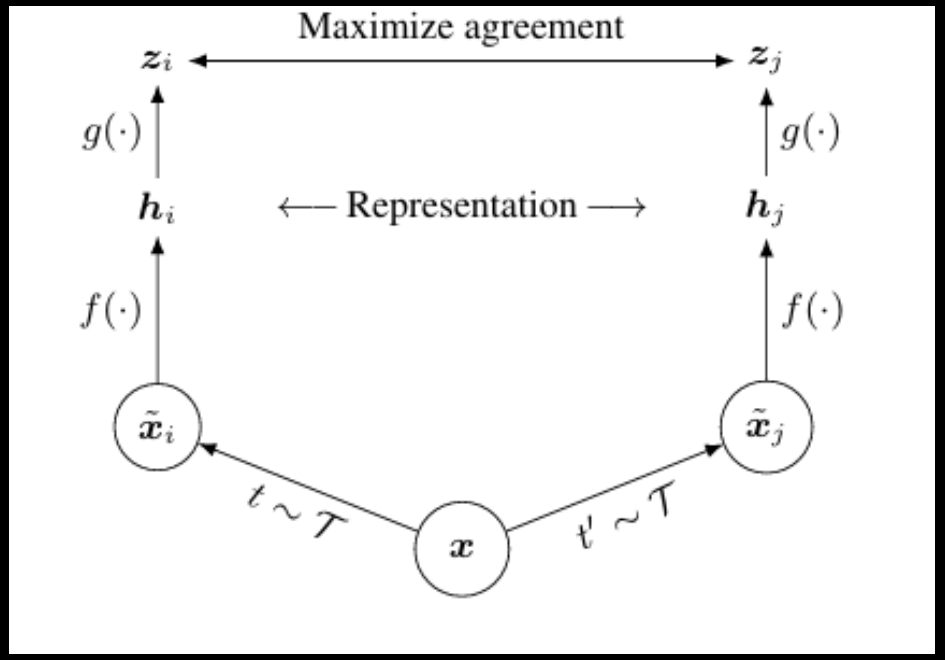
# AIL 862

Lecture 13

# SimCLR



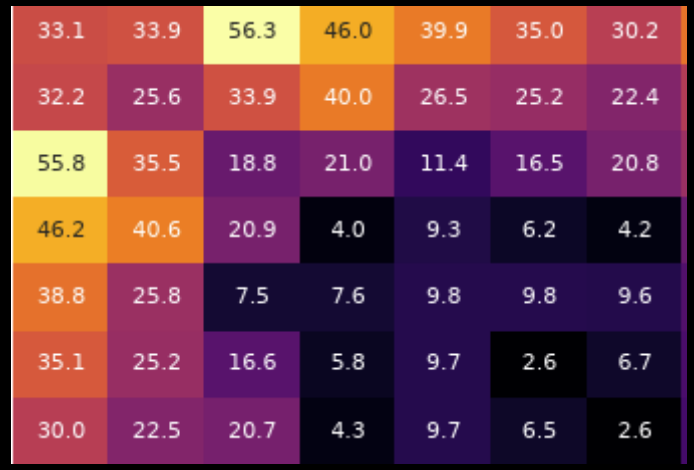

A Simple Framework for Contrastive Learning of Visual Representations, 2020

# SimCLR

Learning algorithm

# Composition of data augmentation

Helps



# Composition of data augmentation

Helps

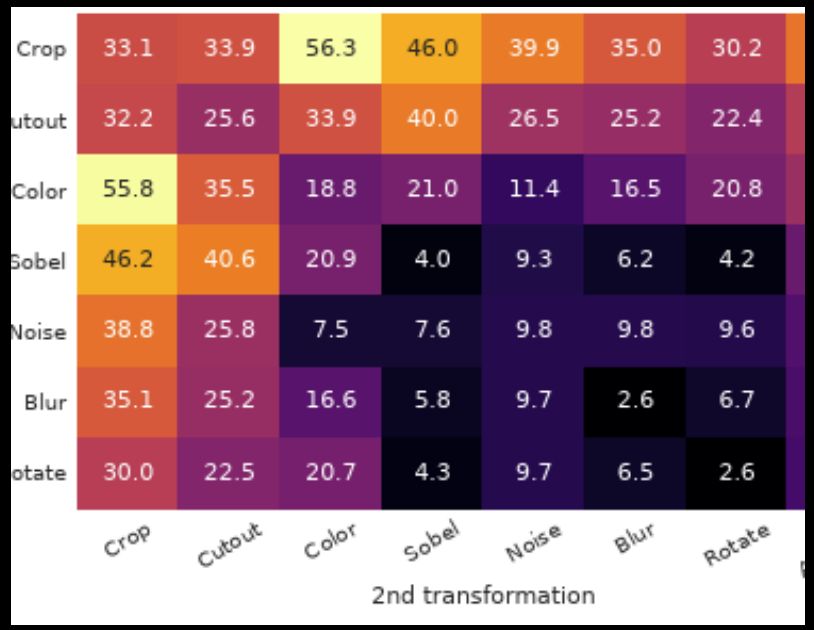
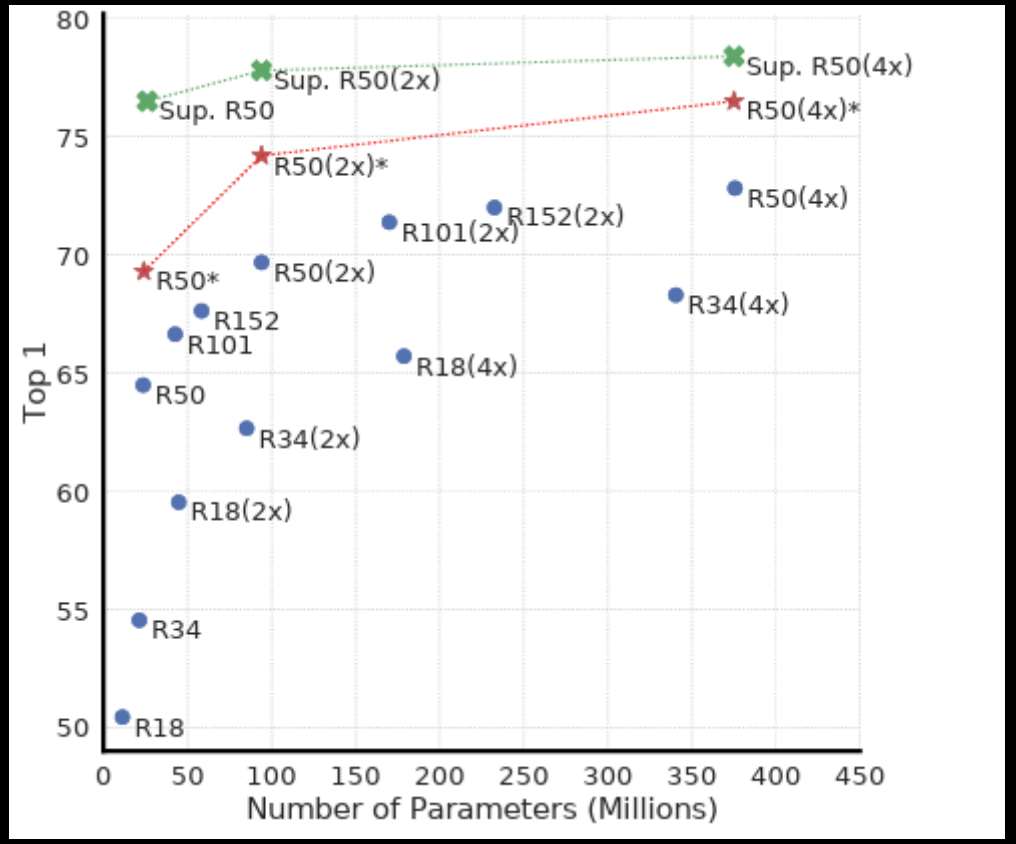



*Figure 5.* Linear evaluation (ImageNet top-1 accuracy) under individual or composition of data augmentations, applied only to one branch. For all columns but the last, diagonal entries correspond to single transformation, and off-diagonals correspond to composition of two transformations (applied sequentially). The

# Stronger data augmentation

- Is needed in comparison to the supervised learning.

