

# Разбор обзорной статьи WaRP и реализация бейзлайнов

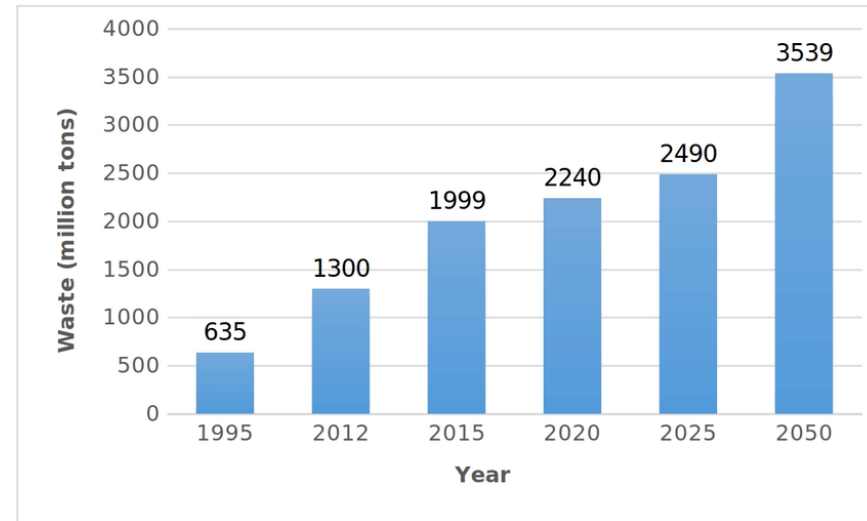
Engineering Applications of Artificial Intelligence 128 (2024) 107542

Докладчик: Семенова Юлия

# Описание задачи

## Актуальность

- Уменьшение загрязнения окружающей среды бытовыми отходами — особенно важная на сегодняшний день задача;
- Используется автоматическая сортировка мусора: конвейеры снабжены камерами, распознающими нужные категории отходов. Для разработки таких систем необходимы алгоритмы, способные осуществлять детекцию, классификацию и сегментацию объектов для дальнейшего их захвата и отправления на переработку.



*Ежегодное количество отходов в мире  
Tran, Binh-Giang & Nguyen, Duy-Linh.  
(2023). Simple and Efficient Convolutional  
Neural Network for Trash Classification.  
10.15439/2022R01.*

# Описание задачи

## Проблема

- Для решения подобных задач лучше всего подходят глубокие нейронные сети;
- Проблема — отсутствие подходящего датасета для обучения моделей детекции, сегментации и классификации отходов на перерабатывающем предприятии.
- Главной задачей авторов статьи было создание нового большого датасета из снимков, сделанных на перерабатывающем предприятии, и исследование современных методов глубокого обучения для решения задачи распознавания ценных фракций отходов.

# Проблемы в сфере распознавания ценных фракций отходов

Основные проблемы, встречающиеся в сфере распознавания отходов:

- Многие модели хорошо работают только с малым количеством категорий отходов и со снимками хорошего качества;
- Естественный дисбаланс классов, датасеты с одинаковым фоном и маленькие размеры изображений;
- В некоторых алгоритмах предполагается использование дополнительных сенсоров, помимо камеры, что трудно осуществить на практике;

# Проблемы в сфере распознавания ценных фракций отходов

- Было предложено использование сверточных сетей с Graph long short-term memory, однако для их обучения нет датасетов в открытом доступе.
- В исследованиях, посвященных использованию моделей семейства YOLO для детекции отходов, отмечается высокая скорость но значительное влияние качества обучающего датасета результат обучения.
- Сверточные нейронные сети или сети-трансформеры для сегментации нуждаются в датасетах с масками объектов: создавать их вручную — очень трудоемкая задача.

# Существующие датасеты



UAVWaste(Kraft et al., 2021)



TrashNet (Thung and Yang, 2017)



TACO (Proença and Simões, 2020)

# Существующие датасеты, плюсы и минусы

- TrashNet (Thung and Yang, 2017)
  - Объекты на нейтральном фоне;
  - Всего 6 категорий;
  - + Один из самых первых датасетов для классификации отходов
- UAVWaste (Kraft et al., 2021)
  - Объекты в природной среде;
  - Нет разбиения по категориям отходов (подходит только для задачи сегментации);
  - + Подойдет для обучения робота-уборщика;
- TACO (Proença and Simões, 2020)
  - Объекты в природной среде;
  - Относительно мало обработанных изображений;
  - + Содержит вручную созданные маски объектов для обучения модели сегментации;
  - + Много категорий (28) и подкатегорий (60)

# Существующие датасеты, плюсы и минусы



Выделяется ZeroWaste Dataset: в нем представлены изображения конвейерной линии на заводе по переработке бумаги.

