

Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time

Рецензия

Денис Козлов

Авторы

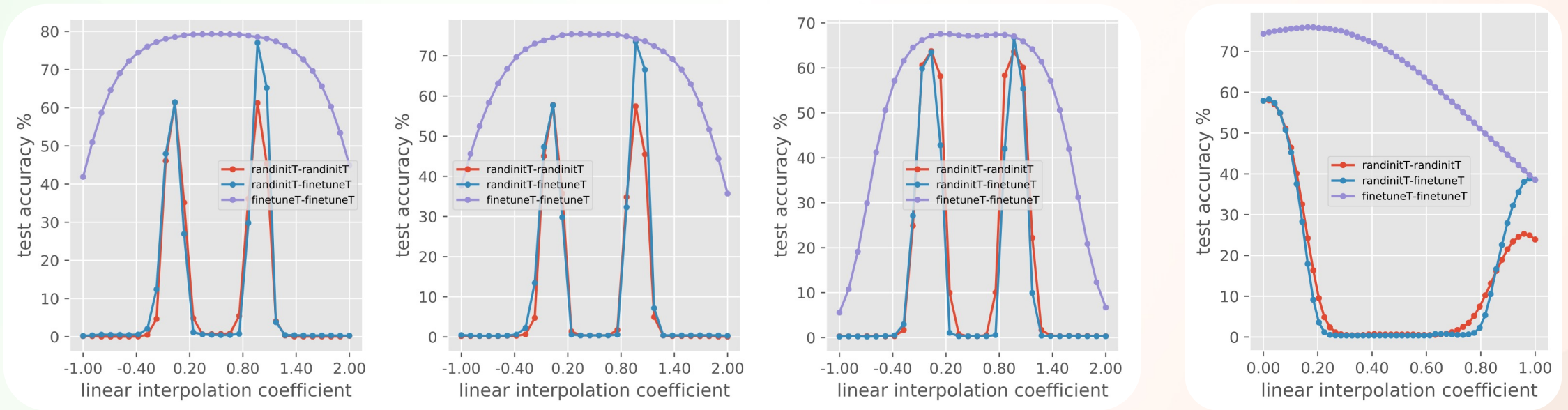
- Washington, Columbia, Tel Aviv Universities
- Google Brain, Meta AI
- Mitchell Wortsman — h-index 10, OpenCLIP, LAION
- Много авторов пересекаются с Robust fine-tuning of zero-shot models
- ICML 2022
- Уже 74 цитирования

Предшественники и мотивация

- Averaging Weights Leads to Wider Optima and Better Generalization, 2018
- Rethinking the Inception Architecture for Computer Vision, 2015
 - Усредняют веса модели на разных точках во время оптимизации

Предшественники и мотивация

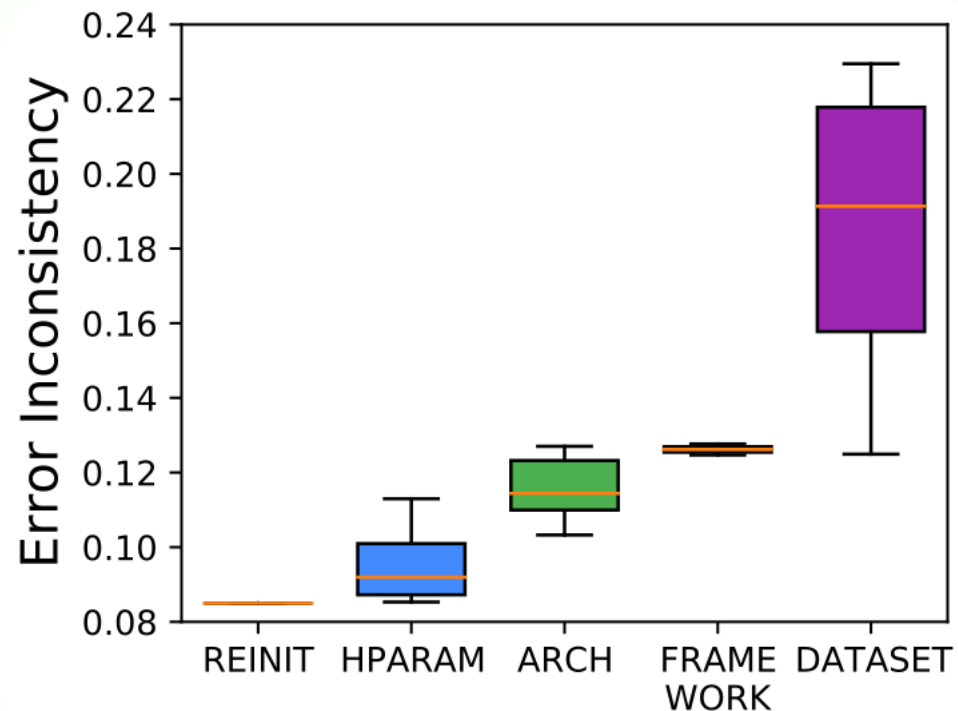
- What is being transferred in transfer learning? 2020
 - При файнтьюне веса модели сходятся к одной «области»



Даже разные данные!

