

*Дели Андрей Георгиевич, студент магистратуры 2 курс, МИРЭА- Российский технологический университет (РТУ МИРЭА), Россия, г. Москва, Институт комплексной безопасности и специального приборостроения, Россия, г. Москва*

## **КЛАССИФИКАЦИЯ ВИДЕО КОНТЕНТА С ПОМОЩЬЮ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ**

**Аннотация:** В данной статье сравниваются алгоритмы классификации видео контента на основе сверточных нейронных сетей, с использованием предобученных моделей ResNet50, VGG16. Цель данной работы –разработать программное обеспечение для классификации запрещенного видео контента на основе сверточных нейронных сетей, при этом рассмотреть проблему энергоэффективности. Данное исследование показало, что наилучший результат демонстрирует модель на основе сети VGG16, вероятность верного распознавания на тестовом наборе при использовании этой модели составила 88,95%. Также рассмотрены алгоритмы предобработки данных и метод переноса обучения.

**Ключевые слова:** классификация, сверточные нейронные сети, глубокое обучение.

**Abstract:** This article compares the algorithms for classifying video content based on convolutional neural networks, using pretrained models ResNet50, VGG16. The purpose of this work is to develop software for the classification of prohibited video content based on convolutional neural networks, while considering the problem of energy efficiency. This study showed that the best result is demonstrated by a model based on the InceptionV3 network, the probability of correct recognition the test set when using this model was 91.51%. The algorithms for pre-processing data and the method of transferring training are also considered.

**Keywords:** classification, convolutional neural networks, deep learning.

## **Введение**

Проблема фильтрации видео контента в Интернете в настоящее время является действительно актуальной. Включить какое-либо видео в современном мире не составляет труда. Трудно встретить человека, который бы не имел выход в глобальную сеть. Кабельное телевидение в настоящий момент уступает иным видео ресурсам. Для этого есть множество причин. Просто лишь имея смартфон и доступ в Интернет люди могут посмотреть видео из любой точки мира.

Одним из быстрорастущих сегментов данных в сети является видеoinформация, доступ к которой может получить любой пользователь. Информация делится на полезную или же неприемлемую. Существует контент, доступ к которому запрещён или должен быть запрещён на территории Российской Федерации.

Целью данного исследования является повышение безопасности пользователей сети Интернет за счет ограничения доступа к запрещенному контенту. Следует разработать программное средство, которое бы соответствовало законодательству РФ, позволило бы автоматизировать процесс фильтрации контента. Так как человеческие ресурсы не позволяют за короткое время просмотреть и найти необходимую информацию в большом объеме видеоматериалов, автоматизация решения данной задачи позволяет существенно снижать временные и трудовые затраты при анализе видеофайлов. Так же снижается влияние человеческого фактора.

Очевидным решением для данной задачи является машинное обучение. В данном исследовании я предлагаю систему глубокого обучения, на основе сверточной нейронной сети, которая автоматически анализирует видеокadres, после чего классифицирует контент как запрещенный или доступный.

На рынке существуют подобные глобальные программы:

- 1) Computer Vision [1] - система анализа контента в изображениях и видео от Microsoft. Позволяет производить визуальную обработку данных, чтобы

маркировать контент объектами и концепциями, извлекать текст, создавать описания изображений, модерировать контент и понимать движения людей в физических пространствах.

2) Американская небезызвестная компания Facebook разработала нейронную сеть Rosetta [1] для распознавания текста на изображениях и видео. Разработка позволяет определять качественный контент и спам, а также умеет находить агрессивный и оскорбительный контент. Нейросеть помечает контент как запрещенный, после чего модераторы в ручном порядке проверяют данные результаты.

3) Масштабная организация Amazon разработала сервис распознавания видеоматериалов - Rekognition Video [2], одной из задач которого является выявление неприемлемого содержимого в видеозаписях.

Все вышеперечисленные системы являются разработками крупномасштабных организаций и показывают по-настоящему действенный результат, однако, они также имеют проблемы. Одна из которых -использование большого количества ресурсов для обучения моделей на крупных наборах данных. Развитие технологий не останавливается, а только набирает обороты. Это влечет за собой новые растраты памяти, энергии и денежных средств. В связи с этой проблемой, предлагается разработать программное обеспечение на основе сверточных нейронных сетей с учетом проблемы энергоэффективности. В этой статье будет проведено исследование обычных нейронных сетей с переносом обучения.

### **Предобработка данных**

Для обучения нейронной сети был модифицирован уже готовый набор данных NSFW (Not Safe For Work), взятый с сайта Kaggle. Эта аббревиатура обозначает материал, который содержит неприемлемый контент откровенного характера. Есть трудность определения такого содержания в различных материалах. Изначальный набор слишком массивный, изображения были хаотичны и не имели необходимых лейблов. В связи с этим были проведены работы по подготовке данных в несколько этапов:

1. Исходные изображения были слишком велики, поэтому был разработан скрипт, который уменьшал размер изображений за счет снижения качества картинки.

