

# GradInit: Learning to Initialize Neural Networks for Stable and Efficient Training

Практик-исследователь: Михненко Наталья

## Публикация

Статья написана в феврале 2021 года и представлена постером на конференции NeurIPS 2021

## Авторы

Всего у статьи 6 авторов, 4 из Мэрилендского университета и 2 из Google Research. Все студенты аспирантуры Мэрилендского университета учатся под руководством Тома Голдштейна.

Исследования Тома Голдштейна лежат на стыке машинного обучения и оптимизации и нацелены на приложения в области компьютерного зрения и обработки сигналов. В числе достижений – в 2012 году получил премию Ричарда ДиПрима, Young Faculty Award (от управления перспективных исследовательских проектов Министерства обороны США) и стипендию Слоана.

Ронни Хуан – ученый-исследователь, специализирующийся на создании надежных и обобщаемых алгоритмов обучения, применяемых к данным речи и зрения.

Всех авторов объединяет интерес к алгоритмам оптимизации и машинному обучению.

## Конкуренты

| Model<br>(# Params)     |              | VGG-19<br>w/o BN<br>(20.03M)     | VGG-19<br>w/ BN<br>(20.04M)      | ResNet-110<br>w/o BN<br>(1.72M)  | ResNet-110<br>w/ BN<br>(1.73M)   | ResNet-1202<br>w/ BN<br>(19.42M) |
|-------------------------|--------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
| Kaiming                 | $Acc_1$      | $29.1 \pm 1.5$                   | $12.6 \pm 0.6$                   | $16.1 \pm 2.1$                   | $23.2 \pm 0.9$                   | $12.9 \pm 2.8$                   |
|                         | $Acc_{best}$ | $94.5 \pm 0.1$                   | $94.4 \pm 0.1$                   | $94.2 \pm 0.1$                   | $95.0 \pm 0.2$                   | $94.4 \pm 0.6$                   |
| +1 epoch<br>(Const. LR) | $Acc_1$      | $37.2 \pm 1.1$                   | $19.6 \pm 4.0$                   | $21.0 \pm 3.8$                   | $32.5 \pm 3.8$                   | $12.6 \pm 2.8$                   |
|                         | $Acc_{best}$ | $94.4 \pm 0.1$                   | $94.5 \pm 0.1$                   | $93.9 \pm 0.4$                   | $94.7 \pm 0.3$                   | $94.0 \pm 0.4$                   |
| +1 epoch<br>(Warmup)    | $Acc_1$      | $37.4 \pm 1.2$                   | $53.5 \pm 2.9$                   | $19.8 \pm 0.5$                   | $48.7 \pm 1.1$                   | $28.1 \pm 1.3$                   |
|                         | $Acc_{best}$ | $94.4 \pm 0.1$                   | $94.7 \pm 0.1$                   | $94.1 \pm 0.1$                   | $95.1 \pm 0.1$                   | $95.4 \pm 0.2$                   |
| MetaInit                | $Acc_1$      | $30.5 \pm 0.9$                   | $35.1 \pm 0.6$                   | $14.6 \pm 2.2$                   | $29.0 \pm 1.5$                   | $11.7 \pm 1.6$                   |
|                         | $Acc_{best}$ | $94.6 \pm 0.1$                   | $94.6 \pm 0.1$                   | $94.2 \pm 0.1$                   | $94.8 \pm 0.1$                   | $95.0 \pm 0.5$                   |
| GradInit                | $Acc_1$      | $29.3 \pm 0.6$                   | $47.8 \pm 1.8$                   | $36.2 \pm 0.8$                   | $38.2 \pm 0.9$                   | $29.0 \pm 1.1$                   |
|                         | $Acc_{best}$ | <b><math>94.7 \pm 0.1</math></b> | <b><math>95.1 \pm 0.1</math></b> | <b><math>94.6 \pm 0.1</math></b> | <b><math>95.4 \pm 0.1</math></b> | <b><math>96.2 \pm 0.1</math></b> |

## Fixup

Статья - Hongyi Zhang, Yann N Dauphin, and Tengyu Ma. Fixup initialization: Residual learning without normalization. In ICLR, 2019.

Масштабирование стандартной инициализации

## MetaInit

Статья - Yann N Dauphin and Samuel Schoenholz. Metainit: Initializing learning by learning to initialize. In NeurIPS, pages 12645–12657, 2019.

Инициализирует таким образом, что градиентный спуск начинается с линейных областей.

Kaiming

Статья - Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In CVPR, 2015.

Инициализация с нормальным распределением (считается что функция активации ReLU)

Работы, оказавшие влияние

Xiao Shi Huang, Felipe Perez, Jimmy Ba, and Maksims Volkovs. Improving transformer optimization through better initialization. In ICML, 2020.

Цитирование

Цитат мало (<10), учитывая, что написана статья год назад, но конференция была недавно, поэтому ситуация может исправиться. Однако наталкивает на мысль, что такая инициализация не применима в действительности.

Дальнейшее исследование

Может ли инициализация не зависеть от оптимизатора?

Применение

Можно применить в любой области, так как инициализация с каждым годом все чаще играет важную роль, потому что архитектуры сетей усложняются.