Self-Supervised Contrastive Learning: SimCLR

What is Contrastive Learning?

The quest for a **representation** $f : X \to Z$ where:

Similar data (positive pairs) $(x, y) \sim p^+$ are close

Dissimilar data (negative pairs) $(x, y) \sim p^-$ are distant

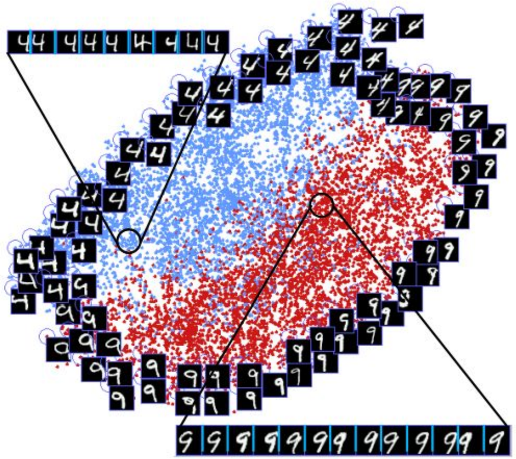
Loss archetype ('siamese loss'):

$$\mathbb{E}_{p^+}\left[d(f(x), f(y))\right] - \mathbb{E}_{p^-}\left[d(f(x), f(y))\right]$$

Two Flavours of Contrastive Learning

Supervised: positive pairs have same class. Dates way back.

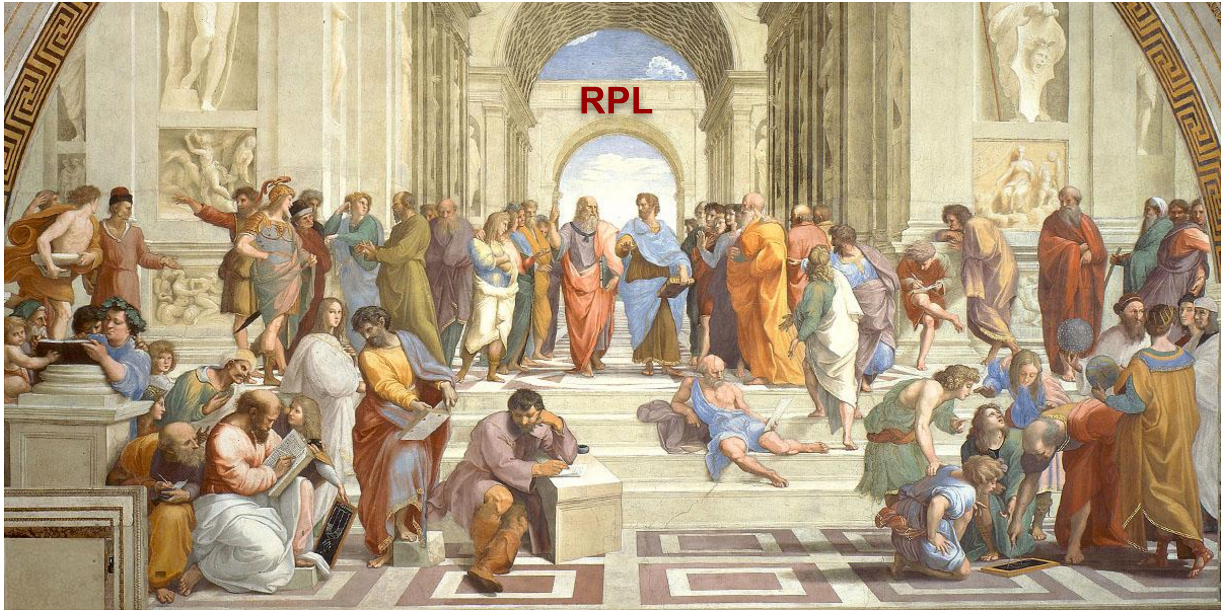
Un/Self -Supervised: positive pairs are related by data augmentation.



