Robust fine-tuning of zero-shot models

Обзор-рецензия

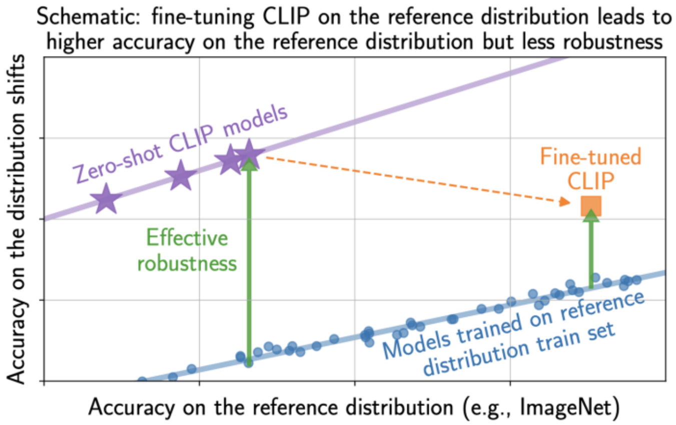
Алексей Шишков, БПМИ192

Основная идея

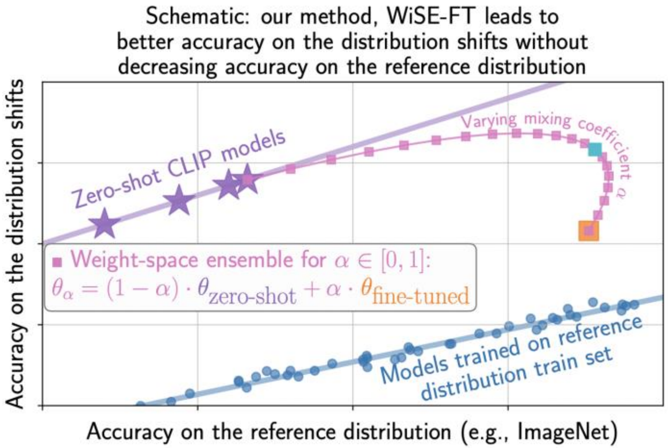
Can zero-shot models be fine-tuned without reducing accuracy under distribution shift?

- Раньше при дообучении zero-shot нейросетей их “устойчивость” (качество на других наборах данных) снижалась
- При дообучении новым способом она повышается

Было:



Стало:



Про статью

- Первая версия – сентябрь 2021
- Финальная – июнь 2022, для конференции CVPR 2022
- Статья-финалист конкурса на лучшую работу

Robust fine-tuning of zero-shot models

Mitchell Wortsman[*†] Gabriel Ilharco[*†] Jong Wook Kim[§] Mike Li[‡]
Simon Kornblith[◇] Rebecca Roelofs[◇] Raphael Gontijo-Lopes[◇]
Hannaneh Hajishirzi[†○] Ali Farhadi[*†] Hongseok Namkoong[*‡] Ludwig Schmidt[†△]

Abstract

Large pre-trained models such as CLIP or ALIGN offer consistent accuracy across a range of data distributions when performing zero-shot inference (i.e., without fine-tuning on a specific dataset). Although existing fine-tuning methods substantially improve accuracy on a given target distribution, they often reduce robustness to distribution shifts. We address this tension by introducing a simple and effective method for improving robustness while fine-tuning: ensembling the weights of the zero-shot and fine-tuned models (WiSE-FT). Compared to standard fine-tuning, WiSE-FT provides large accuracy improvements under distribution shift, while preserving high accuracy on the target distribution. On ImageNet and five derived distribution shifts, WiSE-FT improves accuracy under distribution shift by 4 to 6 percentage points (pp) over prior work while increasing ImageNet accuracy by 1.6 pp. WiSE-FT achieves similarly large robustness gains (2 to 23 pp) on a diverse set of six further distribution shifts, and accuracy gains of 0.8 to 3.3 pp compared to standard fine-tuning on seven commonly used transfer learning datasets. These improvements come at no additional computational cost during fine-tuning or inference.

