

## НАУЧНЫЙ ОБЗОР

УДК 004.93'12  
DOI 10.17513/snt.39865

## МЕТОДЫ И ТЕХНОЛОГИИ ОБНАРУЖЕНИЯ ПОТЕНЦИАЛЬНО ОПАСНЫХ ОБЪЕКТОВ В ВИДЕОПОТОКЕ С ИНТРОСКОПА

Михалев А.С., Кулаков Е.Д., Меньшенин А.Н., Кузнецов А.С.

ФГАОУ ВО «Сибирский федеральный университет», Красноярск, e-mail: asmikhalev@yandex.ru,  
eg.2015j@yandex.ru, menshenin-2001@mail.ru, askuznetsov@sfu-kras.ru

Статья содержит обзор существующих методов и технологий детектирования и реконструкции потенциально опасных объектов в видеопотоке рентгеновских изображений с интроскопа. Для решения задачи детектирования объектов на изображениях на сегодняшний день наиболее эффективной технологией является аппарат нейронных сетей. В результате проведенного литературного обзора было найдено большое количество исследований, в которых для задачи обнаружения опасных предметов используются различные модели детектирования объектов. В большинстве случаев в статьях для задачи обнаружения объектов применяются «классические» нейросетевые модели: Faster R-CNN, SSD и YOLO. Представленные в статье модели показывают достаточно высокие значения по указанным метрикам. Для класса «Взрывные устройства» был найден только один набор данных, которого нет в открытом доступе. Данный набор включает в себя только три компонента для сбора взрывного устройства, при условии, что на сегодняшний момент существует большое число как взрывчатых веществ, так и различных подручных материалов для их сбора. Что касается проблемы 3D-реконструкции, на данный момент существует множество решений для реконструкции единичных объектов, восстановление множества объектов встречается заметно реже. Существующие подходы различаются форматом представления результатов реконструкции: воксели, полигоны и облака точек. Для задачи реконструкции опасных предметов на кадрах, получаемых с интроскопа, наиболее подходящим является воксельное представление из-за его способности точно передавать внутреннюю структуру объекта, а также из-за относительной простоты при обработке и анализе данных. Облака точек также могут быть полезными для детального анализа отдельных объектов, но для быстрой реконструкции на основе данных с интроскопа может оказаться менее эффективным, чем воксельное представление. Перевод изображений с интроскопа в полигональные модели является более сложной задачей, особенно при нечеткости и перекрытии объектов в багаже и ручной клади.

**Ключевые слова:** интроскоп, опасные предметы, обнаружение опасных предметов, 3D-представление опасных предметов, интеллектуальная система поддержки принятия решений, искусственный интеллект

*Данное исследование выполнено за счет гранта Красноярского краевого фонда поддержки научной и научно-технической деятельности в рамках конкурса научно-технических и инновационных проектов, направленных на развитие сферы железнодорожного транспорта на территории Красноярского края, № 2023032509756.*

## METHODS AND TECHNOLOGIES FOR DETECTING POTENTIALLY DANGEROUS OBJECTS IN A VIDEO STREAM FROM INTROSCOPE

Mikhalev A.S., Kulakov E.D., Menshenin A.N., Kuznetsov A.S.

Siberian Federal University, Krasnoyarsk, e-mail: asmikhalev@yandex.ru,  
eg.2015j@yandex.ru, menshenin-2001@mail.ru, askuznetsov@sfu-kras.ru

The article contains a survey of existing methods and technologies of detection and reconstruction of potentially dangerous objects in the video stream of x-ray images from the introscope. To solve the problem of detecting objects in images, the most effective technology today is the neural network apparatus. As a result of the literature review, a large number of studies were found in which different object detection models are used for the problem of detecting dangerous objects. In most cases, the articles use “classical” neural network models for the task of object detection: Faster R-CNN, SSD, and YOLO. The models presented in the articles show quite high values for the mentioned metrics. For the Explosive Devices class, only one dataset was found that is not in the public domain. This dataset contains only three components to assemble an explosive device, as there are currently a large number of both explosives and various improvised materials to assemble them. Regarding the 3D reconstruction problem, there are currently many solutions for single object reconstruction, but reconstruction of multiple objects is much less common. Existing approaches differ in the representation format of the reconstruction results: voxels, polygons, and point clouds. For the task of reconstructing dangerous objects on the images received by the Introscope, the voxel representation is the most suitable.

