



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н. Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н. Э. Баумана)

ФАКУЛЬТЕТ «Информатика, искусственный интеллект и системы управления»

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

К ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЕ

НА ТЕМУ:

*«Метод распознавания надводных объектов с
аэрофотоснимков с использованием нейронных сетей»*

Студент ИУ7-83Б
(Группа)

(Подпись, дата)

Миронов Г. А.
(И. О. Фамилия)

Руководитель ВКР

(Подпись, дата)

Тассов К. Л.
(И. О. Фамилия)

Нормоконтролер

(Подпись, дата)

(И. О. Фамилия)

2023 г.

РЕФЕРАТ

Расчетно-пояснительная записка 28 с., 12 рис., 0 табл., 22 источн., 1 прил.

СОДЕРЖАНИЕ

РЕФЕРАТ	5
ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ	8
ВВЕДЕНИЕ	9
1 Аналитический раздел	10
1.1 Предметная область	10
1.2 Нейронные сети	10
1.3 Сверточные нейронные сети	12
1.3.1 Свертка	13
1.3.2 Двухэтапные CNN	15
1.3.3 Одноэтапные CNN	16
1.4 Ансамбли	18
1.4.1 Бустинг	18
1.4.2 Бэггинг	19
1.5 Существующие методы	20
1.5.1 Разрешение изображений	20
1.5.2 R-CNN	21
1.5.3 YOLO	21
1.5.4 Параметры для сравнения	21
1.5.5 Сравнение существующих методов	22
1.6 Метод распознавания надводных объектов с аэрофотоснимков с использованием нейронных сетей	22
1.7 Формализованная постановка задачи	22
1.8 Выбор данных для обучения модели	22
2 Конструкторский раздел	23
2.1 Метод распознавания	23
ЗАКЛЮЧЕНИЕ	24
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	25

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

В настоящей расчетно-пояснительной записке применяют следующие сокращения и обозначения.

ANN — (англ. Artificial Neural Network) Искусственная нейронная сеть

CNN — (англ. Convolutional Neural Network) Свёрточная нейронная сеть

RoI — (англ. Region Of Interest) Область интереса

VHR — (англ. Very High Resolution) Сверхвысокое разрешение

MR — (англ. Medium Resolution) Среднее разрешение

ВВЕДЕНИЕ

1 Аналитический раздел

1.1 Предметная область

В настоящее время все больше внимания привлекает тема автоматизации судоходства, а так же общее повышение уровня безопасности в процессе эксплуатации судов.

Методы распознавания различных надводных объектов имеют особое значение в данном контексте, так как позволяют решить множество проблем: избежание столкновений судов, осуществление автономного плавания и пр. [1].

Не менее важной является задача распознавания малых надводных объектов, поскольку традиционные методы обнаружения, основанные на использовании радара, не подходят для задачи обнаружения близко расположенных и малых объектов [2].

Кроме того, обнаружение активности рыболовецких судов по-прежнему является сложной задачей для многих стран, расположенных на архипелагах, например — Индонезии. В настоящее время для мониторинга огромной морской акватории используется технология, использующая датчики SAR для обнаружения кораблей, разрабатываемая с 1985 года. Однако стоимость использования данной технологии является одним из препятствий для дальнейшего развития технологии [3].

1.2 Нейронные сети

ANN определяется как массово параллельный распределенный процессор, состоящий из простых процессорных блоков, который обладает естественной склонностью накапливать эмпирические знания [4].

Их название и структура вдохновлены человеческим мозгом, а алгоритм работы основывается на способе, которым биологические нейроны передают сигналы друг другу.

В общем случае ANN включает в себя входной слой, выходной (или целевой) слой и, между ними, скрытый слой. Слои соединены через узлы (искусственные нейроны). Эти соединения образуют «сеть» — нейронную сеть — из взаимосвязанных узлов. Пример приведен на рисунке 1.1.

В общем случае искусственный нейрон можно представить в виде регрессионной модели [5], состоящей из входных данных, весовых коэффициентов,

