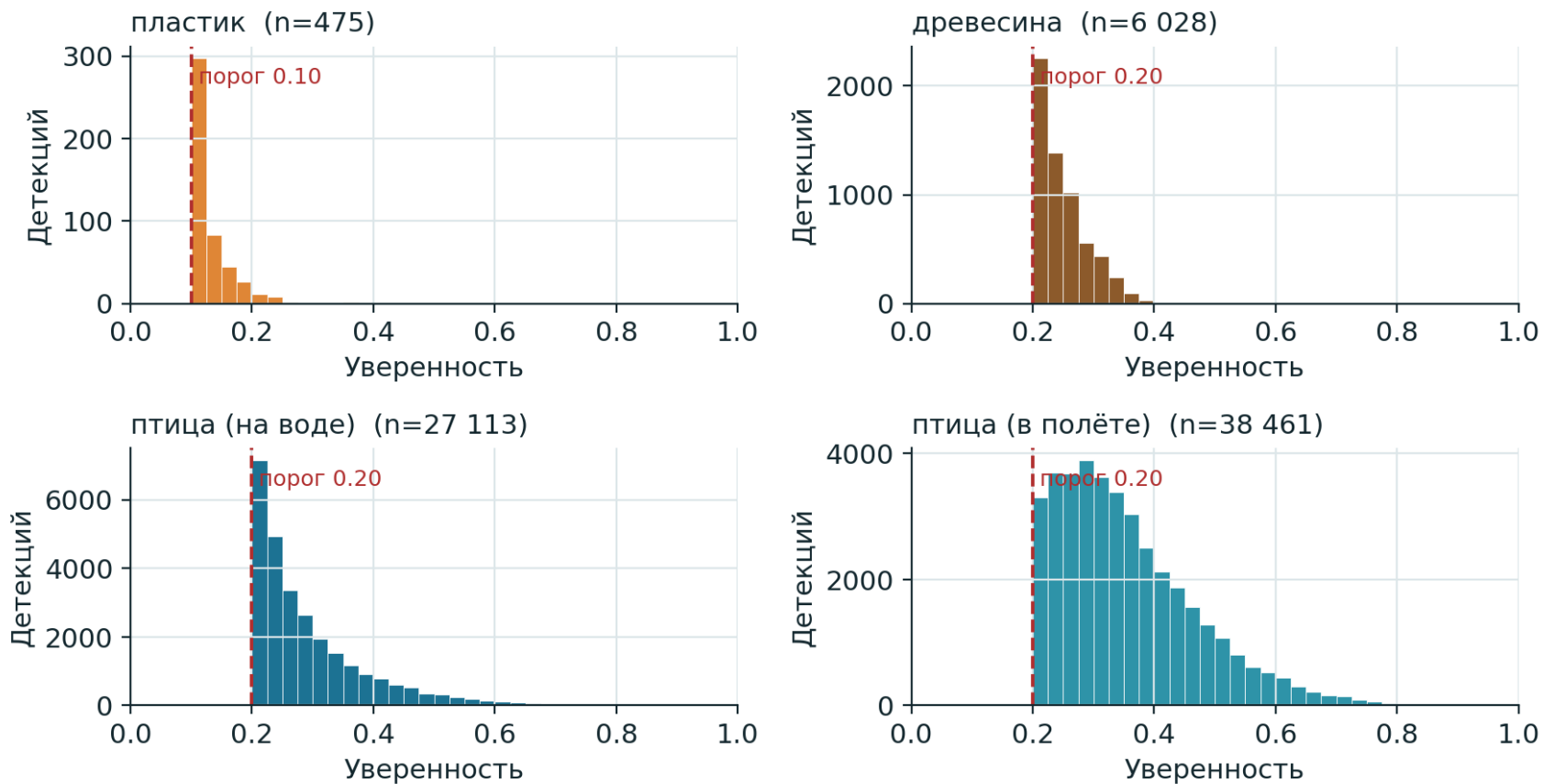
Результаты: детекции по классам



- Всего 72 077 целевых детекций.
- Доминируют птицы (в полёте 38 461, на воде 27 113).
- Пластик редок (475) и с низкой уверенностью.
- «Древесина» (6 028) сильно завышена артефактом (слайд 14).

