



Применение нейронных сетей сиамской архитектуры в задачах классификации продуктов различных категорий на прилавках универсама

Александр Владимирович Смирнов^{1✉}, Игорь Петрович Тищенко²

^{1,2}Институт программных систем им. А. К. Айламазяна РАН, Вельское, Россия

[✉]asmirnov_1991@mail.ru

Аннотация. В настоящей работе представлено исследование на тему применения нейронных сетей сиамской архитектуры в задачах классификации различных продуктов питания на прилавках универсальных магазинов. Сиамские сети – это особый класс нейросетевых архитектур, объединяющий в себе две свёрточные подсети. Его часто используют в задачах сопоставления объектов, поскольку по сравнению с традиционными свёрточными нейронными сетями он не требует большого количества обучающих данных. В ходе работ сгенерирован собственный набор данных, включающий пять различных категорий продуктов. В результате удалось достичь точности в 97.5% при обучении.

Ключевые слова и фразы: сиамские нейронные сети, набор данных, продукты питания

Для цитирования: Смирнов А. В., Тищенко И. П. *Применение нейронных сетей сиамской архитектуры в задачах классификации продуктов различных категорий на прилавках универсама* // Программные системы: теория и приложения. 2024. Т. 15. № 2(61). С. 113–137. https://psta.psiras.ru/read/psta2024_2_113-137.pdf

Введение

Среди подходов к распознаванию объектов на изображении можно выделить два основных направления: непосредственное распознавание объекта по его характерным чертам и особенностям, и сравнение объекта интереса с шаблоном заранее известного класса объектов. С непосредственным распознаванием объектов хорошо справляются классические свёрточные нейронные сети. Тем не менее, для достижения высоких показателей точности распознавания необходимо создать обширный набор данных для каждого из исследуемых классов, а также в зависимости от формы и внешнего вида объектов интереса требуется использовать сеть более сложной архитектуры.

С другой стороны, есть ряд задач, где в непосредственном определении класса/типа объекта нет необходимости. Например, если необходимо определить автомобиль на изображении или мотоцикл, то совершенно не важна марка конкретного автомобиля или мотоцикла. Для решения подобных задач часто используется нейронная сеть сиамской архитектуры.

Нейронные сети сиамской архитектуры были впервые представлены в начале 1990-х годов в работе [1] для решения проблемы проверки подлинности подписей как задачи сопоставления изображений. Авторы использовали набор данных, состоящий из образцов подписей 219 человек, часть из которых была специально подделана. В ходе работ авторы также провели экспериментальное исследование нескольких архитектур используемых подсетей и получили результат в 95.5% успешного определения подлинности подписей.

Сейчас нейронные сети сиамской архитектуры используются для анализа графических данных различных источников. Так в работе [2] авторы поставили цель изучить и объяснить концепцию сходства изображений с использованием сиамских сетей и технологии Grad-CAM¹. В результате им удалось получить область активации нейронной сети на изображении, а также составить ряд рекомендаций по созданию собственного набора данных для более точного обучения сети.

Работа [3] содержит экспериментальное исследование обучения сиамских нейронных сетей для классификации объектов с минимальным набором обучающих данных. Здесь авторы используют различные методы улучшения точности обучения среди которых можно выделить трансферное обучение и имитацию процесса ансамблевого обучения. В итоге им удалось

¹What is the Grad CAM method?^{URL}

добиться точности классификации в 69%–78% при использовании от 5 до 20 обучающих изображений на класс.

Статья [4] также посвящена классификации объектов с применением сиамских нейронных сетей. Однако, в данной работе авторы рассматривают в первую очередь иерархию классов и используют некоторые модификаторы, среди которых можно отметить матрицы затрат и убывающий иерархический множитель. В экспериментах участвовали следующие наборы данных: CIFAR-100², Amazon Product Reviews³ и Fashion MNIST⁴. В результате была получена точность классификации от 64% до 90% в зависимости от набора данных, что сопоставимо с точностью (65%–84%) полученной при стандартной классификации.

В статье [5] сиамские сети используются для распознавания паттернов человеческой деятельности. Авторы применяют такие технологии как тройная функция потерь и Bi-LSTM⁵ для распространения данных внутри сети в обоих направлениях. Таким образом, на наборе данных UCI HAR⁶ авторам удалось получить точность равную 93.9%.

В работе [6] рассматривается применение сиамских нейронных сетей для классификации изображений горных пород. Отличительной особенностью данной работы является разработанная авторами архитектурная надстройка пространственного преобразования (STN), которая используется для извлечения критически важных областей изображений. Благодаря этому, авторы достигли точности классификации в 87.14%–97.75%, что в среднем на 2% больше, чем при использовании классических глубинных нейронных сетей.

Иное применение нейронных сетей сиамской архитектуры представлено в статье [7]. Здесь целью авторов было определение принадлежности сгенерированного изображения какому-либо нейросетевому генератору. В частности, авторы использовали сгенерированные изображения лиц людей, для чего им понадобился набор данных, созданный при помощи 10 различных генеративных нейросетевых архитектур. Полученная в итоге точность составила более 90%.

Также сиамские сети используются для анализа сенсорных данных. Например, в работе [8] сеть используется для идентификации подключённой

²CIFAR-100^{[URL](#)}

³Amazon Product Reviews^{[URL](#)}

⁴Fashion MNIST^{[URL](#)}

⁵Bidirectional LSTM^{[URL](#)}

⁶UCI-HAR^{[URL](#)}

нагрузки. В данном подходе Сеть обучается на изображениях графиков зависимости потребляемого тока от времени различных бытовых приоров (холодильник, настольный ПК, дрель и т.д.). В качестве подсетей авторы используют Lenet-5⁷. Тестирование предложенного метода проводилось на наборах данных WHITED⁸ и PLAID⁹. В результате была получена точность идентификации более 98%.

