Self-supervised learning for zero-shot tracking

Deep Learning meetup June 19th

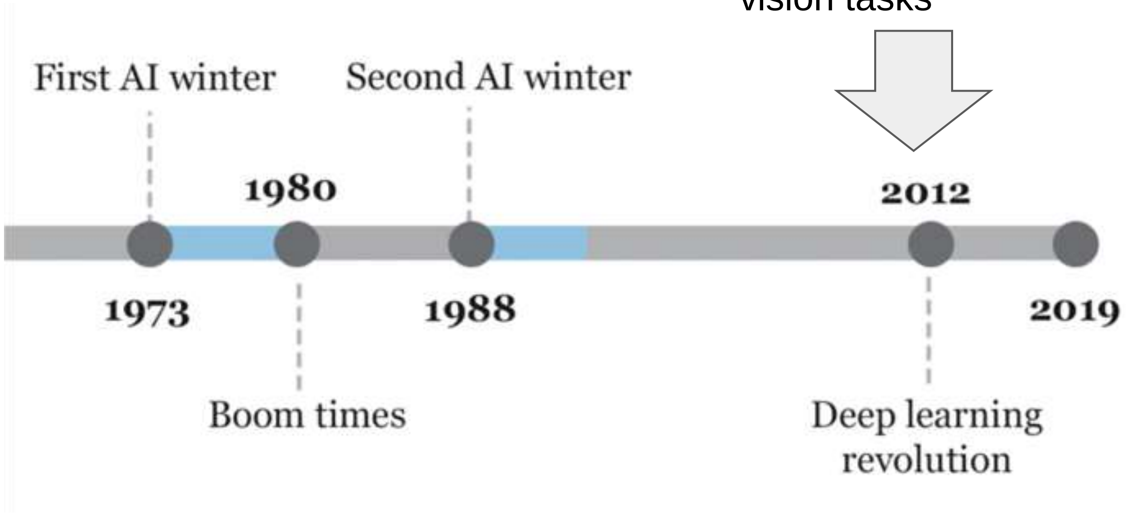
Charles Fieseler

Overview

1. History up until self-supervised learning (SSL)
2. History of SSL
3. Neuroscience connection
4. My work and application

History of neural networks

Neural networks
dominate all computer
vision tasks



https://towardsdatascience.com/history-of-the-first-ai-winter-6f8c2186f80b

“The bitter lesson”

“The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin”

https://www.cs.utexas.edu/~eunsol/courses/data/bitter_lesson.pdf

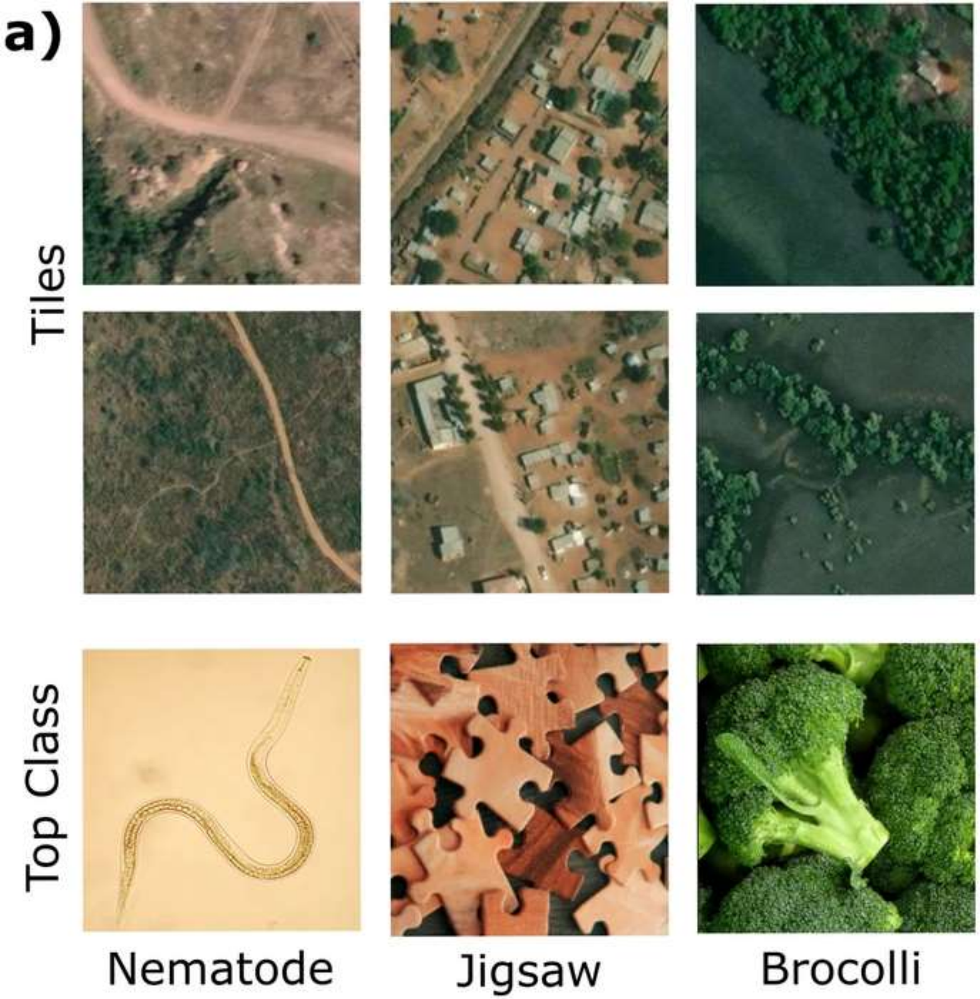
Limitations of supervised learning

1. Tons of annotations required
2. Problems with domain shift

Neal, I., Seth, S., Watmough, G. and Diallo, M.S., 2022. Census-independent population estimation using representation learning. *Scientific Reports*, *12*(1), pp.1-12.

Limitations of supervised learning

1. Tons of annotations required
2. Problems with domain shift



Neal, I., Seth, S., Watmough, G. and Diallo, M.S., 2022. Census-independent population estimation using representation learning. *Scientific Reports*, *12*(1), pp.1-12.

“The bitter lesson”

“The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin”

“The bitter lesson is based on the historical observations that 1) AI researchers have often tried to build knowledge into their agents, 2) this always helps in the short term, and is personally satisfying to the researcher, but 3) in the long run it plateaus and even inhibits further progress”

https://www.cs.utexas.edu/~eunsol/courses/data/bitter_lesson.pdf

“The bitter lesson”

“The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin”

“The bitter lesson is based on the historical observations that 1) AI researchers have often tried to build knowledge into their agents, 2) this always helps in the short term, and is personally satisfying to the researcher, but 3) in the long run it plateaus and even inhibits further progress”

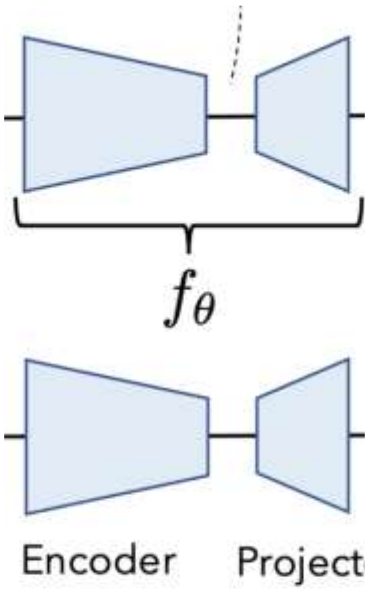
… so how can we “leverage computation” to overcome the supervised learning limitations?

https://www.cs.utexas.edu/~eunsol/courses/data/bitter_lesson.pdf

Transfer learning

Admit: we don’t know what the network learned

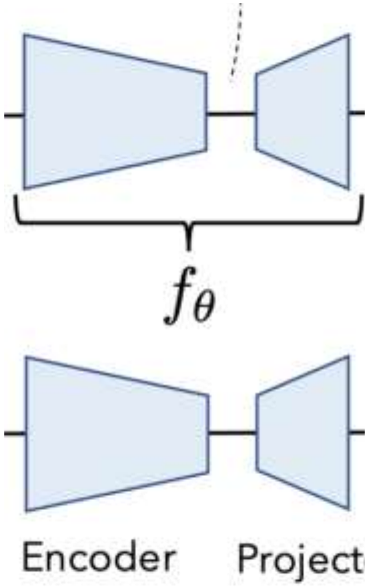
But: use it anyway!



Transfer learning

Admit: we don’t know what the network learned

But: use it anyway!

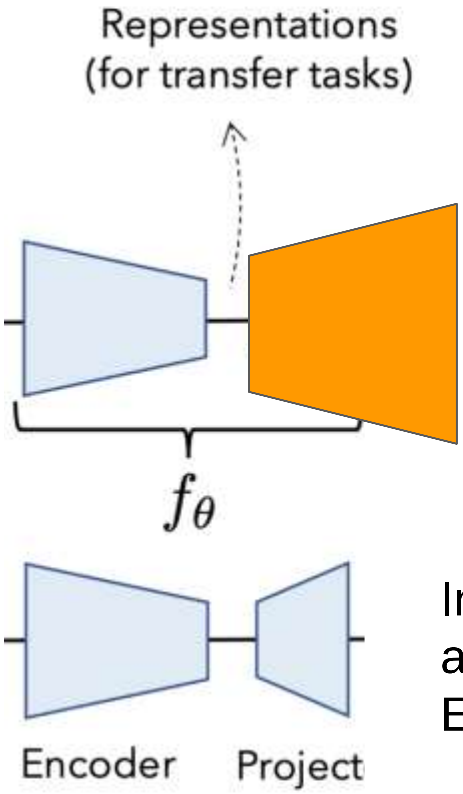


In the original network, solve a standard task
Example: classify images

Transfer learning

Admit: we don't know what the network learned

But: use it anyway!



