

# УЛУЧШЕНИЕ ГЕНЕРАЦИИ ПОЛОЖЕНИЯ ЗАХВАТА С ПОМОЩЬЮ GGCNN И МЕТОДА ГЛАВНЫХ КОМПОНЕНТ

## IMPROVING GRASP POSE GENERATION USING GGCNN AND PRINCIPAL COMPONENT ANALYSIS

**A. Andreev**

*Summary.* The article is dedicated to the development of a modified grasp pose generation algorithm based on GGCNN with the integration of principal component analysis (PCA) to improve the accuracy of robotic object grasping without retraining the neural network.

This problem is highly relevant given the evolving methods of machine vision and neural network approaches in robotics. Working with three-dimensional images is increasingly popular as a replacement for classical machine vision techniques and two-dimensional image processing. As robotic systems and mobile robots advance, a challenge arises in grasping objects that are often randomly located within the robot's operational workspace. This necessitates the use of advanced methods for object detection and subsequent grasping. Equally important is the application of specific algorithms for seemingly simple tasks such as picking up, transferring, and sorting objects in industrial production.

This work proposes an enhanced approach that combines the generative grasping convolutional neural network (GGCNN) with PCA analysis. The methodology includes preprocessing of input data—namely the depth map from an Intel RealSense camera—analyzing the object grasp probability map using PCA, and transforming the data into a format suitable for grasp execution. Experimental data were collected using a pretrained GGCNN on the Cornell dataset to obtain metrics for grasp success.

As a result of this work, an acceptable grasp success probability of approximately 95.6 % was achieved for the test object by applying PCA to the parametric grasp probability map. The study also evaluated grasp probabilities under various point-selection parameters for PCA when dealing with a symmetrical object.

The proposed method increases the object grasp success probability without retraining the neural network, thereby reducing the labor required to create and label training datasets and retrain the network for new objects. It also reduces the number of output parametric maps and, consequently, decreases the dimensionality of the network's output.

*Keywords:* principal component analysis, PCA, robotic grasping, GGCNN, convolutional neural networks, CNN, manipulation tasks, computer vision, robotics.

**Андреев Андрей Валерьевич**

Национальный исследовательский университет ИТМО  
ext13033@gmail.com

*Аннотация.* Статья посвящена разработке модифицированного алгоритма генерации положений захвата на основе GGCNN с интеграцией метода главных компонент (PCA) для повышения точности роботизированного захвата объектов без переобучения нейросети.

Данная проблема актуальна и необходима к рассмотрению в силу развивающихся методов технического зрения и нейросетевых подходов в сфере роботизации. Работа с трёхмерными изображениями всё больше и больше набирает популярность в замену классических подходов машинного зрения и работы с двухмерными изображениями. С развитием робототехнических средств и мобильных роботов возникает проблема, которая заключается в захвате объектов, которые зачастую находятся в операционном пространстве роботизированного захвата хаотично и необходимо использовать продвинутые методы для обнаружения объектов и последующего захвата и что не менее важно, необходимо применять некоторые алгоритмы для таких простых задач, как захват и перенос объектов, а также сортировка объектов, в случае промышленного производства

В работе предложен усовершенствованный подход, сочетающий генеративную свёрточную нейронную сеть GGCNN с PCA-анализом. Методология включает: предварительную обработку входных данных, а именно глубинной карты, которая поступает с камеры RealSense, анализ карты вероятности захвата объекта с помощью метода главных компонент и получение, а также преобразование данных в формат для захвата объекта. Сбор экспериментальных данных проводился на основе предварительно обученной нейросети GGCNN на датасете Корнелла (Cornell dataset) с целью получения метрики успешности захвата.

В результате данной работы удалось получить приемлемый результат вероятности захвата объекта, порядка 95,6 %, для исследуемого объекта с применением метода главных компонент на параметрическую карту вероятности захвата объекта, а также оценить вероятность захвата при различных параметрах выбора точек для работы с методом главных компонент для симметричного объекта.

Применение предложенного метода привело к повышению вероятности захвата объекта без переобучения нейросети, а значит и уменьшению затрат на создание наборов данных для обучения и переобучение сети для необходимых объектов, а также уменьшению выходного числа параметрических карт, а в следствии и уменьшение размерности выхода нейросети.

*Ключевые слова:* метод главных компонент, PCA, роботизированный захват, GGCNN, свёрточные нейронные сети, CNN, манипуляционные задачи, компьютерное зрение, робототехника.

