You Only Need a Good Embeddings Extractor to Fix Spurious Correlations

Raghav Kiranbhai Mehta, Vítor Albiero, Li Chen, **Ivan Evtimov**, Tamar Glaser, Zhiheng Li, Tal Hassner

RCV @ ECCV 2022

Meta AI

Spurious correlations lead to unintended shortcuts in models

- *“(…) we demonstrate that recent deep learning systems to detect COVID-19 from chest radiographs* ***rely on confounding factors rather than medical pathology*** *(…)*”

  DeGrave, Alex J., Joseph D. Janizek, and Su-In Lee. "AI for radiographic COVID-19 detection selects shortcuts over signal." Nature Machine Intelligence 3.7 (2021): 610-619.

- *“(…) models can* ***achieve non-trivial accuracy by relying on the background alone*** *(…)*” and *“(…) models often misclassify images even in the presence of correctly classified foregrounds****—up to 87.5% of the time*** *with adversarially chosen backgrounds (…)*”

  "Noise or signal: The role of image backgrounds in object recognition." arXiv preprint arXiv:2006.09994 (2020).

The Waterbirds dataset

At **training** time: Deliberately introduce spurious correlation between the background and the object through a data imbalance
At **test** time: Measure accuracy in the worst group (WGA) as a signal of whether the model is relying on the shortcut

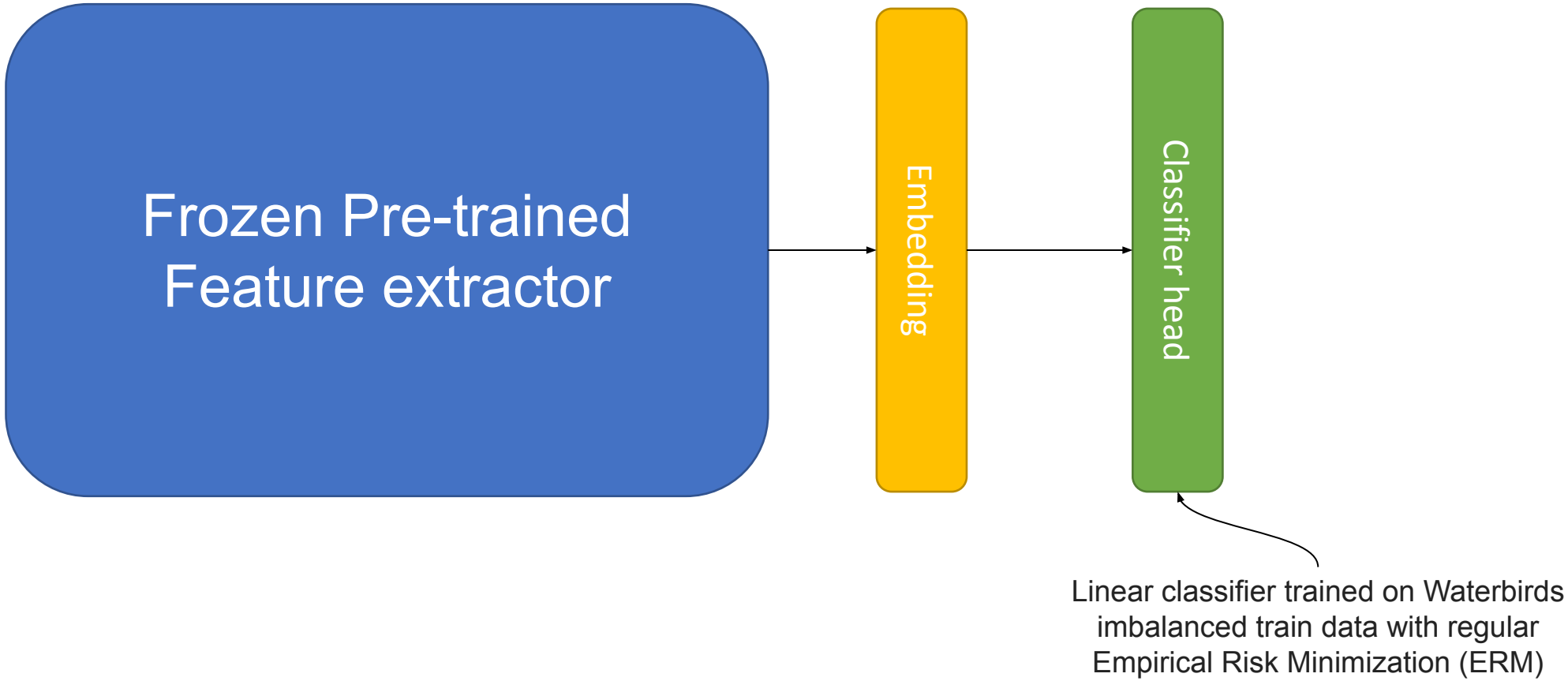
	Waterbird 	**Landbird** 
Water background 	95%	5%
Land background 	5%	95%

Sagawa, S., Koh, P.W., Hashimoto, T.B., Liang, P.: Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. arXiv preprint arXiv:1911.08731 (2019)

