

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра теоретических основ
компьютерной безопасности и
криптографии

**Составление текстовых описаний для изображений с помощью нейронных
сетей**

АВТОРЕФЕРАТ

дипломной работы

студента 6 курса 631 группы

специальности 10.05.01 Компьютерная безопасность

факультета компьютерных наук и информационных технологий

Катимуллина Тимура Даниярбековича

Научный руководитель

к. ф.-м. н., доцент

А. Н. Гамова

22.01.2022 г.

Заведующий кафедрой

д. ф.-м. н., доцент

М. Б.

Абросимов

22.01.2022 г.

Саратов 2022

ВВЕДЕНИЕ

За последние 10 лет технологии глубокого обучения нашли свое повсеместное применение: от прогнозирования погоды до вычислительной фотографии и качественного машинного перевода. Развитие вычислительных мощностей позволяют строить более сложные модели нейронных сетей и решать комплексные проблемы, например, автоматизация беспилотных автомобилей, внедрения систем распознавания лиц, визуальной помощи люди с ограниченными возможностями, описание кадров видеонаблюдения, улучшение качество поиска изображений и т.д. С помощью генерации описаний изображений можно решить или приблизиться к решению вышеперечисленных задач.

Создание подписей к изображению – это процесс создания описания изображения на естественном языке. Данная задача является сложной, поскольку требует идентификации характерных объектов и областей на изображении, понимания их взаимосвязей и создания соответствующих описаний областей изображения на естественном языке.

Целью данной работы является изучение видов нейронных сетей и анализа механизма их работы, изучение алгоритма и способа генерации текстовых описаний, а также его реализация с помощью фреймворка глубокого обучения PyTorch.

Для достижения поставленной цели необходимо выполнение следующих задач:

- исследовать существующие методы генерации описаний изображений;
- реализовать программный, позволяющего генерировать описание для выбранного изображения.

Дипломная работа состоит из введения, 6 разделов, заключения, списка использованных источников и 1 приложения. Общий объем работы – 61

страницы, из них 31 страницы – основное содержания, включая 20 рисунков, список использованных источников из 18 наименования.

КРАТКОЕ СОДЕРЖАНИЕ

Первый раздел содержит необходимые теоретические сведения, касающиеся теории, и принципы работы нейронных сетей. Приводятся основные определения, используемые в дипломной работе.

Искусственная нейронная сеть (ИНС) (также нейронная сеть (НС)) — упрощенная модель биологической нейронной сети, представляющая собой совокупность искусственных нейронов, взаимодействующих между собой.

Основные принципы работы нейронных сетей были описаны еще в 1943 году Уорреном Мак-Каллоком и Уолтером Питтсом. А в 1957 году нейрофизиолог Фрэнк Розенблатт разработал первую нейронную сеть.

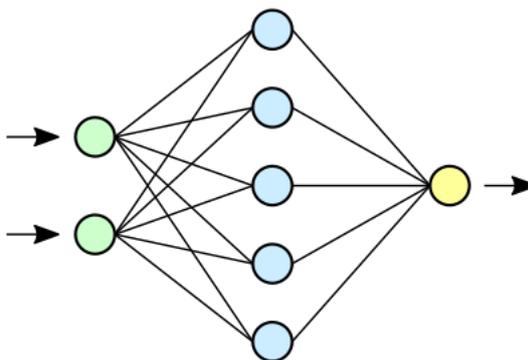


Рисунок 1 – Схема простой нейронной сети

ИНС использует структуру, похожую на структуру человеческого мозга. Как и биологическая нейронная сеть, искусственная состоит из нейронов, взаимодействующих между собой, однако представляет собой упрощенную модель. На рисунке 1 представлена схема простой искусственно нейронной сети. Зелёным цветом обозначены входные нейроны, голубым — скрытые нейроны, жёлтым — выходной нейрон. Так, например, искусственный нейрон, из которых состоит ИНС, имеет намного более простую структуру: у него есть несколько входов, на которых он принимает различные сигналы, преобразует их и передает другим нейронам. То есть, искусственный нейрон — это функция $\mathbb{R}^n \rightarrow \mathbb{R}$, которая преобразует n входных параметров в один выходной.

Во втором разделе раскрываются архитектура, основные понятия и принципы работы сверточных нейронных сетей.