Add a different ½ network to solve a different task

In the original network, solve a standard task
Example: classify images

Small summary

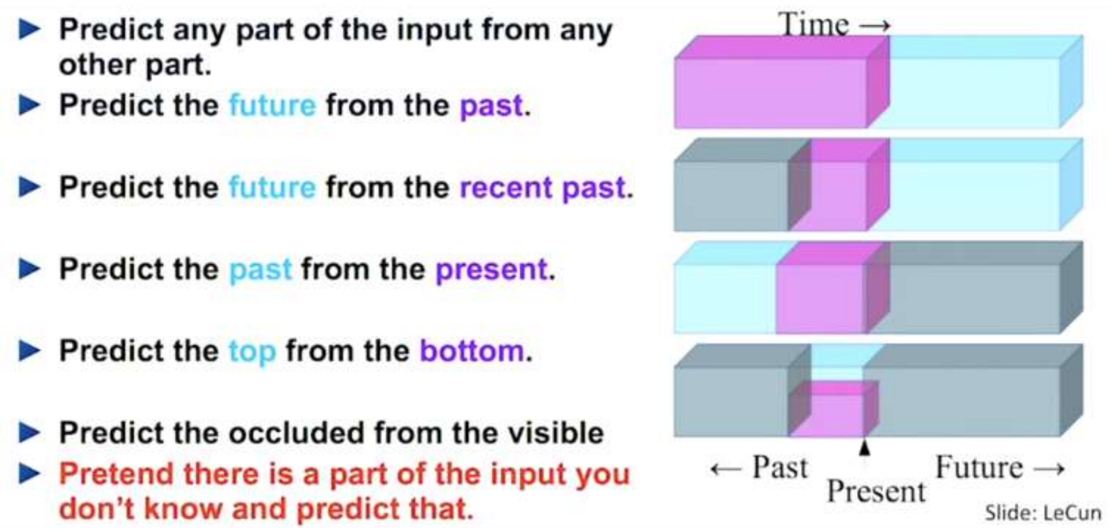
Supervised learning -> amazing but expensive

Transfer learning -> effective if you have a similar dataset

Another idea: what if the network can “self-label” the data?

Idea: self-supervised learning

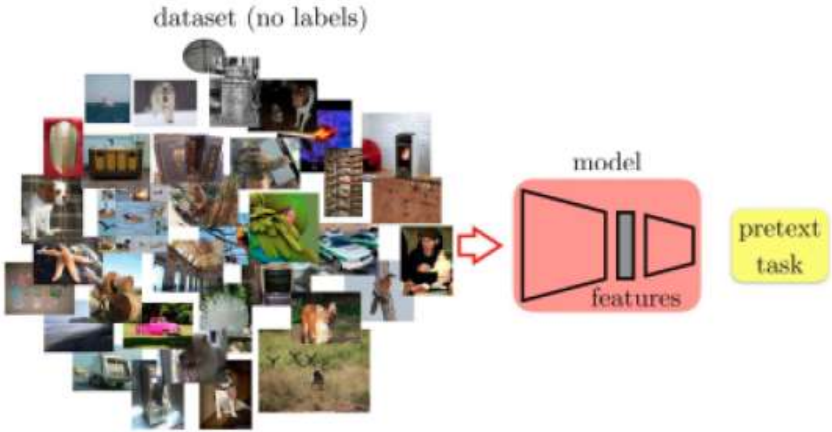
- Example: language processing
- Unclear which transformations really work, or why.
- Prediction is key!
- “Pretext tasks”



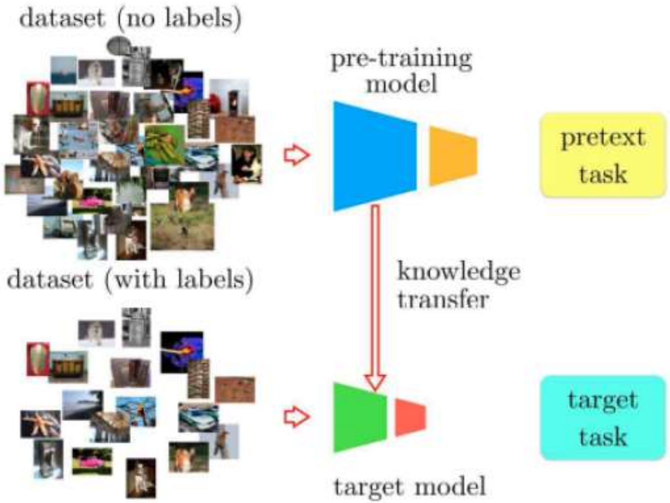
https://neptune.ai/blog/self-supervised-learning

Self-supervised learning is similar to transfer learning

Step 1: build a feature space

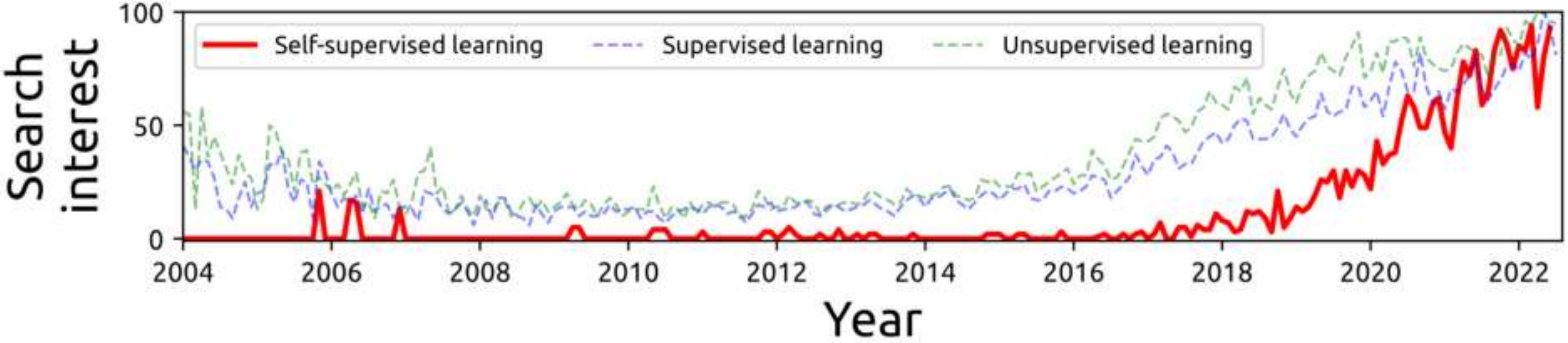


Step 2: fine-tune with few labels



https://neptune.ai/blog/self-supervised-learning

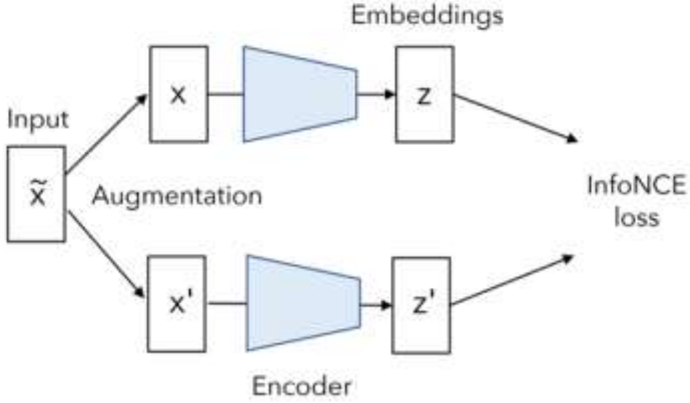
History of SSL



If intelligence is a cake, the bulk of the cake is **self-supervised learning**, the icing on the cake is supervised learning, and the cherry on the cake is reinforcement learning
 - Yann LeCun (2020)

https://openreview.net/pdf?id=Ma25S4ludQ

A specific example, and a major problem



(a) embedding space

Goal: Make the representations similar

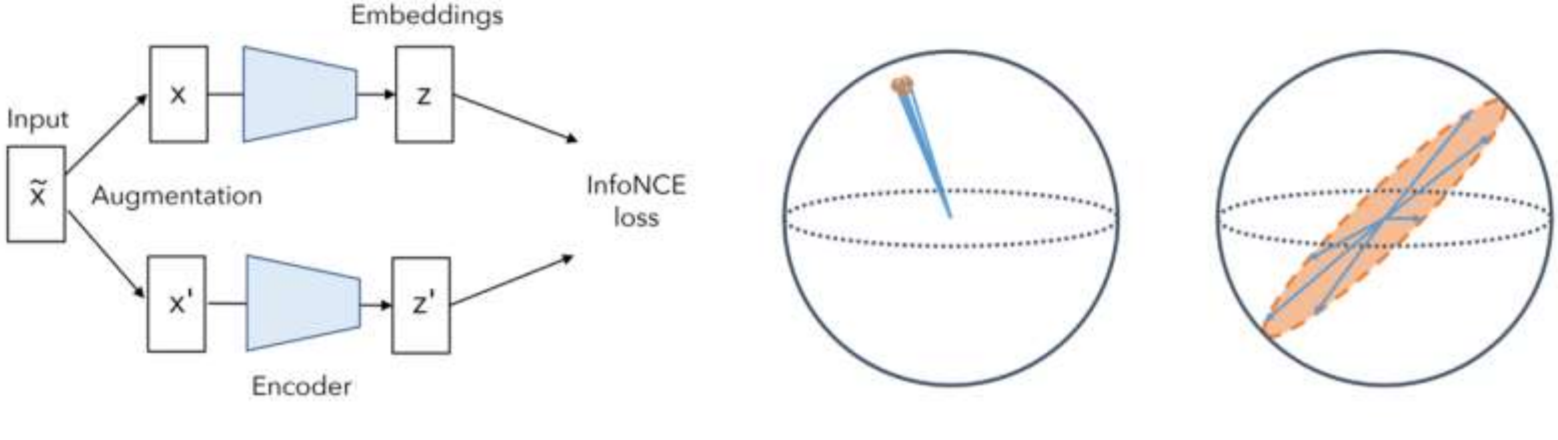
https://arxiv.org/pdf/2110.09348.pdf

(a) Original (b) Crop and resize (c) Crop, resize (and flip) (d) Color distort. (drop) (e) Color distort. (jitter)

