

Название статьи (авторы статьи): **When Vision Transformers Outperform ResNets without Pre-training or Strong Data Augmentations**

(Xiangning Chen, Cho-Jui Hsieh, Boqing Gong)

Автор рецензии: Никита Андреев

1. В статье предлагается применение оптимизатора SAM (sharpness aware minimization) к двум популярным архитектурам, а именно ViT (трансформер для зрения) и MLP-mixer (архитектура для зрения, которая основана исключительно на линейных слоях). Авторы решают применить данный оптимизатор исходя из того, что вышеупомянутые модели зачастую имеют очень “острый” локальный минимум при сходимости, а потому используя сглаживание с помощью SAM получают улучшения в качестве результатов работы моделей.
2. Сильные стороны статьи:
  - a. Статья не очень сложная, хорошо читается, проблем с пониманием содержания не возникает
  - b. Авторы используют SAM и объясняют почему именно он подходит для улучшения обучения моделей
  - c. Также в статье есть теоретический обзор SAM, что важно, так как это единственная теория, которая есть в статье, и наличие формул важно для понимания происходящего
  - d. Статья больше про эксперименты, поэтому их проведено большое количество, все грамотно описано, понятные таблицы, результаты удобно читаемы
  - e. Также в конце статьи есть ablation study, где рассказано про contrastive learning, adversarial learning, а также изменение размера выборки и поведение моделей с SAM в этих контекстах. Очень здорово, что авторы уделили этому время и провели дополнительные эксперименты

### 3. Слабые стороны:

Особо проблем в статье я не обнаружил, основные недочеты:

- a. В статье не предлагается ничего принципиально нового – существующий метод SAM применяется для двух архитектур, и статья построена вокруг экспериментов, однако эти эксперименты по применению и есть суть статьи
- b. В самой статье нет результатов для больших модификаций моделей (например ViT-L, ResNet-152x4) однако авторы объясняют это ограничением ресурсов

Статья написана хорошо, понятным языком, не возникает проблем с пониманием содержания и восприятием.

Также есть выложенные чекпоинты моделей на гитхабе, поэтому, как мне кажется, проблем с воспроизведением статьи быть не должно, так как в статье все описано грамотно и подробно, чтобы завести модели.

Оценки по критериям НИПСa:

Оценка 8, уверенность 4.