

УДК 004.932.75

Т. В. Ермоленко, И. Е. Самородский

Государственное образовательное учреждение высшего профессионального образования
«Донецкий национальный университет», г. Донецк
83001, г. Донецк, ул. Университетская, 24

АНАЛИЗ ЭФФЕКТИВНОСТИ АРХИТЕКТУР ГЛУБОКИХ НЕЙРОСЕТЕЙ ДЛЯ КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ ТОВАРОВ

T. V. Yermolenko, I. Ye. Samorodsky

State Educational Institution of Higher Professional Education «Donetsk National University»
83001, Donetsk, University st, 24

ANALYSIS OF THE EFFICIENCY OF DEEP NEURAL NETWORKS ARCHITECTURES FOR THE CLASSIFICATION OF PRODUCTS IMAGES

Т. В. Єрмоленко, І. Є. Самородський

Державна освітня установа вищої професійної освіти
«Донецький національний університет», м. Донецьк
83001, м. Донецьк, вул. Університетська, 24

АНАЛІЗ ЕФЕКТИВНОСТІ АРХІТЕКТУР ГЛИБОКИХ НЕЙРО- МЕРЕЖ ДЛЯ КЛАСИФІКАЦІЇ ЗОБРАЖЕНЬ ТОВАРІВ

В статье проводится анализ эффективности использования различных архитектур глубоких нейросетей в задаче классификации товаров по их изображениям. Для решения проблем несбалансированных обучающих данных и схожести объектов разных классов предлагается использовать балансировку классов и архитектуру API-Net. Исследования проводятся на наборе данных RP2K dataset.

Ключевые слова: сверточные нейронные сети, ResNet, InceptionV3, алгоритм UMAP, API-Net, алгоритм Class Balanced Loss.

The article analyzes the effectiveness of using various architectures of deep neural networks in the task of classifying products by their images. To solve the problems of unbalanced training data and the similarity of objects of different classes, it is proposed to use class balancing and the API-Net architecture. Research is carried out on the RP2K dataset.

Key words: convolutional neural networks, ResNet, InceptionV3, UMAP algorithm, API-Net, Class Balanced Loss algorithm.

У статті проводиться аналіз ефективності використання різних архітектур глибоких нейромереж у задачі класифікації товарів за їхніми зображеннями. Для вирішення проблем незбалансованих навчальних даних та схожості об'єктів різних класів пропонується використовувати балансування класів та архітектуру API-Net. Дослідження проводяться на наборі даних RP2K dataset.

Ключові слова: згорткові нейронні мережі, ResNet, InceptionV3, алгоритм UMAP, API-Net, алгоритм Class Balanced Loss.

Введение

С ростом цифровизации и массовым переходом в онлайн e-commerce (электронная торговля) стала драйвером ключевых изменений в мировой экономике. Доступ к электронной коммерции сегодня возможен с любого смарт-устройства, что является одной из главных причин роста числа онлайн-покупателей и объема интернет-продаж. Поддержка электронной коммерции в России на период до 2024 года регулируется нацпроектом «Цифровая экономика». Одной из основных целей документа является создание устойчивой, безопасной и общедоступной информационно-телекоммуникационной инфраструктуры высокоскоростной передачи, обработки и хранения больших объемов данных. Не вызывает сомнения тот факт, что цифровые изображения товара занимают далеко не последнее место в этих данных, а повышение качества категоризации товаров и их ранжирование являются крайне востребованными направлениями в сфере e-commerce. Качественная автоматическая классификация товаров позволит избежать участия человека в процессе категоризации, что приведет к понижению серьезных ошибок, вследствие которых товары могут оказаться в неподходящих местах хранения или же отображаться пользователю в некорректных разделах онлайн-магазина.

В настоящее время системы автоматической классификации товаров по их изображениям не получили широкого развития в связи с имеющимися трудностями, которые вызваны большим количеством товарных категорий, а также похожим внешним видом разных продуктов.

Методы классификации изображений, как правило, основаны на глубоком обучении с помощью сверточных нейронных сетей. Данная работа посвящена анализу эффективности нейронных сетей в задаче классификации изображений товаров, что в свете вышесказанного представляется актуальным.

Целью данной работы является сравнительный анализ эффективности различных моделей глубокого обучения в задаче классификации товаров на изображениях.