Результаты: распределение уверенности

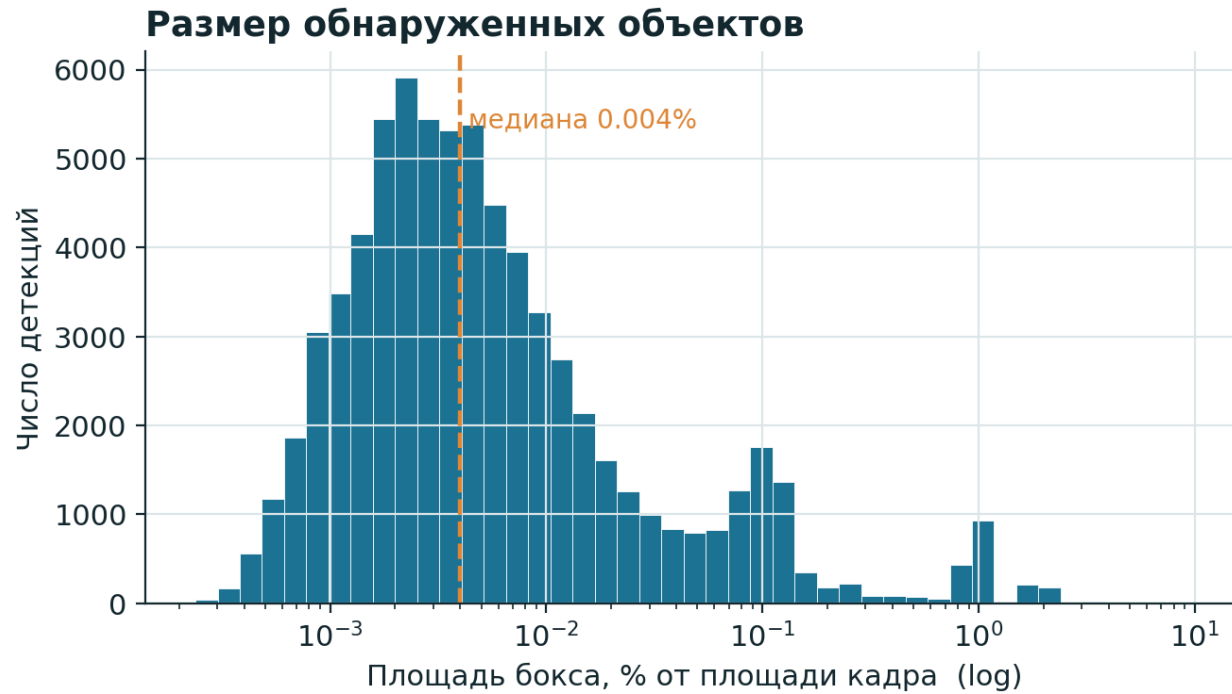
Распределение уверенности по классам



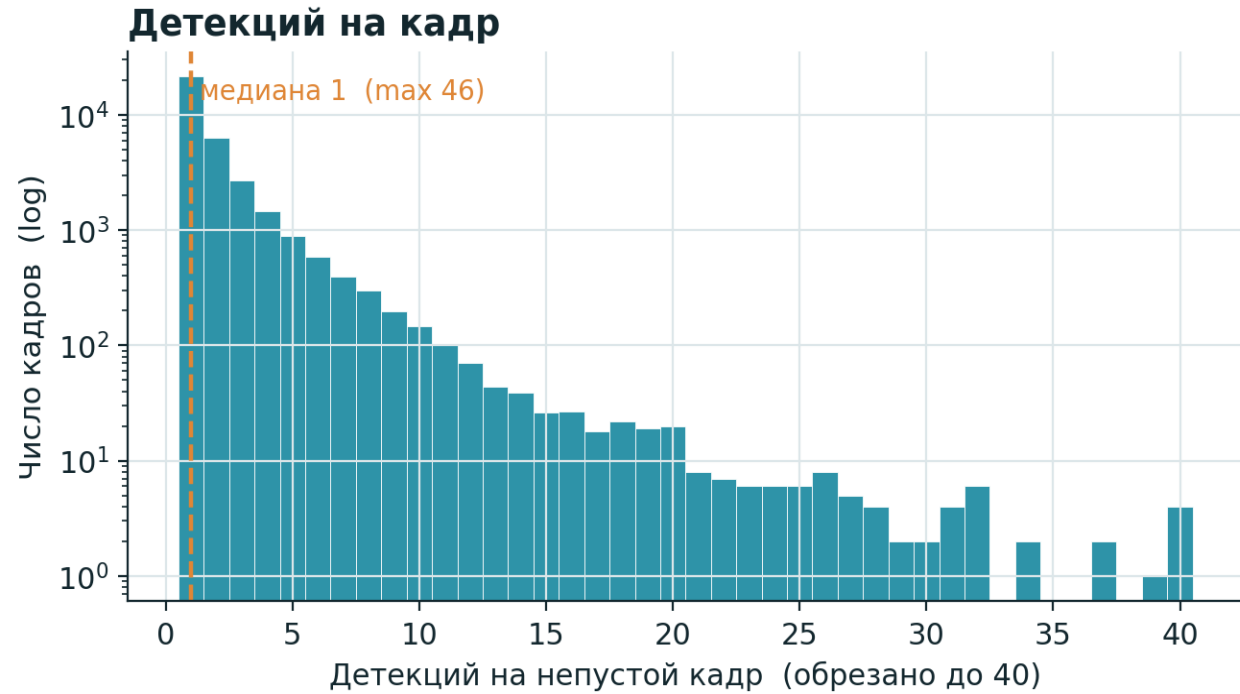
Гистограммы уверенности по классам; штриховая линия — per-class порог.

