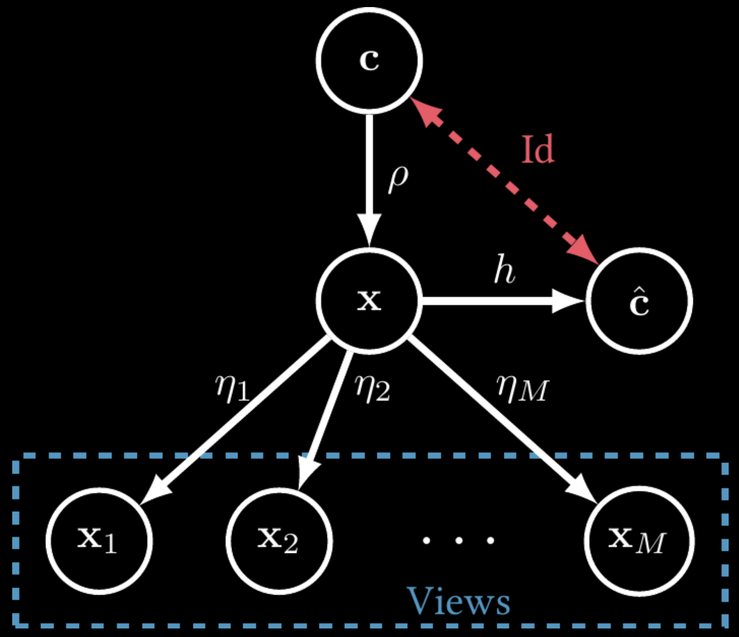
Poly-View Contrastive Learning

Amitis Shidani, Devon Hjelm, Jason Ramapuram, Russ Webb,
Eeshan Gunesh Dhekane, Dan Busbridge



Poly-view generative graph

Amitis Shidani

ICLR 2024 | Apple | Tuesday 7 May

In this talk

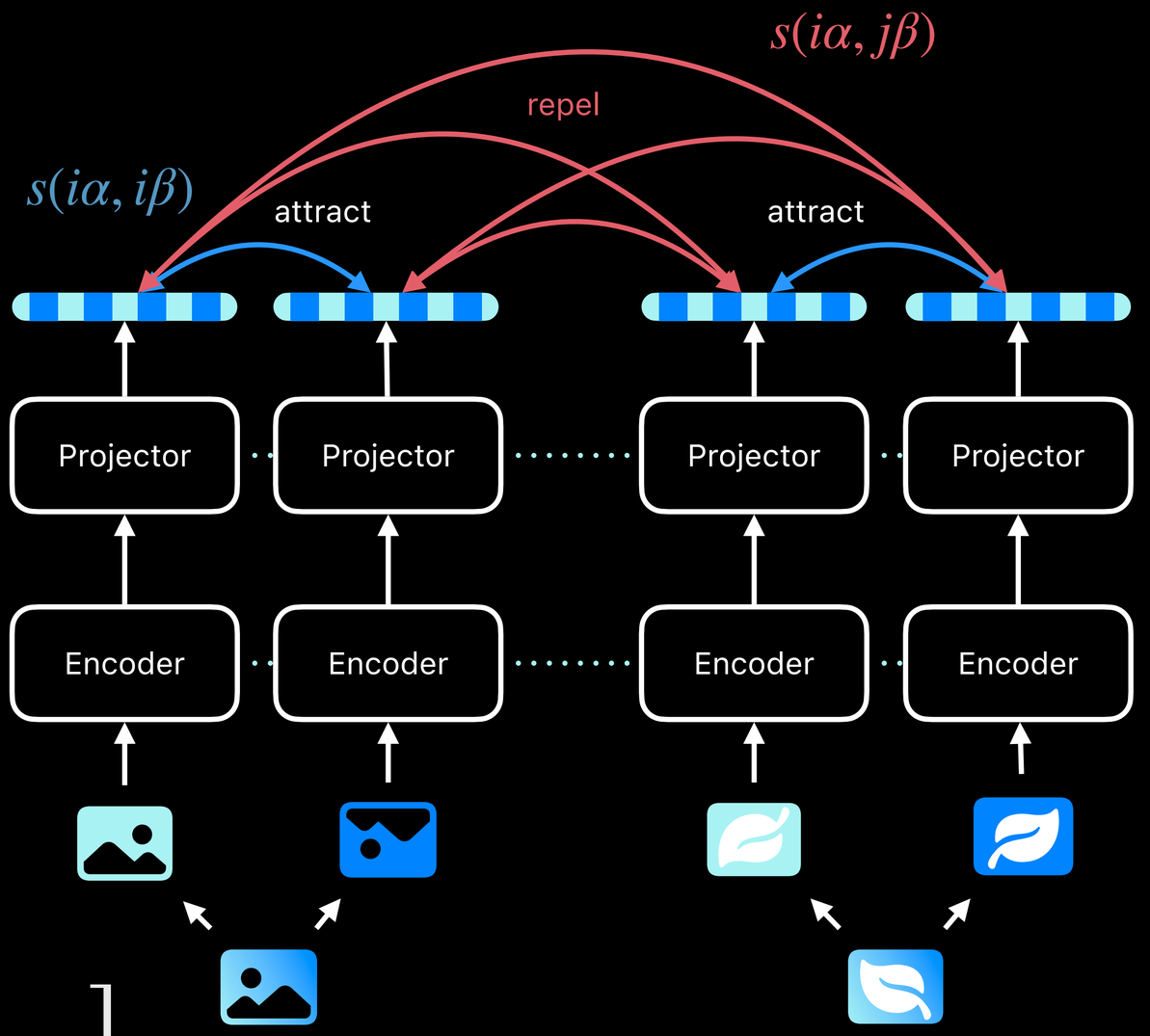
- Background: what is contrastive learning (CL)?
- Motivation: Why view multiplicity?
- Summary of theoretical contributions
- How we expand CL using view multiplicity
- Empirical Results and a new compute Pareto front