Для достижения цели необходимо:

- сформировать обучающую выборку на основе данных RP2K dataset [1] и провести ее балансировку;
- обучить архитектуры InceptionV3 (GoogLeNet) [2], ResNet-34 [3], Attentive Pairwise Interaction Network (API-Net) [4];
- оценить качество полученных моделей на тестовой выборке.

Описание набора данных RP2K dataset

Большинство существующих датасетов имеют несколько существенных недостатков: они либо содержат относительно небольшое количество категорий товаров, либо были собраны в идеальной среде (в лаборатории с отличным освещением и мощной камерой). У RP2K dataset есть ряд преимуществ перед другими коллекциями данных:

- большой охват данных. В данном датасете собрано большое количество категорий товаров, каждая из которых содержит в среднем 160 изображений;
- данные реалистичны. В отличие от большинства других датасетов, собранных в идеальных условиях, текущий максимально приближен к реальным условиям;
- разнообразные метаданные. Для каждого изображения предоставлены несколько полей метаданных. 2388 товаров можно разделить на 7 мета-категорий в зависимости от их типов и еще на 7 мета-категорий в зависимости от их формы.

Изображения в RP2K dataset разбиты на две части: изображения отдельных объектов и фотографии стеллажей товаров.

Для каждого изображения предоставляются обширные аннотации, включая идентификатор SKU, название продукта, бренд, тип продукта, форму, размер, вкус/запах и ссылку на ограничивающую рамку на соответствующее изображение на полке. На рис. 1 показаны некоторые примерные атрибуты изображений объектов. Некоторые атрибуты могут не присутствовать у всех изображений.



Рисунок 1 – Пример атрибутов изображений объектов в RP2K dataset

Каждое изображение объекта принадлежит двум мета-категориям. Первая относит продукт к одному из типов, который отражает размещение продуктов, т.е. продукты одного типа обычно размещают в том же месте торгового ряда. Всего таких типов семь: молочные продукты, ликеры, пиво, косметика, безалкогольные напитки, табак и приправы.

Вторая категория классифицирует продукт по форме. Таких форм всего семь типов: бутылка, жестяная банка, коробка, сумка, банка, бутылка с ручкой и упаковка. Образцы изображений товаров для разных мета-категорий показаны на рис. 2.

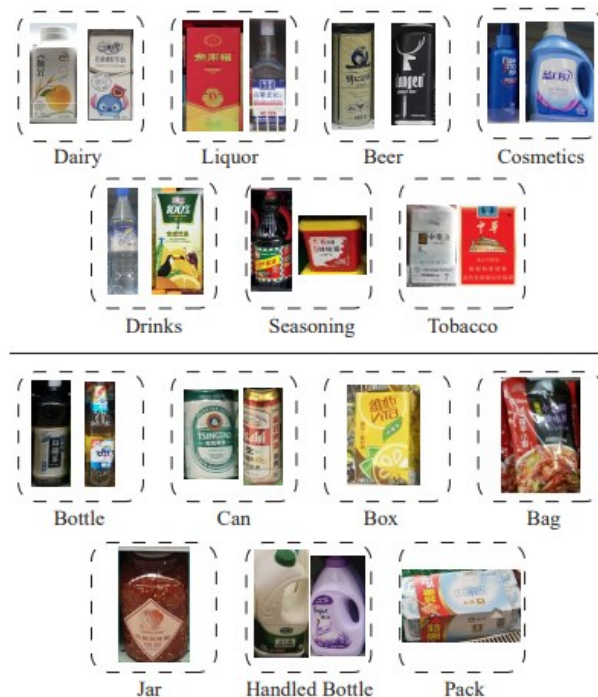


Рисунок 2 – Примеры изображений товаров разных мета-категорий

Общий объем RP2K dataset составляет 384 311 изображений отдельных объектов, разбитых на 2 388 классов товаров. Данные были разбиты на тренировочные/проверочные в соотношении 90/10.

Подробная статистика по количеству образцов по каждой мета-категории для обучения и тестирования моделей представлена в табл. 1.