(f) Rotate $\{90°, 180°, 270°\}$ (g) Cutout (h) Gaussian noise (i) Gaussian blur (j) Sobel filtering

https://arxiv.org/pdf/2002.05709v3

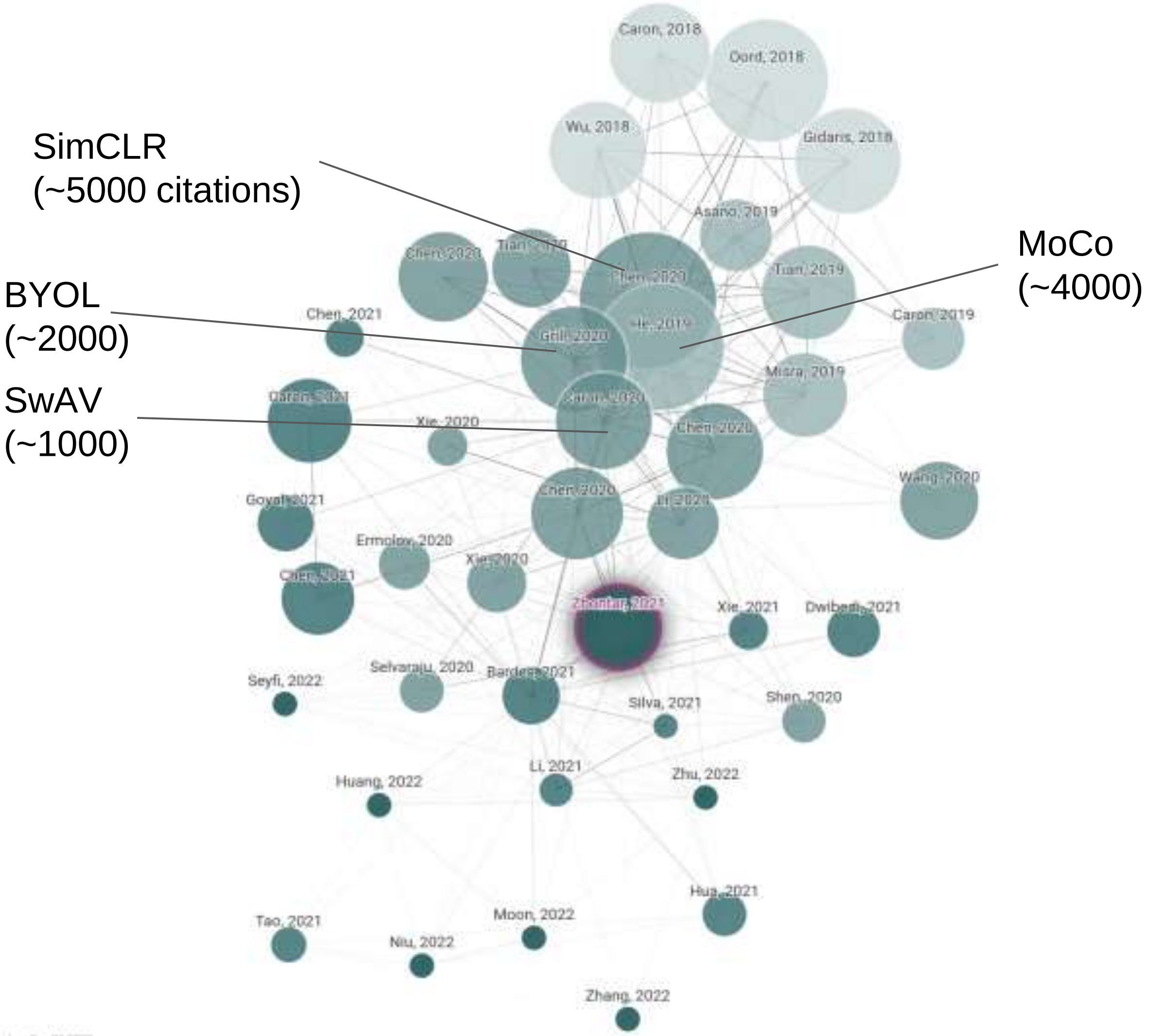
A specific example, and a major problem



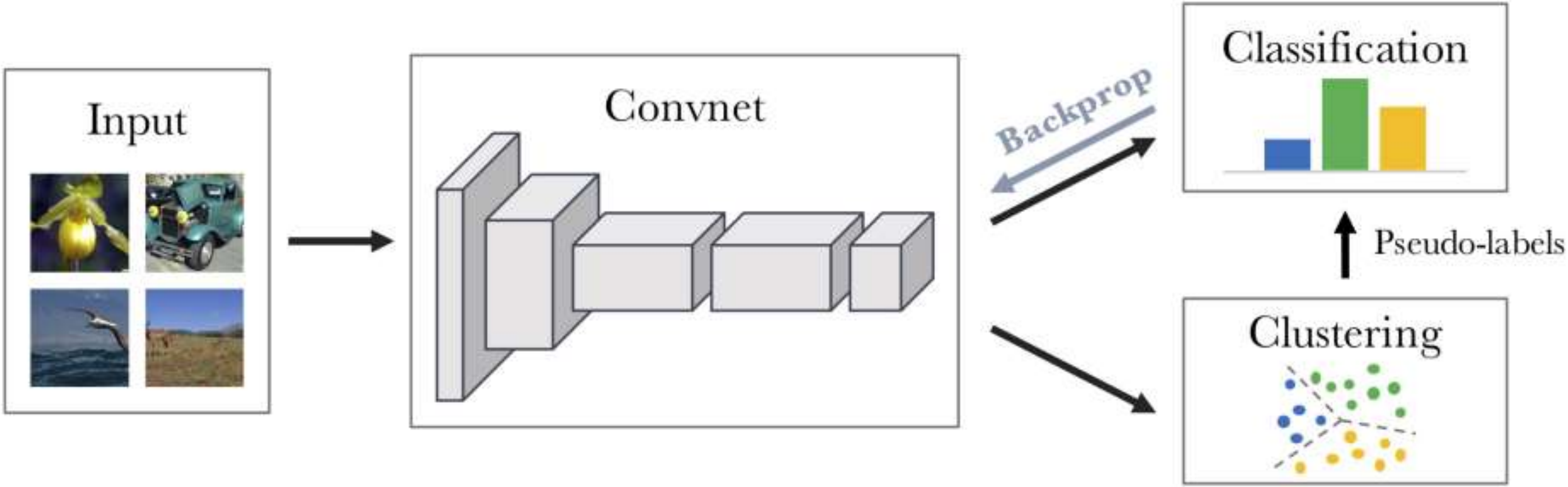
(a) embedding space (b) complete collapse (c) dimensional collapse

https://arxiv.org/pdf/2110.09348.pdf

LOTS of highly influential papers… but (to me) they are very complicated to implement!