Свёрточная нейронная сеть (Convolutional neural network, CNN) — специальная архитектура искусственных нейронных сетей, предложенная Яном Лекуном в 1988 году и нацеленная на эффективное распознавание образов, входит в состав технологий глубокого обучения (deep learning). Использует 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. Таким образом, идея свёрточных нейронных сетей заключается в чередовании свёрточных слоёв (англ. convolution layers) и субдискретизирующих слоёв (англ. subsampling layers или pooling layers). Структура сети — однонаправленная (без обратных связей), многослойная. Для обучения используются стандартные методы, чаще всего метод обратного распространения ошибки. Функция активации нейронов — любая, по выбору исследователя. На рисунке 2 представлена схема сверточной нейронной сети.

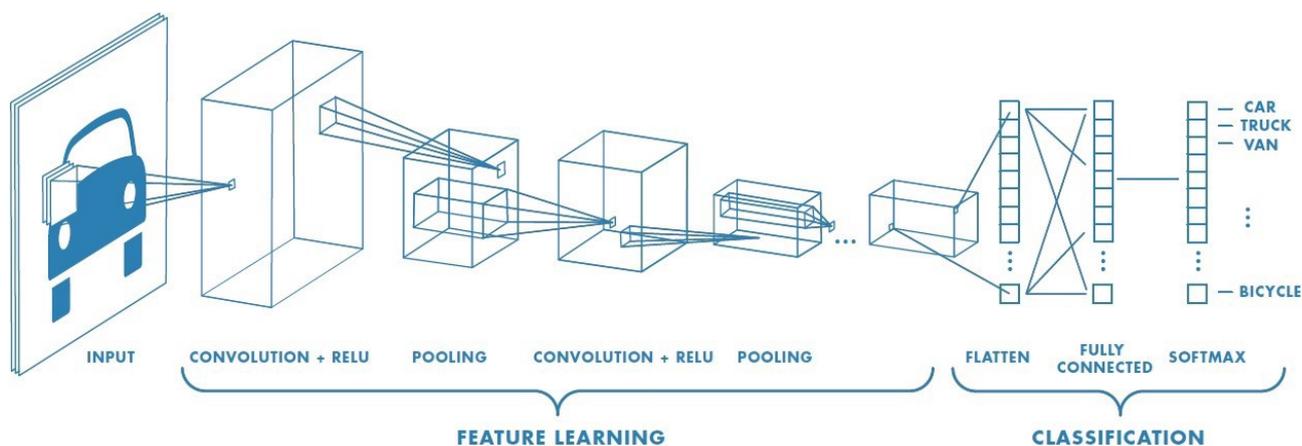


Рисунок 2 – Архитектура сверточной нейронной сети

Название архитектура сети получила из-за наличия операции свёртки, суть которой в том, что каждый фрагмент изображения умножается на матрицу (ядро) свёртки поэлементно, а результат суммируется и записывается в аналогичную позицию выходного изображения.

В следующей главе рассматриваются архитектура, основные понятия и принципы работы рекуррентных нейронных сетей.

Рекуррентные нейронные сети (РНС) — вид нейронных сетей, где связи между элементами образуют направленную последовательность. Общая схема рекуррентной сети представлена на рисунке 3. На рисунке 4 представлен слой сети.

