

СВЁРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ CONVOLUTIONAL NEURAL NETWORKS

Расколотько Владислав Викторович

Свёрточные нейронные сети

Convolutional Neural Networks, CNN

Данные сети являются одним из ключевых инструментов в области глубинного обучения, которые сыграли роль в технологиях компьютерного зрения. CNN были предложены для автоматического выделения признаков на изображениях, и с тех пор они прочно закрепились в большинстве задач, связанных с распознаванием, классификацией и анализом визуальных данных.

Цель данного исследования – рассмотреть архитектуру свёрточных нейронных сетей и объяснить её ключевые особенности, рассказать об областях применения CNN, а также привести конкретные примеры успешных прикладных решений и современные тенденции, касающиеся применения CNN.

Архитектура свёрточной нейронной сети (CNN)

CNN — это класс глубоких нейронных сетей, разработанный для обработки изображений, видео и других пространственно-структурированных данных. Ключевая особенность CNN заключается в том, что они извлекают пространственные признаки из данных с помощью свёрточных операций, что делает их особенно эффективными для анализа изображений.

Основные особенности архитектуры:

Локальные связи и свёртка (Convolution)

В полносвязных нейронных сетях каждый нейрон связан со всеми нейронами предыдущего слоя, что ведёт к значительному увеличению количества параметров при увеличении размера данных (тех же самых изображений). В CNN же используется принцип **локальной рецептивной области**: нейроны каждого свёрточного слоя соединены только с небольшой областью входного изображения.

Ядро свёртки (фильтр): Каждый свёрточный слой имеет набор фильтров (или ядер свёртки), которые "скользят" по изображению и вычисляют свёртку для каждой локальной области. Каждый фильтр выделяет определённые признаки, такие как края, текстуры и формы. Это значительно снижает количество параметров по сравнению с полносвязными слоями.

Математически это можно выразить как:

$$Im^{out}(x, y) = \sum_{i=-d}^d \sum_{j=-d}^d (K(i, j)Im^{in}(x + i, y + j) + b)$$

Где, $Im^{out}(x, y)$ – двумерное изображение на выходе, $K(i, j)$ – ядро свёртки 2×2 , $Im^{in}(x + i, y + j)$ – двумерное входное изображение

Нелинейные активации (ReLU)

После свёртки к выходным данным применяют функцию активации, чтобы внести нелинейность в модель. Самая популярная функция активации в CNN — это ReLU (Rectified Linear Unit), которая задаётся формулой:

$$ReLU(x) = \max(0, x)$$

Она превращает все отрицательные значения в нули, тем самым ускоряя процесс обучения и делая модель менее подверженной проблеме исчезающего градиента, часто встречающейся в нейронных сетях.

Подвыборка (Pooling)

После свёртки обычно выполняется операция подвыборки (Pooling), чтобы уменьшить размер данных и сохранить самые важные признаки. Это также помогает уменьшить вероятность переобучения, делая модель более устойчивой к шумам и незначительным вариациям.

Наиболее распространённый метод — это max pooling, который вычисляет максимум значений в локальной области:

$$y = \max(x_1, x_2, \dots, x_n)$$

Max pooling выбирает наиболее важные признаки, игнорируя менее значимые данные, что снижает размерность и улучшает обобщающую способность модели.

Полносвязные слои (Fully Connected Layers)

После нескольких слоёв свёрток и подвыборки, данные, представляющие высокоуровневые признаки, подаются на полносвязные слои для окончательной классификации или регрессии. Полносвязные слои преобразуют выходные данные в формат, пригодный для предсказания классов.

Регуляризация с помощью Dropout

В современных архитектурах CNN часто используется метод Dropout, чтобы предотвратить переобучение. Dropout случайным образом "отключает" нейроны с определённой вероятностью на этапе обучения, заставляя сеть учиться более устойчивым признакам. Вероятность отключения обычно составляет 50% для полносвязных слоёв.

Идеи, лежащие в основе CNN:

Свёртки позволяют автоматическое извлечение признаков: CNN не требуют явного задания признаков. С помощью слоёв свёрток они автоматически учатся извлекать важные характеристики изображения, начиная от низкоуровневых (границы, углы) до высокоуровневых (формы объектов, текстуры).

Иерархическая структура признаков: Благодаря многоуровневой свёрточной структуре, CNN способны строить иерархию признаков. На начальных слоях извлекаются простейшие признаки (например, контуры), которые затем комбинируются на последующих уровнях для построения более сложных признаков.

Малое количество параметров: По сравнению с традиционными нейронными сетями, CNN имеют гораздо меньше параметров благодаря локальным связям и разделяемым весам фильтров. Это позволяет эффективно обрабатывать данные больших размеров, такие как изображения высокого разрешения.

Пространственная инвариантность: Свёрточные фильтры позволяют CNN быть инвариантными к смещению объектов на изображении, что особенно полезно при распознавании изображений, где важен не конкретный пиксель, а общая форма объекта.