Anti-collapse strategies: clustering (Deep Cluster, 2018; Swav, 2020)

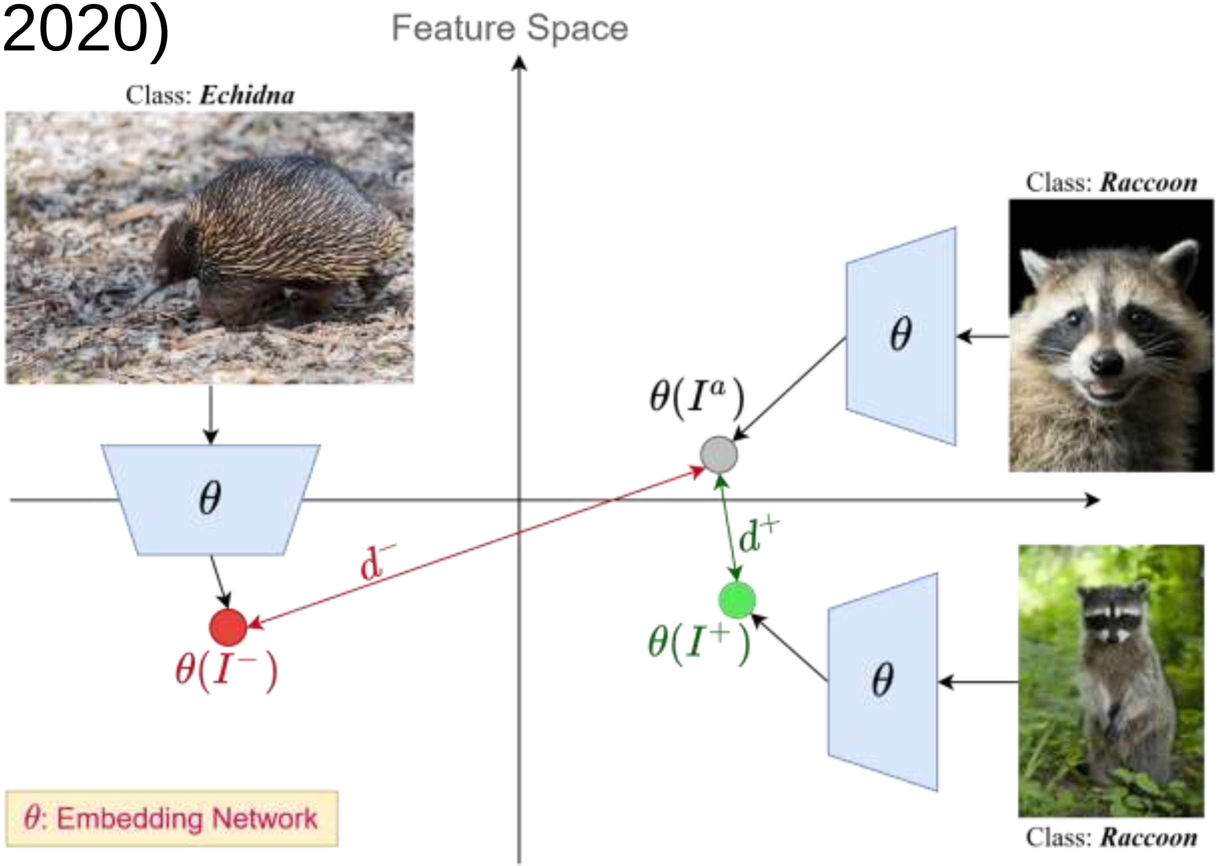


https://arxiv.org/pdf/1807.05520

Anti-collapse strategies: clustering (Deep Cluster, 2018; Swav, 2020)

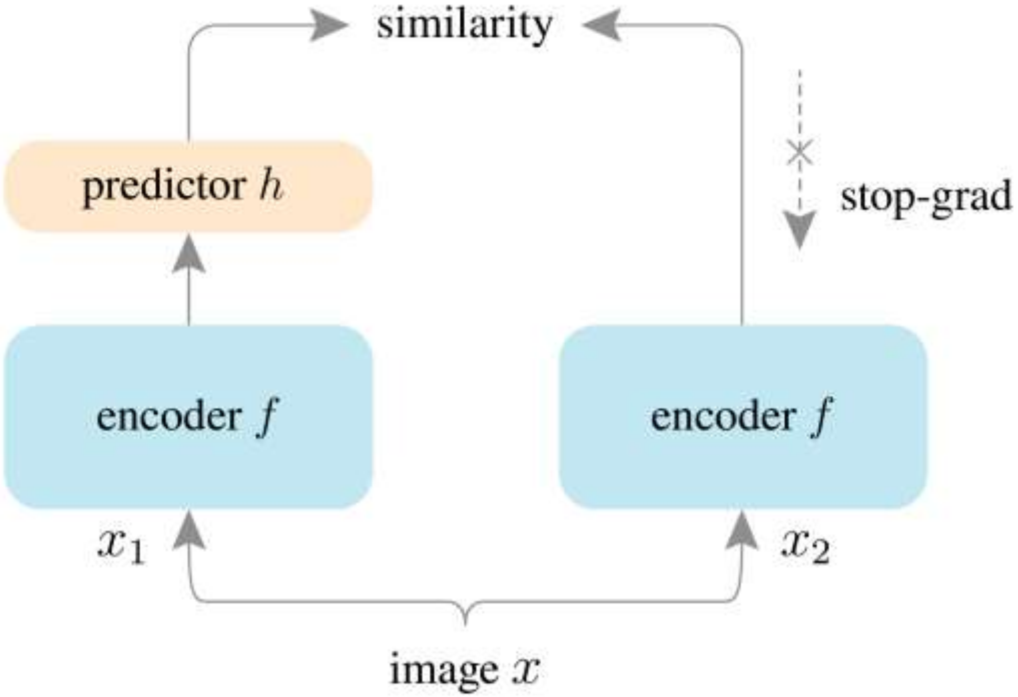
https://github.com/facebookresearch/swav

Anti-collapse strategies: negative examples (MoCo, 2020; SimCLR, 2020)



https://www.v7labs.com/blog/contrastive-learning-guide

Anti-collapse strategies: distillation/asymmetry (BYOL 2020; SimSIAM, 2021; DINO, 2021)



https://github.com/facebookresearch/simsiam

A small problem: resource usage

- “the training of MoCo-v3 with a vision transformer backbone requires approximately 625 TPU days.”
- “a vast majority of the [SSL papers] have at least one author with an industry affiliation”

https://openreview.net/pdf?id=Ma25S4ludQ

Small summary

Supervised learning -> amazing but expensive

Transfer learning -> effective but limited

Self-supervised learning -> Amazing but needs tricks + resources to actually train

Idea: can interdisciplinary ideas help?

Horace Barlow

Theory work:
Redundancy reduction

i.e. given a complex, unlabeled world, try to encode objects by decorrelating the representations in the brain

https://www.theguardian.com/science/2020/aug/23/horace-barlow-obituary
https://www.nature.com/articles/s41593-020-00708-1

“The bitter lesson”

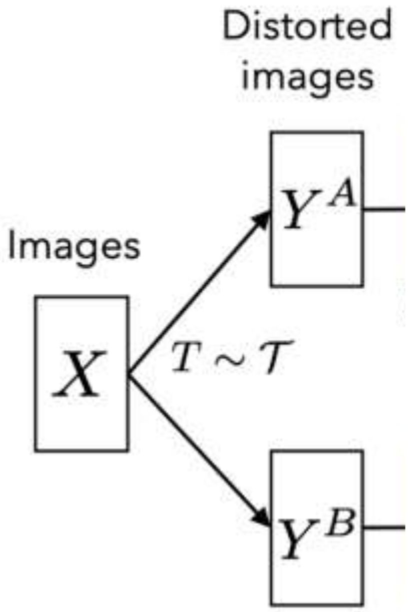
“The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin”

“The bitter lesson is based on the historical observations that 1) AI researchers have often tried to build knowledge into their agents, 2) this always helps in the short term, and is personally satisfying to the researcher, but 3) in the long run it plateaus and even inhibits further progress”

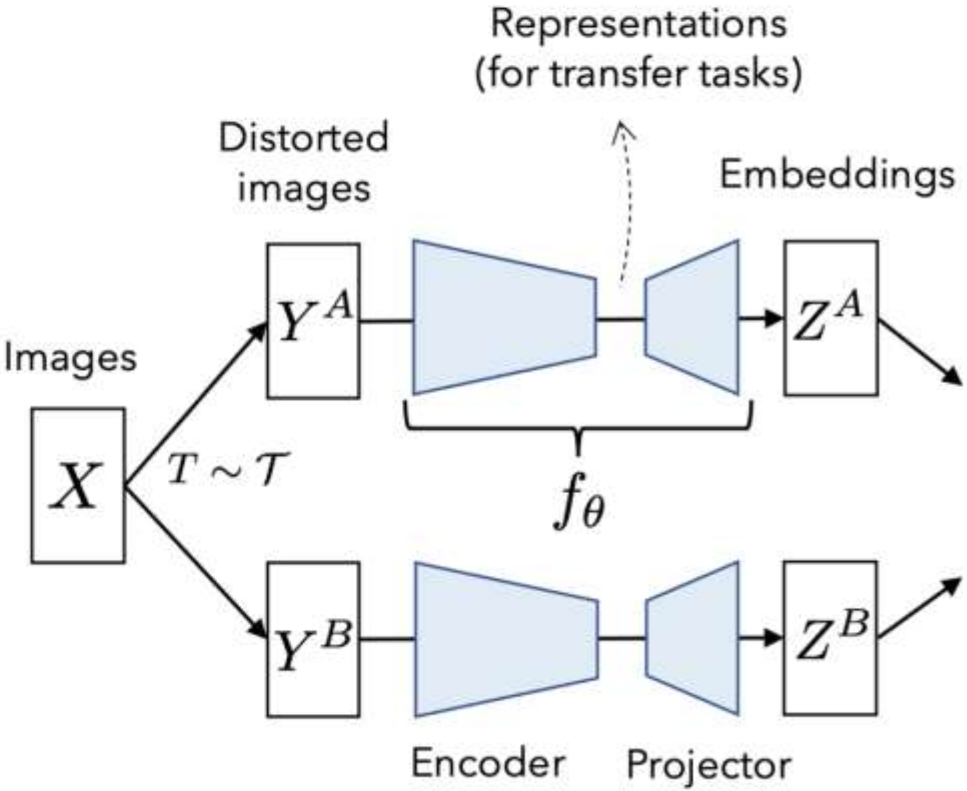
“we should stop trying to find simple ways to think about the contents of minds… instead we should build in only the meta-methods that can find and capture this arbitrary complexity”

