Self-*Soup*ervision: Cooking Model Soups without Labels


Anthony Fuller,

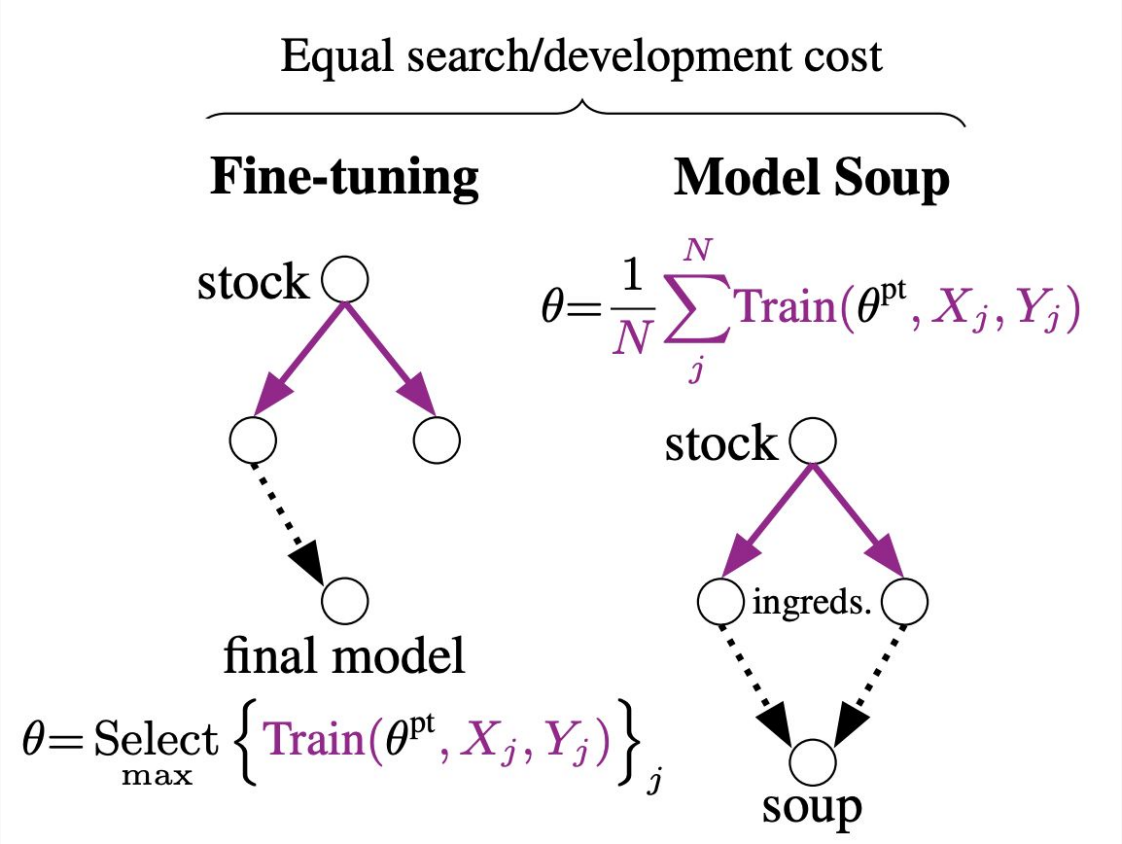
James R. Green,

