

## Нейронные сети с вейвлет-преобразованием в задаче обнаружения надводных объектов при условиях плохой видимости

*А.Б. Филимонов, Нгуен Тхань Конг*

*Российский технологический университет МИРЭА, Москва*

**Аннотация:** В данной работе рассматривается проблема обнаружения и классификации надводных объектов в условиях плохой видимости, таких как дождь, снег и туман. Основное внимание уделяется применению современных алгоритмов глубокого обучения, в частности, архитектуры YOLO, для повышения точности и скорости обнаружения. Введение в проблему включает обсуждение ограничений, связанных с ухудшением видимости, изменением формы и размера объектов в зависимости от угла обзора, а также нехваткой обучающих данных. В работе также представлено использование дискретного вейвлет-преобразования для улучшения качества изображений и повышения устойчивости систем к неблагоприятным условиям. Экспериментальные результаты показывают, что предложенный алгоритм достигает высоких показателей точности и скорости, что делает его подходящим для применения в системах видеомониторинга беспилотниками.

**Ключевые слова:** YOLO, вейвлет- преобразование, надводные объекты, дроны, условие плохой видимости, преобразования Фурье, Наар.

### Введение

Обнаружение надводных объектов в условиях плохой видимости представляет собой сложную задачу, сталкивающуюся с несколькими ключевыми проблемами. Во-первых, ухудшение видимости из-за дождя, снега, тумана значительно снижает контрастность изображений, что затрудняет распознавание и классификацию объектов. Во-вторых, изменение формы и размера судна в зависимости от угла обзора и расстояния также усложняет задачу идентификации. Например, с высоты птичьего полета судна могут выглядеть иначе, чем с обычной точки зрения [1]. Кроме того, нехватка обучающих данных, собранных в условиях плохой видимости, ограничивает возможности алгоритмов, так как создание таких наборов

---

данных является сложной задачей из-за редкости неблагоприятных погодных условий. Сложность моделей также играет важную роль: для достижения обнаружения в реальное время требуется использование моделей, которые могут работать быстро и эффективно, однако сложные модели могут замедлять процесс обнаружения.

Таким образом, для эффективного обнаружения надводных объектов в условиях плохой видимости необходимо разрабатывать новые алгоритмы, собирать специализированные наборы данных и применять методы улучшения качества изображений, чтобы повысить устойчивость систем к неблагоприятным условиям.

### **Алгоритм обнаружения объектов YOLO**

В последние годы, с быстрым развитием искусственного интеллекта, глубокое обучение широко используется в обнаружении целей [2], которое имеет более высокую точность и скорость обнаружения, чем традиционные алгоритмы обнаружения целей.

YOLO (You Only Look Once)— это популярная архитектура для обнаружения объектов в изображениях и видео, разработанная Джозефом Редмоном и его коллегами. Основная идея YOLO заключается в том, чтобы рассматривать задачу обнаружения объектов как задачу регрессии, что позволяет модели предсказывать классы объектов и их координаты в одном проходе по изображению.

В качестве базовой сети мы выбрали YOLOv8 [3], имеющую простую структуру и стабильную точность обнаружения. Существует разница в ракурсе между камерой дрона и камерой возле земли. Угол съемки дрона — вид сверху вниз. С этой точки зрения форма судна сильно отличается от обычной перспективы, и другие плавучие объекты в море можно легко обнаружить как судна, что приводит к невысокой точности обнаружения объектов. Поэтому вопрос о том, как улучшить производительность

обнаружения целей YOLOv8, стал ключевым вопросом при обнаружении на изображениях, сделанных дронами. Стоит отметить, что размеры судов разных типов сильно различаются, и судна на видеороликах дрона будут принимать разные размеры и формы по мере того, как дрон летит на разной высоте и движется судно. Кроме того, вычислительная мощность влияет на эффект обнаружения в реальном времени, что можно решить, уменьшив вес модели. Вышеуказанные проблемы, несомненно, затрудняют обнаружение цели на видео с дронов, поэтому необходим быстрый и точный алгоритм обнаружения.