Mitchell Wortsman

Gabriel Ilharco

Про статью: предыдущие работы авторов

Robust fine-tuning of zero-shot models
Mitchell Wortsman*, Gabriel Ilharco*, Jong Wook Kim, Mike Li, Simon Kornblith, Rebecca Roelofs, Raphael Gontijo-Lopes, Hannaneh Hajishirzi, Ali Farhadi, Hongseok Namkoong, Ludwig Schmidt
CVPR, 2022 **(oral, best paper finalist)**
arxiv / code

OpenCLIP: An open source implementation of CLIP
Gabriel Ilharco*, **Mitchell Wortsman***, Ross Wightman*, Cade Gordon*, Nicholas Carlini, Rohan Taori, Achal Dave, Vaishaal Shankar, Hongseok Namkoong, John Miller, Hannaneh Hajishirzi, Ali Farhadi, Ludwig Schmidt
GitHub, 2021

Learning Neural Network Subspaces
Mitchell Wortsman, Maxwell Horton, Carlos Guestrin, Ali Farhadi, Mohammad Rastegari
ICML, 2021
arxiv / code

Supermasks in Superposition
Mitchell Wortsman*, Vivek Ramanujan*, Rosanne Liu, Aniruddha Kembhavi, Mohammad Rastegari, Jason Yosinski, Ali Farhadi
NeurIPS, 2020
arxiv / code

Soft Threshold Weight Reparameterization for Learnable Sparsity
Aditya Kusupati, Raghav Somani*, Vivek Ramanujan*, **Mitchell Wortsman***, Prateek Jain, Sham Kakade, Ali Farhadi
ICML, 2020
arxiv / code

What's Hidden in a Randomly Weighted Neural Network?
Vivek Ramanujan*, **Mitchell Wortsman***, Aniruddha Kembhavi, Ali Farhadi, Mohammad Rastegari
CVPR, 2020
arxiv / code

Discovering Neural Wirings
Mitchell Wortsman, Ali Farhadi, Mohammad Rastegari
NeurIPS, 2019

Mitchell Wortsman

- 2022.06: I'll be presenting Robust fine-tuning of zero-shot models at CVPR 2022, come say hi!
- 2022.05: I'm very thankful for the recognition of outstanding reviewer at CVPR 2022
- 2022.04: Our open-source repository for training CLIP models has reached 1000 stars!
- 2022.03: Model soups set a new state-of-the-art on ImageNet
- 2022.03: What makes zero-shot CLIP models robust? Find out here
- 2022.03: Check out our work using CLIP for zero-shot object navigation.
- 2021.10: I'm excited to be starting an internship with Jacob Eisenstein at Google Research
- 2021.09: Check our our new work Robust fine-tuning of zero-shot models!
- 2021.08: I'm very thankful for the recognition of outstanding reviewer at ACL 2021
- 2021.06: Our paper exploring the relation between visual and text representations has been accepted to NAACL
- 2021.04: Our MultiModalQA paper on complex QA over text, tables and images is out
- 2021.03: Our paper on contrastive representation learning is out
- 2020.11: We'll be presenting a tutorial on High Performance NLP at EMNLP 2020
- 2020.06: I'm very excited to be starting an internship with Peter Anderson and Ashish Vaswani at Google Research
- 2020.05: Check out our new preprint exploring similarities between vision and language representations
- 2020.04: Our work Evaluating NLP models via contrast sets is out
- 2020.02: Check out our new paper exploring the dynamics of fine-tuning in NLP
- 2020.01: Our paper Toward ML-Centric Cloud Platforms made the cover of the Communications of the ACM
- 2019.12: Don't miss our spotlight presentation on SDTW at ViGIL, NeuRIPS 2019.
- 2019.11: Our CoNLL 2019 paper was awarded Honorable Mention for Best Paper in Research Inspired by Human Language Learning!

