7.5. A Simple Framework for Contrastive Learning of Visual Representations

요약

- Data augmentation이 중요한 역할을 한다.
- Representation과 Contrastive loss사이에 Nonlinear Transformation를 도입한다.
- Supervised보다 훨씬 큰 batch size와 더욱 많은 epochs동안 학습시킨다.

- Contrastive prediction task를 정의할때 Data augmentation의 구성이 중요하고, Unsupervised contrastive learning은 Supervised보다 많은 Data augmentation을 이용해야 한다.
- Representation과 Contrastive Loss 사이에 Nonlinear transformation을 도입하면 성능이 좋아진다.
- Normalized embedding과 Temperature parameter를 사용한 Contrastive cross entropy loss를 이용한다.
- Supervised와 비교해서 더 큰 batch size와 더 긴 training epoch을 사용해야 한다.
- Supervised처럼 네트워크가 깊고 넓을수록 성능이 좋다.

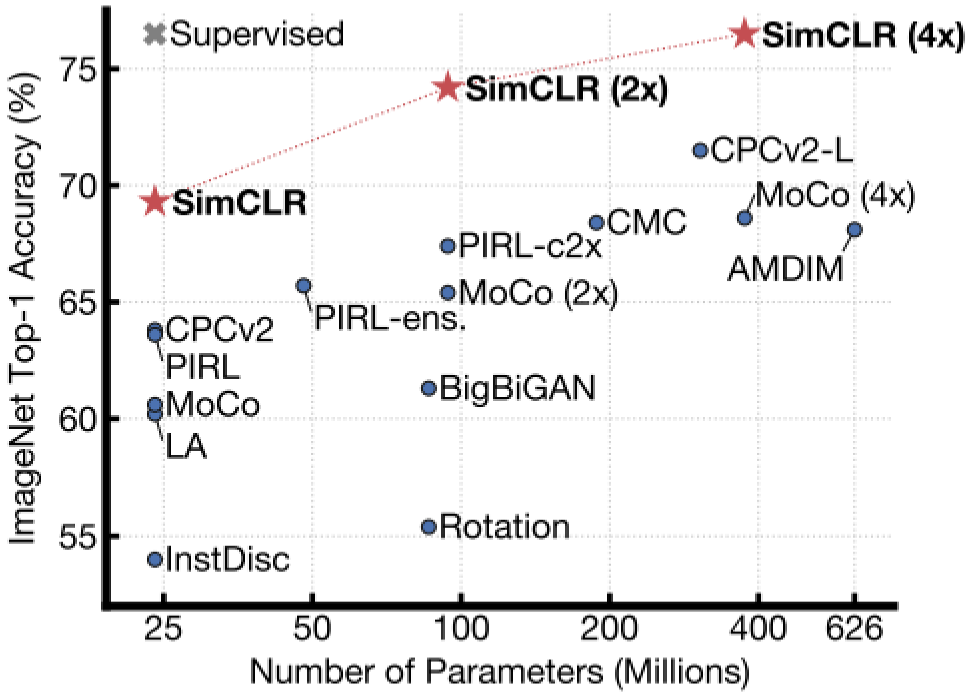


Figure 1. ImageNet Top-1 accuracy of linear classifiers trained on representations learned with different self-supervised methods (pretrained on ImageNet). Gray cross indicates supervised ResNet-50. Our method, SimCLR, is shown in bold.