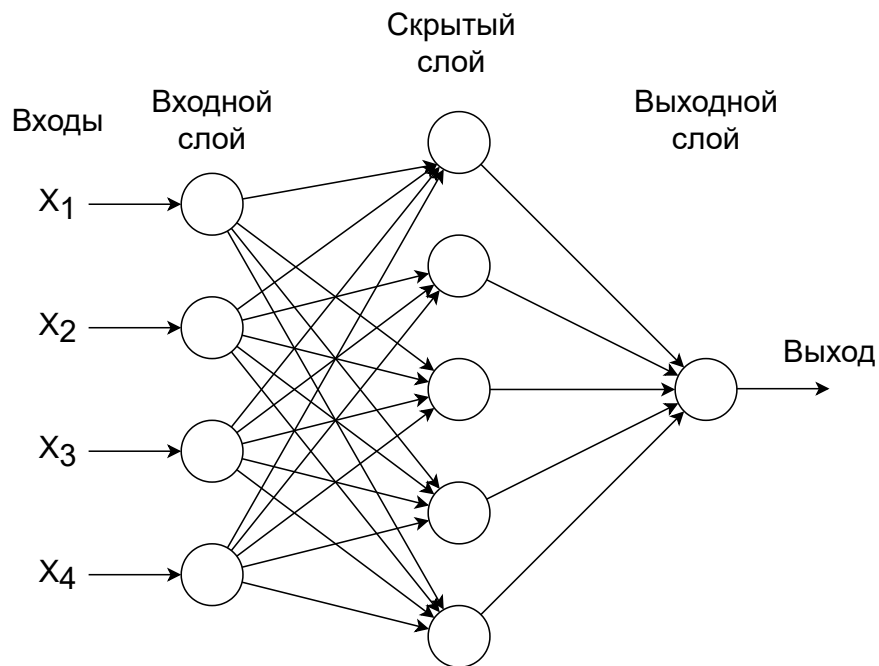


Рисунок 1.1 – Пример схемы ANN

смещения (или порогового значения) и выходных данных. Эту модель можно описать следующей формулой:

$$\hat{y} = \sum_{i=1}^n w_i x_i + w_0, \quad (1.1)$$

В качестве функции активации можно использовать: ступенчатую, линейную функции, сигмоиду, ReLu и другие [5].

Модель искусственного нейрона приведена на рисунке 1.2.

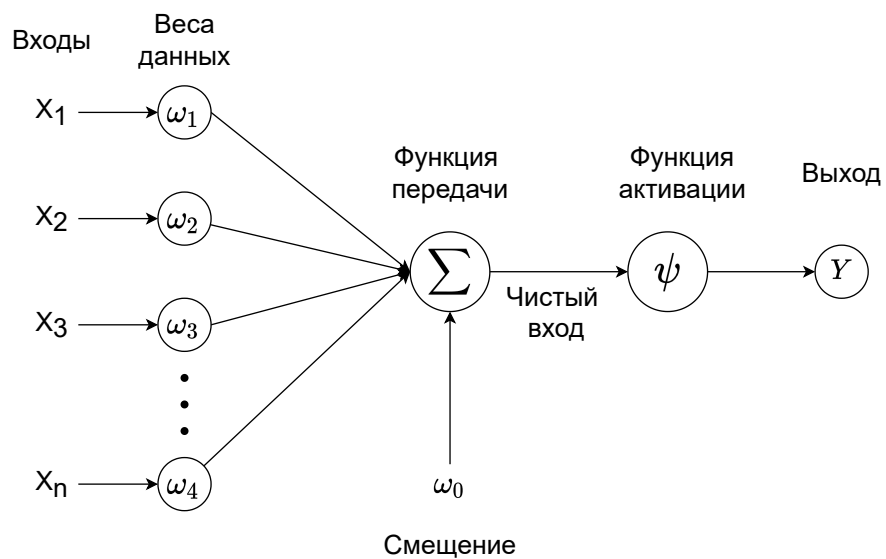


Рисунок 1.2 – Общая схема искусственного нейрона

Данное представление применимо к любому виду нейронных сетей — вне зависимости от типа, нейронные сети реализуются путем упорядочивания нейронов в слои и последующим связыванием соответствующих слоев между собой [4].

В процессе обучения нейронной сети используется так называемая обучающая выборка — заранее подготовленный набор данных, отражающий суть рассматриваемой предметной области [4]. В зависимости от содержимого обучающей выборки результирующие весовые конфигурации нейронной сети (т. е. веса связей между нейронами, а так же смещения отдельно взятых нейронов) могут отличаться [4]. В связи с этим одна и та же структура нейронной сети может переиспользована для работы в различных предметных областях.

1.3 Сверточные нейронные сети

Одним из основных видов нейронных сетей, применяемых для распознавания является CNN [6].

CNN представляет собой тип ANN, которая имеет архитектуру с глубокой обратной связью и выделяется на фоне остальных ANN с полносвязными слоями своей способностью к обобщению. CNN работает с сильно абстрагированными характеристиками объектов, особенно это касается пространственных данных, что позволяет добиться более эффективно идентифицировать объекты в сравнении с другими типами ANN [6]. Одним из отличительных свойств CNN является способность к фильтрации посторонних шумов во входных данных.

