# Механизм внимания в компьютерном зрении

**Attention** (внимание) - это операция, которая рассчитывает релевантность, а затем агрегирует входные данные в соответствии с этой релевантностью.

Attention часто называют **умным пулингом**. Потому что мы не просто усредняем по фичам (или находим максимум), а хитро взвешиваем, и веса у нас каждый раз разные. Таким образом получаем больше вариативности.

Attention может быть локальным и глобальным, может применяться к разным фичам (к каналам, к эмбеддингам токенов и др.). Но, по сути, мы выполняем взвешивание скорами, которые меняются в зависимости от входных данных.

Историю развития механизмов внимания можно условно разделить на четыре периода:

1. Первый период начинается с архитектуры [RAM](https://arxiv.org/pdf/1406.6247.pdf), в которой глубокие сети комбинировали с механизмом внимания. Механизм внимания рекуррентно предсказывал “важные” участки изображений и обновлял веса сети с помощью обучения с подкреплением. В этот период для реализации механизмов внимания всегда использовали RNN.
2. Второй период начался с архитектуры [STN](https://arxiv.org/pdf/1506.02025.pdf), в которой подсеть предсказывает аффинные преобразования для локализации (за счет растяжения и обрезки) и трансформации (сдвиг, поворот) “важной” области изображения. Затем через сеть пропускается только преобразованная важная область входной картинки. Главной отличительной особенностью второго периода является то, что важные области входных данных предсказывались отдельной подсетью, которая училась вместе с классификатором.
3. Третий период начался с [SENet](https://arxiv.org/pdf/1709.01507.pdf), которая предложила поканальный вариант механизма внимания. В нем адаптивно выделяются потенциально важные признаки. SE-блоки используются, например, в EffNet и MobileNet архитектурах.
4. Четвертый период – это эпоха внутреннего внимания, он начался с адаптации работы “[Attention is all you need](https://arxiv.org/pdf/1706.03762.pdf)” к работе с изображениями, породив архитектуры под общим названием *visual transformers*.

Механизмы внимания в компьютерном зрении можно разделить на следующие основные категории:

1. поканальный механизм внимания (Channel attention - НА ЧТО обращать внимание);
2. пространственный механизм внимания (Spatial attention - КУДА обращать внимание);
3. временной (эволюционный) механизм внимания (Temporal Attention - КОГДА обращать внимание);
4. блочный/веточный механизм внимания (Branch attention - ЧТО должно обратить внимание).

Также существуют гибридные архитектуры, которые являются представителями нескольких категорий механизмов внимания одновременно: поканально-пространственный (сверточный модуль внимания) и пространственно-временной механизмы.

**Self-Attention** (внутреннее внимание) — разновидность механизма внимания, задачей которой является выявление закономерности только между входными данными.

Подход Attention применяется для работы с текстом, а также звуком и временными рядами. Для обработки текста широко используются рекуррентные нейронные сети (RNN, LSTM, GRU). Attention может либо дополнять их, либо заменять их, переводя сеть к более простым и быстрым архитектурам. Одно из самых известных применений Attention это применение его для того, чтобы отказаться от рекуррентной сети и перейти к полносвязной модели.

По материалам статей:

Attention Is All You Need (<https://arxiv.org/pdf/1706.03762>);

Attention Mechanisms in Computer Vision: A Survey (<https://arxiv.org/pdf/2111.07624>);