**Тема 3.9 Современные тренды в решение задач компьютерного зрения**

**План**

1. Тренды
2. Метод NAS
3. Аугментация и кросс-аугментация
4. Мета-обучение
5. Исследование методов нормализации
6. Архитектуры трансформеры
7. Достоинства архитектур-трансформеров

**Тренды**

Большинство имеющихся фундаментальных результатов в области глубокого обучения сверточных нейронных сетей в компьютерном зрении были получены в период с конца 1990-х до 2015 годах. В этот период была разработана архитектура современных слоев и их блоков, а также методы регуляризации, типы функций активации скрытых слоев, методы оптимизации и так далее.

На **сегодня можно выделить такие тренды, как:**

* Использование автопоиска архитектур NAS и AutoML
* Выбор методов регуляризации нейронной сети.
* Оптимизация стратегий обучения глубоких нейронных сетей.
* Использование не сверточных слоев или их комбинаций со сверточными (например, слоев трансформеров).
* А также трендом является разработка методов повышения качества набора данных. В конце концов все сводится не только к архитектуре, но и к данным, на которых мы обучаем архитектуру.

**Метод NAS**

Метод NAS сегодня используется для широкой оптимизации архитектуры. Для использования NAS, как правило известны желаемые подходы, но необходимо найти их лучшую комбинацию. Эту комбинацию можно искать как для отельного набора данных, так и для отдельного устройства. Сегодня метод NAS не является решением всех проблем или универсальным инструментом. Результаты работы NAS ограничены опытом и вычислительными ресурсами исследователя.

**Аугментация и кросс-аугментация**

Аугментация данных является одним из методов регуляризации обучения нейронной сети. В сущности же аугментация решает проблему недостаточности разнообразных данных. То есть аугментация это попытка повысить точность и обобщающую способность в ходе обучения, показав сети признак во всем его многообразии. Основная опасность аугментации данных – аугментация не должна создавать дисбаланс или новые признаки, а также искажать распределение существующих признаков. В некоторых случаях также используется **кросс-аугментация** – то есть аугментация путем комбинирования изображений набора данных и их меток.

В ряде подходов к обучению архитектур используется **сглаживание меток**, то есть целевому классу приписывается значение меньше единицы, а остаток размывается среди других классов. Такие подходы также могут способствовать регуляризации. Этот метод также можно назвать аугментацией меток. Такая аугментация может быть как стохастической, так и детерминированной.

**Мета-обучение**

В ряде источников литературы сегодня рассматривается мета-обучение. Мета-обучение - это не прямой подход к обучению целевой архитектуры. Наиболее распространённый пример мета-обучения – это дистилляция знания. Подход предполагает две сети, большую сеть учитель и целевую сеть – поменьше – сеть ученик. В ходе обучения большая сеть учитель создает свою мета-разметку данных. А сеть ученик старается повторить поведение учителя. Таким образом создается архитектура меньшего объема, позволяющая получать такие же результаты, как и сеть учитель. Могу быть и другие подходы, например, подход шумящий студент позволяет повысить точность для случая, когда учитель только мешает обучению.

**Исследование методов нормализации**

Другим трендом в настоящее время является исследование методов нормализации. Один из наиболее популярных методов регуляризации сегодня это батч-нормализация. Однако, этот метод имеет ряд существенных недостатков, которые не позволяют эффективно использовать батч-нормализацию для переобучения нейронной сети на других размерах батча или размерах изображения. Среди альтернатив батч-нормалиазции этих методов наиболее популярен **метод нормализации слоя**. В этом методе мы вычисляем параметры нормализации для каждого слоя, а не батча. Причем нормализация слоя выполняется как на этапе обучения, так и на этапе вывода. Также в ряде случаев нормализацию данных дополняют или замещают **нормализацией, ограничением нормы или стандартизацией весовых параметров нейронных сетей.**

Другим примером альтернативы батч-нормализации является **архитектура NF-Net**. В этом подходе авторы предложили набор техник, заменяющих батч-нормализацию, в том числе нормирующие множители, ограничение нормы весовых параметров и нормализацию весовых параметров. Подход позволил показал точность выше, чем EfficientNet который уже рассматривался нами как пример наиболее эффективной сети.

**Архитектуры трансформеры**

В 2020 году была предложена первая действительно эффективная нейронная сеть без использования операций свертка. Сеть была основана на идеи архитектуры-трансформера из области обработки естественного языка. Эта архитектура **называется visual transformer (ViT)**. В подходе VIT было предложено разделить изображение на ряд частей – так называемых патчей. Каждый патч помечается и трансформируется при помощи слоя **патч-эембединга**. Затем помеченный патч векторизуется и обрабатывается при помощи блоков само внимания и полно связных сетей. Подход VIT достиг весьма высокой точности на наборе данных ImageNet. Однако для этого авторам пришлось предобучить сеть на наборе из порядка 300 миллионов изображений. Иначе такой точности было не достичь.

Подход VIT открыл новый путь решения задач компьютерного зрения. Сегодня можно найти достаточно большое число архитектур трансформеров для решения различных задач компьютерного зрения, как задач классификации, так и задач обнаружения объектов или сегментации, а также генерации объектов.

**Достоинства архитектур-трансформеров**

Достоинства архитектур-трансформеров возможность построения сложных зависимостей между признаками локальными участков изображения независимо от их взаимного расположения. Архитектуры трансформеры позволяют использовать путь повышения точности и обобщающей способности за счет широкомасштабного предобучения. Также некторые архитектуры трансформеры дают возможность предобучения на не размеченных данных. Однако, архитектуры трансформеры куда более избыточны и требуют больших рессурсов для обучения и работы. Эти проблемы остаются актуальными и сегодня.

**Комбинация подходов: сверточные, полносвязные или трансфомеры слои**

Часть предложенных в литературе архитектур- трансформеров полностью полносвязные, а часть имеют и сверточные, и полносвязрые слои. Например, CoAtNet является одним вариантом реализации сети ResNet, где сверточных слоев заменены на слои многоголового само-внимания. При этом авторы получили прирост точности % по сравнению с оригинальной сетью.

Вопрос о том, какой подход или комбинация лучше подходят для решения задач компьютерного зрения остается открытым. Трансформеры быстрее и более точно выучивают крупноразмерные признаки. Однако, архитектурам-трансформерам требуется больше примеров для мелкоразмерных признаков.

**Архитектуры миксеры**

Также ряд авторов предполагают, что слои внимания вообще не нужны, а нужна некоторая перетасовка карт признаков. В 2021 году был представлен ряд трансформеров-подобных полносвязных архитектур для решения задач компьютерного зрения, такие архитектуры называют миксеры. Например, подход **MLP-Mixer** предлагает заменить операцию внимания на транспонирование патчей. Основное преимущество архитектур миксеров — это пониженное число параметров по сравнению с арихтектурами трансформерами. Это получается из-за отсутствия блока многоголового само-внимания. Также архитектуры миксеры имеют более высокую скорость работы так как операции с полносвязными слоями - самые быстрые. В различных архитектурах миксерах могут быть разные механизмы перетасовки признаков. На сегодня миксеры показывают точности ниже классических трансформеров. Исследования в этом направлении еще ведутся.

**Резюме**

Тренды сегодняшнего дня показывают возможности и востребованность использования глубокого обучения нейронных в задачах компьютерного зрения. Сегодня сложно сказать куда приведут эти вектора развития данную область знаний. Мы надеемся, что данный курс позволит слушателю не потеряться в этих реалиях и оставаться в курсе последних событий в изученной области знаний.