**Keywords:** introscope, dangerous objects, detection of dangerous objects, 3D representation of dangerous objects, intelligent decision support system, artificial intelligence

*This research was carried out at the expense of a grant from the Krasnoyarsk Regional Fund for the Support of Scientific and Scientific-Technical Activities within the framework of a competition of scientific, technical and innovative projects aimed at developing the sphere of railway transport in the Krasnoyarsk Territory, No. 2023032509756.*

Досмотр багажа – одна из основных мер обеспечения таможенной безопасности. Основным инструментом досмотра багажа является интроскоп, который с помощью рентгеновских лучей делает снимок содержимого сумки. В дальнейшем этот снимок проверяется сотрудником отдела безопасности на наличие запрещенных предметов. Данная мера предосторожности используется практически на всех крупных КПП Российской Федерации, так как это один из самых быстрых и эффективных способов досмотра.

На данный момент можно выделить две основные проблемы, которые могут возникнуть при досмотре багажа. Во-первых, при анализе видеопотока с интроскопа возможны человеческие ошибки. Операторы интроскопа, хоть и являются профессионалами и справляются с работой, но никто не застрахован от ошибок. Нагромождение сумок также отрицательно влияет на время принятия решения и эффективность обнаружения. В вопросах безопасности цена совершения ошибок чрезвычайно высока. Кроме того, подготовка специалиста может стоить много времени и денег. Второй проблемой при досмотрах является время: чем качественнее досмотр, тем больше на него уходит времени. При большом потоке пассажиров просто невозможно совершить качественный и быстрый досмотр.

В рамках стратегии транспорта Российской Федерации до 2030 г. [1] были утверждены принципы обеспечения безопасности на транспорте и транспортной безопасности, которые направлены на решение широкого спектра задач транспортного комплекса. Они включают в себя развитие систем автоматизации процедур досмотра пассажиров, используя новейшие технологии, технические средства, цифровизацию и системы искусственного интеллекта, для ускорения процессов досмотра, повышения удобства для пассажиров, а также с целью повышения вероятности обнаружения запрещенных предметов и веществ.

На сегодняшний день подавляющее большинство функционирующих на вокзалах интроскопов не имеют интеллектуальных функций обнаружения запрещенных предметов. Лишь только часть производителей начинают внедрять данный функционал в свое оборудование. Однако их решение может работать только с определенным типом интроскопов. В связи с этим возникает необходимость разработки универсального решения, которое позволит работать с любым типом оборудования.

В рамках данной статьи будут рассмотрены существующие достижения в об-

ласти автоматизации процесса обнаружения опасных предметов на видеопотоке с интроскопа.

#### *Обзор существующих подходов к обнаружению опасных предметов*

Для решения задачи детектирования объектов на изображениях на сегодняшний день наиболее эффективной технологией является аппарат нейронных сетей. В результате проведенного литературного обзора было найдено много разных статей, в которых для задачи обнаружения опасных предметов используются различные модели детектирования объектов.

В статье [2] авторы сравнивали модели SSD-InceptionV2, Faster R-CNN-ResNet101, Faster R-CNN-ResNet152, Faster R-CNN-InceptionResNetV2. Для исследования был собран собственный набор данных, состоящий из следующих категорий объектов (жидкости/аэрозоли/гели, острые предметы, тупые предметы, огнестрельное оружие). Наилучший результат показала модель Faster-RCNN-InceptionResNetV2 со значением метрики mAP 0,941. Для холодного оружия значение метрики AP равно 0,818, для огнестрельного оружия – 0,962).

В статье [3] используется модифицированная версия архитектуры Faster R-CNN с целью улучшения детекции объектов на изображениях путем учета и адаптации к сложному фону. Эта модификация призвана бороться с проблемой, когда объекты могут затеряться на сложном фоне, что может привести к неверным детекциям или потере объектов. Для исследования был использован набор данных с категориями: жидкости/аэрозоли/гели, острые предметы, тупые предметы, огнестрельное оружие. Для холодного оружия значение метрики AP равно 0,914, для огнестрельного оружия – 0,986.