Таблица 1 – Количество образцов по мета-категориям

Мета-категории	Тренировочные	Тестовые	Всего
молочные продукты	78,288	8,867	87,155
ликеры	16,753	1,939	18,692
пиво	39,786	4,540	44,326
косметика	7,393	932	8,325
безалкогольные напитки,	29,241	3,405	32,646
приправы	137,082	15,479	152,561
табак	36,311	4,295	40,606
бутылка	164,939	18,327	183,266
жестяная банка	44,461	4,940	49,401
коробка	27,347	3,039	30,386
сумка	15,350	1,705	17,055
банка	13,913	2,657	16,570
бутылка с ручкой	54,895	6,099	60,994
упаковка	23,949	2,690	26,639
Всего	344,854	39,457	384,311

Для визуализации данных использован алгоритм UMAP (*Uniform Manifold Approximation and Projection*) [5]. Алгоритм UMAP уменьшает размерность признакового пространства, что весьма полезно как в целях визуализации (данные можно спроецировать на двумерную или трехмерную плоскость), так и для ускорения обучения нейросети.

На рис. 3 представлены результаты визуализации.

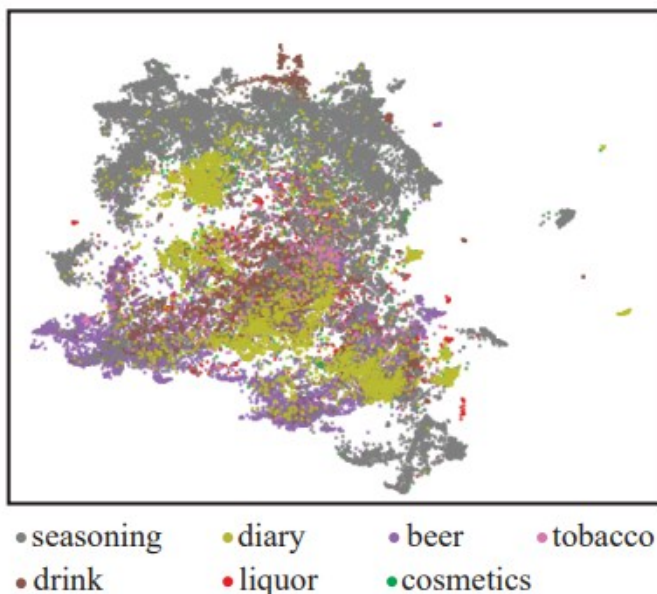


Рисунок 3 – Результаты визуализации с помощью UMAP

Видно, что большинство данных хорошо разделены, если их классифицировать по форме. Хотя некоторые данные, особенно те, которые относятся к категории с относительно небольшим количеством изображений (например, косметика или спиртные напитки), имеют тенденцию распространяться по всему множеству. Кроме того, возникает еще одна проблема: товары из разных мета-категорий, как правило, имеют отличительные визуальные особенности, но продукты из одной категории могут выглядеть очень похожими, даже если они разные.

Описание и оценка различных архитектур для классификации изображений товаров

В работе проведен сравнительный анализ эффективности архитектур сверточных сетей InceptionV3 (GoogLeNet) [2], ResNet-34 [3], а также Attentive Pairwise Interaction Network (API-Net) [4] в задачах классификации товаров по их изображению.

Как известно, сверточные сети типичной архитектуры представляют собой многоступенчатый каскад сетей прямого распространения. Основные типы слоев для таких сетей: несколько сверточных слоев, слоев объединения (pooling) и слоев сетей прямого распространения [6].

Выход сверточного слоя представляет карту признаков: каждый элемент выхода получен применением операции свертки между входным слоем и рецептивным полем с применением определенного фильтра (ядра) и последующим действием нелинейной функции активации. Слои пулинга выполняют операцию понижения уровня дискретизации для карты признаков (например, посредством усреднения или вычисления максимума в пределах некоторой конечной области).

Выбор архитектур ResNet и GoogLeNet для анализа эффективности классификации товаров по изображениям обоснован тем, что указанные архитектуры продемонстрировали большие возможности в решении задач классификации изображений высокого разрешения, поскольку из-за своей большой емкости они обладают большими возможностями обучения, а кроме того, они предобучены на накопленных базах данных, что позволяет компенсировать отсутствие тех данных, которыми в конкретной задаче не располагают. Для ускорения процесса обучения в этих сетях используется в качестве функции активации сетевого нейрона функция ReLU.