## Введение

Основная задача работы сводится к тому, чтобы улучшить вероятность захвата объекта с применением свёрточной нейросети для генерации положения захвата GGCNN без переобучения сети для нового объекта, который не содержится внутри обучающего датасета, а также уменьшить выходную размерность нейросети с целью уменьшения загрузки системы без значимой потери в точности, а также снизить трудозатраты на создание нового размеченного датасета для переобучения нейросети, что является достаточно трудоёмкой задачей.

В робототехнике данная проблематика имеет значение, поскольку нейросетевые алгоритмы внедряются повсеместно, начиная с производственных линий, заканчивая мобильными роботами различного назначения.

Как справедливо отмечается в работе [1], роботизированный захват остаётся ключевым навыком для выполнения сложных операций, при этом точных захвата произвольных объектов в неструктурированных средах продолжает оставаться сложной для решения исследовательской проблемой. Как приводится в этой работе, современные методы захвата включают в себя три ключевых аспекта: детекция захвата, планирование пути и управляющая подсистема. В нашей работе исследование сфокусировано на работе над первым этапом через интеграцию GGCNN и PCA.

## Материалы и методы

В данной статье в качестве объекта изучения использовалась свёрточная нейронная сеть для генерации положения захвата объекта GGCNN (Grasp Generative CNN). Данная сеть используется для получения положения роботизированного двухпальцевого схвата. В текущем случае будет использоваться только одна карта параметров — карта вероятностей захвата.

Метод главных компонент позволяет получить угол захвата объекта, не прибегая к использованию дополнительных карт параметров с выхода GGCNN. В отличие от архитектуры MultiGrasp [2], где используется локально-ограниченный механизм предсказаний (сетка 7×7), наш метод применяет PCA к карте вероятностей GGCNN. Это позволяет: избежать жесткой привязки к сетке, автоматически определять главные оси захвата, однако позволяет работать только с объектами симметричной формы без модификации архитектуры.

## Литературный обзор

В одной из статей приводится утверждение «Большинство методов обеспечивают высокий процент успе-

ха за счет обучения сети на наборах данных с большим количеством аннотаций захватов, что требует значительных трудозатрат и ресурсов. Поэтому достижение высокой точности захвата в стопке объектов без аннотаций остается сложной задачей.» [3, с. 1477–1490]. Данная статья предлагает решение в области робототехники в задаче захвата объекта без обучения нейросети на большом наборе аннотационных или обучающих данных. Проблематика данной статьи поднимается таким образом, что обучение нейросетей с использованием новых датасетов является трудоёмкой и сложной задачей. Данная проблема также имеет решение в текущей статье.

«Использование машинного обучения в роботизированном захвате значительно повышает гибкость и адаптивность систем, открывая новые возможности для автоматизации в динамичных отраслях» [4]. Данная цитата лишний раз подтверждает, что машинное обучение и методы, базированные на искусственном интеллекте, являются достаточно гибкими и добавляют адаптивности системам, что подтверждается в проделанной работе.

«Объединяя перцепционные возможности глубокого обучения с геометрической верификацией, данная работа продвигает развитие автономных роботизированных систем, предлагая масштабируемое решение для промышленной автоматизации, ориентированное на точность и адаптивность» [5]. В текущей работе также используются гибридный метод, обеспечивающий достаточно высокую скорость для систем реального времени с высокой частотой кадров, порядка 20 и более кадров в секунду, однако используется метод из области статистики, а именно метод главных компонент, который позволяет снизить размерность данных системы. Так, например в нашей работе производится переход от двухмерной карты вероятности захвата к одномерной главной компоненте, в следствии чего мы можем получить ориентацию двумерного облака точек в пространстве, а значит и ориентацию объекта для успешного захвата.

Также, в другой научной публикации приводится заключение о том, что раньше роботизированные руки на производствах требовали специального программирования для каждой конкретной задачи, например, захват объекта из фиксированного местоположения с известной заданной ориентацией, однако на текущий момент «современные нейронные сети обладают автономными возможностями, которые активно используются для совершенствования робототехники и повышения точности позиционирования» [6, с. 497–519].

