



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ ИУ «ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ»

КАФЕДРА ИУ7 «ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ЭВМ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ»

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

НА ТЕМУ:

Свёрточные нейронные сети

Студент ИУ7-51Б

Н.А. Беляев

Руководитель

А.С. Кострицкий

2024 г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
1 Аналитический раздел	5
1.1 Основные понятия	5
1.2 Вехи развития	6
1.3 Задача генерации признаков	7
1.4 Некоторые архитектуры свёрточных сетей	9
1.5 Сравнение перечисленных архитектур	10
ЗАКЛЮЧЕНИЕ	11
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	12
ПРИЛОЖЕНИЕ А	13

ВВЕДЕНИЕ

Человеческое внимание – ограниченный и несовершенный ресурс. Существует широкий класс задач, требующий точного обнаружения объектов в потоке входных данных. К таким задачам относится распознавание объектов на фото и видео, определение некоторых признаков в последовательности сигналов. Зачастую человек не способен справиться с подобного рода задачами с требуемой точностью. Для решения этой проблемы были разработаны методы автоматизации обнаружения объектов. Наибольшее применение в данной сфере получили свёрточные нейронные сети.

Цель работы – сравнить архитектуры свёрточных нейронных сетей. Для достижения поставленной цели необходимо решить следующие задачи:

- проанализировать предметную область;
- формализовать задачу генерации признаков;
- перечислить разновидности свёрточных нейронных сетей, сформулировать критерии сравнения;
- сравнить перечисленные разновидности сетей на основании сформулированных критериев.

1 Аналитический раздел

1.1 Основные понятия

Свёрточная нейронная сеть (*СНС*) – это вычислительная модель, предназначенная для эффективного распознавания объектов. Распознавание объекта происходит за счёт выявления присущих ему признаков. Входное данное (изображение, сигнал и т.д.), в котором необходимо определить объект, представляется в виде тензора.

Тензор – это многомерный массив чисел, характеризующийся рангом q [1]. При $q = 0$ тензор – это скаляр ($t \in \mathbb{R}$), при $q = 1$ – вектор ($t \in \mathbb{R}^d$), при $q = 2$ – матрица ($t \in \mathbb{R}^{d_1 \times d_2}$) и так далее.

Веса w свёрточной сети f_w – это числовые коэффициенты, которые отвечают за распознавание признаков объектов. Установка весов происходит на этапе обучения сети на тестовой выборке.

Тестовая выборка $\{(x_i, y_i)\}_{i=1}^N$ – это набор данных где y_i – истинный ответ для входного данного x_i .

Обучение – процесс минимизации разницы между ответом сети f_w при текущих весах w и эталонным ответом из тестовой выборки. Разница описывается функцией ошибки $\mathcal{L}(f_w(x_i), y_i)$, а задача обучения формализуется согласно формуле (1.1):

$$\min_w \frac{1}{N} \sum_{i=1}^N \mathcal{L}(f_w(x_i), y_i). \quad (1.1)$$

Вид функции ошибки зависит от задачи.

Процесс обработки входного тензора свёрточной нейронной сетью состоит из двух этапов:

- 1) определение присущих объекту признаков;
- 2) установление факта принадлежности объекта некоторой группе на основе его признаков.

На рисунке 1.1 показана структура свёрточной нейронной сети.

