

ИНФОРМАЦИОННЫЕ
ТЕХНОЛОГИИ
И ТЕЛЕКОММУНИКАЦИИ

INFORMATION
TECHNOLOGIES
AND TELECOMMUNICATIONS

УДК 004.8

DOI: 10.17212/2782-2001-2025-4-39-52

Исследование методических подходов реализации визуального поиска товаров в интернет-маркетплейсах*

Е.Н. АНТОНЯНЦ^а, А.А. АЛЕТДИНОВА^б, Д.Н. ШИПУНОВ^с,
А.А. СЕДЕЛЬНИКОВ^д, С.В. СИМОНЯК^е

РФ, 630073, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный
технический университет

^а bax201438@gmail.com ^б aletdinova@corp.nstu.ru ^с shipunov_dmitriy@inbox.ru
^д prodaga05@mail.ru ^е SemenSimon21@gmail.com

В статье представлено исследование методов визуального поиска товаров и результаты их программной реализации в системах электронной коммерции, выполнен обзор классических методов машинного обучения и современных нейросетевых архитектур. Для сравнения были выбраны и реализованы следующие подходы: гистограмма направленных градиентов/машина опорных векторов (Histogram of Oriented Gradients, HOG/Support Vector Machine, SVM), локальные бинарные шаблоны/машина опорных векторов (Local Binary Patterns, LBP/SVM), матрица совместной встречаемости уровней серого/случайный лес (Gray Level Co-occurrence Matrix, GLCM/Random Forest) и модель сиамской нейронной сети (Siamese Neural Networks, SNN). Разработанное в рамках исследования программное обеспечение включает методы предобработки изображений, извлечения признаков, сравнения векторных представлений с оптимальным порогом классификации 0,55. Экспериментальные исследования проводились на целевом наборе данных товарных изображений с использованием оценки метрик точности. Наилучшие показатели оказались у сиамской нейронной сети, результаты обучения которой показали следующие значения метрик: точности (accuracy) 0,951, точности (precision) 0,937, полноты 0,967, F1-меры 0,952 и площади под кривой ROC (Area Under the Receiver Operating Characteristic Curve, AUC-ROC) 0,9799. В ходе исследования была отмечена способность SNN-модели демонстрировать высокую устойчивость к изменениям освещения, ракурса съемки и сложного фона, сохраняя при этом стабильную точность идентификации товаров. Полученные результаты демонстрируют практическую значимость предложенного подхода для создания систем визуального поиска товаров в мобильных приложениях интернет-маркетплейсов, позволяющих улучшить пользовательский опыт и повысить уровень продаж. Исследование подтвердило эффективность нейросетевых методов для задач анализа и сопоставления изображений в условиях реального сегмента розничной торговли.

Ключевые слова: визуальный поиск товаров, электронная коммерция, сиамские нейронные сети, гистограмма направленных градиентов, машина опорных векторов, локальные бинарные шаблоны, матрица совместной встречаемости уровней серого, случайный лес

* Статья получена 14 октября 2025 г.

ВВЕДЕНИЕ

В современной сфере мобильной электронной коммерции всё большее значение приобретают интеллектуальные инструменты, повышающие удобство взаимодействия пользователя с приложением. Одним из перспективных направлений является визуальный поиск товаров – технология, позволяющая находить нужные позиции по фотографии вместо текстового запроса. В настоящее время существует несколько подходов к реализации подобных систем, каждый из которых имеет свои сильные и слабые стороны. Среди наиболее распространенных методов можно выделить классические алгоритмы компьютерного зрения, такие как HOG/SVM, LBP/SVM, GLCM/Random Forest, а также современные нейросетевые архитектуры, в частности сверточные (Convolutional Neural Networks, CNN) и сиамские нейронные сети [1, 2].

Классические методы, основанные на извлечении признаков, демонстрируют ограниченную точность и не требуют значительных вычислительных ресурсов. Однако их эффективность существенно снижается в реальных условиях: при изменении освещения, ракурса съемки или наличии сложного фона точность идентификации товаров резко падает. В то же время нейросетевые подходы, особенно основанные на сиамской архитектуре, способны автоматически извлекать высокоуровневые семантические признаки и обучаться на неоднородных изображениях, что делает их более адаптивными к реальным условиям электронной коммерции [3, 4].

Сиамская нейронная сеть, впервые предложенная для решения задач проверки подлинности и сравнения, в последние годы активно применяется в системах визуального поиска благодаря своей способности генерировать компактные и информативные векторные представления изображений. При этом обученная модель SNN может сопоставлять даже сильно отличающиеся снимки одного и того же товара, сохраняя значительную точность распознавания. Важное практическое значение представляет возможность ее оптимизации для мобильных платформ за счет уменьшения размера модели и использования ускоренных библиотек для вычислительного процесса (например, TensorFlow или Core ML), достигая высокой скорости работы даже на устройствах среднего класса [5].

Несмотря на существенные преимущества, внедрение нейросетевого визуального поиска в мобильные приложения для интернет-маркетплейсов связано с некоторыми трудностями: с необходимостью формирования качественного и разнообразного набора данных, с поиском баланса между точностью и скоростью обработки, а также с обеспечением устойчивости к шуму и артефактам, характерным для пользовательских фотографий [6]. Тем не менее реализация такого программного обеспечения открывает возможности для повышения конверсии, сокращения времени поиска товара и улучшения пользовательского опыта.

