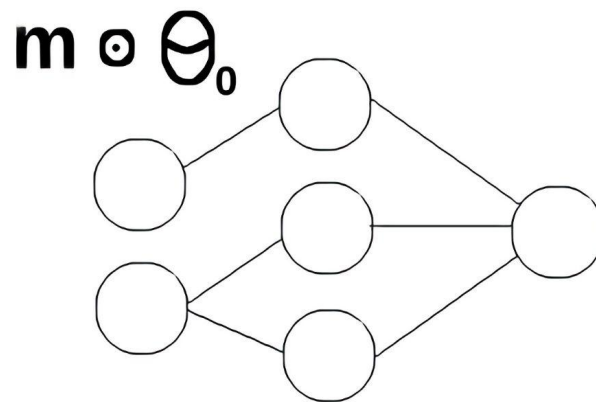
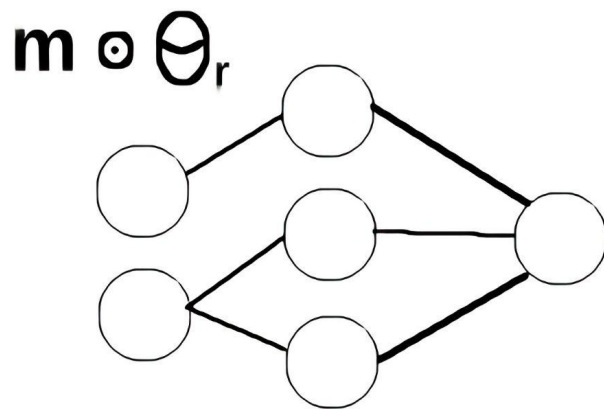
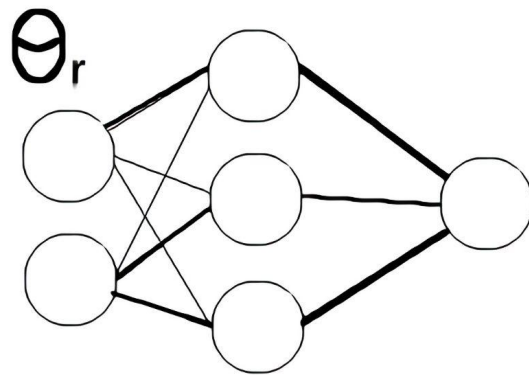
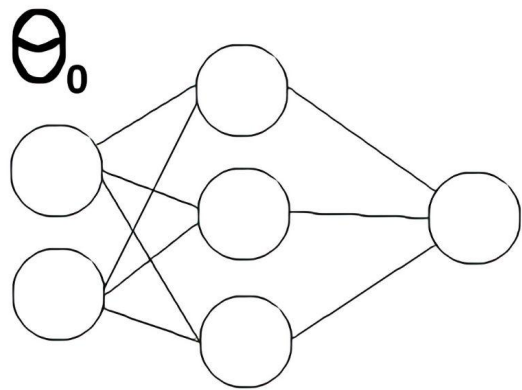