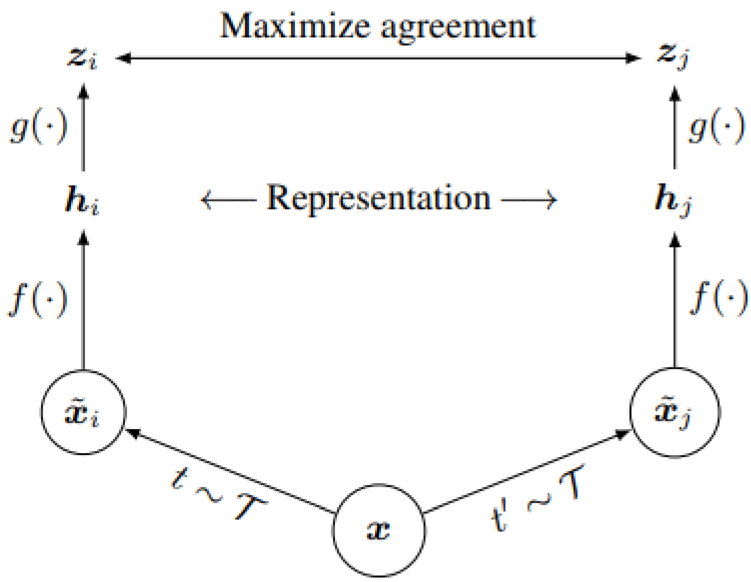
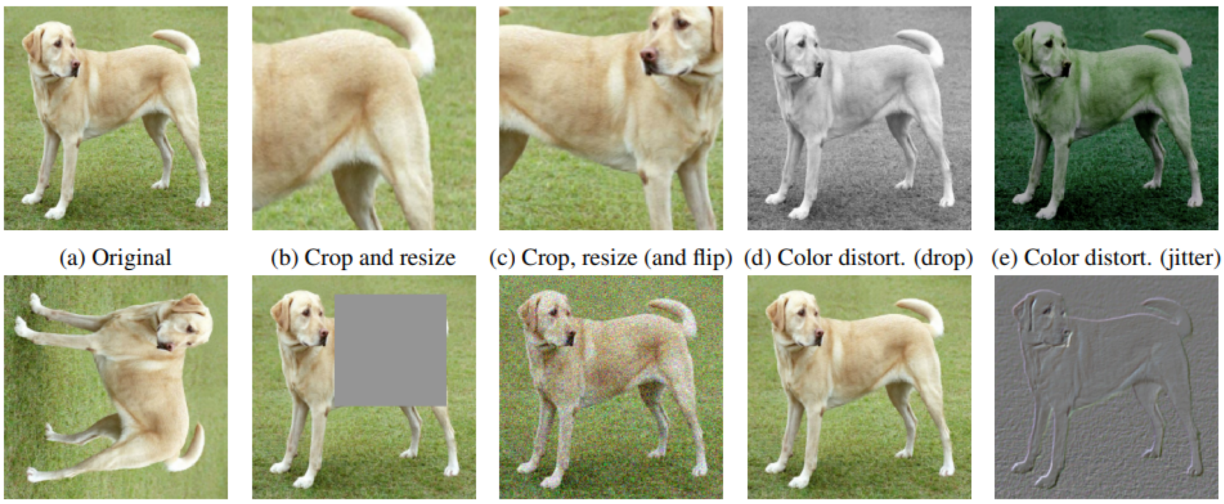


Figure 2. A simple framework for contrastive learning of visual representations. Two separate data augmentation operators are sampled from the same family of augmentations ($t \sim \mathcal{T}$ and $t' \sim \mathcal{T}$) and applied to each data example to obtain two correlated views. A base encoder network $f(\cdot)$ and a projection head $g(\cdot)$ are trained to maximize agreement using a contrastive loss. After training is completed, we throw away the projection head $g(\cdot)$ and use encoder $f(\cdot)$ and representation $\boldsymbol{h}$ for downstream tasks.



(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

Figure 4. Illustrations of the studied data augmentation operators. Each augmentation can transform data stochastically with some internal parameters (e.g. rotation degree, noise level). Note that we *only* test these operators in ablation, the *augmentation policy used to train our models* only includes *random crop (with flip and resize)*, *color distortion*, and *Gaussian blur*. (Original image cc-by: Von.grzanka)

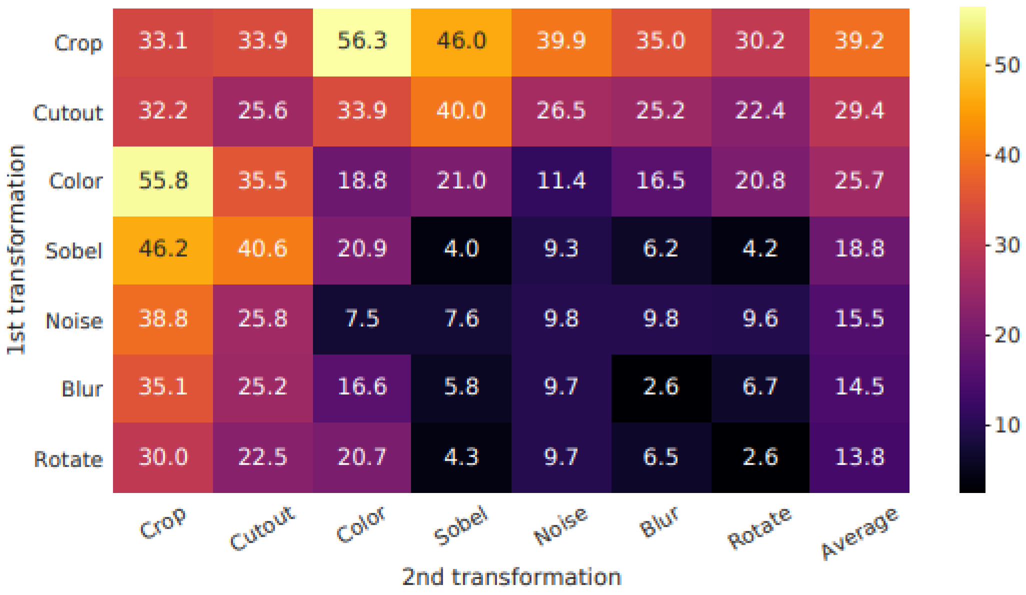


Figure 5. Linear evaluation (ImageNet top-1 accuracy) under individual or composition of data augmentations, applied only to one branch. For all columns but the last, diagonal entries correspond to single transformation, and off-diagonals correspond to composition of two transformations (applied sequentially).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