Нейронные сети сиамской архитектуры уже достаточно широкого применяются для обработки медицинских данных. Любые данные, имеющие эквивалент в виде изображений (данные спектроскопии, снимки МРТ, эндоскопические изображения и т.д.) могут быть проанализированы искусственными нейронными сетями. К примеру, статья [9] посвящена классификации клинически значимых бактерий на основе Рамановская спектроскопии с применением сиамских нейронных сетей. Авторы данной статьи использовали набор спектральных данных Рамана, который содержит 5420 (6 видов бактерий, около 900 спектров на класс). Точность классификации оставила около 83.61%, что сопоставимо с точностью при полученной при использовании глубоких нейронных сетей – 84.13%.

В другой работе [10] анализируются небольшие образцы эндоскопических изображений для классификации лейкоплакии голосовых связок¹⁰ с использованием сиамской нейронной сети. Предложенный авторами метод показал лучшую точность классификации – 97.56%, в сравнении с нейронными сетями архитектур AlexNet¹¹, VGG Net¹², Google Inception¹³, ResNet¹⁴ и др.

Статья [11] затрагивает проблему оценки вариабельности контраста различных МРТ-сканеров. Разные МРТ-сканеры имеют различные параметры конфигурации из-за чего полученные снимки одной и той же части тела могут иметь расхождения, что усложняет диагностику и создаёт проблемы с точки зрения воспроизводимости снимков. Целью авторов данной работы является определить принадлежность снимка конкретному МРТ-сканеру. Для этого они используют сиамскую сеть с тройной

⁷Архитектура LeNet-5^{URL}

⁸WHITED^{URL}

⁹PLAID^{URL}

¹⁰Лейкоплакия гортани^{URL}

¹¹AlexNet — свёрточная нейронная сеть для классификации изображений^{URL}

¹²VGG Very Deep Convolutional Networks^{URL}

¹³Advanced Guide to Inception v3^{URL}

¹⁴ResNet (34, 50, 101): «остаточные» CNN для классификации изображений^{URL}

функцией потерь¹⁵ в купе с классическими методами классификации. В результате, авторам удалось добиться точности в 93.15%.

Некоторое распространение получило применение сиамских нейронных сетей для анализа снимков Дистанционного Зондирования Земли (ДЗЗ)¹⁶. Так в работе [12] применяется глубинная сиамская сеть с ручным извлечением признаков для классификации гиперспектральных изображений. Основной проблемой авторы называют ограниченную доступность меченых данных ДЗЗ. Архитектура сиамских нейронных сетей позволяет решить проблему недостаточного количества обучающих данных. Таким образом, экспериментальные результаты показали точностью 95.17% и 93.25% для наборов данных Pavia U¹⁷ и Indian Pines¹⁸.

Анализируя работы, связанные с применением сиамских нейронных сетей в задачах классификации и распознавания объектов, можно сделать вывод о том, что при использовании данного подхода можно добиться точности сопоставимой с той, что достигается при использовании классических свёрточных нейронных сетей, при этом используя сравнительно небольшие наборы данных.

В настоящей статье будет рассмотрен метод классификации продуктов различных категорий на прилавках универсамов с использованием сиамских нейронных сетей, при достигнутой точности обучения в 97.5%.

1. Постановка задачи. Актуальность

Настоящая статья посвящена исследованию и разработке метода применения нейронных сетей сиамской архитектуры в задачах распознавания различных продуктов питания на прилавках универсальных магазинов. Основная задача данного исследования заключается в разработке работоспособной концепции гибкой подсистемы распознавания различных продуктов/товаров, которая в последствии может быть применена в так называемых автоматизированных роботах-мерчандайзерах.

Исходя из определения профессии, мерчандайзер¹⁹ это — товаровед или помощник товароведа, человек, представляющий производственную или торговую компанию в торговых сетях. Отвечает за выкладку товара, установку сопутствующего необходимого оборудования, размещает

¹⁵Реализация функции потерь *Triplet Loss* в Python^{URL}

¹⁶Дистанционное зондирование Земли^{URL}

¹⁷Pavia University Hyperspectral Dataset^{URL}

¹⁸Indian Pines Hyperspectral Dataset^{URL}

¹⁹Мерчандайзер^{URL}

POS-материалы. Основная задача мерчандайзера — контроль наличия всего ассортимента компании на полках магазина и расположение его в наиболее благоприятных для покупки местах. Очевидно, что мерчандайзер выполняет важную роль в функционировании магазина и формировании прибыли, так как отсутствие на прилавках в нужный момент определённых товаров стоит продавцам миллиардов долларов в год.

Благодаря развитию робототехники и алгоритмов компьютерного зрения часть операций человека-мерчандайзера можно переложить на автоматизированного специального робота. Робот-мерчандайзер представляет собой платформу с вертикальной камерой, передвигающуюся между стеллажами и холодильниками, и выполняющую фото-видео съёмку. Камера робота обычно обладает гораздо более высоким разрешением, чем стационарная камера. С точки зрения экономики, это решение показало себя наиболее эффективным, потому что на магазин достаточно одного робота.

В данный момент применение роботов-мерчандайзеров не носит массовый характер, однако, их уже применяют в крупнейшей мире торговой сети Walmart²⁰ в США (рисунок 1), а также в сети Auchan²¹



Рисунок 1. Робот-мерчандайзер в сети гипермаркетов Walmart

в Португалии (рисунок 2).

