

Проблема обнаружения лиц в видеопотоке: обзор технологий

А. Г. Курдюков¹, А.В. Коваленко¹, Д.М. Теунаев², Ф.М. Узденова²

¹Кубанский государственный университет, Краснодар

²Северо-Кавказская государственная академия

Аннотация: В данной статье представлено исследование использования нейронных сетей (НС) как междисциплинарной модели для классификации объектов на видео, в том числе, для решения задачи поиска лиц. Это подчеркивает универсальность НС в интеграции обученных данных и точной классификации объектов, что критически важно для обеспечения безопасности и эффективности видеонаблюдения.

В исследовании используется анализ различных архитектур нейронных сетей, а также исследование алгоритмов их работы. Данные, полученные из обзора литературы и результатов экспериментов, позволяют оценить эффективность решения задачи классификации объектов на видео с помощью различных архитектур, не привязывая исследование к конкретному набору данных.

Активное применение нейронных сетей в качестве инструмента для видеонаблюдения повышает безопасность объектов инфраструктуры и эффективность работы служб безопасности.

Ключевые слова: нейронные сети, архитектуры нейронных сетей, системы видеонаблюдения, распознавание в реальном времени, повышение безопасности, благополучие общества.

Введение

В современном мире использование нейронных сетей для распознавания объектов на видео становится все более актуальным и значимым не только в контексте улучшения точности и скорости идентификации объектов, но и для повышения эффективности и обеспечения безопасности с помощью видеонаблюдения. С развитием технологий видеонаблюдения и компьютерного зрения появляются новые потребности, которые можно эффективно удовлетворить с помощью нейронных сетей.

Одной из важных потребностей, которую помогают решать нейронные сети, является автоматизация процесса анализа видеоданных. Способность нейронных сетей к распознаванию объектов, классификации действий и обнаружению аномалий на видео, позволяет значительно упростить и ускорить процесс мониторинга больших объемов видеозаписей. Это особенно важно для систем видеонаблюдения в общественных местах, в

транспорте, на предприятиях и других сферах, где требуется постоянный контроль.

Еще одной потребностью, которую решают нейронные сети, является повышение уровня безопасности. Благодаря возможности быстрого и точного распознавания объектов на видео, нейронные сети могут помочь в раннем обнаружении угроз и нежелательных событий, таких, как проникновение на охраняемую территорию, опасное поведение людей или транспортных средств, а также другие потенциально опасные ситуации.

Авторы статьи [1] представляет обзор современных методов глубокого обучения для видеонаблюдения и подчеркивает значимость использования нейронных сетей для повышения эффективности и безопасности систем видеонаблюдения. Также в статье [2] авторы обсуждают эволюцию систем видеонаблюдения и роль нейронных сетей в улучшении их функциональности.

Таким образом, использование нейронных сетей для распознавания объектов на видео не только способствует улучшению точности и скорости анализа видеоданных, но также отвечает на потребности в повышении эффективности и безопасности систем видеонаблюдения в современном мире.

Данная статья посвящена исследованию различных архитектур нейронных сетей для распознавания объектов на видео. Анализируются современные подходы к созданию нейросетей, их применение в задачах компьютерного зрения и обнаружения объектов на видео.

Целью данного исследования является выявление оптимальных архитектур нейронных сетей, способных эффективно обрабатывать видеоданные и проводить точную классификацию объектов. Результаты работы могут быть полезны для улучшения систем видеонаблюдения,

повышения безопасности и эффективности процессов распознавания объектов на видео.

Исследование архитектур нейронных сетей

Выделим основные инструменты, позволяющие качественно осуществить процесс распознавания лица в видеопотоке. В языке программирования Python наиболее популярной библиотекой компьютерного зрения является библиотека OpenCV, предоставляющая широкий набор функций обработки изображений и видеопотока. Центральным ядром можно считать класс `cv2.CascadeClassifier`, загружающий предварительно обученную модель распознавания лиц и применяющий ее в видеопотоке. Библиотека Dlib основана на алгоритме обнаружения лиц, использующего дескриптор функции гистограммы ориентированных градиентов. Эта библиотека также использует предобученную модель распознавания лиц, центральной функцией которой является функция `dlib.get_frontal_face_detector ()`. Глубокое обучение для распознавания лиц является основой функций библиотеки `face_recognition` [3], представляющей простой API для обнаружения нескольких лиц в одном кадре. Крупная библиотека TensorFlow глубокого обучения использует предварительно обученные модели: Single Shot MultiBox Detector (SSD) и Faster R-CNN. Среда PyTorch, реализующая схожий функционал основана на моделях RetinaFace и EfficientDet.