Однако он специализирован для выделения бумажных отходов.

Нет открытых датасетов, содержащих все необходимые категории объектов: пластиковые бутылки, картонные коробки, банки и другие, с разделением некоторых категорий на подклассы.



# WaRP(Waste Recycling Plant) dataset



- Содержит вручную подписанные изображения промышленного конвейера.
- Выделено 28 категорий перерабатываемого мусора.
- Объекты могут пересекаться, быть поврежденными или плохо освещенными.
- Датасет несбалансированный!
- Три части: WaRP-D, WaRP-C, WaRP-S (первые два для обучения и оценки качества детекции и классификации, последний для валидации методов сегментации, содержит бинарные маски распознанных объектов).

# WaRP(Waste Recycling Plant) dataset



WaRP-D



bottle-oil



detergent-box

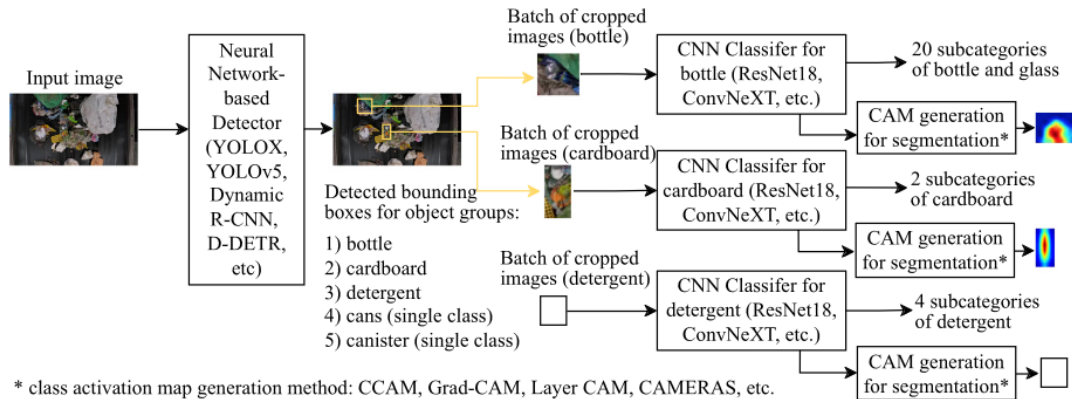
WaRP-C



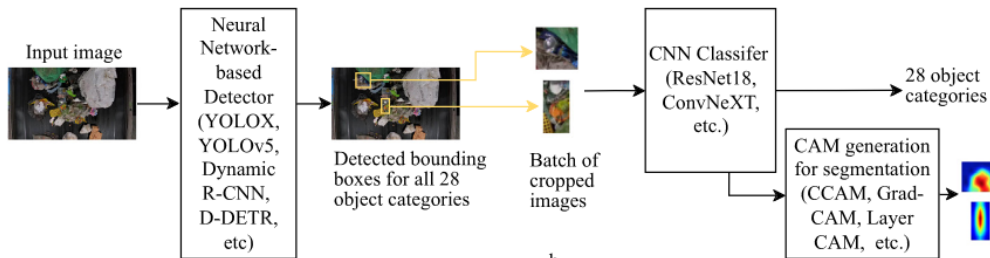
WaRP-S

- WaRP-D: 2452 изображений - training sample и 522 изображений validation sample, full HD.
- WaRP-C: вырезанные изображения из WaRP-D с метками классов. Несбалансированный из-за естественного неравномерного распределения.
- WaRP-S: всего 112 изображений, в каждой категории есть по 4 сильно поврежденных объекта

# Нейронная сеть для иерархической (hierarchical) детекции



a



b

Идея: добавить в предобученную модель детекции дополнительные модули классификации и weakly-supervised сегментации.

В статье рассматриваются два метода иерархического классификатора.

Варианты реализации Н-УС архитектуры (a) Н-УС(5) (b) more simple Н-УС(28)

# Нейронная сеть для иерархической (hierarchical) детекции

## *1 метод: Н-УС(5)*

- Детектор определяет bounding boxes для 5 «суперкатегорий» (бутылки, картонные коробки, упаковки от моющих средств, банки и канистры);
- Независимые классификаторы обучены распознавать подкатегории в первых трех категориях.

## *2 метод: Н-УС(28)*

- Детектор распознает все 28 категорий, классификатор улучшает показания детектора;
- Более индустриальное решение: важна модулярность.

# Нейронная сеть для иерархической (hierarchical) детекции

Исследуется 2 способа обучения иерархического детектора:

- 1) Детектор и классификатор учатся независимо, детектор — на WaRP-D, а классификатор на WaRP-C;
- 2) Классификатор учится на предсказанных детектором вырезках из train набора WaRP-D, категории определяются по пресечению этих вырезок с помеченными GT bounding boxes.