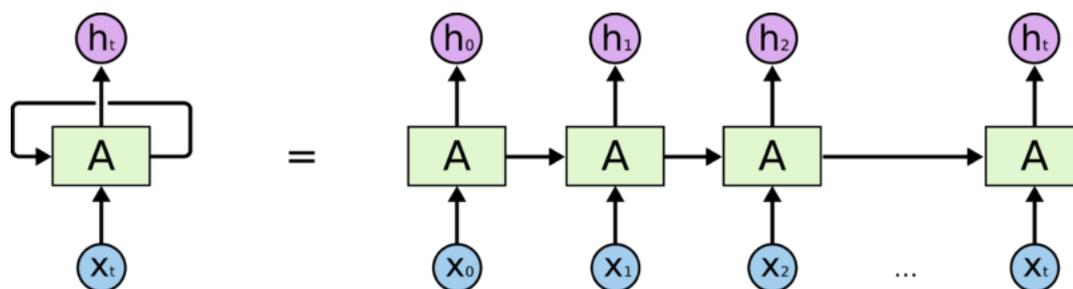


Рисунок 3 – Рекуррентная нейронная сеть

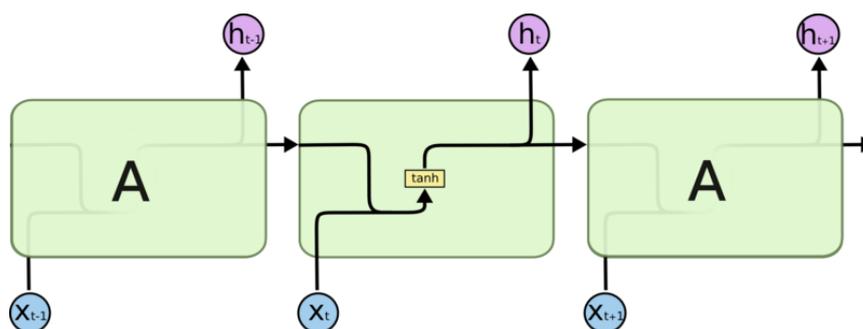


Рисунок 4 – Схема слоя рекуррентной сети

Благодаря этому появляется возможность обрабатывать серии событий во времени или последовательные пространственные цепочки. В отличие от многослойных перцептронов, рекуррентные сети могут использовать свою внутреннюю память для обработки последовательностей произвольной длины. Поэтому сети RNN применимы в таких задачах, где нечто целостное разбито на части, например: распознавание рукописного текста или распознавание речи. Было предложено много различных архитектурных решений для рекуррентных сетей от простых до сложных. В последнее время наибольшее распространение получили сеть с долговременной и кратковременной памятью (LSTM) и управляемый рекуррентный блок (GRU).

В 4 главе описывается механизм внимания.

Механизмы внимания (attention) — это подход в машинном обучении, заключающийся в выделении части входных данных (регионов изображений, фрагментов текста) для более детальной обработки.

Часто для решения задачи генерации описаний изображений не требуется обрабатывать все пиксели изображения: например, в данной задаче фон часто играет незначительную роль. Тем не менее, сверточные сети, являющиеся наиболее популярным методом решения, затрачивают одинаковое количество вычислительных ресурсов на все части изображения. Механизмы внимания могут быть использованы для повышения производительности нейронных сетей.

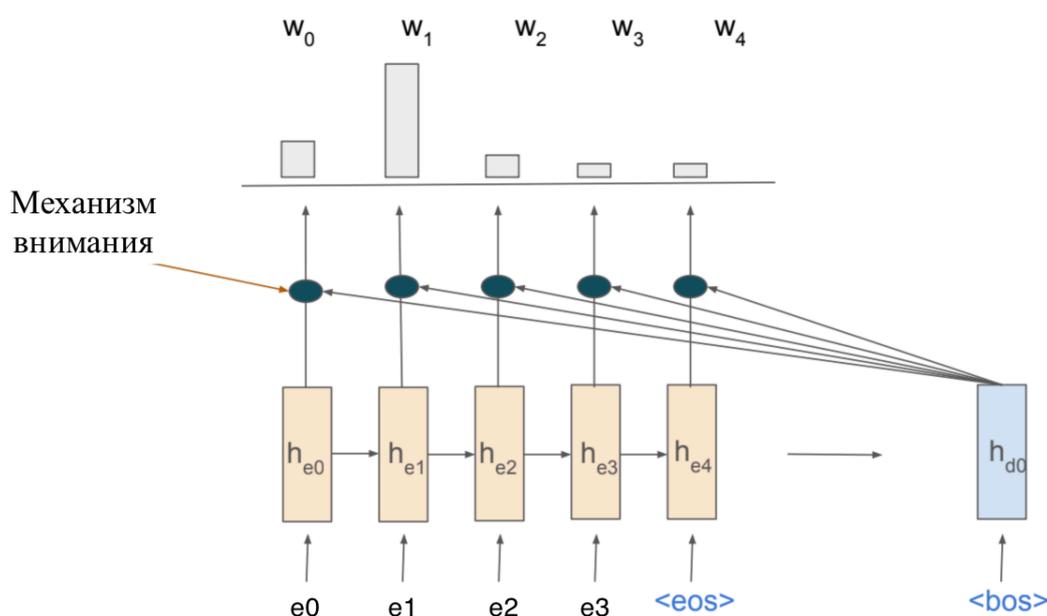


Рисунок 5 – Схема работы механизма внимания

В главе 5 рассматривается модель Энкодер-Декодер.

Данная модель состоит из двух частей: энкодера (Encoder) и декодера (Decoder). Схема отображена на рисунке 6.

