

## Применение сверточных нейронных сетей для обнаружения и распознавания изображений на основе самописного генератора

М. Н. Б. Муаль\*, Д. В. Козырев

ФГАОУ ВО «Российский университет дружбы народов», г. Москва, Российская Федерация

Адрес: 117198, Российская Федерация, г. Москва, ул. Миклухо-Маклая, д. 6

\* bmouale@mail.ru

### Аннотация

Распознавание объектов является ветвью искусственного зрения и одним из столпов машинного зрения. Оно состоит в идентификации форм, заранее описанных в цифровом изображении и, в общем случае, в цифровом видеопотоке. Хотя, как правило, можно выполнять распознавание по видеофрагментам, процесс обучения обычно выполняется на изображениях. В данной работе рассматривается алгоритм классификации и распознавания объектов с использованием сверточных нейронных сетей. Цель работы – реализовать алгоритм обнаружения и классификации различных графических объектов, подаваемых с веб-камеры. Задача состоит в том, чтобы сначала классифицировать и распознавать объект с высокой точностью по заданному набору данных, а затем продемонстрировать способ генерации изображений для увеличения объема обучающего набора данных путем самописного генератора. Используемый алгоритм классификации и распознавания является инвариантным к переносу, сдвигу и повороту. Существенной новизной в этой работе является создание самописного генератора, позволяющего применять различные виды аугментации (искусственное увеличение объема обучающей выборки путем модификации обучающих данных) для формирования каждый раз новых групп измененных изображений.

**Ключевые слова:** обнаружение изображений, распознавание изображений, сверточные нейронные сети, модель R-CNN (Regional Convolutional Neural Networks), аугментации, субдискретизация

**Конфликт интересов:** авторы заявляют об отсутствии конфликта интересов.

**Для цитирования:** Муаль М. Н. Б., Козырев Д. В. Применение сверточных нейронных сетей для обнаружения и распознавания изображений на основе самописного генератора // Современные информационные технологии и ИТ-образование. 2022. Т. 18, № 3. С. 507-515. doi: <https://doi.org/10.25559/SITITO.18.202203.507-515>

© Муаль М. Н. Б., Козырев Д. В., 2022



Контент доступен под лицензией Creative Commons Attribution 4.0 License.  
The content is available under Creative Commons Attribution 4.0 License.



## Using Convolutional Neural Networks for Image Detection and Recognition Based on a Self-Written Generator

M. N. B. Mouale\*, D. V. Kozyrev

Peoples' Friendship University of Russia, Moscow, Russian Federation

Address: 6 Miklukho-Maklaya St., Moscow 117198, Russian Federation

\* bmouale@mail.ru

### Abstract

Object recognition is a branch of artificial vision and one of the pillars of machine vision. It consists of the identification of forms pre-described in a digital image and, in general, in a digital video stream. While it is generally possible to perform recognition on video clips, the learning process is usually performed on images. In this paper, we consider an algorithm for classifying and recognizing objects using convolutional neural networks. The purpose of the work is to implement an algorithm for detecting and classifying various graphic objects fed from a webcam. The task is to first classify and recognize an object with high accuracy according to a given dataset and then demonstrate how to generate images to increase the size of the training data set using a self-written generator. The classification and recognition algorithm used is invariant to translation, translation, and rotation. A significant novelty in this work is the creation of a self-written generator that allows you to apply various types of augmentation (artificial increase in the size of the training sample by modifying the training data) to form new groups (batches) of modified images each time.

**Keywords:** image detection, image recognition, convolutional neural networks, R-CNN (Regional Convolutional Neural Networks) model, augmentations, subsampling

**Conflict of interests:** The authors declare no conflict of interest.

**For citation:** Mouale M.N.B., Kozyrev D.V. Using Convolutional Neural Networks for Image Detection and Recognition Based on a Self-Written Generator. *Modern Information Technologies and IT-Education*. 2022;18(3):507-515. doi: <https://doi.org/10.25559/SITITO.18.202203.507-515>



## 1. Введение

В последнее время значительное внимание уделяется распознаванию объектов. Задача распознавания образов вообще есть задача описания и классификации объектов той или иной природы, это задача выработки понятия о классе объектов [1]. Распознавание объектов человеком состоит в том, что он относит их к классу сходных объектов. Множество всех предметов человек делит на классы и каждому классу присваивает имя: «столы», «птицы», «буква «а» и т.п. Человек распознает как зрительные, так и звуковые сигналы. Однако широкий диапазон вариаций человеческого лица из-за позы, освещения и выражения лица приводит к очень сложному распределению и ухудшает эффективность распознавания. Чтобы решить все эти трудности и исключить фактор человеческой ошибки, ставится задача разработки системы обнаружения лиц, которая должна быть достаточно надежной, чтобы справляться с изменениями освещения или погодных условий. Кроме того, эта система должна обрабатывать движения головы и учитывать тот факт, что пользователь может находиться дальше или ближе к камере. Также система должна иметь возможность обнаруживать голову пользователя с определенным углом поворота по каждой оси. Дело в том, что многие алгоритмы распознавания лиц показывают хорошую производительность, когда голова находится почти фронтально. Однако им не удается обнаружить лицо, если оно не анфас (в профиль). Кроме того, система должна быть способна обнаружить и распознать лицо, даже если пользователь носит оптические или солнцезащитные очки, которые создают окклюзию в области глаз. При разработке проекта компьютерного зрения необходимо учитывать все эти факторы. Поэтому появилась необходимость в алгоритме, который бы хорошо разделял объекты на картинке друг от друга с учетом всех сложностей.

