

## СОВРЕМЕННЫЕ АРХИТЕКТУРЫ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ТЕГИРОВАНИЯ И АННОТИРОВАНИЯ ИЗОБРАЖЕНИЙ: ДОСТИЖЕНИЯ, ВЫЗОВЫ И ПЕРСПЕКТИВЫ

Г. А. Резников • Р. Д. Сеницын • А. М. Шулик

**Аннотация.** В статье исследуется растущая роль тегирования изображений в условиях стремительного увеличения объема визуальных данных, который наблюдается в таких областях, как социальные сети, медицина, безопасность. Тегирование, включая хэштеги и автоматическую маркировку, становится важнейшим инструментом для систематизации, поиска и фильтрации контента, что значительно улучшает пользовательский опыт, облегчает доступ к информации и позволяет эффективно обрабатывать большие массивы данных. Научная новизна работы заключается в комплексном анализе современных архитектур нейронных сетей, применяемых для распознавания и аннотирования изображений. В исследовании рассматриваются не только традиционные методы, такие как сверточные нейронные сети (CNN) и остаточные сети (ResNet), но и новые гибридные модели, а также подходы, основанные на трансформаторах, которые демонстрируют высокую эффективность при решении задач автоматического тегирования и аннотирования изображений. Особое внимание уделяется анализу точности этих моделей, их преимуществ и ограничений в контексте обработки изображений. В статье также обсуждаются актуальные проблемы и риски, связанные с применением нейронных сетей для аннотирования изображений, включая нехватку размеченных данных, низкую интерпретируемость моделей, проблемы обобщенности, а также этические вопросы, такие как предвзятость алгоритмов, защита персональных данных и вопросы конфиденциальности. В заключение делается вывод о значимости нейронных сетей для развития технологий распознавания изображений и подчеркивается необходимость дальнейших исследований для повышения точности, эффективности и этичности существующих подходов в данной области, что открывает новые горизонты для применения технологий в различных отраслях, включая медицину, автомобильную промышленность и другие.

**Ключевые слова:** нейронные сети; маркировка; классификация; архитектура нейронной сети; машинное обучение.

### ВВЕДЕНИЕ

В современную цифровую эпоху большая часть информации воспринимается визуально, а количество фотографий, хранящихся на устройствах пользователей, растет в геометрической прогрессии. По данным Worldmetrics, в 2020 году в мире было сделано 1.4 триллиона фотографий [1]. Этот рост обусловлен повсеместным распространением смартфонов и простой получения и обмена изображениями. Следовательно, растет потребность в эффективной систематизации и категоризации этих огромных визуальных данных.

Одним из способов решения этой задачи стало внедрение тегирования изображений, которое стало важным элементом социальных сетей и мобильных устройств. Различные социальные сети в значительной степени полагаются на хэштеги для упорядочивания контента и облегчения поиска. На мобильных устройствах изображения часто организуются в альбомы не только по категориям, определяемым пользователем, но и с помощью автоматической маркировки на основе метаданных, таких как местоположение и распознавание лиц.

Поисковые системы, такие как Google и Яндекс, активно используют теги и метаданные изображений для получения релевантных изображений в ответ на запросы пользователей. Алгоритмы поиска изображений Google используют теги, подписи и альт-текст для повышения точности результатов поиска. В 2020 году на долю Google Images приходилось 22.6 % всех

поисковых запросов в Интернете [2], что подчеркивает важность эффективного использования тегов изображений при поиске информации.

Тегирование изображений играет важнейшую роль в управлении и осмыслении постоянно растущего объема визуальных данных. Его важность подчеркивают несколько ключевых факторов: расширение возможностей поиска, повышение вовлеченности пользователей, организация контента, соответствие требованиям доступности и аналитика данных. Несмотря на свою важность, ручная маркировка изображений является трудоемкой и зачастую непрактичной, учитывая масштабы данных. В связи с этим возрос интерес к автоматизации процесса тегирования с помощью искусственного интеллекта, что привело к появлению большого количества архитектур нейронных сетей.

Нейронные сети произвели революцию во многих областях, в том числе и в области тегирования изображений, обеспечив более точную и эффективную обработку визуальных данных. Развитие алгоритмов и вычислительных технологий значительно упростило решение сложных задач в обработке изображений, таких как распознавание лиц, идентификация и классификация объектов, создание искусственных изображений, а также восстановление и поиск схожих картинок.