²⁰Новый сотрудник Walmart — робот-сканер^[URL]

²¹Четыре новых способа контролировать товар на полке^[URL]



Рисунок 2. Робот-мерчандайзер в сети гипермаркетов Auchan

Предпосылки к применению подобных роботов также наблюдаются и в России в сети магазинов «Пятёрочка»²².

В связи с ростом популярности роботизированных систем в повседневной жизни, в частности, роботов-мерчандайзеров для крупных торговых сетей, появляется спрос на технологии распознавания и классификации товаров, что подтверждает актуальность исследования, представленного в настоящей работе.

2. Сиамские нейронные сети. Архитектура, обучение и тест

2.1. Архитектура используемой сети

Сиамские нейронные сети (Siamese Neural Network, SNN) – это класс архитектур нейронных сетей, предназначенных для сравнения и измерения сходства между парами входных выборок. Термин «сиамский» происходит от идеи, что архитектура сети (рисунок 3) состоит из парных нейронных сетей (часто свёрточных), которые идентичны по структуре и имеют одинаковый набор весовых коэффициентов. Каждая сеть обрабатывает одну входную выборку из пары, а их выходные данные сравниваются для определения сходства или различия между двумя экземплярами входных данных.

Сиамские сети предназначены для решения задач, где прямое обучение с помеченными выборками ограничено или затруднено, поскольку сеть после обучения способна различать похожие и непохожие экземпляры, не требуя явных меток классов.

²²НПО «Андроидная техника» создает роботов для розничных магазинов 

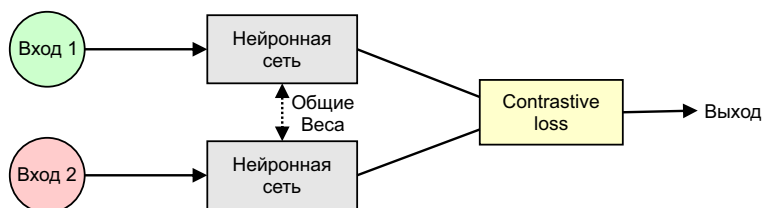


Рисунок 3. Пример архитектуры сиамской нейронной сети

Архитектура сиамской сети обычно состоит из трех основных компонентов: используемая подсеть, метрика сходства и функция контрастных потерь.

Подсеть является основным компонентом архитектуры сиамской сети. Она отвечает за извлечение ключевых признаков и особенностей из входных выборок. Обычно подсети представляют собой свёрточные нейронные сети, состоящие из слоёв свёртки и/или полносвязных слоёв, которые обрабатывают входные данные и создают некоторый дескриптор исследуемого объекта. Распределяя одинаковые веса между идентичными подсетями, модель учится извлекать схожие признаки и особенности для аналогичных входных данных, что позволяет эффективно их сравнивать.

Метрика сходства используется для сравнения сгенерированных дескрипторов и измерения сходства или различия между двумя входными данными. Выбор метрики сходства зависит от конкретной задачи и характера входных данных. Обычно в качестве метрики сходства используется евклидово расстояние, косинусное сходство или коэффициент корреляции.

Функция контрастных потерь — это функция, основанная на подсчёте расстояния, в отличие от более традиционных функций прогнозирования ошибки. Данная функция используется для анализа входных данных, при котором две схожие точки имеют малое евклидово расстояние, а две различные точки имеют большое евклидово расстояние.

Функция контрастных потерь высчитывается по формуле

$$(1 - Y) \frac{1}{2} (D_W)^2 + (Y) \frac{1}{2} \{\max(0, m - D_W)\}^2,$$

где D_W (евклидово расстояние) определяется по формуле:

$$\sqrt{\{G_W(X_1) - G_W(X_2)\}^2}.$$

Среди преимуществ сиамских нейронных сетей можно выделить следующее:

- ✓ Нет необходимости в большом наборе данных для обучения. Сеть способна обучиться на небольшом наборе данных, что также позволяет компенсировать дисбаланс классов.
- ✓ Устойчивость к аффинным преобразованиям изображений, таким как поворот и масштабирование.
- ✓ Семантическое сходство. Нейронная сеть сиамской архитектуры анализирует пространство признаков чтобы сформировать представление о схожести/различии изображений вместо того, чтобы просто извлекать статические признаки с помощью операции свёртки.

К недостаткам сиамских нейронных сетей можно отнести:

- Требуется больше времени на обучение по сравнению с традиционными свёрточными нейронными сетями.
- Не предоставляет данные о вероятности определения объекта к какому-либо классу.

Для выполнения поставленных задач по распознаванию и классификации продуктов различных категорий использовалась нейронная сеть сиамской архитектуры с двумя идентичными свёрточными подсетями, которые содержат три слоями свёртки, три полносвязных слоя и имеют следующую конфигурацию:

```
(cnn1): Sequential
Conv2d(3, 96, kernel_size=(11, 11), stride=(4, 4))
ReLU(inplace=True)
MaxPool2d(kernel_size=3, stride=2, padding=0, dilation=1, ceil_mode=False)
Conv2d(96, 256, kernel_size=(5, 5), stride=(1, 1))
ReLU(inplace=True)
MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
Conv2d(256, 384, kernel_size=(3, 3), stride=(1, 1))
ReLU(inplace=True)

(fc1): Sequential
Linear(in_features=384, out_features=1024, bias=True)
ReLU(inplace=True)
Linear(in_features=1024, out_features=256, bias=True)
ReLU(inplace=True)
Linear(in_features=256, out_features=2, bias=True)
```

2.2. Создание набора данных для обучения и тестирования

Для создания обучающего и тестового набора данных использовались фотографии прилавков и стеллажей (рисунок 4), находящихся