Также в одной из работ приводится реализация системы подобной на реализованную в данной работе, за исключением, что используется связка YOLOv4+GGCNN для захвата объекта [7]. В данной работе показатель обработки одного кадра составляет порядка 0,11 секунды, а ве-

роятность захвата объектов составляет порядка 86,0 %. Данная работа полагается на работу с большим числом различных сред, что в целом обеспечивает достаточно сложные для испытания обстоятельства, однако стоит обратить внимание на параметр скорости обработки кадра. Значение в 0,11 секунды на обработку одного кадра является достаточным значением для систем реального времени, однако в нашем случае скорость обработки одного кадра имеет скорость несколько выше в силу более простой системы обработки изображения.

### Результаты

Система представляет из себя достаточно сложную и многоуровневую систему, представленную на структурной схеме на Рис. 1.

Данная система реализована в имитационной среде с целью обеспечить автономность выполнения экспери-

мента. Алгоритм выполнения эксперимента представлен на алгоритмической блок-схеме на Рис. 2.

Приведём краткое описание данной блок-схемы.

1 — начало выполнения алгоритма.

2 — запуск рабочей среды и инициализация необходимых программных классов, таких как: CSVLogger, RobotControl, ImageProcess и GazeboObjectManager. Все объекты необходимы для решения соответствующих их названию задач.

3 — отведение робота в домашнюю позицию выполняется с целью освобождения рабочей области для последующего создания на ней объектов.

4, 12 — цикл эксперимента, который выполняется необходимое число раз. С целью обеспечения достоверно-