The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks

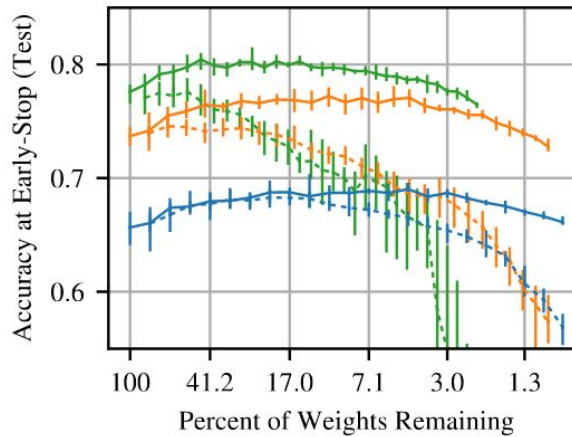
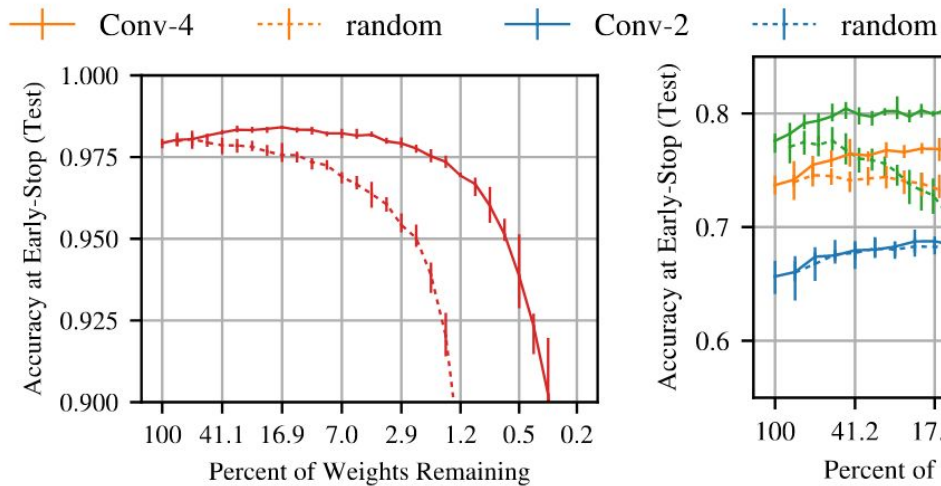
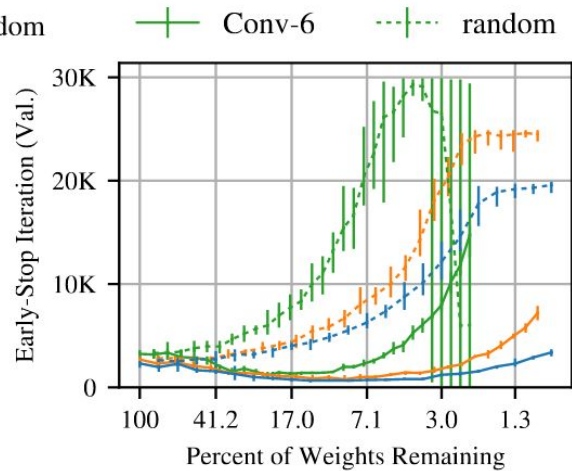
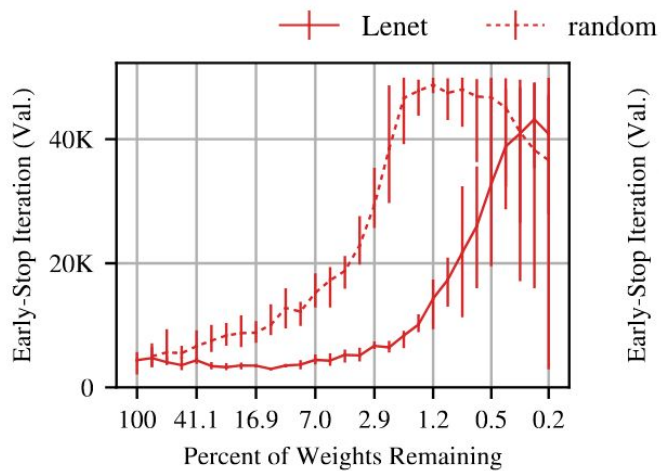
Докладчик: Илья Пахалко
Рецензент: Артем Алекберов
Практик-исследователь: Петр Григорьев
Хакер: Николай Карташев
БПМИ181



One-Shot Pruning

1. Пусть есть нейросеть $f(x, \theta_0)$, где θ_0 - начальная инициализация
2. Обучим её до состояния $f(x, \theta_r)$
3. Зададим бинарную маску m ; тогда $m \odot \theta_0$ - выигрышный билет.

В частности, в алгоритме One-Shot Pruning предлагается задавать m с помощью прунинга $p\%$ самых малых по модулю весов из θ_r



Гипотеза лотерейного билета

Для случайно инициализированной нейросети $f(x, \theta_0)$ существуют подсеть f' и маска m такая, что сеть $f'(x, m \odot \theta_0)$ при обучении с нуля достигает качества на валидации не меньше (обученной) исходной сети за меньшее число итераций процесса обучения.