В настоящей статье рассматривается алгоритм сверточных нейронных сетей, который позволяет классифицировать, обнаруживать и распознавать изображения, сохранив пространственные отношения между пикселями, путем изучения признаков изображения с использованием небольших квадратов входных данных. Показана эффективность разработанного алгоритма и его инвариантность к переносу, сдвигу и повороту, что позволяет достичь хорошей точности обучения.

Работа структурирована следующим образом:

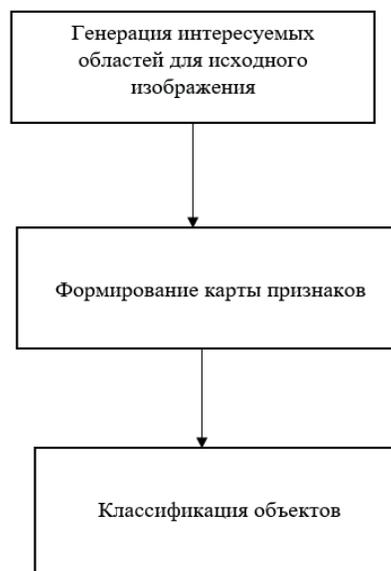
Во втором разделе обсуждается разница между обнаружением и классификацией изображений. В третьем разделе рассматривается метод обработки и классификации изображений с помощью сверточных нейронных сетей. В четвертом разделе приводится практический пример классификации изображения из базы данных «Заландо». Эта часть также посвящена проблеме обнаружения и распознавания изображений. Для проверки работоспособности предложенного метода обнаружения и распознавания приведены результаты экспериментов распознавания неизвестных изображений.

## 2. Разница между обнаружением и классификацией изображения

Важно различать обнаружение и классификацию изображений. Знание того, как обнаружить круг, треугольник, фигуру

или даже лицо, не обязательно относится к области искусственного интеллекта (хотя некоторые алгоритмы способны это делать). Ведь эти задачи возложены на алгоритмы обработки изображения. Однако способность отличить один объект от другого (автомобиль от велосипеда), выполнять распознавание лиц – это искусственный интеллект, в котором обучающиеся алгоритмы обучены распознавать (классифицировать) объекты. Тем не менее, обнаружение форм на изображении может помочь алгоритму классификации в том смысле, что можно легко извлечь лицо из фотографии или видео и предложить его алгоритму классификации, способному выполнять распознавание лиц. Модель, позволяющая одновременно выполнить решение задач поиска объектов (обнаружение) и классификации, называется Модель R-CNN (Regional Convolutional Neural Networks).

При поиске объектов с использованием модели R-CNN выполняется следующая последовательность шагов (рисунок 1):



Р и с. 1. Последовательность шагов при использовании модели R-CNN  
F i g. 1. Sequence of steps when using the R-CNN model

**Генерация интересующих областей для исходного изображения:** выполняется генерация интересующих областей (region proposals), предположительно содержащих в себе искомые объекты с использованием различных алгоритмов (например, Selective search или Regional-proposal function).

**Формирование карты признаков:** на данном шаге выполняется формирование карты признаков для исходного изображения. Проводится масштабирование сформированных интересующих областей в размер, сопоставимый с архитектурой нейронной сети CNN (Convolutional Neural Network или ConvNet). Сформированные данные поступают на вход сверточной нейронной сети (CNN).

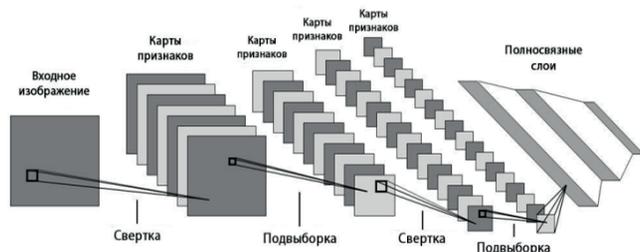
**Классификация объектов:** на данном шаге выполняется классификация объектов для каждой интересующей области с использованием сформированного вектора признаков на основе метода опорных векторов (SVM) [2, 3].