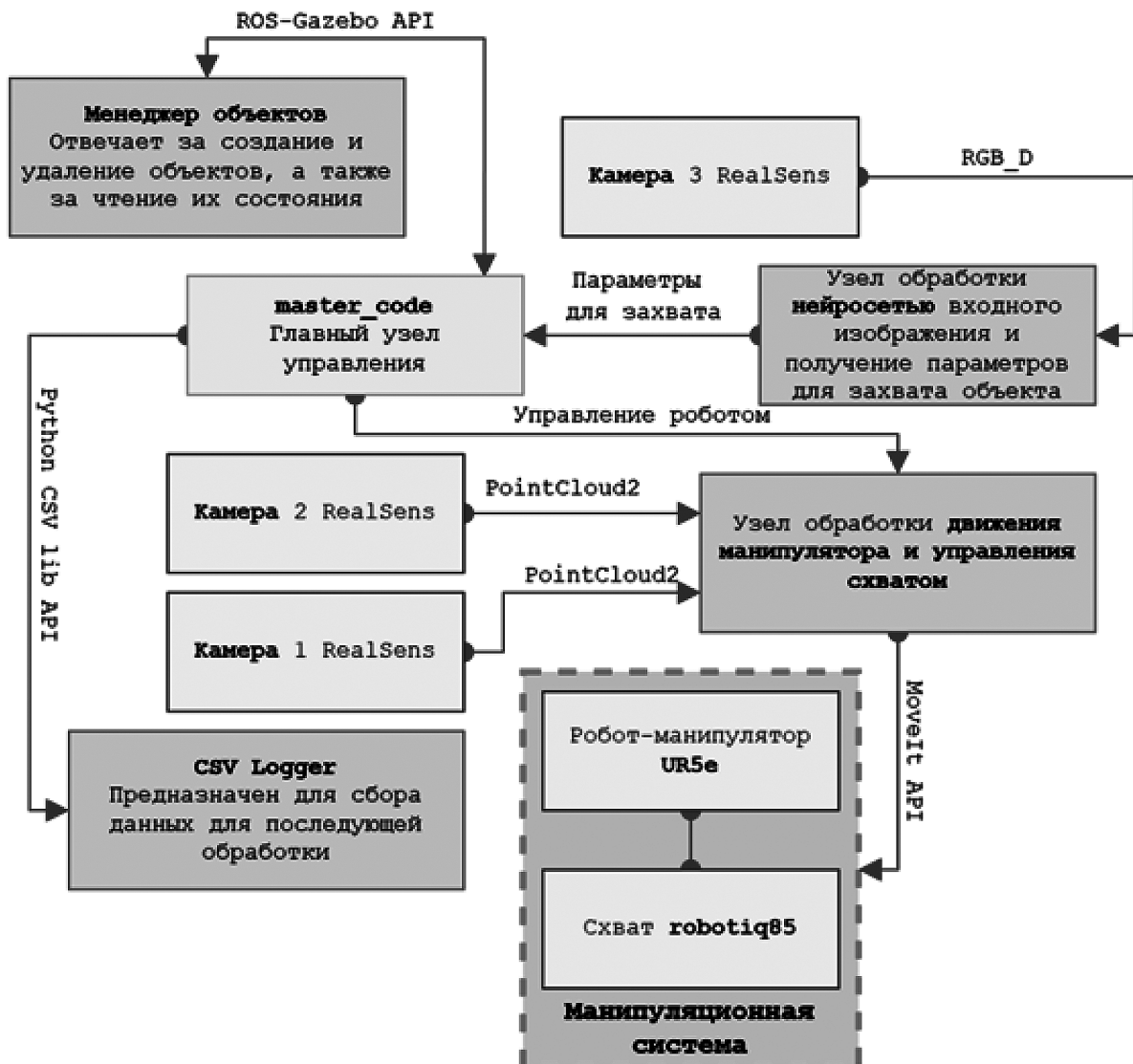


Рис. 1. Структурная схема системы для автоматизированной оценки системы алгоритма захвата объекта

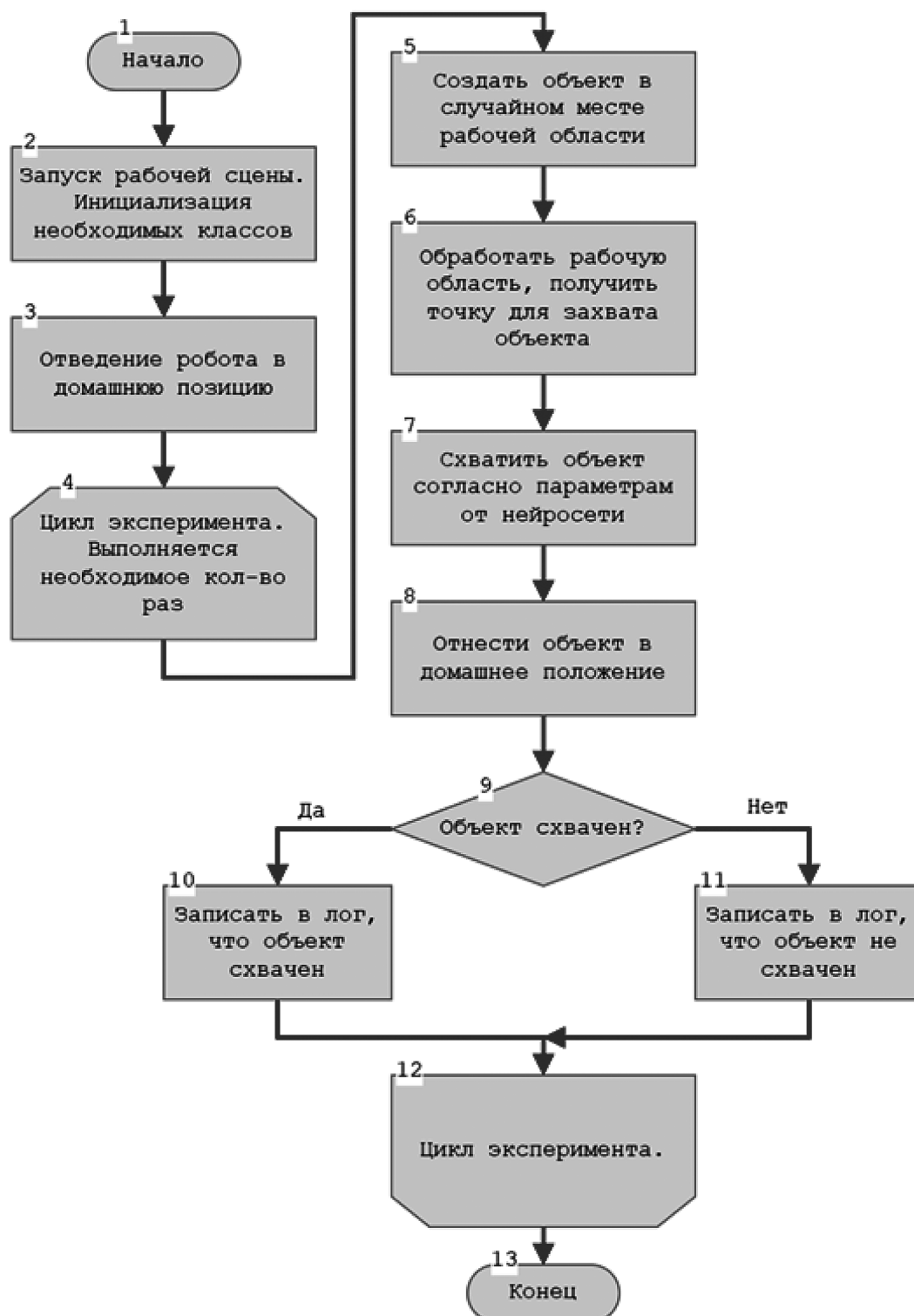


Рис. 2. Алгоритмическая блок-схема проведения эксперимента

сти результатов было выбрано число испытаний порядка сотни циклов эксперимента.

