

СВЕРТОЧНАЯ НЕЙРОННАЯ СЕТЬ В ЗАДАЧАХ ОБРАБОТКИ ИЗОБРАЖЕНИЙ

Адамов Н. А., Шилин Л. Ю., Навроцкий А. А.

Кафедра информационных технологий автоматизированных систем,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: adamov.nikooo@gmail.com, dekgfitu@bsuir.by

В докладе рассмотрены подходы, которые позволяют увеличивать производительность сверточных нейронных сетей путем комбинирования различных эвристик.

ВВЕДЕНИЕ

Современные нейронные сети отличаются высокой производительностью и гибкостью. Высокое качество обуславливается применением различных подходов, которые позволяют значительно увеличить глубину сети, что, в свою очередь, напрямую влияет на конечный результат.

I. МЕТОДЫ ПОВЫШЕНИЯ КАЧЕСТВА ОБУЧЕНИЯ

Затухающий градиент при большой глубине нейронной сети, является проблемой т.к. не позволяет полноценно проводить обучение. Коэффициенты обучения во время обратного распространения ошибки имеют тенденцию к уменьшению, что влечет за собой незначительное изменение в весах нейронной сети и не позволяет полноценно обучать сеть. Данную проблему может решить технология, которая использовалась в сети ResNet на соревновании ImageNet по классификации объектов на 1000 классов. Данная технология называется остаточным обучением. На рисунке 1 представлен график производительности сети в зависимости от глубины. Слева – обычная сеть, справа – сеть ResNet. Обычная сеть показывает меньшую ошибку при меньшем количестве слоев в сети, однако сеть ResNet показала меньшую ошибку при большем количестве слоев, что свидетельствует об эффективности использования остаточного обучения в глубоких сетях.

Для повышения скорости сходимости градиента используются различные алгоритмы оптимизации. Алгоритм Adam позволяет найти глобальный минимум функции потерь, тем самым уменьшая ошибку модели на данных. На рисунке 2 представлены результаты тестирования алгоритмов оптимизации на основе градиентов: слева SGD, справа Adam с добавлением остаточного обучения (RL) и пакетной нормализации (BN),

а также без этих параметров. Алгоритмы SGD и Adam позволяют сети с остаточным обучением и пакетной нормализацией получать хорошие результаты. Остаточное обучение выигрывает в сравнении с пакетной нормализацией. Это связано с тем, что пакетная нормализация снижает для SNN значимость проблемы внутреннего ковариационного сдвига. Остаточное обучение без пакетной нормализации имеет высокую сходимость, однако уступает остаточному обучению с пакетной нормализацией. С другой стороны, пакетная нормализация выигрывает при использовании остаточного обучения. Без остаточного обучения пакетная нормализация имеет определенное неблагоприятное влияние на конвергенцию. При остаточном обучении пакетная нормализация может быть использована для ускорения обучения и повышения производительности. Без остаточного обучения интенсивность входных данных и сверточный признак коррелируются с соседними, а распределение входных данных слоя зависит от содержимого изображений в каждой обучающей батче. Остаточное обучение позволяет улучшить пакетную нормализацию путем уменьшения внутреннего ковариационного сдвига.

Интеграция различных подходов, например, остаточного обучения и пакетной нормализации может не только ускорить и стабилизировать процесс обучения, но и повысить эффективность модели.

СПИСОК ЛИТЕРАТУРЫ

1. Karen Simonyan, Andrew Zisserman, 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
2. Geoffrey Hinton, Nitish Srivasta, Kevin Swersky. 2012. «Lecture 6a Overview of Mini - Batch Gradient Descent) www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf. Accessed 21 Mar. 2023.

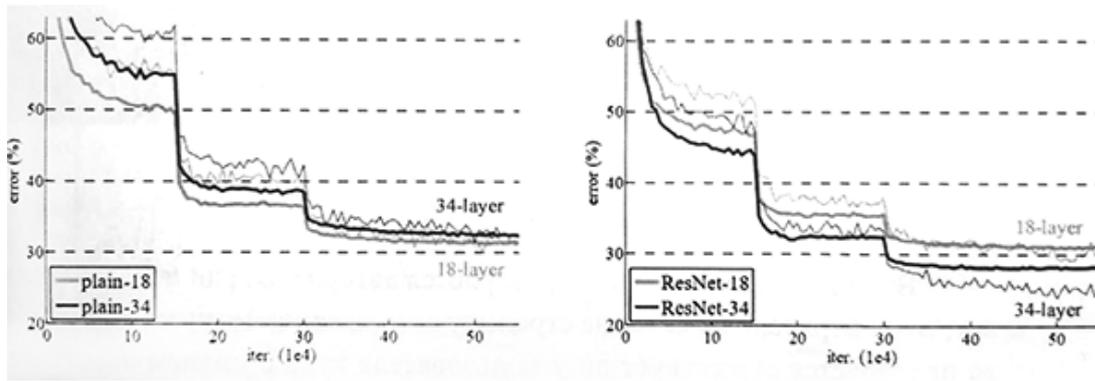


Рис. 1 – Производительность сетей

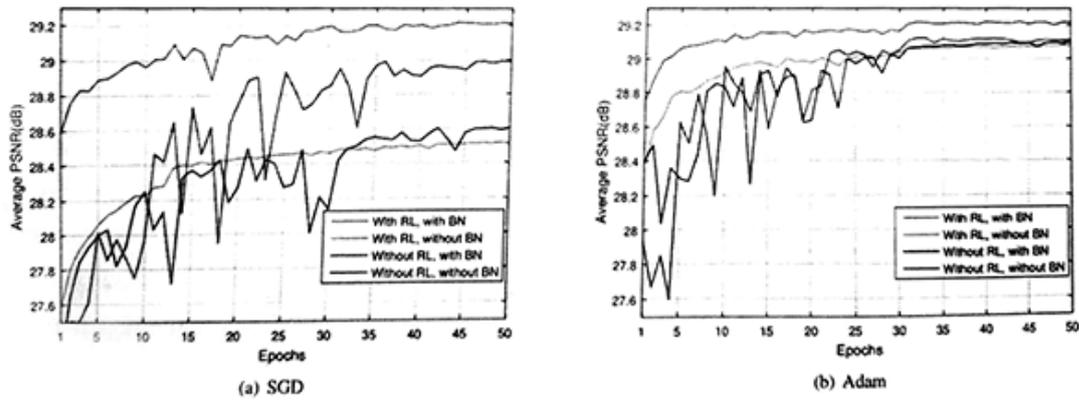


Рис. 2 – Использование алгоритмов оптимизации