### 3. Обработка и классификация изображений с помощью сверточных нейронных сетей

#### 3.1. Сверточные нейронные сети для классификации изображений

Для классификации изображений используется специальная нейронная сеть: сверточная нейронная сеть CNN. Сверточная нейронная сеть – это глубокая искусственная нейронная сеть с прямой связью, в которой нейронная сеть сохраняет иерархическую структуру, изучая представления внутренних функций и обобщая функции в общих задачах изображения, таких как распознавание объектов и другие проблемы компьютерного зрения. Последние CNN слои состоят из нейронных сетей, входные данные которых поступают из предыдущих слоев, называемых слоями свертки. Чтобы нейронная сеть могла сделать качественную классификацию изображений, ей должна быть предоставлена обучающая база. В нашем случае обучение сети проводилось на наборах данных Заландо [4-8]. Применение нового мульти-масштабного подхода к обучению [9] позволило получить одну модель для обработки изображений различных размеров. Представим в виде схемы принципы функционирования сверточной искусственной нейронной сети (рисунок 2) [10].



Р и с. 2. Архитектура сверточной искусственной нейронной сети  
Fig. 2. Architecture of an ultra-precise artificial neural network

CNN – это архитектура глубокой нейронной сети с прямой связью, состоящая из нескольких сверточных слоев, за каждым из которых следует слой пулинга, функция активации и, возможно, батч нормализации (batch normalization). Он также состоит из полносвязных слоев. По мере того, как изображение перемещается по сети, оно становится меньше, в основном из-за MaxПулинга. Последний слой выводит предсказание вероятностей класса) [11-18].

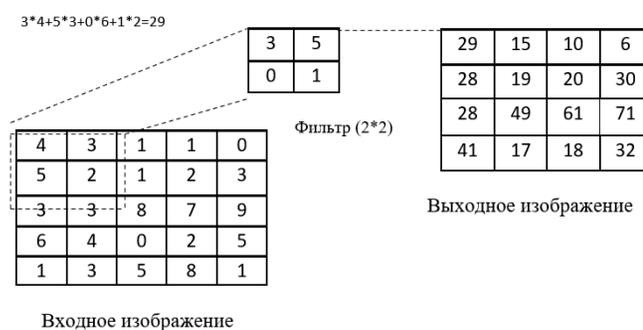
#### 3.2. Обработка изображений

##### 3.2.1. Характеристики изображения и метод свертки изображений

Все начинается с исходного изображения. Представляется, что необходимо классифицировать изображение, содержащее кролика. Цель состоит в том, чтобы, когда мы представляем нашей машине изображение кролика, машина указывала нам, что это кролика или нет. Одно из решений, которое интуитивно приходит на ум, чтобы решить эту проблему, состоит в том, чтобы сравнить попиксельно два изображения, обучающее

изображение и изображение, которое нужно классифицировать. Если эти два изображения совпадают, то можно сказать, что это кролик. Этот метод имеет недостаток, поскольку классифицируемое изображение должно быть строго идентичным эталонному изображению, что совершенно невозможно в реальных случаях применения. Решение этой проблемы заключается в определении признаков на исходном изображении (фрагментах изображений) и поиске их на изображении, которое необходимо классифицировать. Это называется методом свертки.

Если изображение, которое необходимо классифицировать, имеет большое количество общих признаков с исходным классифицируемым изображением, то существует большая вероятность того, что эти изображения были очень похожи. Фаза обучения будет состоять из извлечения этих признаков, а затем подачи их в нейронную сеть, чтобы она научилась устанавливать связь между признаками исходного изображения и маркировкой (имя распознаваемого объекта). В конце обучения нейронная сеть сможет предсказать объект. Чтобы прояснить эту концепцию, рассмотрим изображение размером 5 на 5 пикселей и признаком 2 на 2 пикселя (рисунок 3). Этот признак также называется фильтр или ядро.



Р и с. 3. Извлечение признака из изображения  
Fig. 3. Extracting a feature from an image

Свертка будет состоять из применения фильтра ко всему изображению, чтобы вывести новое изображение. Следующие четыре рисунка поясняют этот процесс:

- Ставим фильтр сверху и слева от изображения;
- Для каждого значения пикселя изображения, находящегося в фильтре, оно умножается на значение пикселя фильтра;
- Сделаем сумму этих произведений;
- Сдвигаем на один пиксель и снова начинаем произведение и сумму. Это движение называется Шаг и в нашем случае имеет значение 1.

Этот процесс должен выполняться до тех пор, пока все изображение не будет отфильтровано, что приведет к созданию нового изображения, меньшего, чем исходное изображение, и имеющего новые значения, как показано на рисунке 3. Эти значения предназначены для выявления определенных признаков изображения.

На этапе свертки мы укажем функцию активации. Обычно используется функция ReLU, позволяющая удалить из изображения свертки отрицательные значения, заменив их значением 0. По сравнению с сигмоидом и tanh, функция активации ReLU



более надежна и ускоряет сходимость в шесть раз. К сожалению, недостатком является то, что ReLU может быть нестабильным во время обучения.