### **Вейвлет-преобразование**

Дискретное вейвлет-преобразование (DWT) представляется собой мощный метод преобразования изображений, созданный для улучшения оконного преобразования Фурье, демонстрирует замечательные возможности по управлению временным и частотным разрешением. Обычно используются два основных подхода: первый придерживается традиционной идеологии глубокого обучения, при которой вейвлет-преобразование применяется непосредственно к исходному изображению, служащему входным признаком для сети обнаружения, применились в работе [4]. И наоборот, второй подход интегрирует вейвлет-методы в саму структуру сети. Эти подходы еще больше увеличивают потенциал глубокого обучения в данной области. Применение двумерного DWT широко распространено в областях обработки и сжатия изображений. Среди различных вариантов, включая Haar, db1, coif1, sym4 и bior, мы используем исключительно вейвлет-функцию Haar из-за ее превосходных характеристик, которые доказали в работе [5].

### **Нейронные сети с вейвлет-преобразованием**

Используя это ценное свойство, мы вводим структуру Wavelet-Conv для эффективного использования этих преимуществ. Модуль Wavelet-Conv

---

состоит из DWT, операции конкатенации и нестридированного сверточного слоя, как показано на рисунке 1. Применение DWT к исходным с размерами  $M \times N$  приводит к получению четырех отдельных, не перекрывающихся поддиапазонов с разным разрешением: LL, LH, HL и HH [6]. Поддиапазон LL представляет собой более грубое приближение исходного изображения, охватывая его низкочастотные компоненты. И наоборот, поддиапазоны HL, LH и HH в совокупности захватывают высокочастотные элементы изображения. В частности, поддиапазоны LH и HL тщательно регистрируют изменения изображения по горизонтальной и вертикальной осям, соответственно, а поддиапазон HH наглядно иллюстрирует изменения по диагонали. Затем четыре поддиапазона были плавно интегрированы вдоль оси канала, в результате чего был получен новый тензор признаков размером  $M/2 \times N/2$ . Этот результирующий тензор был использован в качестве входных данных для последующей нестридовой сверточной операции.

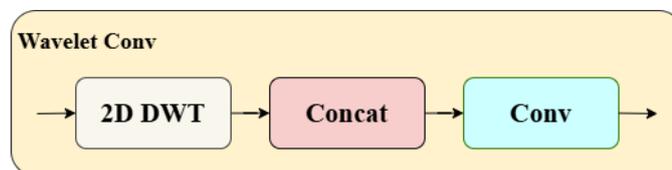


Рис. 1. – Блок-схема Wavelet-Conv

Новый сверточный блок на основе вейвлетов называется Wavelet-Conv, предназначенный для замены слоев strided Conv и pooling. Wavelet-Conv включает в себя слой вейвлет-преобразования и новый слой Conv. Сначала входной сигнал подвергается вейвлет-преобразованию, в результате чего образуются четыре дискретных вейвлет-поддиапазона. Затем эти поддиапазоны объединяются по каналному измерению для создания нового представления признаков. И наконец, сверточный слой обрабатывает объединенное представление. В свете того, что из частотной области можно извлекать сложные характеристики, естественно, следует, что усовершенствование механизма внимания [7]. Для решения этой проблемы

мы вводим структуру Wavelet-Attention, которая расширяет двумерное внимание в пространстве каналов до трехмерного в пространстве частотных полос каналов. Решая проблему, мы, тем не менее, используем другую структуру реализации.