5 — создание объекта в случайном месте рабочей области. Рабочая область представляет из себя область видимости камеры направленной перпендикулярно вниз на платформу. Объект создаётся в случайном месте

таким образом, чтобы оставаться в области видимости камеры полностью. В случае, если объект ранее был создан, то предыдущий объект предварительно будет удалён.

6 — обработка рабочей области происходит в несколько этапов, которые включают нормализацию глу-



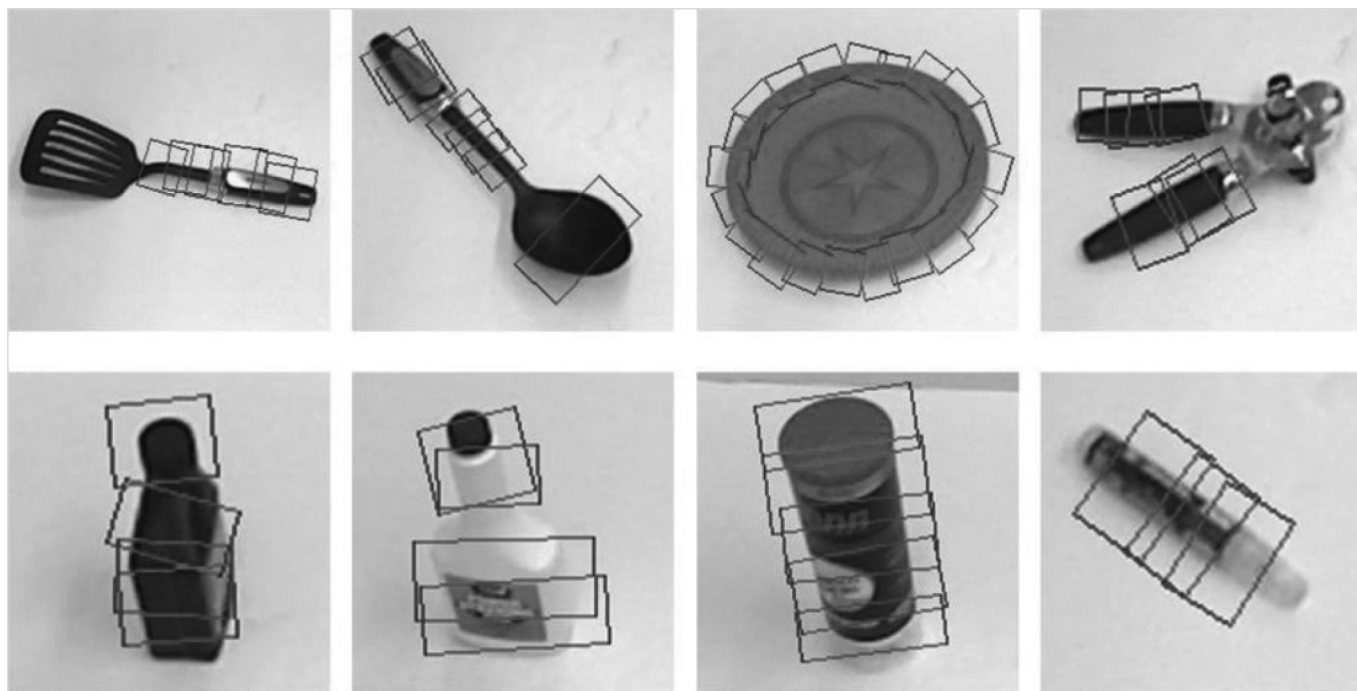


Рис. 3. Пример объектов набора данных Корнелла для обучения нейросети

бинного изображения, фильтрация артефактов и обработка карты через нейросеть, применение алгоритма для получения параметров захвата.

7 — захват объекта происходит также в несколько этапов. Сначала манипулятор разворачивает фланец согласно одному из параметров захвата объекта — углу поворота. Затем происходит позиционирование захвата над объектом в положение захвата и последующее перпендикулярное опускание манипулятора к объекту и последующих захват.