2. Так как изображения не имели лейблов, определяющих их к тому или иному классу, также была разработана программа, которая переименовала все изображения как «nsfw» или «sfw». Эти наименования соответствуют двум классам: «nsfw» - класс, где содержатся материалы откровенного характера, они должны быть помечены программой как запрещенные и «sfw» - класс, где представлены совершенно разнообразные обычные изображения.

3. Распределение данных на три каталога: «train» - выборка для обучения, «test» - тестовая выборка, «val» - валидационная выборка.

На рисунке 1 представлены примеры из класса «sfw».



Рисунок 1. Примеры изображений из класса «sfw»

### **Архитектура нейросетевой модели**

Понятие глубокое обучение (Deep Learning) подразумевает под собой область машинного обучения, которая рассматривает методы решения задач искусственного интеллекта с использованием глубоких нейронных сетей.

Однозначного определения, что такое глубокая нейронная сеть, не

существует, однако, в данной работе под глубокой нейронной сетью будет пониматься такая нейронная сеть, которая содержит более одного скрытого слоя.

Сверточная нейронная сеть является особой архитектурой искусственной нейронной сети, имитирующая особенности зрительной коры головного мозга, состоящая из нескольких многомерных слоев и предназначенная для эффективного распознавания сложных изображений.

Именно такие сети сейчас преимущественно используются на практике для обработки изображений и видео. В данном исследовании был применен метод переноса обучения – использование предварительно обученной сети для решения задачи, которая отличается от той, для которой она обучалась. У уже предобученной сети берется только сверточная часть, отвечающая за выделение характерных признаков у изображения. Часть, отвечающую за классификацию, отключаем, так как она определяет тысячу объектов разных типов, а нам необходимо только два. Поэтому мы берем сверточную часть и добавляем к ней собственный классификатор. Затем обучаем такую составную нейронную сеть на заранее подобранном наборе данных.

Для достижения наилучшего результата был проведен анализ трех сетей: VGG16, ResNet50. Для представленных сетей были заданы одинаковые параметры: 5 эпох обучения,  $lr=1e-5$ , количество изображений для обучения = 17500, количество изображений для проверки = 3750, количество изображений для тестирования = 3750. Веса были получены путем обучения представленных моделей на базе данных ImageNet [4]. В ней представлено более 14 миллионов объектов и более 21 тысячи классов.

### **Экспериментальная часть**

1) Сеть VGG16 как показывает опыт, демонстрирует хороший результат. Архитектура сети представлена на рисунке 2. Часть, отвечающая за классификацию, выделена красным прямоугольником. Архитектура заканчивается полносвязным слоем на тысячу нейронов, что определяет тысячу классов в то время, как необходимо два. Для этого вместо старого слоя с тысячей нейронов добавляем полносвязный слой, на выходе у которого будет только

один нейрон и сигмоидальная функция активации, эта функция плавно меняет своё значение от 0 до 1 и хорошо подходит для решения задач бинарной классификации. По такому же принципу будет произведена замена части классификации при использовании других моделей.

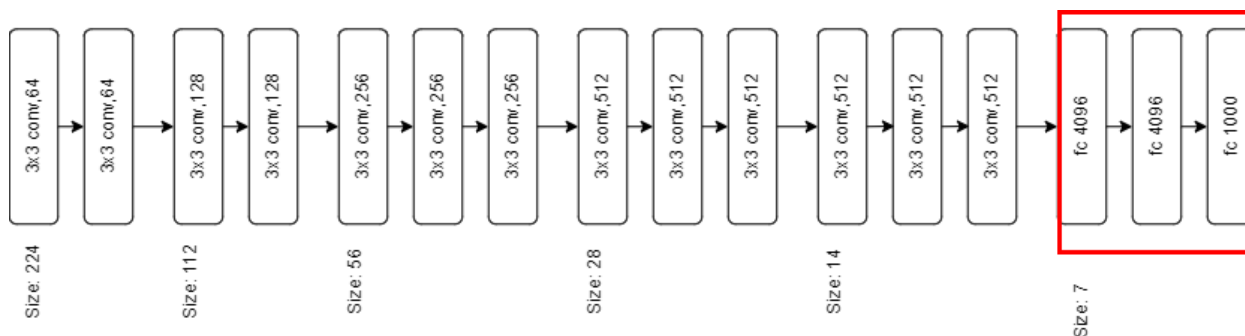


Рисунок 2. Архитектура сети VGG16

Аккуратность на тестовых данных для двух классов составила 88,95%. Время, затраченное на обучение модели – 8 часов 34 минуты в облачной среде для работы с Python – Google Colab. Данный сервер отлично подходит для разработки и тестирования моделей машинного обучения. На рисунке 3 представлены графики аккуратности и ошибки на обучающем и проверочном наборе данных.

