

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра математического анализа

Нейронные сети в задачах обработки изображений

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студента 2 курса 218 группы

направление **01.04.02 – Прикладная математика и информатика**

механико-математического факультета

Нуриева Тимура Тагировича

Научный руководитель

доцент, к. т. н.

А.К. Кузьмин

Заведующий кафедрой

И.о.зав.кафедрой, д.ф.-м.н.

П.А.Терехин

Саратов 2024

Введение. В современном мире обработка изображений играет ключевую роль во многих областях, таких как медицина, безопасность, транспорт, сельское хозяйство и многих других. С развитием технологий и увеличением объёмов данных, обработка изображений становится всё более сложной и актуальной задачей. Для решения таких задач используют нейронные сети. Они также находят свое применения в медицине, безопасности, транспорте, сельском хозяйстве и многих других областях.

Основное содержание работы. Магистерская работа состоит из девяти разделов.

Во введении представлены общие понятия касающиеся нейронных сетей рассмотрены ключевые особенности написания нейронной сети. Представлен разный взгляд на нейронные сети с точки зрения машинного обучения, математики, кибернетики, развития вычислительной техники и программирования и с точки зрения искусственного интеллекта. Введено понятие архитектуры нейронной сети и общее представление архитектуры нейронной сети.

В первом разделе рассмотрена актуальность развития нейронных сетей на примере их применения в современном мире:

- анализ медицинских изображений таких как рентгеновские снимки, МРТ и КТ, для диагностики заболеваний и планирования лечения;
- обнаружение и распознавание лиц, номерных знаков автомобилей, а также для анализа видеопотоков для выявления подозрительного поведения;
- создании автономных транспортных средств, которые могут распознавать дорожные знаки, пешеходов и другие объекты;
- анализ состояния растений, определение заболеваний и прогнозирования урожайности.

Во втором разделе приведена общая постановка задачи для обучения нейронной сети. Приведены базовые типы задач, для которых применяются нейронные сети.

- **Регрессия.** Для определения возраста по фотографии, составления прогноза биржевых курсов, оценки стоимости имущества и других задач, требующих получения в результате обработки конкретного числа.
- **Прогнозирования временных рядов.** Для составления долгосрочных прогнозов на основе динамического временного ряда значений. Например, нейронные сети применяются для предсказания цен, физических явлений, объема потребления и других показателей. По сути, даже работу автопилота Tesla можно отнести к процессу прогнозирования временных рядов.
- **Кластеризация.** Для изучения и сортировки большого объема неразмеченных данных в условиях, когда неизвестно количество классов на выходе, то есть для объединения данных по признакам. Например, кластеризация применяется для выявления классов картинок и сегментации клиентов.
- **Генерация.** Для автоматизированного создания контента или его трансформации. Генерация с помощью нейронных сетей применяется для создания уникальных текстов, аудиофайлов, видео, раскрашивания черно-белых фильмов и даже изменения окружающей среды на фото.
- **Классификация.** Для распознавания лиц, эмоций, типов объектов: например, квадратов, кругов, треугольников. Также для распознавания образов, то есть выбора конкретного объекта из предложенного множества: например, выбор квадрата среди треугольников.

В третьем разделе рассматриваются три наиболее используемых видов нейронных сетей.

- **Нейронные сети прямого распространения** (Feed forward neural networks, FFNN). Прямолинейный вид нейронных сетей, при котором

соседние узлы слоя не связаны, а передача информации осуществляется напрямую от входного слоя к выходному. FFNN имеют малую функциональность, поэтому часто используются в комбинации с сетями других видов.

- **Рекуррентные нейронные сети** (Recurrent neural network, RNN). Используют направленную последовательность связи между узлами. В RNN результат вычислений на каждом этапе используется в качестве исходных данных для следующего. Благодаря этому, рекуррентные нейронные сети могут обрабатывать серии событий во времени или последовательности для получения результата вычислений.
- **Сверточные нейронные сети** (Convolutional neural network, CNN) - это тип глубоких нейронных сетей, которые используются для обработки изображений и видео. Стали популярными из-за своей способности извлекать признаки из изображений.

В четвертом разделе проведено исследование возможностей современной генеративной нейронной сети OpenAI GPT 3 посредством взаимодействия через инструмент Chat GPT v3.