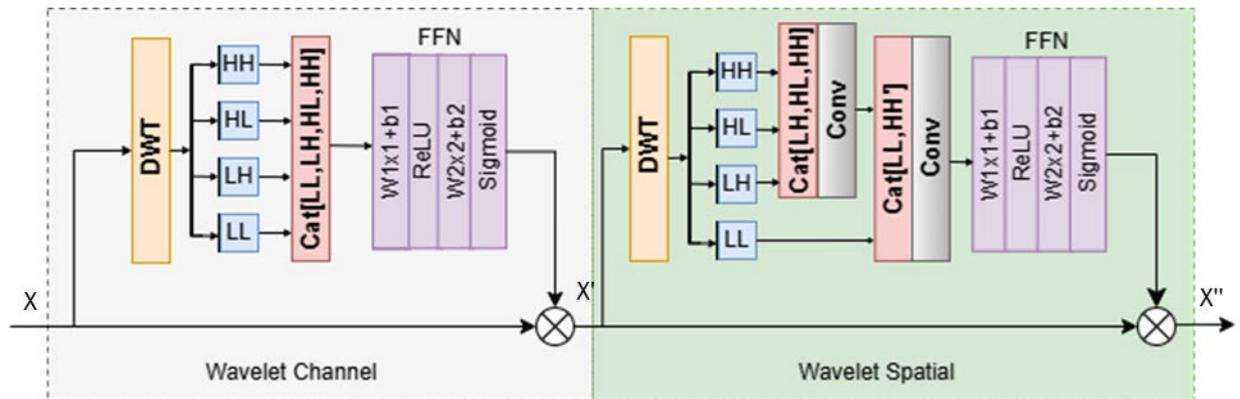


Рис. 3. – Структура Wavelet Attention.

Wavelet-Attention, состоит из двух основных компонентов: модуля вейвлет-канала (Wavelet Channel - WC) [8] и модуля вейвлет-пространства (Wavelet Spatial - WS) [9]. Эта новая конструкция направлена на улучшение канального и пространственного внимания с точки зрения частотного диапазона, подробнее на рисунке 2. Мы используем следующие обозначения:  $X$  представляет входной признак, который может быть либо исходным изображением, либо его отображением.  $X'$  обозначает выход WC, а  $X''$  - выход WS.

$$X' = M_{WC}(X) \odot X \quad (1)$$

$$X'' = M_{WS}(X') \odot X' \quad (2)$$

где  $\odot$  обозначает умножение по элементам,  $M_{WC}, M_{WS}$  - карты внимания WC и WS соответственно.

Модуль внимания WC использует статистическую информацию, полученную из высоко- и низкочастотных характеристик поддиапазонов, разложенных с помощью DWT. Впоследствии статистические признаки  $S_i$

получаются путем агрегирования коэффициентов вышеупомянутых поддиапазонов.

$$S_i = \mathbf{e} \mathbf{e} X_i(x, y), i = (LL, LH, HL, HH) \quad (3)$$

$x=1$   $y=1$

Эти статистические характеристики затем объединяются по измерениям канала, что приводит к созданию скаляров каналов. Затем перед выходом модуля вейвлет-внимания к каналу используется простая сеть с прямой связью (FFN). FFN состоит из двух полностью связанных слоев, включающих функцию активации ReLU, а затем функцию Sigmoid.

Суть внимания к каналу заключается в том, чтобы путем точных вычислений присваивать различные веса признакам из разных каналов, тем самым позволяя элементам с более высокой значимостью оказывать более выраженное влияние. Интеграция вейвлет-технологии позволяет механизму внимания работать на более детальном уровне, что еще больше повышает эффективность подхода, основанного на канальном внимании.

Концептуализация, лежащая в основе модуля WS, направлена на то, чтобы заменить результаты, полученные в результате обычных операций объединения среднего и максимального значений в структуре пространственного внимания. Вместо этого WS использует особенности объединения как высокочастотных, так и низкочастотных диапазонов.

Мы объединяем высокочастотную часть, а именно (LH, HL, HH), и получаем новое представление высокочастотных признаков, обозначаемое HH'. Этот партнер LL и HH' похож на партнеров среднего объединения и максимального объединения, встречающихся в классических механизмах пространственного внимания. Чтобы восстановить размерность признаков, мы подаем сшитый результат LL и HH' через транспонированный сверточный слой. Наконец, FFS используется перед выходом, как WC.

Суть пространственного внимания заключается в отображении пространственно дифференцированной информации через плоскостные объединяющие признаки, создаваемые двумя различными механизмами. Следовательно, теоретически, чем более очевидна физическая значимость этих двух объединяющих признаков и чем больше их диспропорция, тем выше способность отображать пространственно дифференцированную информацию. WS предлагает средства для создания объединяющих элементов с четко определенными физическими последствиями и значительными диспропорциями.