Методы машинного обучения в обработке изображений открывают широкий спектр новых возможностей, оказывая положительное влияние на различные аспекты нашей жизни. Эти технологии способствуют оптимизации процессов, улучшению точности анализа данных и созданию новых инструментов для работы. Благодаря нейронным сетям мы уже способны выполнять точный анализ и интерпретацию изображений, распознавая объекты и решая задачи, которые ранее казались слишком сложными.

## ОБЗОР ОСНОВНЫХ НЕЙРОННЫХ СЕТЕЙ

В настоящее время существуют несколько видов архитектур нейронных сетей для обработки изображений. Ниже рассмотрены основные актуальные и самые распространенные на данный момент архитектуры.

*Сверточные нейронные сети* – это нейронные сети, концептуально разработанные в 1990-х годах. Они используют стек сверточных слоев для извлечения пространственных признаков. Сверточные слои применяют последовательность фильтров для выявления таких ключевых признаков, как края и цвет, простые формы из матрицы пикселей. На основе доработанной сверточной архитектуры были разработаны следующие нейронные сети: MobileNet (2017) – использует разделенные свертки для уменьшения размера модели и вычислений и показывает точность 72 % [3] при использовании датасета ImageNet. SqueezeNet (2016) [4] – использует квантование и отсечение параметров модели для сжатия размеров модели при сохранении точности широко известной AlexNet – 57 %. EfficientNet – архитектура сверточной нейронной сети и метод масштабирования, который равномерно масштабирует все измерения глубины/ширины/разрешения с помощью составного коэффициента. В отличие от обычной практики, которая произвольно масштабирует эти факторы, метод масштабирования EfficientNet равномерно масштабирует ширину, глубину и разрешение сети с помощью набора фиксированных коэффициентов масштабирования и достигает точности в 84.4 % [5].

*Остаточные нейронные сети* – это архитектура нейронных сетей, в которой веса обучаются остаточными функциями с привязкой к входным данным слоя. Чтобы решить проблему исчезающего/исчезающего градиента, в этой архитектуре было введено понятие «остаточные блоки». В этой сети мы используем технику, называемую пропускными соединениями. Пропускное соединение соединяет активации слоя с последующими слоями, пропуская некоторые слои между ними. Таким образом формируется остаточный блок. Сети создаются путем объединения этих остаточных блоков [6].

*Капсульные нейронные сети* – данный тип нейронных сетей эмулирует иерархические связи, как в биологических структурах. В данной архитектуре группы нейронов объединены в капсулы, которые в отличие от обычных нейронов сверточных сетей лучше обрабатывают

иерархические данные и изменения расположения объекта на изображении, так как капсулы объединяют функцию активации и пространственные связи. В результате CapsNet могут лучше распознавать сложные паттерны в данных, улавливать сложные пространственные иерархии и улучшать обобщение и достигают точности в 99.75 % на наборе данных MNIST [7].

*К-ближайших соседей* в сочетании со сверточными нейронными сетями – гибридный подход, сочетающий в себе простоту алгоритма KNN и возможности CNN по извлечению признаков. В этом методе предварительно обученная CNN используется для извлечения высокоуровневых дискриминационных признаков из изображений. Как правило, эти признаки определяются активационными выходами более глубоких слоев CNN, которые отражают сложные паттерны и представления в изображениях. После извлечения признаков применяется алгоритм KNN для классификации новых изображений на основе ближайших векторов признаков в обучающем наборе. Этот подход позволяет использовать богатые представления признаков, изучаемые CNN, без необходимости переобучения всей сети для задач классификации и достигает точности 54.85 % на обучающем наборе Corel-5K [8].

*Глубокое метрическое обучение* (DML) – это подход, направленный на изучение пространства признаков, в котором точки данных располагаются на основе их сходства. В архитектурах DML нейронные сети обучаются отображать входные данные в пространство вкраплений. Задача состоит в том, чтобы обеспечить близкое расположение схожих входных данных, а несхожие отеснить друг от друга. Это достигается за счет использования функций потерь, таких как триплетные потери или контрастные потери, которые явно штрафуют сеть, когда похожие точки данных не выравниваются в пространстве признаков. Они достигают высокой точности, поскольку оптимизируют пространство признаков непосредственно для задачи измерения сходства, а не простой классификации, достигается при этом точность более 94 % на таких наборах данных, как NUS-WIDE [9].