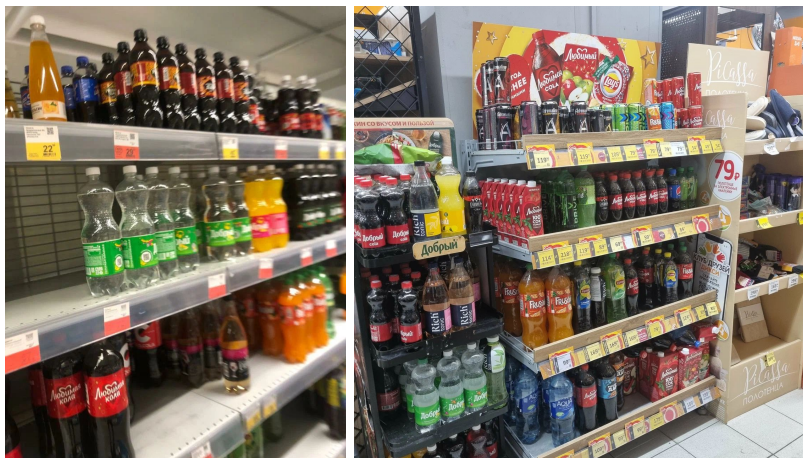


Рисунок 4. Пример фотографий, используемых для создания набора данных

в популярных сетевых универсамах/супермаркетах.

Поиск фотографий осуществлялся в интернете, что является относительно простым способом поиска различных изображений, однако, ракурс съёмки, разрешение и общее качество изображений может сильно варьироваться.

Исходя из характерных особенностей внешнего вида и относительной популярности некоторых продуктов, набор обучающих и тестовых данных содержал следующие классы объектов:

- Напитки типа «Кола» (cola) – напитки популярного бренда и всевозможные аналоги. Имеет вид пластиковой бутылки с тёмным, почти чёрным содержимым и этикеткой красного или оттенков красного цвета.
- Молочные и кисломолочные продукты (milk) – молочные и кисломолочные продукты различных брендов. Имеет вид пластиковой бутылки с белым или кремово-белым содержимым и светлой этикеткой с примесью различных цветов.

- Растительные масла (oil) – растительные масла различных брендов. Имеет вид пластиковой бутылки с золотисто-жёлтым содержимым и этикеткой, также содержащей оттенки жёлтого цвета.
- Колбасные изделия (sausage) – колбасные изделия различных брендов. Имеет продолговатую цилиндрическую форму с оболочкой розового, светло-розового или бежевого цвета.
- Минеральные воды (water) – питьевая и минеральная вода различных брендов. Имеет вид пластиковой бутылки с бесцветным содержимым. Цвет бутылок и этикеток может варьироваться от светло голубого до зелёного.

Выбор данных видов продуктов обусловлен тем, что они встречаются практически в любом современном универсаме или супермаркете и имеют достаточно много схожих особенностей внешнего вида, чтобы их можно было сгруппировать в отдельные классы, которые в свою очередь будут отличаться друг от друга. Группировка была применена в следствии невозможности адекватного распределения отдельных продуктов по классам из-за отсутствия оных в нужном количестве на найденных в интернете фотографиях, а также в следствии их визуального сходства.

Всего было использовано 370 фотографий прилавков/стеллажей с различными продуктами, из которых 102 фотографии содержали объекты представленных ранее классов. Образцы объектов для классов извлекались из найденных изображений методом вырезания соответствующей области с последующем сохранением в виде отдельных изображений. Таким образом количество образцов для каждого из классов составило:

- класс «cola»: 224
- класс «milk»: 240
- класс «oil»: 126
- класс «sausage»: 225
- класс «water»: 227

Несмотря на то, что сиамские сети способны адекватно обучаться при дисбалансе классов, было принято решение о балансировании классов, что потенциально может увеличить точность обучения сети. Процедура балансирования заключалась в добавлении недостающих экземпляров класса, которыми являлись зеркальные копии случайных уже существующих экземпляров. Итого в каждом классе количество экземпляров было выровнено до 252, из которых 220 на обучение и 32

на тест. На рисунке 5 показаны примеры выборок по классам полученного набора данных.



Рисунок 5. Пример экземпляров объектов созданного набора данных

Размер экземпляров варьировался в зависимости от габаритов искомого объекта на исходном изображении. Однако, этот факт не повлиял на процесс обучения сиамской нейронной сети, так как перед подачей экземпляров на вход, они приводились к единому размеру – 100x100 пикселей. Тем не менее, габариты объектов интереса были учтены непосредственно при их распознавании.

2.3. Обучение нейронной сети сиамской архитектуры

Программирование архитектуры используемой сиамской нейронной сети, её обучение и тестирование происходило с использованием язы-

ка программирования Python и фреймворка PyTorch²³. PyTorch – это фреймворк предназначенный для машинного обучения. Он включает в себя набор инструментов для работы с моделями, используется в обработке естественного языка, компьютерном зрении и других похожих направлениях.

Обучение используемой сиамской нейронной сети происходило со следующими параметрами:

- `batch_size`: 16 – размер (количество) данных, посылаемых на вход сети каждую эпоху;
- `epochs`: 80 – количество эпох обучения нейронной сети;
- функция потерь: `ContrastiveLoss` – функция контрастных потерь;
- оптимизатор: `Adam` – один из методов оптимизации обучения включённых в фреймворк.