Модель CNN состоит из конечного набора уровней обработки, которые могут изучать различные характеристики входных данных (например, изображения) с несколькими уровнями абстракции. Начальные уровни изучают и извлекают высокоуровневые свойства, а более глубокие уровни изучают и извлекают более низкоуровневые свойства. Концептуальная модель CNN представлена на рисунке 1.3.

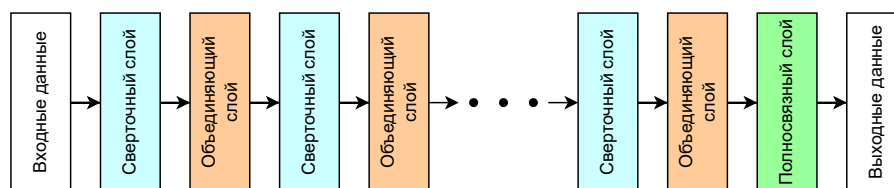


Рисунок 1.3 – Концептуальная модель CNN

Существующие архитектуры CNN для обнаружения объектов на изображениях можно разделить на две категории: одноэтапные (one-stage) и двухэтапные (two-stage) [7].

1.3.1 Свертка

Ядро

Прежде, чем рассматривать процесс свертки, необходимо определить понятие «ядро», используемое при свертке. Ядро представляет собой матрицу из дискретных значений или чисел, где каждое значение известно как вес этого ядра. Пример двумерного ядра приведен на рисунке 1.4.

0	1
-1	2

Рисунок 1.4 – Пример двумерного ядра с размерностью 2×2

Ядро инициализируется случайными значениями, которые изменяются в ходе обучения CNN.

Процесс свертки

Свертка — это операция над парой матриц $A(n_x, n_y)$ и $B(m_x, m_y)$, $m_x \leq n_x$, $m_y \leq n_y$, результатом которой является матрица

$$C(n_x - m_x + 1, n_y - m_y + 1) = A * B, \quad (1.2)$$

каждый элемент которой является скалярным произведением матрицы B (ядра свертки) и некоторой подматрицы A такого же размера.

Т. е. элемент матрицы C вычисляется следующим образом:

$$C_{i,j} = \sum_{u=0}^{m_x-1} \sum_{v=0}^{m_y-1} A_{i+u,j+v} B_{u,v}. \quad (1.3)$$

Пример

Разберем пример свертки для изображения в градациях серого, т.к. такое изображение содержит лишь один канал, передаваемый на вход CNN.

Пусть дано изображение, представленное на рисунке 1.5.

1	0	-2	1
-1	0	1	2
0	2	1	0
1	0	0	1

Рисунок 1.5 – Пример изображения в градациях серого с размерностью 4×4

Далее рассмотрим первые два шага процесса свертки, представленные на рисунках 1.6 и 1.7, соответственно.