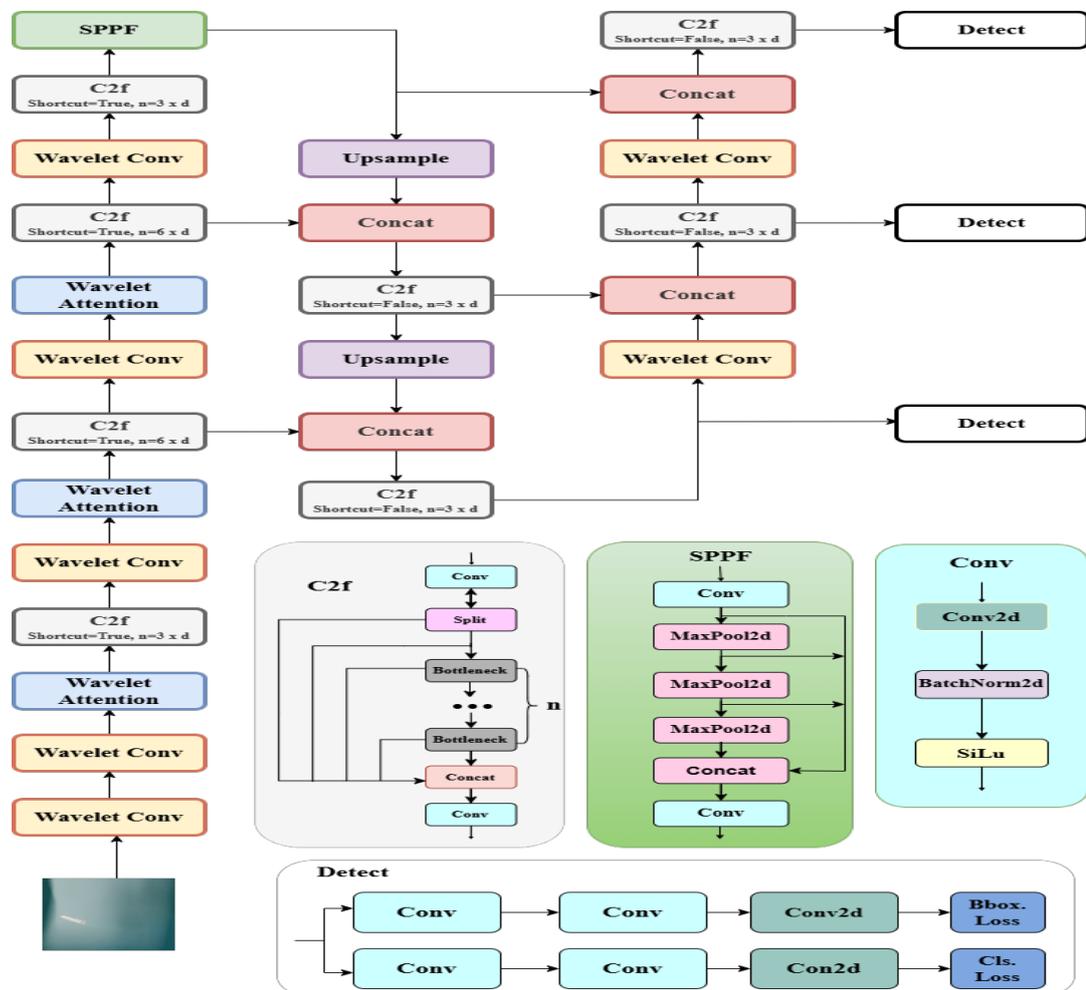


Рис. 3. – Архитектура нейронных сверточных сетей YOLO с вейвлет-преобразованием

Благодаря одновременным преимуществам высокой точности и скорости, мы приняли YOLOv8 в качестве базового алгоритма, показано на рисунке 3.

### Эксперименты и результаты

В этом эксперименте надводные объекты классифицируются на грузовые, военные, транспортные, круизные и нефтяные танкеры в соответствии с их внешним видом и использованием. В это работе собран набор данных, снятые дронами или аналогичные изображения с дронов. Набор данных содержит 7000 изображений реальных судов с размером 256 на 256 в условиях плохой видимости, включают в себя изображения тумана, дождя.

Соотношение обучающего, тестового и проверочного множеств в экспериментальном наборе данных составляет 6:2:2. После настройки размер пакета в нейронной сети в конечном итоге устанавливается равным 8, а количество эпох - 300.

Компьютер использовался в эксперименте, была установлена система Windows 11, а фреймворк глубокого обучения был построен на базе графического процессора NVIDIA GeForce RTX 3050. Эксперимент проводился с использованием инструментального программного обеспечения PyTorch. Версия Python составляла 3.9, а версии Torch - 3.7 и 1.11.0. Для обучения и выводов использовался только один GPU.