Iterative Pruning

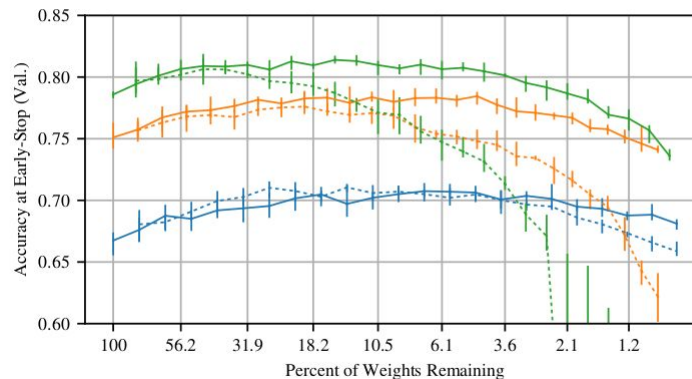
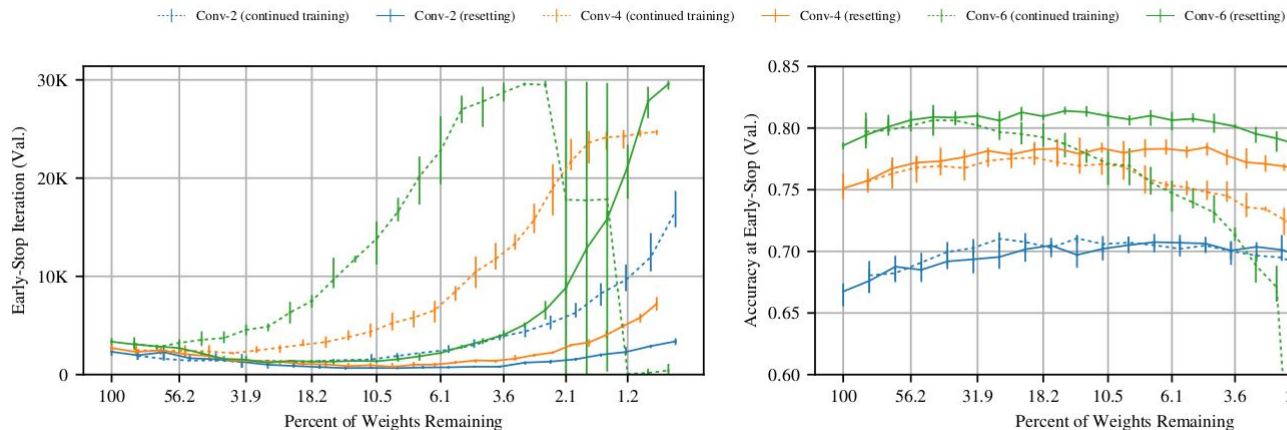
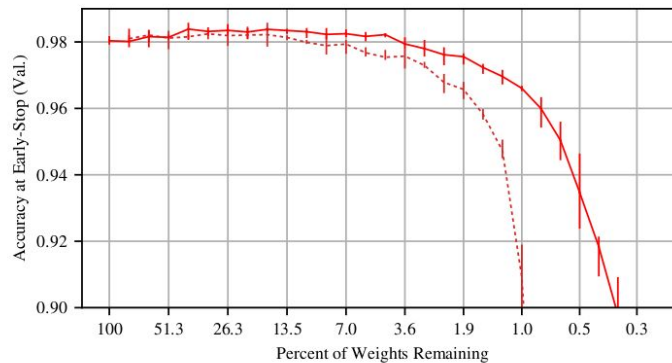
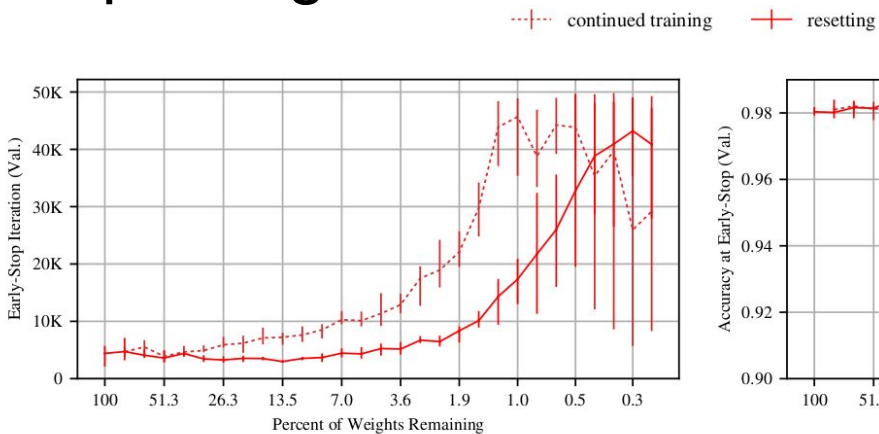
Пусть хотим найти билет за n циклов прунинга, суммарно убрать $p\%$ весов.

