Tradeoffs Between Contrastive and Supervised Learning: An Empirical Study

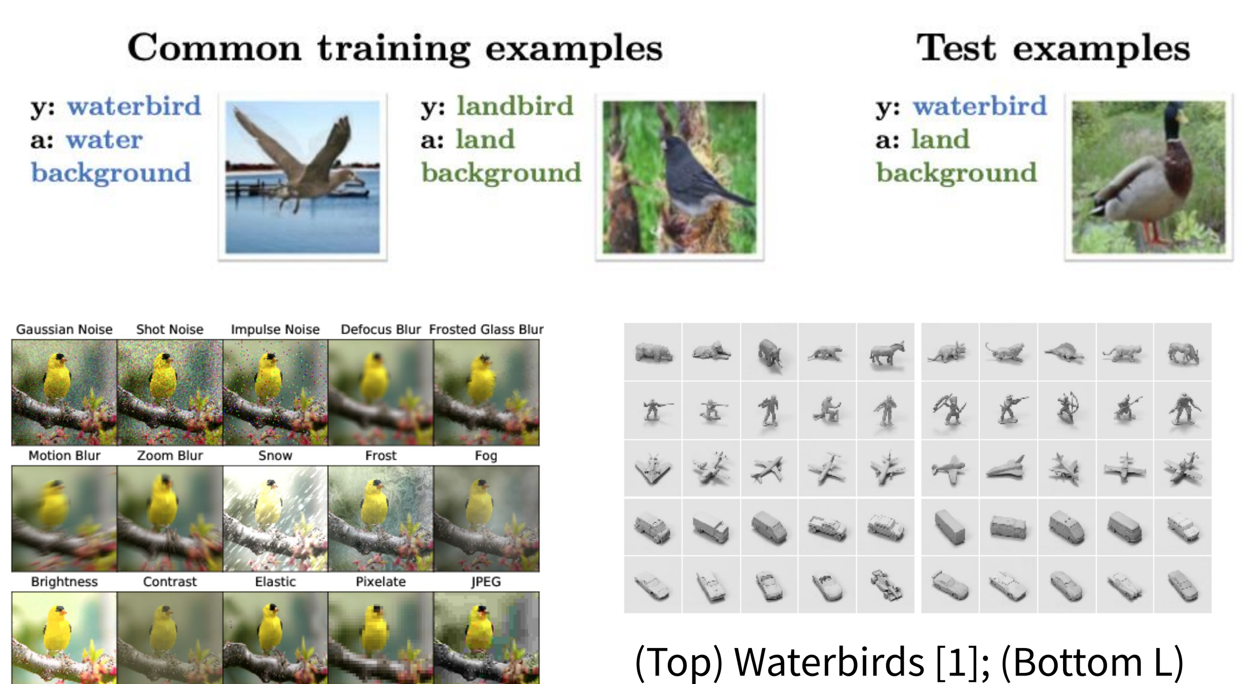
Ananya Karthik, Mike Wu, Noah Goodman, Alex Tamkin
{ananya23, wumike, ngoodman, atamkin} @stanford.edu
Department of Computer Science, Stanford University

Motivation

- Self-supervised pretraining for computer vision has grown in popularity recently due to the cost of annotating data
- Contrastive learning has achieved state-of-the-art results, underscoring the need for studying the real-world tradeoffs between contrastive and supervised pretraining. Specifically:
  1. Is contrastive learning better **across all compute budgets**?
  2. For larger compute budges, is supervised pretraining better **on tasks where an object-centric bias is important**?

Methodology

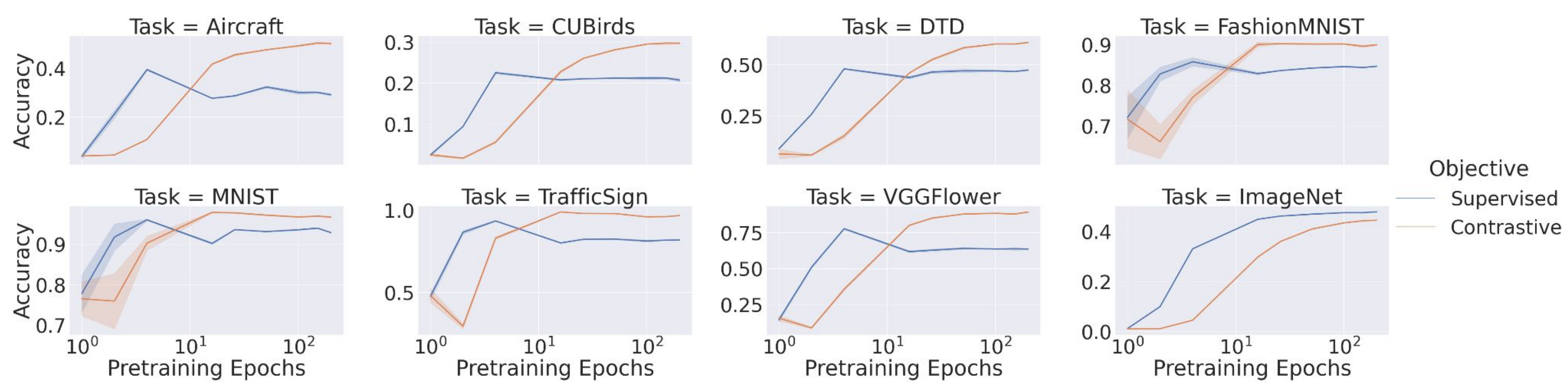
- Experimental Settings
  - Pretraining: 2 ResNet-18 models on ImageNet (200 epochs)
    - Standard cross entropy loss for supervised, InfoNCE objective for the contrastive model
  - Transfer: Linear evaluation protocol (100 epochs)
  - Datasets:
    - Q1 (Transfer across compute budgets): Aircraft, CUBirds, FashionMNIST, DTD, TrafficSign, MNIST, VGGFlower, ImageNet
    - Q2 (Object-centric bias): Waterbirds, Norb, ImageNet-C (below)