Prior work

	Uses externally provided spurious group labels	Infers group labels	End-to-end training?
GroupDRO	✓		✓
Just Train Twice (JTT)		✓	✓
Environment Inference for Invariant Learning (EIIL)		✓	✓
Subsampling Large Groups (SUBG)	✓		✓
Deep Feature Reweighting	✓		✓
Invariant Risk Minimization (IRM)	✓		✓
Our observation	**No**	**No**	**No**

Our approach

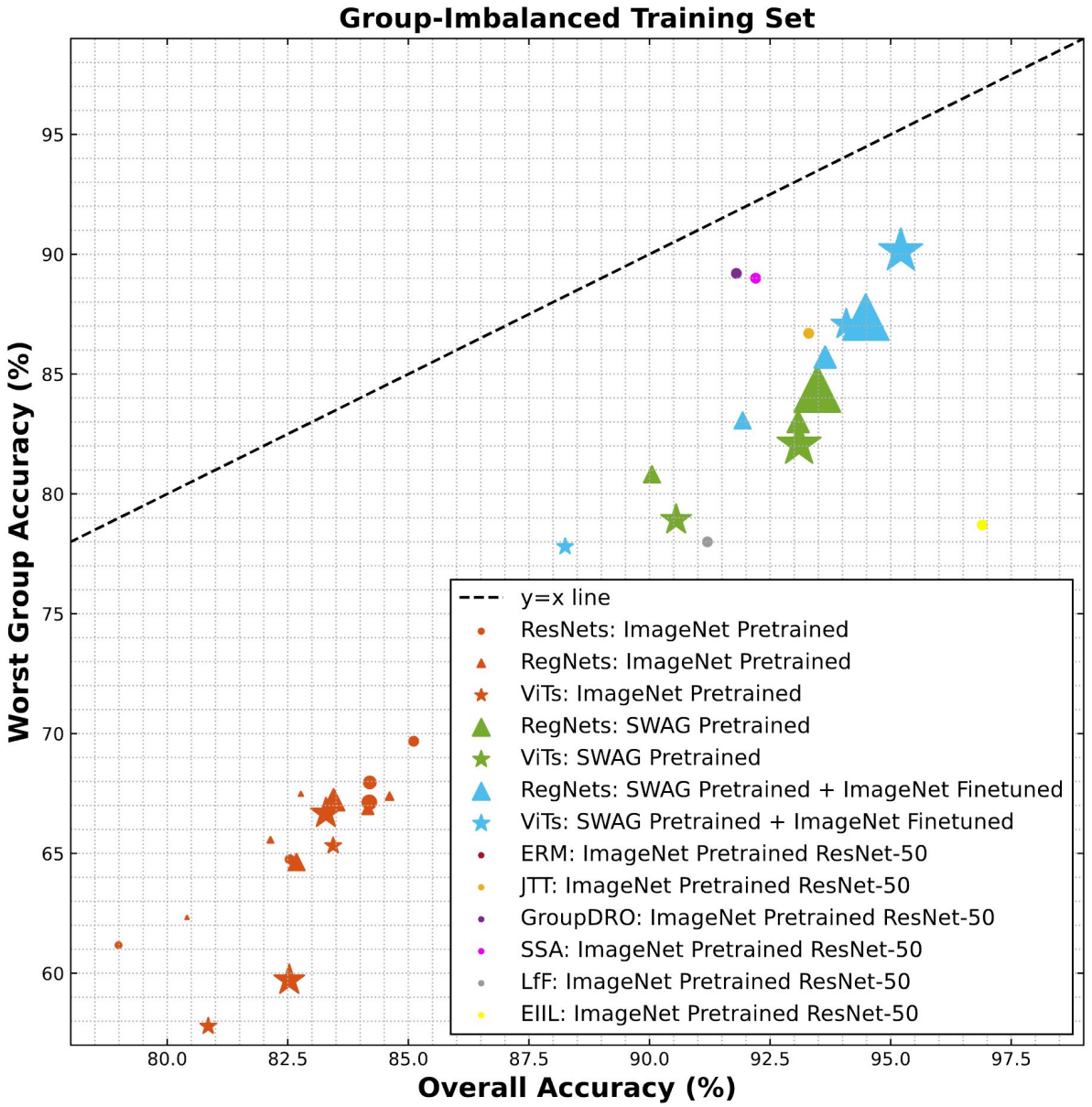


Singh, Mannat, et al. "Revisiting Weakly Supervised Pre-Training of Visual Perception Models." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022. https://github.com/facebookresearch/SWAG

Ex. RegNetY 128 GF

Results

Linear models trained on embeddings extracted from higher capacity networks pretrained on larger datasets perform best.



You Only Need a Good Embeddings Extractor to Fix Spurious Correlations?

Raghav Kiranbhai Mehta, Vítor Albiero, Li Chen, **Ivan Evtimov**, Tamar Glaser, Zhiheng Li, Tal Hassner

RCV @ ECCV 2022

Meta AI