Evan Shelhamer

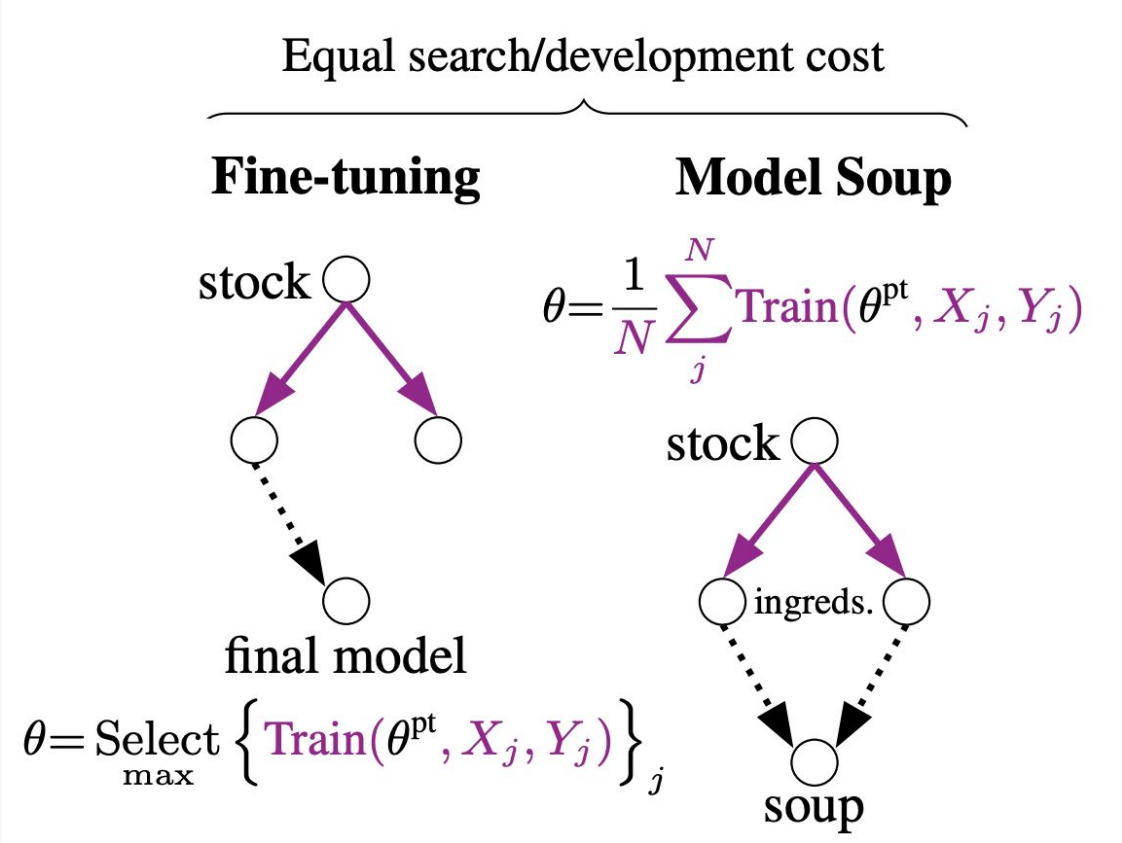
Carleton University, UBC, Vector Institute

ICML 2026 (spotlight)