ResNet и GoogLeNet являются многокаскадными, в таких сетях после каскада сверточных слоев и слоев пулинга обычно следует несколько полносвязных слоев. Таким образом, многокаскадная сверточная сеть преобразует матрицу пиксельных значений входного изображения в матрицу окончательных признаков методом сети прямого распространения. Параметры сети, т.е. веса связей сверточных и полносвязных слоев, как правило настраиваются применением метода обратного распространения ошибки, реализуемого посредством классического градиентного спуска.

GoogLeNet – глубокая архитектура с 22 слоями. Для повышения вычислительной эффективности используют модуль Inception, вся архитектура состоит из множества таких модулей, следующих друг за другом.

Идея основного модуля Inception заключается в том, что он сам по себе является небольшой локальной сетью. Вся его работа состоит в параллельном применении нескольких фильтров на исходное изображение. Данные фильтров объединяются, и создается выходной сигнал, который переходит на следующий слой [7].

Для уменьшения вычислительной сложности введены так называемые «узкие места» – слои с фильтром 1×1 , уменьшающие глубину изображения.

В составе GoogleNet есть небольшая подсеть Stem Network. Она состоит из трёх свёрточных слоёв с двумя pooling-слоями и располагается в самом начале архитектуры. Ещё одной особенностью архитектуры является то, что в GoogleNet нет полносвязных слоёв, и она содержит 5 миллионов параметров, что относительно немного для глубоких сетей.

На схеме нейросети можно увидеть небольшие промежуточные «отростки» – это вспомогательные классификационные выходы для введения дополнительного градиента на начальных слоях. Вспомогательный классификатор действует как регуляризатор.

В версии InceptionV3 использованы свертки малого размера: фильтр 5×5 заменяется на две последовательные свертки 3×3 , которые объединяются. Кроме того, применяются асимметричные свертки: 3×3 можно заменить сверткой 1×3 , за которой следует свертка 3×1 . Это приводит к уменьшению параметров, и как следствие, к более быстрому обучению.

Все вышеперечисленные концепции объединены в окончательную архитектуру, показанную на рис. 4.

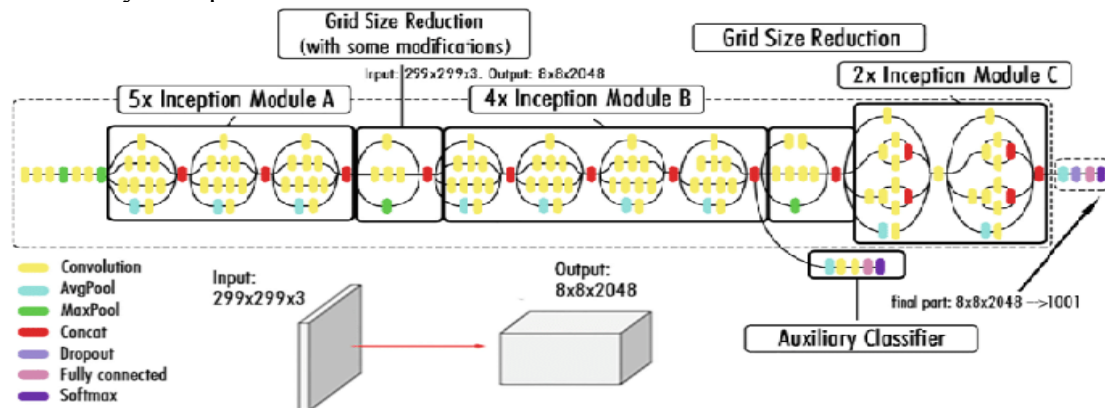


Рисунок 4 – Архитектура InceptionV3

Использование модулей Inception и отсутствие полносвязных слоёв делают GoogLeNet очень эффективной и достаточно точной сетью.

Архитектура ResNet (Residual Network) состоит из 152 слоёв, т.е. намного глубже, чем GoogLeNet. С увеличением глубины сети точность сначала увеличивается, а затем быстро ухудшается, поскольку глубокие модели гораздо хуже поддаются настройке. Для решения проблемы оптимизации и затухающего градиента разработчики ResNet ввели парадигму «остаточного» обучения с помощью соединений быстрого доступа. Вместо каскада свёрточных слоёв для изучения отображения нужной функции напрямую используются остаточные блоки, которые пытаются «подогнать» это отображение, т.е. сеть «перепрыгивает» через каскад слоёв, которые больше не содержат признаков и используются для нахождения остаточной функции $H(x) = F(x) + x$ вместо того, чтобы искать $H(x)$ напрямую (рис. 5).