Gabriel Ilharco

Про статью: предыдущие работы авторов

Robust fine-tuning of zero-shot models
Mitchell Wortsman*, Gabriel Ilharco*, Jong Wook Kim, Mike Li, Simon Kornblith, Rebecca Roelofs, Raphael Gontijo-Lopes, Hannaneh Hajishirzi, Ali Farhadi, Hongseok Namkoong, Ludwig Schmidt
CVPR, 2022 **(oral, best paper finalist)**
arxiv / code

OpenCLIP: An open source implementation of CLIP
Gabriel Ilharco*, **Mitchell Wortsman***, Ross Wightman*, Cade Gordon*, Nicholas Carlini, Rohan Taori, Achal Dave, Vaishaal Shankar, Hongseok Namkoong, John Miller, Hannaneh Hajishirzi, Ali Farhadi, Ludwig Schmidt
GitHub, 2021

Learning Neural Network Subspaces
Mitchell Wortsman, Maxwell Horton, Carlos Guestrin, Ali Farhadi, Mohammad Rastegari
ICML, 2021
arxiv / code

Supermasks in Superposition

- 2022.06: I'll be presenting Robust fine-tuning of zero-shot models at CVPR 2022, come say hi!
- 2022.05: I'm very thankful for the recognition of outstanding reviewer at CVPR 2022
- 2022.04: Our open-source repository for training CLIP models has reached 1000 stars!
- 2022.03: Model soups set a new state-of-the-art on ImageNet
- 2022.03: What makes zero-shot CLIP models robust? Find out here
- 2022.03: Check out our work using CLIP for zero-shot object navigation.
- 2021.10: I'm excited to be starting an internship with Jacob Eisenstein at Google Research

mlfoundations / open_clip Public

Watch 37 | Fork 244 | Star 2.4k

Code | Issues 25 | Pull requests 15 | Discussions | Actions | Projects

Про статью: предыдущие работы авторов

Learning Neural Network Subspaces
Mitchell Wortsman, Maxwell Horton, Carlos Guestrin, Ali Farhadi, Mohammad Rastegari
ICML, 2021
arxiv / code

Supermasks in Superposition
Mitchell Wortsman*, Vivek Ramanujan*, Rosanne Liu, Aniruddha Kembhavi, Mohammad Rastegari, Jason Yosinski, Ali Farhadi
NeurIPS, 2020
arxiv / code

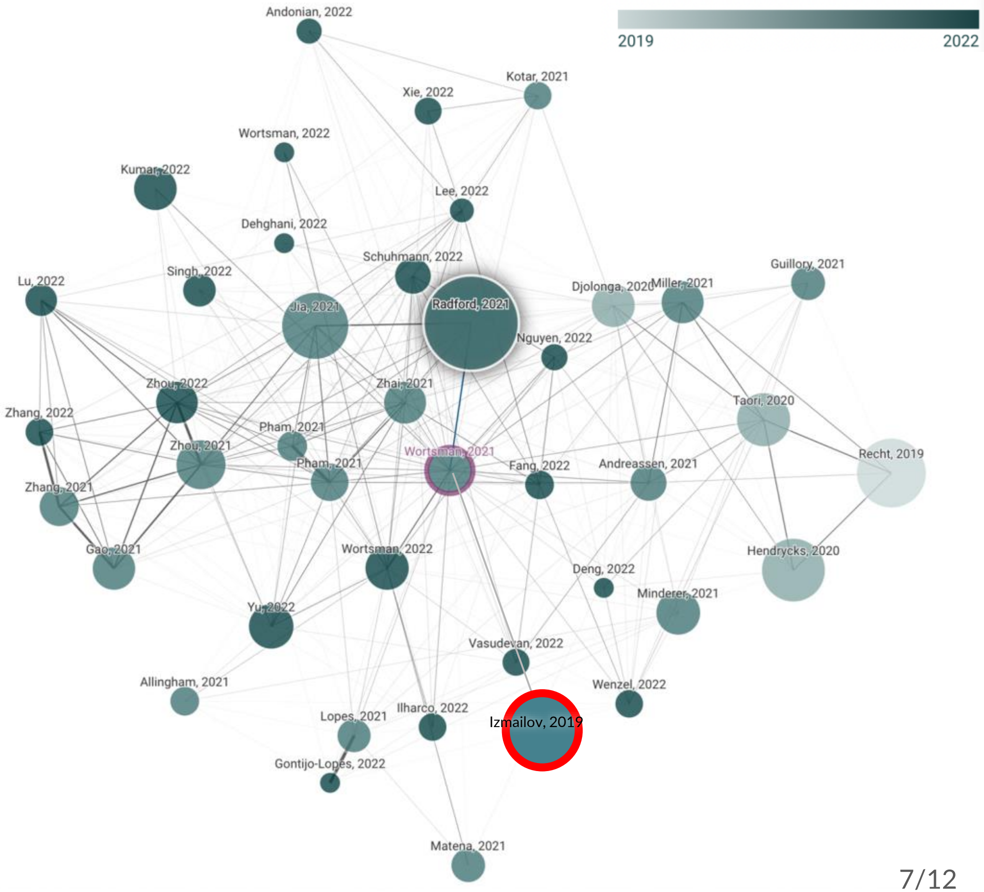
Soft Threshold Weight Reparameterization for Learnable Sparsity
Aditya Kusupati, Raghav Somani*, Vivek Ramanujan*, **Mitchell Wortsman***, Prateek Jain, Sham Kakade, Ali Farhadi
ICML, 2020
arxiv / code