Для обучения был использован 8-ми ядерный CPU AMD Ryzen 7 3700X с частотой 4 ГГц на ядро. На рисунке 6 приведён график изменения

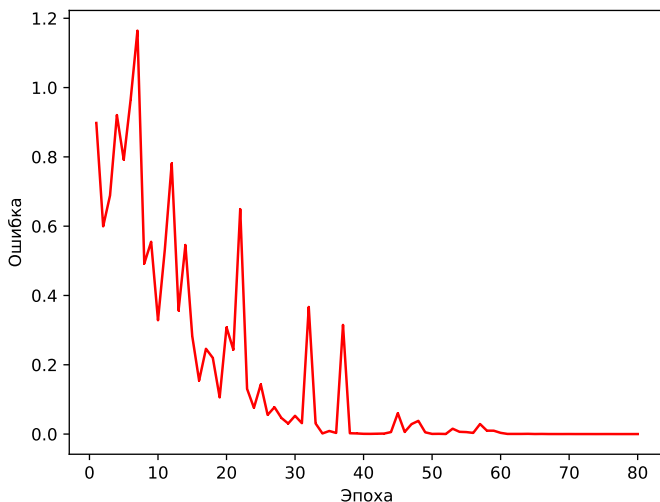


Рисунок 6. График изменения ошибки обучения

ошибки обучения в зависимости от эпох.

²³[PyTorch^{URL}](https://pytorch.org/)

В конце обучения ошибка была близка к 0 и составила 0.0001069. На рисунке 7 показан результат запуска обученной сиамской нейронной сети на тестовой выборке.

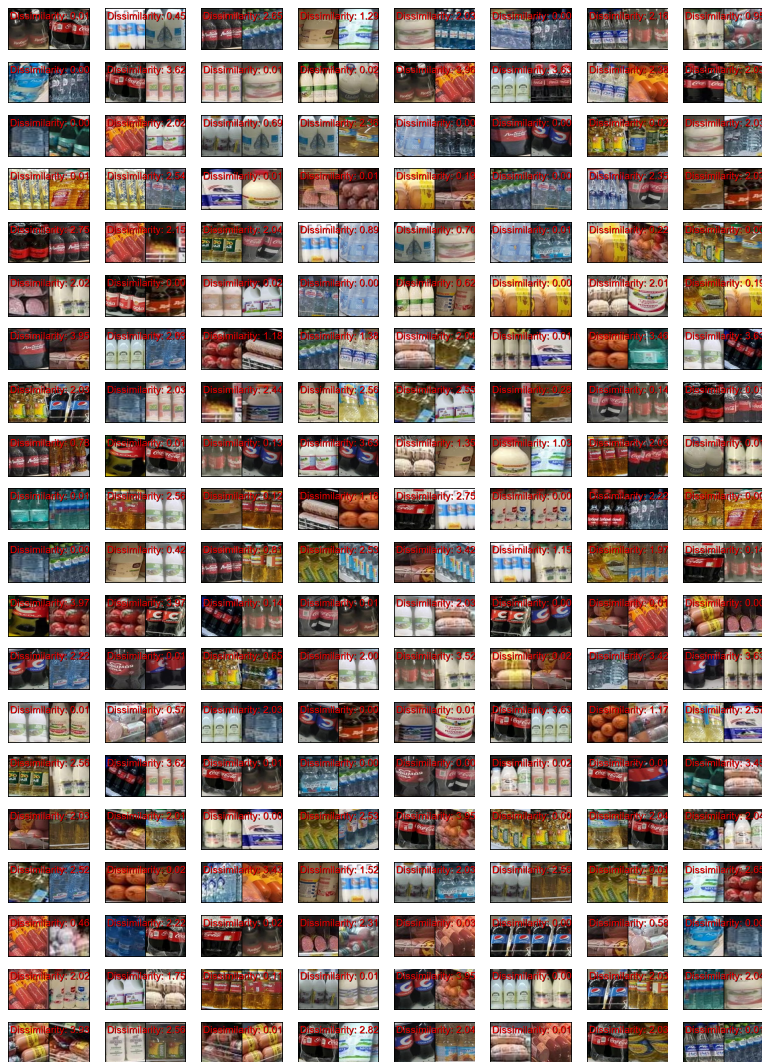


Рисунок 7. Пары входных изображений с рассчитанным показателем «непохожести»

В сеть подавались пары изображений из обучающей выборки. На выходе рассчитывался показатель «непохожести» (dissimilarity), отражающий евклидово расстояние между дескрипторами входных изображений. Таким образом, чем меньше значение данного показателя, тем более схожи объекты на входных изображениях.

2.4. Подсчёт точности обучения

В качестве метрики подсчёта точности обучения сиамской нейронной сети была использован показатель F-score²⁴, который высчитывается по формуле:

$$\begin{aligned}
 (1) \quad Fscore &= \frac{2 * Recall * Precision}{Recall + Precision} \\
 (2) \quad Precision &= \frac{True \ Positive}{True \ Positive + False \ Positive} \\
 (3) \quad Recall &= \frac{True \ Positive}{True \ Positive + False \ Negative} \\
 (4)
 \end{aligned}$$

Здесь следует ввести следующие обозначения:

True Positive(TP) – истинно-положительное решение. Искомый объект обнаружен.

True Negative(TN) – истинно-отрицательное решение. Объект, который не является искомым не был обнаружен.

False Positive(FP) – ложно-положительное решение. Объект, который не является искомым был детектирован как искомый.

False Negative(FN) – ложно-отрицательное решение. Объект, который является искомым не был обнаружен.

Precision (точность) – отношение TP к TP + FP. Это доля объектов, названными классификатором положительными и при этом действительно являющимися положительными.

Recall (полнота) – отношение TP к TP + FN. Это то, какую долю объектов положительного класса из всех объектов положительного класса нашёл алгоритм.

Поскольку сиамская нейронная сеть на выходе даёт информацию об относительном расстоянии объектов друг от друга, что является в некотором понимании мерой схожести (меньше расстояние – более

²⁴Что такое F-score и для чего он используется?^[9]

похожие объекты), то для определения принадлежности объектов к одному классу был введено пороговое значение расстояния. Если расстояние между объектами ниже данного порога, то они считаются принадлежащими одному классу, в противном случае разным классам. В таблице 1 перечислены пороговые значения и полученная точность обучения, рассчитанная на тестовой выборке.