Еще одно исследование [4] посвящено сравнению моделей Faster R-CNN, R-FCN и SSD с разными сверточными моделями (backbone network). Для исследования был выбран набор данных SiXray [5], содержащий 1,059,231 изображений для следующих классов: пистолет, нож, гаечный ключ, плоскогубцы, ножницы, молоток. Для класса оружие лучший результат показала модель R-FCN-ResNet101 (значение метрики AP равно 0.973), для класса нож – SSD-InceptionV2 (AP 0,891).

Результаты распознавания взрывных устройств приведены в статье [6]. В данной работе авторами был собран свой набор данных изображения копий элементов для изготовления самодельных взрывных устройств (датасет IEDXray), который включал 1300 рентгеновских изображений (источни-

ки питания, провода, взрыватель). Для детектирования объектов сравнивались модели Faster R-CNN, SSD и R-FCN с разными классификаторами. Для увеличения точности детектирования применялась техника переноса обучения (transfer learning) и аугментация. Наибольшую точность показала модель Faster R-CNN ResNet-101 со значением метрики mAP 0,7729.

В статье [7] для распознавания опасных объектов была использована модель YOLO v3. В данной статье авторы использовали набор данных датасет IEDXray. В результате обучения была достигнута средняя точность mAP равная 0,524.

В статье [8] авторы исследовали разные версии алгоритма YOLO: YOLOv2, YOLOv3, YOLOv4 и YOLO-T. Для исследования авторы использовали открытый расширенный набор данных GDXray [9], включающий в себя следующие классы: нож, ножницы, пистолет и разные виды гранат (4 типа). Наилучший результат по метрике mAP показала модель YOLO-T (0,9773).

В статье [10] авторы сравнивали модели YOLOv5, SSD и FCOS. Одна из проблем обнаружения объектов в сумках связана с тем, что, если в багаже много предметов, они могут перекрывать друг друга на рентгеновском изображении, что, соответственно, затрудняет их идентификацию. Для борьбы с данной проблемой авторы используют модуль латерального подавления (LIM), который устраняет влияние шумных соседних областей на области интереса объектов и активизирует границы объектов, усиливая их контрастность. Исследование проводилось на двух наборах данных NiXray (портативное зарядное устройство, мобильный телефон, ноутбук, планшет, косметика, вода, зажигалка), включающий порядка 100 000 снимков и OPIXray (складной нож, нож, ножницы, канцелярский нож, швейцарский нож) [11]. Наилучшая средняя точность была получена для модели YOLOv5+LIM: mAP 0,832 для набора данных NiXray и mAP 0,906 для набора данных OPIXray (для холодного оружия значение метрики AP равно 0,776).

В статье [12] авторы сравнивают две архитектуры сверточных нейронных сетей для задачи детектирования объектов: Cascade R-CNN и FreeAnchor. Для апробации были использованы наборы данных SIXray и OPIXray. Результаты апробации показали превосходство архитектуры FreeAnchor как по уровню точности (mAP 0,877 и 0,858), так и по скорости обработки данных. Для набора данных SIXray получено значение метрики AP равное 0,81 для холодного оружия и AP 0,898 для огнестрель-

ного оружия, для набора данных OPIXray значение метрики AP 0,746 для холодного оружия.

В статье [13] авторы предложили новую улучшенную версию алгоритма Mask R-CNN. Mask R-CNN – улучшенная версия алгоритма Faster R-CNN, разработанная для решения задачи сегментации. В Mask R-CNN к традиционным для алгоритмов семейства R-CNN метке класса и координатам ограничивающей рамки добавляется также маска объекта – прямоугольная матрица принадлежности пикселя текущему объекту. Ключевым элементом Mask R-CNN является использование RoIAlign. В качестве основной сети, служащей для извлечения признаков из поступающего на вход изображения, используется FPN (Feature Pyramid Networks). FPN представляет собой структуру, которая генерирует пирамиду признаков на разных масштабах, позволяя нейронной сети работать с объектами разного размера. Это достигается путем создания серии сверточных слоев, каждый из которых работает на разных уровнях разрешения, и затем объединяет их вместе посредством нисходящего пути. Оптимизация процесса извлечения признаков выполняется путем добавления дополнительного пути «снизу вверх», который преобразует низкоуровневые признаки обратно в высокоуровневые признаки для избежания потери информации. Среди других улучшений для данной задачи является поиск отрицательных примеров в реальном времени. Для этого полученные области интереса (ROI) подаются в ONEM-модуль, который отбирает для дальнейшего обучения только области с наибольшими потерями. Для устранения погрешности разметки дополнительно используется оператор Собеля для обнаружения краев. Предложенная модель сравнивалась с моделями Faster R-CNN, SSD и YOLOv3. Предложенная модель позволяет решать ряд возможных проблем, связанных с неправильным размещением, окклюзиями и перекрытиями, а также маленьким размером. В качестве исходных данных использовалась выборка из набора данных SIXray, которая состояла из порядка 7000 рентгеновских снимков. Средняя точность mAP разработанного решения составляет 0,924. Для холодного оружия значение метрики AP равно 0,836, для огнестрельного оружия – 0,967.