*Модели на основе трансформаторов* – данные нейронные сети используют механизмы самовнимания, чтобы улавливать взаимосвязи во всей последовательности входных данных, независимо от расстояния. Это позволяет им более эффективно моделировать зависимости между областями на изображении. Типичная архитектура на основе трансформаторов для решения задач «изображение–текст» состоит из системы кодирования–декодирования. Кодер извлекает визуальные признаки из изображений, часто используя предварительно обученные CNN или трансформаторы зрения (ViT), которые могут обрабатывать фрагменты изображений как лексемы. Затем декодер генерирует текст, обращая внимание на извлеченные особенности и создавая подписи слово за словом. Наиболее продвинутые модели достигают точности в 95 % на наборе MSCOCO [10].

Такое разнообразие архитектур нейронных сетей обусловлено необходимостью повышения производительности, уменьшения вычислительных затрат, улучшения устойчивости к шуму и достижению более высокой точности в сложных задачах обработки данных. Каждая новая архитектура предлагает уникальные преимущества и решает специфические проблемы, что делает их важными для дальнейшего прогресса в области искусственного интеллекта.

### СРАВНИТЕЛЬНЫЙ АНАЛИЗ ВЛИЯНИЯ ДАТАСЕТОВ

Помимо архитектуры, значительное влияние на точность, корректность и качество аннотирования оказывает содержание и разнообразие датасета, используемого при обучении. Нейронные сети без возможности дообучения в процессе эксплуатации сильно ограничены заложенными тегами в начальный датасет, что делает подобные архитектуры более узкими и менее применимыми в условиях меняющихся трендов. Ниже рассмотрен один из самых продвинутых и популярных датасетов.

ImageNet – это огромная база изображений, организованная по иерархии WordNet. В ней более 1.2 миллиона изображений для обучения и 50 тысяч изображений для проверки, распределенных по 1000 категориям.

CIFAR-10 – это набор из 60 000 цветных изображений размером  $32 \times 32$ , распределенных по 10 классам (по 6000 изображений в каждом классе). Он включает 50 000 изображений для обучения и 10 000 для тестирования. CIFAR-10 популярен среди исследователей в задачах распознавания образов из-за своего небольшого размера и разнообразия.

CIFAR-100 также содержит 60 000 цветных изображений  $32 \times 32$ , но уже распределенных по 100 классам, каждый из которых содержит по 600 изображений.

Pascal VOC 2007 включает 9963 изображения с аннотациями для 20 категорий объектов. Этот набор данных предназначен для задач обнаружения, классификации и сегментации объектов, предоставляя стандартные эталоны для оценки алгоритмов.

COCO (Common Objects in Context) – это крупный набор данных с более чем 200 000 изображений, в которых отмечены 80 категорий объектов. Он содержит детализированные аннотации, включая множественные метки для одного изображения.

NUS-WIDE-270K – это коллекция реальных изображений из Интернета, включающая 269 648 снимков, собранных с Flickr. Она аннотирована 81 меткой, что делает её важной для исследований многометочной классификации и анализа социальных сетей.

MIRFlickr-25K содержит 25 000 изображений с Flickr, каждая из которых имеет одну или несколько из 24 семантических меток. Этот набор данных используется для аннотации изображений, поиска и мультимедийных исследований, включая классификацию с несколькими метками и изучение семантических разрывов.

MNIST – это база данных с 70 000 изображениями рукописных цифр (от 0 до 9), каждое изображение имеет размер  $28 \times 28$  пикселей. Она разделена на 60 000 обучающих и 10 000 тестовых изображений и является основополагающим набором для тестирования алгоритмов классификации и начальных задач в компьютерном зрении и машинном обучении.

В табл. 1 на основе ранее проделанных исследований [11–31] приведено сравнение полученных результатов точности различных архитектур нейронных сетей на приведенных выше датасетах.