Contrastive Representation Learning (CL) in a nutshell

1. Generate multiple views of samples (for example, with augmentations)
   - Positives: multiple views of same sample
   - Negatives: views of other samples

2. Train an encoder to recognize *similarities* between positives while making negatives *dissimilar*

Loss is a lower bound to mutual information

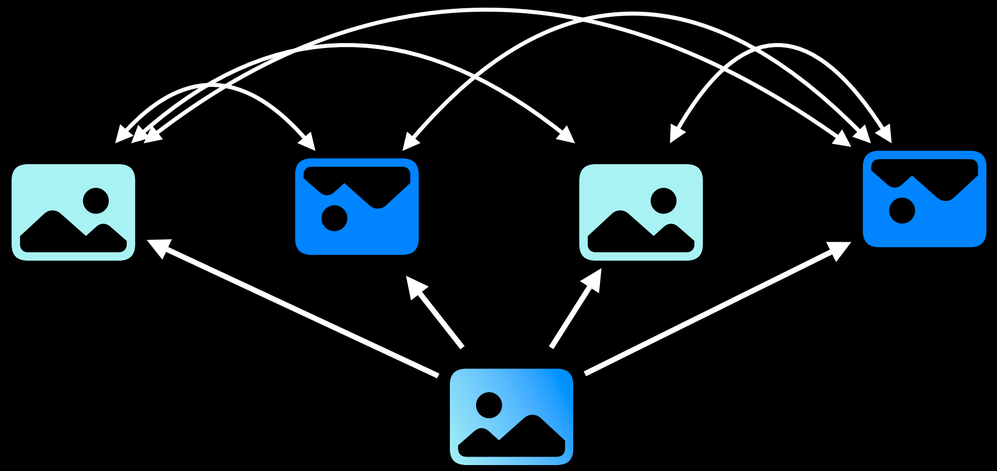


$$\mathbf{I}(x_\alpha; x_\beta) \geq \log K + \mathbb{E}\left[\log \frac{e^{s(i\alpha, i\beta)}}{e^{s(i\alpha, i\beta)} + \sum_{j \neq i} e^{s(i\alpha, j\beta)}}\right]$$