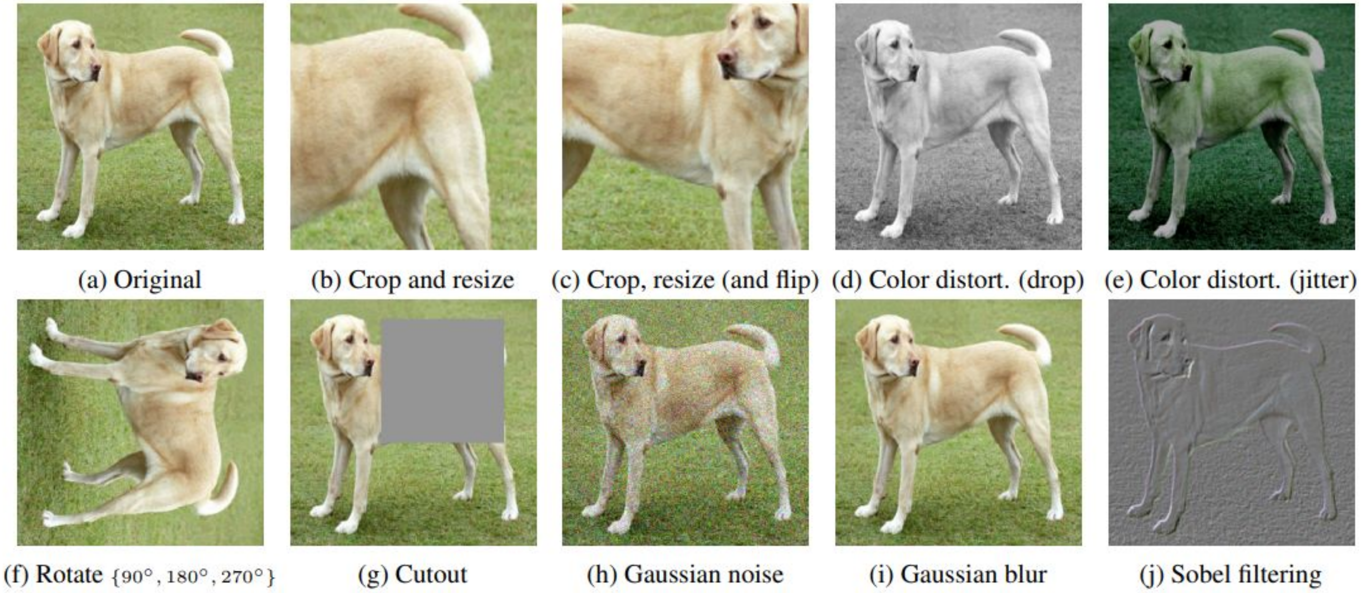
Hadsell, Chopra, **LeCun**,
Dimensionality Reduction by Learning an Invariant Mapping, 2006.

Discussion Point

Comparisons with other un/self -supervised representation learning frameworks, ex. Autoencoders.



Details on SimCLR: Data Augmentation



...And their compositions!

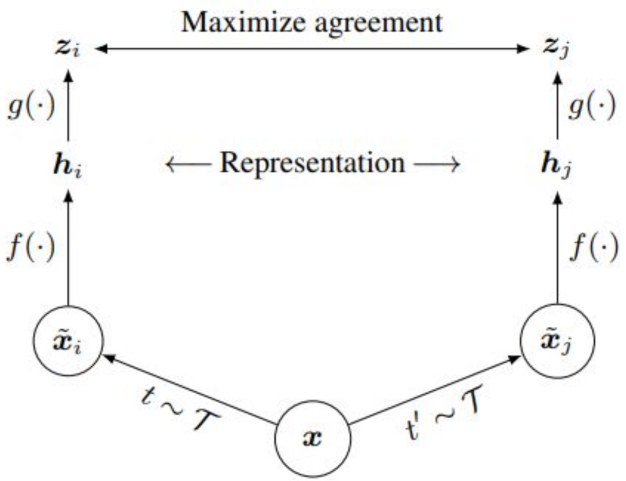
Details on SimCLR: Loss

Softmax'd cosine distance. Everything in the **big batches** is dissimilar. For a similar pair within a batch, the loss is:

$$-\log \frac{e^{\cos(z_i, z_j)/\tau}}{\sum_{k \neq i} e^{\cos(z_i, z_k)/\tau}}$$