Таблица 1

**Сравнение точности архитектур нейронных сетей при различных датасетах**

| Neural Network       | ImageNet | CIFAR-10 | CIFAR-100 | Pascal VOC 2007 | COCO   | NUS-WIDE-270K | MIRFlickr-25K | MNIST  |
|----------------------|----------|----------|-----------|-----------------|--------|---------------|---------------|--------|
| CNN (AlexNet)        | 62.5 %   | 82.5 %   | 54.0 %    | 78.1 %          | 39.6 % | 42.3 %        | 57.2 %        | 99.1 % |
| MobileNet            | 70.6 %   | 90.3 %   | 65.4 %    | 81.3 %          | 21.3 % | 43.5 %        | 58.7 %        | 99.0 % |
| SqueezeNet           | 57.5 %   | 87.4 %   | 61.2 %    | 77.0 %          | 19.2 % | 39.4 %        | 55.0 %        | 99.0 % |
| EfficientNet-B7      | 84.4 %   | 98.9 %   | 91.7 %    | 85.5 %          | 51.3 % | 49.1 %        | 61.3 %        | 99.7 % |
| ResNet-152           | 78.3 %   | 93.8 %   | 78.3 %    | 83.8 %          | 46.9 % | 45.6 %        | 60.1 %        | 99.5 % |
| CapsNet              | 59.0 %   | 89.4 %   | 66.0 %    | 79.8 %          | 34.5 % | 40.2 %        | 56.7 %        | 99.7 % |
| KNN+CNN              | ~61 %    | 91.2 %   | 67.4 %    | 80.5 %          | 37.5 % | 40.2 %        | 56.0 %        | 97.0 % |
| Deep Metric Learning | 65.5 %   | 94.5 %   | 77.2 %    | 84.2 %          | 49.7 % | 47.2 %        | 60.8 %        | 99.0 % |

На основе полученных данных можно сделать следующие выводы: объемные датасеты, такие как ImageNet и COCO, показывают большую вариативность результатов тестирования точности, EfficientNet в среднем показывает наилучшую точность на больших и малых датасетах, что свидетельствует о ее хорошей масштабируемости и общей эффективности, однако данная архитектура требует значительных вычислительных ресурсов. Капсульные нейронные сети показывают лучшую производительность на небольших датасетах, однако при увеличении количества тегов для классификации точность нейронной сети падает, KNN с признаками CNN предлагает простую в реализации комбинацию, но показывает меньшую эффективность

и точность на больших датасетах по сравнению с более продвинутыми архитектурами. Глубокое метрическое обучение эффективно в задачах, требующих измерения сходства между изображениями, таких как поиск и кластеризация, демонстрируя высокую точность на соответствующих датасетах.

По результатам сравнительного анализа можно сделать вывод о том, что выбор датасета для обучения оказывает критическое влияние на получаемую точность распознавания выбранной нейронной сетью, поэтому для отдельных ситуаций применения моделей нужно учитывать не только архитектуру, но и набор данных для обучения.

### ПРИМЕНЕНИЕ РАЗРАБОТАННЫХ МОДЕЛЕЙ

Уже сейчас разработанные нейронные сети используются в самых разных сферах, значительно упрощая и ускоряя процессы обработки и анализа данных. Важно отметить, что для каждой области применяются специфические датасеты.

Одной из наиболее значимых областей применения нейронных сетей является общественная безопасность. Алгоритмы распознавания лиц, основанные на сверточных нейронных сетях, позволяют идентифицировать и отслеживать людей на видеозаписях в реальном времени. Это используется для предотвращения преступлений, поиска пропавших без вести, а также для решения прикладных задач, таких как контроль доступа к важным объектам и мероприятиям. Для тренировки моделей распознавания лиц часто используются такие датасеты, как LFW (Labeled Faces in the Wild) и VGGFace.

В научных исследованиях нейронные сети играют ключевую роль в анализе сложных и больших объёмов данных. Например, в астрономии они успешно применяются для обработки астрономических изображений, что позволяет учёным выявлять новые объекты и аномалии, а эффективная обработка данных из телескопов и других источников помогает глубже исследовать космос и делать новые открытия. Для этих целей используются датасеты, такие как SDSS (Sloan Digital Sky Survey) и Hubble Space Telescope images.