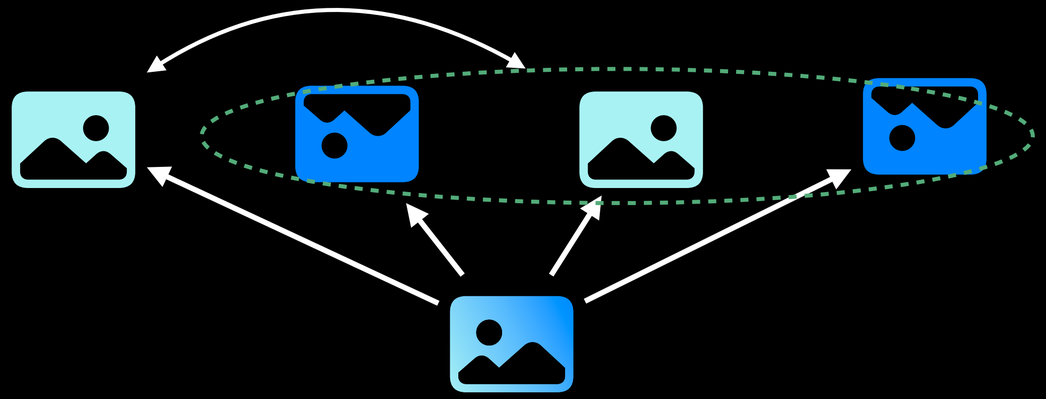
Why *only* two views?

View Multiplicity: how many positives?

- Helps the signal to noise ratio[1]
- In contrastive learning: **Always pairwise "tasks"** (i.e. comparisons)
- In multi-crop[2, 3]: **more pairwise tasks** from the same image (lowers the variance)
- Poly-view (ours): few tasks with lots of positives
- Why? More information in each task: better representations



Multi-Crop CL



Poly-View CL

[1] Stanislav Fort, Andrew Brock, Razvan Pascanu, Soham De, and Samuel L. Smith. Drawing multiple augmentation samples per image during training efficiently decreases test error, CoRR 2021
[2] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding, ECCV 2020
[3] Mathilde Caron, Ishan Misra, Julien Mairal, Prisa Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrastive cluster assignment, NeurIPS 2020

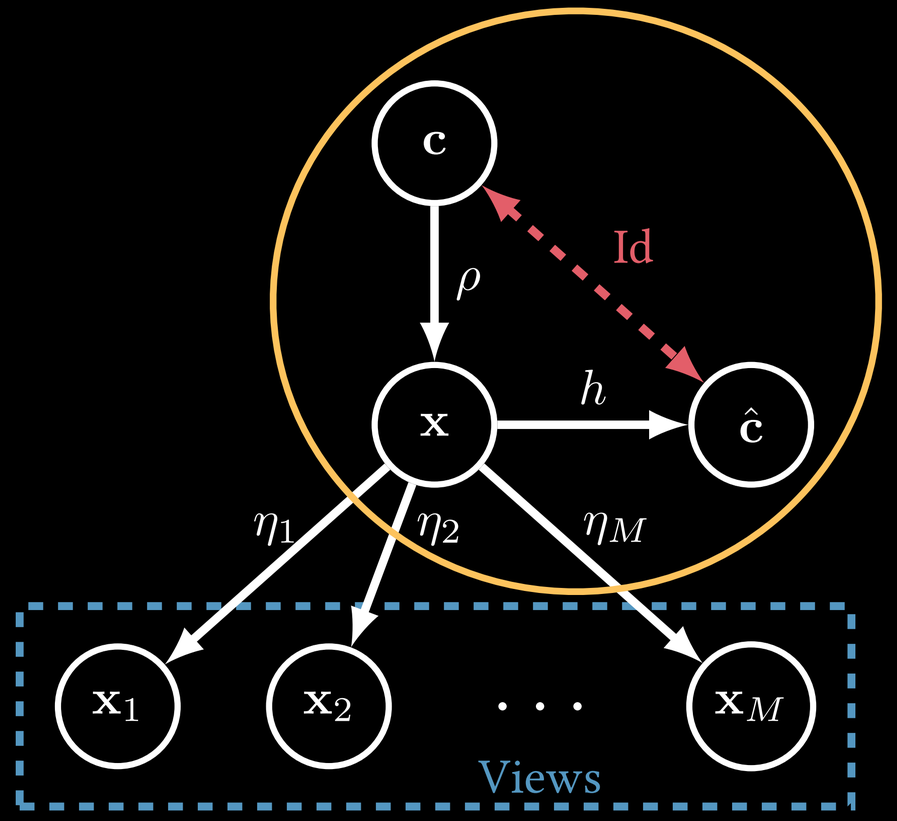
We develop two ways of increasing the number of positives