В результате проведенного литературного обзора можно следующие выводы. Для классов «Холодное оружие» и «Огнестрельное оружие» найдено большое количество статей, представляющих собой результаты исследования нейронных сетей на различных открытых наборах данных.

В большинстве случаев в статьях для задачи обнаружения объектов применяются «классические» нейросетевые модели: Faster R-CNN, SSD и YOLO. Представленные в статьях модели показывают достаточно высокие значения по указанным метрикам. Для класса «Взрывные устройства» был найден только один набор данных, которого нет в открытом доступе. Данный набор включает в себя только три компонента для сбора взрывного устройства, при условии, что на сегодняшний момент существует большое число как взрывчатых веществ, так и различных подручных материалов для их сбора.

#### *Обзор существующих подходов к решению задачи 3D-реконструкции объектов*

На данный момент существует множество разных подходов к решению задачи реконструкции трехмерной структуры объектов. Однако фактически нет работ, посвященных решению задачи генерации 3D-изображений содержимого багажа на основе 2D-изображений с разных проекций. В связи с этим далее будут рассмотрены подходы, которые можно адаптировать под текущую задачу.

Существующие подходы в основном используют похожие принципы работы и полагаются либо на аппаратные и математические инструменты, либо на применение искусственных нейронных сетей. Способ представления результатов при этом может быть разным. Наиболее распространенными форматами представления 3D-объектов являются облака точек, полигональные сетки и воксели (рисунок).

#### *Представление объектов в виде облака точек*

Облако точек – это совокупность большого числа точек поверхности объектов в трехмерной системе координат. Классический способ реконструкции 3D-изображений объектов основан на применении сверточных нейронных сетей к одному RGB изображению. В статье [14] предложена глубокая нейронная сеть Point Set Prediction Network.

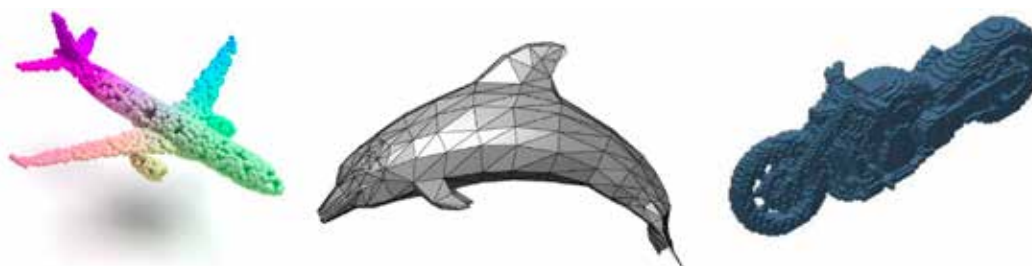
Данная модель включает в себя модули «encoder», «decoder» и «prediction». Операции «кодирования» и «декодирования» выполняются рекурсивно, что позволяет лучше сочетать глобальную и локальную информацию. «Predictor» использует две параллельные ветви: полносвязную нейронную сеть и «decoder». Полносвязная часть предсказывает сами точки, а «decoder» – 3-канальное изображение, где каждый пиксель представляет собой координаты точек. Их объединение представляет собой представление объекта в виде облака точек. Для обучения использовалась выборка из набора данных ShapeNet [15], которая состояла из 220 000 моделей и 2000 категорий объектов.