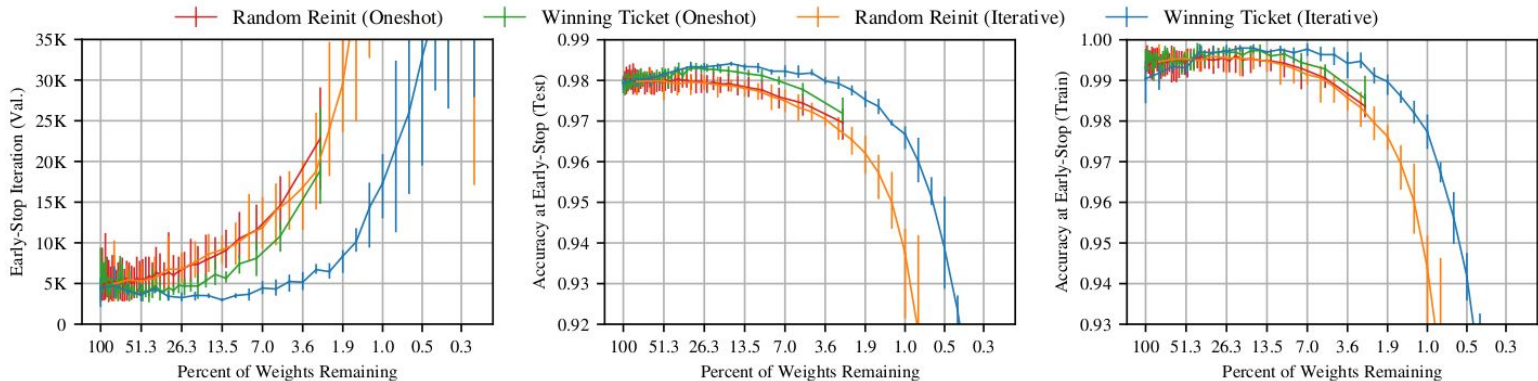
Начальные параметры: $\theta = \theta_0, m = 1^{|\theta|}, s = p^{\frac{1}{n}}$

1. Инициализируем сеть весами $m \odot \theta$
2. Обучаем сеть j итераций
3. Фиксируем $s\%$ от оставшихся весов нулевыми, задавая маску m'
4. Присваиваем $\theta = \theta_0$
5. Присваиваем $m = m'$
6. Повторяем шаги 1)-5) n итераций.
7. Присваиваем $\theta = \theta_0$.

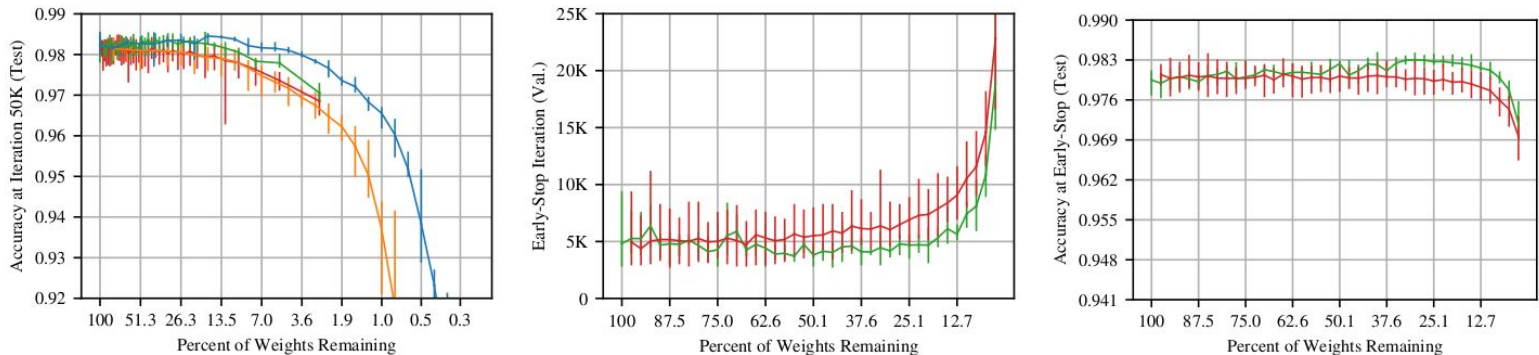
Пункт 4 можно как выполнять (тогда мы начинаем обучение с нуля каждую итерацию цикла), так и не выполнять (тогда мы оставляем выученные веса, меняем лишь маску). Нетрудно догадаться, что от этого зависит сложность процесса обучения модели.

