

УДК 004.853

Научная статья

DOI: 10.35330/1991-6639-2025-27-1-171-180

EDN: ZVKMKN

## Задача обнаружения надводных объектов в условиях плохой видимости

Т. К. Нгуен<sup>✉</sup>, М. Т. Нгуен

Российский технологический университет МИРЭА  
119454, Россия, г. Москва, Проспект Вернадского, 78

**Аннотация.** Статья посвящена задаче обнаружения и распознавания надводных объектов по данным видеонаблюдения в условиях плохой видимости, таких как дождь, снег, туман, сумерки. Наряду с проблемой ухудшения видимости имеются и другие факторы, затрудняющие решение этой задачи: изменение формы и размера изображения при изменении расстояния до объекта наблюдения и угла обзора видеокамеры. Обсуждается один из подходов к проблематике обработки данных видеонаблюдения – он состоит в совместном применении двух технологий: модели глубокого обучения YOLO и дискретного вейвлет-преобразования изображений. Экспериментальные результаты показывают, что предложенный алгоритм достигает высоких показателей точности, что делает его подходящим для применения в системах видеомониторинга беспилотниками.

**Ключевые слова:** задача обнаружения объектов, YOLO, вейвлет-преобразование, надводные объекты, дроны, условие плохой видимости

Поступила 03.02.2025, одобрена после рецензирования 12.02.2025, принята к публикации 13.02.2025

**Для цитирования.** Нгуен Т. К., Нгуен М. Т. Задача обнаружения надводных объектов в условиях плохой видимости // Известия Кабардино-Балкарского научного центра РАН. 2025. Т. 27. № 1. С. 171–180. DOI: 10.35330/1991-6639-2025-27-1-171-180

MSC: 68T07

Original article

## The task of detecting overwater objects in poor visibility conditions

T.C. Nguyen<sup>✉</sup>, M.T. Nguyen

MIREA – Russian Technological University  
119454, Russia, Moscow, 78 Vernadsky avenue

**Abstract.** The article is devoted to the problem of detection and recognition of overwater objects from video surveillance data in poor visibility conditions, such as rain, snow, fog, twilight. Along with the problem of visibility degradation there are other factors that complicate the solution of this problem: changes in the shape and size of the image when changing the distance to the object of observation and the angle of view of the video camera. One of the approaches to the problem of video surveillance data processing is discussed – it consists in the joint application of two technologies: YOLO deep learning model and discrete wavelet image transformation. Experimental results show that the proposed algorithm achieves high accuracy and efficiency, which makes it suitable for application in drone video monitoring systems.

**Keywords:** object detection problem, YOLO, wavelet transform, overwater objects, drones, poor visibility condition

**For citation.** Nguyen T.C., Nguyen M.T. The task of detecting overwater objects in poor visibility conditions. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2025. Vol. 27. No. 1. Pp. 171–180. DOI: 10.35330/1991-6639-2025-27-1-171-180

## ВВЕДЕНИЕ

Обнаружение и распознавание надводных объектов в условиях плохой видимости представляет собой сложную задачу, сталкивающуюся с несколькими ключевыми проблемами. Во-первых, ухудшение видимости из-за осадков (дождя, снега, тумана), а также работа в сумерках. Во-вторых, изменение контура судна вследствие перемены угла обзора и расстояния также усложняет задачу идентификации. Например, с высоты птичьего полета судно будет выглядеть иначе, чем вблизи от видеокамеры [1]. В-третьих, возможна нехватка обучающих данных, собранных в условиях плохой видимости. Наконец, сложность алгоритмов обработки видеоданных также играет важную роль с точки зрения возможности их использования в реальном времени.

Для детекции надводных объектов по данным видеонаблюдения могут применяться сверточные нейронные сети (CNN) [2]. Большой интерес представляют исследования по применению вейвлет-нейронных сетей в области компьютерного зрения [3].

В настоящей статье обсуждаются вопросы детекции надводных объектов по данным видеонаблюдения на основе модели внимания.

## АЛГОРИТМЫ ОБНАРУЖЕНИЯ ОБЪЕКТОВ YOLO

YOLO – это передовая система обнаружения объектов в реальном времени. Благодаря широкому диапазону доступных вариантов можно выбрать версию, наиболее подходящую для ваших нужд. Например, Tiny YOLO – это самый «компактный» вариант, который может работать быстро даже на смартфонах или Raspberry Pi.