Основные характеристики Chat GPT v3:

- Улучшенное понимание контекста. Модель учитывает контекст разговора и может генерировать более релевантные ответы.
- Способность к обучению. Chat GPT v3 обучается на больших объемах данных и постоянно совершенствует свои навыки.
- Широкий спектр применения. Модель может использоваться для создания контента, решения задач, предоставления информации и многого другого.
- Интеграция с другими технологиями. Chat GPT v3 может быть интегрирован с другими технологиями, такими как голосовые помощники и чат-боты.

Chat GPT v3 работает на основе алгоритмов машинного обучения и обработки естественного языка. Модель анализирует запрос пользователя, определяет его контекст и генерирует ответ, который наиболее точно соответствует запросу.

Исследование приведено в Приложении 3 к магистерской работе, по результатам которого было получено заключение, что возможностей нейронной сети не достаточно для написания математической работы, поскольку генеративные сети не могут генерировать новый контент, а только используют известный им.

В пятом разделе подробно описана архитектура сверточных нейронных сетей (Convolutional neural network, CNN). Они состоят из слоев пяти типов в соответствии с Рис. 1:

- ВХОДНОГО;
- свертывающего;
- объединяющего;
- подключенного;
- ВЫХОДНОГО.

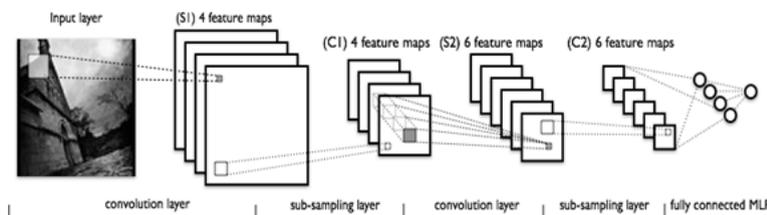


Рис. 1. Архитектура сверточной нейронной сети

Каждый слой выполняет определенную задачу: например, обобщает или соединяет данные. Сверточные нейронные сети применяются для классификации изображений, распознавания объектов, прогнозирования, обработки естественного языка и других задач.

Первые два типа слоев (convolutional, subsampling), чередуясь между собой, формируют входной вектор признаков для многослойного персептрона.

Свое название сверточная сеть получила по названию операции – свертка, суть которой будет описана ниже.

Топология нейронной сети. Определение топологии сети ориентируется на решаемую задачу, данные из научных статей и собственный экспериментальный опыт. Можно выделить следующие этапы влияющие на выбор топологии:

- определить решаемую задачу нейронной сетью (классификация, прогнозирование, модификация);
- определить ограничения в решаемой задаче (скорость, точность ответа);
- определить входные (тип: изображение, звук, размер: 100x100, 30x30, формат: RGB, в градациях серого) и выходные данные (количество классов).

Входной слой учитывает двумерную топологию изображений и состоит из нескольких карт (матриц), карта может быть одна, в том случае, если изображение представлено в оттенках серого, иначе их 3, где каждая карта соответствует изображению с конкретным каналом .

Входные данные каждого конкретного значения пикселя нормализуются в диапазон от 0 до 1, по формуле:

$$f(p, min, max) = \frac{p-min}{max-min}$$

где f - функция нормализации;

p – значение конкретного цвета пикселя от 0 до 255

min – минимальное значение пикселя – 0

max – максимальное значение пикселя - 255

Сверточный слой представляет из себя набор карт, у каждой карты есть синаптическое ядро.

Количество карт определяется требованиями к задаче, если взять большое количество карт, то повысится качество распознавания, но увеличится вычислительная сложность. Исходя из анализа научных статей, в большинстве случаев предлагается брать соотношение один к двум, то есть каждая карта

предыдущего слоя связана с двумя картами сверточного слоя, в соответствии с рис. 2.

