**Тема 2.7 Идеи внимания и слой Squeeze and Excitation**

**План**

1. Идея внимания
2. Идея самовнимания
3. Слой Squeeze-and-Excitation layer
4. Архитектура SE ResNet

**Идея внимания**

Идея «подсвечивания» для нейронной сети наиболее информативных участков входных данных (информативных с точки зрения задачи) присутствовала в сообществе исследователей нейронных сетей начиная с 1990-х. В 2014 Бахданау (Bahdanau) предложил использование данной идеи в задачах машинного перевода, предложенная архитектура была названа механизм внимания (attention). Идея механизма внимания получила широкое развитие и распространение в различных приложениях обработки естественного языка и других схожих задачах. В 2016–2017 годах была развита идея механизма само-внимания (self-attention) для задач обработки естественного языка. Также были предложены т. н. много-головое внимание (multi- head attention) и блок трансформер (transformer).

**Идея самовнимания**

Начиная с 2017 года идеи механизма внимания активно исследуются в приложениях к сверточным нейронным сетям в различных задачах компьютерного зрения. Отметим, что в противовес к само-вниманию, классический механизм внимания называют кросс-внимание (cross-attention). Основной идеей механизма внимания является обучение слоя (блока) выделять наиболее важные участки входных данных при помощи нормализации функцией softmax. В случае само-внимания слой обучается только на основе входных данных. Если речь идет о кросс-внимание то слой обучается на входных и уже предсказанных выходных данных.

Показано, что слой самовнимания позволяет провести перегруппировку пикселей входного слоя – то есть делает результат работ нейронной сети независимым от нерегулярных особенностей каждого экземпляра входных данных. Другими словами, слой внимания дополняет одно из наиболее важных свойств сверки – инвариантность к положению объекта (как по координатам, так и по наклону, повороту и т. д.).

**Слой Squeeze-and-Excitation layer**

В 2017 году концепция внимания для сверточных сетей был значительно переработана. В был предложено слой сжатия-возбуждения (Squeeze-and-Excitation layer, SE layer). В основе работы блока лежит идея выделения наиболее важных карт признаков (каналов) путем их кодирования и декодирования со взвешиванием через фикцию активации сигмоид. В результате такой операции формируются коэффициенты, характеризующие «важность» каждой карты признаков. Таким образом, SE Layer призван помочь нейронной сети принимать во внимание только наиболее важные (регулярные, релевантные задаче) признаки. Отметим, что строго говоря слой SE Layer не является блоком внимания в классическом смысле так как использует логистическую функцию активации для каждого выхода вместо единой нормализации функцией softmax. Однако, часто слой SE Layer рассматривают в качестве варианта межканального внимания.

**Архитектура SE ResNet**

В 2017 году команда авторов архитектуры SE ResNet одержали победу в соревнования по классификации ILSVRC 2017 с ошибкой 2,25 (4,5) % по методике top-5-accuray.

**Резюме**

Слой SE позже был использован во многих успешных архитектурах сверточных нейронных сетей. В частности, с использование слоя SE были синтезированы блоки MobleNet V3 в 2019 году.