8 — относим объект в домашнее положение.

9 — проверка захвата объекта производится с помощью машины состояний, которая позволяет получить положение и ориентацию объекта. В случае, если объект был захвачен, выполняется переход к блоку 10 и в csv файл выполняется запись о том, что объект был захвачен. В случае, если объект захвачен не был, выполняется переход к блоку 11 и соответствующая запись в csv файл.

13 — конец выполнения алгоритма.

Изначально в качестве параметров для захвата использовались параметры сформированные исключительно нейросетью. Таким образом вероятность захвата составляла крайне низкое значение для практического применения, порядка 45 %. Такой низкий результат вероятности захвата обусловлен тем, что в исходном наборе данных для обучения камера располагалась к объектам под некоторым углом относительно нормали рабочей области. Пример датасета приведён на Рис. 3. В нашей

же системе используется объект, который не содержится в обучающем датасете. Задача состояла в том, чтобы без переобучения свёрточной нейросети получить приемлемую вероятность захвата изучаемого объекта.

Наибольший интерес для исследования представляет карта вероятности захвата объекта, которая генерируется с помощью свёрточной нейросети. Данная карта вместе с глубинной картой рабочей области представлена на Рис. 4. Стоит отметить, что мы проводили изучение объекта с размерами 50x50x150 мм.

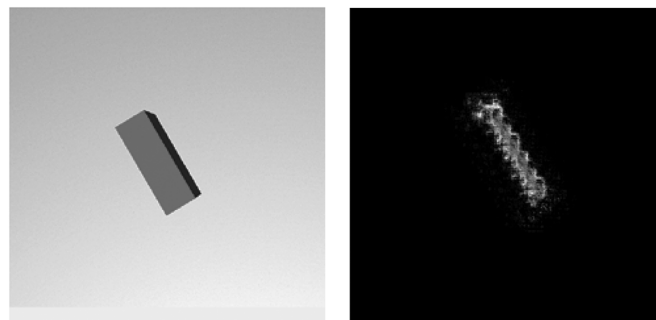


Рис. 4. Глубинная карта рабочей области (слева) и карта вероятности захвата от нейросети (справа)

Карта вероятности захвата представляет собой некоторое облако точек, которое отражает вероятность успешного захвата в каждой точке изображения. Значения от 0 до 1, где 1 — наилучшее место для захвата. Проводя аналогию с Рис. 4, более светлые пиксели соответствуют наибольшей вероятности успешного захвата объекта, а более тёмные соответствуют обратному. Как это должно работать на практике: на вход GGCNN по-

даётся RGB-D карта рабочей области, нейросеть предсказывает три карты параметров (Quality/Q-map, Angle, Width), алгоритм выбирает точку с наибольшим значением Q-map и затем смотрит соответствующие значения Angle и Width на соответствующих картах для точки с наибольшей вероятностью захвата [5]. Однако такой подход показывает не самые лучшие результаты, поскольку камера в нашем случае имеет нормальную рабочей области ориентацию, а угол и ширина захвата зачастую определяются некорректно. В следствии чего было принято решение найти альтернативный способ увеличить вероятность захвата объекта без переобучения нейросети.

В качестве решения было принято использовать метод главных компонент PCA, как описано в одной из работ «Чтобы справиться с высокой размерностью пространства конфигурации рук, авторы предлагают отображать траектории рук в более низкое скрытое пространство с использованием анализа основных компонентов (PCA)» [8]. Было принято решение перейти к анализу карты вероятности захвата Q-map с помощью метода PCA для получения двух важных параметров для захвата объекта: ориентация симметричного объекта и центра массы объекта.

Таким образом, новый алгоритм определения параметров захвата объекта включил в себя дополнительно обработку карты вероятности захвата с помощью метода главных компонент, однако было дополнительно принято решение отфильтровать точки с низкими вероятностями захвата и оставить исключительно точки в высокими вероятностями захвата. Так на рисунке 5 приведена сравнительная характеристика для выявления вероятности захвата объекта.



Рис. 5. Графическая интерпретация отсеечения точек с вероятностью захвата выше 0,5 слева и 0,9 справа (серые точки), центром массы объекта (светло-серая точка) и главной компонентой (темно-серая линия)

Также были собраны данные о том, насколько реальный геометрический центр отличается от центра массы, рассчитанного по средним координатам отфильтрованных точек, объекта и абсолютное среднее арифметическое отклонение главной компоненты от истинной ориентации объекта. Данные представлены в таблице 1.