В работе [16] для реконструкции 3D-изображений был предложен подход на основе сверточной нейронной сети и метода свободной деформации. Данный метод позволяет выполнять деформацию точек, полученных из высококачественной сетки, для представления объекта в виде облака точек. Для настройки метода авторами использовалось подмножество набора данных ShapeNet Core. Оценка предлагаемой структуры проводилась как на синтетических, так и на реальных данных. В рамках апробации алгоритм доказал свою конкурентоспособность.

В работе [17] для задачи реконструкции предлагается использовать модель вариационного автокодировщика, которая позволяет получать 3D-представление как с одной, так и по нескольким изображениям. 3D-изображение объекта получается на основе многоточечных силуэтов и карт глубины. Для обучения в работе был использован набор данных NYU-D [18].

#### *Представление объектов в виде вокселей*

Результат реконструкции объектов также можно представить в виде пространственной сетки, которая разбивает объем сцены на небольшие прямоугольные ячейки – воксели. Воксел – это аналог пиксела в трехмерном пространстве.



*Примеры форматов представления 3D-объектов*

Пример решения задачи преобразования 2D-изображения в воксельное представление представлен в статье [19]. В данной работе была предложена рекуррентная нейронная сеть 3D-R2N2. Данная модель состоит из 2D-кодировщика, 3D-сверточной сети долгой краткосрочной памяти и 3D-декодировщика. Данное решение способно выполнять реконструкцию объекта как по одному, так и по нескольким изображениям. Для обучения модели был собран набор данных 3D-моделей взятых из ShapeNet, PASCAL 3D [20] и Online Products [21].

В работе [22] для решения задачи 3D-реконструкции применяется проективная генеративная состязательная сеть PrGAN. Предложенная в статье генеративная модель позволяет создавать 3D-формы объектов, используя 2D-изображения этих объектов без каких-либо дополнительных аннотаций. Для этого в алгоритме используется генератор ракурса и модуль проекции. Модуль проекции позволяет получать информацию о базовом распределении трехмерных форм без использования какой-либо трехмерной информации. Для обучения были использованы 3D-модели, взятые из ModelNet [23] и ИКЕА [24].

В статье [25] был предложен метод Pix2Vox, основанный на концепции объединения кодировщика и декодировщика с использованием предварительного обучения сети VGG. Дальнейшим развитием данного решения стал Pix2Vox++ [26]. В данной версии метода была заменена нейронная сеть для извлечения признаков изображений, а также выполнена модификация контекстно-ориентированного слияния до мультиразмерного варианта, отличающегося наличием в оценочной сети остаточных связей для учета информации от слоев разной глубины.

#### *Представление объектов в виде полигонов*

Другая группа решений использует в качестве представления реконструированных объектов полигональную форму (меш). Как правило, такие методы полагаются на некий начальный примитив, вершины которого деформируются так, чтобы наиболее точно соответствовать форме объекта на изображении, что несколько ограничивает круг восстанавливаемых объектов.

В статье [27] предложено решение SurfNet на основе глубокой остаточной нейронной сети. Основной идеей подхода является генерация поверхности 3D-объекта за счет объединения различных представлений объекта в одно геометрическое изо-

бражение с учетом их формы и геометрии. Авторами для обучения была использована выборка из наборов данных ShapeNet и PASCAL 3D+.

В работе [28] применяется сверточная нейронная сеть для реконструкции 3D-модели по цветным изображениям и изображениям глубины. Предложенный подход сначала выполняет поиск похожего изображения глубины из обучающего набора полигональных форм. Далее выполняется деформация выбранной формы. На последнем этапе выполняется предсказание полной формы целевого объекта на основе деформированной формы и значения глубины. Для обучения был собран свой набор данных 3D-моделей, взятых из SHREC'12 [29], ShapeNetCore и PASCAL 3D+.

В статье [30] предложена сверточная нейронная сеть AtlasNet. В данном подходе процесс реконструкции поверхности объекта реализован по принципу атласа: форма объекта выводится из множества мелких 2D-примитивов (карт), представляющих собой единичные квадраты. Данное решение способно получать на вход как единичное изображение, так и облако точек. Особенностью же архитектуры является использование набора многослойных перцептронов, где каждый перцептрон отвечает за поведение одной карты. В процессе обучения сеть запоминает способ трансформации точек 2D-карт, который позволит сложить из них объект.