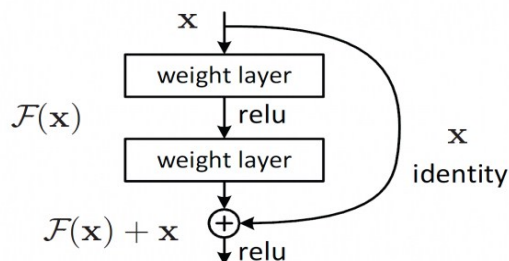


Рисунок 5 – Остаточный блок в ResNet

Нейросеть состоит из большого стека одинаковых остаточных блоков, каждый из которых имеет два свёрточных слоя 3×3 . Периодически число фильтров удваивается, а их размерность уменьшается с шагом 2 (/ 2 в каждом измерении). В самом начале архитектуры присутствует дополнительный свёрточный слой. Также у ResNet нет полносвязных слоёв в конце – используется только один слой с выходными классами. С увеличением числа слоёв для уменьшения размерности изображения применяются точно такие же дополнительные слои, как и в GoogleNet.

ResNet-34 имеет архитектуру ResNet с 34 слоями (3,6 миллиарда параметров). Благодаря «остаточному» обучению ResNet является очень глубокой сетью, которую можно обучить без ухудшения точности. Нейросеть достигла наименьшей ошибки в задачах классификации, которая превзошла даже человеческий результат.

Проведено сравнение стандартно используемых архитектур для классификации изображений в задаче классификации товаров. Результаты оценки эффективности этих моделей представлены в табл. 2. Для оценки точности использовались метрики, широко применяющиеся при оценивании систем компьютерного зрения – top-1 и top-5. Точность top-n – это доля тестовых изображений, для которых правильный ответ модели попал в один из n наиболее вероятных ответов.

Таблица 2 – Точность моделей для разных мета-категорий товаров

Мета-категории	ResNet-34		InceptionV3	
	Top-1	Top-5	Top-1	Top-5
молочные продукты	93.68%	98.45%	93.38%	98.31%
ликеры	76.99%	96.44%	79.57%	97.31%
пиво	96.76%	99.16%	95.19%	98.98%
косметика	87.87%	93.71%	86.58%	96.56%
безалкогольные напитки	94.86%	98.88%	93.65%	98.53%
приправы	97.96%	99.56%	97.28%	99.53%
табак	91.89%	98.06%	91.73%	97.88%
Среднее	95.18%	99.01%	94.69%	98.96%

Из таблицы видно, что ResNet-34 превосходит по качеству InceptionV3, достигая средней точности более 95%, однако, для некоторых категорий, таких как «ликеры» и «косметика», точность гораздо ниже, чем средняя. Это можно объяснить следующими факторами:

- 1) количество обучающих данных этих категорий относительно меньше, чем других;
- 2) внешний вид товаров в этих категориях обычно очень похож.

Для балансировки классов в работе предложено использовать алгоритм Class Balanced Loss (CBL) [8], для решения проблемы похожести объектов разных классов, но одной категории, т.е. детальной классификации внутри одной мета-категории, – архитектуру API-Net [4].

Классификация изображений товаров одной категории

Как показали авторы метода CBL, он обеспечивает значительный рост производительности свёрточных сетей на больших сильно несбалансированных наборах данных по сравнению с часто используемыми методами стратификации и не требует предположений о распределении данных [5]. В связи с чем для задачи классификации товаров по их изображениям выбран для балансировки классов в работе используется CBL.

Метод представляет собой упрощенный вариант случайного покрытия. Ключевая идея CBL – связать каждый образец с небольшой соседней областью вместо единственной точки.

Эффективное количество образцов класса рассчитывается по простой формуле:

$$E_n = \frac{1 - \beta^n}{1 - \beta},$$

где n – количество экземпляров класса, $\beta \in [0, 1)$ – гиперпараметр, контролирующий, насколько быстро E_n растет с увеличением n . В данной работе значение β составило 0.99.