Разработчики в области человеко-машинных интерфейсов предлагают различные инновационные решения. Анализ подобных решений позволил выявить наиболее удачные. Так, авторы статьи [4] предлагают использовать систему обнаружения и повторной идентификации лиц с использованием глубокого обучения и библиотеки OpenVINO. В результате детектор достигает самой современной производительности на нескольких наборах

эталонных данных по сравнению с другими моделями, оптимизированных для использования на базе ЦП. При этом эффективность позволяет обрабатывать видео со скоростью 53 кадра в секунду.

Анализ научных исследований современных разработчиков обнаружения различных объектов в видеопотоке и на фото, в том числе и лиц, позволил определить наиболее используемые архитектуры нейронных сетей.

Сверточные нейронные сети (CNN) эффективны для извлечения признаков из изображений и видео. Они могут быть использованы для обнаружения объектов, таких как лица, автомобили, животные и т.д. Основная идея свёрточных нейронных сетей заключается в чередовании свёрточных слоёв и субдискретизирующих слоёв. Структура сети — однонаправленная (без обратных связей), принципиально многослойная. Важно отметить, что свёрточные нейронные сети используют некоторые особенности зрительной коры, в которой были открыты так называемые простые клетки, реагирующие на прямые линии под разными углами, и сложные клетки, реакция которых связана с активацией определённого набора простых клеток. Так, исследователи Викулова А.А., Гришина Л.С., Болодурин И.П., предлагают использование архитектуры CNN для классификации снимков рентгенограмм с Covid-19. Авторы, проведя серию экспериментов, остановились на архитектуре CNN на основе f1-меры - наиболее точной метрики оценки качества классификатора, выбрана модель Net1 с пятью слоями свёртки, показавшая результаты с точностью 98 % [5]. Ермоленко Т. В., Самородский И. Е. выполнили анализ использования различных архитектур глубоких нейросетей в задаче классификации товаров по их изображениям. Для решения проблем несбалансированных обучающих данных и похожести объектов разных классов предложено использовать балансировку классов и архитектуру API-Net [6]. В статье [7] на основе

нейросети архитектуры CNN реализован метод Виолы-Джонсона, при обучении которого функция потерь стала равна 0,1366, что является показателем качественного распознавания.

Другим эффективным решением проблемы обнаружения лиц в видеопотоке являются региональные сверточные нейронные сети (R-CNN), эти нейронные сети могут обнаруживать объекты и создавать ограничивающие рамки вокруг них. Архитектура R-CNN (Region-Based Convolutional Neural Network) разработана в 2014 году Россом Гиршиком и его коллегами [8, 9]. Она предназначена для обнаружения объектов на изображениях.

Основной алгоритм R-CNN заключается в следующем.

- Нахождение потенциальных объектов на изображении и разбиение их на регионы с помощью метода selective search.
- Применение сверточной нейронной сети к каждому предложенному региону для извлечения признаков.
- Классификация каждого региона с использованием классификаторов, обученных на извлеченных признаках.

R-CNN является качественным прорывом в области обнаружения объектов, но он имеет некоторые недостатки, включая высокую вычислительную сложность и медленное время обучения и тестирования. Эти проблемы были улучшены в последующих итерациях, таких как Fast R-CNN и Faster R-CNN. Важно отметить, что R-CNN и его вариации являются основой для многих современных систем обнаружения объектов.