# Benefits from bigger models



# Batch size

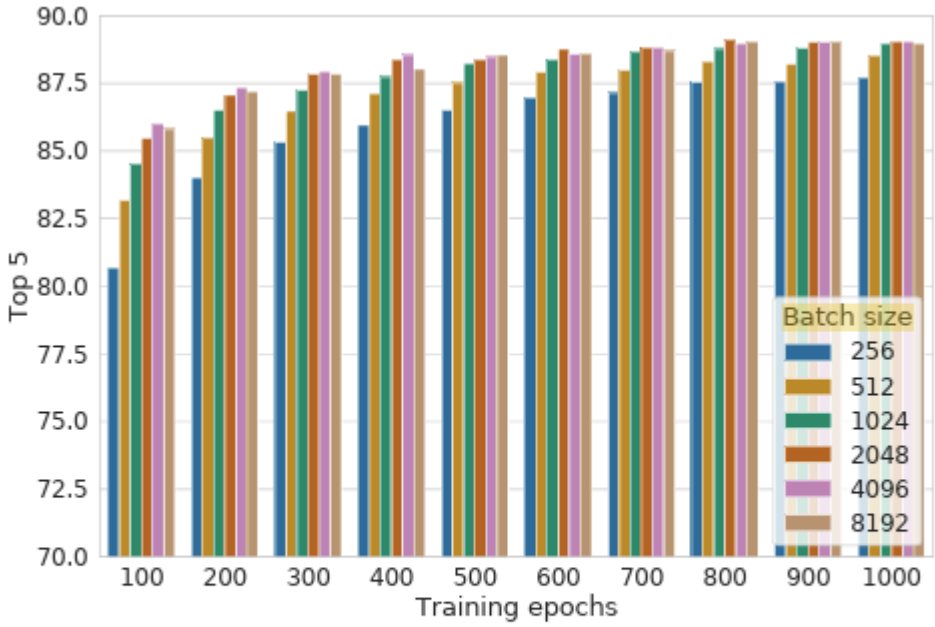



*Figure B.1.* Linear evaluation (top-5) of ResNet-50 trained with different batch sizes and epochs. Each bar is a single run from scratch. See Figure 9 for top-1 accuracy.

Default batch size - 4096

# Not always beneficial

- SSL-based pretraining need not be always beneficial (when we are going to finetune it later on some labeled dataset)
- SimCLR authors trained the network from random initialization using the same procedure as for fine-tuning, but for longer
- On Food-101, Stanford Cars, and FGVC Aircraft datasets, fine-tuning provides only a small advantage over training from random initialization. However, on the remaining 8 datasets, pretraining has clear advantages

# Pre-text invariant representation learning

Main working principle similar to the SimCLR, however uses memory bank

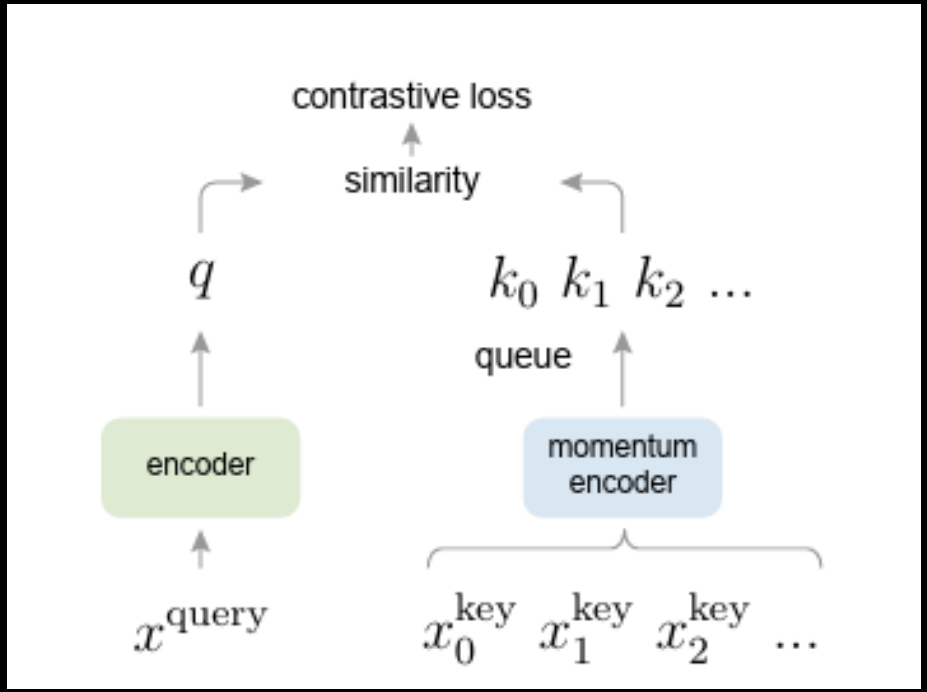
# PIRL memory bank

- The memory bank contains feature representation of each original image (without transformation) in the dataset.

# PIRL memory bank

- The memory bank contains feature representation of each original image (without transformation) in the dataset.
- Memory bank allows us to replace negative terms in the loss function with their memory bank representation, without increasing training batch size.

# MoCo




Momentum Contrast for Unsupervised Visual Representation Learning