



Обнаружение объектов в аэронавигации с использованием вейвлет-преобразования и сверточных нейронных сетей: первый подход

¹Х.М. Фортуна-Сервантес, ORCID: 0000-0002-9229-3159
<juan.manuel.fortuna@hotmail.com>

¹М.Т. Рамирес-Торрес, ORCID: 0000-0002-7457-7318 <tulio.torres@uaslp.mx>

²Х. Мартинес-Карранса, ORCID: 0000-0001-8123-0008 <carranza@inaoe.mx>

¹Х.С. Мургуа-Ибарра, ORCID: 0000-0001-7239-8968 <ondeleto@uaslp.mx>

¹М. Мехиа-Карлос, ORCID: 0000-0003-2872-9461 <marce.mejia@uaslp.mx>

¹Автономный университет Сан-Луис-Потоси,
Мексика, 78000, Сан-Луис-Потоси, Альваро Обрегон, 64

²Национальный институт оптической и электронной астрофизики,
Мексика, 72840, Пуэбла, Тонанцинтла, Луис Энрике Эрро, 1

Аннотация. Предлагается первый подход, основанный на применении вейвлет-анализа при обработке изображений с целью обнаружения объектов с повторяющимися чертами и двойной классификации в плоскости изображения, в частности, для навигации в симулируемых средах. На сегодняшний день стало привычным использовать алгоритмы на основе сверточных нейронных сетей (Convolutional Neural Networks, CNN) для обработки изображений, полученных с бортовых камер беспилотных летательных аппаратов (Unmanned Aerial Vehicles, UAV), в пространственной области, что способствует решению задач обнаружения и классификации. Архитектура CNN позволяет обучать сеть, используя в качестве входных данных изображения без предварительной обработки. Это позволяет извлекать характерные признаки изображения. Тем не менее, в этой работе мы утверждаем, что спектральные характеристики изображений на разных частотах, низких и высоких, также влияют на производительность CNN во время обучения. Мы предлагаем архитектуру CNN, дополненную двумерным дискретным вейвлет-преобразованием как методом выделения признаков. Такая информация улучшает способность сети к обучению, устраняет переобучение и обеспечивает более высокую эффективность при обнаружении цели.

Ключевые слова: сверточная нейронная сеть; вейвлет-анализ; обнаружение объектов; дрон; классификация объектов; среда симуляции Gazebo

Для цитирования: Фортуна-Сервантес Х.М., Рамирес Торрес М.Т., Мартинес-Карранса Х., Мургуа-Ибарра Х.С., Мехиа Карлос М. Обнаружение объектов в аэронавигации с использованием вейвлет-преобразования и сверточных нейронных сетей: первый подход. Труды ИСП РАН, том 33, вып. 2, 2021 г., стр. 149-162. DOI: 10.15514/ISPRAS-2021-33(2)-9

Благодарности. Х.М. Фортуна-Сервантес – докторант CONACYT (Мексика) по программе «Ciencias Aplicadas» в IICO-UASLP. Мы благодарим INAOE за предоставление условий для проведения исследовательской стажировки, в рамках которой была выполнена часть этой работы.

Object Detection in Aerial Navigation using Wavelet Transform and Convolutional Neural Networks: A first Approach

¹J.M. Fortuna-Cervantes, ORCID: 0000-0002-9229-3159 <juan.manuel.fortuna@hotmail.com>

¹M.T. Ramírez-Torres, ORCID: 0000-0002-7457-7318 <tulio.torres@uaslp.mx>

²J. Martínez-Carranza, ORCID: 0000-0001-8123-0008 <carranza@inaoe.mx>

¹J.S. Murguía-Ibarra, ORCID: 0000-0001-7239-8968 <ondeleto@uaslp.mx>

¹M. Mejía-Carlos, ORCID: 0000-0003-2872-9461 <marce.mejia@uaslp.mx>

¹Universidad Autónoma de San Luis Potosí,
Álvaro Obregón #64, Col. Centro, San Luis Potosí, C.P. 78000, México

²Instituto Nacional de Astrofísica Óptica y Electrónica,
Luis Enrique Erro # 1, Tonantzintla, Puebla, C.P. 72840, México

Abstract. This paper proposes a first approach based on wavelet analysis inside image processing for object detection with a repetitive pattern and binary classification in the image plane, in particular for navigation in simulated environments. To date, it has become common to use algorithms based on convolutional neural networks (CNNs) to process images obtained from the on-board camera of unmanned aerial vehicles (UAVs) in the spatial domain, being useful in detection and classification tasks. CNN architecture can receive images without pre-processing, as input in the training stage. This advantage allows us to extract the characteristic features of the image/ Nevertheless, in this work, we argue that characteristics at different frequencies, low and high, also affect the performance of CNN during training. Thus, we propose a CNN architecture complemented by the 2D discrete wavelet transform, which is a feature extraction method. The information improves the learning capacity, eliminates the overfitting, and achieves a better efficiency in the detection of a target.

Keywords: CNN; Wavelet Analysis; Object Detection, Drone, Object Classification, Gazebo Simulation Environment.

For citation: Fortuna-Cervantes J.M., Ramírez-Torres M.T., Martínez-Carranza J., Murguía-Ibarra J.S., Mejía-Carlos M. Object Detection in Aerial Navigation using Wavelet Transform and Convolutional Neural Networks: A first Approach. *Trudy ISP RAN/Proc. ISP RAS*, vol. 33, issue 2, 2021, pp. 149-162 (in Russian). DOI: 10.15514/ISPRAS-2021-33(2)-.

Acknowledgments. J.M. Fortuna-Cervantes is a doctoral fellow of CONACYT (México) in the program of «Ciencias Aplicadas» at IICO-UASLP. We thank INAOE for giving the facilities to carry out the research internship where part of this work was done.