### 3.2.2. Принцип сверток

Как было рассмотрено выше, количество используемых фильтров соответствует признакам, которые необходимо обучить. Следовательно, чем больше различных фильтров используется на этапе обучения, тем эффективнее будет этот этап обучения; однако это может привести к переобучению алгоритма и эффект будет обратным к ожидаемому.

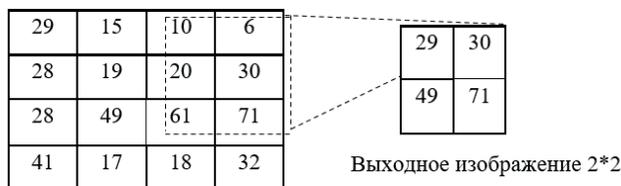
К сожалению, не существует метода для определения оптимального количества используемых фильтров, а также их размера, и часто приходится определять эти значения эмпирическим путём. Если мы применим этот метод обработки свертки к одному изображению, мы увидим, что количество вычислений может быстро стать значительным, если размер изображения большой и, если мы умножаем количество фильтров. Этот расчет требует много ресурсов на уровне процессора и видеокарты. Поэтому в большинстве случаев используются изображения с маленькими размерами (28x28 пикселей или 32x32 пикселей).

### 3.2.3. Пулинг или подвыборка

После получения отфильтрованного изображения мы можем применить к нему новую обработку под названием «Пулинг» (Pooling), что позволит извлечь важные признаки, полученные в результате свертки.

Эта обработка состоит в перемещении окна на изображении, полученного в результате свертки. В контексте «MaxПулинг» мы определяем максимальное значение, содержащееся в этом окне, в контексте «Средний Пулинг» вычисляется среднее значение значений, содержащихся в окне, и, наконец, в контексте «Стохастический Пулинг», мы сохраним значение, выбранное случайным образом.

В качестве примера возьмем изображение из нашей свертки (рисунок 4) и применим к нему окно объединения 2 \* 2 пикселя со строгим смещением 2.



Входное изображение 4\*4

Р и с. 4. Применение MaxПулинга к большому изображению  
F i g. 4. Applying MaxPooling to a Large Image

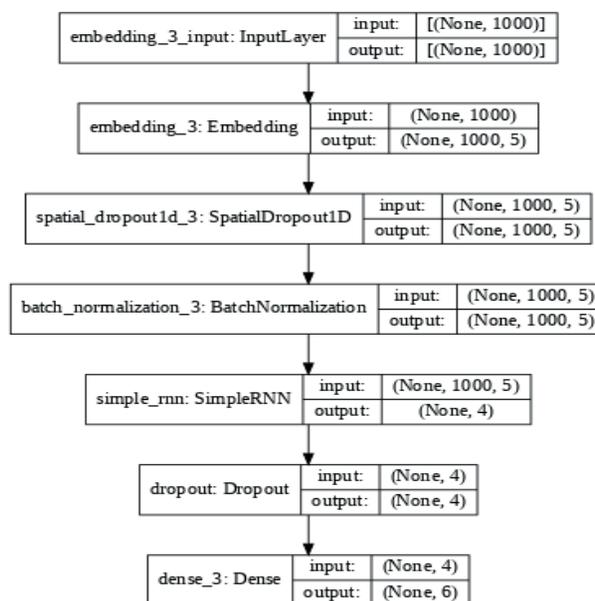
После свертки и пулинга исходное изображение размером 4x4 пикселей сводится к изображению 2x2 с сохранением признаков изображения, полученного в результате свертки. Это уменьшение размера особенно полезно для различных будущих вычислений. Цель состоит в том, чтобы уменьшить изображение, сохранив важную информацию [19-23].

## 4. Практический пример: Классификация изображения из Базы данных «Заландо»

### 4.1. Обучение нейронной сети на обучающих и тестовых данных Базы данных «Заландо»

Одним из наиболее распространенных примеров использования сверточных нейронных сетей является классификация рукописных букв с использованием коллекции наблюдений MNIST (Mixed National Institute of Standards and Technology) [10]. Однако в 2018 году Zalando выпустила собственную коллекцию изображений под названием Zalando-MNIST [10]. Именно этот набор данных используется в нашем практическом случае для обучения сети. Классификация изображений проводится с помощью сверточной нейронной сети, реализованной на основе библиотеки Tensorflow и Keras. Применение нового мульти-масштабного подхода к обучению позволило получить единую модель для обработки изображений различных размеров. Результат работы сети представлен на рисунках 6 и 7.

