A Simple Framework for Contrastive Learning of Visual Representations

SimCLR review

YU MO, 2022/06/27

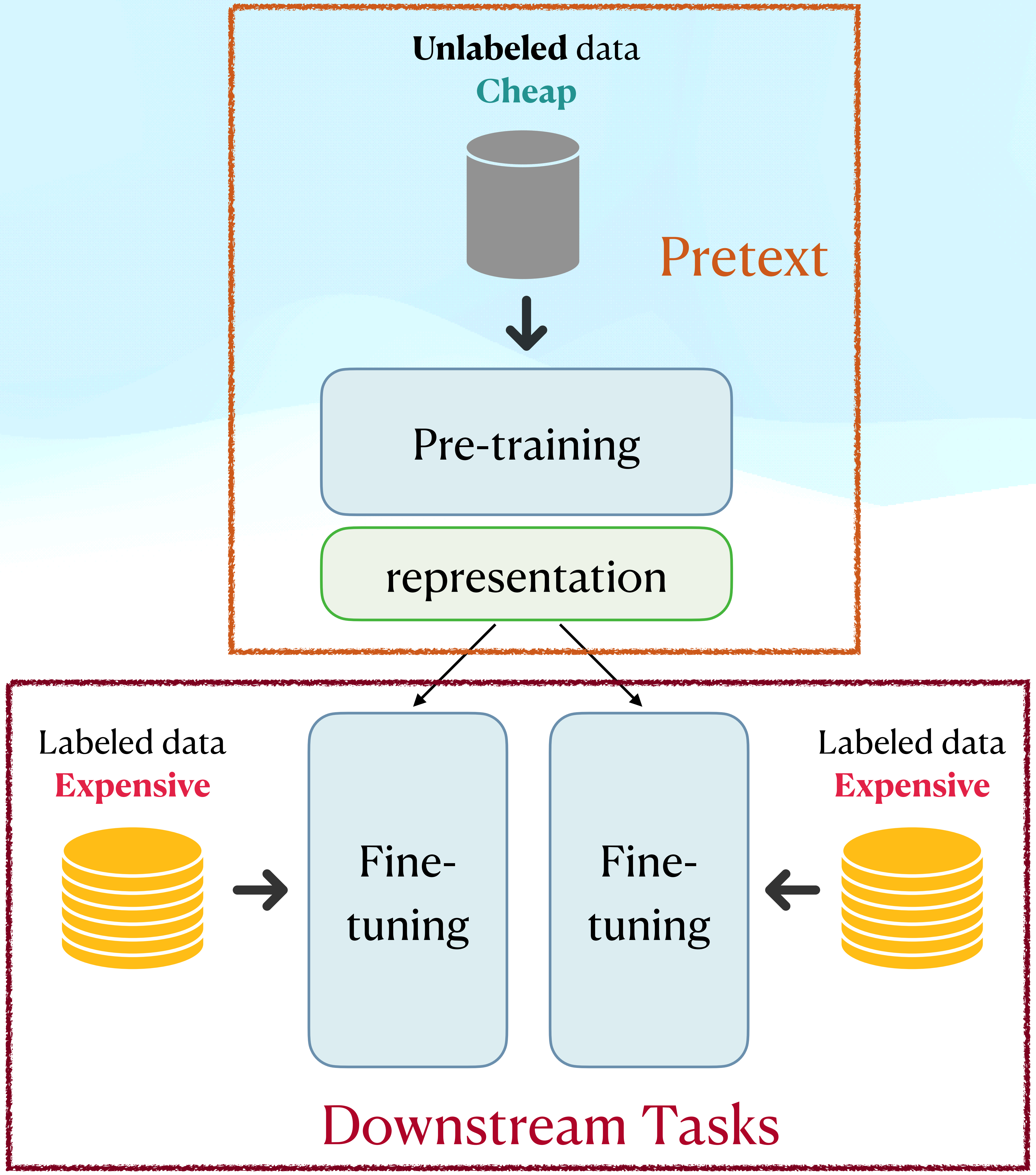
Background

Self-supervised Learning

- **“Unsupervised pre-training, Supervised fine-tuning”**

- **Pretext** is task-agnostic, we train the model (Encoder) with a **large** amount of unlabeled data