Третьей архитектурой, способной обнаруживать объекты в реальном времени на изображениях и видео можно отметить Single Shot Multibox Detector (SSD). SSD (Single Shot MultiBox Detector) представлена в 2016 году и с тех пор стала одной из самых популярных архитектур для обнаружения объектов [10]. Основное преимущество SSD заключается в том, что она

может обнаруживать объекты различных масштабов благодаря использованию сверточных слоев различных размеров. Это делает SSD более эффективной по сравнению с другими методами, такими как R-CNN, которые требуют нескольких этапов обработки для обнаружения объектов различных размеров. SSD использует метод, известный как non-maximum suppression, для уменьшения количества ложных срабатываний, что позволяет SSD работать быстрее и точнее. Однако, несмотря на преимущества, SSD имеет некоторые недостатки, например, возникают проблемы с обнаружением очень маленьких объектов. Кроме того, SSD требует большого количества вычислительных ресурсов, что может быть проблемой для устройств с ограниченной вычислительной мощностью.

Архитектура SSD применена в исследовании Чен Ченга для обнаружения масок в период пандемии [11]. Авторы статьи используют SSD (Single Shot MultiBox Detector) в качестве локатора и классификатора масок. Используется сеть MobileNetV2 для извлечения признаков изображения и снижения количества параметров. Благодаря этому система может быть развернута на мобильных устройствах. В исследовании представлены методы обучения, которые используются для переноса предварительно обученных моделей из других областей в модель SSD-MobileNetV2. Методы увеличения данных, такие, как MixUp, эффективно предотвращают переобучение и снижают зависимость от больших наборов данных. Эксперименты в практических сценариях показали, что предложенная система хорошо справляется с обнаружением масок в реальном времени.

Последняя архитектура, предназначенная для обнаружения объектов в реальном времени на изображениях и видео - You Only Look Once (YOLO). Это метод идентификации и распознавания объектов на изображениях в реальном времени, предложенный Джозефом Редмондом и Али Резой в 2015 году на конференции IEEE/CVF по компьютерному зрению и распознаванию

образов (CVPR). В отличие от предыдущих методов идентификации объектов, которые использовали классификаторы для обнаружения, YOLO использует сквозную нейронную сеть, которая одновременно предсказывает ограничивающие рамки и вероятности классов. Метод YOLO делит изображение на N сеток, каждая из которых имеет сектор одинакового размера $S \times S$. Каждая из этих N сеток отвечает за обнаружение и определение местоположения объекта, который она содержит. YOLO значительно сокращает вычисления, поскольку и обнаружение, и распознавание обрабатываются ячейками из изображения. Однако, этот метод производит много повторяющихся прогнозов. Чтобы решить эту проблему, YOLO использует немаксимальное подавление. Существуют различные версии YOLO, включая YOLOv8, каждая из которых предлагает улучшения в скорости и точности.

В современных исследованиях архитектура YOLO является достаточно популярной. Х. Тервен, Д. Кордова-Эспарза выполняют всесторонний анализ эволюции YOLO, рассматривают инновации и преимущества каждой итерации, от оригинального YOLO до YOLOv8, YOLO-NAS и YOLO с трансформаторами, обсуждаются основные изменения в сетевой архитектуре и приемы обучения для каждой модели [12]. Т. Диван, Г. Анирудх, Джитендра В. В своих исследованиях предлагают всесторонний обзор одноступенчатых детекторов объектов, особенно YOLO, их архитектурных достижений и статистики производительности. В исследовании М. Махешвари, М. С. Жозефина, В. Джеябалараджа [13] предлагается использование сверточной нейронной сети (CNN), которая снижает вычислительные потребности для задач обнаружения объектов в видео. CNN использует улучшенную платформу YOLO для классификации и обнаружения объектов, создавая новую архитектуру CNN. Предложенная модель обеспечивает точность классификации объектов в 96,7%.

Таким образом, выделенные пять основных архитектур являются базой любого эффективного решения задачи обнаружения лиц, а перечисленные инструменты языка программирования Python способны качественно выполнить реализацию предложенных архитектур. Посредством эксперимента возможно улучшение качества обнаружения лиц, основным вектором которого может быть модификация и интеграция выделенных базовых нейросетевых архитектур.