Для детальной классификации внутри одной метакатегории в работе использована архитектура Attentive Pairwise Interaction [6]. API-Net имитирует способность человека сравнивать похожие изображения, определяя контрасты. API-Net сначала изучает вектор общих признаков, чтобы фиксировать семантические различия во входной паре. Затем он сравнивает этот общий вектор с отдельными векторами каждого изображения пары, чтобы сгенерировать гейты – «вентили», пропускающие или не пропускающие информацию. Эти вентильные векторы наследуют взаимный контекст по семантическим различиям, которые позволяют API-Net внимательно улавливать контрастные сигналы путем попарного взаимодействия между двумя изображениями.

Пара мелкозернистых изображений проходят через базовую сверточную сеть (backbone) для извлечения их векторов признаков x_1, x_2 , которые подаются на API-Net в качестве входных данных. Сеть состоит из последовательно связанных модулей: выделения общего вектора признаков (*Mutual Vector Learning*); генерации вентильных векторов или векторов гейтов, (*Gate Vector Generation*) и парного взаимодействия (*Pairwise Interaction*) [6].

Для повышения точности распознавания товаров категорий «ликеры» и «косметика» была использована сеть API-Net, в качестве backbone применялась ResNet-34. Результаты классификации до и после использования API-Net для этих категорий сведены в табл. 3.

Таблица 3 – Точность моделей для мета-категорий «ликеры» и «косметика»

Мета-категории	ResNet-34		ResNet-34 + API-Net	
	Top-1	Top-5	Top-1	Top-5
ликеры	76.99%	96.44%	96.47%	99.33%
косметика	87.87%	93.71%	93.88%	99.03%

Как видно из табл. 3, использование API-Net для детальной классификации существенно повысила точность распознавания: для категории «ликеры» точность стала лучше более чем на 20%, для категории «косметика» – на 6%.

Повышение точности классификации с помощью техник аугментации и *transfer learning*

Аугментация изображений направлена на борьбу с переобучением и повышение обобщающей способности сети. Аугментация позволяет расширить тренировочный набор данных благодаря внесению в них различных искажений.

Чтобы имитировать реальную сценарную среду с нечеткими изображениями, была использована следующая схема увеличения данных:

- добавление постоянной границы с произвольно выбранной шириной от 0 до 30 пикселей;

- обрезка произвольно до 10 пикселей;
- применение трансформации перспективы с произвольными параметрами;
- затемнение / повышение яркости изображения в диапазоне до 20%.

Для качественного обучения свёрточных сетей необходим большой набор данных. Но на таких наборах для обучения требуются огромные ресурсы и долгое время. Техника переноса обучения (англ. *transfer learning*) позволяет использовать нейронные сети, уже обученные на крупных наборах данных, для инициализации обучения для решения других задач. При переносе обучения из предварительно обученной нейронной сети удаляется часть, отвечающая за классификацию объектов, и добавляется новая часть, которая обеспечивает классификацию объектов в поставленной задаче [9], [10]. Техника *transfer learning* ускоряет и упрощает обучение глубоких сетей, при этом повышая точность модели.

В рамках задачи классификации товаров проверена эффективность использования аугментации и *transfer learning* для обучения модели ResNet-34. Результаты сведены в табл. 4.

Таблица 4 – Результаты применения различных техник обучения модели

Используемые техники	Top-1 Acc	Top-5 Acc
Обучение с нуля	95.18%	99.01%
Перенос обучения	95.54%	99.04%
Обучение с нуля с аугментацией	90.41%	94.74%
Перенос обучения с аугментацией	90.89%	95.01%

Как видно из табл. 4, аугментация не увеличивает точность распознавания. Это связано с тем, что текущий датасет уже содержит изображения под различными углами обзора и условиями освещения, поэтому дальнейшее увеличение данных путем аугментации не приносит никаких значительных изменений в точности.

Использование *transfer learning* повысило точность классификации на величину около 0.4%. Небольшой прирост можно объяснить тем, что в наборе данных ImageNet содержатся изображения, похожие на используемый датасет, и сверточная часть нейронной сети ResNet-34 уже хорошо выделяет их характерные признаки.