1. Введение

В последние годы в различных областях робототехники, в частности для автономной навигации беспилотных летательных аппаратов (Unmanned Aerial Vehicles, UAV), или дронов, активно используются алгоритмы визуального сервоуправления [1, 2]. В некоторых случаях возникает необходимость обеспечивать более высокую надежность при решении задач обнаружения, поскольку это оказывает существенное влияние на автономность робота. Для визуального восприятия, подобно цвету и форме, важна текстура изображения, так как она обеспечивает информацию о структуре поверхности и объектов на изображении [3]. Однако иногда обнаружение и отслеживание объекта на сцене бывает осложнено изменениями в освещении, масштабе и перспективе камеры [4, 5]. Поэтому использование таких методов обработки изображений, как вейвлет-анализ, глубокие нейронные сети или сверточные нейронные сети (Convolutional Neural Networks, CNN), стало очевидной альтернативой традиционным подходам [6-8].

В этой работе мы предлагаем подход, основанный на вейвлет-анализе как методе извлечения спектральных характеристик в сочетании с архитектурой CNN. Архитектура CNN явно допускает, что входными данными являются изображения в вейвлет-домене. Использование изображений в вейвлет-домене повышает способность к обучению на этапе обучения по

сравнению со случаем, когда для обучения используются изображения только в пространственном домене. Кроме того, если обучающий датасет невелик, то это позволяет избежать переобучения при обобщении. В результате при использовании для навигации оказывается возможна валидация модели обучения, в которой дрон может распознавать объекты с повторяющимися чертами (например, текстурой) и обучаться на этом. Модель обнаружения изображения с бортовой камеры дрона кадр за кадром и классифицирует изображения; предопределены два выходных класса: объект с текстурой (Texture) или не объект с текстурой (NotTexture).

Статья организована следующим образом: в разд. 2 описываются работы по родственной тематике; в разд. 3 представлена предлагаемая методология; разд. 4 демонстрирует экспериментальную часть и результаты; в разд. 5 представлено заключение.

2. Работы по родственной тематике

В последние годы проблема обнаружения объектов в приложениях UAV стала объектом активного исследования [8, 9]. Техника визуальной обработки, которая в настоящее время дает отличные результаты, основывается на архитектуре глубокого обучения. В некоторых работах по автономной навигации предлагается использовать CNN в реальных и моделируемых средах; например, в работе [10] авторы предлагают методологию обнаружения и избегания препятствий. Они используют предварительно обученную сеть AlexNet [11], обучение которой производит нейронная сеть меньшего размера.

Метод, предложенный в [12], – это нейронная архитектура YOLO, которая хорошо зарекомендовала себя в области обнаружения объектов, обрабатывая изображения в реальном времени со скоростью 45 кадров в секунду. Помимо этого, для решения задач распознавания объектов авторы работы [13] предлагают подход, основанный на глубоком обучении, для надежного определения центра объекта. Генерация линии прямой видимости в качестве направляющей позволяет избежать столкновений с другими объектами, которые могут происходить вследствие изменения таких условий, как освещение, геометрия объекта и наложение в плоскости изображения.

С другой стороны, в ряде проектов в области визуальной обработки используются методы глубокого обучения и вейвлет-анализа. Например, применительно к классификации изображений, метод, предложенный в [14], преобразует изображения из базы данных CIFAR-10 и KDEF в вейвлет-домен, получая таким образом временные и частотные характеристики. Различные представления изображений используются в нескольких архитектурах CNN; эта комбинация информации в вейвлет-домене обеспечивает более высокую эффективность обнаружения и более короткое время выполнения по сравнению с использованием только пространственного домена.

Авторы [15] предлагают другую альтернативу – вейвлет-пулинг как слой в архитектуре CNN. Этот метод разделяет карты признаков на два поддиапазона, отбрасывая карты первого уровня, чтобы уменьшить размер карт признаков. Преобразование карты признаков в вейвлет-домен улучшает классификацию изображений базы данных MNIST. Кроме того, этот подход поддерживает структурное сжатие данных, уменьшает образование неровных краев и других дефектов в изображении.

Объединение в архитектуре CNN инфракрасных и видимых фотографий обеспечивает эффективный метод обнаружения. Объединение основывается на декомпозиции изображения на основе вейвлет-анализа, и реконструируемое изображение лучше воспринимается зрительной системой человека [16].

В работе [17] представлены два метода выделения границ изображений с целью их классификации. Первый метод декомпозирует изображение на основе вейвлет-преобразования, а затем реконструирует их ограниченным образом. Второй метод, который

создает улучшенные изображения для ввода в нейронную сеть с использованием модулей локальных максимумов вейвлет-коэффициентов. Оба метода применяются для предварительной обработки изображений.

Говоря о классификации текстур в приложениях обработки изображений, авторы [18] предлагают вейвлет-CNN для обеспечения возможности обобщения спектральной информации, которая теряется в обычных CNN. Эта информация полезна для классификации текстур, поскольку обычно содержит достаточно сведений о форме объекта. Модель позволяет иметь меньше параметров, чем в традиционных CNN, и поэтому проводить обучение с меньшим объемом памяти.

Таким образом, обзор современных публикаций показывает, что алгоритмы вычислительного интеллекта улучшают стратегии обнаружения в приложениях UAV. Достигается приспособляемость к изменениям окружающей среды, освещенности, масштаба и проч. В отличие от работ, упомянутых выше, в настоящей работе основное внимание уделяется архитектуре CNN в сочетании с вейвлет-анализом. В результате такого подхода дрон лучше обнаруживает объекты с повторяющимися чертами, например, текстурой. Кроме того, дрон использует спектральную информацию о форме объекта, что увеличивает способность к обучению. Также исключается переобучение на этапе обучения, в отличие от случая наличия только пространственных данных и использования традиционной архитектуры CNN.

3. Материалы и методы

3.1 Кратномасштабный анализ

Алгоритм кратномасштабного анализа Малла (Stéphane Georges Mallat, Multiresolution Analysis, MA) обеспечивает связь между вейвлетами и наборами фильтров [19-21]. Кратномасштабная декомпозиция двумерной функции или изображения представляется сериями приближений и деталей во вспомогательных изображениях. На первом уровне декомпозиции применяются два фильтра соответственно – низкочастотный (h) и высокочастотный (g), за каждым из которых следует операция субдискретизации с коэффициентом 2, как показано на рис. 1.