Целью настоящего исследования является разработка и сравнительный анализ методов визуального поиска товаров в интернет-маркетплейсах. Научная новизна работы заключается в реализации и обучении модели сиамской нейронной сети для задач поиска товаров, а также в экспериментальной оценке ее эффективности против классических подходов на специализированном наборе данных. Для решения указанных задач и реализации программного обеспечения необходимо проанализировать обзор существующих методов.

1. ОБЗОР МЕТОДОВ ДЛЯ ВИЗУАЛЬНОГО ПОИСКА ТОВАРОВ

Современные системы визуального поиска товаров в мобильных приложениях реализуются с использованием традиционных методов анализа изображений и глубокого обучения. Некоторые классические алгоритмы машинного обучения используют в качестве признаков гистограммы ориентированных градиентов, локальные бинарные шаблоны или текстурные характеристики, применяя для классификации методы опорных векторов или случайный лес. Несмотря на свою распространенность, такие подходы показывают высокую эффективность исключительно в контролируемых условиях: при стабильном освещении, неизменном фоне и фиксированном ракурсе [7]. Однако в реальной эксплуатации, когда пользователь делает снимок на улице, в магазине или при слабом свете, такие подходы быстро теряют эффективность из-за высокой чувствительности к внешним искажениям.

Метод гистограмм ориентированных градиентов описывает локальную структуру объекта через распределение направлений градиентов яркости. Он особенно эффективен для объектов с четкими контурами, таких как одежда или обувь, при стабильном освещении и фиксированном ракурсе. Гистограмма направленных градиентов (HOG) делит изображение на ячейки, в каждой из которых строится гистограмма ориентаций градиентов; затем вектор признаков нормализуется по блокам для повышения устойчивости к изменениям освещения. В сочетании с машиной опорных векторов достигается удовлетворительная разделимость классов в контролируемых условиях. Однако метод плохо переносится на реальные сценарии с шумом, размытием или частичным перекрытием объекта. Направление градиента в точке (x, y) в данном методе рассчитывается по формуле

$$\theta(x, y) = \arctan\left(\frac{G_y(x, y)}{G_x(x, y)}\right), \quad (1)$$

где G_x и G_y – градиенты яркости по горизонтали и вертикали.

Локальные бинарные шаблоны фокусируются на анализе текстуры путем сравнения яркости центрального пикселя с его окружением. Результат кодируется в виде целого числа, что позволяет строить гистограмму текстурных паттернов по всему изображению. Благодаря инвариантности к монотонным изменениям освещения и высокой вычислительной эффективности метод LBP широко использовался в ранних системах распознавания. Однако он не учитывает глобальную форму объекта и чувствителен к поворотам и масштабу, что ограничивает его применимость для поиска товаров по произвольным снимкам. При использовании с SVM обеспечивается умеренная точность на однородных фонах. Значение LBP для центрального пикселя рассчитывается по формуле:

$$\text{LBP}(x_c, y_c) = \sum_{p=1}^{p-1} s(i_p - i_c) 2^p, \quad (2)$$

где $s(z) = 1$ при $z \geq 0$, иначе 0; i_c – яркость центра, i_p – яркость p -го соседа.

Матрица совместной встречаемости уровней серого оценивает статистическую зависимость между парами пикселей на заданном расстоянии и под

определенным углом. На основе GLCM вычисляются скалярные текстурные признаки (контраст, энергия, однородность, корреляция), которые отражают характер поверхности объекта. Эти признаки подаются на вход ансамблевому алгоритму случайного леса, который устойчив к переобучению и сохраняет работоспособность даже при небольших наборах данных. Метод эффективен для различения материалов, но требует тщательной настройки параметров и теряет точность при наличии сложного фона или неоднородного освещения. Контраст как мера текстурной неоднородности вычисляется по формуле

$$Contrast = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - j)^2 P(i, j), \quad (3)$$

где $P(i, j)$ – нормализованная GLCM, L – количество уровней серого.

Несмотря на свою прозрачность и низкие требования к вычислительным ресурсам, все рассмотренные классические подходы уступают современным нейросетевым методам в условиях реального использования. Это ограничение становится особенно критичным в мобильных приложениях, где ключевыми являются три взаимосвязанных требования: 1) высокая скорость процесса вывода для работы в реальном времени, 2) максимально возможная точность распознавания и 3) устойчивость к значительному разнообразию и зашумленности входных данных.

Именно требования удовлетворяются уровнями архитектуры на основе глубокого обучения. В частности, сиамские нейронные сети обладают ключевым преимуществом: они способны извлекать устойчивые семантические закономерности из сильно зашумленных и нестандартных изображений. Такие модели обучаются на парах снимков и формируют единое метрическое пространство, в котором схожие товары оказываются ближе друг к другу, независимо от условий съемки [8]. Практическая ценность этого подхода подтверждена внедрением в коммерческие программные продукты. Например, система Pinterest Lens использует схожий подход для поиска товаров по фото и демонстрирует высокую точность даже при работе с пользовательскими изображениями [9].

Несмотря на различия в реализации, большинство систем визуального поиска строятся по единым этапам обработки: сначала изображение проходит этап предварительной обработки (нормализация, изменение размера, удаление артефактов), затем из него извлекаются признаки, а на завершающем этапе выполняется сравнение с эталонными векторами из каталога. Ключевое отличие заключается в том, как именно формируются эти признаки: вручную в классических методах или автоматически в нейросетевых моделях, что напрямую влияет на их адаптивность и масштабируемость [8]. Тем не менее даже передовые нейросетевые решения сталкиваются с практическими трудностями. Одна из главных проблем – это противоречие между точностью и ресурсоемкостью. Модели, показывающие наилучшие результаты в лабораторных условиях, часто требуют значительных вычислительных ресурсов и работают медленно, что ограничивает их использование на мобильных устройствах. Кроме того, многие обучающие наборы, такие как DeepFashion, состоят преимущественно из профессиональных фотографий, что ограничивает способность модели корректно распознавать товары на реальных пользовательских снимках с размытием, перекосами или частичным перекрытием [10].