Таблица 1. Точность обучения, полученная на тестовой выборке с различными пороговыми значениями расстояния

Пороговое значение	Точность по F-score
0.5	0.961
0.8	0.975
1.0	0.975
1.5	0.969

3. Экспериментальное тестирование обученной сиамской нейронной сети

3.1. Подготовка эталонных данных

Тестирование обученной сиамской нейронной сети проводилось на фотографиях прилавков и стендов популярных сетей универсамов и супермаркетов, которые содержали продукты исследуемых классов, то есть тех, что были представлены в созданном ранее наборе данных.

Помимо тестового набора изображений, также были подготовлены так называемые Ground Truth²⁵ данные, то есть эталонные данные необходимые для подсчёта результирующей точности классификации. Ground Truth данные представляют собой список с координатами габаритных прямоугольников искомых объектов с указанием принадлежности к определённому классу и названием тестового изображения, на котором находится искомый объект. Каждая строка этого списка содержит:

label_name – название класса;

bbox_x – координата X левого верхнего угла габаритного прямоугольника объекта;

bbox_y – координата Y левого верхнего угла габаритного прямоугольника объекта;

²⁵Ground Truth[®]

bbox_width – ширина габаритного прямоугольника объекта;
bbox_height – высота габаритного прямоугольника объекта;
image_name – название содержащего объект изображения;
image_width – ширина содержащего объект изображения;
image_height – высота содержащего объект изображения.

Разметка изображений для получения Ground Truth данных происходила с использованием онлайн сервиса Make Sense²⁶. На рисунке 8 показан



Рисунок 8. Пример размеченных изображений. Зелёные рамки – габаритные прямоугольники Ground Truth для искомых объектов

пример размеченных изображений.

При тестировании использовался подход на основе сканирующего окна²⁷. Размер сканирующего окна определялся для каждого класса отдельно. Высотой и шириной окна являлись значения среднего геометрического по высоте и ширине экземпляров класса. Шаг окна по горизонтали

²⁶Make Sence^{URL}

²⁷Принцип сканирующего окна^{URL}

и вертикали составлял $1/3$ от его ширины и высоты соответственно.

В процессе тестирования, область исходного изображения под сканирующим окном подавалась на вход обученной сиамской нейронной сети в паре с эталонным экземпляром класса, выбранным случайно из тестовой выборки. Затем сеть рассчитывала показатель «непохожести» (dissimilarity) для пары входных данных, который показывал на сколько образец схож с эталоном. Если показатель dissimilarity был ниже некоего порогового значения (подбиралось экспериментально), то образец считался экземпляром того же класса, что и эталон, и его координаты на исходном изображении, а также его ширина и высота сохранялись в отдельный список для последующего визуального отображения и подсчёта точности распознавания/классификации. На рисунке 9 показан результат работы сиамской нейронной сети.



Рисунок 9. Исходные изображения после классификации продуктов: зелёные прямоугольники – объекты Ground Truth; жёлтые рамки – объекты, определённые сиамской нейронной сетью

3.2. Подсчёт точности классификации продуктов сиамской нейронной сетью

Точность классификации продуктов рассчитывалась с использованием метрики F-score, что была описана в п. 2.4 настоящей статьи. В данном случае для определения принадлежности полученных габаритных прямоугольников к эталонным данным (Ground Truth) было использовано значение IoU. Intersection over Union (IoU, пересечение по объединению), также известное как коэффициент Жаккара²⁸ – это число от 0 до 1, показывающее, насколько у двух объектов (эталонного и текущего) совпадает внутренний «объём» (рисунок 10).

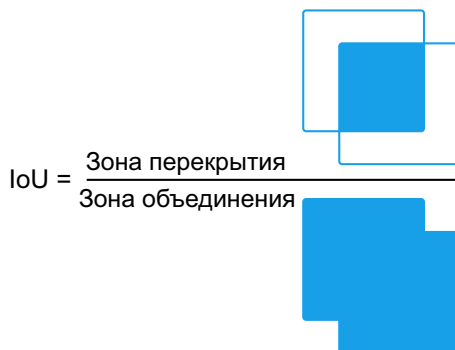


Рисунок 10. Демонстрация расчёта значения IoU

Таким образом, объект считался обнаруженным если габаритный прямоугольник его эталонных данных и габаритный прямоугольник, полученный от нейронной сети, имели значение пересечения по объединению (IoU) не менее 0.5. В виду особенностей метода сканирующего окна, одному эталонному объекту (габаритному прямоугольнику) могло соответствовать несколько габаритных прямоугольников, сгенерированных сиамской нейронной сетью. Все габаритные прямоугольники сгенерированные сетью и имеющие значение $\text{IoU} \geq 0.5$ с эталонным, считались за успешное детектирование искомого объекта.

В таблице 2 представлены данные о точности классификации продуктов сиамской нейронной сетью, обученной на наборе данных из п. 2.2 настоящей статьи.