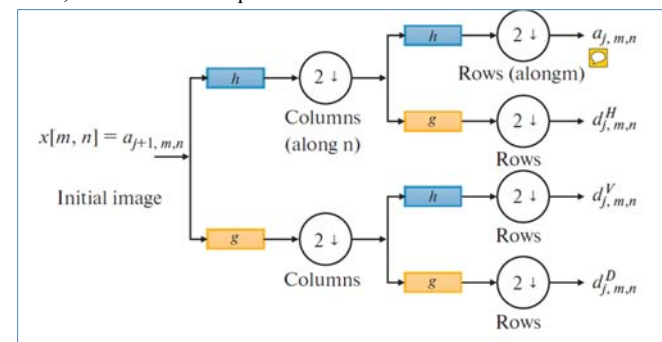


Рис 1. Первый уровень декомпозиции, примененный к изображению с использованием набора фильтров [22]

Fig 1. The first level of decomposition applied to an image using the filter bank [22]

Результат применения трех уровней вейвлет-декомпозиции к изображению ($x[m, n]$) размером $M \times N$ пикселей показан на рис. 2. После этого двумерный сигнал проходит через структуру набора фильтров, показанную на рис. 1. Получаются четыре вспомогательных

изображения с $M/2$ строками и $N/2$ столбцами; то есть каждое из четырех подизображений имеет четверть пикселей исходного изображения. Аппроксимация вспомогательного изображения достигается аппроксимационными вычислениями сначала по строкам, а затем по столбцам исходного изображения. Это подизображение представляет собой осредненную версию изображения ($x[m,n]$) с одной четвертой разрешения и аналогичными статистическими свойствами, аналогичными исходному сигналу [23]. Остальные подизображения показывают специфические характеристики исходного изображения в определенном направлении, то есть обеспечивают горизонтальный, вертикальный и диагональный коэффициенты детализации. Такое же преобразование формы сигнала применяется аппроксимированному подизображению для определения следующего уровня декомпозиции. Снова получаются четыре подизображения, но теперь с $M/2^2$ строками и $N/2^2$ столбцами. Эта итерация повторяется до достижения желаемого уровня разрешения или до уровня, допускаемого размерами изображения [22].

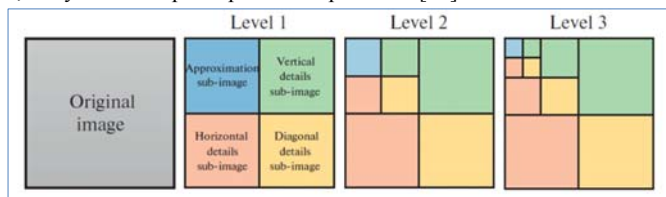


Рис. 2. Процесс декомпозиции с применением трех уровней набора фильтров, результатом которой является некоторая аппроксимация и детализация подизображений

Fig 2. Decomposition process applying three levels of the filter bank, which results are some approximation and detail sub-images

В общем случае кратномасштабное разложение двумерного сигнала выявляет различия в уровнях разрешения. Детали показываются в различных ориентациях, из чего следует, что метод двумерного дискретного вейвлет-преобразования (Two-Dimensional Discrete Wavelet Transform, 2D-DWT) хорошо подходит для обнаружения важной информации из исходного двумерного сигнала или изображения. Это существенно для таких задач обработки изображений, как обнаружение границ, распознавание изображений, классификация текстур и повышение качества изображений [3].

3.2 Сверточные нейронные сети

CNN широко используются при решении задач компьютерного зрения. CNN формируются из нейронов и обладают параметрами в виде весов и смещений, которые позволяют сети обучаться [24-26]. Эти сети состоят из входного и выходного слоев, а также нескольких скрытых слоев, некоторые из которых являются сверточными, откуда и происходит название этого вида нейронных сетей [24]. Операция свертки выполняется внутри сети на всех картах признаков сверточного слоя. Кроме того, при распространении весов и смещений эти операции применяются в направлении от первого входного слоя к последнему скрытому слою [14]. Формула (1) задает математическое представление распределения веса желаемого фильтра:

$$y_{ij} = \sigma \left(b + \sum_{l=0}^{n-1} \sum_{m=0}^{n-1} W_{l,m} a_{j+l,k+m} \right), \quad (1)$$

где $W_{l,m}$ представляет распределенные веса, b – смещение, $a_{j+l,k+m}$ – функция активации в заданном положении, n – размер окна фильтра.

Как следствие, использование сверточных слоев позволяет CNN обучаться разным уровням абстракции. Отличительной особенностью сверточной нейронной сети является наличие

явного предположения, что входными данными являются изображения; это позволяет закладывать в архитектуру специальные возможности распознавания конкретных элементов [27]. В общем случае сети с несколькими слоями могут выявлять во входных данных более сложные структуры. При наличии в CNN нескольких слоев увеличивается число параметров для обучения и поиска наилучшего решения. Имеются архитектуры глубоких нейронных сетей, подобные VGG16 [28], VGG 19 [28], AlexNet [11], ConvNet [29], SD [30], YOLO [12], с положительными результатами в областях классификации изображений и обнаружения объектов [31,32].

Тем не менее, в данном исследовании, в котором мы хотели обойтись небольшим датасетом для обучения модели, мы решили использовать архитектуру CNN, на которой выполнялась классификация изображений собак и кошек [33]. Использованный нами датасет не размещен в библиотеке Keras и поэтому нам пришлось создать наш собственный экспериментальный датасет. CNN представляет собой стек двумерных сверточных слоев с функцией активации блока линейной ректификации (Rectified Linear Unit, ReLU), чередующихся со слоями MaxPooling 2D. Кроме того, значение глубины скрытых слоев постепенно увеличивается с 32 до 128, в то время как размер карт признаков уменьшается с 62×62 до 2×2 , как показано на рис. 3. При использовании бинарной классификации сеть завершается одним блоком (сжатый слой размером 1) и сигмоидальной функцией активации.