Самое большое преимущество модели YOLO, собственно, отражено в названии – You Only Look Once. Эта модель накладывает на изображение сетку, разделяя его на ячейки. Каждая ячейка пытается предсказать координаты зоны обнаружения с оценкой уверенности для этих полей и вероятностью классов. Затем оценка уверенности для каждой зоны обнаружения умножается на вероятность класса, чтобы получить окончательную оценку.

Сверточные НС находят применение для обнаружения объектов на изображениях. Существующие архитектуры таких НС можно разделить на две категории: одноэтапные (one-stage) и двухэтапные (two-stage).

В двухэтапных нейросетевых алгоритмах на первом этапе осуществляется поиск подозрительных зон на изображении, которые могут содержать интересующие нас объекты, а на втором этапе выполняется классификация зон, найденных на первом этапе. Среди двухэтапных алгоритмов детектирования объектов на изображении наиболее известны Fast-RCNN и Faster-RCNN.

Одноэтапные нейросетевые алгоритмы детектирования объектов на изображении сразу нацелены на обнаружение объектов и отличаются простотой и высокой скоростью работы. К алгоритмам данного вида относится YOLO (You Only Look Once). Основная идея YOLO заключается в том, что на входном изображении наносится сетка с фиксированным размером ячейки и осуществляется предсказание ограничивающих рамок – якорей всех объектов.

## ВЕЙВЛЕТ-НЕЙРОННЫЕ СЕТИ

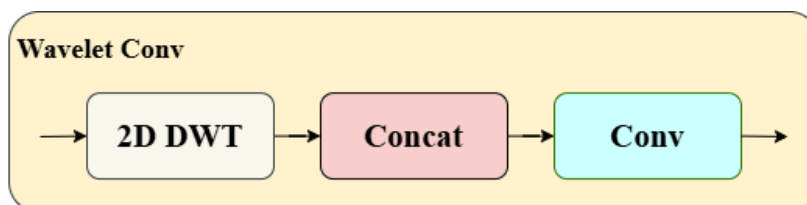
Поскольку растровое изображение является двумерным дискретным сигналом, то к нему применимы двумерные дискретные вейвлет-преобразования (DWT). В настоящее время вейвлеты широко применяются в области обработки растровых изображений, включая задачи фильтрации, реставрации и идентификации изображений.

Двумерное дискретное вейвлет-преобразование (DWT) получается в результате применения одномерного преобразования последовательно к строкам и столбцам изображения. Для изображения размером  $M \times N$  пикселей DWT дает четыре не перекрывающихся поддиапазона с разным разрешением: LL, LH, HL и HH. Поддиапазон LL представляет более грубое приближение исходного изображения, охватывая его низкочастотные компоненты. Поддиапазоны LH, LH и HH в совокупности охватывают высокочастотные элементы изображения: поддиапазоны LH и HL регистрируют изменения изображения по горизонтальной и вертикальной осям соответственно, а поддиапазон HH – изменения изображения по диагонали.

Посредством вейвлетов решают задачу подавления шумовой части сигнала без ухудшений его качества с последующим восстановлением исходной формы сигнала. В этих вейвлет-технологиях находят применение вейвлеты Хаара. В настоящей работе мы используем вейвлеты с базисом Наар.

Развитие методов вейвлет-анализа изображений и CNN закономерно приводит к идее их интеграции. Здесь возможны два подхода: в первом исходное изображение подвергается вейвлет-преобразованию и затем применяется сеть CNN [4], а во втором вейвлеты интегрируют в структуру сети посредством вейвлет-нейронной сети [2].

В вейвлет-нейронных сетях (Wavelet Neural Networks, WNN) в качестве функции активации используются вейвлеты [4]. Такие нейронные сети обладают хорошими показателями скорости и качества обучения и получают все более широкое распространение.



*Рис. 1. Блок вейвлет-нейронных сетей*

*Fig. 1. Block of wavelet neural networks*

После завершения вейвлет-преобразования данные проходят через сверточный слой, как показано на рисунке 1. Этот слой специально разработан для обработки увеличенного количества каналов, где количество входных каналов в четыре раза больше исходных. Это позволяет нейронной сети изучать признаки из различных поддиапазонов, используя информацию на разных частотах и с разных углов. В результате сеть способна обнаруживать признаки на разных масштабах – от мелких деталей до общей структуры объекта.