- **Downstream task** is task-specific, we train the model with a **few** amount of labeled data (1% to 10%)

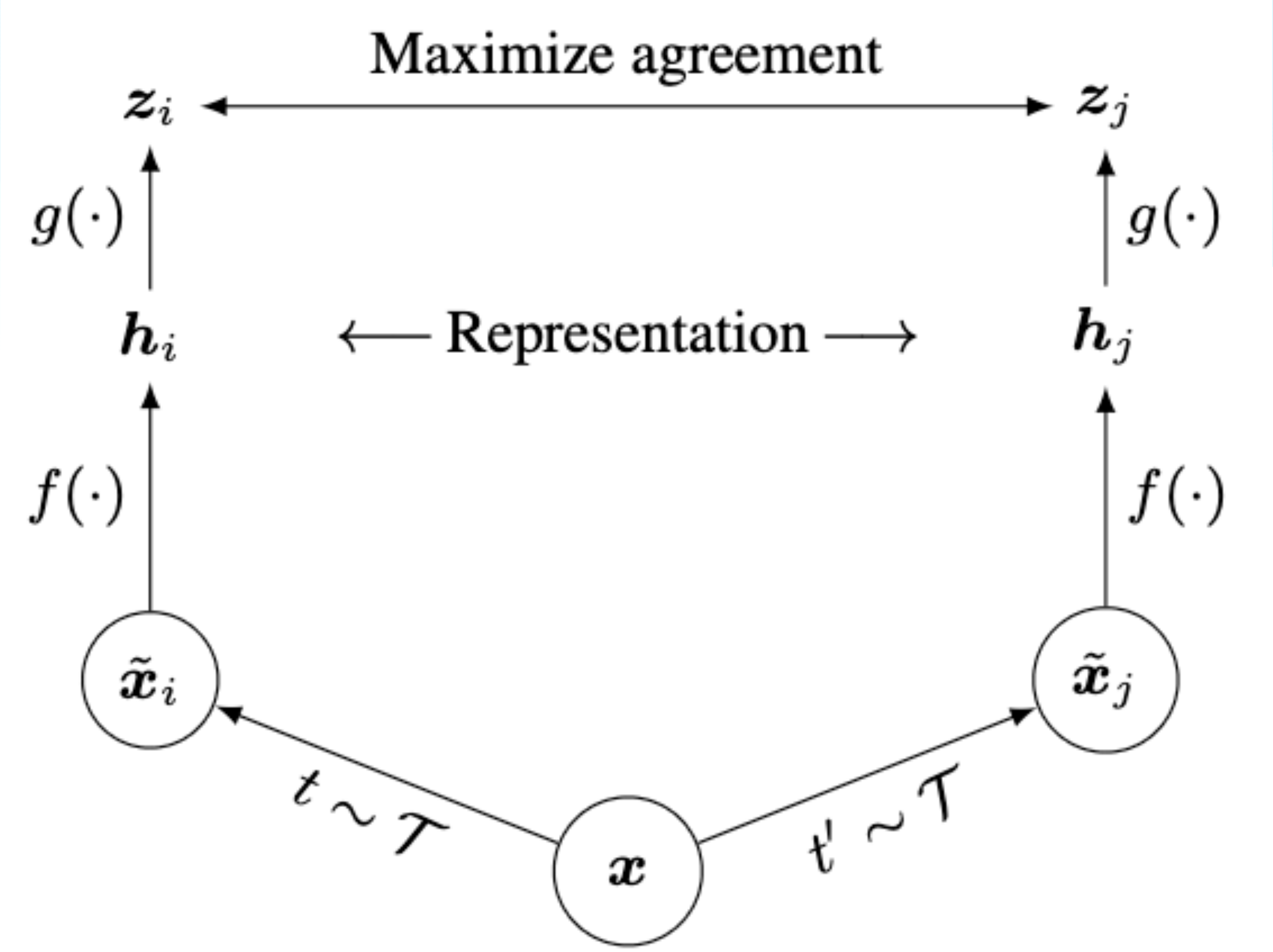


Framework

Paper abstract

Findings/Contributions

- Composition of data augmentations is important
- Functionality of projection head (nonlinear transformation)
- Larger batch size and longer training steps benefits SSL more than supervised model.