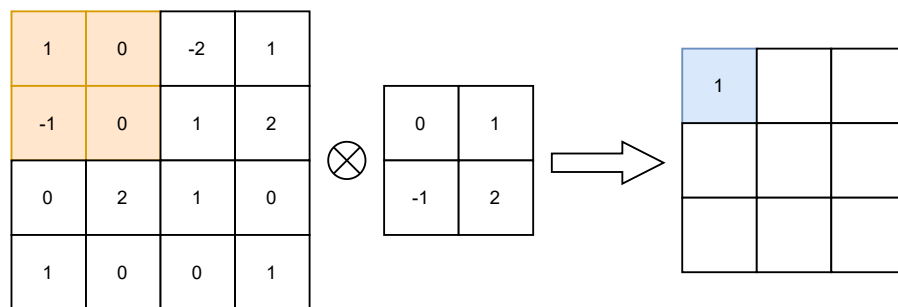


Рисунок 1.6 – Пример свертки изображения в градациях серого с размерностью 4×4

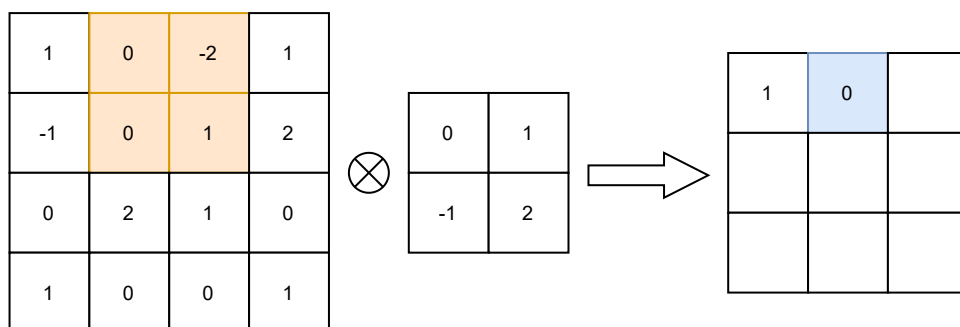


Рисунок 1.7 – Пример свертки изображения в градациях серого с размерностью 4×4

Аналогичным образом свертка продолжается до полного заполнения результирующей матрицы. Стоит отметить, что, в зависимости от размеров окна и перекрытия окон, будет меняться размер результирующей матрицы.

1.3.2 Двухэтапные CNN

В таких нейросетевых алгоритмах выделяют два этапа: поиска RoI (англ. Candidate Region Extraction) на изображении и последующей классификации RoI, найденных на первом этапе. При этом под RoI на изображении подразумеваются зоны, потенциально содержащие искомые объекты [8].

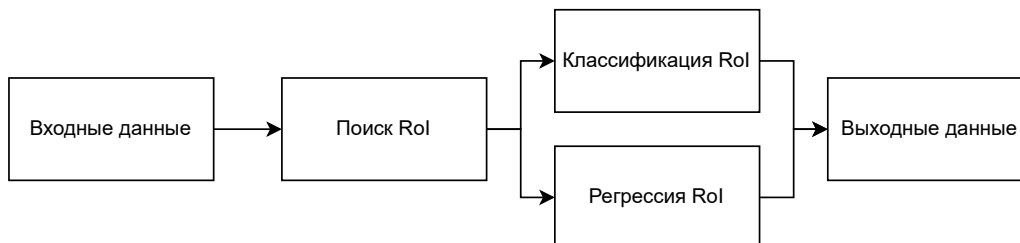


Рисунок 1.8 – Схема работы двухэтапного алгоритма

Стоит отметить, что первый этап может происходить без использования нейронных сетей. Для этого можно использовать информацию о контрасте, ключевые точки или перебор всех возможных положений объекта с помощью процедуры **selective search** [9].

RoI, полученные вышеперечисленными методами, могут обладать серьезными недостатками, например:

- содержать слишком большое количество фона;
- содержать лишь небольшую часть объекта;
- содержать более одного объекта.

В связи с этим на первом этапе более предпочтительным методом является применение CNN, не содержащих полносвязных слоев [9].

На втором этапе CNN применяются к обнаруженным RoI.

Преимуществом данных алгоритмов является высокая точность распознавания объектов, однако, платой за это является время, необходимое для выделения «подозрительных» зон на изображении [8].

R-CNN

Первой CNN, разработанной для обнаружения объектов, является модель Region-based CNN, которая использует подходы на основе скользящего окна (sliding window) для обнаружения объектов [10].

Здесь авторы разделяют всю задачу на три модуля. В первом модуле из каждого входного изображения извлекаются RoI, которые могут содержать какой-либо объект (с помощью процедуры selective search), затем во втором модуле авторы используют аффинное искажение изображения, чтобы сделать все извлеченные RoI фиксированного размера (или фиксированного соотношения сторон), а затем пропускают эти искаженные RoI через AlexNet CNN для извлечения конечных признаков (векторов признаков фиксированного размера). Наконец, третий этап представляет собой набор линейных SVM (support vector machine), которые причисляют каждый вектор какому-либо классу и отдельный регрессор обрамляющих окон [10].

Последующие версии алгоритма — Fast R-CNN [11] и Faster R-CNN [12] — призваны оптимизировать время работы алгоритма, а так же повысить точность распознавания. Так, в Faster R-CNN процедура selective search была заменена на сеть предложений регионов (RPN). RPN — это CNN, используемая для создания высококачественных RoI.

1.3.3 Одноэтапные CNN

Данные нейросетевые алгоритмы не включают в себя этап поиска RoI на изображении.



Рисунок 1.9 – Схема работы одноэтапного алгоритма

Преимуществами одноэтапных алгоритмов являются их простота и относительно высокая скорость работы. К недостаткам же можно отнести более низкую точность детектирования объектов по сравнению с двухэтапными алгоритмами, а также меньшую гибкость алгоритма с точки зрения рассматриваемых изображений [7].

YOLO

Алгоритм YOLO (You Only Look Once) является одноэтапным и может непосредственно распознавать объекты, а также их местоположение с помощью сквозной обученной модели CNN [13].

