

# Momentum Residual Neural Networks

Исследование вокруг статьи  
Барановская Дарья

У статьи [1] есть три опубликованные версии: 15 февраля, 13 мая и 22 июля 2021, она была представлена в виде постера на ICML 2021. Статью писали 4 автора, основным из которых был Michael Sander - PhD студент École Normale Supérieure. Данная работа является первой публикацией автора, однако совсем недавно вышла вторая публикация «Sinkformers: Transformers with Doubly Stochastic Attention» от все той же группы исследователей, которая хоть и не является продолжением предыдущей, однако тоже связана с оптимизацией существующих известных методов. Вторым автором Pierre Ablin - исследователь из CNRS, имеющий 23 публикации и 248 цитирований. В авторах также указаны Matieu Blondel - сотрудник Brain team в Google Research и один из создателей scikit-learn и Gabriel Peyré - сотрудник Ecole и CNRL, у которого на счету 371 статья и 12882 цитирования.

Понятно, что основная архитектура ResNet, также как и оптимизация Momentum не новы, однако авторам удалось успешно совместить две идеи. Ранее momentum уже применялся к различным архитектурам: например, для оптимизации [2], для простой RNN [3], для линейных слоев [4], однако все эти архитектуры не обратимы, в отличие от Momentum Resnet. При этом также существует обратимая RevNet [5] с похожей идеей, однако уступающая в удобстве архитектуры и качестве.

Работа цитируется 12 раз на Google Scholar, при чем некоторые цитирования - продолжения работы авторов или исследования качества Momentum Resnet, но уже от других лабораторий. Так, например, заинтересованным в данной теме могут понравиться статьи [6], [7]

В будущем было бы интересно посмотреть, можно ли заменить momentum на какие-то более сложные оптимизации, например, Adadelta, Adagrad и Adam и какой от этого будет эффект.

Рассказывая про данную статью, нельзя не упомянуть скандал с плагиатом. 12 августа 2021 впервые вышла статья с названием m-RevNet: Deep Reversible Neural Networks with Momentum от Duo Li и Shang-Hua Gao с максимально похожим текстом и идеей с точностью до цветов графика и переобозначения переменных. Статья была подана на ICCV, однако Duo Li обвинили в плагиате после того как авторы оригинальной статьи опубликовали доказательства плагиата. Duo Li признался в плагиате, а второй автор Shang-Hua Gao написал

пост с извинениями, сказал, что Duo Li прислал ему pdf статью, попросил комментарии. В ответ Shang-Hua Gao порекомендовал добавить пару вещей, а впоследствии был добавлен в соавторы, хотя его вклада в статью практически не было. После, когда он увидел обвинения в плагиате, то написал Li, который не дал ему однозначного ответа. Shang-Hua Gao отозвал статью с ICCV, предупредил комиссию о плагиате, и теперь извиняется что не сделал достаточное исследование о подобных существующих работах и не предотвратил всю эту историю.

#### References:

- [1] Sander, M. E., Ablin, P., Blondel, M., and Peyr e, G. Momentum residual neural networks. In Proceedings of the 38th International Conference on Machine Learning, volume 139 of Proceedings of Machine Learning Research, pages 9276–9287. PMLR. 2021
- [2] Chun, I. Y., Huang, Z., Lim, H., and Fessler, J. Momentum-net: Fast and convergent iterative neural network for inverse problems. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.
- [3] He, K., Fan, H., Wu, Y., Xie, S., and Girshick, R. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9729– 9738, 2020.
- [4] Li, H., Yang, Y., Chen, D., and Lin, Z. Optimization algorithm inspired deep neural network structure design. In Asian Conference on Machine Learning, pp. 614–629. PMLR, 2018.
- [5] Gomez, A. N., Ren, M., Urtasun, R., and Grosse, R. B. The reversible residual network: Backpropagation without storing activations. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc., 2017.
- [6] B. Wang, H. Xia, T. Nguyen, S. Osher How Does Momentum Benefit Deep Neural Networks Architecture Design? A Few Case Studies. 2021
- [7] D. Ruiz-Balet, E. Affili, E. Zuazua. Interpolation and approximation via Momentum ResNets and Neural ODEs. 2021