Для преодоления этих ограничений активно применяются облегченные архитектуры (например, MobileNetV3 или EfficientNet-Lite), а также методы постобработки моделей, включая квантование весов и удаление избыточных слоев [11]. Эти методы позволяют значительно сократить размер сети и ускорить процесс вывода, сохраняя при этом приемлемый уровень точности.

2. РЕАЛИЗАЦИЯ И ОБУЧЕНИЕ МОДЕЛИ СИАМСКОЙ НЕЙРОННОЙ СЕТИ ДЛЯ ВИЗУАЛЬНОГО ПОИСКА ТОВАРОВ

В рамках настоящего исследования была реализована и обучена модель визуального поиска товаров, предназначенная для интеграции в мобильные приложения интернет-маркетплейсов. В основе решения лежит сиамская нейронная сеть, специально адаптированная для задачи сравнения изображений в условиях реального использования при переменном освещении, нестандартных ракурсах и сложном фоне [12].

В отличие от традиционных классификаторов, которые присваивают изображению метку категории, сиамская архитектура обучена оценивать степень визуального сходства между двумя снимками. Это особенно важно в ситуациях, когда пользователь фотографирует товар в магазине или на улице и ожидает получить подходящие товары только на основе изображения [13].

Архитектурно сеть, представленная на рис. 1, состоит из двух параллельных ветвей с идентичной структурой и общими весами. Каждая ветвь представляет собой сверточный энкодер, преобразующий входное изображение в компактное векторное представление, которое называется эмбедингом. Благодаря общей параметризации один и тот же товар, даже снятый под разными углами или при разном освещении, будет отображаться в близких точках многомерного пространства признаков. На заключительном этапе вычисляется евклидово расстояние между двумя эмбедингами, которое служит мерой несходства объектов.

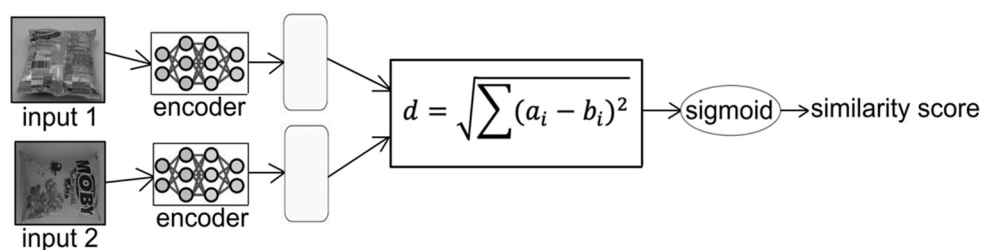


Рис. 1. Структура разработанной сиамской нейронной сети

Fig. 1. Architecture of the developed Siamese neural network

Научная и инженерная новизна предложенного архитектурного решения заключается в разработке специализированной конфигурации сиамской сети, ориентированной на задачу визуального поиска товаров в условиях ограниченных вычислительных ресурсов. В отличие от стандартных подходов, где в качестве энкодера часто используются избыточно глубокие и требовательные к памяти сети (например, ResNet-50), в настоящем исследовании была предло-

жена облегченная сверточная архитектура. За основу энкодера был взят паттерн MobileNetV3, однако его структура была существенно упрощена: вместо 28 сверточных слоев реализована конфигурация из трех последовательных сверточных блоков с постепенным увеличением числа фильтров (32 – 64 – 128). Такая модификация была обусловлена необходимостью снижения времени процесса вывода и объема модели без критической потери точности для предметной области, где визуальные признаки товаров часто локальны и не требуют экстремально глубокой иерархии для своего распознавания.

Исходные данные представляли собой RGB-изображения, масштабированные до разрешения 128×128 пикселей. На последнем сверточном слое применяют глобальное усреднение по пространственным координатам и находят линейный проекционный слой, формирующий 64-мерный вектор признаков. Такой подход позволяет достичь компромисса между выразительностью признакового представления и вычислительной эффективностью.

Обучение модели, представленное на рис. 2, осуществлялось с использованием контрастной функции потерь, которая направлена на сближение эмбеддингов изображений одного товара и отдаление изображений разных товаров. Формула контрастной функции потерь имеет следующий вид:

$$L = \frac{1}{2N} \sum_{i=1}^N \left[y_i d_i^2 + (1 - y_i) \max(\text{margin} - d_i, 0)^2 \right], \quad (4)$$

где y – бинарная метка пары (1 – совпадающие товары, 0 – не совпадающие); d – евклидово расстояние между эмбеддингами; margin – порог разделения, установленный на определенном уровне.

