

КЛАССИФИКАЦИЯ ПАЛАТОК НА УЛИЧНЫХ БАЗАРАХ С ИСПОЛЬЗОВАНИЕМ CNN

А. ИБРАГИМОВ¹, Р. ИСАЕВ¹

¹Международный университет «Ала-Тоо»
(г. Бишкек, Кыргызстан)

E-mail: ruslan.isaev@alato.edu.kg

Аннотация. В этом исследовании представлена модель сверточной нейронной сети для классификации палаток на уличных базарах, особенно в Центральной Азии, где продавцы используют различные конструкции палаток для продажи таких товаров, как продукты питания, сухофрукты и чистящие средства. Разнообразие типов палаток создает проблемы для эффективной организации и анализа рыночной конъюнктуры, а традиционные методы каталогизации вручную оказываются трудоемкими и неэффективными. Чтобы решить эту проблему, был создан пользовательский набор данных из 130 уникальных изображений палаток, каждое из которых было сделано с разных ракурсов, который был расширен до 2000 изображений за счет увеличения объема данных. Этот набор данных позволил нам обучить модель сверточной нейронной сети, способную разделить палатки на шесть различных типов, достигнув точности 90-95%. За пределами Центральной Азии этот подход может быть адаптирован к аналогичным рынкам в других развивающихся регионах. В этой статье подробно описывается разработка набора данных, архитектура сверточной нейронной сети, результаты экспериментов и потенциальные приложения, направленные на то, чтобы заложить основу для более широкого применения искусственного интеллекта в анализе региональных рынков и экономического развития.

Abstract. This study presents a convolutional neural network model for classifying tents in street bazaars, especially in Central Asia, where vendors use various tent designs to sell goods such as groceries, dried fruits and cleaning products. The variety of tent types creates problems for effective organization and analysis of market conditions, and traditional manual cataloging methods turn out to be time-consuming and inefficient. To solve this problem, a custom dataset of 130 unique images of tents was created, each of which was taken from different angles, which was expanded to 2,000 images by increasing the amount of data. This data set allowed us to train a convolutional neural network model capable of dividing tents into six different types, achieving an accuracy of 90-95%. Outside of Central Asia, this approach can be adapted to similar markets in other developing regions. This article describes in detail the development of the dataset, the architecture of convolutional neural network, experimental results and potential applications aimed at laying the foundation for the wider application of artificial intelligence in the analysis of regional markets and economic development.

Введение

Уличные базары, особенно в Центральной Азии - это оживленные рынки, где продавцы продают товары в палатках. Такие рынки, которые часто встречаются в развивающихся регионах где размещается множество палаток с товарами, такие как сухофрукты, мясо, овощи и фрукты, продукты питания и чистящие средства. Палатки, используемые продавцами, сильно различаются по структуре, размеру и планировке в зависимости от продаваемых товаров, что приводит к разнообразию и динамичности обстановки. Однако это разнообразие также создает уникальную проблему, когда речь заходит о категоризации и организации рыночной планировки.

Эффективная классификация палаток на основе их визуальных характеристик может дать ценную информацию о динамике рынка, помогая понять распределение поставщиков и размещение продуктов. Традиционные методы каталогизации и категоризации палаток в основном выполнялись вручную, с опорой на наблюдения человека, что может быть трудоемким. Автоматизация процесса классификации может привести к лучшему пониманию рыночных операций, предоставлению данных для управления рынком и позволит поставщикам и организаторам принимать решения, основанные на данных.

В этом исследовании рассматривается задача классификации палаток на уличных базарах с использованием модели сверточной нейронной сети (CNN), специально обученной распознавать шесть различных типов палаток. Отсутствие общедоступного набора данных для таких палаток привело к разработке пользовательского набора данных, палатки снятые под разными углами, чтобы максимально разнообразить характеристики. Для дальнейшего расширения набора данных для эффективного обучения модели были применены методы увеличения объема данных, которые позволили расширить набор данных до 2000 изображений.

Целью данной работы является разработка и оценка классификатора уличных палаток на базе CNN.

Благодаря возможности быстрой и автоматизированной классификации палаток, модель может быть использована для создания цифровых классификационных карт базаров. Эти карты, основанные на данных, поступающих от таких устройств, как камеры с обзором на 360 градусов, могут позволить администраторам рынка отслеживать расположение палаток и распределение продукции с минимальным ручным вмешательством. Кроме того, эта модель может помочь предприятиям в организации и структурировании крупномасштабных рынков, а ее применение распространится за пределы Центральной Азии на другие развивающиеся регионы с аналогичными рынками сбыта.

Методология. Сбор данных

Для создания набора данных, необходимого для классификации палаток на уличных базарах, были сделаны фотографии палаток в одном из базаров Кыргызстана. Было собрано 126 фотографий различных палаток, каждая из которых была заснята с трех различных ракурсов, чтобы обеспечить разнообразие углов съемки. Эти ракурсы включали виды спереди и сбоку, что позволило захватить особенности каждой палатки. Примеры фотографий вы можете наблюдать ниже (Рис. 1, Рис. 2, Рис. 3, Рис. 4, Рис. 5, Рис. 6).



Рис. 1. dried_fruits



Рис. 2. dried_fruits



Рис. 3. dried_fruits



Рис. 4. household_goods



Рис. 5. household_goods



Рис. 6. household_goods

Обработка

Для подготовки данных к обучению модели были предприняты следующие шаги:

1. Увеличение объема данных: Методы аугментации с помощью библиотеки Augmentor, такие как случайные искажения, повороты и отражения, применялись для увеличения объема данных и повышения устойчивости модели. Эти техники подробно рассмотрены в обзоре методов глубокого обучения для распознавания категорий еды [3]. Были выполнены следующие операции:

- Случайная дисторсия: Применена с вероятностью 100% для изменения формы изображений и создания более разнообразных примеров.