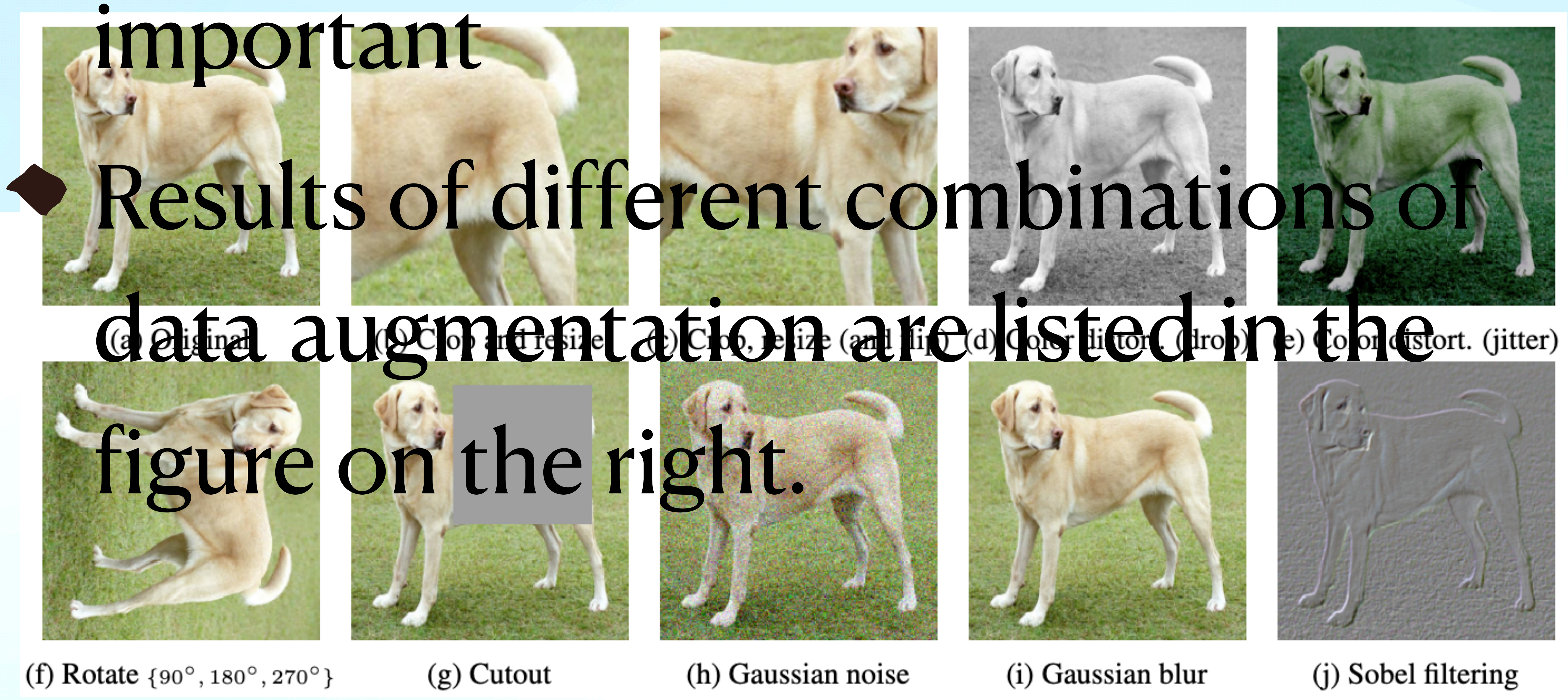


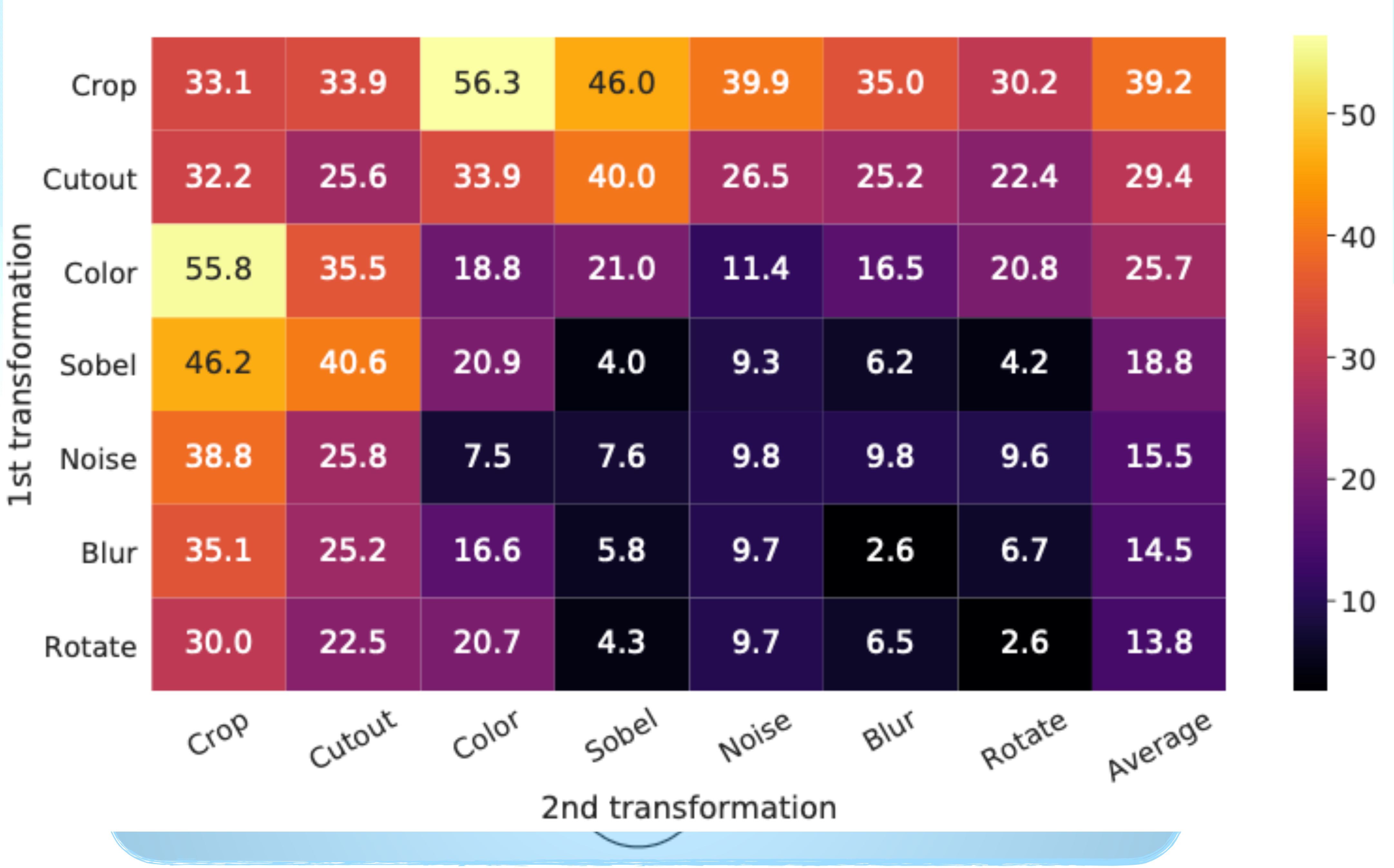
simCLR framework

Data Augmentations

Finding 1

- Composition of data augmentations is important
- Results of different combinations of data augmentation are listed in the figure on the right.