Вейвлет-нейронные преобразования предоставляют несколько важных преимуществ. Во-первых, они обеспечивают многомасштабный анализ, позволяющий сети обрабатывать информацию на разных уровнях детализации. Во-вторых, способность разделять информацию по частотам помогает сети лучше различать шум и полезную информацию. В-третьих, этот процесс создает форму естественного сжатия информации, уменьшая

сложность данных без потери важной информации. Наконец, значительно улучшается способность обнаружения признаков на разных масштабах, что способствует повышению эффективности сети в распознавании объектов разных размеров.

### МЕХАНИЗМЫ ВНИМАНИЯ В КОМПЬЮТЕРНОМ ЗРЕНИИ

Концепция внимания в контексте нейронных сетей и глубокого обучения была представлена в статьях [5, 6]. Механизм позволяет моделям обращать внимание на разные части входных данных с различной степенью важности. Основная идея механизма внимания в компьютерном зрении состоит в том, чтобы научить систему фокусироваться на важной информации, обращать внимание и игнорировать нерелевантную информацию. Внедрение механизмов внимания позволило добиться значительных успехов в компьютерном зрении.

Сверточный модуль внимания (*convolutional block attention module*) применяется для задач детектирования объектов на изображениях. Компонентку данного модуля представляет рис. 2. Он состоит из двух последовательно применяемых подмодулей – канального (применяется ко всем каналам одного пикселя с изображения) и пространственного (применяется ко всему изображению с фиксированным каналом).

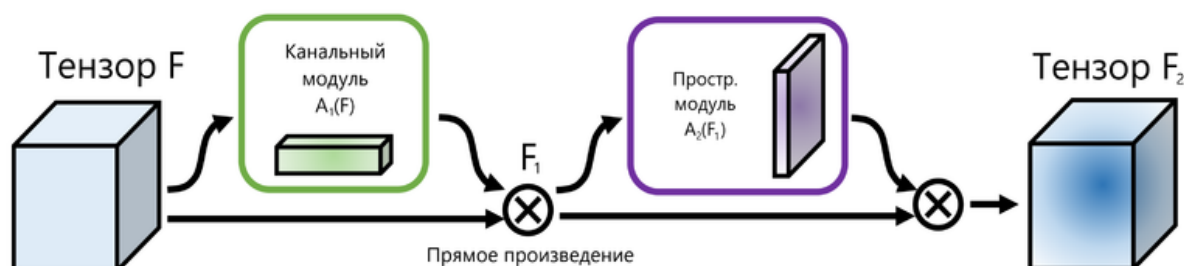


Рис. 2. Сверточный модуль внимания

Fig. 2. Convolutional attention module

Действие модуля иллюстрирует рис. 2. На вход блока подается множество признаков  $F \in \mathbf{R}^{C \times H \times W}$ , где  $C$  – число каналов,  $H$  – высота,  $W$  – длина изображения. Канальный подмодуль  $A_1(F)$  принадлежит множеству  $\mathbf{R}^{C \times 1 \times 1}$ , а пространственный  $A_2(F_1)$  принадлежит множеству  $\mathbf{R}^{1 \times H \times W}$ . Применение модуля можно описать соотношениями

$$F_1 = A_1(F) \otimes F, \quad F_2 = A_2(F_1) \otimes F_1. \quad (1)$$

Здесь символ  $\otimes$  обозначает поэлементное произведение, а тензоры  $A_1(F)$  и  $A_2(F_1)$  копируются вдоль недостающих измерений.  $F_1$  – тензор после применения канального модуля внимания,  $F_2$  – выходное множество признаков.

### СВЕРТОЧНЫЕ СЕТИ WAVELET-ATTENTION

Сети CNN используются в задачах классификации изображений. Однако на эффективность их работы существенно влияют шум, помехи, условия плохой видимости и другие негативные факторы. В этой связи интерес представляют сверточные сети Wavelet-Attention (WA-CNN), также реализующие механизм внимания [8].

WA-CNN декомпозирует карты признаков на низкочастотные и высокочастотные компоненты. Низкочастотные компоненты хранят основную информацию, а высокоча-

стотные – детали и шум. Discrete Wavelet Transform (DWT) применяется в обработке изображений и может быть интегрирован в CNN. DWT используется для подавления шума и сохранения структур признаков.

WA-CNN использует WA блок для получения детальной информации, содержащейся в высокочастотных компонентах.