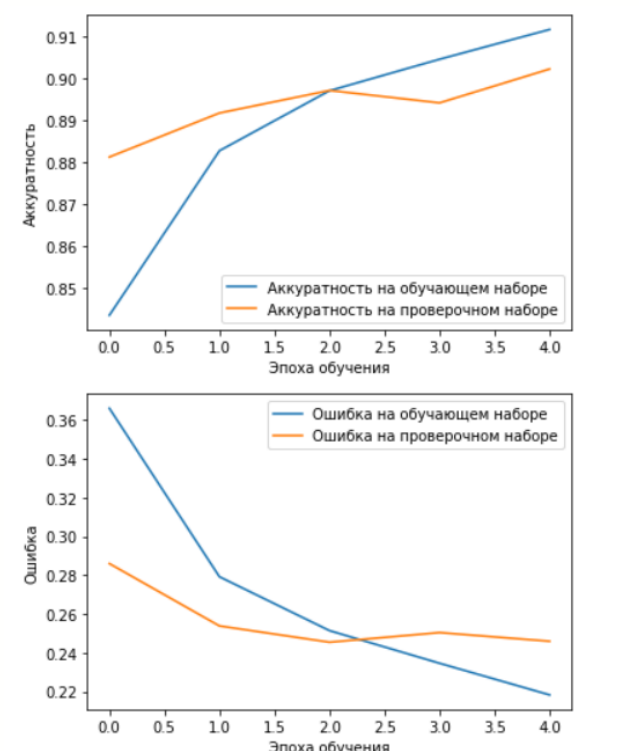


Рисунок 3. График аккуратности вероятности правильного распознавания и вероятность ошибки для составной сети с использованием VGG16

2) Сеть ResNet (Residual Network) была впервые представлена миру в 2015 году и с этого момента пользуется большой популярностью. С ростом количества слоев сети точность сначала увеличивается, а затем быстро ухудшается. Причина этого явления связана с эффектом затухания градиента. Для решения этой проблемы и были выведены остаточные сети. Ключевая идея, лежащая за ResNet – это введение остаточных блоков, которые содержат "обходную связь идентичности", обходящую один или большее количество слоев. ResNet имеет множество вариантов, которые работают по одной и той же концепции, но имеют разное количество слоев. ResNet50 используется для обозначения варианта, который может работать с 50 уровнями нейронной сети. На рисунке 4 представлена отличительная черта архитектуры – концепция пропускного соединения, где добавляется исходный ввод к полученному выводу.

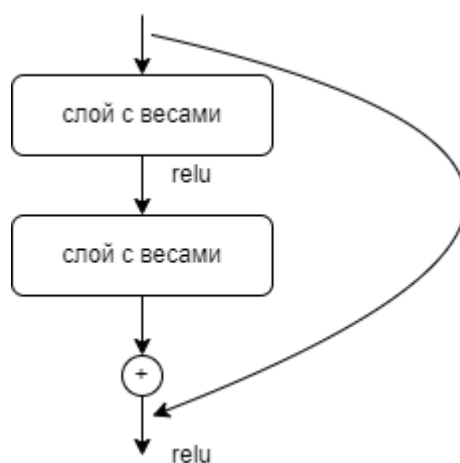


Рисунок 4. Соединение быстрого доступа

Аккуратность на тестовых данных составила 68,27%. Время обучения – 3 часа 27 минут в Google Colab. Графики, полученные в результате обучения представлены на рисунке 5. Аналогично с VGG16 встроенные слои классификации были заменены на собственные слои.