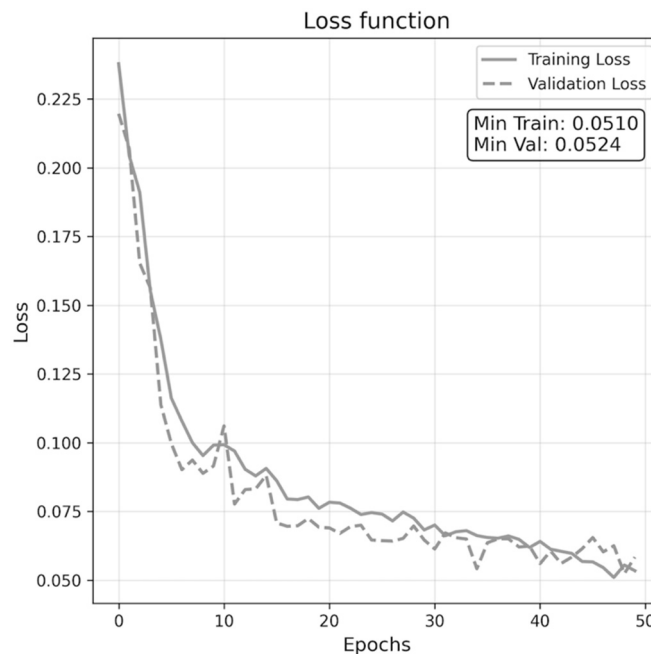


Рис. 2. Обучение SNN-модели по эпохам

Fig. 2. Training an SNN model by epochs

Такой подход позволяет модели сосредоточиться на «сложных» парах, где расстояние между разными классами ещё не превышает заданного порога, что повышает устойчивость к переобучению и улучшает способность к обобщению.

Евклидово расстояние между двумя векторами a_i и b_i рассчитывалось по формуле

$$d = \sqrt{\sum (a_i - b_i)^2}. \quad (5)$$

Это расстояние служит мерой близости векторов: чем меньше значение d , тем более похожими считаются исходные изображения [15]. Для принятия итогового решения рассчитанное расстояние сравнивается с заданным порогом, что позволяет отнести пару изображений к одному и тому же или к разным товарам.

3. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ РАЗЛИЧНЫХ МЕТОДОВ ДЛЯ ВИЗУАЛЬНОГО ПОИСКА ТОВАРОВ

Процесс обучения модели сиамской нейронной сети проводился с использованием оптимизатора Adam (начальная скорость обучения – 10^{-3}) и включал в себя набор преобразований: случайные повороты (до $\pm 15^\circ$), изменение яркости ($\pm 20\%$), горизонтальное отражение.

Для обучения и оценки модели были сформированы обучающая и тестовая выборки. Обучающая выборка включала 1376 изображений, охватывающих 10 товарных категорий, и содержала как профессиональные студийные фотографии, так и пользовательские снимки, сделанные в реальных условиях. Тестовая выборка объемом 344 изображения была сформирована по принципу случайной стратифицированной выборки из исходного набора данных, что обеспечило пропорциональное представление всех товарных категорий как в обучающем, так и в тестовом наборе. Таким образом, общий объем размеченного датасета составил 1720 изображений, разделенных в соотношении 80/20.

По результатам тестирования на независимом наборе данных модель продемонстрировала следующие показатели качества: точность – 0,951, достоверность – 0,937, полнота – 0,967, F1-мера – 0,952, AUC-ROC – 0,9799.

Анализ распределения евклидовых расстояний, продемонстрированный на рис. 3, выявил разделение между парами «схожие товары» и «разные товары». На основе ROC-анализа был выбран оптимальный порог принятия решения, равный 0,55, при котором достигается максимальная F1-мера.

Матрица ошибок, представленная на рис. 4, подтверждает надежность метода: доля ложных срабатываний минимальна.

Для практического внедрения в мобильное приложение был реализован следующий процесс.

1. Предварительная обработка изображения включает нормализацию цветовых каналов, масштабирование до разрешения 128×128 пикселей и коррекцию геометрических искажений.