Fine-tuning and Souping

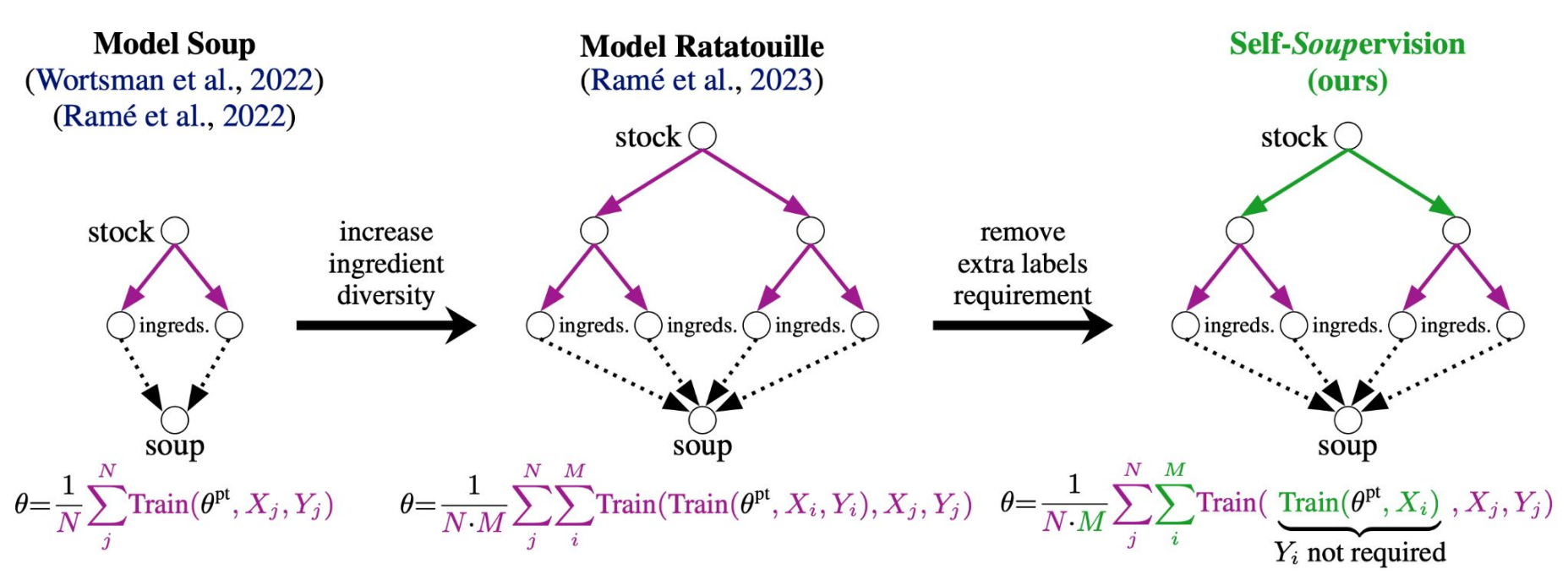


Fine-tuning and Souping



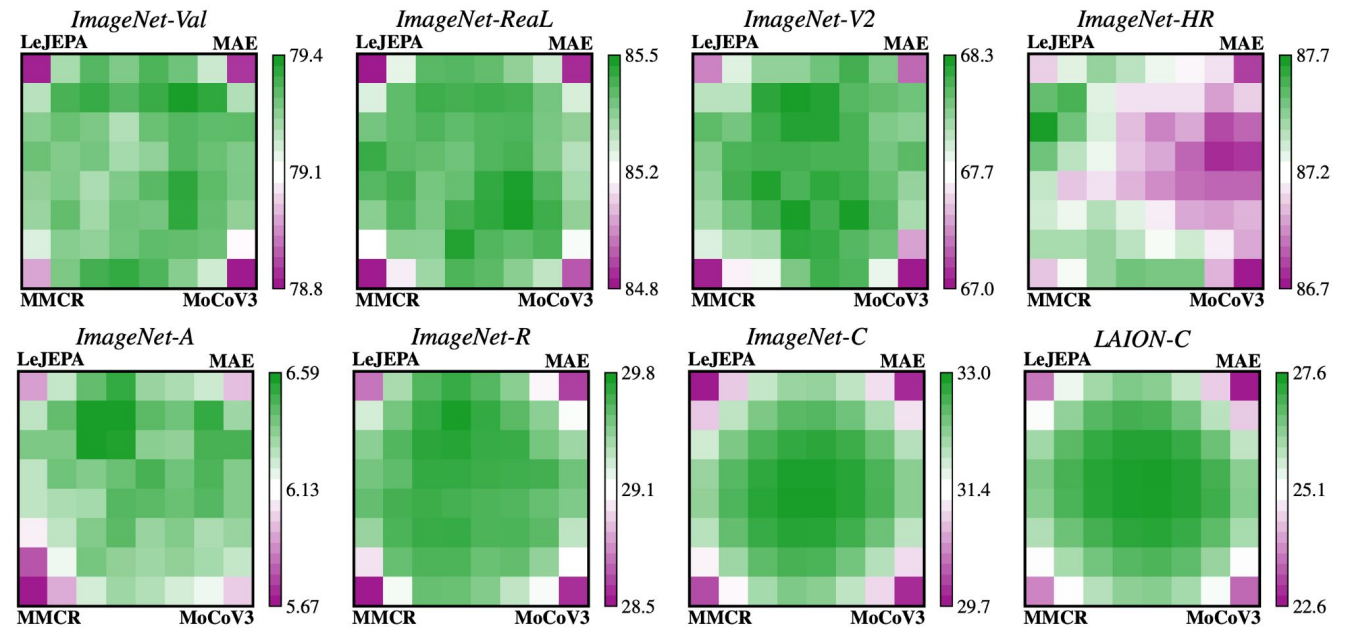
Fine-tuning and model souping have equal deployment costs because they are both just a single model / set of weights

Soups and Supervision



Self-Souping is possible and productive!