В алгоритме YOLO входное изображение разбивается на фиксированное число сеток, а затем из каждой сетки предсказывается фиксированное число местоположений обрамляющих окон (bounding boxes) и вероятностей. Затем используется пороговое значение для выбора и определения местоположения объекта на изображении [13].

Основной проблемой YOLO является более низкая точность при распознавании больших и малых объектов, а так же присущая всем одноэтапным алгоритмам потеря точности распознавания в сравнении с двухэтапными алгоритмами [13].

Данное семейство алгоритмов включает в себя множество оптимизаций оригинального YOLO, среди которых:

- YOLOv2 (YOLO9000) [14];
- YOLOv3 [15];
- YOLOv3 tiny [16].

Последующие варианты алгоритма имеют малозначительные изменения, призванные улучшить его в конкретных задачах [17]. Последней опубликованной версией является YOLOv7 [18].

SSD

Алгоритм SSD (Single Shot MultiBox Detector) является альтернативным взглядом на YOLO — от последнего взяты наиболее удачные, с точки зрения авторов, решения, а так же привнесены свои идеи для повышения mAP и снижения времени работы алгоритма [19].

Отличительной особенностью является распознавание объектов за один прогон с помощью заданной сетки окон (default box) на пирамиде изображений, закодированной в сверточных тензорах. Таким образом алгоритм распознает как большие, так и маленькие объекты за один прогон сети [19].

Второй версией алгоритма является MobileSSD, который является комбинацией двух нейросетевых алгоритмов — MobileNet и SSD [20].

1.4 Ансамбли

Ансамбль — это набор слабых экспертов, выполняющих классификацию произвольного объекта $x \in X$, конечный результат которого рассчитывается на основе результатов работы составляющих слабых экспертов. Слабыми экспертами в ансамбле могут выступать как классификаторы, так и другие ансамбли.

Тем не менее, ансамбль на основе нескольких моделей, построенных независимо друг от друга, в общем случае будет иметь более низкое качество, чем ансамбль на основе моделей, построенных с использованием специальных алгоритмов [21].

К этим алгоритмам относятся бустинг и бэггинг, позволяющие существенно повысить качество классификации [21].

1.4.1 Бустинг

При использовании бустинга, каждый последующий алгоритм, входящий в ансамбль, стремится компенсировать недостатки композиции всех предыдущих алгоритмов [21].

Формальная постановка задачи

Пусть $h(a, \vec{x})$ — базовый классификатор, где \vec{x} — это вектор параметров. Необходимо найти следующий алгоритм:

$$H_T(a) = \sum_{t=1}^T b_t h(a, \vec{x}), \quad (1.4)$$

где $b_i \in \mathbb{R}$ — коэффициенты, при которых

$$Q = \sum_i L(H_T(a_i), y_i) \rightarrow \min, \quad (1.5)$$

где L — функция потерь.

Так как в общем случае процесс вычисления $\{(\vec{x}_t, b_t)\}_{t=1}^T$ нетривиален, решение находят пошагово:

$$H_t(a) = H_{t-1}(a) + b_t h(a, \vec{x}). \quad (1.6)$$

Схема бустинга представлена на рис. 1.10.