#### *Распознавание объектов через подмену готовыми моделями*

Можно выделить еще один подход к восстановлению сцены, который опирается на использовании заранее подготовленной базы данных с трехмерными моделями, которые подставляются на сцену вместо обнаруженного объекта. В этом случае понимание пространственной структуры объекта не требуется, достаточно научиться хорошо определять форму и внешний вид объекта, т.е. определить, к какому классу относится данный объект. Примером такого решения служит статья [31], в которой производится распознавание множества объектов по одному изображению. В основе архитектуры предлагаемой модели лежит детектор центральных точек объектов. Ключевая идея состоит в том, чтобы найти все объекты на изображении и предсказать для них форму и 9D ограничивающую рамку (3D-положение, 3D-вращение, 3D-масштаб). После чего для каждого объекта необходимо предсказать индекс трехмерной модели из базы данных, которая лучше всего подходит форме, и поместить ее на сцене, выровняв в со-

ответствии с предсказанной рамкой. Это позволяет получить наиболее качественный результат с визуальной точки зрения. Все работает хорошо до тех пор, пока объекты в кадре соответствуют моделям в базе данных. В противном случае на место объекта будет поставлена одна из моделей, даже если она на самом деле слабо на него похожа. Кроме того, базу объектов необходимо хранить и обслуживать, что вносит дополнительные расходы.

### Заключение

В свете растущих требований к безопасности, особенно в транспортных узлах, таких как аэропорты и железнодорожные станции, возрастает необходимость в более детализированном анализе содержимого багажа и грузов. В связи с программой импортозамещения в России и отсутствием программных систем автоматического обнаружения потенциально опасных объектов, которые поставляются без соответствующего оборудования, возможным решением является разработка универсальной программной системы, которая позволит работать с различными моделями интроскопов, что обеспечит широкий охват потенциальных пользователей и сценариев применения. Данная система должна позволять обнаруживать опасные предметы и создавать трехмерное представление опасных предметов на основе двумерных проекций. 3D-моделирование на основе данных видеопотока с рентгеновской установки позволяет получить более детализированное представление о внутренней структуре объекта, что может способствовать лучшему выявлению запрещенных или опасных предметов.

В рамках данной статьи был проведен литературный обзор исследований в области автоматического обнаружения опасных предметов на изображениях, получаемых с интроскопа и 3D-реконструкции объектов. На основании этого обзора можно сделать несколько выводов относительно будущих направлений в области разработки системы поддержки принятия решений по обнаружению потенциально опасных объектов в видеопотоке с интроскопа. В открытом доступе отсутствуют большие, хорошо сбалансированные наборы данных, содержащие основные классы опасных предметов (оружие, взрывчатые вещества и взрывные устройства). Таким образом, во-первых, стоит задача сбора собственного набора данных для указанных классов опасных предметов. Во-вторых, в большинстве статей для решения данной задачи применяются «классические» подходы к решению задачи обнаружения объектов. В рамках создания

собственного алгоритма требуется провести исследование современных нейросетевых архитектур, таких как EfficientDet [32], M2Det [33], CornerNet [34], DetectoRS [35] и новые версии YOLO [36].

Что касается задачи 3D-реконструкции, на данный момент существует множество решений для реконструкции единичных объектов, восстановление множества объектов встречается заметно реже. Существующие подходы различаются форматом представления результатов реконструкции: воксели, полигоны и облака точек. Для задачи реконструкции опасных предметов на кадрах, получаемых с интроскопа, наиболее подходящим является воксельное представление из-за его способности точно передавать внутреннюю структуру объекта, а также из-за относительной простоты при обработке и анализе данных. Облака точек также могут быть полезными для детального анализа отдельных объектов, но для быстрой реконструкции на основе данных с интроскопа могут оказаться менее эффективными, чем воксельное представление. Перевод изображений с интроскопа в полигональные модели является более сложной задачей, особенно при нечеткости и перекрытии объектов в багаже и ручной клади.

