

ПРИМЕНЕНИЕ СВЕРТОЧНЫХ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ КЛАССИФИКАЦИИ ТОВАРОВ ЦЕЛЕВОЙ ГРУППЫ ПО ВЫДЕЛЕННЫМ ПРИЗНАКАМ

Поляков Филипп Алексеевич¹, Поляков Алексей Павлович²

¹Аспирант;

Государственный университет «Дубна»;

Россия, 141980, Московская обл., г. Дубна, ул. Университетская, д. 19;

e-mail: as_evristic@mail.ru.

²Кандидат физико-математических наук, ведущий специалист;

Научно-исследовательский институт точных приборов;

Россия, 127490, Москва, ул. Декабристов, с. 51;

e-mail: alpoliakov@gmail.com.

В работе рассматривается один из этапов определения кода товарной номенклатуры внешнеэкономической деятельности для товаров, входящих в целевую группу «обувь», состоящий в анализе изображений товарных позиций, присутствующих в сопроводительных документах. Приведено обоснование применения сверточных нейронных сетей для классификации изображений. Рассмотрены возможные подходы к построению специализированных нейросетевых классификаторов. Проведен сравнительный анализ эффективности подходов, основанных на дообучении существующих классификаторов (transfer learning) и на построении сверточных сетей, обученных только на размеченных данных выбранного товарного ассортимента. Исследованы вопросы получения обучающей выборки путем парсинга специализированных сайтов и получения элементов выборки с помощью систем искусственного интеллекта, специализирующихся на генерации изображений по запросу.

Ключевые слова: классификатор, С-ИНС, классификационный код, искусственный интеллект, нейронные сети, преобразование матриц, стохастический градиентный спуск, визуальные признаки.

Для цитирования:

Поляков Ф. А., Поляков А. П. Применение сверточных искусственных нейронных сетей для классификации товаров целевой группы по выделенным признакам // Системный анализ в науке и образовании: сетевое научное издание. 2025. № 1. С. 1–7. EDN: ARKVMZ. URL: <https://sanse.ru/index.php/sanse/article/view/646>.

CLASSIFICATION OF COMMODITY ITEMS BY THEIR IMAGES USING CONVOLUTION NEURAL NETWORKS

Polyakov Philip A.¹, Polyakov Aleksey P.²

¹PhD student;

Dubna State University;

19 Universitetskaya Str., Dubna, Moscow region, 141980, Russia;

e-mail: as_evristic@mail.ru.

²PhD in Physical and Mathematical Sciences, leading specialist;

Research Institute of Precision Instruments;

51 Dekabristov Str., Moscow, 127490, Russia;

e-mail: alpoliakov@gmail.com.

The paper considers stage of determining the full HS code of the target group "shoes" by considering images of goods, attached to electronic documentation. Reasons of using convolutional neural networks (CNNs) for image classification is presented. Possible approaches to constructing specialized neural network



Статья находится в открытом доступе и распространяется в соответствии с лицензией Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная (CC BY 4.0) <https://creativecommons.org/licenses/by/4.0/deed.ru>

classifiers are examined. A comparative analysis is conducted on the effectiveness of methods based on fine-tuning pre-existing classifiers (transfer learning) versus building convolutional networks trained exclusively on labeled data from a selected product assortment. The challenges of acquiring training datasets through scraping specialized websites and generating dataset elements via artificial intelligence systems specialized in on-demand image generation are explored.

Keywords: classifier, CNN, classification code, artificial intelligence, neural networks, matrix transformation, stochastic gradient descent, visual signs.

Для цитирования:

Polyakov P. A., Polyakov A. P. Classification of commodity items by their images using convolution neural networks. *System analysis in science and education*, 2025;(1):1-7 (in Russ). EDN: ARKVMZ. Available from: <https://sanse.ru/index.php/sanse/article/view/646>.