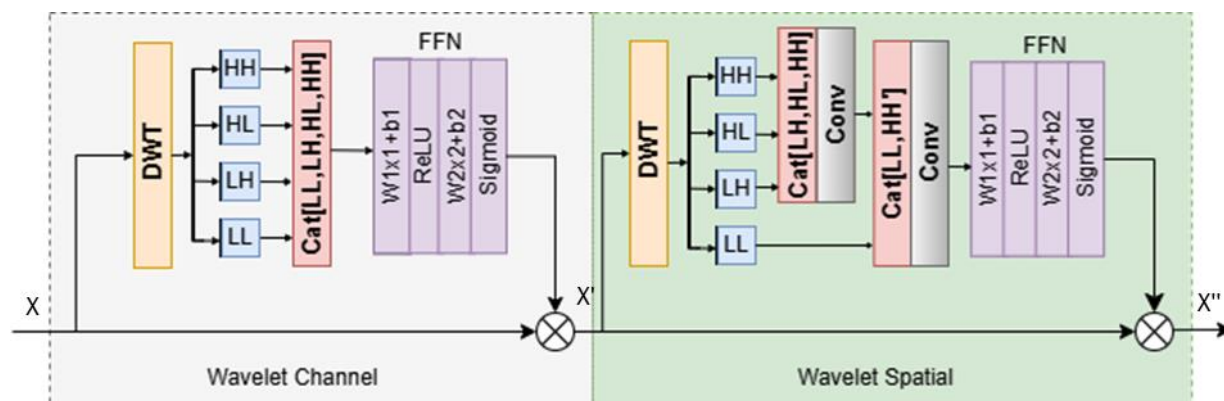


Рис. 3. Структура Wavelet Attention

Fig. 3. Wavelet Attention structure

Wavelet-Attention состоит из двух основных компонентов: модуля Wavelet Channel (WC) и модуля Wavelet Spatial (WS). Мы используем следующие обозначения:  $X$  является входом, который может быть либо исходным изображением, либо его отображением.  $X'$  обозначает выход WC, а  $X''$  – выход WS.

Модуль WC использует статистическую информацию, полученную из высоко- и низкочастотных характеристик поддиапазонов, разложенных с помощью DWT. Впоследствии статистические признаки получают путем агрегирования коэффициентов вышеупомянутых поддиапазонов.

Суть WC заключается в том, чтобы путем точных вычислений присваивать различные веса признакам из разных каналов вследствие чего элементы с более высокой значимостью оказывают большее влияние. Интеграция с вейвлет-технологией позволяет механизму внимания работать на более детальном уровне, что еще больше повышает эффективность подхода, основанного на канальном внимании.

А модуль WS использует преимущества как высокочастотных, так и низкочастотных компонентов сигнала. В процессе работы модуля происходит объединение высокочастотных поддиапазонов (LH, HL, HH) для формирования нового представления высокочастотных признаков, обозначаемого как HH'. Полученная пара поддиапазонов LL и HH' концептуально аналогична паре результатов среднего и максимального объединения, используемых в традиционных механизмах WS. Для восстановления исходной размерности признакового пространства объединенный результат LL и HH' пропускается через транспонированный сверточный слой. Завершающим этапом, как и в модуле WC, является применение механизма FFS (Feed Forward Network) перед формированием выходного сигнала.

Благодаря одновременным преимуществам высокой точности и скорости мы приняли YOLOv8 [7] в качестве базового алгоритма, показанного на рисунке 4. При этом заменили класс Conv, который имеет блок Conv2d с размером stride 2, на Wavelet Conv и добавили еще 3 модели Wavelet Attention, как показано на рис. 4.

Кроме классов Wavelet Conv и Wavelet Attention, архитектура сети включает в себя классы C2f и SPPF, подробнее структуры представлены на рис. 4.