### Список литературы

1. Распоряжение Правительства Российской Федерации от 23.05.2020 № 1388-р. 52 с. [Электронный ресурс]. URL: <https://customs.gov.ru/activity/programmy-razvitiya/strategiya-razvitiya-fts-rossii-do-2030-goda> (дата обращения: 30.09.2023).
2. Liang K.J., Sigman J.B., Spell G.P. and others. Toward automatic threat recognition for airport X-ray baggage screening with deep convolutional object detection // arXiv preprint arXiv:1912.06329. 2019. [Электронный ресурс]. URL: <https://arxiv.org/abs/1912.06329> (дата обращения: 30.09.2023).
3. Sigman J.B., Spell G.P., Liang K.J. and others. Background adaptive faster R-CNN for semi-supervised convolutional object detection of threats in x-ray images // Anomaly Detection and Imaging with X-Rays Vol. 2020. P. 5. DOI: 10.1117/12.2558542.
4. Wang R., Wu S., Lan Y. Object Detection in X-ray Images [Электронный ресурс]. URL: <https://medium.com/sfucspmp/object-detection-in-x-ray-images-414a4fb06dff> (дата обращения: 30.09.2023).
5. Miao C., Xie L., Wan F., Su C., Liu H., Jiao J. & Ye Q. (2019). Sixray: A large-scale security inspection x-ray benchmark for prohibited item discovery in overlapping images // In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019. C. 2119–2128. DOI: 10.48550/arXiv.1901.00303.
6. Galvez R.L., Dadios E.P., Bandala A., Vicerra R.R.P. Object Detection in X-ray Images Using Transfer Learning with Data Augmentation // International Journal on Advanced Science Engineering and Information Technology. 2019. P. 2147–2153. DOI: 10.18517/ijaseit.9.6.9960.
7. Galvez R.L., Dadios E.P., Bandala A., Vicerra R.R.P. YOLO-based Threat Object Detection in X-ray Images // International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, En-