Введение

В процессе присвоения товарам классификационного кода товарной номенклатуры внешнеэкономической деятельности (ТН ВЭД) анализируются относящиеся к товарной позиции данные различных типов, в том числе прилагаемые в электронных сопроводительных документах изображения товара. При наличии сомнений в правильности выбора классификационного кода ТН ВЭД таможенный инспектор может осуществить визуальный осмотр и сфотографировать объект. В работе исследуются вопросы автоматизации получения классификационных признаков, требующихся для корректного определения кода, путем применения методов искусственного интеллекта (ИИ) к цифровым изображениям товаров.

В области классификации изображений наилучшим образом себя зарекомендовали методы, основанные на применении сверточных искусственных нейронных сетей (С-ИНС или *CNN* от *Convolution neural networks*). С-ИНС различной архитектуры решают задачи классификации изображений в астрономии, медицине, радиоэлектронике и прочих областях науки и техники [1,4].

При выборе оптимальной архитектуры классифицирующей сверточной нейронной сети требуется сравнение эффективности таких вариантов выбора специализированного классификатора, как дообучение (*transfer learning*) предварительно обученных С-ИНС общего назначения, имеющих открытую архитектуру, или построение специализированного классификатора оригинальной сетевой архитектуры, обученного только на представителях классифицируемой целевой группы. При наличии в сопроводительных электронных документах нескольких вариантов изображений товара более высокий уровень точности обеспечивает полученная по байесовской схеме итоговая оценка признаков по результатам применения классифицирующей С-ИНС ко всей совокупности изображений.

Для качественного обучения ИНС, специализирующихся на классификации товаров целевой группы, требуются размеченные, т.е. предварительно классифицированные данные, достаточно полно представляющие варианты проявления квалификационных признаков на объектах. Для достижения высокого уровня точности классификации и во избежание переобучения следует обеспечить присутствие в обучающей выборке достаточно большого количества представителей каждого класса объектов, обладающих заданными признаками (порядка нескольких тысяч на класс), что представляет собой затратную по времени, но реализуемую задачу.

Ниже следует более подробное изложение алгоритмов решения предлагаемых задач.

1. Построение сверточных нейронных сетей, предназначенных для выявления квалификационных признаков в изображениях товаров целевой группы

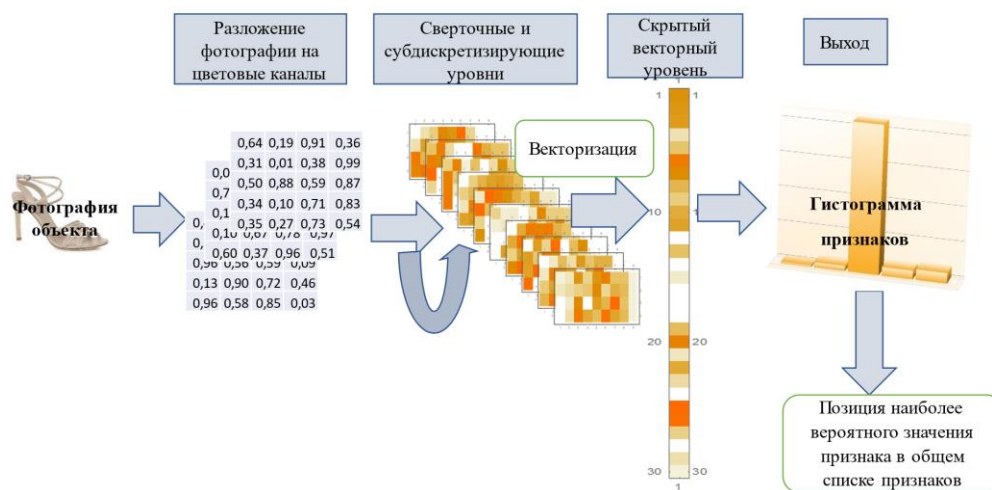
Сверточные нейронные сети широко применяются для классификации данных, имеющих матричную структуру, и являются наиболее распространенным типом ИНС, применяемых для решения задач распознавания объектов, которые могут быть представлены в виде графических образов. Понятие сверточных сетей ввел в научный оборот Ян Лекун [1]. Применение С-ИНС упростило анализ изображений методами искусственного интеллекта, так как анализ данных

