

Название статьи: [Momentum Residual Neural Networks](#)

Авторы статьи: Michael E. Sander, Pierre Ablin, Mathieu Blondel, Gabriel Peyre

Автор исследования: Каратаева Екатерина

О самой статье

Статья впервые была представлена 15 февраля 2021 года на arXiv, с тех пор статью два раза корректировали. Финальная версия была представлена на 38-ой конференции ICML (International Conference on Machine Learning), которая проходила с 18-24 июля онлайн, и была выложена на arXiv 22 июля 2021.

Статья представлена на конференции в виде [poster](#) и [spotlight](#).

[Здесь](#) вы найдете запись с выступлением на ICML одного из авторов данной статьи – Michael E. Sander.

Как было написано ранее, статья появилась на arXiv 15 февраля 2021, спустя месяц эту статью переписали и выдали за свою Duo Li, Shang-Hua Gao [“m-RevNet: Deep Reversible Neural Networks with Momentum”](#) ([статья](#)) на конференции ICCV (International Conference on Computer Vision), дедлайн подачи на которую был 17 марта 2021. Переписанную статью впервые выложили мошенники на arXiv 12 августа 2021 года, поэтому судя по твиттеру о плагиате узнали совсем недавно. Причем эти авторы участвовали в плагиате ранее (CVPR2020). По [ссылке](#) приведены доказательства в плагиате. По итогу статью отменили, виновные извинились.

Об авторах

Первый автор статьи – [Michael E. Sander](#). Является Ph. d. студентом в ENS (Ecole Normale Supérieure), DMA (Département de Mathématiques et applications). Согласно Википедии ENS – самое престижное высшее учебное заведение во Франции. А также Michael является сотрудником CNRS – Национальный центр научных исследований во Франции.

Данная статья является первой его публикацией. Согласно твиттеру его коллег, он работал над ней усердно год.

Второй автор статьи – [Pierre Ablin](#). Является [постдоком](#) в ENS DMA, а также связан с CNRS. В отличие от Michael E. Sander у Pierre Ablin насчитывается около 16 статей, начиная с 2015 года. В основном его

работы связаны с методами оптимизаций и теоретическими свойствами нейронных сетей.

Третий автор статьи – [Mathieu Blondel](#). Является senior research scientist at Google Research, Brain team во Франции. Интересуется дифференциальным программированием и методами оптимизаций. Тоже опытный researcher, он автор около 25 работ, и многие из них связаны с дифференциальным программированием.

И последний автор – [Gabriel Peyre](#). Является senior researcher в CNRS и работает в ENS (DMA). Самый опытный из всех авторов, он является основным автором или соавтором около 188 статей (в том числе и книг) с 2003 года. Список его интересов очень широк. Во многих статьях он выступает консультантом.

Я практически не нашла никакой информации в английском интернете об интересах главного автора статьи – Michael E. Sander. Мое предположение о создании статьи Momentum Residual Neural Networks: Michael заинтересовался способами улучшений ResNet, изучил существующие решения: RevNet, i-RevNet, Neural ODE. И со своими идеями обратился к более опытным коллегам из ENS и CNRS, работающим в области дифференциальных вычислений и методов оптимизаций. Этим можно объяснить математическую подкованность статьи Momentum Residual Neural Networks. Авторы не только эмпирически доказывают эффективность предложенного метода, но и строго теоретически.

Об основных статьях (на кого ссылаются)

Так как данная статья об улучшении ResNet, то в качестве первой важной работы стоит выделить статью, без которой ничего бы не было – [Deep Residual Learning for Image Recognition](#).

Следующая важная статья про [RevNet](#).

Идея об обратимых архитектурах: т.е. моделях, которые позволяют делать backprop без хранения активаций – не нова. Их можно разделить на два подхода: дискретный. Этот подход заключается в поиске правила, которое позволило из $(n+1)$ -ой активации получить n -ую. RevNet как раз про это. Но минус этого подхода, по сравнению с Momentum ResNet, в том что не совсем очевидно, как из модели сделать сделать “обратимую” без потери качества и скорости.

Подход, основанный на momentum, же легко можно применить к любым моделям, где есть residual блоки.

Статья [Neural Ordinary Differential Equations](#):

Следующий подход заключается в том, чтобы рассматривать ResNet как непрерывную систему и применять к ней математический аппарат ОДУ (обыкновенных дифференциальных уравнений). Получаем константное потребление памяти, но метод сложно применять к нейронным сетям и оптимизировать, особенно из-за того, что точность решения зависит от накопленной численной ошибки.

Я выделила эти две статьи, потому что авторы Momentum ResNet на протяжении всей своей статьи не раз обращаются к перечисленным выше статьям и доказывают эмпирически и теоретически превосходство своего метода.

О том, кто ссылается

[HeunNet: Extending ResNet using Heun's Methods](#) - предлагают улучшение ResNet, основанное на Heun's методе для решения ОДУ.

[Neural ODE control for classification, approximation and transport](#) - провели исследование Neural Ordinary Differential Equations (NODEs).

Применение в индустриальных приложениях

Сложность внедрения NN в индустриальных приложениях связана с тем, что модели NN требуют много ресурсов. Предложенный метод в статье Momentum ResNet, который урезает количество потребляемой памяти.