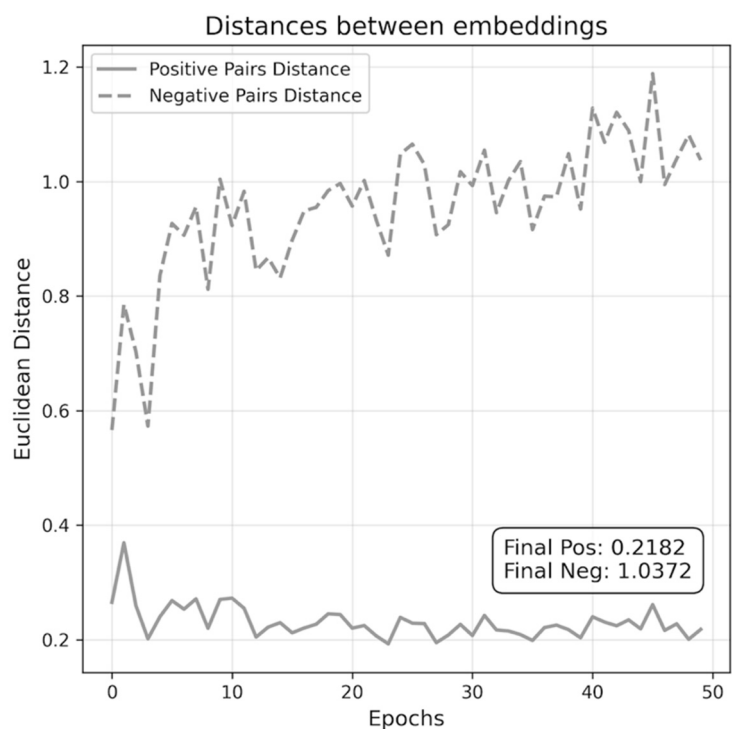


Рис. 3. Распределение евклидовых расстояний между эмбедингами

Fig. 3. Distribution of Euclidean distances between embeddings

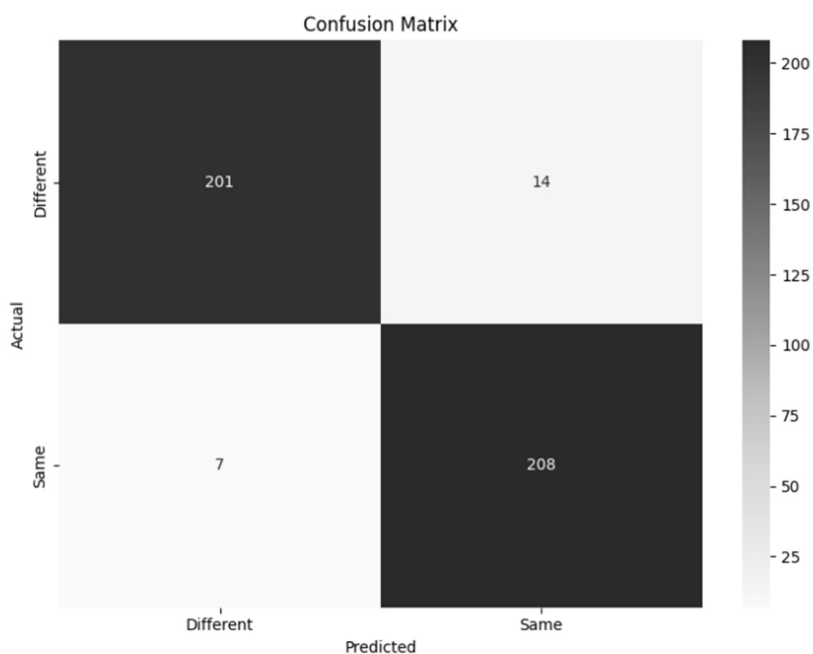


Рис. 4. Матрица ошибок для тестовой выборки

Fig. 4. Confusion matrix for the test set

2. На этапе генерации эмбединга изображение подается на вход оптимизированной сиамской нейронной сети (SNN), реализованной в формате TensorFlow.

3. Поиск осуществляется путем вычисления попарных расстояний между полученным векторным представлением и эталонными векторными представлениями всех товаров в каталоге.

4. Формирование итогового ответа выполняется с помощью отбора до пяти наиболее близких позиций, для которых расстояние до запроса не превышает порогового значения 0,55.

Для проведения сравнительного анализа были реализованы и обучены классические методы: HOG/SVM, LBP/SVM, GLCM/Random Forest. Исходя из результатов, представленных в таблице, можно утверждать, что классические методы показали более низкую эффективность: лучший показатель $F1$ -меры среди них не превысил 0,924. Это подчеркивает преимущество обученной модели глубокого обучения в задачах, требующих устойчивости к внешним искажениям [14].

Значения метрик обученных моделей

Metric values of trained models

Метрики	Методы			
	SNN	HOG/SVM	LBP/SVM	GLCM/Random Forest
Accuracy	0,951	0,739	0,842	0,923
Точность	0,937	0,75	0,854	0,929
Полнота	0,967	0,739	0,842	0,925
F1-мера	0,952	0,739	0,839	0,924

Предложенная SNN-модель продемонстрировала наилучшие результаты по всем метрикам, превзойдя традиционные методы машинного обучения. Наибольшее преимущество наблюдается в метрике полноты (96,7 % в сравнении с 92,5 % у GLCM/Random Forest и 84,2 % у LBP/SVM). Это подтверждает гипотезу о том, что методы метрического обучения на основе глубоких нейронных сетей более эффективно улавливают семантические особенности объектов.

Классические методы машинного обучения показали относительно высокую, но меньшую эффективность. Метод GLCM, основанный на текстурных статистических признаках, показал себя лучше методов, основанных на HOG и LBP.

Значения точности и полноты для SNN-модели (0,937 и 0,967 соответственно) указывают на то, что модель не только точно идентифицирует целевые объекты (меньше ложных срабатываний), но и пропускает мало фактических целей (меньше ложных пропусков). Это имеет критически важное значение для систем визуального поиска, где пропуск целевого объекта (False Negative, FN) может быть столь же нежелателен, как и ложная рекомендация (False Positive, FP).

Таким образом, разработанный модуль визуального поиска сочетает в себе высокую точность распознавания, устойчивость к реальным условиям эксплуатации и эффективность на устройствах с ограниченными ресурсами, что делает его перспективным инструментом для повышения удобства пользователей и увеличения конверсии в мобильных приложениях интернет-маркетплейсов.

ЗАКЛЮЧЕНИЕ

В настоящей статье проведено исследование методических подходов к решению задачи визуального поиска товаров для мобильных приложений интернет-маркетплейсов. В ходе анализа было установлено, что классические методы компьютерного зрения, основанные на выделении признаков вручную, демонстрируют ограниченную эффективность в условиях реальной эксплуатации, в то время как подходы, основанные на глубоком обучении, позволяют напрямую выявлять устойчивые семантические закономерности. В качестве основного методического подхода была выбрана и архитектура сиамской нейронной сети, реализующая метрическое обучение для сравнения изображений через их векторные представления.

На основе выбранного подхода была разработана и реализована SNN-модель, способная сравнивать пользовательские фотографии с изображениями из каталога товаров без необходимости текстового запроса. Экспериментально подтверждено, что предложенный метод, основанный на метрическом обучении и формировании инвариантных векторных представлений, демонстрирует высокую эффективность в условиях, приближенных к реальным условиям: изменчивость освещения, изменение ракурса съемки и наличие сложного фона. На специализированном наборе данных товарных изображений разработанная SNN-модель достигла следующих показателей эффективности: ассурасу – 0,951, точность – 0,937, полнота – 0,967, F1-мера – 0,952 и AUC-ROC – 0,9799. Этот результат значительно превосходит значения показателей качества классических методов машинного обучения (HOG/SVM, LBP/SVM, GLCM/Random Forest), лучший из которых не превысил F1-меру в 0,924. Это свидетельствует о высокой практической применимости нейросетевого подхода в электронной коммерции, где качество пользовательских снимков часто непредсказуемо.

