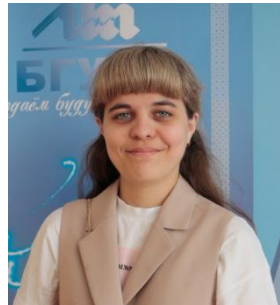


УДК 004.932: 004.451.2

ОБРАБОТКА ВХОДНЫХ ИЗОБРАЖЕНИЙ ДЛЯ КЛАСТЕРИЗАЦИИ И КЛАССИФИКАЦИИ ПРОДУКТОВ



А.В. Деркач
Магистрант кафедры
электронных вычислительных
машин БГУИР
i.angelika897@gmail.com



И.И. Фролов
Доцент кафедры
электронных вычислительных
машин БГУИР, кандидат
технических наук, доцент
frolov@bsuir.by

А.В. Деркач

Окончила Белорусский государственный университет информатики и радиоэлектроники по специальности «Вычислительные машины, системы и сети». Обучается в магистратуре БГУИР по специальности «Компьютерная инженерия», выполняет исследования в рамках магистерской диссертации на тему «Автоматизированная система анализа продуктов по изображениям для рецептурного подбора».

И.И. Фролов

Окончил Белорусский государственный университет информатики и радиоэлектроники. Доцент кафедры электронных вычислительных машин БГУИР. Проводит научные исследования в области машинного обучения и компьютерного зрения, участвует в проектировании и разработке систем технического зрения.

Аннотация. В работе выполнен комплексный анализ методов обработки входных изображений в контексте модели нейронной сети, разработанной для решения задачи классификации продуктов. В данном исследовании рассмотрены основные этапы обработки, включая шаги масштабирования изображений, их нормализации, преобразования данных в необходимый формат, а также применение дополнительных фильтров с целью повышения точности классификации.

Ключевые слова: обработка изображений, машинное обучение, нейронные сети, классификация продуктов.

Введение. В современном цифровом окружении, характеризующемся непрерывным увеличением информационных потоков, вопросы эффективной обработки и анализа данных приобретают первостепенное значение. С усовершенствованием технологий обработки изображений и искусственного интеллекта, автоматизированные системы анализа по изображениям предоставляют новые перспективы для более эффективного решения задач. Способность к автоматическому распознаванию и классификации продуктов приобретает стратегическое значение для бизнес-процессов, улучшая логистические процессы, оптимизируя управление инвентаризацией и обогащая пользовательский опыт.

Однако, перед тем как приступить к классификации, необходимо уделить внимание важности предварительной обработки входных изображений. В свете разнообразия условий съемки, освещения и ракурсов, изображения могут содержать шумы и негативные аспекты, влияющие на точность классификации. В этом контексте применение методов обработки изображений становится неотъемлемым этапом,

направленным на стандартизацию / нормализация и предобработку данных перед их подачей на входы нейронной сети.

Для выполнения задачи обнаружения продуктов на изображениях в рамках выполняемой работы выбраны следующие нейронные сети: *CNN*, *RCNN* и *YOLO* – как одни из наиболее популярных и распространенных при решении задач данного класса (детекция объектов на изображении) [1].

Глубокие слои *CNN* обеспечивают высокую степень извлечения признаков, но требуют сложной обработки входного изображения на различных масштабах. *RCNN* разработана для улучшения проблемы обработки различных масштабов изображений, используя региональные предложения, но характеризуется высокими вычислительными требованиями и долгим процессом обучения из-за необходимости обработки каждого региона независимо. *YOLO* представляет собой подход, который обрабатывает изображение в единой сети и делает прогнозы в несколько сеток, что делает его намного быстрее по сравнению с *RCNN* и классическими *CNN*, но не таким точным в обнаружении маленьких объектов.

Для обучения модели нейронной сети необходимо подобрать изображения продуктов с высоким качеством и разнообразием, чтобы обеспечить эффективное обучение и обобщение модели. Важно учесть следующие аспекты: разнообразие и баланс классов продуктов, различные углы и фоны изображений. Предварительная обработка изображений для каждого из алгоритмов направлена на подготовку входных данных для обеспечения эффективного обучения и более точного обнаружения объектов.