Для оценки эффективности алгоритмов обнаружения объектов использовались различные показатели оценки, включая среднюю точность (Average Precision - AP), точность (Precision - P), Полнота (Recall - R) и F1-score. Среднее значение точности (mAP) была принята для оценки точности алгоритмов обнаружения объектов. Показатель P использовался для измерения точности классификации алгоритмов, а показатель R - для измерения способности алгоритмов обнаружения к запоминанию. F1-score

может учитывать как P, так и R. Кадры в секунду (FPS) - важный показатель для оценки скорости работы алгоритма обнаружения целей, который указывает на количество кадров в секунду, обрабатываемых алгоритмом обнаружения. Формулы расчета представлены следующим образом:

$$P = \frac{TP}{TP + FP} \quad (4)$$

$$R = \frac{TP}{TP + FN} \quad (5)$$

$$AP = \frac{1}{T_0} \sum PR \quad (6)$$

$$F1 = 2 \frac{P \cdot R}{P + R} \quad (7)$$

$$mAP = \frac{\sum_{i=1}^n AP_i}{n} \quad (8)$$

где P - коэффициент точности, R - коэффициент отзыва, TP (True Positive) - ситуация, когда предсказание и метка являются суднами, FP (False Positive) - ситуация, когда предсказание - судно, а метка - фон; FN (False Negative) - ситуация, когда предсказание - фон, а метка - судно. И n - количество классов.

Таблица № 1

Результаты тестирования обнаружения объектов различными  
алгоритмами

Алгоритм	mAP0.5 (%)	mAP@0.5:0.95 (%)	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
--------------	-------	-------	-------	-------	-------	-----

Конкретные результаты тестирования, приведенные в таблице 1, показывают, что предложенный алгоритм достигает наилучших результатов по скорости и точности обнаружения, а его точность обнаружения выше, чем у Faster R-CNN, RetinaNet, YOLOv5, YOLOv7 и YOLOv8. Значение FPS нашего алгоритма составило 138. Скорость обнаружения нашего алгоритма выше, чем остальные. Результаты показывают, что алгоритм обнаружения предложенного алгоритма достигает оптимальных результатов между скоростью и точностью. Таким образом, алгоритм обнаружения судов, предложенный в нашем исследовании, подходит для применения на БПЛА.

На рисунке 4 показаны кривые точности и возврата (P-R) для YOLOv5, YOLOv7, YOLOv8 и предложенного алгоритма. Кривые P-R представляют собой предсказания положительных образцов тестового набора при различных пороговых значениях, при этом получают различные показатели точности и отзыва. Чем больше площадь, занимаемая кривой P-R и осью координат, тем выше точность и отзыв алгоритма обнаружения. После сравнения видно, что область, которую охватывает алгоритм, используемый в данном исследовании, больше, чем у других алгоритмов обнаружения объектов. Следовательно, алгоритм в данной работе лучше, чем три алгоритма YOLOv5, YOLOv7 и YOLOv8 с точки зрения эффективности обнаружения.