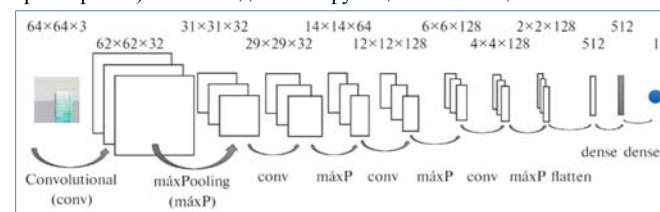


Рис. 3. Архитектура глубокой нейронной сети (ConvNet) с двоичным выходом
Fig 3. Architecture of deep neural network (ConvNet) with binary output

4. Эксперименты и результаты

В этом разделе представлена экспериментальная система с двумя моделями обнаружения. В первой модели используются изображения в пространственном домене, то есть исходные изображения без предварительной обработки. Вторая модель использует изображения в вейвлет-домене, поэтому перед входом в нейронную сеть изображения предварительно обрабатываются на трех уровнях декомпозиции. Эти два датасета ранее были получены на этапе распознавания и навигации в среде моделирования Gazebo. Кроме того, необходимо отметить, что в обеих моделях была использована архитектура ConvNet, показанная на рис. 3. Нейронная сеть обучалась с использованием среды машинного обучения Keras, а в качестве бэкенда использовалась библиотека TensorFlow.

Как правило, оценка обучающей способности модели и эффективности обнаружения производится сравнением показателей точности и потерь. Это две статистики обычно собираются на этапе обучения, валидации и тестирования. Наконец, наша модель оценивалась в аэронавигационном приложении с использованием ROS и среды моделирования Gazebo.

4.1 Датасеты

Две модели обнаружения имеют бинарный выход, так что задача обнаружения сосредоточена на двух классах: первый класс обнаруживает присутствие текстурированного объекта в плоскости изображения, а во второй класс попадают изображения, для которых объект

находится вне сцены. Для каждого класса датасет содержит 700 изображений для этапа обучения, 150 для валидации и 105 для тестирования (955 изображений на класс). Архитектура ConvNet для двух моделей обнаружения параметризована для обучения и прогнозирования только изображений с размером 64×64 пикселя и тремя каналами RGB (красный, зеленый, синий). Поэтому для первой модели обнаружения размер исходных изображений (640×380 пикселей) изменяются до 64×64 пикселей. Во второй модели используются вейвлет-изображения; поэтому, во-первых, исходные изображения (640×380 пикселей) приводятся к размеру 512×512 пикселей для кратномасштабного анализа. В этом процессе генерируются одно аппроксимационное и три детализированных (горизонтальное, диагональное и вертикальное) вспомогательных изображения. Как показано на рис. 4, вейвлет-процесс применяется к двум классифицированным датасетам (классам). В этом случае предлагается использовать только аппроксимационные подмножества (изображения размером 64×64 пикселей), поскольку именно в этом месте можно сэкономить больше всего энергии.

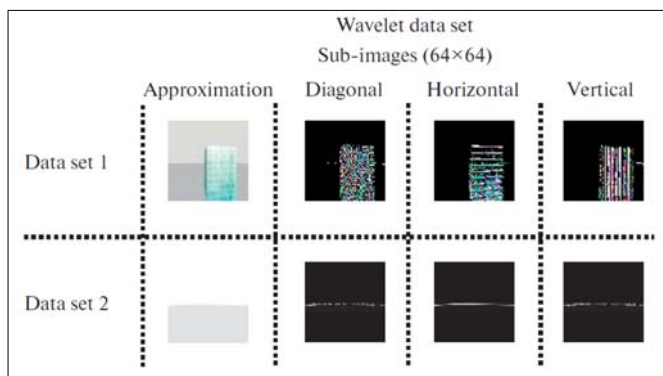


Рис. 4. Набор данных вейвлет-подизображения; в этом случае мы сосредотачиваемся только на аппроксимированных подизображениях

Fig 4. Wavelet sub-image dataset for the second detection model; in this case, we only focus on the approximation sub-images

4.2 Оценка модели

Эти две модели обнаружения используются для демонстрации вклада вейвлет-анализа в сочетании с архитектурой CNN. Кроме того, экспериментальная разработка позволяет наблюдать за поведением обеих моделей при обучении, поэтому показатели точности и потерь выбираются за десять эпох (количество итераций, в которых должно производиться обучение на основе датасета) на этапах обучения и валидации. Таким образом, обе модели обнаружения обучаются с использованием 504 001 параметра на компьютере с процессором Intel Core i5-2450M.

На рис. 5 приведены результаты, демонстрируемые во всех десяти эпохах первой моделью обучения, для которой характерно использование только исходных изображений и ConvNet для бинарной классификации [33]. Точность на этапе обучения (зеленая линия) начинается с 78%, затем достигает почти 100% во вторую эпоху; с этого момента поведение обучения является случайным между 98% и 100%. Что касается точности на этапе валидации (синяя линия), то обобщение обучения уменьшается при обучении на новых данных. Этот эффект вызван переобучением (после трех эпох); то есть сеть начинает обучаться паттернам, которые характерны для обучающих данных, но некорректны или неуместны по отношению к новым данным. Между тем, на рис. 6 показаны показатели потерь для оценки модели

обучения. Значение потерь, близкое к нулю, очень быстро достигается на этапе обучения (зеленая линия); в результате этого эффекта сеть становится более восприимчивой к переобучению. Таким образом, в случае новых данных потери на этапе валидации (синяя линия) уменьшаются и значительно увеличиваются (три и семь эпох), сохраняя эффект переобучения.