Основные области применения CNN

Свёрточные нейронные сети (CNN) применяются в различных задачах глубинного обучения, особенно там, где требуется анализ структурированных данных, таких как изображения и видео. Рассмотрим основные области применения:

Компьютерное зрение (CV)

Задачи:

Классификация изображений: Определение, к какому классу относится изображение (например, определение породы собаки или классификация рукописных цифр).

Обнаружение объектов (Object Detection): Нахождение и локализация объектов на изображении (например, идентификация автомобилей или пешеходов на дороге).

Сегментация изображений: Разделение изображения на отдельные области или объекты (например, распознавание органов на медицинских снимках).

Примеры:

Классификация с использованием ImageNet: CNN, такие как AlexNet и VGG добились прорыва в задаче классификации изображений на ImageNet. AlexNet впервые смогла существенно превзойти другие алгоритмы, показав высокую точность и автоматическую способность выделения признаков с изображений.

Обнаружение объектов (YOLO, Faster R-CNN): YOLO (You Only Look Once) и Faster R-CNN — это модели, которые применяются для задачи обнаружения объектов. Faster R-CNN использует двухэтапный процесс (обнаружение регионов и их классификация), а YOLO работает в режиме "в один проход", что делает её очень быстрой.

Сегментация изображений (U-Net): U-Net — это архитектура CNN, которая используется для сегментации изображений, особенно в медицине. Она позволяет разделять изображение на объекты и фон, что необходимо, например, для выделения опухолей на МРТ или КТ снимках.

Обработка естественного языка (NLP)

Задачи:

Классификация текстов: Определение темы или тональности текста (например, классификация новостных статей или отзывов).

Извлечение сущностей: Выделение ключевых объектов в тексте (например, имен, дат, организаций).

Анализ последовательностей: Извлечение паттернов из текстов и их классификация.

Примеры:

Классификация текста (TextCNN): CNN используются для анализа текста, например, для классификации отзывов по тональности (позитивные или негативные) или для выявления ключевых сущностей. Например, TextCNN применяют свёртки для анализа последовательностей символов и слов.

Актуальность CNN

Компьютерное зрение (CV)

Актуальность:

CNN остаются очень актуальными в задачах компьютерного зрения, таких как классификация изображений, обнаружение объектов и сегментация. За последние годы они стали основой большинства коммерческих и научных решений в этой области.

Однако, начиная с 2020-х годов, в некоторых задачах появились альтернативы, например, Vision Transformers (ViT). Эти модели, основанные на архитектуре трансформеров, изначально разработанной для обработки текста, показали высокую точность в задачах компьютерного зрения, часто превосходя CNN.

Причины этого перехода:

Лучшее внимание к глобальным связям: Vision Transformers лучше справляются с захватом глобальной структуры изображения, что может быть проблемой для CNN, которые больше ориентированы на локальные признаки.

Снижение зависимости от свёрточных операций: ViT работают без свёрток, опираясь на механизм **самовнимания (self-attention)**, что позволяет лучше обрабатывать изображения с высокой разрешающей способностью.

Тем не менее, CNN остаются стандартом для большинства коммерческих приложений, особенно там, где важна эффективность и производительность на более простых устройствах.

Обработка естественного языка (NLP)

Актуальность:

В области обработки естественного языка (NLP) CNN уступили место другим архитектурам, в частности, трансформерам. Например, модели на основе BERT, GPT и других трансформеров доминируют в большинстве задач NLP.

Причины:

Трансформеры лучше справляются с длинными последовательностями: В отличие от CNN, которые работают с фиксированной длиной данных, трансформеры могут лучше захватывать долгосрочные зависимости в тексте благодаря механизму внимания.

Гибкость обработки текста: CNN ограничены размером окна свёртки, тогда как трансформеры способны эффективно моделировать взаимосвязи на любом уровне абстракции.

CNN ещё используются в некоторых узко специализированных задачах, таких как классификация коротких текстов или анализ тональности, но их роль в NLP заметно сократилась с ростом популярности трансформеров.

ИСТОЧНИКИ ИНФОРМАЦИИ

1. Расколотько В.В Лекция 2. Свёрточные сети / Расколотько В.В // Личный конспект курса по глубинному обучению ФКН НИУ ВШЭ. — Москва, 2024.
2. Eugenio Culurciello Neural Network Architectures / Eugenio Culurciello [Электронный ресурс] // towardsdatascience : [сайт]. — URL: <https://towardsdatascience.com/neural-network-architectures-156e5bad51ba> (дата обращения: 02.10.2024).
3. Сверточная нейронная сеть, часть 1: структура, топология, функции активации и обучающее множество / [Электронный ресурс] // Habr : [сайт]. — URL: <https://habr.com/ru/articles/348000/> (дата обращения: 02.10.2024).
4. Иванов Г. Свёрточные нейросети / Иванов Г. [Электронный ресурс] // Яндекс.образование : [сайт]. — URL: <https://education.yandex.ru/handbook/ml/article/svyortochnye-nejroseti> (дата обращения: 02.10.2024).