Литература

1. Duong H-T, Le V-T, Hoang VT. Deep Learning-Based Anomaly Detection in Video Surveillance: A Survey. 23(11):5024. URL: doi.org/10.3390/s23115024 (Дата обращения: 03.03.2024).
2. Porikli, Fatih & Bremond, Francois & Dockstader, Shiloh & Ferryman, James & Hoogs, Anthony & Lovell, Brian & Pankanti, Sharath & Rinner, Bernhard & Tu, Peter & Venetianer, Peter. Video Surveillance: Past, Present, and Now the Future [DSP Forum]. IEEE signal processing magazine, 2013, 30. Pp. 190-198. 10.1109/MSP.2013.2241312.
3. Официальный репозиторий разработчиков библиотеки. URL: github.com/ageitgey/face_recognition (Дата обращения: 10.01.2024).
4. Muhamad D. P., Laksono K., Kang-Hyun J. High Performance and Efficient Real-time Face Detector on CPU Based on Convolutional Neural Network URL: researchgate.net/publication/346486312_High_Performance_and_Efficient_Real-time_Face_Detector_on_CPU_Based_on_Convolutional_Neural_Network (Дата обращения 10.01.2024).
5. Викулова А. А., Гришина Л. С., Болодурин И. П. Разработка эффективной архитектуры CNN для классификации снимков рентгенограмм с Covid-19 // Шаг в науку. 2022. №4. URL: cyberleninka.ru/article/n/razrabotka-

effektivnoy-arhitektury-cnn-dlya-klassifikatsii-snimkov-rentgenogramm-s-covid-19 (Дата обращения: 12.01.2024).

6. Ермоленко Т. В., Самородский И. Е. Анализ эффективности архитектур глубоких нейросетей для классификации изображений товаров // Проблемы искусственного интеллекта. 2022. №1 (24). URL: cyberleninka.ru/article/n/analiz-effektivnosti-arhitektur-glubokih-neyrosetey-dlya-klassifikatsii-izobrazheniy-tovarov (Дата обращения: 12.01.2024).

7. Д.Ю. Романюта, А.В. Коваленко, М.В. Шарпан Исследование алгоритма Виолы-Джонса для разработки системы распознавания лиц с помощью нейронных сетей // Инженерный вестник Дона, 2024, №1. URL: ivdon.ru/uploads/article/pdf/IVD_6__1y24_Romanyuta_Kovalenko_Sharpan.pdf_08e8b88e0c.pdf.

8. Komissarenko N. 3 метода детектирования объектов с Deep Learning: R-CNN, Fast R-CNN и Faster R-CNN. URL: medium.com/@bigdataschool/3-%D0%BC%D0%B5%D1%82%D0%BE%D0%B4%D0%B0%D0%B4%D0%B5%D1%82%D0%B5%D0%BA%D1%82%D0%B8%D1%80%D0%BE%D0%B2%0%B0%D0%BD%D0%B8%D1%8F%D0%BE%D0%B1%D1%8A%D0%B5%D0%BA%D1%82%D0%BE%D0%B2-c-deep-learning-r-cnn-fast-r-cnn-%D0%B8-faster-r-cnn-acdf6380fd33 (Дата обращения: 12.01.2024).

9. Mask R-CNN: архитектура современной нейронной сети для сегментации объектов на изображениях. URL: habr.com/ru/articles/421299/ (Дата обращения: 14.01.2024).

10. Нейросети. Куда это все движется. URL: habr.com/ru/articles/482794/ (Дата обращения: 14.01.2024).

11. Cheng C. Real-Time Mask Detection Based on SSD-MobileNetV2. URL: arxiv.org/abs/2208.13333 (Дата обращения: 15.01.2024).

12. Terven J., Cordova-Esparza D. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS URL: arxiv.org/abs/2304.00501 (Дата обращения: 17.01.2024).

13. Maheswari M., Josephine M. S., Jeyabalaraja V. YOLO Architecture-based Object Detection for Optimizing Performance in Video Streams. URL: ijettjournal.org/Volume-70/Issue-11/IJETT-V70I11P220.pdf (Дата обращения: 17.01.2024).