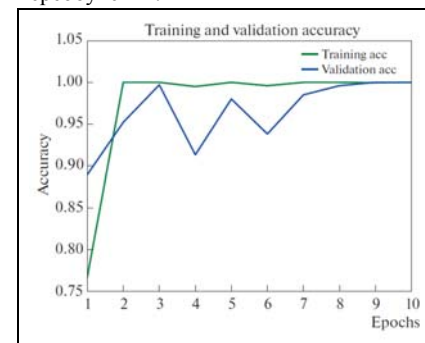


Рис. 5. Точность обучения и валидации модели без предварительной обработки входных изображений в Convnet
Fig 5. Capacity of the model in the accuracy of training and validation, without pre-processing of the input images to the ConvNet

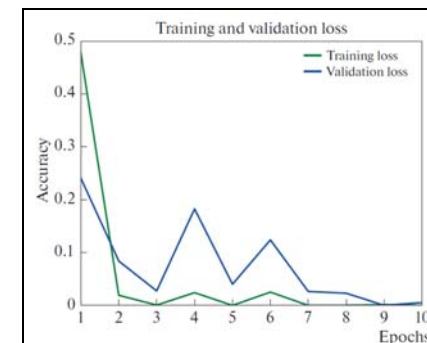


Рис. 6. Потери обучения и валидации модели без предварительной обработки входных изображений в Convnet
Fig 6. Capacity of the model in the loss of training and validation, without pre-processing of the input images to the ConvNet

Вторая модель обучения находится в тех же условиях архитектуры ConvNet, но с добавлением изображений в вейвлет-домене (только аппроксимационные подизображения). Ниже представлены результаты нашего подхода с использованием вейвлет-анализа за десять эпох обучения и валидации. Как показано на рис. 7, модель имеет лучшую производительность при обобщении обучения, а также позволяют избежать эффекта переобучения по отношению к новым изображениям датасета валидации. На этапе обучения сначала достигается 68% точности (зеленая линия), но на четвертой эпохе достигается почти 100%.

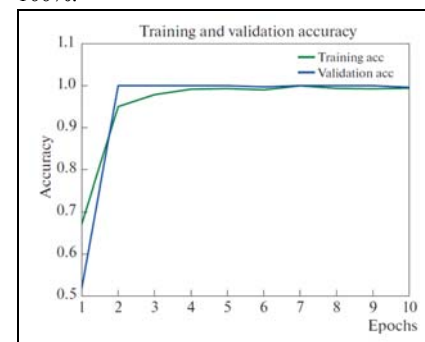


Рис. 7. Точность обучения и валидации модели при использовании вейвлет-датасета
Fig 7. Capacity of the model in the accuracy of training and validation, with the wavelet dataset

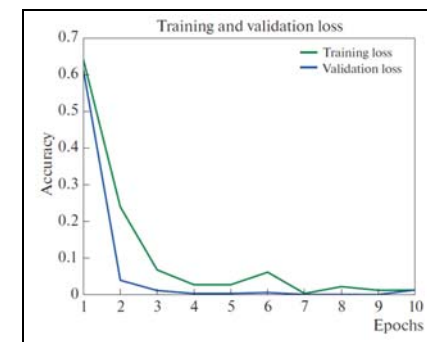


Рис. 8. Потери обучения и валидации модели при использовании вейвлет-датасета
Fig 8. Capacity of the model in the loss of training and validation, with the wavelet dataset

Показатель точности работает лучше на этапе валидации (синяя линия), поскольку обобщение обучения выше, чем обучающие данные, что позволяет избежать переобучения для новых данных и достичь почти 100% на второй эпохе. Более того, рис. 8 показывает, что модель достигает нулевых потерь на этапе обучения (зеленая линия), очень медленно приближаясь к нулю на седьмой эпохе; этот эффект позволяет нам иметь модель, которая не поддается переобучению для новых данных. На стадии валидации значение потерь, близкое к нулю, получается очень быстро (синяя линия); это связано с обучаемостью и качеством изображений в вейвлет-доме.

Таким образом, мы имеем модель с более высокой производительностью обнаружения и классификации по сравнению с первой моделью обучения, которая использует исходный набор данных. Обобщение знаний позволяет нам адекватно обучать новой информации; в результате мы получаем меньшую разницу между потерями на этапе обучения и на этапе валидации. Наша работа позволяет оптимизировать обучение для обнаружения объектов в аэронавигационных приложениях. Кроме того, некоторые преимущества преобразования данных в вейвлет-дом заключаются в возможности обучения физических характеристикам объектов, ориентированным на детали текстуры, и устранении переобучения при наличии небольшого обучающего набора данных. В табл. 1 приведены показатели результативности модели при использовании вейвлет-датасета для валидации. Показатели точности и потерь валидируют модель обучения, полученную слиянием методов CNN и вейвлет-анализа.

Табл. 1. Результаты валидации модели обучения, созданной с применением вейвлет-анализа и глубокого обучения.

Показатель	Значение [%]
Точность тестирования	100
Потери тестирования	0.61

4.3 Эксперименты с симуляцией

В нашей работе использовался симулятор Gazebo, который применяется для проектирования UAF типа Parrot AR.Drone 2.0 и разработки реалистичных 3D-сценариев для симуляции [34]. Симулятор, в частности, обеспечивает быстрое выполнение алгоритмов, предоставляет пользовательский интерфейс и гибко контролирует навигацию дрона. Кроме того, Gazebo обеспечивает такое же управление, как и в реальных мобильных аппаратах.

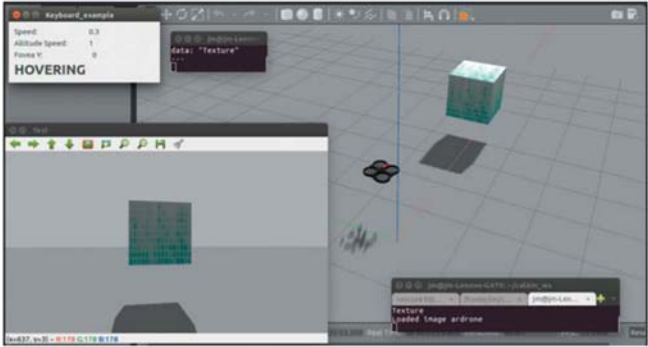


Рис. 9. Тест 1: Обнаружение объектов на сцене предложенным методом с применением подхода вейвлет-анализа и глубокого обучения
Fig 9. Test 1: Object detection in scene applying the proposed method with the approach of wavelet analysis and deep learning

При разработке приложений симулятор позволяет подключать операционную систему ROS с открытым исходным кодом, созданную под лицензией Berkeley Software Distribution (BSD).

Обеспечивается функционирование операционной системы в гетерогенном компьютерном кластере [35]. Узлы обмениваются сообщениями, что позволяет программировать их на любом языке, имеющем клиентские библиотеки для ROS (например, C, C++, Python, Java, Matlab) [36].