изображения с помощью многослойного персептрона предполагал преобразование матриц, содержащих значения интенсивности цвета в пикселях изображения, во входной одномерный вектор, длина которого равна произведению размерностей матрицы по горизонтали и вертикали на количество цветовых каналов что вызывало проблемы с размещением данных в памяти. Посредством преобразований цифрового цветного изображения с помощью сверточных (*convolution*) и субдискретизирующих (*pooling*) уровней совокупная размерность исходных данных снижается на порядок. После операции векторизации – преобразования набора матриц, полученного на последнем сверточном уровне, в вектор – остаются закодированные в числовые значения признаки, совокупность которых с высокой степенью достоверности определяет объект, содержащийся в изображении [1].

Уровни С-ИНС, предшествующие выдаче результата, представляют собой, как правило, многослойный персептрон, состоящий из одной или нескольких пар линейных и активаторных уровней, на входе которого находится результат векторизации матриц предыдущего уровня, а на выходе – вектор, размерность которого равна количеству распознаваемых классов. Посредством активаторного уровня типа softmax этот вектор преобразуется к вектору, который может интерпретироваться как вектор вероятностей классов [4].

Применение именно С-ИНС для классификации изображений в выбранной товарной группе представляется естественным. Нормализация данных, содержащихся в изображении, предполагается в виде его преобразования к набору матриц заданной размерности, содержащих значения интенсивностей цветовых каналов в пикселях изображения в интервале от нуля до единицы. В случае если данные, предназначенные для обучения или классификации, имеют несовпадающие размерности, происходит их преобразование к заданной размерности, для чего используются алгоритмы, реализованные в среде *PyTorch* [2].

На рис. 1 приведена принципиальная схема работы сверточной нейронной сети.



5

Рис. 1. Принципиальная схема работы С-ИНС. Количество сверточных уровней может быть произвольным, скрытый векторный уровень может быть преобразован к гистограмме также за несколько шагов

В процессе построения синтетического алгоритма классификации по кодам ТНВЭД товаров целевой группы «Обувь» возникла необходимость построения классификаторов на базе С-ИНС по следующим визуально распознаваемым признакам (табл. 1) [5].

Табл. 1. Визуальные признаки товаров целевой группы, требующие выявления с помощью С-ИНС

Признак, обозначение	Подробное описание	Товарная позиция
ПЕРФОРАЦИЯ_СОЮЗКА	Обувь с союзкой из ремешков или имеющая одну или более перфораций	6402, 6403
РЕМЕШКИ_СОЮЗКА	Обувь с союзкой из ремешков или имеющая одну или более перфораций	6402, 6403
РЕМЕШКИ_ЗАКЛЕПКИ	Обувь с верхом из ремешков или полосок, прикрепленных к подошве заклепками	6402
РЕМЕШКИ_ПАЛЕЦ	Обувь с подошвой из натуральной кожи и верхом из ремешков из натуральной кожи, проходящих через подъем и охватывающих большой палец стопы	6403
ВОДА	Обувь водонепроницаемая	6401
ПЛАТФОРМА_ДЕРЕВО	С основанием или платформой из дерева, без внутренней стельки	6402, 6403
СПОРТ	Спортивная обувь (шипы)	6402, 6403, 6404
ДОМ	Домашняя обувь	6402, 6403, 6404, 6405
КАБЛУК	Высокий каблук, для обуви с союзкой из ремешков или имеющая одну или более перфораций	6402, 6403
Высота обуви: Н	Обувь, не закрывающая лодыжку	6402, 6403
Высота обуви: С	Обувь, закрывающая лодыжку, но не часть икры	6403
Высота обуви: В	Обувь, закрывающая лодыжку, но не закрывающая колено/ закрывающая лодыжку до колена и после	6401, 6403
Высота обуви: СВ	Обувь, закрывающая лодыжку, выше колена	6401
НЕ_ОБУВЬ	Некоторые виды: обувь без подошв, чехлы, бахилы, роликовые коньки и т.д.	