Iterative pruning





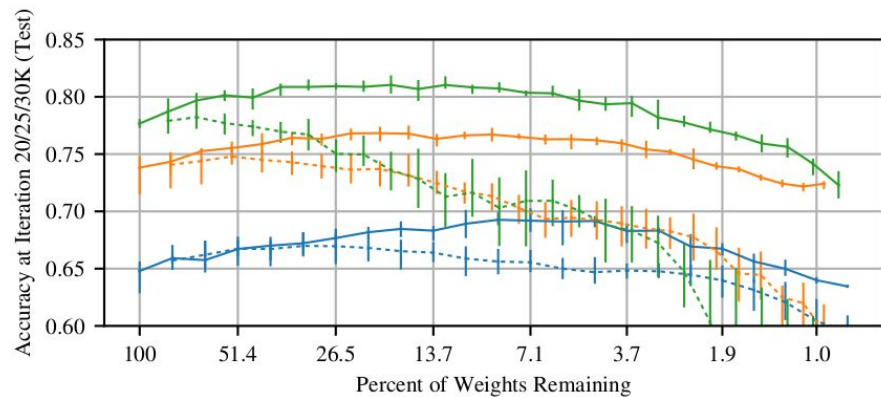
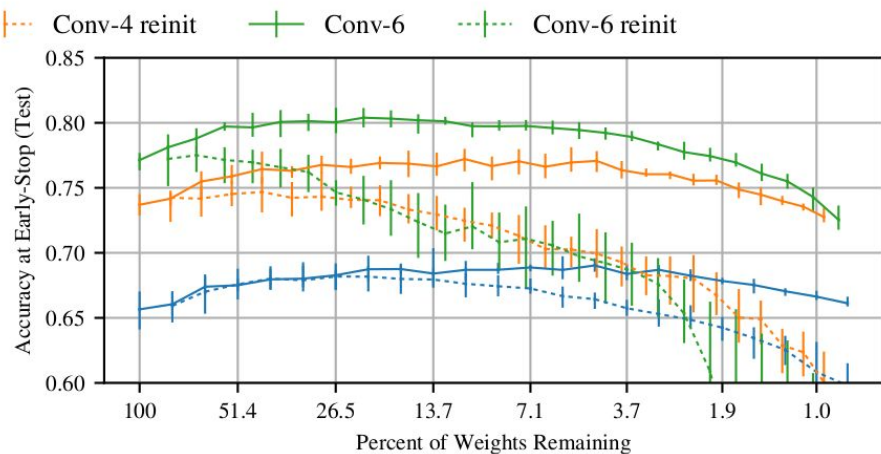
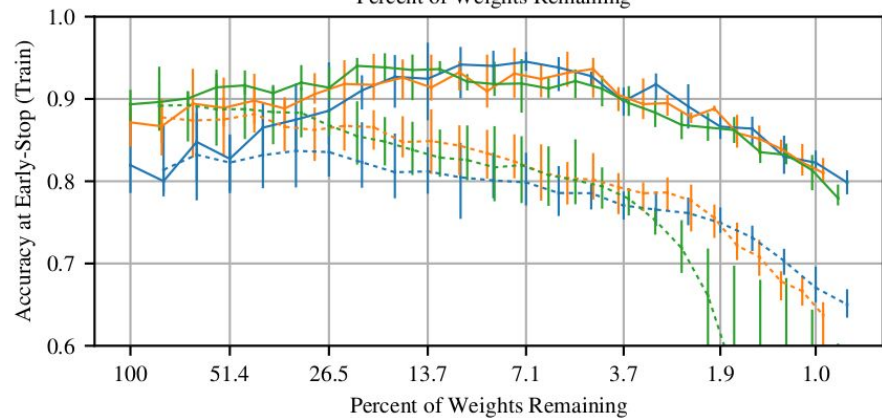
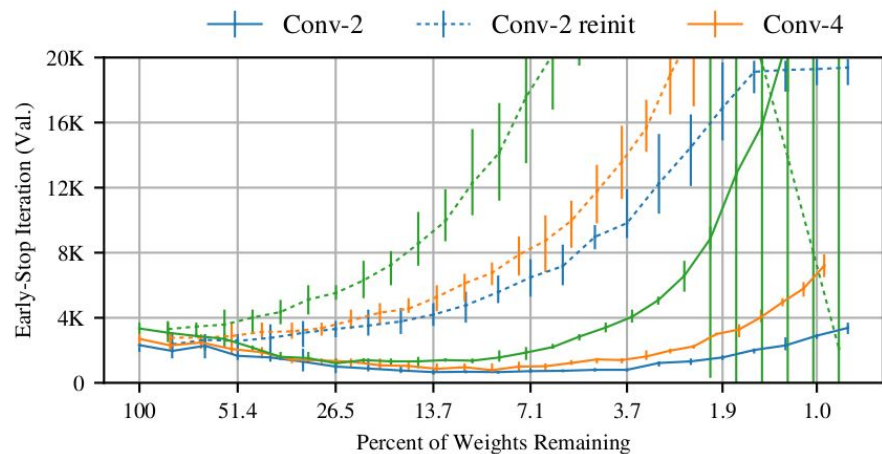
(a) Early-stopping iteration and accuracy for all pruning methods.