Реализованный метод визуального поиска представляет собой перспективное решение для интеграции в коммерческие мобильные приложения интернет-маркетплейсов. Его внедрение позволяет сократить время поиска нужного товара, улучшить пользовательский опыт и, как следствие, повысить конверсию продаж. Кроме того, оптимизация модели под мобильные платформы (с использованием TensorFlow и квантования после обучения) может обеспечить приемлемую скорость работы даже на устройствах среднего класса, что делает решение масштабируемым и экономически целесообразным.

В качестве направлений будущих исследований рассматриваются расширение функциональности модуля за счет поддержки поиска по частичным или размытым изображениям, интеграция механизмов непрерывного дообучения на основе пользовательских запросов, а также изучение применимости архитектур на основе трансформеров (Vision Transformers, ViT) для дальнейшего повышения точности и устойчивости к экстремальным условиям съемки.

СПИСОК ЛИТЕРАТУРЫ

1. Siamese neural network based multimodal biometrics recognition / H.L. Aung, C. Pluempitiriyawej, S. Wangsiripitak, K. Hamamoto // The 14th Regional Conference on Electrical and Electronics Engineering (RC-EEE 2021). – Thailand, 2022. – P. 147–150.
2. Class-balanced siamese neural networks / S. Berlemont, G. Lefebvre, S. Duffner, C. Garcia // *Neurocomputing*. – 2018. – Vol. 273. – P. 47–56. – DOI: 10.1016/j.neucom.2017.07.060.
3. Khalifa T., Sengul G. The integrated usage of LBP and HOG transformations and machine learning algorithms for age range prediction from facial images // *Tehnički vjesnik*. – 2018. – Vol. 25 (5). – P. 1356–1362. – DOI: 10.17559/TV-20170308030459.
4. Колосов М.И., Дерюгина Е.О. Сравнительный анализ типов нейронных сетей для эффективного распознавания образов, полученных с использованием оптической лазерной триангуляции // Лучшая студенческая работа 2024: сборник статей XIII Международного научно-исследовательского конкурса. – Пенза, 2024. – С. 15–21. – ISBN 978-5-00236-454-1.
5. Новиков А.В. Системное исследование инструментов внедрения машинного обучения на платформе Android // *Universum: технические науки*. – 2024. – Т. 1, № 5 (122). – С. 49–52. – ISBN 978-5-00236-454-1.
6. Гаспарян Д.Э., Стырин Е.М. Прикладные проблемы внедрения этики искусственного интеллекта в России: отраслевой анализ и судебная система. – М.: ВШЭ, 2020. – 112 с. – ISBN 978-5-7598-2351-3.
7. Classification of brain tumor based on shape and texture features and machine learning / M.A. Rizki, M.R. Faisal, A. Farmadi, T.H. Saragih, D.T. Nugrahadi, A.M. Bachtiar, R.R. Keswani // *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*. – 2024. – Vol. 6 (4). – P. 240–251. – DOI: 10.35882/27236g49.
8. Guarino A., Spagnuolo G. Automatic features extraction of faults in PEM fuel cells by a siamese artificial neural network // *International Journal of Hydrogen Energy*. – 2021. – Vol. 46 (70). – P. 34854–34866. – DOI: 10.1016/j.ijhydene.2021.08.024.
9. Клейман А.А., Бабанчикова О.А. Стратегия управления адаптацией продвижения турпродукта к новым трендам туристского рынка // *Инновационные подходы в современной науке* / отв. ред. С.А. Белозеров. – М., 2021. – С. 912–918. – ISBN 978-5-00197-009-5.
10. Метод обнаружения фактов обхода блокировок ресурсов сети Интернет / С.М. Ишкuvatov, А.Н. Бегаев, И.И. Комаров, И.В. Левко // *Вопросы кибербезопасности*. – 2024. – № 3 (61). – С. 76–84. – DOI: 10.21681/2311-3456-2024-3-76-84.
11. Combining lightweight deep learning models with data augmentation for analysis of cervical cells / N. Le-Thi-Thu, P. Dat Doan, A. Phan-Nguyen-Thanh, T. Tran-Thi, D. Tran-Tuan, K. Nguyen-Hoang // *International Conference on Smart Objects and Technologies for Social Good*. – Cham: Springer, 2024. – P. 173–182. – DOI: 10.1007/978-3-032-01497-9_16.
12. Feng K., Chaspari T. Few-shot learning in emotion recognition of spontaneous speech using a siamese neural network with adaptive sample pair formation // *IEEE Transactions on Affective Computing*. – 2021. – Vol. 14 (2). – P. 1627–1633. – DOI: 10.1109/TAFFC.2021.3109485.
13. Bos H., Muir D. Sub-mW Neuromorphic SNN audio processing applications with Rockpool and Xylo // *Embedded Artificial Intelligence: Applications and Advanced Engineering* / ed. by K.R. Kumar. – River Publishers, 2023. – P. 69–78. – ISBN 9781003394440.
14. Optimization of HOG algorithm for computer vision in industrial robots / N.M. Sawai, A.R. Mankar, R. Mahant, K. Ingole, Y.R. Mahulkar, S.M. Asutkar // *2024 IEEE International Conference on Computing, Power and Communication Technologies (ICCPCT)*. – IEEE, 2024. – P. 572–577. – DOI: 10.1109/IC2PCT60090.2024.10486349.
15. Antonyants E.N., Aletdinova A.A. Research and development of software for comparison of visual data with hashing methods // *Progress through innovations: труды XIII Всероссийской научно-практической конференции аспирантов и магистрантов*, Новосибирск, 29 апр. 2025 г. – Новосибирск: Изд-во НГТУ, 2025. – С. 19–21. – ISBN 978-5-7782-5484-8.

