

Авторы статьи предлагают модификацию ResNet, которая направлена на уменьшение потребления памяти без потери качества. Для этого предлагается в residual блок добавить momentum. Это позволяет сделать residual блок обратимым - т.е. активации блока можно вычислить по активации следующего блока, что позволяет не хранить активации, а вычислять динамически во время backpropagation. Таким образом достигается значительный выигрыш по памяти - Momentum ResNet практически константен по памяти, в то время как обычный ResNet имеет линейный рост памяти в зависимости от глубины.

Плюсы статьи:

- 1) Все утверждения авторов о преимуществах оформлены в виде чётко сформулированных теорем и подробно доказаны в приложении
- 2) Блок Momentum ResNet может использоваться в качестве drop-in замещения обычного residual блока
- 3) Авторы демонстрируют возможности применения модели в других задачах, к примеру в оптимизации

Минусы статьи:

- 1) Авторы не проводят экспериментов на сетях меньше ResNet-101, и как показывают наши эксперименты, для моделей меньшей глубины Momentum ResNet не улучшает (а иногда и ухудшает) потребление памяти
- 2) В статье очень много теории, что осложняет её чтение и понимание

В целом статья хорошая и предлагает значительные улучшения, и поэтому я считаю заслуживает оценки 8.