Выводы

В результате проведенных исследований можно сделать вывод о высокой эффективности использования сверточных сетей в задаче классификации товаров. Сравнительный анализ архитектур нейросетей ResNet-34 и InceptionV3 для классификации товаров на изображениях, обученных и протестированных на RP2K dataset, показал преимущество архитектуры ResNet, которая достигла средней точности более 95%. Для повышения результатов классификации похожих товаров использован ансамбль двух сетей: ResNet-34 и API-Net, что заметно повысило точность классификации (от 6% до 20%). Применение техники аугментации не привело к повышению точности, использование *transfer learning* повысило качество классификации незначительно, менее чем на 0.5%.

Список литературы

1. RP2K: A Large-Scale Retail Product Dataset for Fine-Grained Image Classification [Электронный ресурс]. – URL: https://www.pinlandata.com/rp2k_dataset (дата обращения 16.12.2021).
2. Rethinking the Inception Architecture for Computer Vision [Электронный ресурс] / Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna – URL: <https://arxiv.org/pdf/1512.00567.pdf> (дата обращения 15.12.2021).

3. Deep residual learning for image recognition [Электронный ресурс]. – URL: https://rajatvikramsingh.github.io/media/DeepLearning_ImageNetWinners.pdf (дата обращения 15.12.2021).
4. Peiqin Zhuang. Learning. Attentive Pairwise Interaction for Fine-Grained Classification [Электронный ресурс] / Peiqin Zhuang, Yali Wang, Yu Qiao. – URL: <https://arxiv.org/pdf/2002.10191.pdf> (дата обращения 16.12.2021).
5. Leland McInnes. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction [Электронный ресурс] / Leland McInnes, John Healy, James Melville. – URL: <https://arxiv.org/pdf/1802.03426.pdf> (дата обращения 16.12.2021).
6. Басс Л. П. Сверточные нейронные сети с глубоким обучением в задачах обработки гиперспектральных спутниковых данных [Электронный ресурс] / Л. П. Басс, М. Г. Кузьмина, О. В. Николаева // Препринты ИПМ им. М.В. Келдыша. – 2018. – № 282. – 32 с. – URL: <http://library.keldysh.ru/preprint.asp?id=2018-282> (дата обращения 15.12.2021).
7. Стэнфордский курс: лекция 9. Архитектуры CNN [Электронный ресурс]. – URL: <https://www.reg.ru/blog/stehnfordskij-kurs-lekcija-9-arhitektury-cnn/> (дата обращения 15.12.2021).
8. Class-Balanced Loss Based on Effective Number of Samples. [Электронный ресурс] / Yin Cui, Menglin Jia et al. – URL: <https://arxiv.org/pdf/1901.05555.pdf> (дата обращения 16.12.2021).
9. Tan, C. A Survey on Deep Transfer Learning / C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, C. Liu // In Proc. 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4–7, 2018, Part III. – 2018. – DOI: 10.1007/978-3-030-01424-7_27.
10. Weiss K. A Survey of Transfer Learning / K. Weiss, T.M. Khoshgoftaar, D.Wang // Journal of Big Data. 2016. Vol. 3, №. 1. P. 1-9. DOI: 10.1186/s40537-016-0043-6