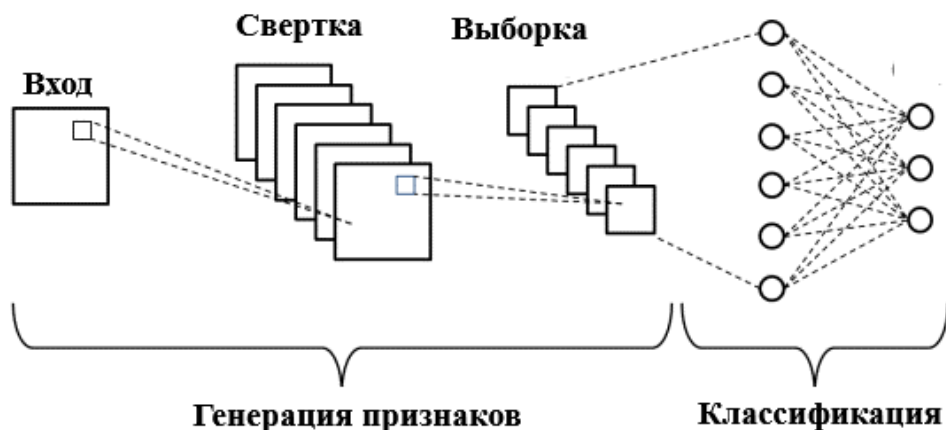


Рисунок 1.1 – Структура свёрточной нейронной сети

1.2 Вехи развития

Эволюция свёрточных нейронных сетей берет свое начало в работе 1959 года нейрофизиологов Визеля Торстена и Дэвида Хьюбела. Учёные обнаружили, что зрительная кора головного мозга содержит сложные клетки. Данный тип клеток активируется при зрительном распознавании некоторой геометрической формы вне зависимости от её положения в обозримой части пространства.

В 1980 году Кунихико Фукусима на основе открытых сложных клеток коры головного мозга построил математическую модель двухслойной системы «Neocognitron», которая могла определять объект на изображении без привязки к его положению на изображении и незначительных изменениях в форме.

В 1990 году Ян Лекун продемонстрировал то, что свёрточные сети, основанные на принципах модели «Neocognitron», способны решать широкий класс задач, связанных с распознаванием изображений. Первой значимой решённой задачей стала классификация рукописных символов.

В 2010-2013 году начало появляться аппаратное обеспечение, которое обладало достаточной для работы нейронных сетей вычислительной мощностью. Появились новые архитектуры свёрточных сетей: *AlexNet* (2012), *VGG* (2014), *ResNet* (2015) и другие.

1.3 Задача генерации признаков

Генерация признаков – процесс извлечения признаков объекта из входного тензора. Процесс генерации признаков состоит в поочерёдном применении операций свёртки и выборки [2].

Свёртка

Свёртка выполняется путем применения тензора фильтра $K \in \mathbb{R}^{k_H \times k_W \times C}$ к входному тензору $X \in \mathbb{R}^{H \times W \times C}$. Веса, составляющие фильтр, устанавливаются в процессе обучения. Фильтр скользит по входному тензору с заданным шагом. На каждом шаге вычисляется скалярное произведение значений фильтра и соответствующих элементов входного тензора согласно формуле (1.2):

$$Y_{i,j,d} = \sum_{c=1}^C \sum_{u=1}^{k_H} \sum_{v=1}^{k_W} X_{i+u-1,j+v-1,c} \cdot K_{u,v,c,d}, \quad (1.2)$$

где $Y \in \mathbb{R}^{H' \times W' \times D}$ – выходной тензор. Каждый фильтр извлекает один признак, представленный на выходе. Рисунок 1.2 демонстрирует пример вычисления выходного тензора.



Рисунок 1.2 – Пример вычисления выходного фильтра

Расчёт значения для текущего фрагмента приведён в выражении (1.3):

$$x = 1 \cdot 1 + 0 \cdot 0 + 0 \cdot 1 + 1 \cdot 0 + 1 \cdot 1 + 0 \cdot 0 + 1 \cdot 1 + 1 \cdot 0 + 1 \cdot 1 = 4. \quad (1.3)$$

Ключевое свойство свёртки – обнаружение признаков без привязки к их положению в тензоре.

Выборка

Задача выборки – понизить размерность тензора, полученного после выполнения свёртки. Данная операция позволяет оставить для последующей обработки только наиболее значимые признаки. Пренебрежение выборкой может привести к переобучению – ситуации, когда признаков становится слишком много, и ошибка в ходе последующей классификации увеличивается.

Критерий выборки зависит от реализации сети, наиболее широко применяется выборка максимума и выборка среднего.

Выборка максимума происходит на основе формулы (1.4):

$$Y_{i,j,d} = \max_{(u,v) \in P} X_{i+u-1,j+v-1,d}. \quad (1.4)$$

Выборка среднего происходит на основе формулы (1.5):

$$Y_{i,j,d} = \frac{1}{|P|} \sum_{(u,v) \in P} X_{i+u-1,j+v-1,d}. \quad (1.5)$$

$P = k_H \times k_W$ – количество элементов в фильтре.