ИНФОРМАЦИОННЫЕ РАДИОСИСТЕМЫ И РАДИОТЕХНОЛОГИИ 2024»

Открытая республиканская научно-практическая интернет-конференция,
21-22 ноября 2024 г., Минск, Республика Беларусь

- Вращение: С вероятностью 70% изображения поворачивались влево или вправо на случайный угол до 25 градусов.
 - Генерация новых изображений: было создано 335 новых изображений для каждого типа палаток, чтобы обеспечить баланс и увеличение набора данных.
 -
 - 2. Изменение размера изображений: После аугментации все изображения были изменены до размера 256x256 пикселей с использованием библиотеки PIL и метода интерполяции LANCZOS. Это обеспечило единообразие входных данных для модели CNN.
- В окончательном итоге пример фотографий вы можете наблюдать ниже (Рис. 7, Рис. 8, Рис. 9, Рис. 10, Рис. 11, Рис. 12).



Рис. 7. dried_fruits



Рис. 8. dried_fruits



Рис. 9. dried_fruits



Рис. 10. household_goods



Рис. 11. household_goods



Рис. 12. household_goods

Построение модели

Для классификации палаток на уличных базарах была разработана сверточная нейронная сеть (CNN), используя библиотеки TensorFlow и Keras. Модель прошла несколько этапов подготовки и обучения, как описано ниже:

1. **Импорт библиотек и подготовка данных:**
 - Импортированы необходимые библиотеки, такие как PIL для обработки изображений, matplotlib для визуализации, numpy для работы с массивами, и инструменты из библиотеки scikit-learn для разбиения данных и кодирования меток.
 - Данные были загружены из директории с аугментированными изображениями, организованными по классам: мясо, сухофрукты, специи, бытовые товары, фрукты и овощи, продукты питания. Каждый класс был представлен в формате массива изображений X, а метки классов были сохранены в y.
 - Изображения были нормализованы путем деления на 255.0 для масштабирования значений пикселей в диапазон [0, 1].
2. **Разделение данных на тренировочные и тестовые:**
 - Данные были случайно разделены на тренировочные и тестовые наборы в соотношении 80/20.
 - Метки классов были закодированы в числовой формат с помощью LabelEncoder.

3. Архитектура модели CNN:

Архитектура модели включает несколько сверточных и полносвязных слоев с функциями активации ReLU и слоями Dropout для регуляризации. Подобные архитектурные решения обсуждаются в исследовании по обучению глубоких признаков для распознавания сцен с использованием базы данных Places [1]

- **Входной слой:** Модель использует сверточный слой с 32 фильтрами размером (3, 3) и функцией активации ReLU, принимающий изображения размером 256x256 с тремя каналами (RGB).
- **Скрытые слои:**
 - Второй сверточный слой с 64 фильтрами и функцией активации ReLU.
 - Первый слой подвыборки (MaxPooling) с размером ядра (2, 2), за которым следует слой Dropout с вероятностью 0.25 для предотвращения переобучения.
 - Третий сверточный слой с 128 фильтрами и функцией активации ReLU, за которым следует еще один слой MaxPooling.
 - Четвертый сверточный слой с 256 фильтрами и функцией активации ReLU, с последующим слоем MaxPooling.
- **Полносвязные слои:**
 - Слой Flatten для преобразования данных в одномерный массив.
 - Полносвязный слой с 512 нейронами и функцией активации ReLU.
 - Слой Dropout с вероятностью 0.5 для регуляризации.
 - Выходной слой с 6 нейронами и функцией активации softmax для многоклассовой классификации.

4. Компиляция и обучение модели:

- Модель была скомпилирована с использованием оптимизатора Adam, функции потерь sparse_categorical_crossentropy, и метрики точности.
- Модель обучалась на тренировочном наборе в течение 20 эпох с размером пакета 64. 20% тренировочных данных были использованы для проверки на каждом этапе обучения.

5. Оценка производительности модели:

- Точность модели оценивалась на тестовом наборе данных, что соответствует стандартным практикам в области распознавания изображений, как описано в исследовании по обучению глубоких признаков для распознавания сцен [1].

Заключение

Наша модель достигла точности (1) классификации 92%. Эти результаты подчеркивают потенциал использования глубокого обучения для автоматизации анализа рыночных условий, что может значительно упростить управление рыночными операциями и повысить их эффективность.

$$\text{Точность} = \frac{\text{количество правильно предсказанных классов}}{\text{общее количество предсказаний}} \quad (1)$$

Возможности применения разработанной модели выходят за пределы рынка в Центральной Азии и могут быть адаптированы к аналогичным рынкам в других развивающихся регионах. Разработанная система также.

В дальнейшем планируется расширить исследование путем интеграции более сложных архитектур нейронных сетей и применения методов для улучшения обобщающей способности модели, а также рассмотреть возможность использования дополнительных данных с различных рынков для повышения точности классификации и устойчивости модели к изменениям условий.

Список использованных источников

1. Zhou B. et al. Learning deep features for scene recognition using places database //Advances in neural information processing systems. – 2014. – Т. 27.
2. Kim D. et al. Understanding tourists' urban images with geotagged photos using convolutional neural networks //Spatial Information Research. – 2020. – Т. 28. – С. 241-255.
3. Zhang Y. et al. Deep learning in food category recognition //Information Fusion. – 2023. – Т. 98. – С. 101859.