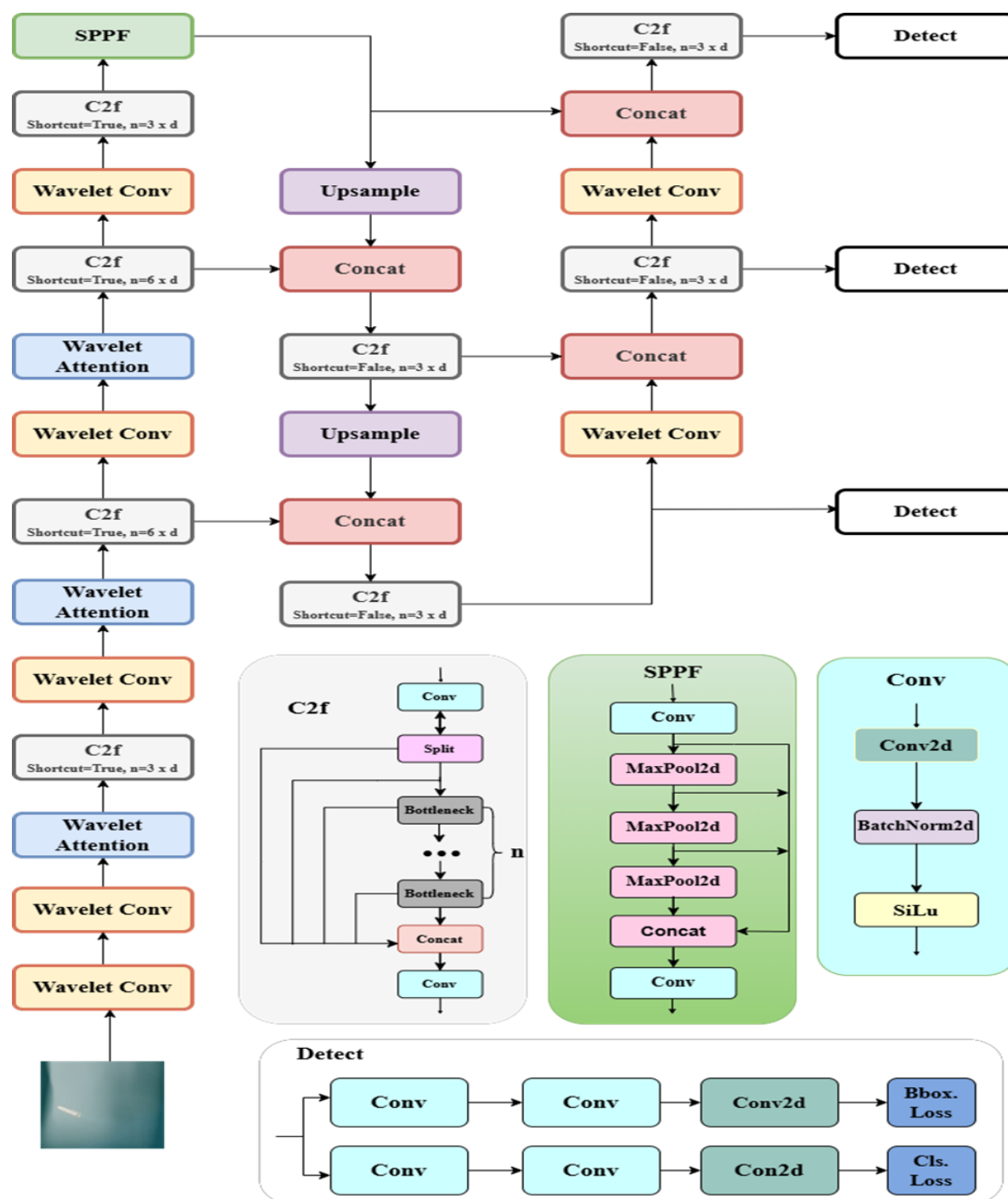


Рис. 4. Архитектура нейронных сверточных сетей с вейвлет-преобразованием

Fig. 4. Architecture of neural convolutional networks with wavelet transform

C2f (Cross Stage Partial Network – Faster) в этой архитектуре используется как важный компонент основы для эффективной обработки и извлечения признаков. В частности, C2f работает путем разделения входного потока на две ветви: одна ветвь проходит через серию блоков bottleneck, повторяющихся  $n$  раз для изучения сложных признаков, в то время как

другая ветвь проходит напрямую для сохранения исходной информации, после чего обе ветви объединяются. В модели блоки C2f настроены с различными параметрами, такими как `shortcut=True/False`, для управления короткими соединениями и `n=3xd` для определения количества повторений Bottleneck, что помогает сбалансировать сложность модели и способность извлечения признаков, одновременно оптимизируя скорость обработки по сравнению с традиционной версией CSP. Такая конструкция не только значительно снижает вычислительные затраты, но и сохраняет способность к изучению различительных признаков, что особенно полезно при обработке признаков из Wavelet Conv и подготовке их для последующих этапов детектирования.