https://www.cs.utexas.edu/~eunsol/courses/data/bitter_lesson.pdf

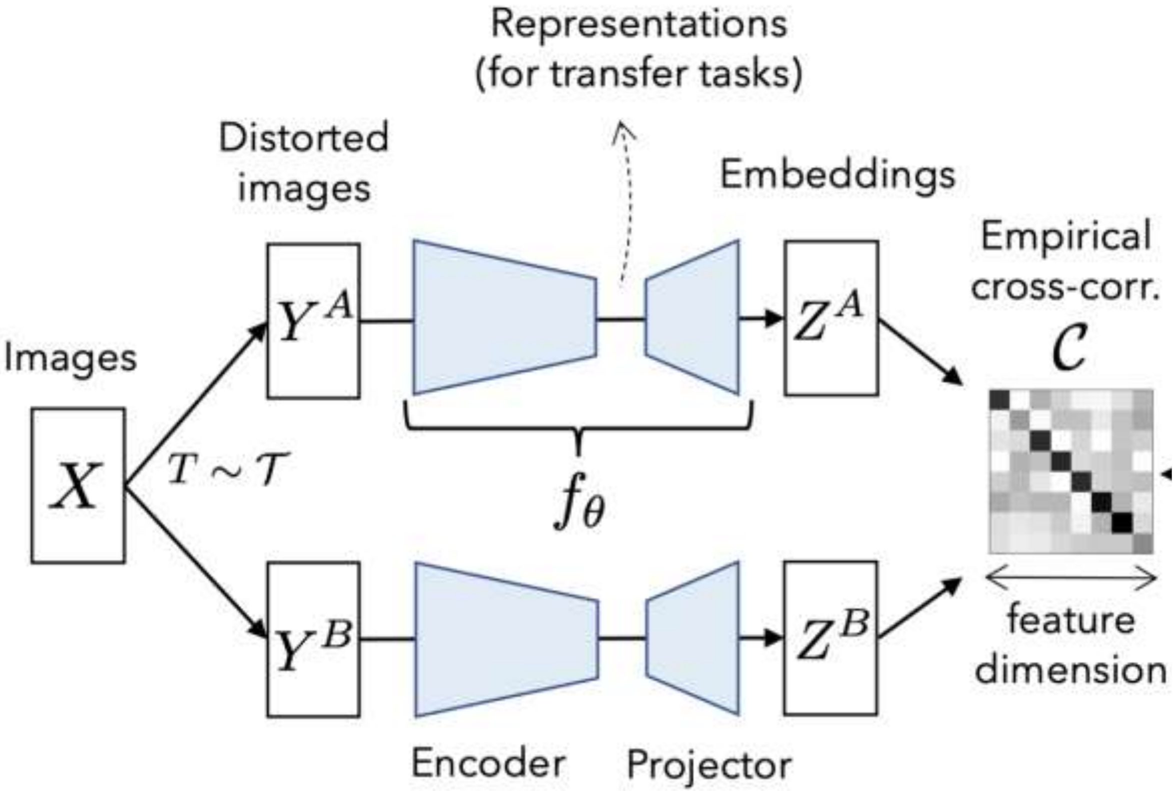
Barlow Twins



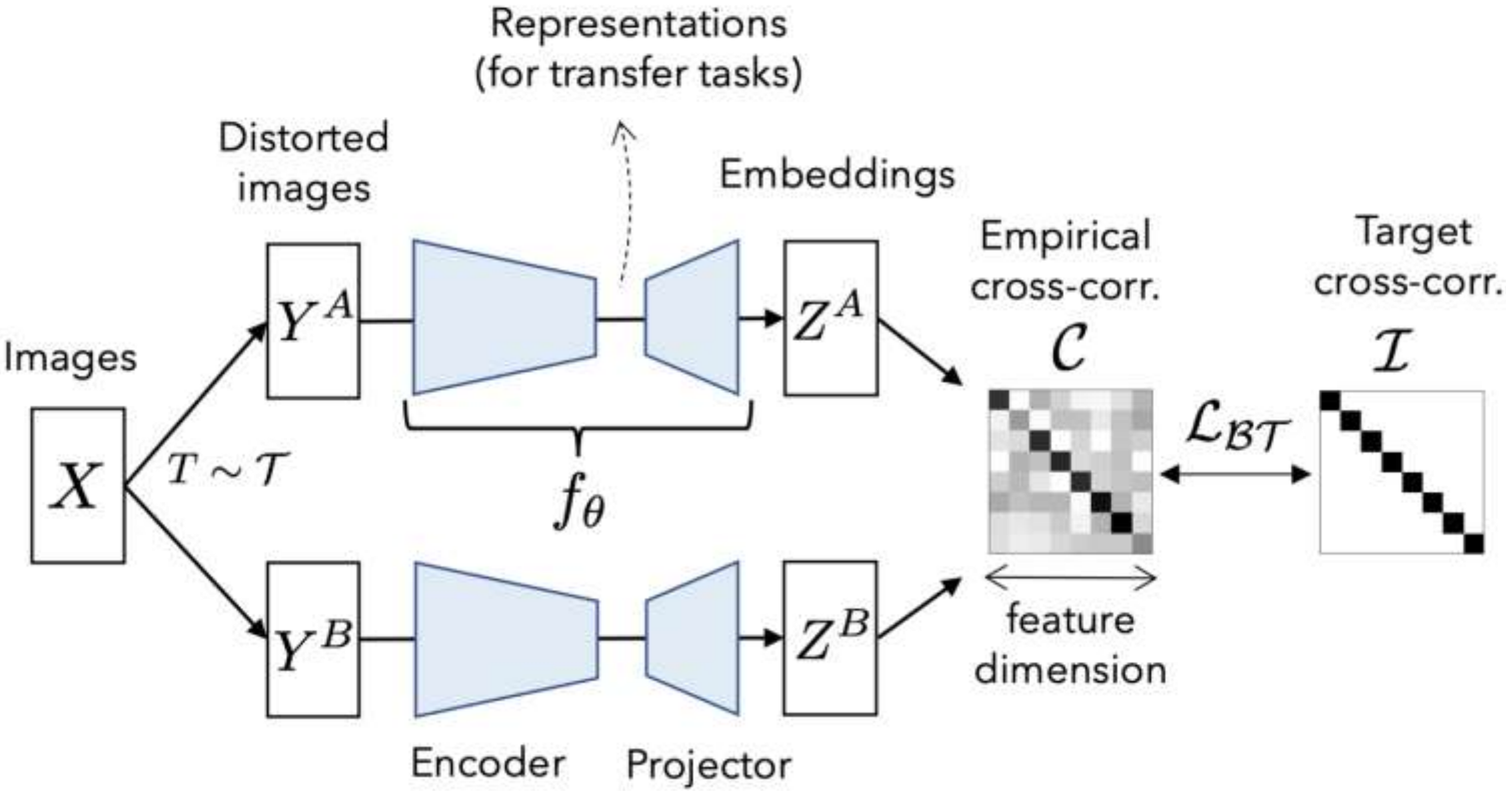
Barlow Twins



Barlow Twins



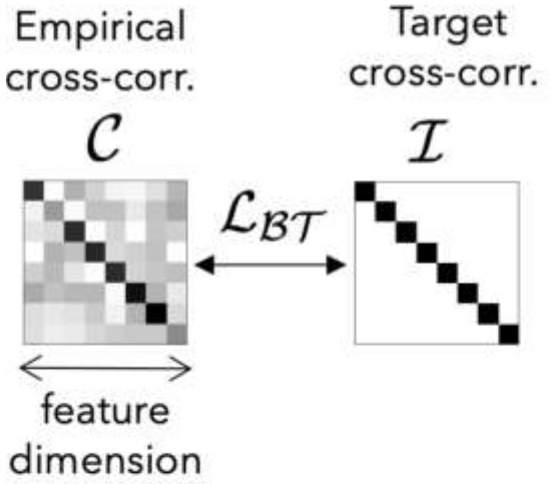
Barlow Twins



Barlow Twins

Two parts to the loss function:

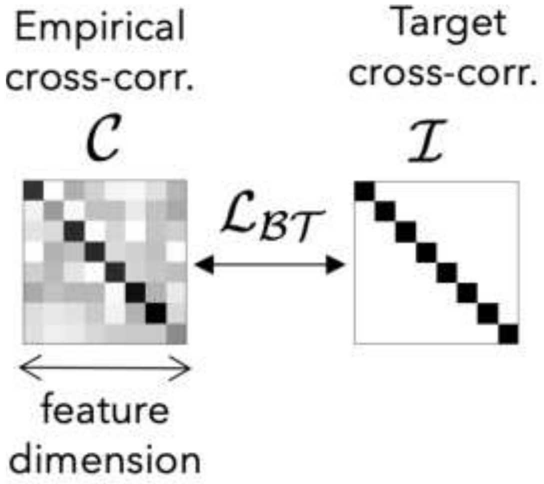
1. Diagonal - Representation should be invariant to (chosen) transformations
   a. “Invariance”
2. Off-diagonal - Feature i and feature j should not be the same (across input set)
   a. “High-dimensionality”