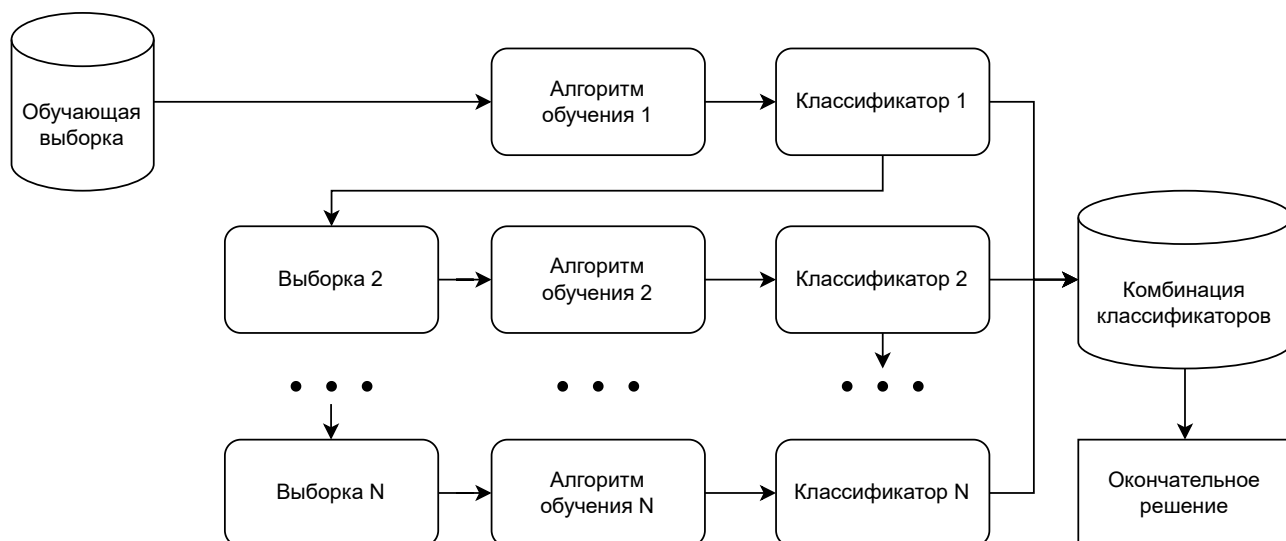


Рисунок 1.10 – Схематическое представление алгоритма бустинга

Алгоритмы

Среди наиболее часто используемых алгоритмов бустинга: AdaBoost и BrownBoost [21].

1.4.2 Бэггинг

В бэггинге все элементарные классификаторы обучаются и работают параллельно, т. е. независимо друг от друга. При этом обучающая выборка X разделяется на n выборок X_1, X_2, \dots, X_n , причем X_i и X_j могут пересекаться при любых $i, j \in 1 \dots n$.

Идея данного подхода заключается в том, что в отличие от бустинга классификаторы не исправляют ошибки друг друга, а компенсируют их при голосовании [21].

При этом результат голосования определяется посредством:

- консенсуса — все классификаторы дают одинаковый ответ;
- простого большинства;
- взвешивания — каждому классификатору присваивается вес, учитываемый при принятии решения.

Схема бэггинга представлена на рис. 1.11.

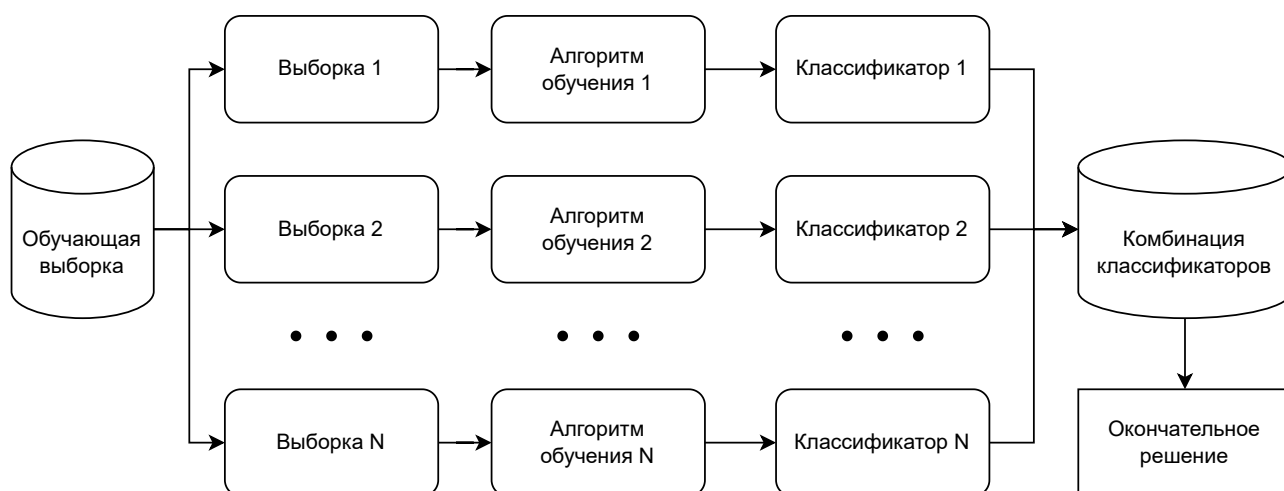


Рисунок 1.11 – Схематическое представление алгоритма бэггинга

1.5 Существующие методы

По результатам проведенного исследования рынка, в настоящий момент не представлены общедоступные системы обнаружения надводных объектов, в связи с чем невозможно определить методы, использующиеся в них.

Тем не менее, в последние годы тема распознавания надводных объектов является объектом множества исследований, предлагающих разнообразные подходы к решению поставленных задач. Данные методы и будут рассмотрены далее.

1.5.1 Разрешение изображений

В настоящее время существующие методы можно разделить на две группы по разрешению снимков:

- VHR — достигает разрешения в 0.5 метра и менее на 1 пиксель;
- MR — достигает разрешения в несколько метров на 1 пиксель.