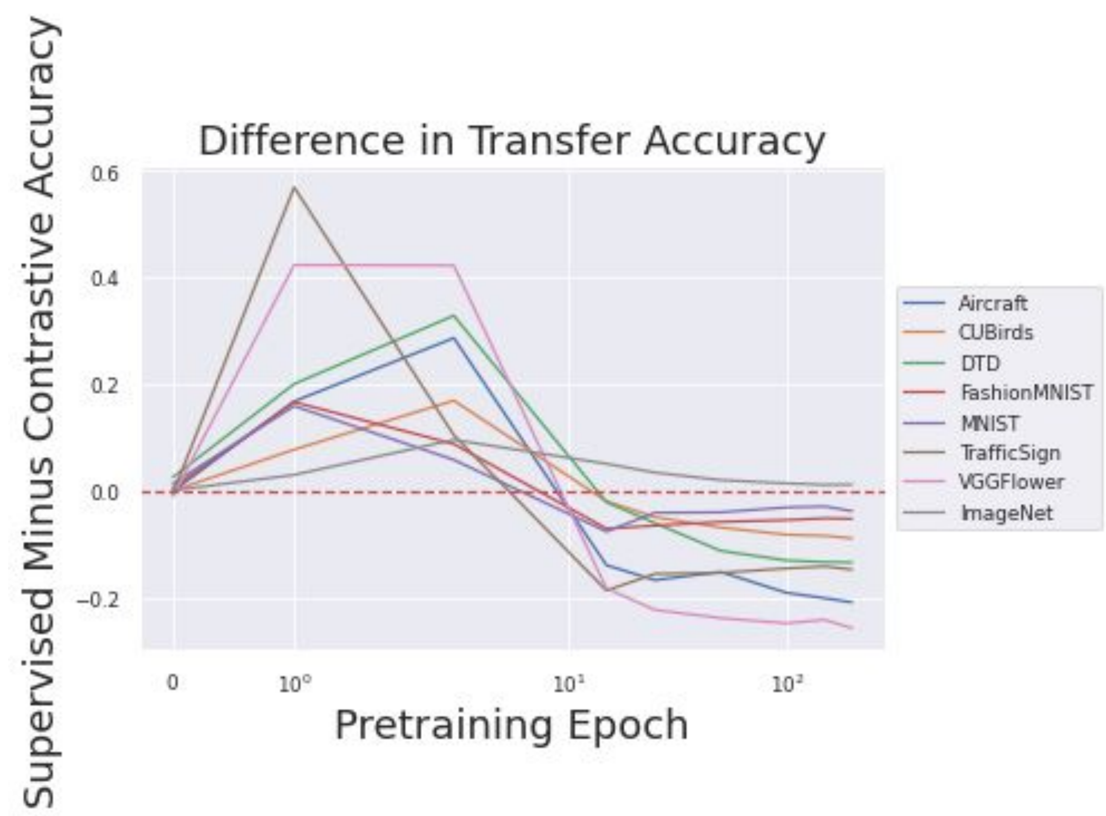


(Top) Waterbirds [1]; (Bottom L) ImageNet-C [2]; (Bottom R) NORB [3]