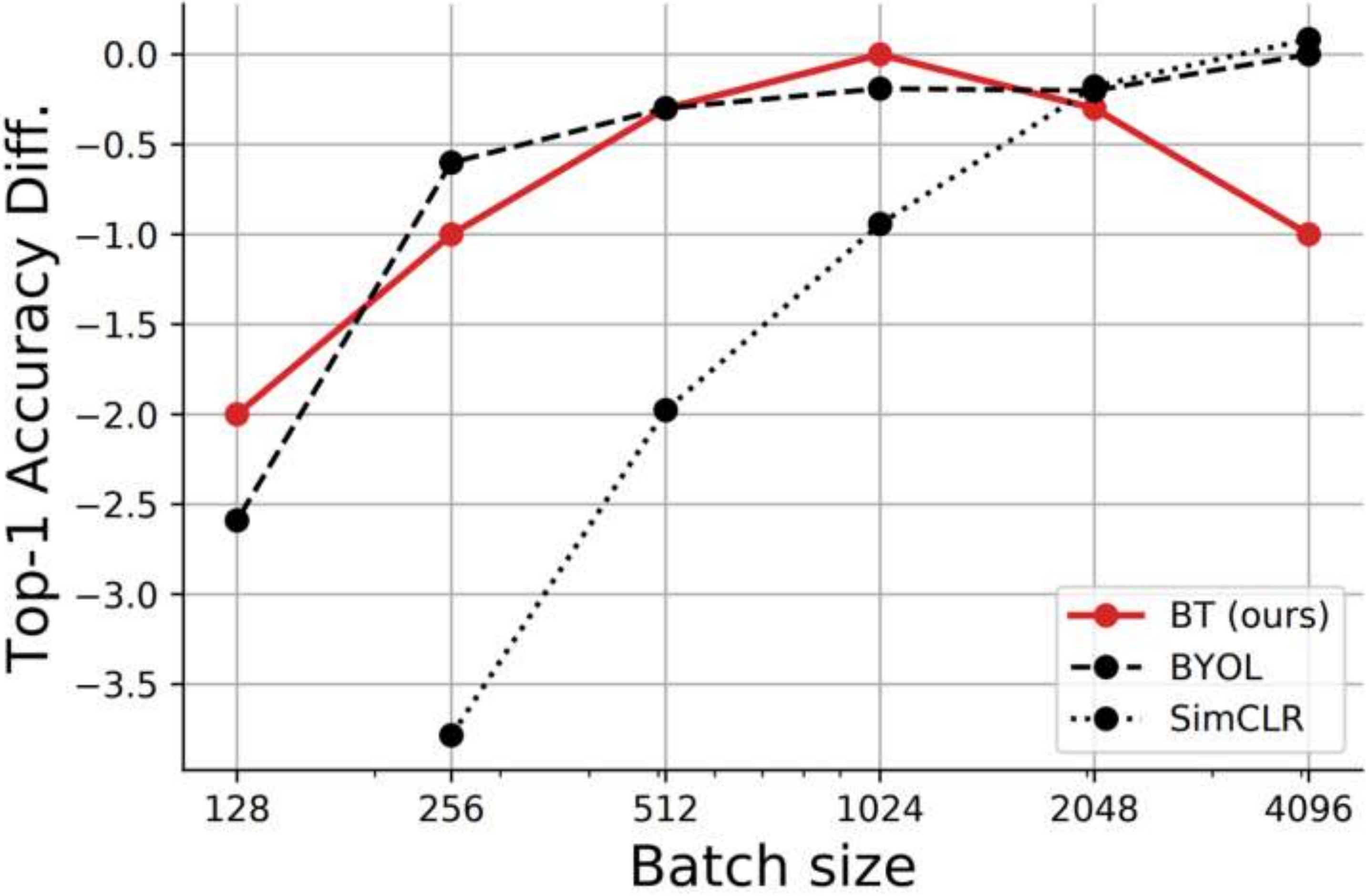


Anti-collapse strategies: covariance regularization (Barlow Twins, 2021; VicReG, 2022)

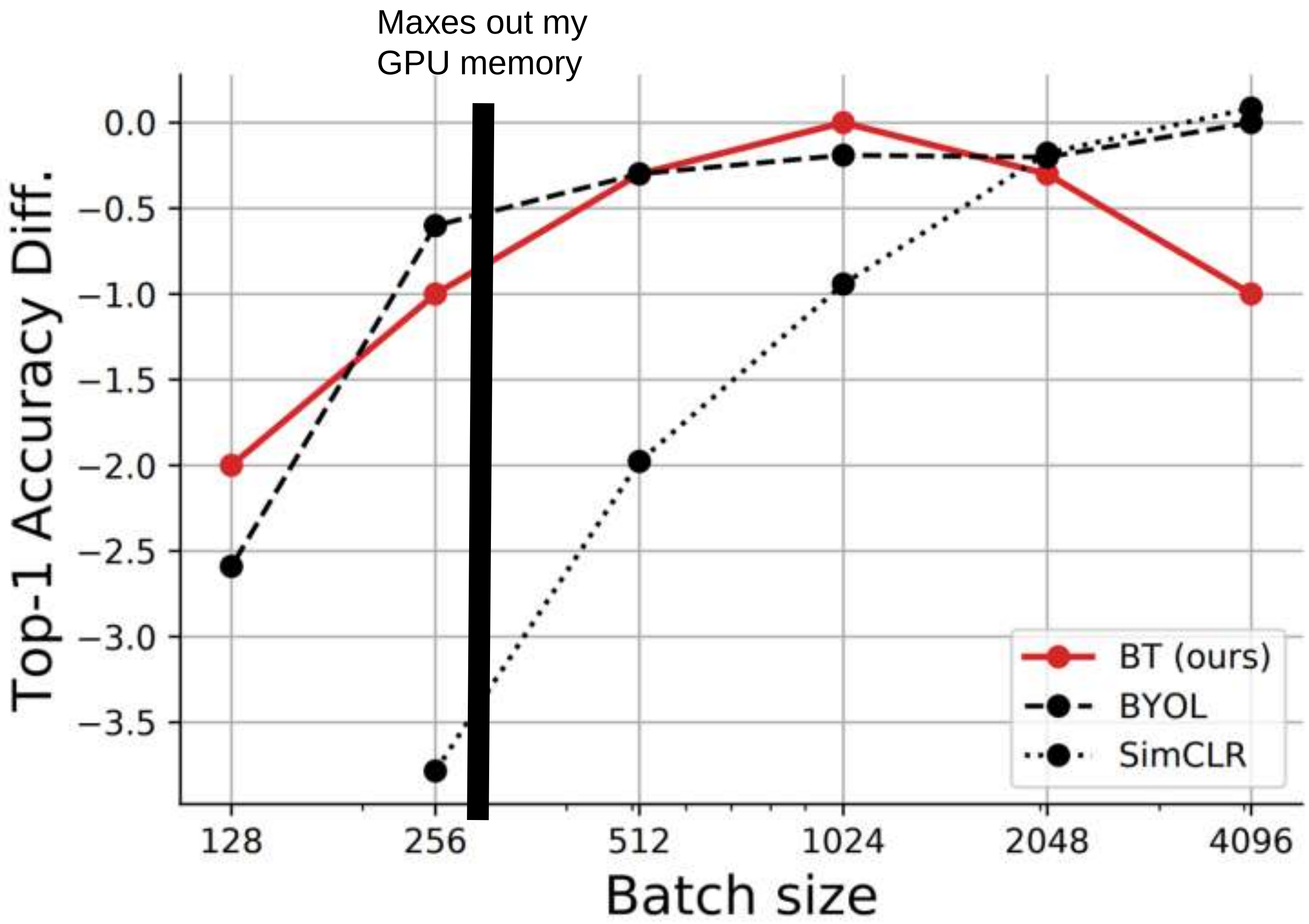
Two parts to the loss function:

1. Diagonal - Representation should be invariant to (chosen) transformations
   a. “Invariance”
2. Off-diagonal - Feature i and feature j should not be the same (across input set)
   a. “High-dimensionality”





Top-1 Accuracy Diff.
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
−3.5
128
256
512
1024
2048
4096
Batch size
BT (ours)
BYOL
SimCLR