Details on SimCLR: Projection Head

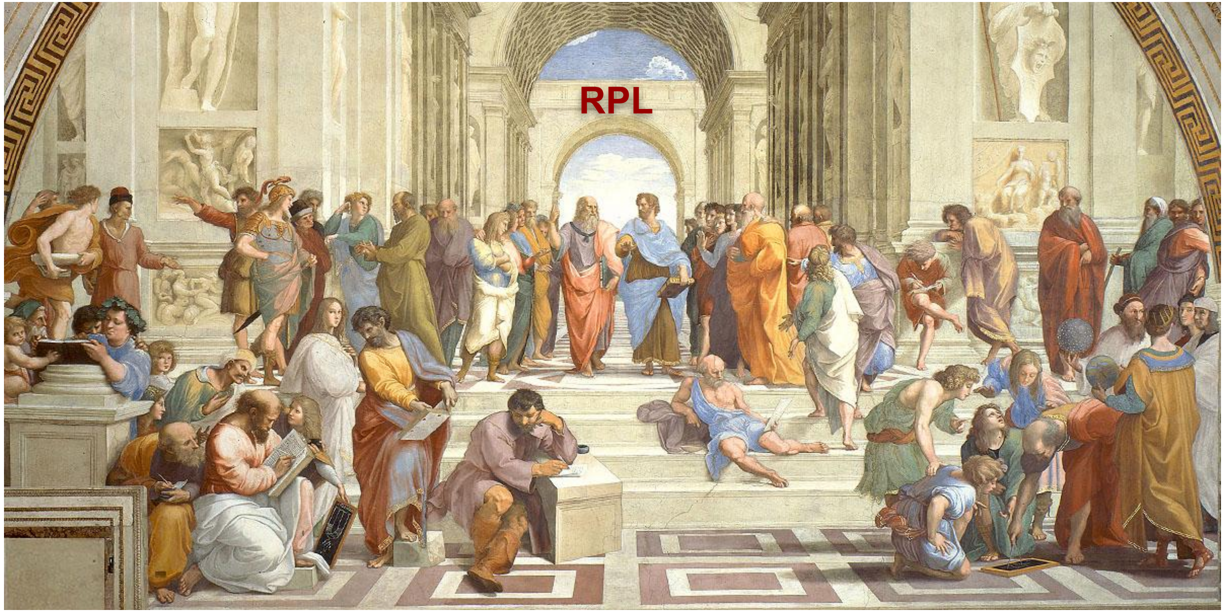
The representation one should use is actual an intermediate one. In other words, a **projection head *g*** is attached on top of the model.



Discussion Points

Why composing transformations?

Why attaching a (complex) projection head?



Weak Points and How to Solve Them

- Memory cost of batch gradients.

  A solution: MoCO. He et al., *Momentum Contrast for Unsupervised Visual Representation Learning*

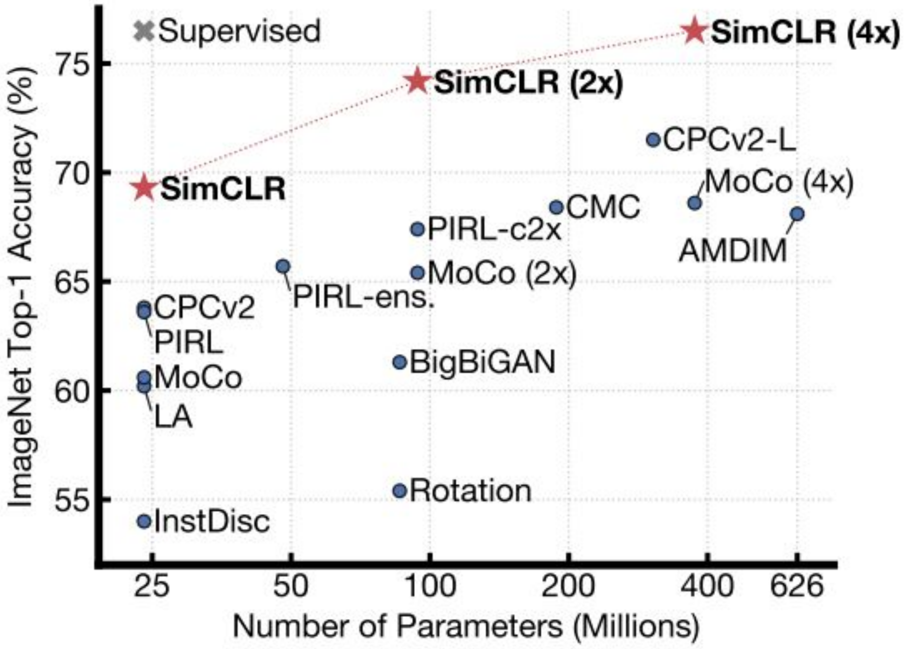
  It is incorporated in SimCLRv2: Chen et al., *Big Self-Supervised Models are Strong Semi-Supervised Learners*

- Too much invariance, even with projection head.