1. Does more views share more information about the generative factor?

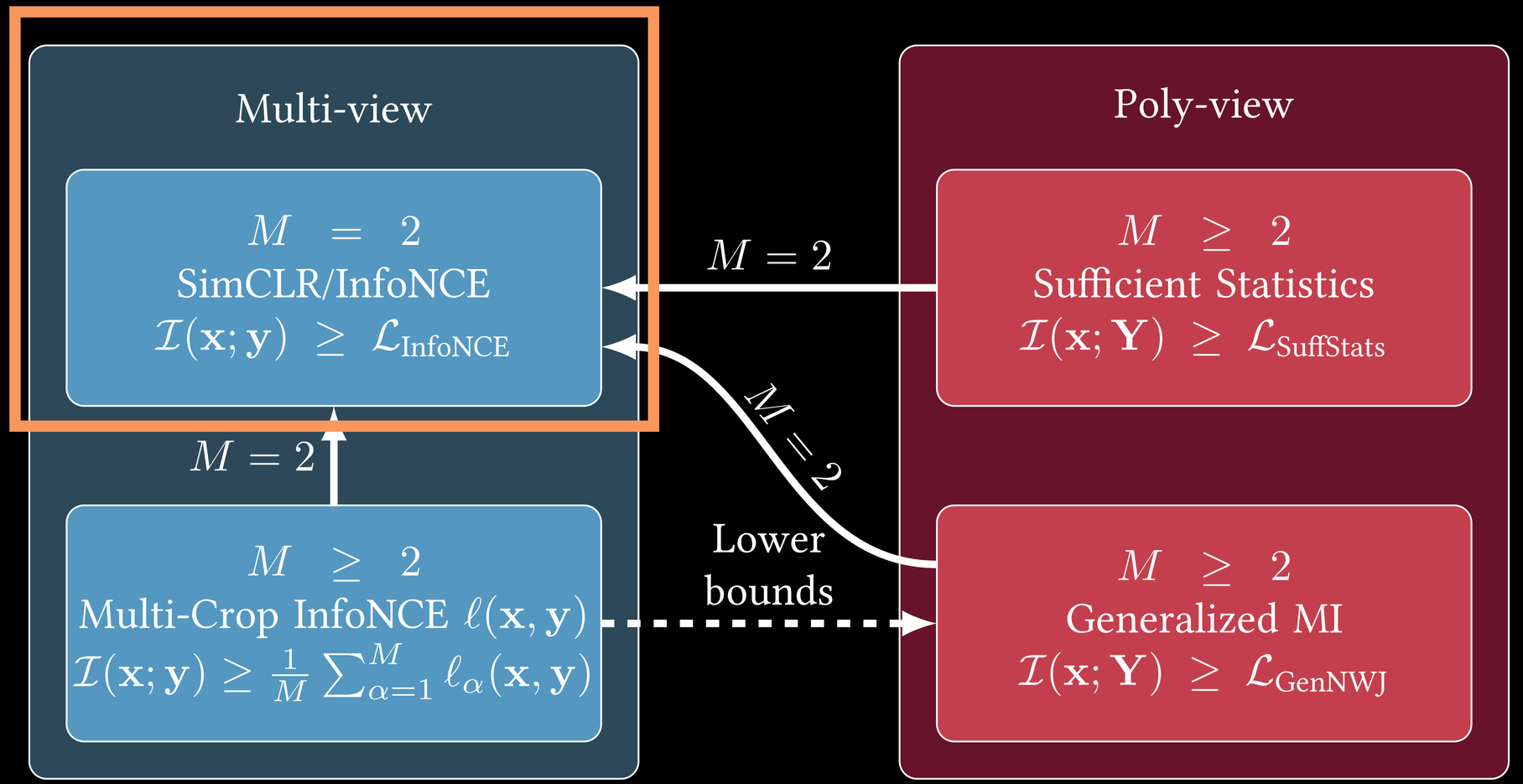
   Generalizing the InfoMax proxy

2. Does more views help to estimate the conditional distribution of x given c better?

   Connecting Sufficient Statistics to InfoMax

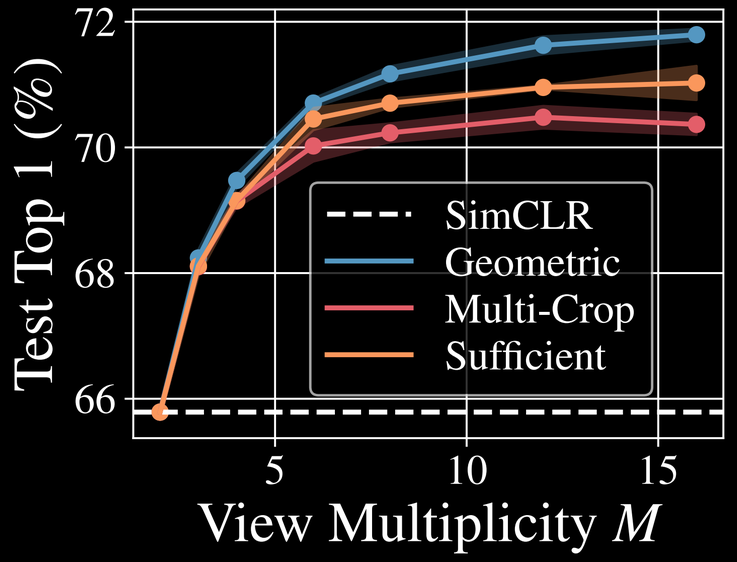


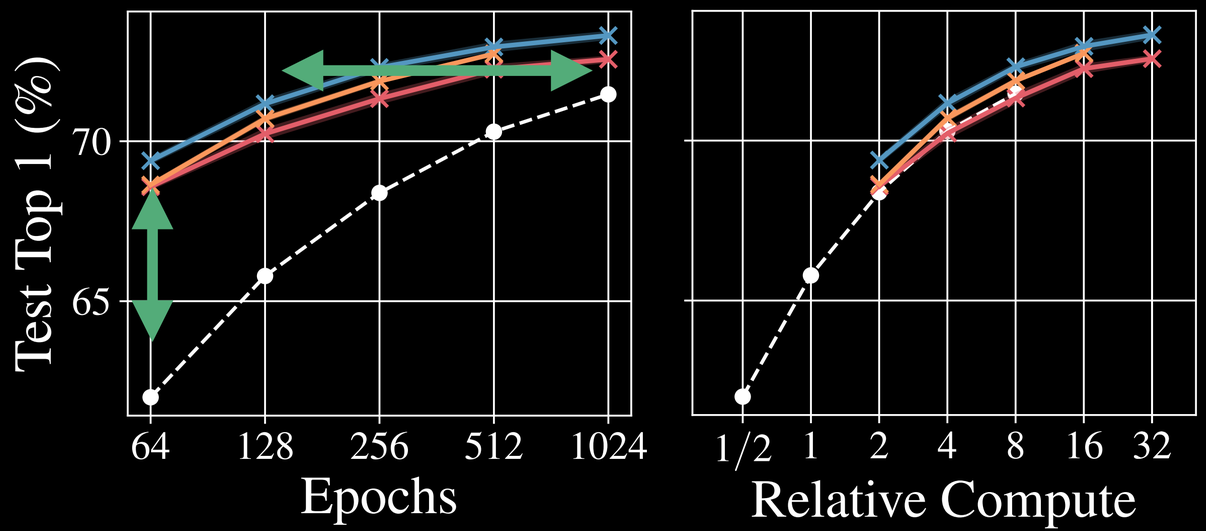
Multi-view and poly-view estimators are all related



View Multiplicity improves ImageNet1k Performance

- Higher multiplicity means better results, but especially for poly-view
- For Poly-view: reduce compute by trading batch size for multiplicity





Conclusions

- We studied the role of view multiplicity, giving rise to a new families of contrastive learning algorithms
- View multiplicity theoretically and empirically improves representations
- View multiplicity reveals that SimCLR (two views) is not compute optimal and increased multiplicity gives a new compute Pareto front for contrastive learning

https//iclr.cc/virtual/2024/poster/18069

Tuesday May 7
Afternoon Poster Session

TM and © 2024 Apple Inc. All rights reserved.