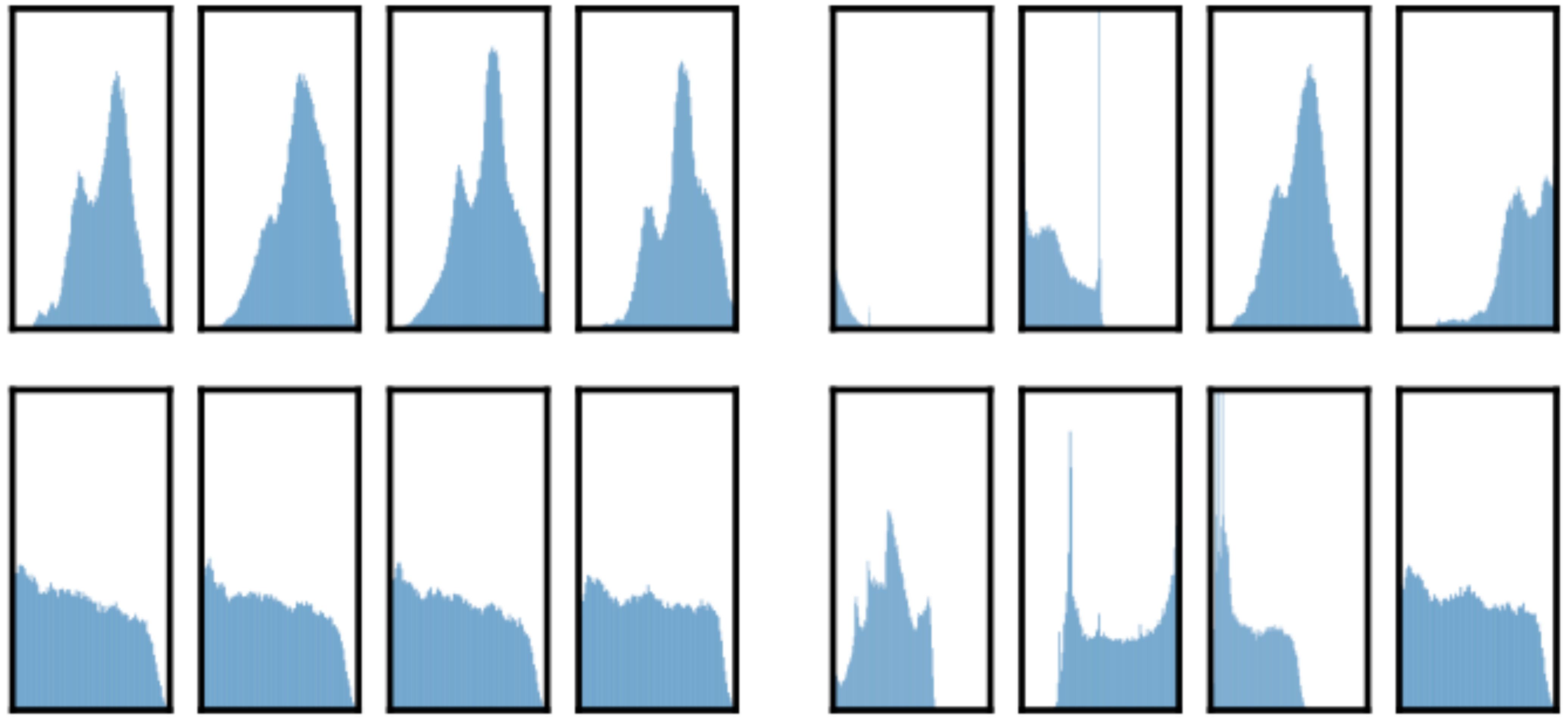


simCLR framework

Data Augmentations

Finding 1

- Color distortion suffice to distinguish the image with similar color distributions after random cropping.
- SSL need stronger data augmentations



(a) Without color distortion. (b) With color distortion.

Color distortion

Methods	1/8	1/4	1/2	1	1 (+Blur)	AutoAug
	Color distortion strength					
SimCLR	59.6	61.0	62.6	63.2	64.5	61.1
Supervised	77.0	76.7	76.5	75.7	75.4	77.1