²⁸[Jaccard index URL](#)

ТАБЛИЦА 2. Средняя точность обученной сиамской сети по классам, рассчитанная по метрике F-score в результате тестирования, а также показатели Precision (точность) и Recall (полнота)

Класс	F-score	Precision	Recall	Кол-во эталонных объектов	Кол-во найденных объектов
cola	0.401	0.298	0.890	265	230
milk	0.425	0.306	0.836	255	220
oil	0.577	0.477	0.787	138	120
sausage	0.277	0.177	0.786	244	200
water	0.413	0.304	0.842	298	241

3.3. Анализ результатов экспериментального тестирования

Средняя точность классификации различных категорий продуктов с применением сиамской нейронной сети по метрике F-score составила 0.419 (41.9%). Минимальная точность была зафиксирована на классе «sausage» и составила 0.277 (27.7%), максимальная на классе «oil» – 0.577 (57.7%). Следует отметить показатель Precision, среднее значение которого – 0.312, что свидетельствует о большом количестве ложных срабатываний, причиной которых могло стать присутствие объектов, которые визуально были схожи с искомыми. Например, на исходных изображения для класса «cola» были зафиксированы объекты (рисунок 11) визуально похожие на объекты класса, но по факту являющиеся другими напитками.

Однако показатель Recall находится на достаточно высоком уровне и в среднем составляет 0.828, что в свою очередь говорит о высоком проценте детектирования искомых объектов на исходном изображении.

Также стоит перечислить ряд факторов, которые повлияли на точность классификации. К ним можно отнести следующее:

- Низкое качество используемых изображений.
- Различное разрешение используемых изображений.
- Ракурс съёмки и дистанция до искомых объектов на используемых изображениях.

Точно определить влияние каждого из перечисленных факторов достаточно затруднительно. Тем не менее, стоит учесть тот факт, что



Рисунок 11. Ложные срабатывания при выполнении классификации: 1) напитки типа «Байкал»; 2) напитки типа «Квас»

из-за специфики использования роботов-мерчендайзеров, изображения, полученные с их камер, будут лишены практически всех перечисленных недостатков. Это объясняется тем, что робот передвигается по определённой траектории и выполняет съёмку всех продуктов с одного ракурса и дистанции, используя одну или несколько камер со схожим качеством и разрешением съёмки.

Вывод

Исходя из всего вышеперечисленного, можно сделать вывод о том, что подход к классификации продуктов основанный на использовании сиамских нейронных сетей имеет по крайней мере теоретическую работоспособность, о чём свидетельствует точность по F-score полученная при обучении и равная 97.5%. Основными факторами низкой точности, полученной при экспериментальном тестировании, является разнородность и низкое качество используемых тестовых данных. При исключении данных факторов, подход, описанный в настоящей статье, может быть

использован в качестве составной части системы компьютерного зрения робота-мерчандайзера.

Список использованных источников

- [1] Bromley J., Guyon I., LeCun Y., Säckinger E., Shah R. *Signature verification using a "Siamese" time delay neural network* // *Proceedings of the 6th International Conference on Neural Information Processing Systems, NIPS'93*, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc..– 1993.– Pp. 737–744.  [↑114](#)
- [2] Livieris I. E., Pintelas E., Kiriakidou N., Pintelas P. *Explainable image similarity: integrating siamese networks and grad-CAM* // *Journal of Imaging*.– 2023.– Vol. **9**.– No. 10.– Pp. 224.  [↑114](#)
- [3] Valero-Mas J. J., Gallego A. J., Rico-Juan J. R. *An overview of ensemble and feature learning in few-shot image classification using siamese networks* // *Multimedia Tools and Applications*.– 2024.– Vol. **83**.– No. 7.– Pp. 19929–19952.  [↑114](#)
- [4] Понамарёв В. В., Китов В. В., Китов В. А. *Учёт иерархии классов при классификации объектов с помощью сиамских нейронных сетей* // *Прикладная математика и информатика*.– 2023.– № 73.– С. 38–57.  [↑115](#)
- [5] Byung-Rae C., Binod V. *Enhancing human activity recognition with siamese networks: a comparative study of contrastive and triplet learning approaches* // *Electronics*.– 2024.– Vol. **13**.– No. 9.– Pp. 1739.  [↑115](#)
- [6] Qiqi Z., Sai W., Shun T., Liangbin Y., Kunlun Q., Qingfeng G. *RockS² Net: Rock image classification via a spatial localization siamese network* // *Computers & Geosciences*.– March 2024.– Vol. **185**.– id. 105560.  [↑115](#)
- [7] Abady L., Wang J., Tondi B., Barni M. *A siamese-based verification system for open-set architecture attribution of synthetic images* // *Pattern Recognition Letters*.– April 2024.– Vol. **180**.– Pp. 75–81.  [↑115](#)
- [8] Lingxia L., Ju-Song K., Fanju M., Miao Y. *Non-intrusive load identification based on retrainable Siamese network* // *Sensors*.– 2024.– Vol. **24**.– No. 8.– Pp. 2562.  [↑115](#)
- [9] Contreras J., Mostafapour S., Popp J., Bocklitz T. *Siamese networks for clinically relevant bacteria classification based on Raman spectroscopy* // *Molecules*.– 2024.– Vol. **29**.– No. 5.– Pp. 1061.  [↑116](#)
- [10] Zhenzhen Y., Botao H., Zhenghao S., Minghua Z., Shuangli D., Haiqin L., Xinhong H., Xiaoyong R., Yan Y. *Vocal cord leukoplakia classification using Siamese network under small samples of white light endoscopy images* // *Otolaryngology–Head and Neck Surgery*.– 2024.– Vol. **170**.– No. 4.– Pp. 1099–1108.  [↑116](#)
- [11] Polsinelli M., Li H. B., Mignosi F., Zhang L., Placidi G. *Siamese network to assess scanner-related contrast variability in MRI* // *Image and Vision Computing*.– May 2024.– Vol. **145**.– id. 104997.  [↑116](#)

- [12] Ranjan P., Girdhar A. *Deep Siamese network with handcrafted feature extraction for hyperspectral image classification* // Multimedia Tools and Applications.– 2024.– Vol. **83**.– No. 1.– Pp. 2501–2526.  [↑117](#)

Поступила в редакцию 22.03.2024;
одобрена после рецензирования 10.06.2024;
принята к публикации 10.06.2024;
опубликована онлайн 22.06.2024.