В соответствии с приведенной таблицей, обучающая выборка должна быть размечена по каждому из выявляемых признаков. В процессе ручной или автоматической разметки каждому изображению из обучающей выборки был сопоставлен тринадцатимерный вектор, каждая координата которого может быть равной нулю, если признак не присутствует, единице если признак присутствует и 0.5, если данная фотография непригодна для обучения по заданному признаку. Благодаря векторной разметке одни и те же фотографии могут участвовать в нескольких обучающих выборках. Обученные бинарные классификаторы применяются к фотографиям товара. В результате применения всех классификаторов к одной фотографии на выходе получается вектор размерности 13, состоящий из величин в интервале от нуля до единицы, а применение классификаторов к M фото даст матрицу размерности $M \times 13$.

Необходимость классификации по всем признакам, изложенным в таблице 1, предполагает проведение большой аналитической работы с исходными данными. Анализ проявления различных признаков показал, что существует определенная иерархия признаков, согласно которой все обучающие данные могут быть разбиты на несколько кластеров, в каждом из которых проявляются не 13, а от двух до пяти признаков. Это упрощает задачу разметки данных для пополнения обучающей выборки. Классификация изображения по визуальным признакам с использованием иерархии признаков не требует выявления каждого из 13 признаков, а к выбору с помощью С-ИНС номера кластера и последующему выявлению признаков, характерных для выбранного кластера. При маловероятном выборе С-ИНС неправильного кластера последующие некорректные результаты выявления несвойственных признаков подтвердят ошибочность первоначального выбора.

2. Выбор архитектуры сверточной нейронной сети, осуществляющей визуальную классификацию

Каждый из 13 бинарных классификаторов может иметь индивидуальную сетевую архитектуру, однако предпочтительнее выглядит выбор универсального варианта сетевой архитектуры классификатора. В этом случае для каждого признака будут сформированы свои наборы весовых матриц. Были рассмотрены различные варианты стандартной размерности, к которой должны быть приведены все изображения, в диапазоне от $64 \times 64 \times 1$ до $256 \times 256 \times 3$. В итоге для совместимости с СНС-классификаторами VGG-16/VGG-19 была выбрана размерность входных данных, равная $224 \times 224 \times 3$.

В качестве базовых рассматривались два варианта построения классификаторов:

- использование предварительно обученных С-ИНС общего назначения (*Transfer Learning*);
- построение и обучение сети с нуля [4].

Преимущества первого варианта состоят в том, что предварительно обученные С-ИНС общего назначения, например, VGG-16 и VGG-19, уже умеют выделять информативные признаки на огромных наборах данных (например, *ImageNet*), что позволяет рассчитывать на то, что и в узком классе изображений информативные признаки будут выделены.

Дообучение сети на собственной обучающей выборке состоит в замене последних слоев, начиная со скрытого векторного уровня (рис. 1) на слои, выдающие классификацию по запрашиваемым признакам и обучаемые только на данных целевой товарной группы. Практически такой подход означает, что вся обученная на больших массивах данных последовательность сверточных, активаторных и субдискретизирующих слоев сети является функцией, осуществляющей нормализацию индивидуальных данных, предназначенных для обучения, в скалярный вектор, который в дальнейшем может быть классифицирован многослойным персептроном. Обучение выделенного многослойного персептрона на обучающей выборке, состоящей из представителей целевой группы, следует провести, используя такие алгоритмы обучения, как стохастический градиентный спуск (*SGD*) или адаптивная оценка момента (*ADAM*). Такой подход обеспечивает экономию времени и вычислительных ресурсов за счет «замораживания» большей части весовых коэффициентов и обучения только в одном или двух линейных уровнях классификатора.