А SPPF (Spatial Pyramid Pooling Fast) в YOLOv8 играет важную роль в улучшении способности модели к обнаружению объектов. Он позволяет модели обрабатывать пространственные характеристики на разных масштабах, что помогает более эффективно обнаруживать объекты, особенно в ситуациях с разнообразными размерами объектов. SPPF уменьшает размер входных данных, собирая информацию из различных областей изображения, не теряя при этом важные детали. Используя уровни пулинга, SPPF объединяет извлеченные характеристики с разных размеров, создавая более обобщенный выход. Это не только улучшает точность обнаружения объектов, но и оптимизирует время и ресурсы вычислений, что делает YOLOv8 более эффективным.

Основные преимущества такого подхода заключаются в следующем. Во-первых, интеграция вейвлет-преобразований позволяет эффективно разделять изображение на частотные поддиапазоны, сохраняя важные детали, такие как границы объектов, даже в условиях низкого разрешения. Это особенно полезно для обнаружения неопознанных объектов, которые часто теряются при использовании традиционных методов свертки и пулинга. Во-вторых, предложенный Wavelet-Attention использует информацию из частотных поддиапазонов для улучшения как Wavelet Channel, так и Wavelet Spatial, что позволяет модели более эффективно фокусироваться на важных областях изображения. В-третьих, интеграция этих модулей в YOLOv8 значительно улучшает точность обнаружения, особенно для неопознанного объекта, а также сохраняет высокую скорость обработки.

## ЭКСПЕРИМЕНТЫ И РЕЗУЛЬТАТЫ

В проведенных экспериментах решалась задача распознавания морских судов по их внешнему виду. В качестве базовой нейросети использовалась YOLOv8.

Рассматривались следующие типы судов: грузовые, военные, транспортные, круизные и нефтяные танкеры. Использовались аэрофотоснимки с БПЛА. Набор данных содержит 7000 изображений реальных судов 256 x 256 px в условиях плохой видимости (туман, дождь). Соотношение обучающего, тестового и проверочного множеств в экспериментальном наборе данных составляет 6 : 2 : 2. Число эпох при обучении сети – 300.

Программно-аппаратные средства: ОС Windows 11, графической процесс NVIDIA GeForce RTX 3050, библиотека глубокого обучения PyTorch (Python 3.9, версии Torch – 3.7 и 1.11.0).

Для оценки эффективности алгоритмов обнаружения объектов использовались показатели mAP – средняя точность, точность (Precision – P), полнота (Recall – R), F1 – мера (взвешенное среднее между точностью и полнотой), а также FPS – кадры

в секунду. FPS – важный показатель для оценки скорости работы алгоритма обнаружения целей, который указывает на количество кадров в секунду, обрабатываемых алгоритмом обнаружения.

Precision – это показатель правильности предсказания, отражающий долю точно предсказанных исходов. Recall представляет собой отношение числа правильно идентифицированных категорий к общему числу категорий, присутствующих в тестовом наборе. Метрика mAP0.5 рассчитывает среднее значение точности по всем категориям с учетом порога IoU (intersection over union), равного 0,5. Она служит для оценки эффективности модели при определенном пороге IoU. С другой стороны, mAP0.5:0.95 обозначает среднее значение mAP при различных пороговых значениях IoU в диапазоне от 0,5 до 0,95 с шагом в 0,05. Эта комплексная метрика отражает производительность модели в диапазоне пороговых значений IoU, обеспечивая более широкую оценку ее точности. Показатель F1 основан на среднем значении точности и полноты модели. Его значение варьирует от 0 до 1, где значение ближе к 1 указывает на то, что модель имеет лучший баланс между точностью и полнотой. Если одно из значений P и R смещено в сторону 0, F1 также будет близок к 0, что указывает на плохую работу модели. При совместном рассмотрении значений P и R показатель F1 помогает оценить точность модели в предсказании положительных категорий и ее чувствительность к положительным категориям.