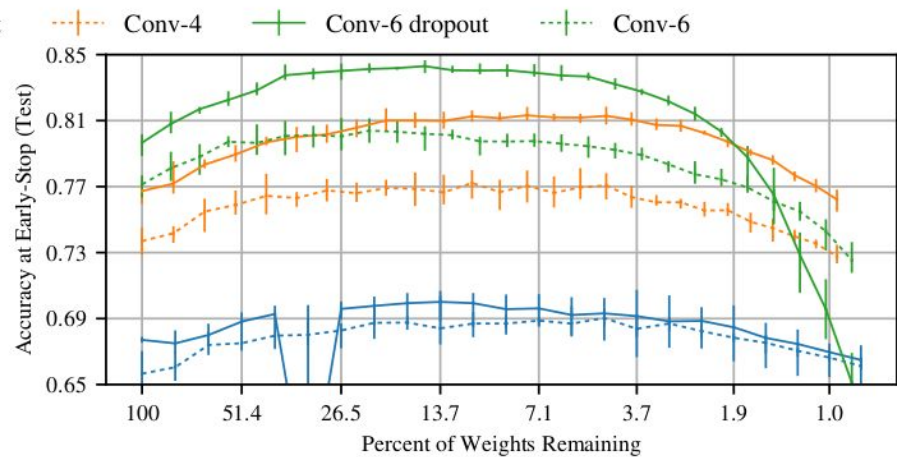
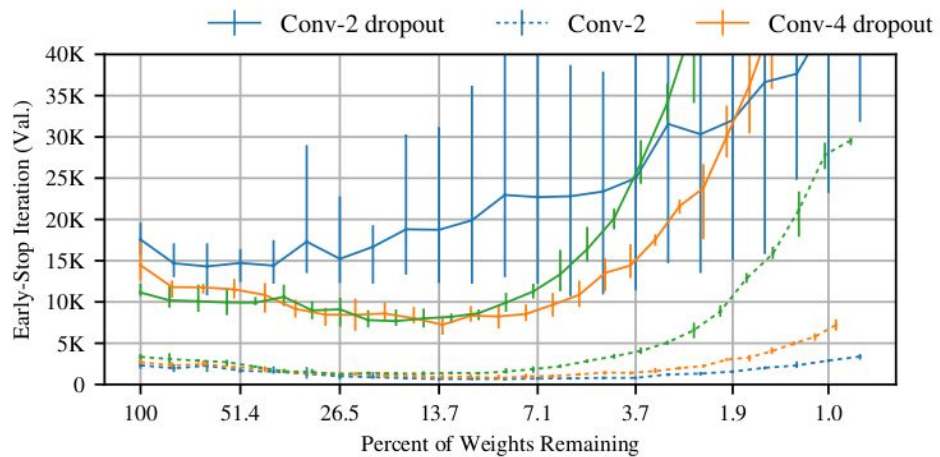


(b) Accuracy at end of training.

(c) Early-stopping iteration and accuracy for one-shot pruning.

Figure 4: Early-stopping iteration and accuracy of Lenet under one-shot and iterative pruning. Average of five trials; error bars for the minimum and maximum values. At iteration 50,000, training accuracy $\approx 100\%$ for $P_m \geq 2\%$ for iterative winning tickets (see Appendix [D](#) Figure [12](#)).