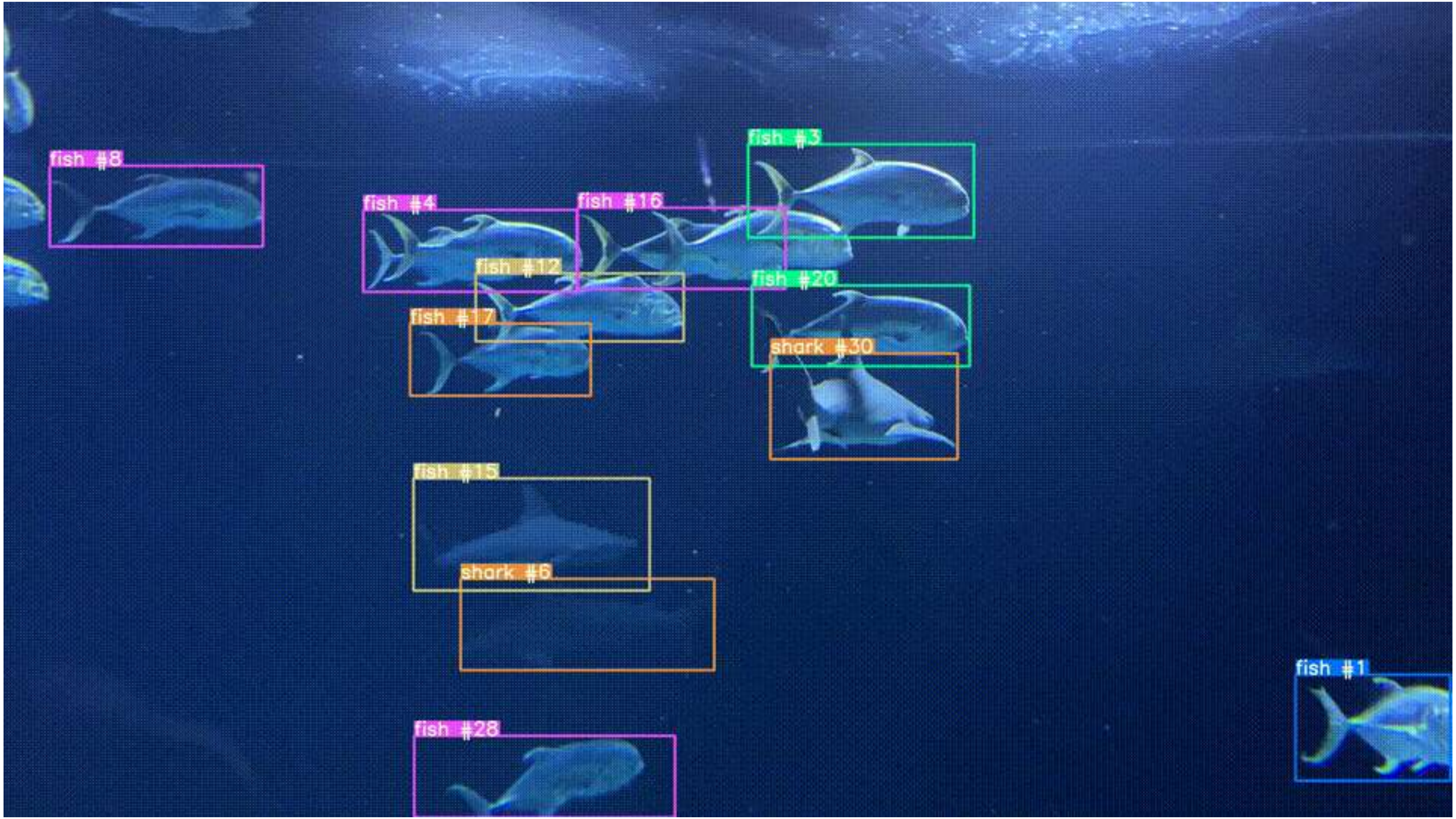

Maxes out my
GPU memory
Top-1 Accuracy Diff.
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
−3.5
128
256
512
1024
2048
4096
Batch size
BT (ours)
BYOL
SimCLR

Small summary

1. Self-supervised solves (partially) the annotation problem
2. BT is implementing a neuroscience hypothesis!
3. For academics:
   a. We are in a “low-labeled data” setting…
   b. We have tons of unlabeled data…
   c. … we need a method that doesn’t require Google/Meta/Huawei resources!
4. Next: applications in my own work

Small summary

1. Self-supervised solves (partially) the annotation problem
2. BT is implementing a neuroscience hypothesis!
3. For academics:
   a. We are in a “low-labeled data” setting…
   b. We have tons of unlabeled data…
   c. … we need a method that doesn’t require Google/Meta/Huawei resources!
4. Next: applications in my own work
5. … What is zero-shot tracking??



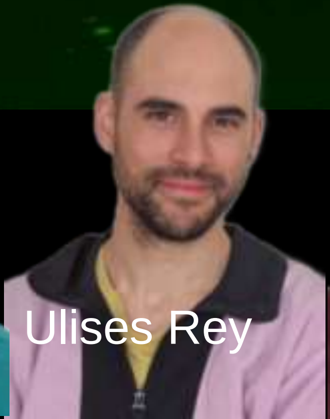
https://github.com/roboflow/zero-shot-object-tracking

Whole brain activity from freely moving animals

Behavior (80Hz)

Activity (3.5 Hz): NLSGCamp7b

Reference (3.5 Hz): NLSmScarlet

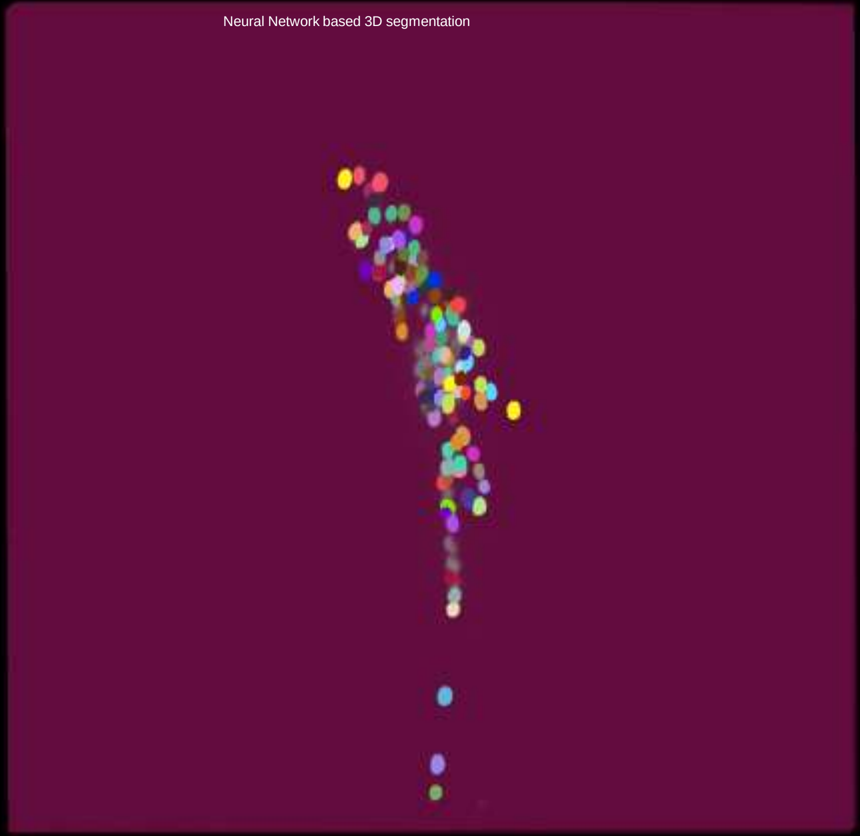


Live demo ☺

Live demo ☺

- Original goal: get ground truth
- … hundreds of hours of manual annotation
- … difficulties with domain adaptation

A pipeline to extract neuronal traces from freely moving recordings



Two challenging problems: segmentation + tracking

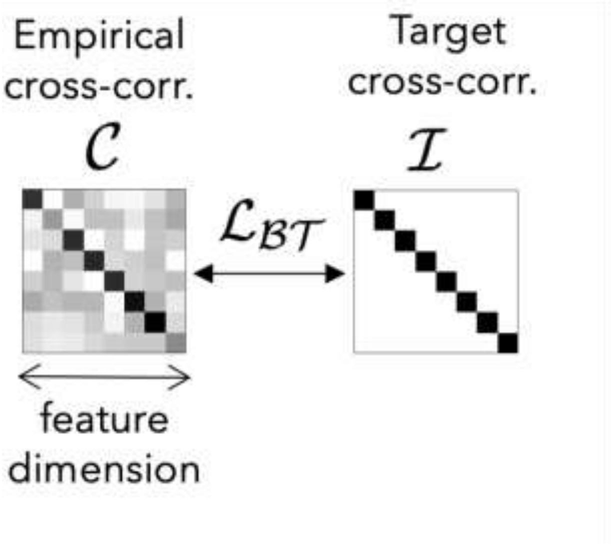
(Sarlin et al., 2020, CVF)

(Weigert et al., 2020, CVF)

"BarlowTrack"

In a video, you know all the objects are unique

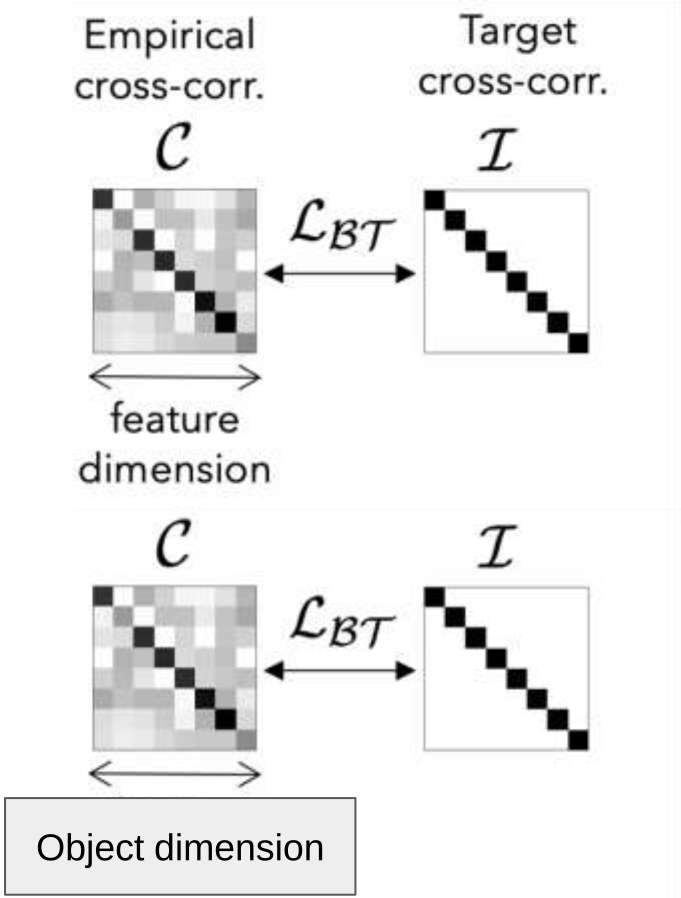
Strategy: add an additional loss term to decorrelate them!