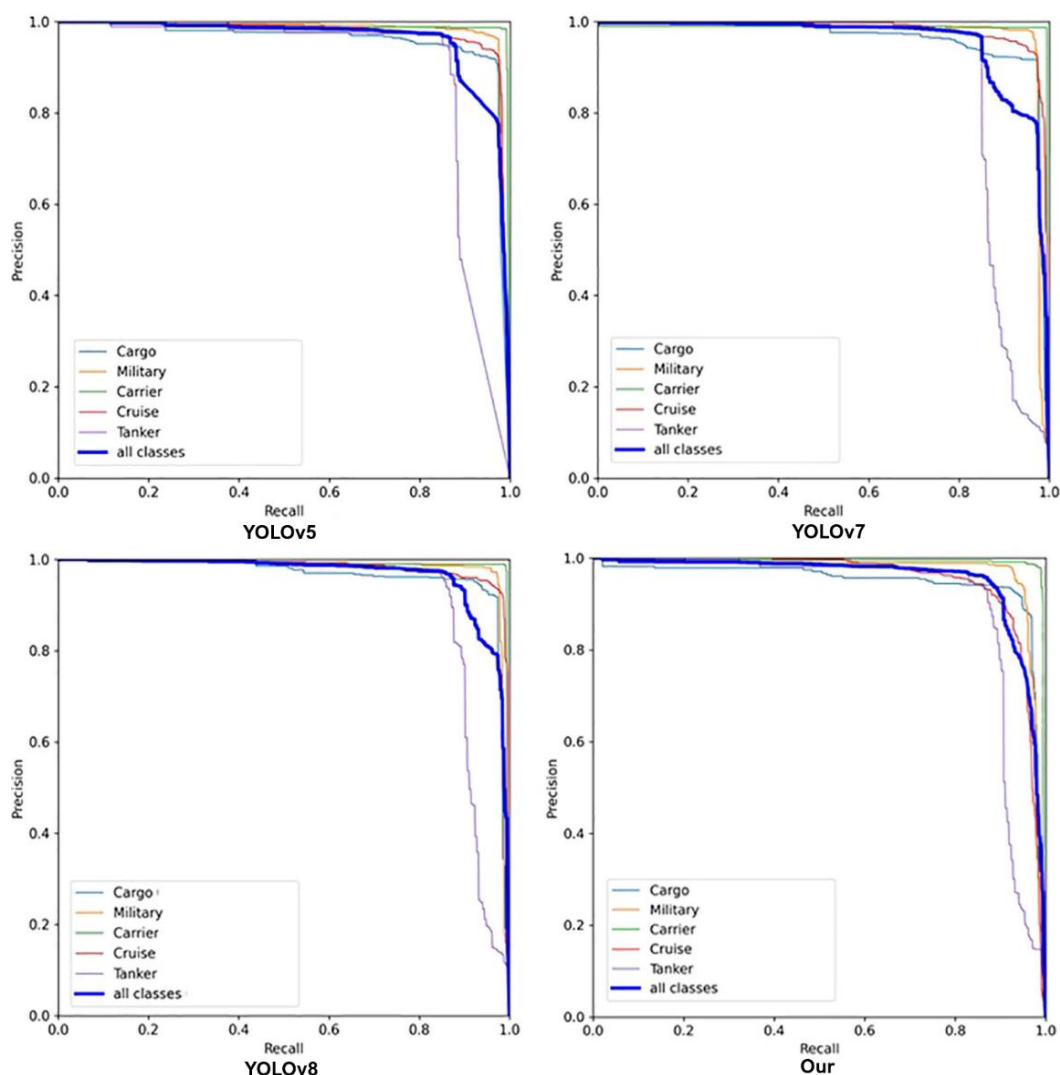
Результаты тестирования алгоритмов распознавания приведены в таблице 1. Они показывают, что предложенный алгоритм дает наилучшие результаты по скорости и точности обнаружения.

**Таблица 1.** Результаты тестирования алгоритмов обнаружения объектов

**Table 1.** Test results of object detection algorithms

| Алгоритм     | mAP0.5 (%) | mAP0.5:0.95 (%) | P (%) | R (%) | F1 (%) | FPS |
|--------------|------------|-----------------|-------|-------|--------|-----|
| Faster R-CNN | 83.16      | 45.09           | 86.49 | 79.00 | 82.58  | 83  |
| RetinaNet    | 80.08      | 42.20           | 80.88 | 74.47 | 77.54  | 58  |
| YOLOv5       | 85.87      | 49.43           | 86.53 | 80.51 | 83.41  | 120 |
| YOLOv7       | 86.54      | 50.35           | 87.24 | 83.69 | 85.43  | 98  |
| YOLOv8       | 89.15      | 54.47           | 89.3  | 86.73 | 88.00  | 131 |
| Наш алгоритм | 91.96      | 55.43           | 92.14 | 88.27 | 90.16  | 138 |

На рисунке 5 показаны кривые (P-R) для YOLOv5, YOLOv7, YOLOv8 и предложенного алгоритма. Кривые P-R представляют собой предсказания положительных образцов тестового набора при различных пороговых значениях. Чем больше площадь, занимаемая кривой P-R и осью координат, тем выше точность и отзыв алгоритма обнаружения. Приведенные данные показывают, что предложенный алгоритм в данной работе лучше, чем три алгоритма YOLOv5, YOLOv7 и YOLOv8 с точки зрения эффективности процесса обнаружения.