Другим примером является медицина. Нейронные сети используются для классификации тканей и опухолевых клеток на основе микроскопических изображений. Это содействует медикам в постановке более точных диагнозов и разработке индивидуализированных лечебных планов, что, в свою очередь, может значительно улучшить результаты лечения. Например, алгоритмы глубокого обучения способны распознавать паттерны, которые могут быть не заметны человеческому глазу, что повышает точность диагностики. В этой области активно применяются такие датасеты, как CAMELYON (для обнаружения метастазов в лимфатических узлах) и MIT-BIH Arrhythmia Database (для классификации сердечных заболеваний).

В сфере автономных транспортных средств нейронные сети играют ключевую роль в обеспечении безопасности и эффективности движения. Алгоритмы глубокого обучения помогают автомобилям распознавать такие объекты на дороге, как пешеходы, другие транспортные средства и дорожные знаки. Это позволяет автомобилям принимать обоснованные решения в реальном времени, что критически важно для предотвращения аварий. Например, системы автоматического вождения используют сверточные нейронные сети для анализа видеопотоков с камер, обеспечивая высокую степень безопасности и надежности. Для обучения этих моделей часто используются такие датасеты, как KITTI и Cityscapes.

В промышленности и производстве нейронные сети также находят своё применение. Они используются для автоматизации процессов контроля качества на производственных линиях, позволяя быстро выявлять дефекты и сокращать количество бракованных изделий. Алгоритмы глубокого обучения могут анализировать изображения продукции, сравнивая их с эталонными стандартами. Это не только повышает эффективность производства, но и позволяет значительно сократить затраты. Кроме того, нейронные сети могут оптимизировать производственные процессы, анализируя данные о работе оборудования и выявляя возможности для улучшения. В этой сфере часто применяются такие датасеты, как CIFAR-10 (для распознавания объектов) и MVTeс AD (для обнаружения дефектов в продукции).

## ТЕКУЩИЕ ПРОБЛЕМЫ

Тем не менее, несмотря на все достижения, существующие архитектуры нейронных сетей также сталкиваются с рядом вызовов и проблем.

Во-первых, большинство нейронных сетей требуют больших объёмов данных для эффективного обучения. Однако в реальных сценариях часто недостаточно размеченных данных, особенно для узкоспециализированных задач. Это может привести к переобучению модели, когда она хорошо работает на тренировочных данных, но имеет плохую обобщающую способность на новых, неразмеченных данных.

Во-вторых, изображения могут содержать шум или искажения, которые негативно влияют на производительность модели.

В-третьих, как было продемонстрировано в данной работе, не все архитектуры в своей нынешней реализации одинаково эффективно работают в различных ситуациях (при использовании различных датасетов), а также у некоторых моделей отмечается деградация при росте данных тренировки, что свидетельствует о проблемах с масштабируемостью.

Отдельно стоит выделить этические и правовые аспекты, ведь нейронные сети могут унаследовать предвзятость, существующую в данных, на которых они обучаются. Это может привести к дискриминации определённых групп людей или некорректным результатам работы.

Также крайне важной проблемой являются необходимые для обучения вычислительные ресурсы, так как нейронные сети требуют значительных вычислительных мощностей для обучения и работы, что может стать препятствием для отдельных исследователей и малых компаний. Высокие затраты на оборудование и электроэнергию делают внедрение этих технологий сложным и дорогостоящим. Кроме того, время, необходимое для обучения больших моделей, также может быть значительным.

## ЗАКЛЮЧЕНИЕ

В ходе исследования была рассмотрена важность тегирования изображений в условиях стремительного увеличения визуальных данных и проанализированы современные архитектуры нейронных сетей, применяемые для распознавания и аннотирования изображений. Технологии, основанные на нейронных сетях, демонстрируют значительные преимущества в точности и эффективности обработки визуальных данных, что делает их ключевыми инструментами в самых различных областях, от социальных сетей до медицины и автономного транспорта.

Научная новизна данной работы заключается в комплексном анализе как традиционных, так и новых гибридных архитектур нейронных сетей, включая подходы на основе трансформаторов. Были выделены их основные преимущества и недостатки, а также проанализирована их точность в задачах аннотирования изображений. Это позволяет более глубоко понять возможности и ограничения современных технологий в контексте автоматизации тегирования.