Масштабирование изображений. Модели нейронных сетей проходят обучение на наборе данных с изображениями фиксированного размера. В контексте практических приложений неизбежно возникает проблема обработки изображений с различными размерами, что поднимает необходимость приведения изображений к единому размеру входного слоя модели. Масштабирование, в данном контексте, представляет собой процесс изменения размеров изображения таким образом, чтобы оно соответствовало заранее заданным размерам входного слоя нейронной сети.

Размер входного слоя может варьироваться в зависимости от конкретной модели; однако, в рамках широкоиспользуемых нейронных сетей данного класса, наиболее часто выбираемый диапазон размеров входного слоя колеблется от 150×150 до 512×512 пикселей, как отмечено в обзорах литературы. В данном исследовании осуществляется использование модели, предварительно обученной на изображениях с размерами 512×512 пикселей, поскольку сохранение более высокого разрешения изображения существенно для точного определения продуктов. Иллюстрация процесса приведения изображений к единообразному размеру представлена на рисунке 1.

При обсуждении увеличения масштаба изображения под масштабированием в большую сторону подразумевается увеличение его физических размеров. Этот процесс может быть осуществлен двумя методами: либо путем преобразования изображения в целевой размер с использованием методов, включающих в себя растяжение изображения, либо путем заполнения изображений. В данном контексте заполнение предполагает добавление дополнительных строк и столбцов, не несущих смысловой нагрузки, с использованием нулевых значений. Рекомендуется применять методы интерполяции и генерации новых пикселей для увеличения размера с сохранением деталей изображения.

Обратный процесс – уменьшение размера изображения. Среди распространенных методов уменьшения размера изображения [2] выделяются следующие: билинейная интерполяция – метод, в основе которого лежит использование линейных комбинаций значений пикселей в окрестности целевой позиции для вычисления новых значений; бикубическая интерполяция – метод, который включает в себя более сложные вычислительные процессы, учитывающие значения пикселей вблизи точки интерполяции;

свертка – применение сверточных фильтров, таких как фильтр Гаусса, которые способны размывать изображение, сопровождая уменьшение его размера, но при этом сохраняя существенные детали и уменьшая риск потери информации.



Рисунок 1. Пример приведения размеров изображений к единым значениям

В рамках задачи по анализу продуктов на изображениях будет использован метод бикубической интерполяции для увеличения и уменьшения размера изображения к ожидаемым значениям входа нейронной сети. Выбор метода обусловлен тем, что он обеспечивает хорошую производительность и лучшую точность, по сравнению с билинейной интерполяцией, для разработки системы.

Процесс преобразования данных в соответствующий формат. Распространенные цветные изображения обычно сохраняются в формате *RGB*, где цвет изображения представлен комбинацией трех основных цветов: красного (*Red*), зеленого (*Green*) и синего (*Blue*). Каждый из указанных цветов характеризуется численным значением, представляющим интенсивность соответствующего цвета. Такие числовые значения обычно охватывают диапазон от 0 до 255, где 0 соответствует отсутствию цвета (черный), а 255 представляет максимальную интенсивность цвета (полностью насыщенный цвет).

Формат *RGB* может быть представлен в виде массива, состоящего из трех чисел, которые представляют интенсивность красного, зеленого и синего цветов соответственно. Например, при наличии цвета с интенсивностью $R=255$, $G=0$, $B=128$, его можно записать в виде массива $[255, 0, 128]$. Такой подход облегчает хранение и передачу информации о цвете, а также проведение разнообразных операций с изображениями.