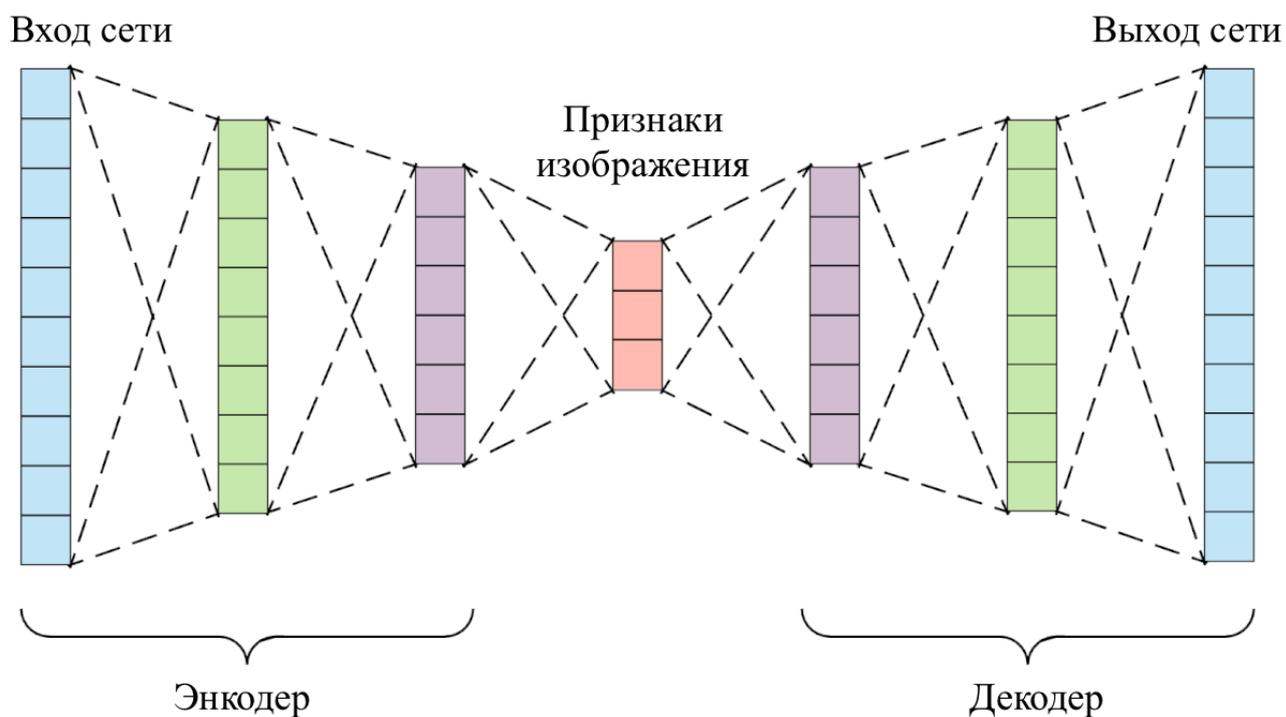


Рисунок 6 – Схема энкодер-декодер

Энкодер кодирует поданное на вход изображение с 3 цветовыми каналами (rgb) и размерами $H \times W$ в изображение с "изученными" каналами размером $h' \times w'$. Такое закодированное изображение представляет собой выученные и важные по мнению нейронной сети признаки в исходном изображении.

Декодер отвечает за то, чтобы с помощью признаков, полученных из энкодера, сгенерировать описание изображения слово за словом. В решении задачи генерации последовательности в большинстве случаев лучше всего справляется рекуррентная нейронная сеть. В ходе работы будет использоваться ее модификация сеть с долговременной и кратковременной памятью (LSTM).

В шестом разделе приводится описание реализованной в ходе выполнения дипломной работы программы, состоящей из двух подпрограмм, одна из которых реализует веб-сервис, а вторая отвечает за обучение и вывод описания.

Представленная нейронная сеть был создана и обучена с использованием облачной платформы исполнения кода Google Colab на одной GPU NVIDIA Tesla P100. Обучение длилось 150 эпох, что примерно 80 часов. Обучено проводилось с использованием датасета MSCOCO 2014 Captions с объемом тренировочной выборки 13 гигабайт и валидационной в 6 гигабайт.

Создание и отладка веб-приложения производилась на устройстве с операционной системой MacOS 10.15.7 и интегрированной средой разработки PyCharm. В качестве клиента интерфейсной части приложения использовался браузер Google Chrome 94.0.4606.81.

Также раздел включает в себя скриншоты графического интерфейса программы.

Полный листинг программы приведен в приложении А.

ЗАКЛЮЧЕНИЕ

В ходе выполнения работы были рассмотрены различные виды нейронных сетей и методов работы с изображениями и текстов, такие как:

- Сверточные нейронные сети
- Рекуррентные нейронные сети
- Механизм внимания
- Модель энкодер-декодер

В практической части работы реализована программа, позволяющая обучить нейронную сеть для генерации описаний изображений, а также веб-приложение для использования данной модели.

Программный комплекс может быть использована для автоматизации создании тегов изображений, распознавания объектов для систем безопасности, для улучшения работы поиска, помощи людям с нарушениями зрения.