# Нейронная сеть для иерархической (hierarchical) детекции

Для weakly supervised segmentation исследуются популярные CAM (Class Activation Map) методы создания карт активации; для прехода от карт активации к маскам сегментации, используется алгоритм, предложенный авторами в другой статье. Модель сегментации обучается без Ground Truth feature masks.

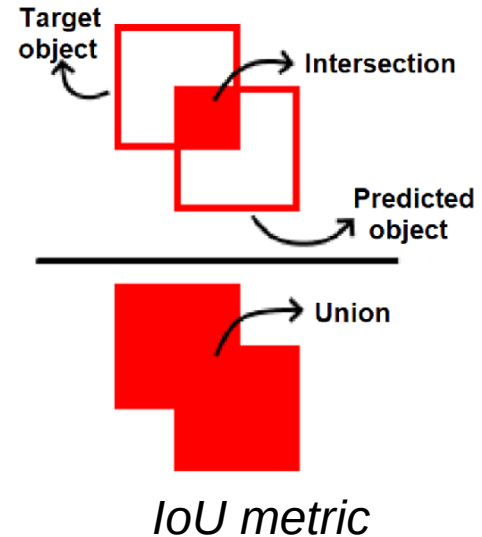
# Используемые метрики для детекции

Сравнивая различные модели для задачи детекции, авторы сравнивали метрики  $mAP_{50}$  ( $AP_{50}$  для категорий, не содержащих подкатегории; и  $mAP_{50}$  по всем классам) и  $mAP_{50..95}$  (для всех классов).

$mAP_{50}$ : средняя точность (mean precision), рассчитанная при пороге intersection over union (IoU), равном 0,50.

$mAP_{50-95}$ : среднее значение средней точности, рассчитанное при различных пороговых значениях IoU, варьирующихся от 0,50 до 0,95.

Mean precision: среднее значение площади под pr-кривой для всех классов.



# Используемые метрики для детекции

$mAP_{50}$  - наиболее универсальная метрика для задачи детекции, отражает правильность угаданных классов.

$mAP_{50-95}$  также отражает, насколько точно удалось выделить bounding boxes.

Также оценивался FPS (Frame Per Second) моделей, поскольку они должны работать в реальном времени.

Лучшими по  $mAP$  метрике были выбраны TOOD (60.2%  $mAP$ ) и YOLOX-m (58.6%  $mAP$ ).

Самые быстрые модели: YOLOV5-x (66.6 FPS) и YOLOX-m (64.9 FPS)

В качестве основного детектора выбрана YOLOX-m.



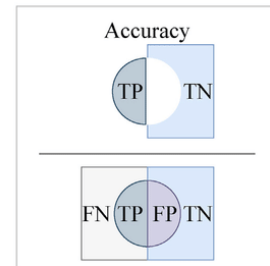
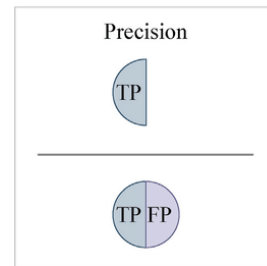
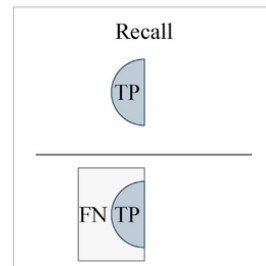
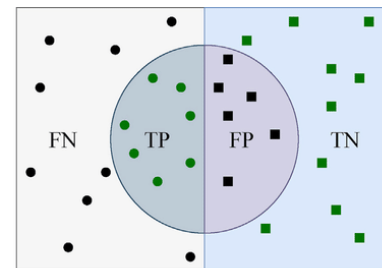
# Используемые метрики для классификации

Для моделей классификации были рассмотрены метрики recall и mean recall (для категорий, содержащих несколько подкатегорий), а также считалось accuracy.

Ассурасу — наиболее универсальная метрика для задач классификации, когда нет цели значительно понизить FP или FN.

В качестве классификатора была выбрана архитектура ConvNeXt, показавшая accuracy 81.8% и performance 48FPS.