Антонянц Егор Николаевич, аспирант, ассистент кафедры автоматизированных систем управления Новосибирского государственного технического университета. Основное направление научных исследований – машинное обучение и системы компьютерного зрения. Имеет 32 печатные работы. E-mail: bax201438@gmail.com

Алетдинова Анна Александровна, доктор экономических наук, кандидат технических наук, профессор кафедры автоматизированных систем управления Новосибирского государственного технического университета. Основное направление научных исследований – анализ данных и системы искусственного интеллекта. Имеет 344 печатные работы. E-mail: aletdinova@corp.nstu.ru

Шипунов Дмитрий Николаевич, студент бакалавриата факультета автоматики и вычислительной техники Новосибирского государственного технического университета. Основное направление научных исследований – машинное обучение и глубокое обучение. E-mail: shipunov_dmitriy@inbox.ru

Седельников Александр Александрович, студент бакалавриата факультета автоматики и вычислительной техники Новосибирского государственного технического университета. Основное направление научных исследований – нейронные сети и системы компьютерного зрения. E-mail: prodaga05@mail.ru

Симоняк Семен Вячеславович, студент бакалавриата факультета автоматики и вычислительной техники Новосибирского государственного технического университета. Основное направление научных исследований – нейронные сети и глубокое обучение. E-mail: SemenSimon21@gmail.com

Antonyants Egor N., postgraduate student, assistant lecturer at the Department of Automated Control Systems, Novosibirsk State Technical University. His research interests are currently focused on machine learning and computer vision systems. He has more than 32 publications. E-mail: bax201438@gmail.com

Aletdinova Anna A., D. Sc. (Econ.), Ph.D. (Eng.), full professor at the Department of Automated Control Systems, Novosibirsk State Technical University. His research interests are currently focused on data analysis and artificial intelligence systems. She has more than 45 publications. E-mail: aletdinova@corp.nstu.ru

Shipunov Dmitry N., undergraduate student at the Faculty of Automation and Computer Engineering, Novosibirsk State Technical University. His research interests are currently focused on machine learning and deep learning. E-mail: shipunov_dmitriy@inbox.ru

Sedelnikov Alexander A., undergraduate student at the Faculty of Automation and Computer Engineering, Novosibirsk State Technical University. His research interests are currently focused on neural networks and computer vision systems. E-mail: prodaga05@mail.ru

Simonyak Semyon V., undergraduate student at the Faculty of Automation and Computer Engineering, Novosibirsk State Technical University. His research interests are currently focused on neural networks and deep learning. E-mail: SemenSimon21@gmail.com

DOI: 10.17212/2782-2001-2025-4-39-52

Investigation of methodological approaches to implementing visual product search in online marketplaces*

*E.N. ANTONYANTS^a, A.A. ALETDINOVA^b, D.N. SHIPUNOV^c,
A.A. SEDELNIKOV^d, S.V. SIMONYAK^e*

Novosibirsk State Technical University, 20 K. Marx Prospekt, Novosibirsk, 630073, Russian Federation

^a bax201438@gmail.com ^b aletdinova@corp.nstu.ru ^c shipunov_dmitriy@inbox.ru

^d prodaga05@mail.ru ^e SemenSimon21@gmail.com

Abstract

The article presents the research of methods of visual product search and the results of their software implementation in e-commerce systems. The work provides an overview of classical machine learning methods and modern neural network architectures. For comparison, the following approaches were selected and implemented: Histogram of Oriented Gradients/Support Vector

* Received 14 October 2025.

Machine (HOG/SVM), Local Binary Patterns/Support Vector Machine (LBP/SVM), Gray Level Co-occurrence Matrix/Random Forest (GLCM/Random Forest) and the model of Siamese Neural Network (SNN). The software developed within the research includes methods for image preprocessing, feature extraction, comparison of vector representations with optimal classification threshold of 0,55. Experimental study was conducted on the target dataset of product images using the evaluation of accuracy metrics. The best result was shown by the Siamese Neural Network, which achieved an accuracy value equal to 0,951, precision – 0,937, completeness – 0,967, F1-score – 0,952 and AUC-ROC – 0,9799. During the research, the ability of the SNN-model to demonstrate high resistance to changes in lighting, shooting angles and complex background was noted, while maintaining stable accuracy of product identification. The obtained results demonstrate the practical significance of the proposed approach for creating visual product search systems in mobile applications of online marketplaces, enabling the improvement of user experience and increasing sales levels. The research confirmed the effectiveness of neural network methods for the tasks of image analysis and matching in the conditions of the real retail segment.