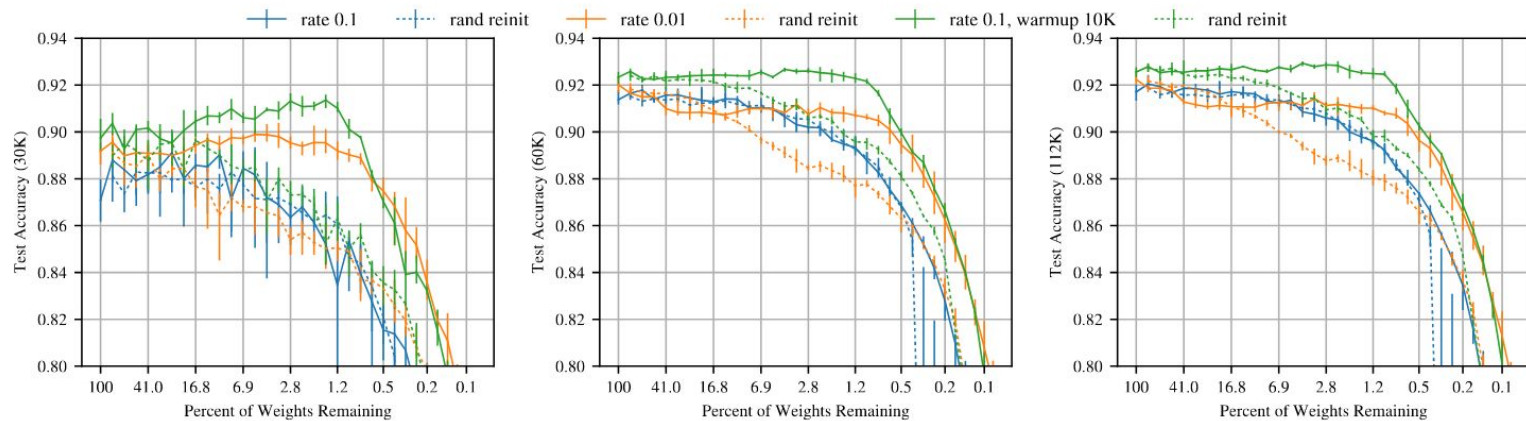


Figure 7: Test accuracy (at 30K, 60K, and 112K iterations) of VGG-19 when iteratively pruned.

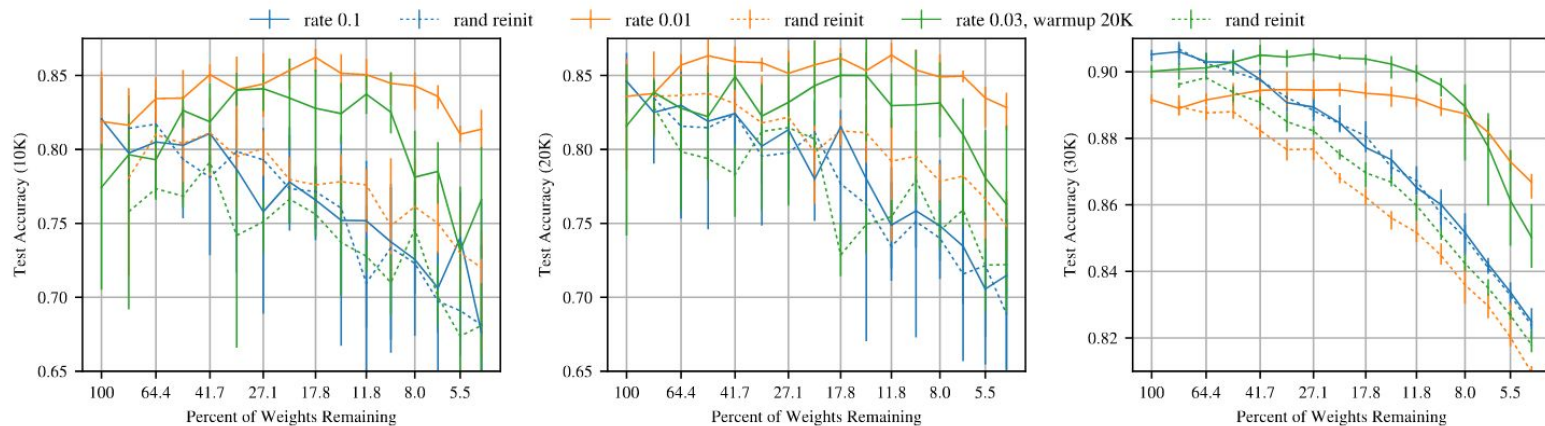


Figure 8: Test accuracy (at 10K, 20K, and 30K iterations) of Resnet-18 when iteratively pruned.