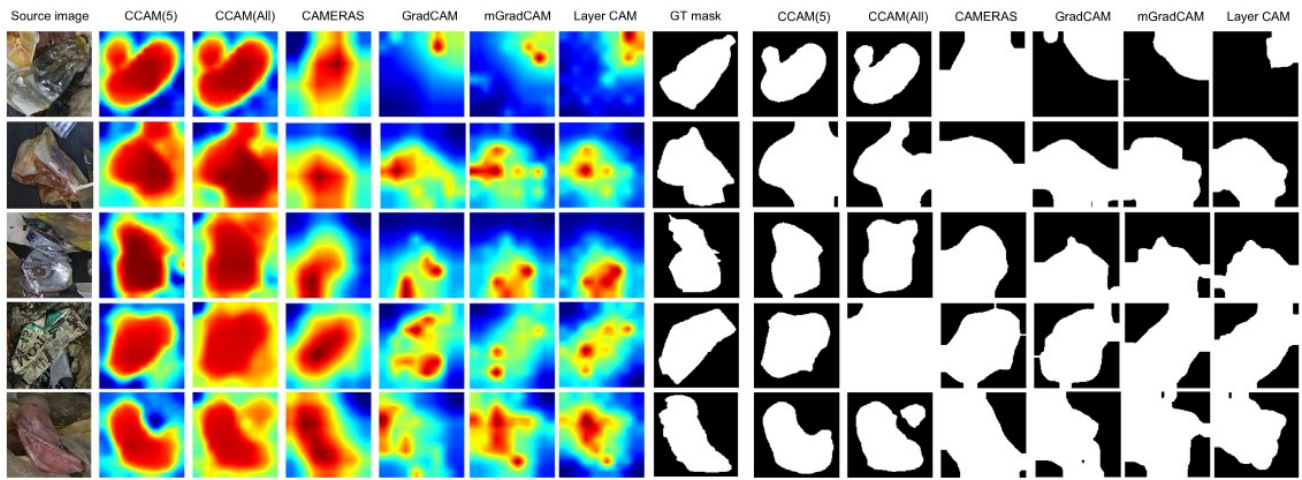
Получили hierarchical neural networks H-YS с YOLOX-m детектором и ConvNeXt классификатором.



# Используемые метрики для weakly supervised сегментации

Для решения задачи сегментации использовались CAM (Class Activation Map) методы. Для проверки предсказаний использовалась ResNet-18.

Для оценки качества сегментации использовалась mIoU (mean Intersection over Union) метрика: она хорошо подходит, поскольку нет необходимости особенно четко выделить границы или мелкие детали.



*Сгенерированные карты активации и построенные по ним бинарные маски*

# Оценка полученной модели

Приведена также метрика F1-score для полученной модели на основании трех экспериментов, проведенных в разные дни. Получены следующие результаты:

63% для упаковок от моющих средств, 73% для картонных коробок, 79% для бутылок и 81% для банок (канистры замечены не были).

$$\text{F1 Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

F1-score используется, как наиболее универсальная метрика, когда одинаково важны precision и recall.

# Заключение

Составленный WaRP датасет является наиболее полным из существующих открытых датасетов для задачи распознавания ценных фракций отходов. Его главным преимуществом является то, что изображения сделаны непосредственно на перерабатывающем предприятии и содержит снимки в плохом освещении, с пересекающимися и сильно поврежденными предметами.

Минусом стало то, что снимки в датасете сняты с одного ракурса. В контрольном эксперименте модель тестировалась на снимках с камеры, расположенной по-другому, фон и освещение предметов были другими. Тем не менее, модель показала хорошее качество детекции и классификации, что говорит о хорошей способности к обобщению (Generalization ability) модели.

# Заключение

Модель Н-УС(28) показала лучший результат по метрикам  $mAP_{50}$  и  $mAP_{50..95}$ , чем отдельный детектор YOLOX-m.

Модель Н-УС(5), напротив, не улучшила метрики; независимое обучение детектора на 5 категориях и трех классификаторов для различных подкатегорий показало результат хуже, чем отдельный детектор YOLOX-m. Предложенный иерархический подход к детекции отходов помог улучшить качество предобученных моделей, а также позволил производить weakly supervised сегментацию с приемлемым качеством. Метод weakly supervised сегментации позволил сэкономить ресурсы на создании feature masks.

# Повторяемость результатов

[https://github.com/Iluciola/WaRP\\_project/](https://github.com/Iluciola/WaRP_project/)

# ИСТОЧНИКИ

- <https://www.sciencedirect.com/science/article/pii/S0952197623017268>
- [https://www.researchgate.net/figure/The-chart-of-the-annual-amount-of-trash-on-the-globe\\_fig1\\_369299857](https://www.researchgate.net/figure/The-chart-of-the-annual-amount-of-trash-on-the-globe_fig1_369299857)
- <https://www.mdpi.com/2072-4292/13/5/965>
- [https://www.researchgate.net/figure/Illustration-of-TrashNet-data-from-different-classes\\_fig2\\_343841421](https://www.researchgate.net/figure/Illustration-of-TrashNet-data-from-different-classes_fig2_343841421)
- <https://encord.com/blog/taco-dataset-guide/>
- <https://github.com/AIRI-Institute/WaRP>
- <https://github.com/Megvii-BaseDetection/YOLOX>
- <https://github.com/fcjian/TOOD>
- <https://github.com/ultralytics/yolov5>