Ниже представлена схема обучаемой модели с разными параметрами (рисунок 5):



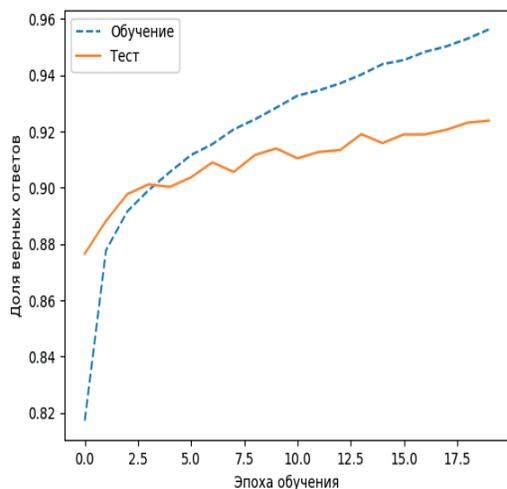
Р и с. 5. Схема обучаемой модели и оценка время обучения

F i g. 5. A diagram for the trainable model and estimation of training time

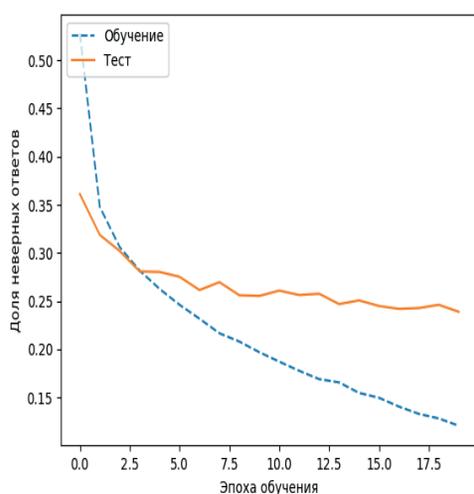
После завершения обучения и формирования обучающей модели мы получаем следующие значения точности на обуча-

Доля верных ответов на обучающем наборе, в процентах: 91.62%  
Время обработки: 9.733333333333333минут  
Модель сохранена !





Р и с. 6. Точность нейронной сети на обучающих и тестовых данных  
F i g. 6. Neural network accuracy on training and test data



Р и с. 7. Ошибки нейронной сети на обучающих и тестовых данных  
F i g. 7. Neural network errors on training and test data

**Вывод:** Фаза обучения заняла около 9 минут в нашем случае с ПК (ноутбук), оснащенным процессором Intel Core i5 2,6 ГГц с 8 ГБ памяти и графической картой Intel NVIDIA GEFORCE 940MX CORE i5 с 64 МБ ОЗУ. После завершения на 20 эпохе обучения и построения нашей обучающей модели мы получили точность 91,62% на тестовом наборе (они могут различаться в зависимости от используемого оборудования).

Графики на рисунках 6 и 7 демонстрируют увеличение точности на этапе обучения до 92%, а также уменьшение ошибки. Это ожидаемый и логичный результат.

#### 4.2. Повышение точности классификации с помощью генератора аргументируемых данных

Иногда бывает так, что в тестовом наборе данных есть изображения, которых нет в обучающем наборе. Следовательно, это влияет на точность классификации модели. Хорошая по объе-

му и качеству база данных – вечная проблема AI-разработчика. Недостаточное количество обучающих данных может привести как к низкой точности работы нейронной сети, так и к переобучению. Таким образом, для решения проблемы, придётся увеличить количество изображений в нашем тренировочном наборе т.е. провести аугментацию данных. Аугментация – это способ генерации новых изображений, похожих на образцы, с изменением позиции, масштаба, яркости и некоторых дополнительных параметров. Но цель аугментации не сгенерировать совсем новые изображения; аугментация помогает нейронной сети просмотреть больше вариантов одних и тех изображений и лучше выделить общие признаки, когда база данных слишком мала.

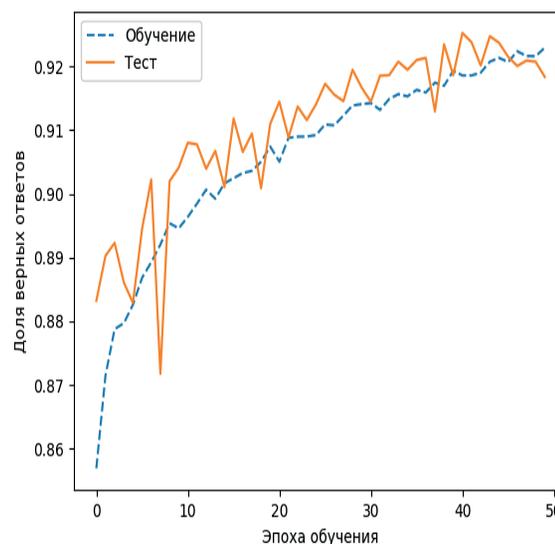
Для этого не нужно искать фото в интернете, мы будем использовать функцию *ImageDataGenerator* модуля Keras, предоставленную для этой цели.

Принцип работы этой функции, следующий:

На основе изображения, содержащегося в обучающем наборе, функция будет выполнять действия случайного поворота, масштабирования и перемещения для создания новых изображений [24, 25].