1. Inter-train 4 models on ImageNet using different self-supervised learning (SSL) algorithms
   a. LeJEPA, MAE, MoCoV3, and MMCR
2. Fine-tune on ImageNet with labels
3. Mix ingredients to make our tasty soup!



Self-Soups vs Supervised Soups

Soup Type	Mix Method	IN-Val	IN-ReaL	IN-V2	IN-HR	IN-A	IN-R	IN-C	LAION-C
Supervised Soup	Best Ingredient	79.05	85.07	67.51	87.34	7.49	28.84	28.74	18.67
	Greedy Search	79.34	85.58	68.09	87.40	7.91	29.92	30.88	21.02
	Uniform Mix	79.12	85.54	68.01	87.46	7.75	30.05	30.91	22.05
	Ensemble	80.09	86.17	68.74	88.44	7.64	30.24	30.04	20.10
Continued SSL + Supervised Soup	Best Ingredient	78.98	85.10	67.59	87.24	7.16	29.01	28.56	18.39
	Greedy Search	79.38	85.66	68.22	**87.60**	8.15	29.70	31.04	20.66
	Uniform Mix	79.21	85.68	68.16	87.20	7.99	29.93	31.44	21.50
	Ensemble	80.05	86.09	68.92	88.16	7.76	30.27	30.38	19.83
Self-Soup (ours)	Best Ingredient	79.18	85.10	67.24	87.28	7.20	29.04	28.55	18.67
	Greedy Search	**79.41**	**85.74**	**68.45**	87.38	**8.25**	29.90	31.75	21.20
	Uniform Mix	79.09	85.64	68.07	87.30	7.83	**30.15**	**32.34**	**22.92**
	Ensemble	80.49	86.51	69.60	88.50	7.84	30.85	31.41	20.82

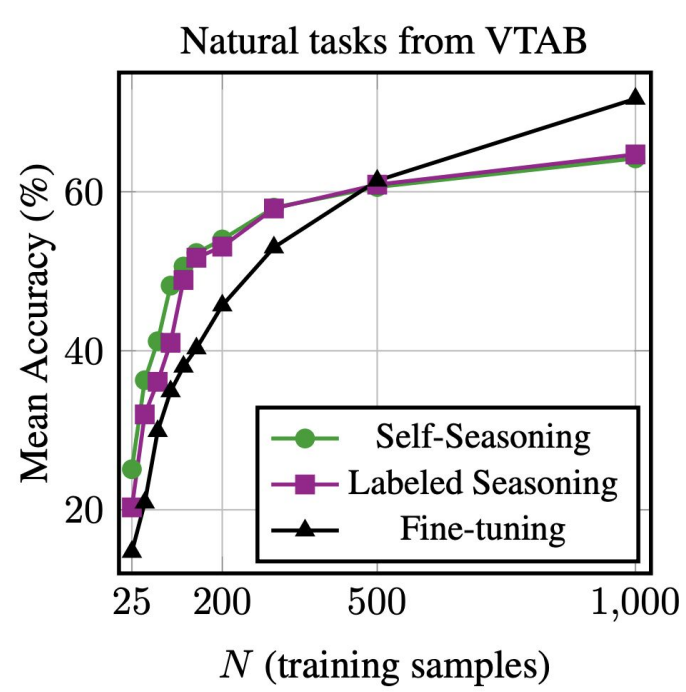
~1% gains against corrupted test sets without adding deployment cost!

Self-Soups vs Model Ratatouille

We now compare our Self-Soups to Model Ratatouille, which needs several ***labeled*** auxiliary datasets for its inter-training

Method	Needs extra labels	IN-Val	IN-C	LAION-C
Ratatouille	✔	79.05	32.20	22.89
Self-Soup	✘	79.09	32.34	22.92

Mixing Soups entirely without labels



We now mix SSL ingredients directly!

Our Self-Seasoning learns task-specific ingredient-mixture coefficients without labels by minimizing kNN entropy

Conclusion: Let's continue cooking

We show that model soups can be made without labels but there is plenty more to cook

Code: https://github.com/antofuller/self_soupervision