Что касается второй модели обнаружения, симулятор обеспечивает поддержку оценки и выполнения в реальном времени. Важно подчеркнуть, что изображения на этапе оценки – это изображения с камеры на борту AR.Drone (вид спереди), как если бы это была настоящая камера. Изображения в системе изначально имеют размер 640 × 380 пикселей, поэтому они преобразуются к размеру 512 × 512 пикселей. Потом эти изображения преобразуются в вейвлет-дом посредством кратномасштабного анализа на уровне масштабирования, равном трем. В результате мы имеем четыре подизображения с разрешением 64 × 64 пикселя (значение, допустимое для модели обнаружения). При исполнении в реальном времени мы показываем разные ракурсы с бортовой камеры. Показаны два класса предсказания модели обнаружения: первый класс, в котором объект с текстурой полностью появляется на сцене, как показано на рис. 9, и второй класс, в котором он не появляется в плоскости изображения, см. рис. 10.

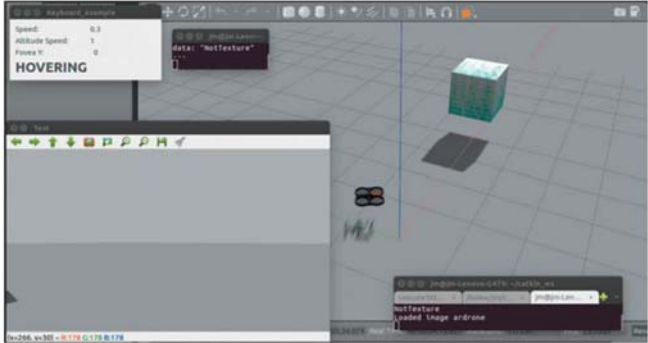


Рис. 10. Тест 2: Объект не обнаруживается на сцене, так как он находится вне поля зрения бортовой камеры дрона. Видео по работе доступно на <https://youtu.be/MOSrJyf14T8>
Fig 10. Test 2: Object is not detected on the scene, as it is out of view from the on-board camera of the drone. A video of this work for review purposes is available at <https://youtu.be/MOSrJyf14T8>

Средняя частота обнаружения достигает 98% в различных ракурсах (вариации масштаба сцены и освещения), как показано на рис. 11. Модель обнаружения продемонстрировала отличные значения времени прогнозирования. Это время значительно меньше того, через которое становятся видимыми исходные изображения, как показано в табл. 2. Время обнаружения зависит от сети, глубины, размера изображения, количества нейронов в последнем слое и возможностей используемой аппаратуры.

Табл. 2. Экспериментальные результаты выполнения нашего приложения в операционной системе ROS и симуляторе Gazebo

Время	Значение [с]
Функция предсказания	0.01608
Видимое исходное изображение	0.02587

Мы показали, что взаимодействие нейронной сети ConvNet с вейвлет-набором данных в роботизированных приложениях может дать многообещающие результаты на этапе обучения. Среда моделирования обеспечивает полный анализ предложенного метода, следовательно, он может быть адаптирован к реальным условиям для классификации и

обнаружения объектов с текстурами. Однако следует заметить, что при этом целесообразно использовать компьютер с высокой вычислительной мощностью, поскольку для моделирования требуется много ресурсов.

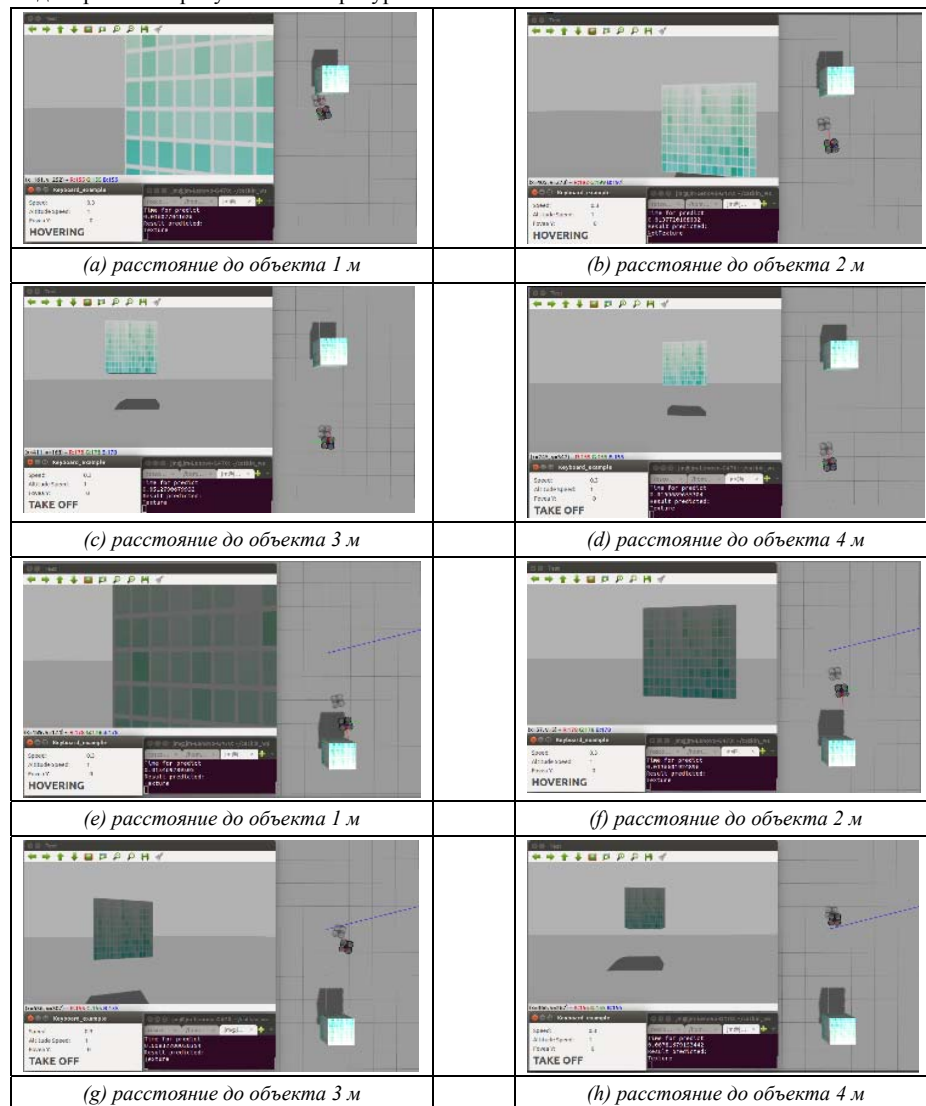


Рис. 11. Результаты обнаружения на четырех различных расстояниях от цели до камеры дрона и при разном освещении; (a)-(d) основаны на снимках в одном ракурсе, а (e)-(h) – на снимках в противоположном ракурсе (изображение цели содержит тень)