При применении предварительно обученной СНС VGG-19 последний линейный уровень с весовой матрицей 4096×1000 и вектором *bias* размерностью 1000 был заменен на уровень с весовой матрицей 4096×2 и двумерный вектор *bias*.

Обучение сети с нуля в целевой товарной группе также имеет свои преимущества, поэтому были проведены компьютерные эксперименты по построению классифицирующей С-ИНС с нуля. Поскольку речь идет о классификации специфических данных и количество изображений, накопленных в результате работы с клиентами достаточно велико, были использованы преимущества построения индивидуальной сети, настроенной для обработки данных своей товарной категории. Фотографии обуви, в отличие от снимков всевозможных объектов живой и неживой природы, обладают определенными специфическими качествами, которые могут быть выделены индивидуальной С-ИНС меньшей глубины, чем универсальные сверточные сети. Для достижения лучшего результата были проведены эксперименты с архитектурой сети и ее тонкой настройкой. По результатам тестирования различных вариантов архитектуры С-ИНС был сделан вывод о необходимости пополнения существующей обучающей выборки (порядка 3000 изображений товарных

позиций в целевой категории). Поскольку в процессе обучения появлялись признаки переобучения нейросети, рассматриваются возможности пополнения обучающей выборки способами, описанными ниже.

При построении нейросети, классифицирующей целевую группу по признакам, перечисленным в таблице 1, использовались входные данные размерности $224 \times 224 \times 3$, преобразованные в *PyTorch* по алгоритмам, повышающим яркость изображения [2].

Наилучшую среднюю точность по всем классифицируемым признакам при обучении С-ИНС в течение 13 часов удалось достигнуть при использовании в качестве ядра свертки двух сверточных уровней матрицы размерности 5×5 (таблица 2). Подобная архитектура близка к архитектуре сетей *LeNet-5*, предложенной Яном Лекуном, являющимся автором понятия сверточной нейронной сети. Тестировались варианты с ядрами свертки меньшей (3×3 и 4×4) и большей (до 7×7) размерности, но эти варианты показали худший.

Архитектура С-ИНС, выбранной в качестве оптимальной по соотношению времени обучения и точности, классифицирующей фотографии товаров, относящихся к товарной категории «Обувь», символически отображена на рис. 2.



Рис. 2. Схематическое отображение архитектуры специализированной С-ИНС, классифицирующей изображения представителей целевой товарной группы по различным бинарным признакам.

Архитектура используемой ИНС содержит 13 уровней (отдельными уровнями являются линейное преобразование данных с помощью весовых матриц и применение к результату преобразования векторной или матричной активаторной функции *ReLU* [4].

При проведении предварительных оценок точности были получены следующие результаты: средняя точность (accuracy) классификации с использованием предварительно обученных сетей общего назначения составила 93%, а при классификации с использованием индивидуальной сетевой архитектуры – 90.4%. Итак, промежуточный итог исследований в области выбора оптимального классификатора состоит в том, что определенное преимущество продемонстрировал первый способ нейросетевой классификации, основанный на дообучении нейросетей, обученных на больших датасетах, но и индивидуально обученная сеть лишь незначительно отстала по показателям точности. Большее время, затраченное на обучение построенной «с нуля» нейросети, не имеет существенного значения, так как процедура обучения осуществляется лишь при существенном изменении данных или структуры классифицирующих признаков. Учитывая консервативность ТН ВЭД, не следует ожидать больших изменений в рассматриваемой товарной категории.

3. Способы пополнения обучающей выборки

Несмотря на то, что в целом результаты классификации обеспечивают точность, необходимую для правильного функционирования синтетического алгоритма классификации выбранной товарной группы по кодам ТНВЭД [5], лучший результат может быть достигнут при пополнении обучающей выборки.