Для объективной оценки нашего подхода сравним его с современными аналогами. Система Gu et al. [9], те-

Таблица 1.  
Отклонение полученных алгоритмом параметров объекта от реальных параметров объекта

Отсечение Q	$Q \geq 0,5$	$Q \geq 0,75$	$Q \geq 0,9$
Среднее арифметическое отклонение рассчитанного центра массы от истинного, мм	2,82	2,92	2,63
Максимальное отклонение рассчитанного центра массы объекта от истинного, мм	5,02	5,44	5,37
Среднее арифметическое отклонение главной компоненты от реальной ориентации объекта, °	2,28	1,89	1,92
Максимальное отклонение главной компоненты от реальной ориентации объекта, °	6,09	6,08	5,64

стировавшаяся на тех же датасетах, показывает точность 97.8 % (Cornell, image-wise), однако требует предварительного обучения и человеческого участия. Наш метод, хотя и демонстрирует несколько меньшую точность (95.6 %), работает полностью автономно и не нуждается в переобучении для новых объектов.

Обсуждение

Как можно заметить из таблицы 2 с отсечением  $Q \geq 0,9$ , которая показывает наибольшую вероятность захвата объекта среднее отклонение имеет значение между отсеечениями  $Q \geq 0,5$  и  $Q \geq 0,75$ . Это обосновано тем, что при отсечении  $Q \geq 0,5$  большое число точек имеет достаточно низкую вероятность захвата объекта и находится на краю объекта, что и вызывает такое большее отклонение рассчитанного центра массы объекта от истинного центра массы.

В случае с отсечением  $Q \geq 0,75$  получаем лучшие результаты с точки зрения отклонения рассчитанного геометрического центра от реального. Это связано с тем, что в данном случае отсекаются точки с достаточно малыми вероятностями захвата, которые расположены около краёв объекта, но при этом сохраняется достаточное число точек, которые при усреднении дают координату самую близкую к реальному геометрическому центру объекта.

В случае с отсечением  $Q \geq 0,9$  начинает немного увеличиваться отклонение рассчитанного центра массы объекта от реального центра массы, при этом значение не превышает значение в случае с отсечением  $Q \geq 0,5$ . Стоит отметить, что при этом отсечение  $Q \geq 0,9$  показывает наименьшее отклонение главной компоненты объекта, которая отражает ориентацию объекта в пространстве, от реальной ориентации объекта по оси нор-

мальной к рабочей области. Следовательно, можно сделать вывод о том, что отклонение рассчитанного центра массы объекта имеет меньшее влияние, чем отклонение рассчитанной ориентации объекта от реальной, однако, важно, что оба фактора оказывают влияние на вероятность захвата объекта.

Хотя метод [10, с. 1149–1161] демонстрирует более высокую точность (98.9 % против 95.6 % в нашем случае), ключевое преимущество нашего подхода — в его универсальности: не требует переобучения для новых объектов; позволяет снизить размерность данных (переход от Q-тар к главным компонентам), что снижает вычислительную нагрузку ещё и за счёт снижения числа выходов нейросети.

Один из недостатков нашего метода в том, что на текущий момент данный подход применим к объектам симметричной, вытянутой вдоль одного из измерений, формой, однако в случае производственных задач зачастую встречаются симметричные вытянутые объекты: небольшие коробки, различные скобяные изделия (болты, винты) и подобные объекты.

### Заключение

В данной работе предложен модифицированный алгоритм генерации положений захвата на основе GGCNN с интеграцией метода главных компонент (PCA), что позволило повысить точность роботизированного захвата объектов без необходимости переобучения нейронной сети. Основные результаты исследования заключаются в следующем:

1. Повышение вероятности успешного захвата: Применение PCA к карте вероятности захвата (Q-тар) позволило увеличить вероятность успешного захвата объекта до 95,6 % при отсечении точек с вероятностью захвата  $Q \geq 0,9$ . Это существенно выше исходного показателя в 45 %, полученного при использовании только GGCNN.
2. Снижение вычислительной нагрузки: Использование PCA позволило уменьшить размерность

выходных данных нейросети, перейдя от двумерной карты вероятностей к одномерной главной компоненте, что снизило нагрузку на систему без значительной потери точности.