**Рис. 5.** Отношение значений P-R различных алгоритмов

**Fig. 5.** The ratio of P-R values of different algorithms

### ЗАКЛЮЧЕНИЕ

В статье осуждается задача распознавания надводных объектов по данным видеонаблюдения. Предполагается, что видеонаблюдение осуществляется с бортовой камеры БПЛА в условиях плохой видимости. Предлагается новый алгоритм обработки видеоданных, использующий технологии вейвлетных нейронных сетей, YOLO и механизм внимания. Проведенные вычислительные эксперименты показывают эффективность предложенных алгоритмических решений.

### СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Wang Z., Wang G., Yang W. Aircraft detection in remote sensing imagery with lightweight feature pyramid network. *MIPPR 2019: Automatic Target Recognition and Navigation*. 2020. Vol. 11429. Pp. 365–369. DOI:10.1117/12.2539372
2. Бондаренко В. А., Павлова В. А., Туликов В. А., Холод Н. Г. Алгоритм нейросетевого распознавания надводных объектов в реальном времени // Известия ТулГУ. Технические науки. 2021. № 1. С. 19–33. EDN: LBWTUH

Bondarenko V.A., Pavlova V.A., Tupikov V.A., Kholod N.G. Algorithm for neural network recognition of surface objects in real time. *Izvestiya TulGU. Tekhnicheskiye nauki* [Bulletin of Tula State University. Technical sciences]. 2021. No. 1. Pp. 19–33. EDN: LBWTUH. (In Russian)

3. Zhang Q., Benveniste A. Wavelet networks. *IEEE Transactions on Neural Networks*. 1992. Vol. 3. No. 6. Pp. 889–898.

4. De Silva D.D.N., Fernando S., Piyatilake I.T.S., Karunarathne A.V.S. Wavelet based edge feature enhancement for convolutional neural networks. *Eleventh International Conference on Machine Vision (ICMV 2018)*. 2019. Vol. 11041. DOI: 10.1117/12.2522849

5. Bahdanau D., Cho K., Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate. *International Conference on Learning Representations*. ICLR 2015. DOI: 10.48550/arXiv.1409.0473

6. Niu Z., Zhong G., Yu H. A review on the attention mechanism of deep learning. *Neurocomputing*, 2021. Vol. 452. Pp. 48–62. DOI: 10.1016/j.neucom.2021.03.091

7. Muhammad Y. What is Yolov8: an in-depth exploration of the Internal features of the next-generation object detector. *Computer Vision and Pattern Recognition*. August 29, 2024. DOI: 10.48550/arXiv.2408.15857

**Вклад авторов:** все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

**Contribution of the authors:** the authors contributed equally to this article. The authors declare no conflicts of interests.

**Финансирование.** Исследование проведено без спонсорской поддержки.

**Funding.** The study was performed without external funding.

### Информация об авторах

**Нгуен Тхань Конг**, аспирант кафедры автоматических систем, Российский технологический университет МИРЭА;

119454, Россия, Москва, Проспект Вернадского, 78;

congvietaimail.ru, ORCID: <https://orcid.org/0009-0005-9719-8731>

**Нгуен Минь Тьонг**, канд. техн. наук, доцент кафедры информатики, Российский технологический университет МИРЭА;

119454, Россия, Москва, Проспект Вернадского, 78;

nguen\_m@mirea.ru, ORCID: <https://orcid.org/0009-0002-7267-1121>, SPIN-код: 5480-9970

### Information about the authors

**Thanh Cong Nguyen**, Post-graduate Student of the Department of System Automatic, MIREA – Russian Technological University;

119454, Russia, Moscow, 78 Vernadsky avenue;

congvietaimail.ru, ORCID: <https://orcid.org/0009-0005-9719-8731>

**Minh Tuong Nguyen**, Candidate of Engineering Sciences, Associate Professor of the Department of Informatics, MIREA – Russian Technological University;

119454, Russia, Moscow, 78 Vernadsky avenue;

nguen\_m@mirea.ru, ORCID: <https://orcid.org/0009-0002-7267-1121>, SPIN-code: 5480-9970