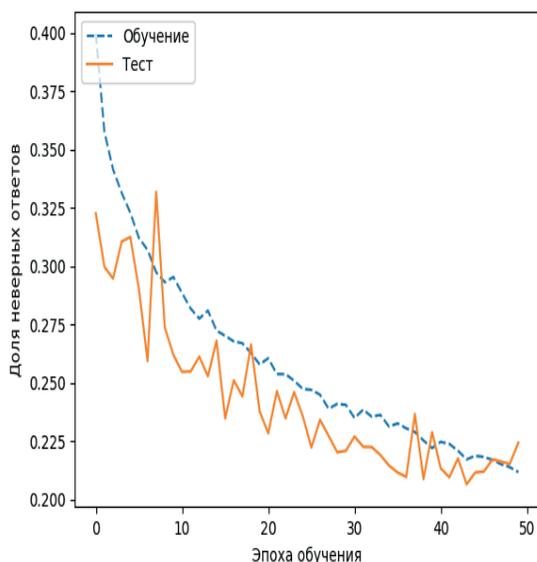
Теперь будем генерировать новые изображения и проводить новое обучение. После завершения обучения и формирования нашей обучающей модели мы получаем следующие значения точности на тестовом наборе.

Доля верных ответов на обучающем наборе, в процентах: 93.09%  
Время обработки: 184.2минут



Р и с. 8. Точность нейронной сети на обучающих и тестовых данных  
F i g. 8. Neural network accuracy on training and test data





Р и с. 9. Ошибки нейронной сети на обучающих и тестовых данных  
F i g. 9. Neural network errors on training and test data

После обучения нашей модели мы сохранили ее с помощью нового скрипта, чтобы делать предсказания на новых неизвестных изображениях. Этот подход аналогичен использованию модели в новом приложении. Были получены результаты, которые представлены на рисунке 10

```

Нейронная сеть распознавала: Кросовки

Футболка/топ: 0.018240550707560033%
Брюки: 0.002725721242313739%
Свитер : 0.03270065353717655%
Платье: 0.030631612753495574%
Пальто: 0.042272411519661546%
сандалии: 1.0207144543528557%
Рубашка: 0.02135587128577754%
Кросовки: 88.91431093215942%
Сумка: 9.671255946159363%
ботильоны: 0.24579542223364115%
Время распознавания после обучения : 0.15минут
    
```

Р и с. 10. Результаты предсказания модели на новых неизвестных изображениях

F i g. 10. Model prediction results on new unknown images

**Вывод:** Замечено, что новое обучение занимает гораздо больше времени (184 минут, т.е. около 3 часов против 9 минут при предыдущем обучении). Это связано с тем, что в дополнение к обучению при каждой партии (batch) генерируются новые изображения, что требует больших аппаратных ресурсов и, следовательно, замедляет весь процесс. Но, несмотря на это время обучения, мы видим небольшое улучшение точности,

приближающееся к 93%. При этом скорость распознавания после обучения значительно мала – 0.15 минут. Также можно заметить снижение частоты ошибок между этапом обучения и этапом тестирования. Поэтому можно считать, что эта модель более эффективна, чем предыдущая.

В этой части мы реализовали эффективную модель классификации изображений с точностью более 93% с использованием сверточных нейронных сетей. Обучение такого типа алгоритма требует большое число изображений, значительных затрат времени и хороших вычислительных ресурсов.

**Предложенный метод:**

- на малых датасетах выступает как способ регуляризации и снижает переобучение на малочисленных классах;
- позволяет сделать классификацию устойчивой к некоторым изменениям тестовых данных (например, при подаче изображения большей яркости относительно обучающей выборки);
- улучшает обобщающую способность сети.

Это очень мощный метод средство расширения обучающего набора данных. Теперь можно управляемо увеличивать объем набора данных, добавляя в него любое нужное количество аугментированных изображений.

Проведено обучение модели на основе базового метода (**самописный генератор**). Эта функция совместно с генераторами данных и выборок позволила применить различные виды аугментации для формирования каждый раз новых батчей с измененными изображениями. В итоге та же модель, которая была обучена ранее без аугментации, может учиться глубже и лучше. За весь период обучения мы подавали намного больше данных, аугментируя изображения для обучения на каждом шаге. Удобство генератора в том, что создание аугментированных изображений на лету занимает память только на один батч. Если попробовать сразу вычислить всю большую аугментированную базу, памяти может просто не хватить. На проверку модели на каждой эпохе мы подавали каждый раз обновленные версии проверочного набора. Как показывают графики (рисунки 8, 9 и 10), модель лучше с ними справлялась. В финале мы оценили работу модели на тестовых данных, которые ей никогда не предъявлялись.

**Заключение**

Распространение и развитие технологий компьютерного зрения влечет за собой изменение других профессиональных областей жизнедеятельности человека.