- vironment, and Management. 2019. P. 1–5. DOI: 10.1109/HNICEM48295.2019.9073599.
8. Wang M., Yang B., Wang X., and others. YOLO-T: Multitarget Intelligent Recognition Method for X-ray Images Based on the YOLO and Transformer Models // *Applied Sciences*. 2022. № 12. P. 1–18. DOI: 10.3390/app122211848.
9. Mery D., Riffo V., Zscherpel U. and others. GDXray: The database of X-ray images for nondestructive testing // *Journal of Nondestructive Evaluation*. 2015. T. 34, № 4. P. 42.
10. Tao R., Wei Y., Jiang X. and others. Towards real-world X-ray security inspection: A high-quality benchmark and lateral inhibition module for prohibited items detection // *Proceedings of the IEEE/CVF international conference on computer vision*. 2021. P. 10923–10932.
11. Wei Y., Tao R., Wu Z., Ma Y., Zhang L., Liu X. Occluded prohibited items detection: An x-ray security inspection benchmark and de-occlusion attention module // *Proceedings of the 28th ACM international conference on multimedia*. 2020. P. 138–146.
12. Webb T.W., Bhowmik N., Gaus Y.F.A. & Breckon T.P. Operationalizing convolutional neural network architectures for prohibited object detection in x-ray imagery // *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2021. P. 610–615.
13. Zhang J., Song X., Feng J., Fei J. X-Ray Image Recognition Based on Improved Mask R-CNN Algorithm // *Mathematical Problems in Engineering*. 2021. № 2. P. 1–14. DOI: 10.1155/2021/6544325.
14. Fan H., Su H., Guibas L. A Point Set Generation Network for 3D Object Reconstruction from a Single Image // *Conference on Computer Vision and Pattern Recognition*. 2017. P. 1–12. DOI: 10.1109/CVPR.2017.264.
15. Chang A.X., Funkhouser T., Guibas L. and other Shapenet: An information-rich 3d model repository // *arXiv preprint arXiv:1512.03012*. 2015. [Электронный ресурс]. URL: <https://arxiv.org/abs/1512.03012v1> (дата обращения: 30.09.2023).
16. Jack D., Pontes J.K., Sridharan S., and others. Learning free-form deformations for 3D object reconstruction // *Asian Conference on Computer Vision*. 2018. P. 1–16. DOI: 10.1007/978-3-030-20890-5\_21.
17. Soltani A.A., Huang H., Wu J. and others. Synthesizing 3D Shapes via Modeling Multi-view Depth Maps and Silhouettes with Deep Generative Networks // *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017. P. 2511–2519. DOI: 10.1109/CVPR.2017.269.
18. Silberman N., Hoiem D., Kohli P. and Fergus R. Indoor Segmentation and Support Inference from RGBD Images // *European Conference on Computer Vision*. 2012. P. 746–760. DOI: 10.1007/978-3-642-33715-4\_54.
19. Choy C.B., Xu D., Gwak J. and others. 3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction // *European Conference on Computer Vision*. 2016. P. 1–17. DOI: 10.1007/978-3-319-46484-8\_38.
20. Xiang Y., Mottaghi R., Savarese S. Beyond PASCAL: A benchmark for 3D object detection in the wild // *Winter Conference on Applications of Computer Vision*. 2014. P. 75–82. DOI: 10.1109/WACV.2014.6836101.
21. Song H.O., Xiang Y., Jegelka S., Savarese S. Deep Metric Learning via Lifted Structured Feature Embedding // *Conference on Computer Vision and Pattern Recognition*. 2015. P. 4004–4012. DOI: 10.1109/CVPR.2016.434.
22. Gadelha M., Maji S., Wang R. 3D Shape Induction from 2D Views of Multiple Objects // *International Conference on 3D Vision (3DV)*. 2017. P. 402–411. DOI: 10.1109/3DV.2017.00053.
23. Wu Z., Song S., Khosla A. and others. 3D ShapeNets: A Deep Representation for Volumetric Shape Modeling // *Computer Vision and Pattern Recognition*. 2015. P. 1912–1920. DOI: 10.48550/arXiv.1406.567.
24. Lim J.J., Pirsiavash H., Torralba A. Parsing IKEA Objects: Fine Pose Estimation // *International Conference on Computer Vision*. 2013. P. 1–8. DOI: 10.1109/ICCV.2013.372.
25. Haozhe X., Hongxun Y., Xiaoshuai S., Shangchen Z., Shengping Z. Pix2Vox: Context-aware 3D Reconstruction from Single and Multi-view Images // *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2019. P. 2690–2698.
26. Xie H. and others. Pix2Vox++: Multi-scale context-aware 3D object reconstruction from single and multiple images // *International Journal of Computer Vision*. 2020. T. 128. № 12. C. 2919–2935.
27. Sinha A., Unmesh A., Huang Q., Ramani K. SurfNet: Generating 3D shape surfaces using deep residual networks // *Computer Vision and Pattern Recognition*. 2017. P. 1–10. DOI: 10.48550/arXiv.1703.04079.
28. Rock J., Gupta T., Thorsen J., Gwak J., Shin D., and Hoiem D. Completing 3d object shape from one depth image // *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015. P. 2484–2493.
29. Shin D., Fowlkes C.C., Hoiem D. Pixels, Voxels, and Views: A Study of Shape Representations for Single View 3D Object Shape Prediction // *Conference on Computer Vision and Pattern Recognition*. 2018. P. 3061–3069. DOI: 10.1109/CVPR.2018.00323.
30. Groueix T., et al. A papier-mache approach to learning 3d surface generation // *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018. C. 216–224.
31. Engelmann F., et al. From points to multi-object 3D reconstruction // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021. P. 4588–4597.
32. Tan M., Pang R., Le Q.V. Efficientdet: Scalable and efficient object detection // *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020. P. 10781–10790.
33. Zhao Q., Sheng T., Wang Y., Tang Z., Chen Y., Cai L., & Ling H. M2det: A single-shot object detector based on multi-level feature pyramid network // *Proceedings of the AAAI conference on artificial intelligence*. 2019. T. 33, № 01. P. 9259–9266.
34. Law H., Deng J. Cornernet: Detecting objects as paired keypoints // *Proceedings of the European conference on computer vision (ECCV)*. 2018. P. 734–750.
35. Qiao S., Chen L.C., Yuille A. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution // *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021. P. 10213–10224.
36. Terven J., Cordova-Esparza D. A comprehensive review of YOLO: From YOLOv1 and beyond. *arXiv 2023* // *arXiv preprint arXiv:2304.00501*. [Электронный ресурс]. URL: <https://arxiv.org/abs/2304.00501> (дата обращения: 30.09.2023).