Тем не менее, несмотря на достижения в данной области, существуют и значительные недостатки: проблемы нехватки размеченных данных, интерпретируемость моделей и потенциальные этические риски остаются актуальными и требуют дальнейшего внимания. Важно отметить, что выбор датасета для обучения нейронных сетей существенно влияет на их производительность и способность обобщать результаты, что подчеркивает необходимость в комплексном подходе к обучению и оценке моделей.

С учетом текущих тенденций и прогнозов роста рынка компьютерного зрения можно уверенно утверждать, что будущее нейронных сетей в сфере тегирования изображений является многообещающим. Ожидается, что этот сектор будет продолжать развиваться, создавая новые возможности для применения технологий в различных областях. Можно уверенно сказать, что дальнейшие исследования и инновации в области искусственного интеллекта и машинного обучения помогут преодолеть существующие преграды и откроют новые горизонты для автоматизации обработки визуальных данных.

## БЛАГОДАРНОСТИ

Авторы выражают признательность преподавателям кафедры инструментального и прикладного программного обеспечения РТУ МИРЭА за ценные консультации по аспектам рассматриваемой предметной области. Также следует отметить публикации в смежных областях применения нейронных сетей [32–35], оказавшие влияние на данное исследование.

## СПИСОК ЛИТЕРАТУРЫ / REFERENCES

- [1] Global Photography Statistics: Booming Market, Smartphone Dominance, Social Media Influence / Worldmetrics [Online]. Available: <https://worldmetrics.org/photography-statistics> [Accessed Oct. 30, 2024].
- [2] Insights from Googles Remarkable Word Search Statistics Worldwide / Worldmetrics [Online]. Available: <https://worldmetrics.org/google-word-search-statistics/> [Accessed Oct. 30, 2024].
- [3] Andrew G. Howard, Menglong Zhu, Bo Chen, et al. "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications". arXiv preprint arXiv:2101.12491, 2017.
- [4] Forrest N. Iandola, Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, Kurt Keutzer. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size". arXiv preprint arXiv:1602.07360, 2016.
- [5] Mingxing Tan, Quoc V. Le. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks". arXiv preprint arXiv:1905.11946, 2019.
- [6] Mazzia V., Salvetti F., Chiaberge M. "Efficient-CapsNet: capsule network with self-attention routing". arXiv preprint arXiv:2101.12491, 2019.
- [7] Bensaci R., Khaldi B., Aiadi O., Benchabana A. "Deep Convolutional Neural Network with KNN Regression for Automatic Image Annotation" // Applied Sciences. 2021. Vol. 11(21).
- [8] Myasar Mundher Adnan, Mohd Shafry Mohd Rahim, et al. "Automatic Image Annotation Based on Deep Learning Models: A Systematic Review and Future Challenges" // IEEE Access. 2021. Vol. 9.
- [9] Ondeng O., Ouma H., Akuon P. "A Review of Transformer-Based Approaches for Image Captioning" // Applied Sciences. 2023. Vol. 13(19).
- [10] He K., Zhang X., Ren S., Sun J. "Deep Residual Learning for Image Recognition". arXiv preprint arXiv:1512.03385, 2015.
- [11] Andrew G. Howard, Menglong Zhu, Bo Chen, et al. "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications". arXiv preprint arXiv:1704.04861, 2017.
- [12] Tan M. and Le Q. V. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks". arXiv preprint arXiv:1905.11946, 2020.
- [13] Sabour S., Frosst N., Hinton G. E. "Dynamic Routing Between Capsules". arXiv preprint arXiv:1710.09829, 2017.
- [14] Sander van der Hoog. Deep Learning in Agent-Based Models: A Prospectus. Bielefeld University Working Papers in Economics and Management No. 02-2016, January 2016 ISSN 2196–2723
- [15] Chao-Yuan Wu, R. Manmatha, Alexander J. Smola, Philipp Krähenbühl. "Sampling Matters in Deep Embedding Learning". arXiv preprint arXiv: 1706.07567, 2017.
- [16] Gong Y., Jia Y., et al. "Deep Convolutional Ranking for Multilabel Image Annotation". arXiv preprint arXiv:1312.4894, 2013.
- [17] Deldjoo Y., Schedl M., Hidasi B., Wei Y., He X. "Multimedia Recommender Systems: Algorithms and Challenges" // Ricci F., Rokach L., Shapira B. (eds). Recommender Systems Handbook Springer. New York: NY, 2022.
- [18] Ren S., He K., Girshick R., Sun J. "Faster R-CNN. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". arXiv preprint arXiv:1506.01497, 2015. Nam J., Kim J., Eneldo Loza Mencía, Gurevych I., Fürnkranz J. "Large-scale Multilabel Text Classification - Revisiting Neural Networks". arXiv preprint arXiv: 1312.5419, 2014.
- [19] Xu Y., Mo T., Feng Q., Zhong P., Lai M., Eric I-Chao Chang. "Deep learning of feature representation with multiple instance learning for medical image analysis" // IEEE International Conference on Acoustics, Speech and Signal Processing '05, 2014. Pp. 1626–1630.
- [20] He K., Gkioxari G., Dollár P., Girshick R. "Mask R-CNN". arXiv preprint arXiv: 1703.06870, 2017.
- [21] Chen Y., Li J., Xiao H., Jin X., Yan Sh., Feng J. "Dual Path Networks". arXiv preprint arXiv: 1707.01629, 2017.
- [22] Tan M., Pang R., Le Q. V. "EfficientDet: Scalable and Efficient Object Detection". arXiv preprint arXiv:1911.09070, 2020.
- [23] He K., Zhang X., Ren S., Sun J. "Identity Mappings in Deep Residual Networks". arXiv preprint arXiv:1603.05027, 2016.
- [24] Xie S., Girshick R., Dollár P., et al. "Aggregated Residual Transformations for Deep Neural Networks". arXiv preprint arXiv:1611.05431, 2017.
- [25] Prakash A., Moran N., et al. "Deflecting Adversarial Attacks with Pixel Deflection". arXiv preprint arXiv:1801.08926, 2018.
- [26] Xi E., Bing S., Jin Y. "Capsule Network Performance on Complex Data". arXiv preprint arXiv:1712.03480, 2017.
- [27] Wang Zh., Chen T., Li G., Xu R., Lin L. "Multi-label Image Recognition by Recurrently Discovering Attentional Regions". arXiv preprint arXiv: 1711.02816, 2017.
- [28] Noor Kh. T., Robles-Kelly A., Kusy B. "A Capsule Network for Hierarchical Multi-Label Image Classification". arXiv preprint arXiv:2209.05723, 2022.
- [29] Liu X., Yang H., Ravichandran A., Bhotika R., Soatto St. "Multi-Task Incremental Learning for Object Detection". arXiv preprint arXiv:2002.05347, 2020.
- [30] Girshick R., Donahue J., Darrell T., Malik J. "Rich feature hierarchies for accurate object detection and semantic segmentation". arXiv preprint arXiv:1311.2524, 2014.