References

1. *RP2K: A Large-Scale Retail Product Dataset for Fine-Grained Image Classification* [Elektronnyj resurs]. URL: https://www.pinlandata.com/rp2k_dataset (data obrashcheniya 16.12.2021).
2. Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna. *Rethinking the Inception Architecture for Computer Vision* [Elektronnyj resurs]. URL: <https://arxiv.org/pdf/1512.00567.pdf> (data obrashcheniya 15.12.2021).
3. *Deep residual learning for image recognition* [Elektronnyj resurs]. URL: https://rajatvikramsingh.github.io/media/DeepLearning_ImageNetWinners.pdf (data obrashcheniya 15.12.2021).
4. Peiqin Zhuang, Yali Wang, Yu Qiao. *Learning. Attentive Pairwise Interaction for Fine-Grained Classification* [Elektronnyj resurs]. URL: <https://arxiv.org/pdf/2002.10191.pdf> (data obrashcheniya 16.12.2021).
5. Leland McInnes, John Healy, James Melville. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction* [Elektronnyj resurs]. URL: <https://arxiv.org/pdf/1802.03426.pdf> (data obrashcheniya 16.12.2021).
6. Bass L.P., Kuz'mina M.G., Nikolaeva O.V. Cvertochnye nejronnye seti s glubokim obucheniem v zadachah obrabotki giperspektral'nyh sputnikovyh dannyh. *Preprinty IPM im. M.V. Keldysha*. 2018. № 282. 32 s. [Elektronnyj resurs]. URL: <http://library.keldysh.ru/preprint.asp?id=2018-282> (data obrashcheniya 15.12.2021).
7. *Стэнфордский курс: лекция 9. Архитектуры CNN* [Elektronnyj resurs]. URL: <https://www.reg.ru/blog/stehnfordskij-kurs-lekcija-9-arhitektury-cnn/> (data obrashcheniya 15.12.2021).
8. Yin Cui, Menglin Jia et al. *Class-Balanced Loss Based on Effective Number of Samples* [Elektronnyj resurs]. URL: <https://arxiv.org/pdf/1901.05555.pdf> (data obrashcheniya 16.12.2021).
9. Tan, C. A Survey on Deep Transfer Learning / C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, C. Liu. In *Proc. 27th International Conference on Artificial Neural Networks*, Rhodes, Greece, October 4–7, 2018, Part III. 2018. DOI: 10.1007/978-3-030-01424-7_27.
10. Weiss K., Khoshgoftaar T.M., Wang D. A Survey of Transfer Learning. *Journal of Big Data*. 2016. Vol. 3, №. 1. P. 1-9. DOI: 10.1186/s40537-016-0043-6.

RESUME

T. V. Yermolenko, I. Ye. Samorodsky

Analysis of the Efficiency of Deep Neural Networks Architectures for the Classification of Products Images

In this article, an analysis of the effectiveness of various architectures of deep neural networks in the task of classifying goods was carried out.

The RP2K dataset was chosen as the dataset, as it contains a large number of product images that are very close to real ones.

Among the existing architectures of neural networks, 2 were considered: InceptionV3 (GoogLeNet) and ResNet-34, since they have proven themselves well in classification problems. In the classification of goods, ResNet-34 performed better, reaching an average accuracy of more than 95%. To improve the accuracy, the classes were balanced using the Class Balanced Loss algorithm.

To solve the problem of similarity of products belonging to the same category, in particular, for the categories "liquors" and "cosmetics", within which the classification accuracy indicators are noticeably lower and did not reach 90%, an ensemble of ResNet and API-Net neural networks was used in the work, which significantly increased classification accuracy (from 6% to 20%). The choice of API-Net for detailed classification is explained by the fact that the network simulates a person's ability to compare similar images by identifying contrasts.

The use of the augmentation technique did not lead to an increase in accuracy; the use of transfer learning improved the classification quality insignificantly, by less than 0.5%.

Implementation of the proposed models in systems for automatic categorization of goods will allow avoiding human participation in this process.

РЕЗЮМЕ

Т. В. Ермоленко, И. Е. Самородский

Анализ эффективности архитектур глубоких нейросетей для классификации изображений товаров

В данной статье был проведен анализ эффективности различных архитектур глубоких нейросетей в задаче классификации товаров по их изображениям.

В качестве датасета был выбран RP2K dataset, так как он содержит большое количество изображений товаров разных категорий, которые очень близки к реальным, а также метаданные для каждого изображения.

Среди существующих архитектур нейросетей были рассмотрены две: InceptionV3 (GoogLeNet) и ResNet-34, так как они хорошо себя зарекомендовали в задачах классификации. В классификации товаров лучше показала себя ResNet-34, достигая средней точности более 95%. Для улучшения точности проведена балансировка классов с помощью алгоритма Class Balanced Loss.

Для решения проблемы похожести товаров, принадлежащих одной категории, в частности, для категорий «ликеры» и «косметика», внутри которых показатели точности классификации заметно ниже и не достигли 90%, в работе использовался ансамбль нейросетей ResNet и API-Net, что заметно повысило точность классификации (от 6% до 20%). Выбор API-Net для детальной классификации объясняется тем, что сеть имитирует способность человека сравнивать похожие изображения, определяя контрасты.

Применение техники аугментации не привело к повышению точности, использование transfer learning повысило качество классификации незначительно, менее чем на 0.5%.

Реализация предложенных моделей в системах автоматической категоризации товаров позволит избежать участия человека в этом процессе.

Статья поступила в редакцию 14.12.2021.