Keywords: visual product search, e-commerce, siamese neural networks, histogram of oriented gradients, support vector machine, local binary patterns, gray level co-occurrence matrix, random forest

REFERENCES

1. Aung H.L., Pluempitiwiriyaew C., Wangsiripitak S., Hamamoto K. Siamese neural network based multimodal biometrics recognition. *The 14th Regional Conference on Electrical and Electronics Engineering (RC-EEE 2021)*, Thailand 2022, pp. 147–150.
2. Berlemont S., Lefebvre G., Duffner S., Garcia C. Class-balanced siamese neural networks. *Neurocomputing*, 2018, vol. 273, pp. 47–56. DOI: 10.1016/j.neucom.2017.07.060.
3. Khalifa T., Sengul G. The integrated usage of LBP and HOG transformations and machine learning algorithms for age range prediction from facial images. *Tehnički vjesnik*, 2018, vol. 25 (5), pp. 1356–1362. DOI: 10.17559/TV-20170308030459.
4. Kolosov M.I., Deryugina E.O. [Comparative analysis of neural network types for effective pattern recognition using optical laser triangulation]. *Luchshaya studentcheskaya rabota 2024* [Best Student Work, 2024]. Collection of articles from the XIII International Research Competition. Penza, 2024, pp. 15–21. ISBN 978-5-00236-454-1. (In Russian).
5. Novikov A.V. Sistemnoe issledovanie instrumentov vnedreniya mashinnogo obucheniya na platforme Android [System research of machine learning implementation tools on the Android platform]. *Universum: tekhnicheskie nauki = Universum: Technical Sciences*, 2024, vol. 1, no. 5 (122), pp. 49–52. ISBN 978-5-00236-454-1.
6. Gasparyan D.E., Styryn E.M. *Prikladnye problemy vnedreniya etiki iskusstvennogo intellekta v Rossii: otraslevoi analiz i sudebnaya sistema* [Applied problems of implementing artificial intelligence ethics in Russia: industry analysis and judicial system]. Moscow, HSE Publ., 2020. 112 p. ISBN 978-5-7598-2351-3.
7. Rizki M.A., Faisal M.R., Farmadi A., Saragih T.H., Nugrahadi D.T., Bachtiar A.M., Keswani R.R. Classification of brain tumor based on shape and texture features and machine learning. *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 2024, vol. 6 (4), pp. 240–251. DOI: 10.35882/27236g49.
8. Guarino A., Spagnuolo G. Automatic features extraction of faults in PEM fuel cells by a siamese artificial neural network. *International Journal of Hydrogen Energy*, 2021, vol. 46 (70), pp. 34854–34866. DOI: 10.1016/j.ijhydene.2021.08.024.
9. Kleiman A.A., Babanchikova O.A. Strategiya upravleniya adaptatsiei prodvizheniya turprodukta k novym trendam turistskogo rynka [Strategy for managing the adaptation of tourism product promotion to new trends in the tourist market]. *Innovatsionnye podkhody v sovremennoi nauke* [Innovative approaches in modern science]. Moscow, 2021, pp. 912–918. ISBN 978-5-00197-009-5.
10. Ishkuvatov S.M., Begaev A.N., Komarov I.I., Levko I.V. Metod obnaruzheniya faktov obkhoda blokirovok resursov seti Internet [Method for detecting facts of circumvention of internet resource locks]. *Voprosy kiberbezopasnosti = Cybersecurity Issues*, 2024, no. 3 (61), pp. 76–84. DOI: 10.21681/2311-3456-2024-3-76-84. (In Russian).

11. Le-Thi-Thu N., Dat Doan P., Phan-Nguyen-Thanh A., Tran-Thi T., Tran-Tuan D., Nguyen-Hoang K. Combining lightweight deep learning models with data augmentation for analysis of cervical cells. *International Conference on Smart Objects and Technologies for Social Good*. Cham, Springer, 2024, pp. 173–182. DOI: 10.1007/978-3-032-01497-9_16.
12. Feng K., Chaspari T. Few-shot learning in emotion recognition of spontaneous speech using a siamese neural network with adaptive sample pair formation. *IEEE Transactions on Affective Computing*, 2021, vol. 14 (2), pp. 1627–1633. DOI: 10.1109/TAFFC.2021.3109485.
13. Bos H., Muir D. Sub-mW Neuromorphic SNN audio processing applications with Rockpool and Xylo. *Embedded Artificial Intelligence: Applications and Advanced Engineering*. Ed. by K.R.R. Kumar. River Publishers, 2023, pp. 69–78. ISBN 9781003394440.
14. Sawai N.M., Mankar A.R., Mahant R., Ingole K., Mahulkar Y.R., Asutkar S.M. Optimization of HOG algorithm for computer vision in industrial robots. *2024 IEEE International Conference on Computing, Power and Communication Technologies (ICCPCT)*. IEEE, 2024, pp. 572–577. DOI: 10.1109/IC2PCT60090.2024.10486349.
15. Antonyants E.N., Aletdinova A.A. Research and development of software for comparison of visual data with hashing methods. *Proceedings 2025 XIII All-Russian Academic and Research Conference of Graduate and Postgraduate Students*, Novosibirsk, April 29, 2025. Novosibirsk, NSTU Publ., 2025, pp. 19–21. ISBN 978-5-7782-5484-8.

Для цитирования:

Исследование методических подходов реализации визуального поиска товаров в интернет-маркетплейсах / Е.Н. Антонянц, А.А. Алетдинова, Д.Н. Шипунов, А.А. Седельников, С.В. Симоняк // Системы анализа и обработки данных. – 2025. – № 4 (100). – С. 39–52. – DOI: 10.17212/2782-2001-2025-4-39-52.

For citation:

Antonyants E.N., Aletdinova A.A., Shipunov D.N., Sedelnikov A.A., Simonyak S.V. Issledovanie metodicheskikh podkhodov realizatsii vizual'nogo poiska tovarov v internet-marketpleisakh [Investigation of methodological approaches to implementing visual product search in online marketplaces]. *Sistemy analiza i obrabotki dannykh = Analysis and data processing systems*, 2025, no. 4 (100), pp. 39–52. DOI: 10.17212/2782-2001-2025-4-39-52.