*Масса детекций пластика имеет явный пик 0.10 → преимущественно пограничные срабатывания;
у птиц распределение шире и увереннее.*

Результаты: размер объектов и плотность

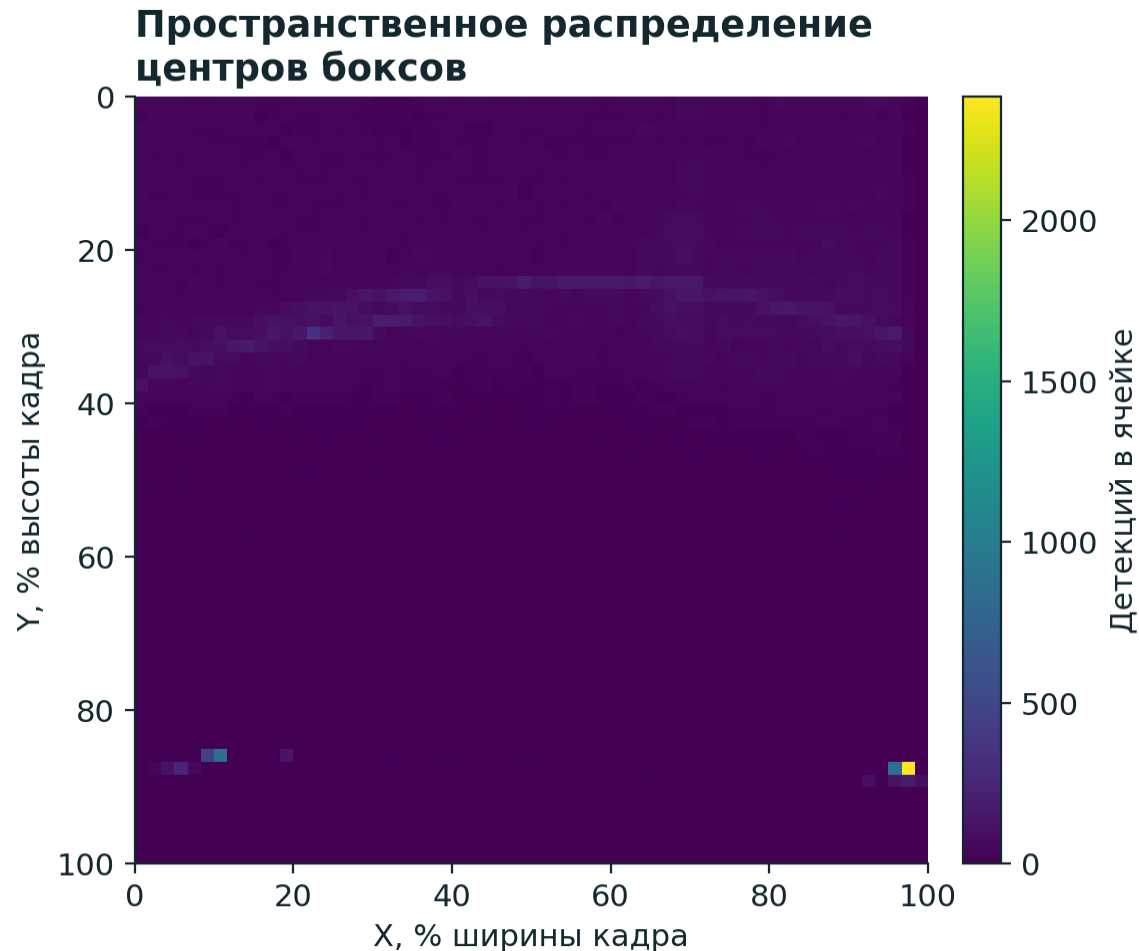


Медиана площади бокса ≈ 0.004 % кадра.



Медиана 1 объект/кадр, редкие «стайи» до 46.

Результаты:пространственный анализ



Центры боксов по кадру (все детекции).

- Дуга у $Y \approx 25\text{--}35\%$ — линия горизонта:
- Резкие пики в нижних углах кадра — фиксированный элемент судна.
- Угол $X \approx 97\%$: 2 380 срабатываний как «wooden log»;
- В нижних углах $\approx 4\,700$ из 6 028 «wooden log».

Выводы

- Реализована автоматическая система пред-разметки на базе zero-shot **OWLv2** и тайлового инференса, без дообучения на целевом домене.
- На пилотной валидации OWLv2 превзошла Florence-2 и YOLO-World по полноте и выбрана базовой моделью.
- Авторазделение исключает ≈ 79 % пустых кадров выборки из очереди ручного просмотра.
- Анализ результатов выявил систематический ложный класс «древесина»
- Перспективы: экспертная верификация → обучающая выборка → специализированная модель → повторная разметка.

Спасибо за внимание

И. А. Герасимов, М. А. Борисов, М. П. Погожева, В. А. Спирина, М. А. Криницкий