What's Hidden in a Randomly Weighted Neural Network?
Vivek Ramanujan*, **Mitchell Wortsman***, Aniruddha Kembhavi, Ali Farhadi, Mohammad Rastegari
CVPR, 2020
arxiv / code

Связанные статьи

Pavel Izmailov et al., Averaging Weights Leads to Wider Optima and Better Generalization

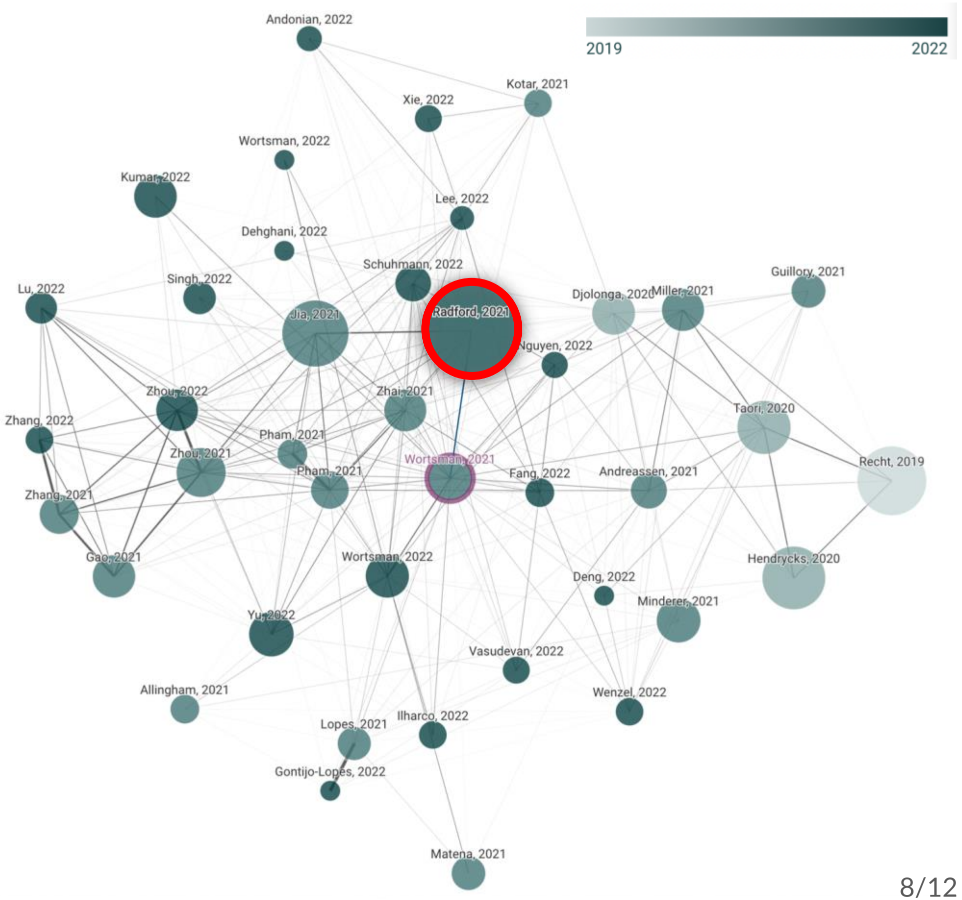
Усреднение весов чекпоинтов моделей для повышения качества