Одна из проблем методов, основанных на VHR — доступность данных. В настоящее время изображения с таким разрешением могут быть получены лишь в нескольких источниках, использующих спутниковую съемку.

В тоже время, MR изображения могут быть получены, помимо прочего, с помощью съемки с дронов, что значительно удешевляет процесс сбора данных.

1.5.2 R-CNN

CNN данного семейства активно используются в задачах, допускающих повышение времени работы системы ради повышения точности распознавания, т.е. в задачах, не относящихся к системам реального времени.

Faster R-CNN

1.5.3 YOLO

Данное семейство CNN активно используется в задачах распознавания надводных объектов. В первую очередь, это связано с малыми затратами времени на обработку изображения, в сравнении с другими алгоритмами. Также, достоинством YOLO является, распознавать большое число объектов на одном изображении (вплоть до 9000 для версии YOLO 9000).

YOLOv3

YOLOv3 tiny

1.5.4 Параметры для сравнения

Для оценки работы нейронной сети используется величина AP (Average Precision), вычисляемая следующим образом:

$$AP = \frac{\text{кол-во верно распознанных объектов}}{\text{общее кол-во распознанных объектов}}. \quad (1.7)$$

Отметим, что AP вычисляется для объектов одного класса, в связи с чем для классификаторов с несколькими возможными классами объектов используется величина mAP (англ. mean AP) — среднее значение AP.

Кроме того, данную величину принято оценивать в зависимости от IoU (Intersection over Union), вычисляемую по формуле:

$$IoU = \frac{\text{площадь пересечения областей}}{\text{площадь объединения областей}}. \quad (1.8)$$

IoU описывает то, насколько предсказанные CNN обрамляющие окна близки к «истинным». Данная величина принимает значения в диапазоне $[0; 1]$, соответственно.

1.5.5 Сравнение существующих методов

1.6 Метод распознавания надводных объектов с аэрофотоснимков с использованием нейронных сетей

1.7 Формализованная постановка задачи

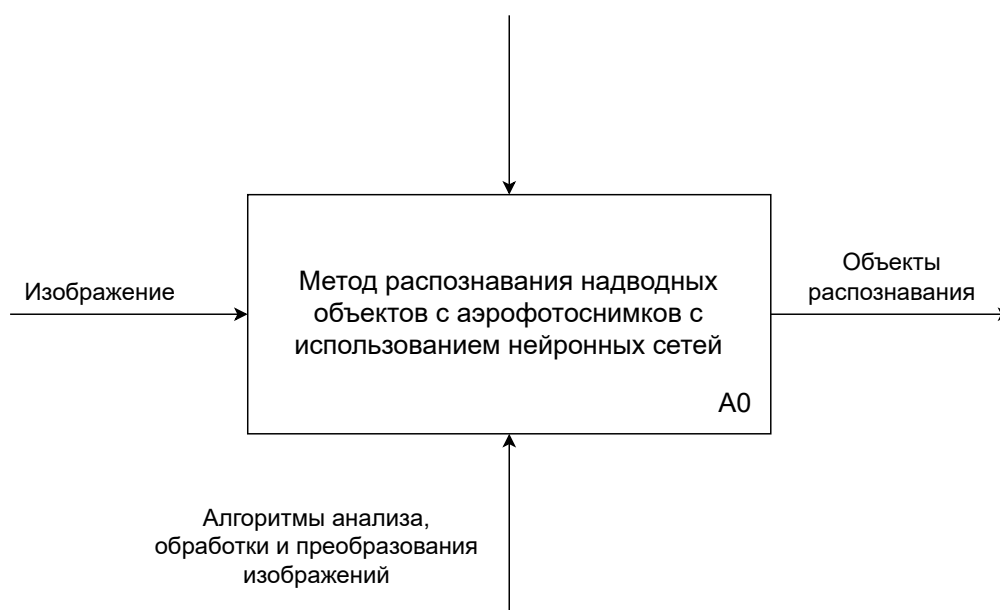


Рисунок 1.12 – Общий вид метода

1.8 Выбор данных для обучения модели

[22]