- [31] Зиновьев М. С., Нургаянова О. С. Прогнозирование вероятности развития диабетической ретинопатии у пациентов с сахарным диабетом: анализ методов машинного обучения // СИИТ. 2024. Т. 6. № 3(18). С. 95–101. EDN VLFFLP. [[Zinoviev M. S., Nurgayanova O. S. “Predicting the probability of developing diabetic retinopathy in patients with diabetes mellitus: analysis of machine learning methods” // SIIT. 2024. Vol. 6, No. 3(18), pp. 95–101. EDN VLFFLP. (In Russian.)]
- [32] Сулавко А. Е. Высоконадежная биометрическая аутентификация на основе защищенного исполнения нейросетевых моделей и алгоритмов искусственного интеллекта // СИИТ. 2024. Т. 6. № 2(17). С. 11–32. EDN VNIEEV. [[Sulavko A. E. “Highly reliable biometric authentication based on secure execution of neural network models and artificial intelligence algorithms” // SIIT. 2024. Vol. 6, No. 2(17), pp. 11–32. EDN VNIEEV. (In Russian.)]
- [33] Кучкарова Н. В. Оценка актуальных угроз и уязвимостей объектов критической информационной инфраструктуры с использованием технологий интеллектуального анализа текстов // СИИТ. 2024. Т. 6. № 2(17). С. 50–65. EDN NLDWBE. [[Kuchkarova N. V. “Assessment of current threats and vulnerabilities of critical information infrastructure objects using text mining technologies” // SIIT. 2024. Vol. 6, No. 2(17), pp. 50–65. EDN NLDWBE. (In Russian.)]
- [34] Елизарова А. В., Сaitова Г. А., Момзилов Н. В. Выбор архитектуры нейронной сети для прогнозирования состояния заряда аккумулятора // СИИТ. 2023. Т. 5. № 4(13). С. 123–131. EDN RXSDIA. [[Elizarova A. V., Saitova G. A., Momzikov N. V. “Selecting a neural network architecture for predicting the battery charge state” // SIIT. 2023. Vol. 5, No. 4(13), pp. 123–131. EDN RXSDIA. (In Russian.)]
- [35] Шалфеева Е. А. Методология производства жизнеспособных систем доверительного искусственного интеллекта // СИИТ. 2023. Т. 5. № 4(13). С. 28–49. EDN CJTKQH. [[Shalfeeva E. A. “Methodology to produce viable systems of trustworthy artificial intelligence” // SIIT. 2023. Vol. 5, No. 4(13), pp. 28–49. EDN CJTKQH. (In Russian.)]