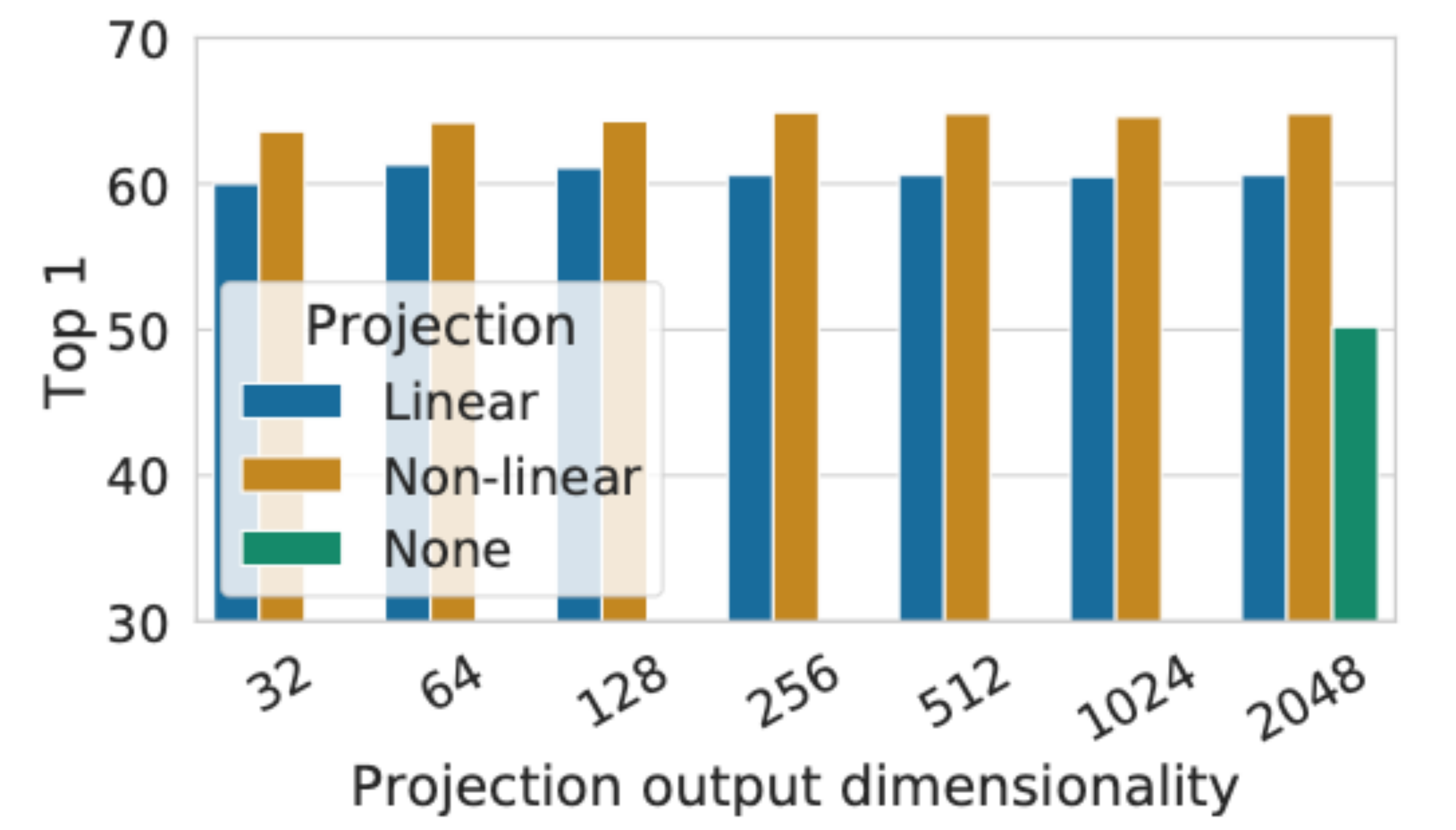
Different level of color distortion

Projection Head

Finding 2

A nonlinear projection head improves the representation quality of the layer before it.

- Default setting is MLP and nonlinear ReLU.
- More information is contained before projection than after, e.g, the augmentation operation.



What to predict?	Random guess	Representation $\boldsymbol{h}$	$g(\boldsymbol{h})$
Color vs grayscale	80	99.3	97.4
Rotation	25	67.6	25.6
Orig. vs corrupted	50	99.5	59.6
Orig. vs Sobel filtered	50	96.6	56.3

Intensive Training

Finding 3

- Contrastive learning benefits more from larger batch size and longer training.