Для пополнения обучающей выборки, изначально основанной на каталогах продукции поставщиков, предполагается использование аугментации данных, алгоритмы которой реализованы в пакете *PyTorch*, которая без использования новых данных поможет увеличить объем данных и улучшить обобщающую способность модели [2].

Также получение дополнительных фотографий для пополнения обучающей выборки планируется следующими способами:

- обработка результатов специальных запросов в поисковых системах;
- парсинг сайтов интернет-магазинов;
- использование фотостоков, предпочтительно бесплатных.

Для парсинга сайтов предполагается использование имеющегося в *Python* инструментария *BeautifulSoup* (для анализа *HTML*), *requests* (для отправки HTTP-запросов) и *Scrapy* [3].

В процессе пополнения обучающей выборки часть данных может быть получена в результате использования ИИ-генераторов изображений, реализованных в таких моделях *Kandinski*, *DALL·E*, *MidJourney*, *Stable Diffusion* и проч. Результаты запросов к таким моделям, подвергнутые предварительной обработке, отсеивающей фантазийные и неточные варианты, показали свою пригодность для обучения нейросетевых классификаторов. Опасения ухода от реальности при обучении на результатах работы ИИ не оправдываются, так как уже в настоящее время дизайнеры обуви используют ИИ для разработки новых моделей. В определенной степени результаты генерации изображений методами ИИ позволяют быть уверенным в том, что при изменении моды показатели точности классификаторов существенно не изменятся.

Заключение

Проведенные исследования в области классификации изображений в рамках одной целевой товарной группы с целью выявления признаков, определяющих таможенные коды товарной номенклатуры показали, что применение таких методов ИИ, как С-ИНС, дает результат, точность которого обеспечивает уверенную классификацию, не уступающую результатам досмотра опытным таможенным инспектором.

При построении классифицирующей С-ИНС практически равноценными по качеству классификации являются С-ИНС, основанные на дообучении классификаторов образов общего назначения с открытым кодом, так и сверточные сети индивидуальной архитектуры, обученные исключительно на данных из определенной товарной группы.

Для повышения качества определения классифицирующих признаков с помощью С-ИНС необходимо пополнение обучающей выборки с помощью дополнительных ресурсов, парсинга интернет-сайтов и генерации изображений методами ИИ.

Публикуемые результаты получены в процессе исследований в рамках диссертационной работы, целью которой является разработка программного обеспечения, решающего актуальную задачу автоматизации процесса таможенной идентификации товаров определенных товарных групп [5].

Список источников

1. Лекун Ян. Как учится машина : революция в области нейронных сетей и глубокого обучения / Ян Лекун при участии Каролины Бризар ; перевод с французского Е. Арсеновой. – Москва : Альпина ПРО, 2021. – 334 с.
2. Стивенс Эли. PyTorch. Освещающая глубокое обучение / Эли Стивенс, Лука Антига, Томас Виман ; предисловие С. Чинталы. – Санкт-Петербург [и др.] : Питер, 2022. – 572 с
3. Рамальо Лучано. Python. К вершинам мастерства / Пер. с англ. Слинкин А. А. – Москва : ДМК Пресс, 2016. – 768 с.
4. Гудфеллоу Я. Глубокое обучение // Я. Гудфеллоу, И. Бенджио, А. Курвилль ; [пер. с англ. А. А. Слинкина]. - 2-е изд., испр. – Москва : ДМК Пресс, 2018. – 652 с.
5. Поляков Ф. А., Задорожный А. М. Систематизация признаков идентификации товаров для таможенных целей при построении интеллектуальных систем классификации // Системный анализ в науке и образовании: сетевое научное издание. – 2024. – № 2. – С. 30-38. –EDN: GNNNFJ. – URL: <https://sanse.ru/index.php/sanse/article/view/617>.