Results

1. Learning Dynamics Across Compute Budgets



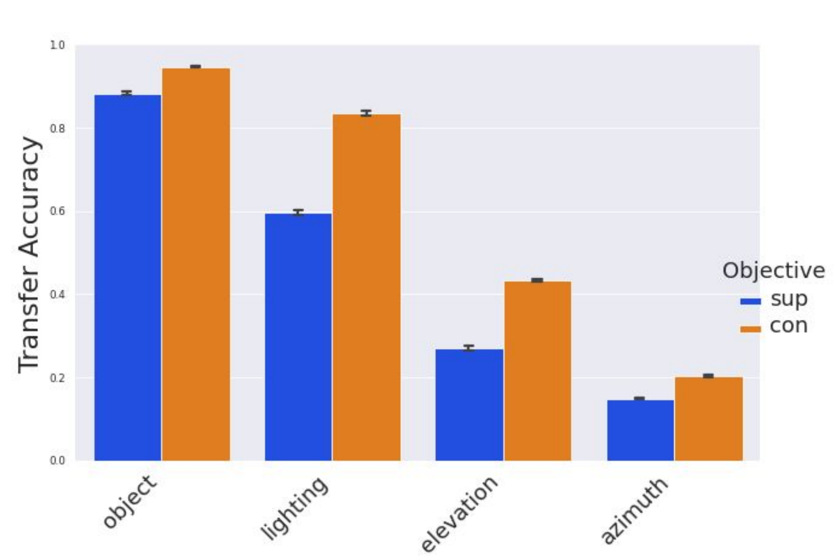


Task	Supervised	Contrastive
Aircraft	29.1 ± 0.3	$\mathbf{50.0} \pm 0.3$
CUBirds	20.7 ± 0.3	$\mathbf{29.7} \pm 0.2$
FashionMNIST	84.6 ± 0.1	$\mathbf{89.9} \pm 0.1$
DTD	47.4 ± 0.3	$\mathbf{60.8} \pm 0.2$
TrafficSign	81.8 ± 0.2	$\mathbf{96.6} \pm 0.1$
MNIST	92.8 ± 0.1	$\mathbf{96.7} \pm 0.1$
VGGFlower	63.6 ± 0.2	$\mathbf{89.4} \pm 0.1$
ImageNet	$\mathbf{47.8} \pm 0.0$	44.4 ± 0.0

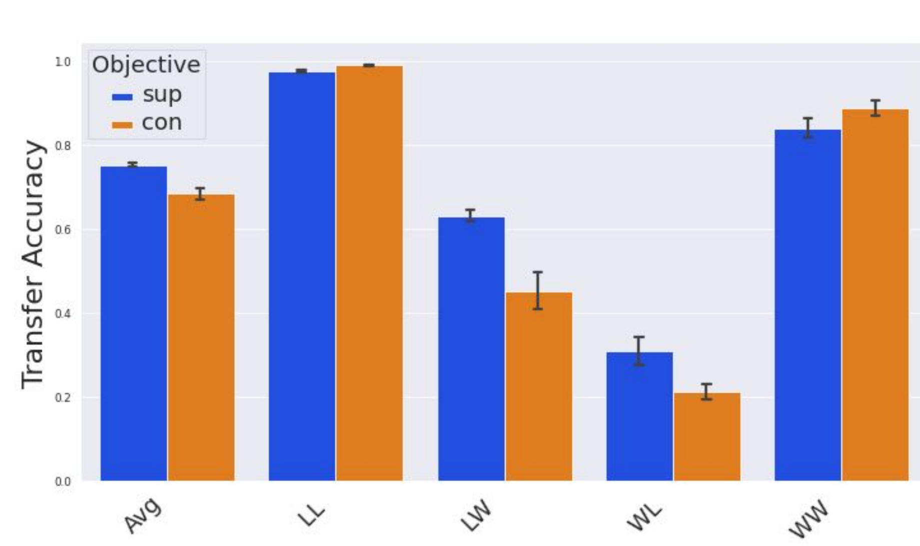
Downstream accuracy of contrastive and supervised models on 8 transfer tasks for different pretraining budgets

2. Downstream Effects of Biases Acquired During Pretraining

NORB



Waterbirds



ImageNet-C

Supervised	Contrastive
91.08 +/- 0.279%	95.41 +/- 0.157%

Relative mCE (Mean Corruption Error): performance degradation from clean to corrupted data, lower is better

Does the object-centric bias of supervised learning improve downstream performance on transfer tasks?
We find strong effects for Waterbirds and ImageNet-C, but weaker effects for NORB.

Conclusions

- Contrastive learning is **not necessarily better across all compute budgets**: different pretraining algorithms produce better representations at different budgets
  - Transfer performance does not increase monotonically across pretraining → potential misalignment between representations learned for pretraining vs transfer
  - While the contrastive model eventually achieves higher performance, for the first 10-15 epochs the supervised model yields better representations for downstream tasks → potential differences in the two representation learning processes
  - We encourage developers of new pretraining techniques to release learning dynamics curves
- Contrastive learning is **not necessarily better across all tasks**: the supervised model eventually achieves worse downstream accuracy on most tasks, but the object-centric bias of ImageNet pretraining aids transfer on some tasks, especially WaterBirds (reliance on spurious correlations) and ImageNet-C (robustness to common corruptions)

Future Work

- Investigating whether these conclusions hold across a wide range of architectures, hyperparameters, datasets, and training objectives
- Exploring other dimensions along which pretraining algorithms differ (e.g. Cole et al. 2021 and Horn et al. 2021 find that supervised learning tends to perform better on fine-grained classification tasks)
- Studying how pretraining objectives shape the behavior of models in ambiguous scenarios

[1] Sagawa et al. 2019
[2] Hendrycks and Dietterich 2019
[3] LeCun et al. 2004