*Поступила в редакцию 23 ноября 2024 г.*

#### МЕТАДААННЫЕ / METADATA

**Title:** Modern neural network architectures for image tagging and annotating: achievements, challenges, and prospects.

**Abstract:** This paper explores the growing role of image tagging in the context of the rapid increase in visual data volume observed in such areas as social networks, medicine, and security. Tagging, including hashtags and automatic labeling, is becoming an essential tool for systematization, search, and filtering of content, which significantly improves user experience, facilitates access to information, and allows for the efficient processing of large data sets. The scientific novelty of the work lies in the comprehensive analysis of modern neural network architectures used for image recognition and annotation. The study considers not only traditional methods, such as convolutional neural networks (CNN) and residual networks (ResNet), but also new hybrid models and transformer-based approaches that demonstrate high efficiency in solving problems of automatic image tagging and annotation. Particular attention is paid to the analysis of the accuracy of these models, their advantages and limitations in the context of image processing. The article also discusses the current challenges and risks associated with the use of neural networks for image annotation, including the lack of labeled data, low interpretability of models, generalizability issues, as well as ethical issues such as algorithm bias, privacy and data protection. The conclusion draws a conclusion on the importance of neural networks for the development of image recognition technologies and emphasizes the need for further research to improve the accuracy, efficiency and ethics of existing approaches in this area, which opens up new horizons for the application of technologies in various industries, including medicine, the automotive industry and others.

**Key words:** neural networks; tagging; classification; neural network architecture; machine learning

**Язык статьи / Language:** русский / Russian.

#### Об авторах / About the authors:

##### РЕЗНИКОВ Григорий Анатольевич

МИРЭА – Российский технологический университет, Россия.  
Студ. института информационных технологий по направлению «Программная инженерия».  
E-mail: kave7771@outlook.com  
ORCID: <https://orcid.org/0009-0009-0461-5831>

##### REZNIKAU Ryhor Anatolievich

MIREA – Russian Technological University, Russia.  
Student of the Institute of Information Technology in the field of software engineering.  
E-mail: kave7771@outlook.com  
ORCID: <https://orcid.org/0009-0009-0461-5831>

##### СЕНИЦЫН Роман Дмитриевич

МИРЭА – Российский технологический университет, Россия.  
Студ. института информационных технологий по направлению «Программная инженерия».  
E-mail: sinicnr2003@mail.ru  
ORCID: <https://orcid.org/0009-0003-6750-5222>

##### SINITSYN Roman Dmitrievich

MIREA – Russian Technological University, Russia.  
Student of the Institute of Information Technology in the field of software engineering.  
E-mail: sinicnr2003@mail.ru  
ORCID: <https://orcid.org/0009-0003-6750-5222>

##### ШУЛИК Анатолий Михайлович

МИРЭА – Российский технологический университет, Россия.  
Студ. института информационных технологий по направлению «Программная инженерия».  
E-mail: shulik.2003@mail.ru  
ORCID: <https://orcid.org/0009-0006-4258-0666>

##### SHULIK Anatoly Mihailovich

MIREA - Russian Technological University, Russia.  
Student of the Institute of Information Technology in the field of software engineering.  
E-mail: shulik.2003@mail.ru  
ORCID: <https://orcid.org/0009-0006-4258-0666>