“BarlowTrack”

In a video, you know all the objects are unique

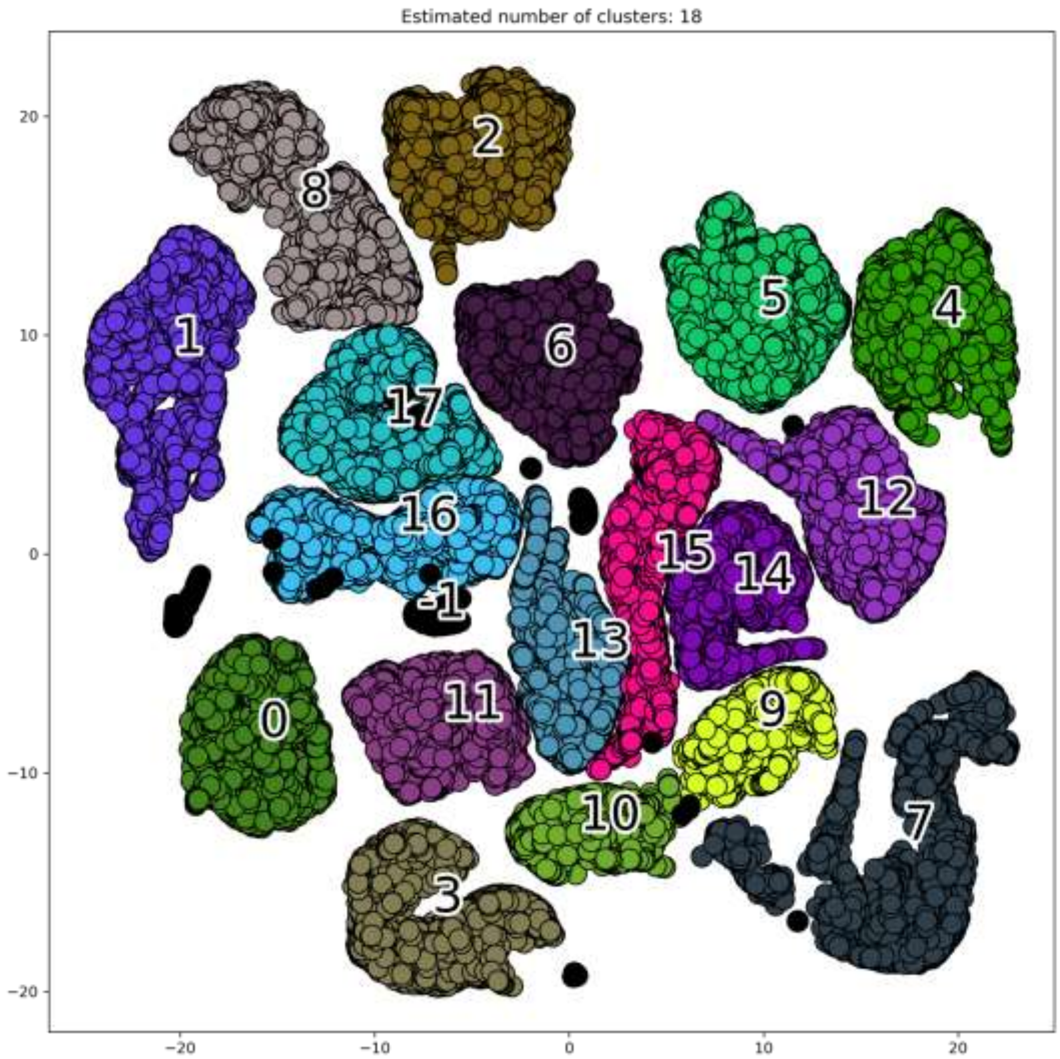
Strategy: add an additional loss term to decorrelate them!



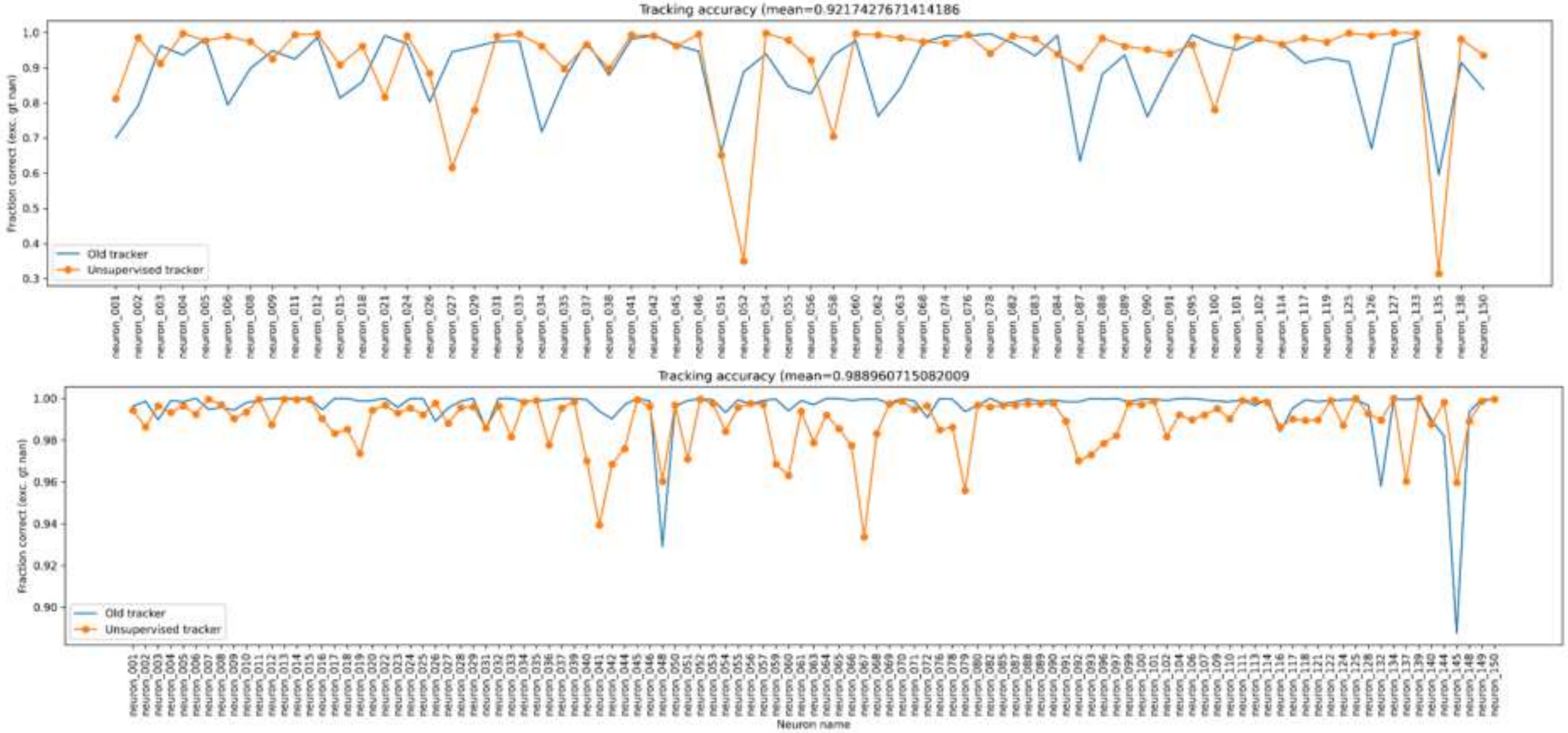
On our data: crops around neurons

T-sne embedding

Not all neurons, only the first 17 (just for display purposes)



Preliminary results as a tracker



Summary

1. SSL is a big deal
    1. And may be related to how humans learn
2. SSL is relevant in the low-label setting
    a. And there are robust methods that work at the laptop-scale
3. Lots of creative uses possible

Thank you!

Other resources

1. Explanations of SSL:
    1. https://lilianweng.github.io/posts/2019-11-10-self-supervised/#contrastive-predictive-coding
2. Explanation of Simsiam
    1. https://arxiv.org/pdf/2203.16262
3. History and collection of papers
    1. https://github.com/jason718/awesome-self-supervised-learning?tab=readme-ov-file

When data is limited, self-supervised is MUCH better

Table 2. **Semi-supervised learning on ImageNet** using 1% and 10% training examples. Results for the supervised method are from (Zhai et al., 2019). Best results are in **bold**.

Method	Top-1		Top-5	
	1%	10%	1%	10%
Supervised	25.4	56.4	48.4	80.4
PIRL	-	-	57.2	83.8
SimCLR	48.3	65.6	75.5	87.8
BYOL	53.2	68.8	78.4	89.0
SwAV	53.9	**70.2**	78.5	**89.9**
Barlow Twins (ours)	**55.0**	69.7	**79.2**	89.3