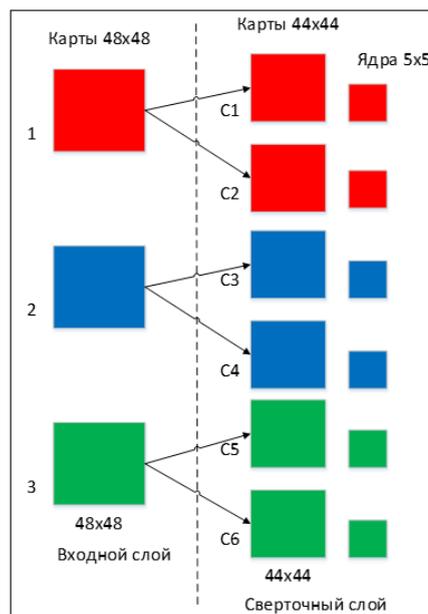


Рис. 2. Пример выбора карт

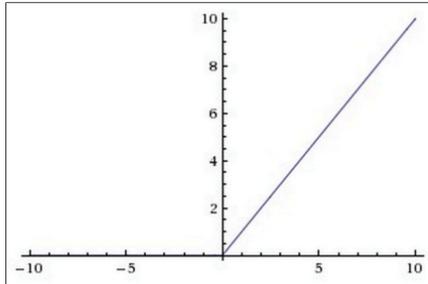
Подвыборочный слой также, как и сверточный имеет карты, но их количество совпадает с предыдущим (сверточным) слоем, их 6. Цель слоя – уменьшение размерности карт предыдущего слоя. Если на предыдущей операции свертки уже были выявлены некоторые признаки, то для дальнейшей обработки настолько подробное изображение уже не нужно, и оно уплотняется до менее подробного.

Полносвязный слой - последний из типов слоев это слой обычного многослойного персептрона. Цель слоя – классификация, моделирует сложную нелинейную функцию, оптимизируя которую, улучшается качество распознавания.

Выходной слой связан со всеми нейронами предыдущего слоя. Количество нейронов соответствует количеству распознаваемых классов, то есть 2 – лицо и не лицо. Но для уменьшения количества связей и вычислений для бинарного случая можно использовать один нейрон и при использовании в качестве функции активации гиперболический тангенс.

Также приведено сравнение **функций активации**. Вид функции активации во многом определяет функциональные возможности нейронной сети и метод обучения этой сети.

ReLU (rectified linear unit)



$$f(s) = \max(0, s)$$

$$f'(s) = \begin{cases} 1, & s > 0 \\ \text{rand}(0.01, 0.05), & s = 0 \end{cases}$$

Достоинства

- лишена ресурсоемких операций
- отсекает ненужные детали
- отсутствует разрастание/затухание градиента

Недостатки

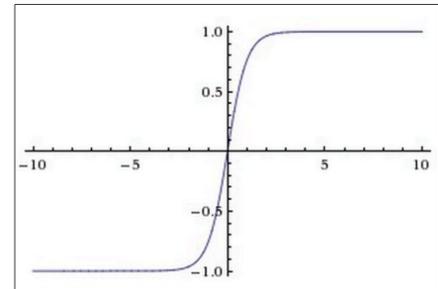
- не всегда надежна, в процессе обучения может «умирать»
- сильно зависима от инициализации весов

Функция активации сигмоиды относится к классу непрерывных функций и принимает на входе произвольное вещественное число, а на выходе дает вещественное число в интервале от 0 до 1. Сигмоида (sigmoid) выражается формулой:

$$f(s) = \frac{1}{1 + e^{-s}}$$

График сигмоидальной функции в соответствии с Рис. 3:

Гиперболический тангенс



$$f(s) = \frac{e^{2s} - 1}{e^{2s} + 1}$$

$$f'(s) = 1 - f(s)^2$$

Достоинства

- простое вычисление производной через значение своей функции
- область значений от -1 до 1

Недостатки

- затухание или увеличение градиента
- ресурсоемкая по сравнению с ReLU

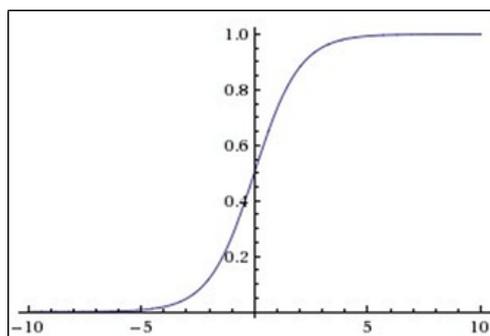


Рис. 3. Сигмоидальная функция

Функция активации гиперболический тангенс.

- симметричные активационные функции, типа гиперболического тангенса обеспечивают более быструю сходимость, чем стандартная логистическая функция;
- функция имеет непрерывную первую производную;
- функция имеет простую производную, которая может быть вычислена через ее значение, что дает экономию вычислений.