Fig 11. Detection results at four different distances from the target to the drone camera, and different scales and illumination (a)-(d) are results based on a perspective in the environment, while (e)-(f) are results based on an opposite perspective in the environment (the image of the target contains shadow)

5. Заключение

Мы представили новый метод обнаружения объектов с текстурами. Наше предложение основано на использовании методов вейвлет-анализа и глубоких нейронных сетей, которые были адаптированы к области аэронавигации. Для этого мы кадр за кадром классифицировали изображения, полученные с бортовой камеры дрона. В сочетании с системой ROS и моделью обучения полученная информация преобразуется в вейвлет-домен, в котором это трехуровневое преобразование проявляет важные характеристики аппроксимации и детализации. В получаемом аппроксимационном датасете сохраняется большинство свойств исходного изображения. Результаты оценки показывают, что наличие изображений в вейвлет-доме и с более низким разрешением значительно повышает эффективность обнаружения по сравнению с использованием изображений в пространственном домене в качестве входных данных архитектуры ConvNet. Кроме того, вейвлет-датасет был полезен для смягчения эффекта переобучения при обобщении обучения на этапе обучения.

В заключение отметим, что приложение показывает высокую точность в случае, когда в результате прогноза на сцене обнаруживается объект с текстурой (Texture) или не объект с текстурой (NotTexture). Существенно и то, что система предсказывает почти вдвое быстрее, чем передаются кадры камеры. Мы считаем, что полученные результаты перспективны, поэтому было бы целесообразно экспериментировать с новыми семействами вейвлетов для расширения возможностей обнаружения объекта, улучшения характеристики текстур, особенно использования нашего подхода в более сложной автономной навигационной системе.

Список литературы / References

- [1]. L.O. Rojas-Pérez. Autonomous Navigation System for Micro Aerial Vehicles. Dissertation, Instituto Tecnológico Superior de Atlixco, Puebla, México, 2018.
- [2]. J. Martínez-Carranza, L. Valentin, F. Márquez-Aquino et al. Obstacle Detection during Autonomous Flight of Drones Using Monocular SLAM. *Research in Computing Science*, vol. 114, 2016, pp. 111-124.
- [3]. Н.С. Васильева. Методы поиска изображений по содержанию. Программирование, том 35, no. 3, 2009 г., стр. 51-80 / N.S. Vassilieva. Content-based image retrieval methods. *Programming and Computer Software*, vol. 35, no. 3, 2009, pp. 158-180.
- [4]. А.Д. Жданов, Д.Д. Жданов, Н.Н. Богданов и др. Проблемы дискомфорта зрительного восприятия в системах виртуальной и смешанной реальности. Программирование, том 45, no. 4, 2019 г., стр. 9-18 / A.D. Zhdanov, D.D. Zhdanov, N.N. Bogdanov et al. Discomfort of Visual Perception in Virtual and Mixed Reality Systems. *Programming and Computer Software*, vol. 45, no. 4, 2019, pp. 147-155.
- [5]. В.В. Санжаров, В.А. Фролов. Уровень детализации для предрасчитанных процедурных текстур. Программирование, том 45, no. 4, 2019 г., стр. 54-63 / V.V. Sanzharov, V.A. Frolov. Level of Detail for Precomputed Procedural Textures. *Programming and Computer Software*, vol. 45, no. 4, 2019, pp. 187-195.
- [6]. C. Vargas-Olmos. Procesamiento de imágenes con métodos de ondelete. Dissertation, Facultad de Ciencias, UASLP, San Luis Potosí, México, 2010 (in Spanish).
- [7]. J. D. Cárdenas-Amaya. Extracción y análisis de características con la Transformada Wavelet para el reconocimiento de imágenes. Dissertation, Instituto de Investigación en Comunicación Óptica, UASLP, San Luis Potosí, México, 2018 (in Spanish).
- [8]. Y. Bazi and F. Melgani, Convolutional SVM Networks for Object Detection in UAV Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, 2018, pp. 3107-3118.
- [9]. Y. Pi, N.D. Nath, A.H. Behzadan, Convolutional neural networks for object detection in aerial imagery for disaster response and recovery. *Advanced Engineering Informatics*, vol. 43, 2020, Article 101009.
- [10]. S. Dionisio-Ortega, L.O. Rojas-Perez, J. Martinez-Carranza, I. Cruz-Vega. A deep learning approach towards autonomous flight in forest environments. In *Proc. of the International Conference on Electronics, Communications and Computers (CONIELECOMP)*, 2018, pp. 139-144.