  A solution: LooC. Xiao et at., *What Should not be Contrastive in Contrastive Learning*

Results: Comparisons

Representations are evaluated by training a linear classifier on top.



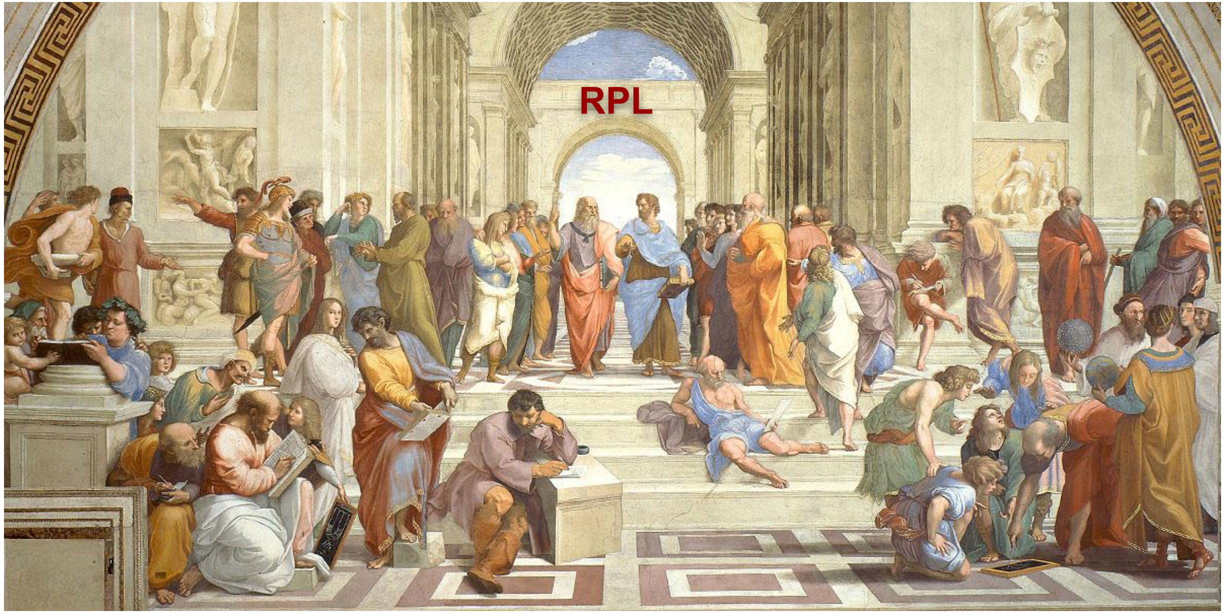
Dataset: ImageNet

Results: Transfer Learning

Models are pretrained on ImageNet.

	Food	CIFAR10	CIFAR100	Birdsnap	SUN397	Cars	Aircraft	VOC2007	DTD	Pets	Caltech-101	Flowers
Linear evaluation:												
SimCLR (ours)	**76.9**	**95.3**	80.2	48.4	**65.9**	60.0	61.2	**84.2**	**78.9**	89.2	**93.9**	**95.0**
Supervised	75.2	**95.7**	**81.2**	**56.4**	64.9	**68.8**	**63.8**	83.8	**78.7**	**92.3**	**94.1**	94.2
Fine-tuned:												
SimCLR (ours)	**89.4**	**98.6**	**89.0**	**78.2**	**68.1**	**92.1**	**87.0**	**86.6**	**77.8**	92.1	**94.1**	97.6
Supervised	88.7	98.3	**88.7**	**77.8**	67.0	91.4	**88.0**	86.5	**78.8**	**93.2**	**94.2**	**98.0**
Random init	88.3	96.0	81.9	**77.0**	53.7	91.3	84.8	69.4	64.1	82.7	72.5	92.5

Free Discussion



Opinions About the Paper

- Making contrastive learning self-supervised: elegantly simple and useful idea.

- Great results.

- The paper contains **a lot** of ablations. Maybe too many?

Tack!