References

1. Duong H-T, Le V-T, Hoang VT. Deep Learning-Based Anomaly Detection in Video Surveillance: A Survey. *Sensors*. 23(11):5024. URL: doi.org/10.3390/s23115024 (Date accessed: 03.03.2024).

2. Porikli, Fatih & Bremond, Francois & Dockstader, Shiloh & Ferryman, James & Hoogs, Anthony & Lovell, Brian & Pankanti, Sharath & Rinner, Bernhard & Tu, Peter & Venetianer, Peter. Video Surveillance: Past, Present, and Now the Future [DSP Forum]. *IEEE Signal Processing Magazine*, 2013. 30. Pp 190-198. [10.1109/MSP.2013.2241312](https://doi.org/10.1109/MSP.2013.2241312).

3. Oficial`ny`j repozitorij razrabotchikov biblioteki [Official repository of the library developers]. URL: github.com/ageitgey/face_recognition (Date accessed 10.01.2024).

4. Muhamad D. P., Laksono K., Kang-Hyun J. High Performance and Efficient Real-time Face Detector on CPU Based on Convolutional Neural Network. URL: researchgate.net/publication/346486312_High_Performance_and_Efficient_Real-time_Face_Detector_on_CPU_Based_on_Convolutional_Neural_Network (Date accessed: 10.01.2024).

5. Vikulova A. A., Grishina L. S., Bolodurina I. P. Shag v nauku. 2022. №4. URL: cyberleninka.ru/article/n/razrabotka-effektivnoy-arhitektury-cnn-dlya-klassifikatsii-snimkov-rentgenogramm-s-covid-19 (Date accessed: 12.01.2024).
 6. Ermolenko T. V., Samorodskij I. E. 2022. №1 (24). URL: cyberleninka.ru/article/n/analiz-effektivnosti-arhitektur-glubokih-neyrosetey-dlya-klassifikatsii-izobrazheniy-tovarov (Date accessed: 12.01.2024).
 7. D.Yu. Romanyuta, A.V. Kovalenko, M.V. Sharpan Inzhenernyj vestnik Dona, 2024, No. 1. URL: ivdon.ru/uploads/article/pdf/IVD_6__1y24_Romanyuta_Kovalenko_Sharpan.pdf_08e8b88e0c.pdf.
 8. Komissarenko N. 3 metoda detektirovaniya ob`ektov s Deep Learning: R-CNN, Fast R-CNN и Faster R-CNN [3 methods of object detection using Deep Learning: R-CNN, Fast R-CNN and Faster R-CNN]. URL: medium.com/@bigdataschool/3%D0%BC%D0%B5%D1%82%D0%BE%D0%B4%D0%B0%D0%B4%D0%B5%D1%82%D0%B5%D0%BA%D1%82%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D1%8F%D0%BE%D0%B1%D1%8A%D0%B5%D0%BA%D1%82%D0%BE%D0%B2-c-deelearning-r-cnn-fast-r-cnn-%D0%B8-faster-r-cnn-acdf6380fd33 (Date accessed: 12.01.2024).
 9. Mask R-CNN: arxitektura sovremennoj nejronnoj seti dlya segmentacii ob`ektov na izobrazheniyax [Mask R-CNN: modern neural network architecture for object segmentation in images]. URL: habr.com/ru/articles/421299/ (Date accessed: 14.01.2024).
 10. Nejroseti. Kuda e`to vse dvizhetsya [Neural networks. Where is this all going?]. URL: habr.com/ru/articles/482794/ (Date accessed: 14.01.2024).
 - 11 Cheng C. Real-Time Mask Detection Based on SSD-MobileNetV2. URL: arxiv.org/abs/2208.13333 (Date accessed: 15.01.2024).
-



12. Terven J., Cordova-Esparza D. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. URL: arxiv.org/abs/2304.00501 (Date accessed: 17.01.2024).

13. Maheswari M., Josephine M. S., Jeyabalaraja V. YOLO Architecture-based Object Detection for Optimizing Performance in Video Streams. URL: ijettjournal.org/Volume-70/Issue-11/IJETT-V70I11P220.pdf (Date accessed: 17.01.2024).

Дата поступления: 13.03.2024

Дата публикации: 18.04.2024