- [11]. A. Krizhevsky, I. Sutskever, and G.E. Hinton. Imagenet classification with deep convolutional neural networks. In Proc. of the 26th Annual Conference on Neural Information Processing Systems (NIPS), 2012, pp. 1097-1105.
- [12]. J. Redmon, S. Divvala, R. Girshick, A. Farhadi. You Only Look Once: Unified, Real-Time Object Detection. In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.
- [13]. S. Jung, S. Hwang, H. Shin, D. H. Shim. Perception, Guidance, and Navigation for Indoor Autonomous Drone Racing Using Deep Learning. IEEE Robotics and Automation Letters, vol.3, issue 3, 2018, pp. 2539-2544.
- [14]. T. Williams, R. Li. An Ensemble of Convolutional Neural Networks Using Wavelets for Image Classification. Journal of Software Engineering and Applications, vol. 11, no 2, 2018, pp. 69-88.
- [15]. T. Williams, R. Li. Wavelet Pooling for Convolutional Neural Networks. International Conference on Learning Representations, 2018, 12 p.
- [16]. J. Piao, Y. Chen, H. Shin. A New Deep Learning Based Multi-Spectral Image Fusion Method. entropy, vol. 21, issue 6, 2019, 16 p.
- [17]. D. N. De Silva, S. Fernando, I.T.S. Piyatilake, A.V.S. Karunaratne. Wavelet based edge feature enhancement for convolutional neural networks. In Proc. of the Eleventh International Conference on Machine Vision (ICMV 2018), 2018.
- [18]. S. Fujieda, K. Takayama, and T. Hachisuka. Wavelet convolutional neural networks for texture classification. arXiv preprint arXiv:1707.07394, 2017.
- [19]. C.S. Burrus, R.A. Goniath and H. Guo. Introduction to Wavelets and Wavelet Transforms: A Primer. Pearson, 1998, 288 p.
- [20]. S. Mallat. A Wavelet Tour of Signal Processing: The Sparse Way. 3rd ed. Academic Press, 2009, 832 p.
- [21]. G. Strang and T. Nguyen. Wavelets and Filter Banks. 2nd revised ed. Wellesley-Cambridge Press, 1996 500 p.
- [22]. C. Vargas-Olmos. Procesamiento de señales y solución de problemas con la transformada wavelet. Ph.D. dissertation. Instituto de Investigación en Comunicación Óptica, UASLP, San Luis Potosi, México, 2017 (in Spanish).
- [23]. J.S. Walker. A Primer on Wavelets and Their Scientific Applications. 2nd ed. Chapman & Hall/CRC, 2008, 320 p.
- [24]. I. Goodfellow, Y. Bengio, and A. Courville. Deep Learning. MIT Press, 2016, 800 p.
- [25]. И.В. Степанян. Методология и инструментальные средства проектирования бинарных нейронных сетей. Программирование, том 46, no. 1, 2020 г., стр. 54-62 / I.V. Stepanyan, Methodology and Tools for Designing Binary Neural Networks. Programming and Computer Software, vol. 46, no. 1, 2020, pp. 49–56.
- [26]. Ю.Г. Сметанин, Л.Е. Карпов, Ю.Л. Карпов. Адаптация общих концепций тестирования программного обеспечения к нейронным сетям. Программирование, том 44, no. 5, 2020 г., стр. 43-56 / Y.L. Karpov, L.E. Karpov, and Y.G. Smetanin. Adaptation of General Concepts of Software Testing to Neural Networks. Programming and Computer Software, vol. 44, no. 5, 2018, pp. 324–334.
- [27]. J. Torres. Convolutional Neural Networks for Beginners. Practical Guide with Python and Keras, 2018. URL: <https://towardsdatascience.com/convolutional-neural-networks-for-beginners-practical-guide-with-python-and-keras-dc688ea90dca>, accessed 27 April 2019.
- [28]. K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
- [29]. A. Yayık, K. Yakup, and G. Altan. Deep Learning with ConvNET Predicts Imagery Tasks Through EEG. arXiv preprint arXiv:1907.05674, 2019.
- [30]. W. Liu, D. Anguelov, D. Erhan et al. SSD: Single shot multibox detector. Lecture Notes in Computer Science, vol. 9905, 2016, pp. 21-37.
- [31]. H. Kaiming, G. Gkioxari, P. Dollár, and R. Girshick. Mask R-CNN. In Proc. of the IEEE international conference on computer vision, 2017, pp. 2961-2969.
- [32]. R. Shaoqing, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In Proc. of the 28th Annual Conference on Neural Information Processing Systems (NIPS), 2015, pp. 91-99.
- [33]. F. Chollet. Deep Learning with Python. Manning Publications, 2018, 384 p.

- [34]. N. Koenig, A. Howard. Design and use paradigms for Gazebo, an open-source multi-robot simulator, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), vol. 3, 2004, pp. 2149-2154.
- [35]. E. Rodriguez-Martin. Sistema de posicionamiento para un drone. Dissertation, Universidad de La Laguna, España, 2015 (in Spanish).
- [36]. L. Joseph. Robot Operating System for Absolute Beginners: Robotics Programming Made Easy. Apress, 2018, 295 p

Информация об авторах / Information about authors

Хуан Мануэль ФОРТУНА-СЕРВАНТЕС – аспирант. Область научных интересов: компьютерное зрение, цифровая обработка изображений, глубокое обучение, беспилотные летательные аппараты, робототехника.

Juan Manuel FORTUNA-CERVANTES, PhD Student. Research interests include Computer Vision, Digital Image Processing, Deep Learnig, Unmanned Aerial Vehicles, Robotics.

Марко Тулио РАМИРЕС-ТОРРЕС – кандидат прикладных наук, профессор. Область научных интересов: криптография, вейвлет-анализ, виртуальные приборы.

Marco Tulio RAMÍREZ-TORRES, Ph.D. in Applied Sciences, Full time Professor. Research interests include Cryptography, Wavelet Analysis, Virtual Instrumentation.

Хуан МАРТИНЕС-КАРРАНСА – кандидат прикладных наук, профессор. Область научных интересов: обработка сигналов, вейвлет-анализ, системы шифрования и прикладная математика.

Juan MARTINEZ-CARRANZA, Ph.D. in Applied Sciences, Full Professor. Research interests include signal processing, wavelet and scaling analysis, encryption systems and applied mathematics.

Хосе Саломе МУРГУИЯ-ИБАРРА, кандидат наук, профессор-исследователь кафедры электроники факультета естественных наук. Область научных интересов: обработка сигналов, нелинейные динамические системы.

José Salomé MURGUÍA-IBARRA, Ph.D., Professor Researcher of the Electronic Department of the Sciences Faculty. Research interests include Signal Processing, Nonlinear Dynamical Systems.

Марсела МЕХИЯ-КАРЛОС, кандидат наук, профессор-исследователь. Область научных интересов: коммуникации и обработка сигналов.

Marcela MEJÍA-CARLOS, Ph.D., Professor Researcher. Research interests: Communication and Signal Processing.