Формат *BGR* (*Blue, Green, Red*) часто применяется в области компьютерного зрения и обработки изображений, особенно при использовании специфических библиотек и фреймворков, таких как *OpenCV* и *TensorFlow*[3]. В контексте моделей нейронных сетей важно учесть необходимость адаптации изображений из формата *RGB* в формат, совместимый с конкретной моделью. Процесс преобразования из формата *RGB* в формат *BGR* и обратно реализуется путем перестановки значений цветовых каналов. Данное преобразование можно описать следующим образом: для каждого пикселя изображения $[R, G, B]$ выполняется замена значения R на позицию B , а значения B на позицию R .

Нейронные сети обычно оперируют с тензорами входных данных, где тензор представляет собой математическую абстракцию в виде многомерного массива данных. В рамках глубокого обучения и нейронных сетей, тензоры представляют собой основной формат данных. Они могут быть представлены в виде скаляров (тензор нулевого ранга), векторов (тензор первого ранга), матриц (тензор второго ранга) или более высокоразмерных массивов.

Для преобразования изображения в тензор требуется провести его конвертацию в массив числовых значений, представляющих интенсивности пикселей изображения с использованием чисел с плавающей запятой. В случае цветного изображения происходит формирование трехмерного массива, где высота и ширина определяются соответственно высотой и шириной изображения (512×512), а глубина отражает количество цветовых каналов в изображении (в данном контексте – 3 канала RGB). Графическое представление полученного массива представлено на рисунке 2.

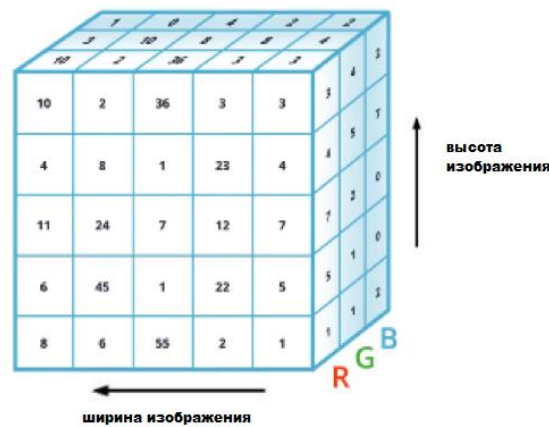


Рисунок 2. Модель тензора для передачи на вход нейронной сети

Нормализация данных. В рамках нейронных сетей процесс нормализации данных представляет собой преобразование значений характеристик с целью их выравнивания в соответствии с установленным стандартом или распределением [4]. В данном контексте нормализация применяется к значениям пикселей изображения с целью оптимизации производительности модели. Этот процесс включает в себя масштабирование значений пикселей из исходного диапазона 0-255 до нового диапазона 0-1. Нормализация осуществляется путем деления значения каждого пикселя на максимальное значение в изначальном диапазоне (255). Иллюстрация примера нормализации, применяемой в данной модели, представлена на рисунке 3.

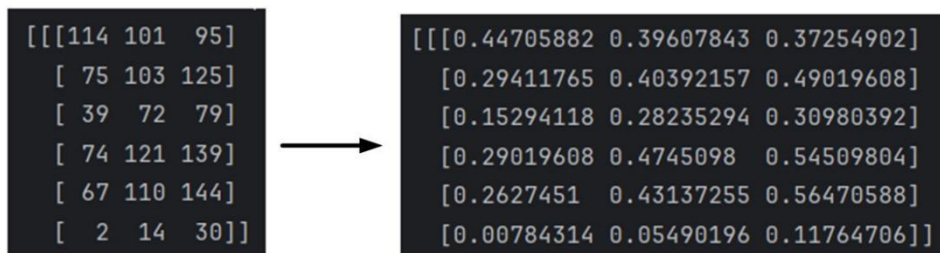


Рисунок 3. Нормализация значений пикселей

Применение дополнительных фильтров. Для повышения точности классификации продуктов с применением нейронной сети на входных цветных изображениях предполагается использовать разнообразные фильтры. Эти фильтры спроектированы с целью улучшения контрастности и яркости, а также для снижения воздействия шумов и выделения контуров. Повышение контрастности способствует выделению деталей и характеристик продуктов, в то время как применение контурных фильтров помогает сети более эффективно распознавать форму продуктов.