Сверточные нейронные сети применяются в системах распознавания объектов и лиц, специальном медицинском ПО для анализа снимков, навигации автомобилей, оснащенных автономными системами, в системах защиты, и других сферах. С ростом вычислительной мощности компьютеров, появлением баз изображений стало возможным обучать глубокие нейронные сети. Одной из главных задач машинного обучения является задача классификации и распознавания объектов. СНС применяются для оптического распознавания образов и объектов, детектирования предметов, семантической сегментации и т.д. В данной статье были рассмотрены наиболее распространенные алгоритмы сверточных нейронных сетей для задачи классификации и распознавания изображений с использованием



нового метода, который является инвариантным к переносу, сдвигу, повороту и другим случайным модификациям изображений. Этот метод устойчив к некоторым изменениям тестовых данных (например, при подаче изображения большей яркости относительно обучающей выборки).

В результате проведенных обучений выявлено, что сверточная нейронная сеть показала наилучшие результаты в задаче

классификация и распознавания изображений.

Анализ графических результатов (рисунки 8 и 9 и 10) обучения и результатов работы модели на тестовых данных, можно сделать вывод, что предложенную модель можно учить и далее, но можно искать и более эффективные конфигурации нейронной сети.

## References

- [1] Arulogun O.T., Omidiora E.O., Olaniyi O.M., Ipadeola A.A. Development of Security System using Facial Recognition. *The Pacific Journal of Science and Technology*. 2008;9(2):377-385.
- [2] Viola P., Jones M. Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. Kauai, HI, USA: IEEE Computer Society; 2001. p. 511-518. doi: <https://doi.org/10.1109/CVPR.2001.990517>
- [3] Viola P., Jones M.J. Robust real-time face detection. *International Journal of Computer Vision*. 2004;57(2):137-154. Available at: <https://www.face-rec.org/algorithms/boosting-ensemble/16981346.pdf> (accessed 19.07.2022).
- [4] Canny J. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1986;PAMI-8(6):679-698. doi: <https://doi.org/10.1109/TPAMI.1986.4767851>
- [5] Alblushi A. Face Recognition Based on Artificial Neural Network: A Review. *Artificial Intelligence & Robotics Development Journal*. 2021;1(2):116-131. doi: <https://doi.org/10.52098/airdj.202125>
- [6] Girshick R. Fast R-CNN. In: 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE Computer Society; 2015. p. 1440-1448. doi: <https://doi.org/10.1109/ICCV.2015.169>
- [7] Nguyen D.T., Khachumov V.M., Khachumov M.V., Salpagarov S.I., Yakovlev K.S. On the Estimation of Accuracy and Stability of 3D Face Modeling Using a Pair of Stereo Cameras. *CEUR Workshop Proceedings*. 2018;2236:56-64. Available at: <https://ceur-ws.org/Vol-2236/paper-07-007.pdf> (accessed 19.07.2022).
- [8] He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE Computer Society; 2016. p. 770-778. doi: <https://doi.org/10.1109/CVPR.2016.90>
- [9] Redmon J., Farhadi A. YOLO9000: Better, Faster, Stronger. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE Computer Society; 2017. p. 6517-6525. doi: <https://doi.org/10.1109/CVPR.2017.690>
- [10] Mouale M.N.B., Kozыrev D.V., Houankpo H.G.K., Nibasumba E. Development of a Neural Network Method in the Problem of Classification and Image Recognition. *Modern Information Technologies and IT-Education*. 2021;17(3):507-518. (In Russ., abstract in Eng.) doi: <https://doi.org/10.25559/SITITO.17.202103.507-518>
- [11] Erhan D., Szegedy C., Toshev A., Anguelov D. Scalable Object Detection Using Deep Neural Networks. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE Computer Society; 2014. p. 2155-2162. doi: <https://doi.org/10.1109/CVPR.2014.276>
- [12] He K., Zhang X., Ren S., Sun J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2015;37(9):1904-1916. doi: <https://doi.org/10.1109/TPAMI.2015.2389824>
- [13] Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*. 2017;60(6):84-90. doi: <https://doi.org/10.1145/3065386>
- [14] Gao Z., Guan S., Guo M. Semantic Segmentation of Street View Image Based on Fully Convolutional Neural Networks. In: 2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT). Shanghai, China: IEEE Computer Society; 2021. p. 400-404. doi: <https://doi.org/10.1109/AINIT54228.2021.00084>
- [15] Gao H., Zhang Y., Lv W., Yin J., Qasim T., Wang D. A Deep Convolutional Generative Adversarial Networks-Based Method for Defect Detection in Small Sample Industrial Parts Images. *Applied Sciences*. 2022;12(13):6569. doi: <https://doi.org/10.3390/app12136569>
- [16] Mrong S., Jormanainen I., Toivonen T. Visualization tool for teaching and learning Artificial Neural Networks. In: Ninth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'21) (TEEM'21). New York, NY, USA: Association for Computing Machinery; 2021. p. 58-62. doi: <https://doi.org/10.1145/3486011.3486419>
- [17] Houankpo H.G.K., Kozыrev D.V., Nibasumba E., Mouale M.N.B., Sergeeva I.A. A Simulation Approach to Reliability Assessment of a Redundant System with Arbitrary Input Distributions. In: Vishnevskiy V.M., Samouylov K.E., Kozыrev D.V. (eds.). Distributed Computer and Communication Networks. DCCN 2020. Lecture Notes in Computer Science. Vol. 12563. Cham: Springer; 2020. p. 380-392. doi: [https://doi.org/10.1007/978-3-030-66471-8\\_29](https://doi.org/10.1007/978-3-030-66471-8_29)
- [18] Panetto H., Cecil J. Information systems for enterprise integration, interoperability and networking: theory and applications. *Enterprise Information Systems*. 2013;7(1):1-6. doi: <https://doi.org/10.1080/17517575.2012.684802>
- [19] Baby C.T., Baby C.J. Image Detection and Recognition Using Convolutional Neural Networks. In: Kumar S., Hiranwal S., Purohit S.D., Prasad M. (eds.). Proceedings of International Conference on Communication and Computational Technologies. Algorithms for Intelligent Systems. Singapore: Springer; 2023. p. 255-266. doi: [https://doi.org/10.1007/978-981-19-3951-8\\_20](https://doi.org/10.1007/978-981-19-3951-8_20)