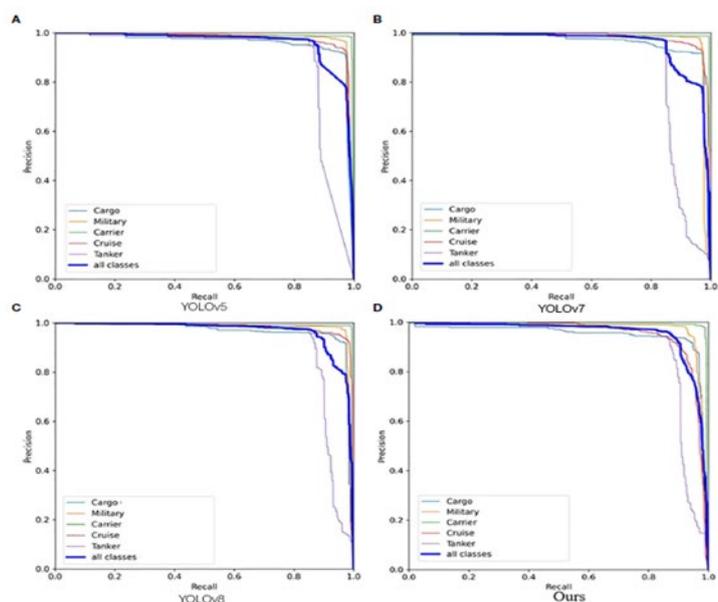


Рис. 4. – Отношение значений P-R различных алгоритмов

### Заключение

Исследование подтверждает, что применение алгоритмов глубокого обучения, таких как YOLOv8, в сочетании с вейвлет-преобразованием значительно улучшает результаты обнаружения надводных объектов в условиях плохой видимости. Проведенные эксперименты демонстрируют, что разработанный алгоритм превосходит традиционные методы по точности и скорости работы, что делает его эффективным инструментом для реальных приложений. В будущем необходимо продолжить работу над сбором специализированных наборов данных и оптимизацией моделей для дальнейшего повышения производительности систем обнаружения в сложных условиях.

### Литература

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, V. 11429, pp. 365-369.

2. Zhao X. B., Qin Y. Y. A review of object detection algorithms based on deep learning.// Automobile Applied Technology, 2021, V. 46, pp. 207-209.

3. 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. URL: doi.org/10.48550/arXiv.2408.15857

4. Yan B., Li J., Yang Z., Zhang X., Hao X. AIE-YOLO: Auxiliary information enhanced YOLO for small object detection //Sensors, 2022 , V. 22, №. 21, P. 8221.

5. Luyuan W., Yankui S. Image classification using convolutional neural network with wavelet domain inputs// IET Image Processing, 2022, V. 16, Is. 8. pp. 2037-2048.

6. Тхи Т. Ч. Б., Спицын В.Г. Разложение цифровых изображений с помощью двумерного дискретного вейвлет преобразования и быстрого преобразования Хаара. Известия Томского политехнического университета. 2011. Т. 318. № 5. С. 73-76.

7. Механизм внимания. URL: neerc.ifmo.ru/wiki/index.php?title=Механизм\_внимания. (Дата обращения: 11/12/2024)

8. Young J. C., Young W. L., Byung G. K. Wavelet Attention Embedding Networks for Video Super-Resolution // 25th International Conference on Pattern Recognition (ICPR) Milan, Italy. 2021. pp. 7314-7320.

9. Xibin S., Dingfu Z., Wei L., Haodong D., Yuchao D., Liangjun Z. WSAMF-Net: Wavelet Spatial Attention-Based MultiStream Feedback Network for Single Image Dehazing // IEEE Transactions on Circuits and Systems for Video Technology. 2023. Vol. 33. № 2. pp. 575-588.

## References

1. Wang Z., Wang G., Yang W. MIPPR 2019: Automatic Target Recognition and Navigation, 2020, V. 11429, pp. 365-369.



2. Zhao X.B., Qin Y.Y. Automobile Applied Technology. 2021. Т. 46. pp. 207-209.
3. Muhammad Y. Computer Vision and Pattern Recognition. August 29, 2024. URL: [doi.org/10.48550/arXiv.2408.15857](https://doi.org/10.48550/arXiv.2408.15857)
4. Yan B., Li J., Yang Z., Zhang X., Hao X. Sensors, 2022 , V. 22, №. 21, P. 8221.
5. Luyuan W., Yankui S. IET Image Processing. 2022, V 16, Issue 8. pp. 2037-2048.
6. Txi T. C. B., Spicyn V.G. Izvestiya Tomskogo politexnicheskogo universiteta , 2011, Vol. 318, №. 5, pp. 73-76.
7. Механизм внимания [Mechanism of attention]. URL: [neerc.ifmo.ru/wiki/index.php?title=Механизм\\_внимания](http://neerc.ifmo.ru/wiki/index.php?title=Механизм_внимания). (Accessed 11/12/2024).
8. Young J. C., Young W. L., Byung G. K. 25th International Conference on Pattern Recognition (ICPR) Milan, Italy. 2021. pp. 7314-7320.
9. Xibin S., Dingfu Z., Wei L., Haodong D., Yuchao D., Liangjun Z. IEEE Transactions on Circuits and Systems for Video Technology. 2023. Vol. 33. № 2. P. 575-588.

**Дата поступления: 19.11.2024**

**Дата публикации: 2.01.2025**