В разрабатываемой системе используется медианный фильтр для уменьшения шума на изображениях [5]. Принцип работы медианного фильтра заключается в замене значения каждого пикселя на медиану значений пикселей в окрестности. Обычно медианный фильтр применяется к окну фиксированного размера, которое перемещается по всему изображению. Размер окна определяет количество соседних пикселей, участвующих в процессе усреднения. При применении медианного фильтра резкие перепады значений яркости на изображениях не изменяются. Это очень важно, поскольку, контуры на изображениях несут основную информацию. В то же время импульсные помехи, протяженность которых составляет менее половины окна, подавляются. Данный фильтр часто используется для устранения шумов на изображениях, таких как “соль и перец”, сохраняя при этом края объектов. Применительно к RGB изображению, медианный фильтр применяется к каждому цветовому каналу отдельно. Для реализации фильтра использовалась библиотека *OpenCV*, в результате был получен результат, представленный на рисунке 4.



Рисунок 4. Результат обработки входного изображения медианным фильтром

Как видно по результатам приведенного образца, медианный фильтр эффективно удаляет импульсные и аддитивные шумы, полученные при некачественной съемке, при этом сохранив резкие края объектов. Качественные изображения, полученные после применения медианного фильтра, впоследствии подаются на вход модели нейронной сети для обнаружения продуктов и способствуют более эффективному и точному результату классификации.

В настоящей работе проведен комплексный анализ методов предварительной обработки входных изображений в контексте модели нейронной сети, разработанной для решения задачи классификации продуктов. Рассмотрены ключевые этапы обработки, включая масштабирование изображений, их нормализацию, преобразование данных в соответствующий формат, а также использование дополнительных фильтров с целью повышения точности классификации.

Список литературы

- [1] Козак, А. В. Object Detection для автоматизации обработки документов / Козак А. В. // Компьютерные системы и сети : сборник тезисов докладов 56-й научной конференции аспирантов, магистрантов и студентов, Минск, апрель-май 2020 года / Белорусский государственный университет информатики и радиоэлектроники. – Минск : БГУИР, 2020. – С. 139-141.
- [2] Трубаков А.О., Селейкович М.О. Сравнение интерполяционных методов масштабирования растровых изображений // Научно-технический вестник Брянского государственного университета, 2017. – С. 92-97.
- [3] Фролов, И. И. Эффективные библиотеки машинного обучения / И. И. Фролов // Технические средства защиты информации : тезисы докладов XVI Белорусско-российской научно – технической конференции, Минск, 5 июня 2018 г. – Минск: БГУИР, 2017. – С. 94.
- [4] Старовойтов, В. В. Нормализация данных в машинном обучении / Старовойтов В. В., Голуб Ю. И. // Информатика. – 2021. – Т. 18, № 3. – С. 83–96. – DOI : <https://doi.org/10.37661/1816-0301-2021-18-3-83-96>.
- [5] Chan, Raymond H., Chungwa Ho, and Mila Nikolova. "Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization." IEEE Transactions on Image Processing 14.10 (2005): 1479-1485.

Авторский вклад

Деркач Анжелика Валерьевна – проведение исследования, анализ существующих методов обработки изображений, формирование структуры статьи.

Фролов Игорь Иванович – руководство исследованием, постановка задачи исследования, рецензирование статьи.

PROCESSING OF INPUT IMAGES FOR PRODUCT CLUSTERING AND CLASSIFICATION

A.V. Dziarkach

*Master's student of the Department
of Electronic Computing Machines
of BSUIR*

I.I. Frolov

*Associate Professor of the
Department of Electronic
Computing Machines of BSUIR,
Candidate of Technical Sciences,
Associate Professor*

Abstract. A comprehensive analysis of input image processing methods for a neural network model designed for product classification has been carried out. The main stages of processing are highlighted, including image scaling, data conversion and normalization into the required format, as well as the use of additional filters to improve classification accuracy.

Keywords: image processing, machine learning, neural networks, product classification.