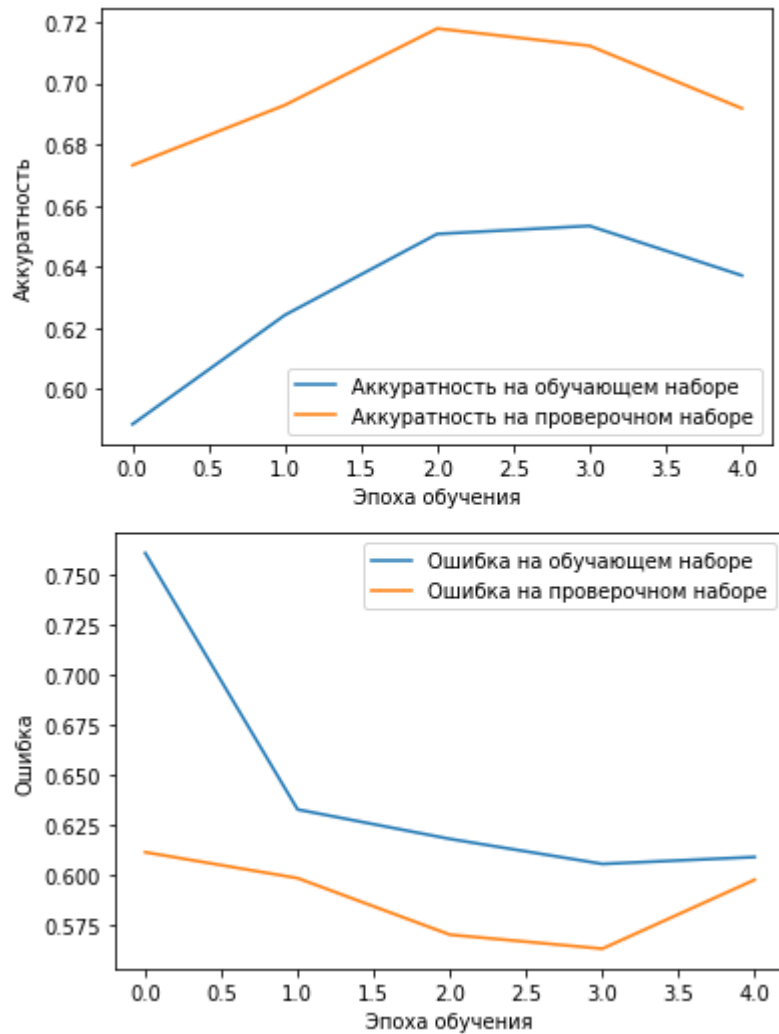


Рисунок 5 – График аккуратности вероятности правильного распознавания и вероятность ошибки для составной сети с использованием ResNet50

Исходя из проведенного анализа были получены следующие результаты (таблица 1).

Таблица 1. Результаты обучения и тестирования сверточной нейронной сети

Наименование модели	Количество эпох, шт	Количество параметров сети, млн	Алгоритм оптимизации	Аккуратность на тестовых данных, %	Время обучения, часы
VGG16	5	138	Adam	88,95	8
ResNet50	5	25	Adam	68,27	3



## **Заключение**

В представленной статье был проведен анализ различных моделей для включения их в составную сверточную нейронную сеть, которая будет использоваться для классификации запрещенного графического контента в сети Интернет. Анализ экспериментальной работы, проведенный с использованием различных тестов, показал, что наиболее удачной моделью является Inception-V3, так как именно эта модель показала высший результат аккуратности на тестовом наборе данных и наименьшее время обучения сети. В дальнейшем плане развития исследования будет рассматриваться автоматический подбор гиперпараметров для улучшения эффективности работы сети, решение проблемы энергоэффективности и встраивание модели в программный модуль для однозначной классификации видео контента.

## **Библиографический список:**

1. Rosetta: Understanding text in images and videos with machine learning, Режим доступа: свободный [Электронный ресурс] – URL - <https://ai.facebook.com/blog/rosetta-understanding-text-in-images-and-videos-with-machine-learning/> (Дата обращения: 04.11.2021).
2. Вопросы и ответы по Amazon Rekognition, Режим доступа: свободный [Электронный ресурс] – URL - <https://aws.amazon.com/ru/rekognition/faqs/> (Дата обращения: 08.11.2021).
3. Computer Vision, Режим доступа: свободный [Электронный ресурс] – URL - <https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/#overview> (Дата обращения: 09.11.2021).
4. ImageNet, Режим доступа: свободный [Электронный ресурс] – URL - <https://image-net.org/> (Дата обращения: 09.11.2021).
5. Wikipedia. Convolutional neural network, Режим доступа: свободный [Электронный ресурс] – URL - [https://en.wikipedia.org/wiki/Convolutional\\_neural\\_network](https://en.wikipedia.org/wiki/Convolutional_neural_network) (Дата обращения: 15.11.2021).

6. ResNet (34, 50, 101): «остаточные» CNN для классификации изображений, Режим доступа: свободный [Электронный ресурс] – URL - <https://neurohive.io/ru/vidy-nejrosetej/resnet-34-50-101/> (Дата обращения: 23.11.2021).

7. Исследуем архитектуры сверточных нейронных сетей, Режим доступа: свободный [Электронный ресурс] – URL - <https://proglib.io/p/issleduem-arhitektury-svertochnyh-neyronnyh-setey-s-pomoshchyu-fast-ai-2020-12-28> (Дата обращения: 30.11.2021).

8. LeCun Y., Kavukvuoglu K., Farabet C. // Proc. Int. Symposium on Circuits and Systems. 2010. P 253–256.