Предшественники и мотивация

- No One Representation to Rule Them All: Overlapping Features of Training Methods
 - Обученные по-разному модели ошибаются в разных местах.

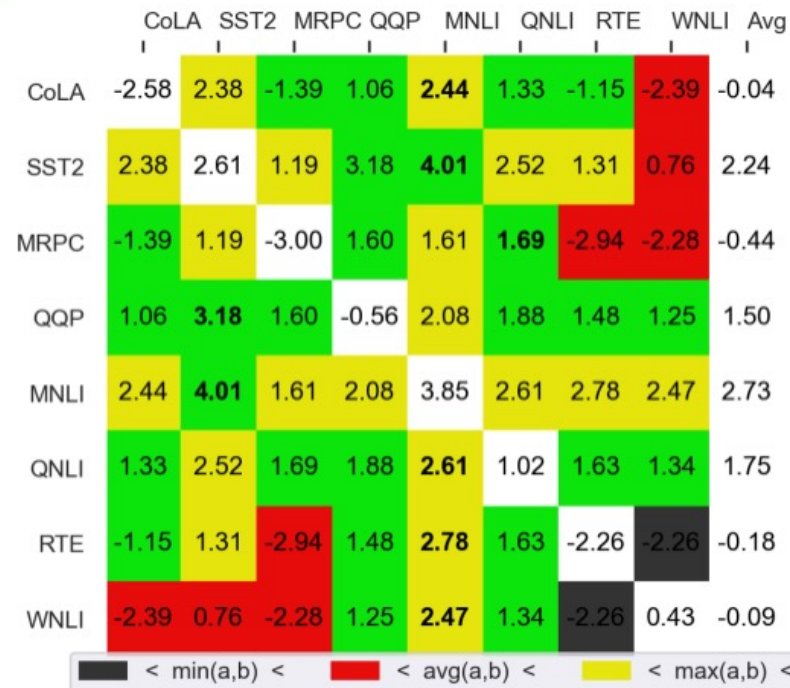


Результаты

Task	Dataset	Model	Metric Name	Metric Value	Global Rank	Uses Extra Training Data
Image Classification	ImageNet	Model soups (BASIC-L)	Top 1 Accuracy	90.98%	# 2	✗
			Number of params	2440M	# 782	✗
Image Classification	ImageNet	Model soups (ViT-G/14)	Top 1 Accuracy	90.94%	# 3	✗
			Number of params	1843M	# 778	✗
Domain Generalization	ImageNet-A	Model soups (BASIC-L)	Top-1 accuracy %	94.17	# 1	✓
Domain Generalization	ImageNet-A	Model soups (ViT-G/14)	Top-1 accuracy %	92.67	# 2	✓
Domain Generalization	ImageNet-R	Model soups (BASIC-L)	Top-1 Error Rate	3.90	# 1	✓
Domain Generalization	ImageNet-R	Model soups (ViT-G/14)	Top-1 Error Rate	4.54	# 2	✓
Unsupervised Domain Adaptation	ImageNet-R	Model soups (ViT-G/14)	Top 1 Error	4.54	# 1	✓
Image Classification	ImageNet-Real	Model soups (ViT-G/14)	Accuracy	91.20%	# 2	✓

Открытия рядом

- Fusing finetuned models for better pretraining, 2022
 - Усреднение весов нескольких finetuned моделей является хорошей базовой моделью для файнтьюна



Открытия рядом

- Building Machine Learning Models like Open-Source Software, 2022
 - Colin Raffel: T5, Librosa, Theano

Комментарии к работе

- Код и веса — есть!
 - Кода мало, но выглядит прилично
- Автор отвечает на issues на гитхабе!
- В работе приведены не только экспериментальные результаты, но и теоретические доводы
- Рассмотрено не так много способов смешения супов
 - «Супы второго порядка», более подробные результаты обучаемых супов...

Мне не очень понравились графики