Количество слоёв свёртки и выборки, размеры и количество фильтров, а также критерий для выборки зависят от реализации конкретной СНС.

1.4 Некоторые архитектуры свёрточных сетей

LeNet

Архитектура *LeNet* (1998) стала первой реализованной свёрточной нейронной сетью, показавшей выдающиеся результаты в классификации изображений. Обучение и тестирование на наборе данных *MNIST*, содержащем 60.000 изображений рукописных цифр, показало ошибку предсказания порядка 0.95 %. Данная сеть содержала два слоя свёртки и выборки. В ходе выборки применялась выборка среднего. Архитектура *LeNet* заложила базовые принципы построения свёрточных сетей, на которых были основаны поздние модели [lenet].

AlexNet

Архитектура *AlexNet* (2012) совершила прорыв в области машинного обучения за счёт возможности обработки объемных входных данных. Архитектура основана на использовании 96 фильтров размера 5×5 и 11×11 и выборке максимума на этапе выборки. Использование больших фильтров позволяет определять объемные признаки, за счёт получения большей информации от соседних значений тензора. В данной архитектуре используется пять слоёв свёртки и применяются методы для нейтрализации переобучения [3].

VGG-16

Архитектура *VGG* (2014) выступает антагонистом *AlexNet* в вопросе количества слоёв свёртки и размера фильтров. *VGG* использует 16 слоёв свёртки и фильтры размером 3×3 . Очередной слой свёртки содержит информацию о предыдущих, а информация из первого слоя уточняется с ростом числа слоёв. Данный факт позволил ещё сильнее повысить точность предсказаний, однако потребовал большее количество вычислительных мощностей [4].

ResNet-50

На сегодняшний день *ResNet* (2015) является наиболее совершенной свёрточной нейронной сетью. Впервые именно данная сеть смогла превзойти человека в способности распознавать изображения. Данная сеть имеет 50 слоёв свёртки и порядка 25 миллионов весов. Стоит отметить, что прогресс в

области разработки нейронных сетей идёт в ногу с развитием аппаратного обеспечения. Существование приложений для таких моделей обусловлено возможностью аппаратной поддержки в виде графических процессоров *GPU*, которые оптимизированы для выполнения матричных операций [5].

1.5 Сравнение перечисленных архитектур

Сравнение перечисленных архитектур будет проведено на основе следующих критериев:

- средний процент ошибки из пяти лучших результатов;
- число весов;
- число операций умножения и сложения *MAC*;
- количество слоёв свёртки.

В таблице 1.1 приведена сравнительная таблица описанных методов [6]:

Таблица 1.1 – Сравнение рассмотренных архитектур

Архитектура	Ошибка	Количество весов, млн.	<i>MAC</i> , млн.	Количество слоёв свёртки, шт.
LeNet	23.4%	0.431	2.3	2
AlexNet	15.3%	61	724	5
VGG-16	7.3%	138	15500	16
ResNet-50	3.6%	25.5	3900	50

ЗАКЛЮЧЕНИЕ

В ходе выполнения научно-исследовательской работы были выполнены задачи:

- предметная область проанализирована;
- задача генерации признаков формализована;
- перечислены некоторые разновидности СНС, сформулированы критерии сравнения;
- перечисленные методы СНС сравнены по сформулированным критериям.

Цель работы достигнута: проведено сравнение различных архитектур свёрточных нейронных сетей.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Шарипов Р.* Введение в тензорный анализ. — БХВ, 2004.
2. *Katiyar S.* Comparative evaluation of CNN architectures for image caption generation // IJACSA. — 2020.
3. *Alex Krizhevsky I. S.* ImageNet classification with deep convolutional neural networks // University of Toronto. — 2012.
4. *Karen Simonyan A. Z.* Very deep convolutional networks for large-scale image recognition // ICLR. — 2015.
5. *Kaimig He Xiangyu Zhang S. R.* Deep residual learnin for image recognition // Microsoft research. — 2015.
6. *Patel S.* A comperhensive analysis of convolutional neural network models // Charotar University of Science and Technology. — 2020.

ПРИЛОЖЕНИЕ А

Презентация к научно-исследовательской работе состоит из 3 слайдов.