Связанные статьи

Alec Radford et al., Learning transferable visual models from natural language supervision

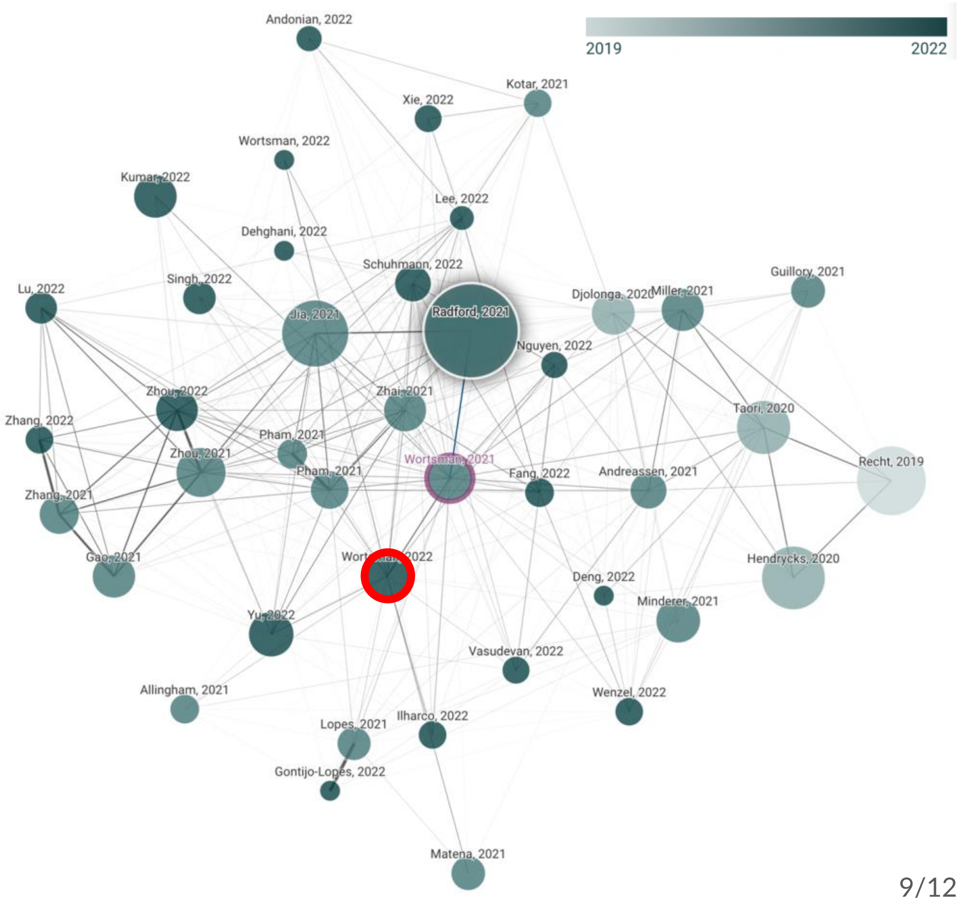
Способ повышения устойчивого дообучения предобученной модели