3. Уменьшение трудозатрат: Предложенный метод исключает необходимость переобучения нейросети для новых объектов, что значительно сокращает затраты на создание и разметку дополнительных датасетов.
4. Точность определения параметров захвата: Эксперименты показали, что отклонение рассчитанного центра массы объекта от реального составило в среднем 2,63 мм, а отклонение главной компоненты от истинной ориентации объекта —  $1,92^\circ$ , что является приемлемым для практического применения.

Таким образом, предложенный гибридный подход, сочетающий генеративную сверточную нейронную сеть GGCNN и метод главных компонент, демонстрирует высокую эффективность в задачах роботизированного захвата. Он обеспечивает адаптивность системы к новым объектам без дополнительного обучения, снижает вычислительные затраты и увеличивает точность позиционирования.

Дальнейшие исследования могут быть направлены на:

- Оптимизацию алгоритма для работы с несимметричными объектами,
- Интеграцию дополнительных методов фильтрации для повышения устойчивости к шумам,
- Тестирование метода в реальных условиях (не только в симуляции) с различными типами захватов.

Применение данного подхода может быть полезно в промышленной автоматизации, логистике и сервисной робототехнике, где требуется высокая скорость и точность захвата разнообразных объектов без постоянной перенастройки системы.

### ЛИТЕРАТУРА

1. Dong M., Zhang J. A review of robotic grasp detection technology // *Robotica*. — 2023. — Vol. 41, № 12. — P. 3846–3885. — DOI: 10.1017/S0263574723001285.
2. Redmon J., Angelova A. Real-Time Grasp Detection Using Convolutional Neural Networks [Электронный ресурс] // arXiv. — 2014. — arXiv:1412.3128v2 [cs.RO]. — Режим доступа: <https://doi.org/10.48550/arXiv.1412.3128> (дата обращения: 16.05.2025). — Принято к публикации в ICRA 2015.
3. Shi M., Hou J., Li Z. et al. NG-Net: No-Grasp annotation grasp detection network for stacked scenes [Электронный ресурс] // *Journal of Intelligent Manufacturing*. — 2025. — Vol. 36. — P. 1477–1490. — Режим доступа: <https://doi.org/10.1007/s10845-024-02321-6> (дата обращения: 22.06.2025).
4. Xie Z., Liang X., Canale R. Learning-based robotic grasping: A review [Электронный ресурс] // *Frontiers in Robotics and AI*. — 2023. — Vol. 10. — Режим доступа: <https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2023.1038658> (дата обращения: 2.07.2025).
5. Xiao Y. Integrating CNN and RANSAC for improved object recognition in industrial robotics [Электронный ресурс] // *Systems and Soft Computing*. — 2025. — Vol. 7. — P. 200–240. — Режим доступа: <https://www.sciencedirect.com/science/article/pii/S2772941925000584> (дата обращения: 2.07.2025).
6. Edeh M.O., Dalal S., Iwendi C., Seth B., Odinakachi N., Chichi A.M. Management and prediction of navigation of industrial robots based on neural network [Электронный ресурс] // *International Journal of Systems Engineering and Management*. — 2024. — P. 497–519. — Режим доступа: <https://doi.org/10.1504/IJSEM.2024.140940> (дата обращения: 6.07.2025).

7. Morrison D., Corke P., Leitner J. Closing the Loop for Robotic Grasping: A Real-time, Generative Grasp Synthesis Approach [Электронный ресурс] // arXiv.org. — 2018. — Режим доступа: <https://doi.org/10.48550/arXiv.1804.05172> (дата обращения: 18.06.2025).
8. Arruda E. Generative and predictive models for robust manipulation [Электронный ресурс] // дис. ... д-ра философии (PhD) / University of Birmingham. — 2021. — Режим доступа: <http://etheses.bham.ac.uk/id/eprint/10305> (дата обращения: 14.07.2025).
9. Gu Y., Wei D., Du Y., Cao J. Cooperative Grasp Detection using Convolutional Neural Network // Journal of Intelligent & Robotic Systems. — 2023. — Vol. 110, № 1. — P. 5. — DOI: 10.1007/s10846-023-02028-5.
10. Shi M., Lu H., Li Z.-X. et al. Accurate Robotic Grasp Detection with Angular Label Smoothing // Journal of Computer Science and Technology. — 2023. — Vol. 38, № 5. — P. 1149–1161. — DOI: 10.1007/s11390-022-1458-5.

---

© Андреев Андрей Валерьевич (ext13033@gmail.com)

Журнал «Современная наука: актуальные проблемы теории и практики»