2 Конструкторский раздел

2.1 Метод распознавания

ЗАКЛЮЧЕНИЕ

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Lee S.-J., Roh M.-I., Oh M.-j.* Image-based ship detection using deep learning // Ocean Systems Engineering. — 2020. — Т. 10. — С. 415–434. — DOI: 10.12989/ose.2020.10.4.415.
2. *Z. Chen* Deep learning for autonomous ship-oriented small ship detection / Z. Chen [и др.] // Safety Science. — 2020. — Т. 130. — С. 104812.
3. *M. Marzuki* Fishing boat detection using Sentinel-1 validated with VIIRS Data / M. Marzuki [и др.] // IOP Conference Series: Earth and Environmental Science. — 2021. — Т. 925. — С. 012058. — DOI: 10.1088/1755-1315/925/1/012058.
4. *Sharkawy A.-.-N.* Principle of Neural Network and Its Main Types: Review // Journal of Advances in Applied & Computational Mathematics. — 2020. — Т. 7. — С. 8–19. — DOI: 10.15377/2409-5761.2020.07.2.
5. *Sharma S., Sharma S., Athaiya A.* ACTIVATION FUNCTIONS IN NEURAL NETWORKS // S. Sharma International Journal of Engineering Applied Sciences and Technology. — 2020. — Т. 4. — С. 310–316. — DOI: 10.33564/IJEAST.2020.v04i12.054.
6. *A. Ghosh* Fundamental Concepts of Convolutional Neural Network / A. Ghosh [и др.]. — 2020. — DOI: 10.1007/978-3-030-32644-9_36.
7. *Zhang H., Cloutier R.* Review on One-Stage Object Detection Based on Deep Learning // EAI Endorsed Transactions on e-Learning. — 2022. — Т. 7. — С. 174–181. — DOI: 10.4108/eai.9-6-2022.174181.
8. *Du L., Zhang R., Wang X.* Overview of two-stage object detection algorithms // Journal of Physics: Conference Series. — 2020. — Т. 1544. — С. 012033. — DOI: 10.1088/1742-6596/1544/1/012033.
9. *А. В. Бондаренко* АЛГОРИТМ НЕЙРОСЕТЕВОГО РАСПОЗНАВАНИЯ НАДВОДНЫХ ОБЪЕКТОВ В РЕАЛЬНОМ ВРЕМЕНИ / А. В. Бондаренко [и др.] // Известия Тульского государственного университета. Технические науки. — 2021. — Т. 1. — С. 19–33.

10. *R. Girshick* Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation / R. Girshick [и др.] // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. — 2013. — DOI: 10.1109/CVPR.2014.81.
11. *Girshick R.* Fast R-CNN. — 2015. — DOI: 10.1109/ICCV.2015.169.
12. *S. Ren* Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks / S. Ren [и др.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2015. — Т. 39. — DOI: 10.1109/TPAMI.2016.2577031.
13. *P. Jiang* A Review of Yolo Algorithm Developments / P. Jiang [и др.] // Procedia Computer Science. — 2022. — Т. 199. — С. 1066—1073. — ISSN 1877-0509. — DOI: <https://doi.org/10.1016/j.procs.2022.01.135>.
14. *Redmon J., Farhadi A.* YOLO9000: Better, Faster, Stronger // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2017. — С. 6517—6525. — DOI: 10.1109/CVPR.2017.690.
15. *Redmon J., Farhadi A.* YOLOv3: An Incremental Improvement. — 2018.
16. *D. Li* Yolo-tiny-MSD: A tiny neural network for object detection / D. Li [и др.] // Journal of Physics: Conference Series. — 2021. — Т. 1873. — С. 012073. — DOI: 10.1088/1742-6596/1873/1/012073.
17. *L. Hao* Enhanced YOLO v3 Tiny Network for Real-time Ship Detection from Visual Image / L. Hao [и др.] // IEEE Access. — 2021. — Т. PP. — С. 1—1. — DOI: 10.1109/ACCESS.2021.3053956.
18. *Wang C.-Y., Bochkovskiy A., Liao H.-y.* YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. — 2022. — DOI: 10.48550/arXiv.2207.02696.
19. *W. Liu* SSD: Single Shot MultiBox Detector / W. Liu [и др.]. — 2016. — DOI: 10.1007/978-3-319-46448-0_2.
20. *A. Younis* Real-Time Object Detection Using Pre-Trained Deep Learning Models MobileNet-SSD / A. Younis [и др.] // International Journal for Research in Applied Science & Engineering Technology (IJRASET). — 2020. — С. 44—48. — DOI: 10.1145/3379247.3379264.

21. *M. Ganaie* Ensemble deep learning: A review / M. Ganaie [и др.]. — 2021. — DOI: 10.48550/arXiv.2104.02395.
22. Kaggle. Airbus Ship Detection Challenge [Электронный ресурс]. — Режим доступа: <https://www.kaggle.com/c/airbus-ship-detection> (Дата обращения: 30.03.2023).

ПРИЛОЖЕНИЕ А