Связанные статьи

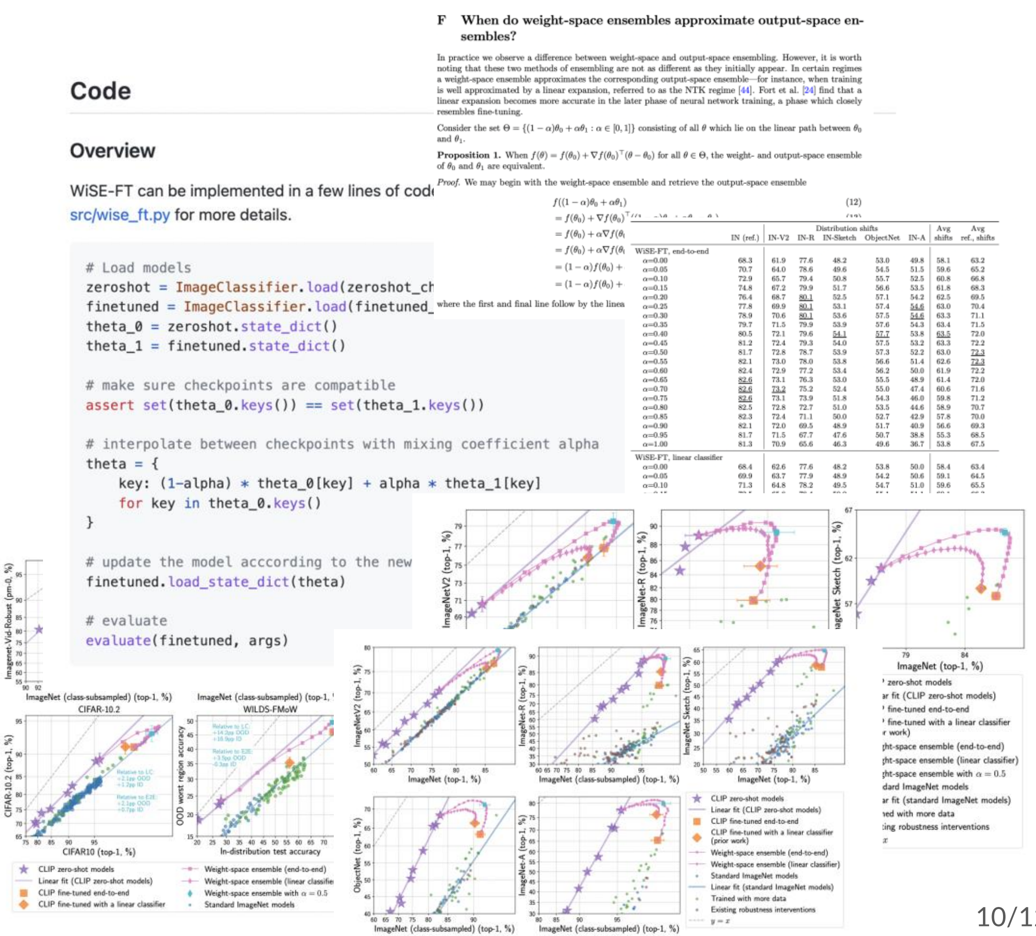
Mitchell Wortsman et al., Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time

Смотрите в следующих сериях!



Сильные стороны статьи

- Проведено большое число обучений и экспериментов, много моделей
- Приложен код
- Идея несложная, но описано, как к ней пришли авторы, сама идея описана хорошо



Слабые стороны статьи

- Новизна подхода
- Похоже на обучение на двух датасетах
- Выбор θ

Сравнить качество с обучением на двух датасетах?

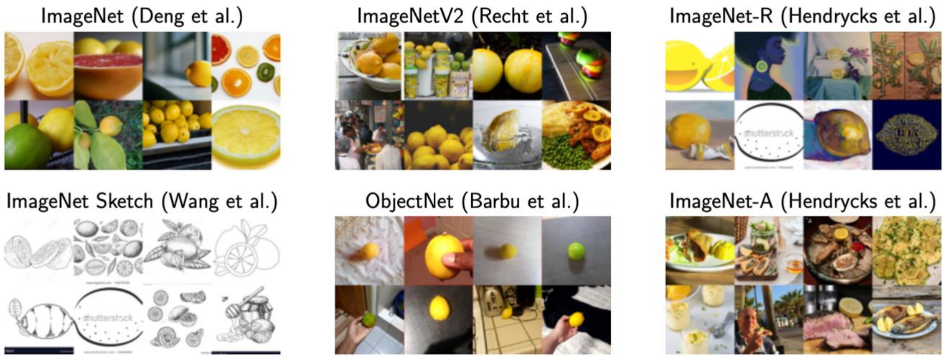


Figure 2: Samples of the class *lemon*, from the reference distribution ImageNet [17] and the derived distribution shifts considered in our main experiments: ImageNet-V2 [83], ImageNet-R [37], ImageNet Sketch [100], ObjectNet [4], and ImageNet-A [38].

Выводы

- Умеем хорошо дообучать сетки, получаем лучшее качеств
- Простые идеи неплохо работают
- Ещё одно применение усреднений весов моделей
- В статье важна не только идея, но и оформление результатов