Рекомендовал к публикации

д.ф.-м.н. А. М. Елизаров

Информация об авторах:



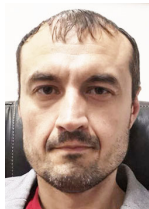
Александр Владимирович Смирнов

Младший научный сотрудник Лаборатории методов обработки и анализа изображений, Институт Программных Систем имени А. К. Айламазяна РАН. Научные интересы: компьютерное зрение; нейронные сети; робототехника; автоматизация и управление



0000-0002-7104-1462

e-mail: asmirnov_1991@mail.ru



Игорь Петрович Тищенко

Кандидат технических наук, зав. Лабораторией методов обработки и анализа изображений, Институт Программных Систем имени А. К. Айламазяна РАН. Научные интересы: компьютерное зрение; нейронные сети; робототехника; автоматизация и управление



0000-0002-0369-0524

e-mail: igor.p.tishchenko@gmail.com

Вклад авторов: *А. В. Смирнов* – 95% (идея, методология, программное обеспечение, валидация, формальный анализ, расследование, сбор материала, курирование данных, написание черновой версии, доработка и редактирование); *И. П. Тищенко* – 5% (наставничество, администрирование).

Декларация об отсутствии личной заинтересованности: *благополучие авторов не зависит от результатов исследования.*



Application of neural networks of Siamese architecture in problems of classifying products of various categories on supermarket shelves

Alexander Vladimirovich **Smirnov**¹, Igor Petrovich **Tishchenko**²

^{1,2} Ailamazyan Program Systems Institute of RAS, Ves'kovo, Russia

¹ asmirnov_1991@mail.ru

Abstract. This paper presents a study on the application of Siamese architecture neural networks in problems of classifying various food products on the shelves of department stores. Siamese networks are a special class of neural network architectures that combine two convolutional subnets. This type of neural networks is often used in object matching problems and has an important advantage over traditional convolutional neural networks, namely the absence of the need for a large amount of training data. During the work, we generated our own data set, including five different product categories. As a result, it was possible to achieve a tonality of 97.5% during training. (*In Russian*).

Key words and phrases: Siamese neural networks, dataset, foodstuffs

2020 Mathematics Subject Classification: 68T10; 68T45

For citation: Alexander V. Smirnov, Igor P. Tishchenko. *Application of neural networks of Siamese architecture in problems of classifying products of various categories on supermarket shelves*. Program Systems: Theory and Applications, 2024, **15**:2(61), pp. 113–137. (*In Russ.*).
https://psta.psiras.ru/read/psta2024_2_113-137.pdf

References

- [1] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, R. Shah. “Signature verification using a "Siamese" time delay neural network”, *Proceedings of the 6th International Conference on Neural Information Processing Systems*, NIPS’93, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993, pp. 737–744. 
- [2] I. E. Livieris, E. Pintelas, N. Kiriakidou, P. Pintelas. “Explainable image similarity: integrating siamese networks and grad-CAM”, *Journal of Imaging*, **9**:10 (2023), pp. 224. 
- [3] J. J. Valero-Mas, A. J. Gallego, J. R. Rico-Juan. “An overview of ensemble and feature learning in few-shot image classification using siamese networks”, *Multimedia Tools and Applications*, **83**:7 (2024), pp. 19929–19952. 
- [4] V. V. Ponamaryov, V. V. Kitov, V. A. Kitov. “Accounting for class hierarchy in object classification using Siamese neural networks”, *Computational Mathematics and Modeling*, 2024. 
- [5] C. Byung-Rae, V. Binod. “Enhancing human activity recognition with siamese networks: a comparative study of contrastive and triplet learning approaches”, *Electronics*, **13**:9 (2024), pp. 1739. 
- [6] Z. Qiqi, W. Sai, T. Shun, Y. Liangbin, Q. Kunlun, G. Qingfeng. “RockS²Net: Rock image classification via a spatial localization siamese network”, *Computers & Geosciences*, **185** (March 2024), id. 105560. 
- [7] L. Abady, J. Wang, B. Tondi, M. Barni. “A siamese-based verification system for open-set architecture attribution of synthetic images”, *Pattern Recognition Letters*, **180** (April 2024), pp. 75–81. 
- [8] L. Lingxia, K. Ju-Song, M. Fanju, Y. Miao. “Non-intrusive load identification based on retrainable Siamese network”, *Sensors*, **24**:8 (2024), pp. 2562. 
- [9] J. Contreras, S. Mostafapour, J. Popp, T. Bocklitz. “Siamese networks for clinically relevant bacteria classification based on Raman spectroscopy”, *Molecules*, **29**:5 (2024), pp. 1061. 
- [10] Y. Zhenzhen, H. Botao, S. Zhenghao, Z. Minghua, D. Shuangli, L. Haiqin, H. Xinhong, R. Xiaoyong, Y. Yan. “Vocal cord leukoplakia classification using Siamese network under small samples of white light endoscopy images”, *Otolaryngology–Head and Neck Surgery*, **170**:4 (2024), pp. 1099–1108. 
- [11] M. Polsinelli, H. B. Li, F. Mignosi, L. Zhang, G. Placidi. “Siamese network to assess scanner-related contrast variability in MRF”, *Image and Vision Computing*, **145** (May 2024), id. 104997. 
- [12] P. Ranjan, A. Girdhar. “Deep Siamese network with handcrafted feature extraction for hyperspectral image classification”, *Multimedia Tools and Applications*, **83**:1 (2024), pp. 2501–2526. 