Итоги

- Эмпирически подтверждено, что для наилучших результатов важна как сама структура подсети, так и её инициализация
- Предложен способ нахождения “выигрышных билетов”
- Изучение свойств “выигрышных билетов” потенциально может помочь создать новый способ инициализации нейросетей

Рецензия

Положительные качества

1. Статья написана хорошим языком, гипотеза сформулирована просто и доходчиво
2. Новый подход к пониманию работы нейросети во время обучения, обратили внимание не только на саму архитектуру, как делали предыдущие статьи, но и на важность инициализации
3. Подтверждение гипотезы множеством экспериментов для разных архитектур, оптимизаторов, значений прунинга и гиперпараметров.
Очень большой объем дополнительных материалов, где детально описаны эксперименты, а также некоторые дополнительные исследования.

Недостатки

1. Графики в статье достаточно тяжело читаемы и отрисованы не очень аккуратно с большим количеством перекрещиваний и слияний графиков.
2. Авторы рассматривают только неструктурированный прунинг, что является на данный момент не применимым на практике способом ускорения нейросетей.
3. Авторы исследуют только Glorot инициализацию, однако возникает вопрос, является ли данная гипотеза специфичной под эту инициализацию. Возникают некоторые сомнения, что при Uniform инициализации эффект будет так заметен из-за ограничения значений аргумента распределения.
4. Также авторы не исследуют достаточный момент остановки обучения нейросети с полным количеством параметров, что может быть достаточно критично при обучении больших моделей.

Важное замечание от [рецензентов на openreview](#)

Большинство экспериментов проведено на достаточно маленьких и не используемых на практике датасетах (MNIST, Cifar 10) и моделях (Lenet и уменьшенные VGG варианты). При этом авторы замечают, что при увеличении глубины моделей находить «счастливые билеты» становится намного сложнее – только используя Warm-up получилось находить их более стабильно при высоких скоростях обучения. Однако более этого данный момент не был исследован.

Практический анализ

История публикаций



Опубликована на **ICLR | 2019**, Oral, Best Paper

- 5 версий статьи
- Изменилось оформление
- Убрали некоторые примеры (**XOR**)
- Применили к другим моделям (**CNN, ResNet**)
- Добавили аппендикс

Авторы статьи

- Jonathan Frankle, Michael Carbin, MIT
- Первое соавторство
- Первый автор занимался в основном распознаванием объектов и проблемами приватности
- Второй автор занимался в том числе оптимизацией сетей (с 2017)
- Продолжали работу по теме

Основные источники

- Применяемый метод прунинга
 - Learning both weights and connections for efficient neural network (2015)
 - [Song Han, Jeff Pool, John Tran, William Dally](#)
- Градиентный спуск на подсетях
 - Understanding dropout. In Advances in neural information processing systems (2013)
 - [Pierre Baldi and Peter J Sadowski](#)
- Конкуренты: описаны альтернативные подходы
 - Deep Rewiring: Training very sparse deep networks (2018)
 - [Guillaume Bellec, David Kappel, Wolfgang Maass, Robert Legenstein](#)

Продолжения

Улучшение качества, применение к другим архитектурам

Stabilizing the Lottery Ticket Hypothesis / The Lottery Ticket Hypothesis at Scale (2020)

Linear Mode Connectivity and the Lottery Ticket Hypothesis (2020)

The Lottery Ticket Hypothesis for Pre-Trained BERT Networks (2020)

Цитирования

~1500 цитирований. Мета обзоры, работы про прунинг

Explaining explanations: An overview of interpretability of machine learning (2019) [Gilpin et al.](#)

The state of sparsity in deep neural networks (2019) [Gale et al.](#)

Importance estimation for neural network pruning (2019) [Molchanov et al.](#)

Идеи исследований

- Аналогичный анализ для сетей другого рода (например, авторы уже делали для bert)
- Выявить выигрышные билеты без обучения

Исследование переноса выигрышных билетов на другие наборы данных