График функции гиперболического тангенса показан на Рис. 4:

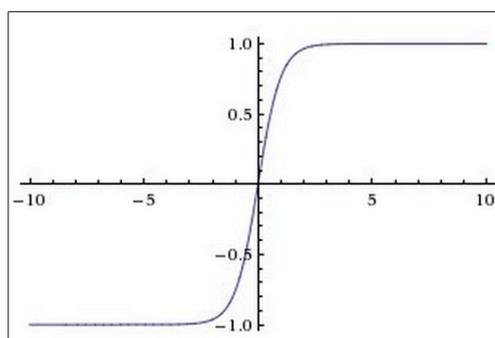


Рис. 4. Гиперболический тангенс

Функция активации ReLU - выпрямленная линейная функция активации (rectified linear unit, ReLU), которая выражается формулой:

$$f(s) = \max(0, s)$$

График функции ReLU в соответствии с Рис. 5:

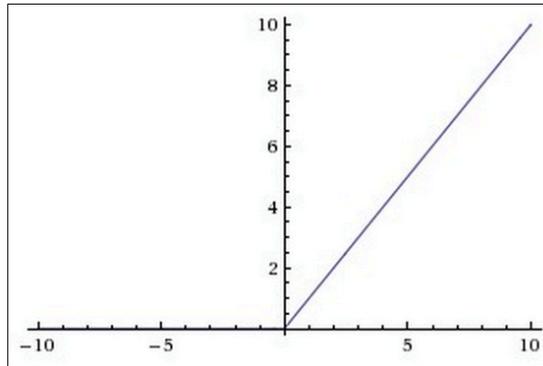


Рис. 5. Функция ReLU

В шестом разделе представлено решение задачи классификации и обучение нейронной сети на открытом датасете Fashion MNIST состоящим из 70 000 изображений предметов одежды в черно-белом формате, разделенных на 10 категорий. Подробно рассмотрены инструменты tensorflow - библиотеки применяемой в создании архитектуры нейронных сетей. Код программы приведен в Приложении 1 магистерской работы.

В седьмом разделе определяется задача обнаружения объектов на изображении, и приводятся аргументы в пользу использования сверточных нейронных сетей для ее решения.

Сверточные нейронные сети имеют ряд преимуществ для обнаружения объектов:

+Высокая точность и производительность. Сверточные нейронные сети позволяют достичь высокой точности обнаружения объектов.

+Возможность работы с различными типами изображений. Сверточные нейронные сети могут быть использованы для обнаружения объектов на различных типах изображений, таких как медицинские изображения, изображения дистанционного зондирования и т. д.

Однако сверточные нейронные сети также имеют ряд недостатков:

–Необходимость большого количества данных для обучения. Сверточные нейронные сети требуют большого количества данных для обучения.

–Сложность интерпретации результатов. Результаты обнаружения объектов, полученные с помощью сверточных нейронных сетей, могут быть сложными для интерпретации.

В восьмом разделе рассмотрены различные архитектуры нейронных сетей для обнаружения объектов. Среди них R-CNN, Fast R-CNN, Faster R-CNN, YOLO, SSD, Feature Pyramid Networks, RetinaNet.

Подробно рассматривается используемая в следующем разделе архитектура YOLOv3 — это усовершенствованная версия архитектуры YOLO. Она состоит из 106-ти сверточных слоев и лучше детектирует небольшие объекты по сравнению с её предшественницей YOLOv2. Особенность YOLOv3 состоит в том, что на выходе есть три слоя каждый из которых рассчитан на обнаружения объектов разного размера.

Девятый раздел посвящена написанию кода для решения задачи определения объектов как на изображения так и на видеозаписи. Для написания кода была выбрана облачная среда программирования Google Colaboratory, поскольку данный ресурс предоставляет дополнительные облачные мощности, что стало ключевым моментом.

Также рассматривается инструментарий библиотеки imageAI, которая позволяет без труда подключиться к предобученным сверточным нейронным сетям RetinaNet, YOLOv3, TinyYOLOv3.

Заключение. В ходе работы были рассмотрены основные аспекты функционирования нейронных сетей, их применение и перспективы развития.

Нейронные сети представляют собой мощный инструмент для решения сложных задач, таких как распознавание образов, классификация данных, прогнозирование и многие другие. Они способны обучаться на больших объёмах данных и автоматически настраивать свои параметры для достижения оптимальных результатов.

Несмотря на эти ограничения, нейронные сети продолжают развиваться и находить новые применения. Они становятся всё более доступными и

эффективными, что открывает новые возможности для их использования в различных областях, таких как медицина, финансы, транспорт и другие.

Таким образом, можно сделать вывод, что нейронные сети являются перспективным направлением развития искусственного интеллекта. Они требуют значительных ресурсов для обучения и развития, но их потенциал огромен. Дальнейшее развитие нейронных сетей может привести к созданию более эффективных и интеллектуальных систем, способных решать сложные задачи и адаптироваться к изменяющимся условиям.