- [20] Poniszewska-Maranda A. Management of access control in information system based on role concept. *Scalable Computing: Practice and Experience*. 2011;12(1):35-49. Available at: <https://www.scpe.org/index.php/scpe/article/view/686> (accessed 19.07.2022).
- [21] Kharchevnikova A.S., Savchenko A.V. Neural Networks in Video-Based Age and Gender Recognition on Mobile Platforms. *Optical Memory and Neural Networks*. 2018;27(4):246-259. doi: <https://doi.org/10.3103/S1060992X18040021>
- [22] Parkhi O.M., Vedaldi A., Zisserman A. Deep Face Recognition. In: Xie X., Jones M.W., Tam G.K.L. (eds.). Proceedings of the British Machine Vision Conference (BMVC). BMVA Press; 2015. p. 41.1-41.12. doi: <https://dx.doi.org/10.5244/C.29.41>
- [23] Cao Q., Shen L., Xie W., Parkhi O.M., Zisserman A. VGGFace2: A Dataset for Recognising Faces across Pose and Age. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). IEEE Computer Society; 2018. p. 67-74. doi: <https://doi.org/10.1109/FG.2018.00020>
- [24] Sharma R., Kim M., Gupta A. Motor imagery classification in brain-machine interface with machine learning algorithms: Classical approach to multi-layer perceptron model. *Biomedical Signal Processing and Control*. 2022;71(A):103101. doi: <https://doi.org/10.1016/j.bspc.2021.103101>
- [25] Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In: 3rd International Conference on Learning Representations (ICLR 2015). Computational and Biological Learning Society; 2015. p. 1-14. doi: <https://doi.org/10.48550/arXiv.1409.1556>

*Поступила 19.07.2022; одобрена после рецензирования 10.09.2022; принята к публикации 24.09.2022.  
Submitted 19.07.2022; approved after reviewing 10.09.2022; accepted for publication 24.09.2022.*

#### Об авторах:

**Муаль Мутуама Нда Бьенвеню**, аспирант кафедры прикладной информатики и теории вероятностей факультета физико-математических и естественных наук, ФГАОУ ВО «Российский университет дружбы народов» (117198, Российская Федерация, г. Москва, ул. Миклухо-Маклая, д. 6), ORCID: <https://orcid.org/0000-0002-7230-5714>, [bmouale@mail.ru](mailto:bmouale@mail.ru)

**Козырев Дмитрий Владимирович**, доцент кафедры прикладной информатики и теории вероятностей факультета физико-математических и естественных наук, ФГАОУ ВО «Российский университет дружбы народов» (117198, Российская Федерация, г. Москва, ул. Миклухо-Маклая, д. 6), кандидат физико-математических наук, доцент, ORCID: <https://orcid.org/0000-0003-0538-8430>, [kozyrev-dv@rudn.ru](mailto:kozyrev-dv@rudn.ru)

*Все авторы прочитали и одобрили окончательный вариант рукописи.*

#### About the authors:

**Moutouama N'dah B. Mouale**, Postgraduate Student of the Department of Applied Probability and Informatics, Faculty of Science, Peoples' Friendship University of Russia (6 Miklukho-Maklaya St., Moscow 117198, Russian Federation), ORCID: <https://orcid.org/0000-0002-7230-5714>, [bmouale@mail.ru](mailto:bmouale@mail.ru)

**Dmitry V. Kozyrev**, Associate Professor of the Department of Applied Probability and Informatics, Faculty of Science, Peoples' Friendship University of Russia (6 Miklukho-Maklaya St., Moscow 117198, Russian Federation), Cand.Sci. (Phys.-Math.), Associate Professor, ORCID: <https://orcid.org/0000-0003-0538-8430>, [kozyrev-dv@rudn.ru](mailto:kozyrev-dv@rudn.ru)

*All authors have read and approved the final manuscript.*

