南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Supervised Contrastive Learning

Prannay Khosla *
Google Research

Piotr Teterwak *
Boston University

Chen Wang †
Snap Inc.

Aaron Sarna ‡
Google Research

Yonglong Tian
MIT

Phillip Isola
MIT

Aaron Maschinot
Google Research

Ce Liu
Google Research

Dilip Krishnan
Google Research

NIPS 2020

Contrastive Learning

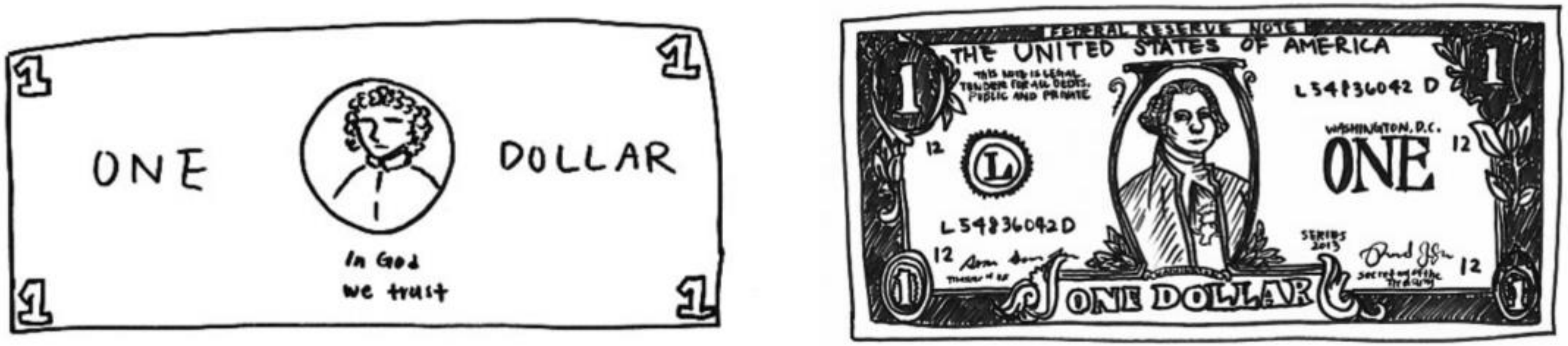


Fig. Left: Drawing of a dollar bill from memory.
Right: Drawing subsequently made with a dollar bill present

- 通过数据之间的对比进行表示学习
- 让像的样本所得表示差异小，让不像的样本所得表示差异大

Contrastive Learning---SimCLR

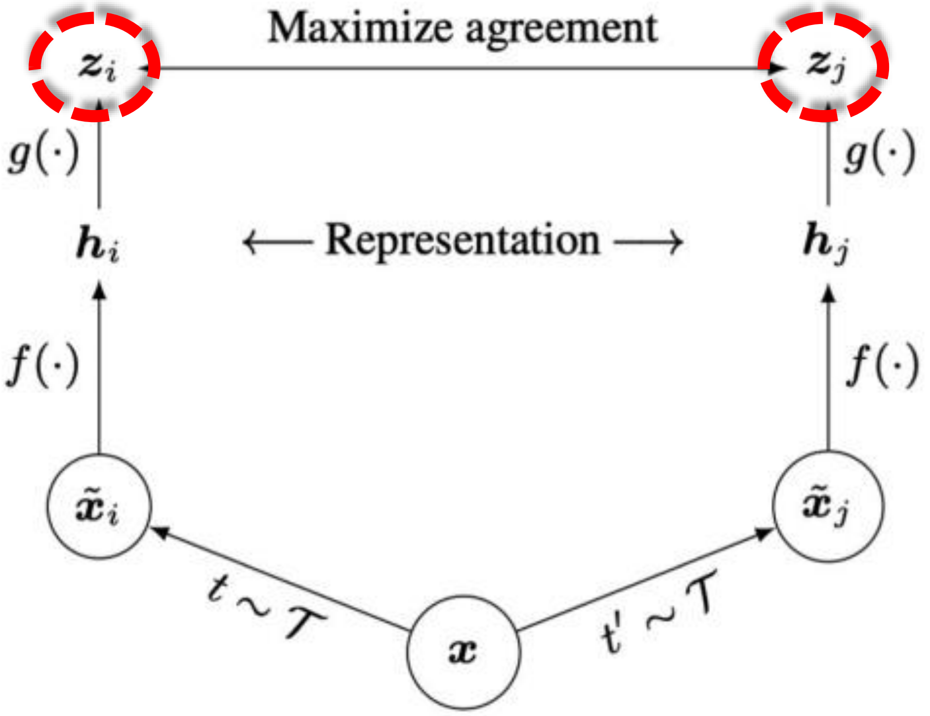
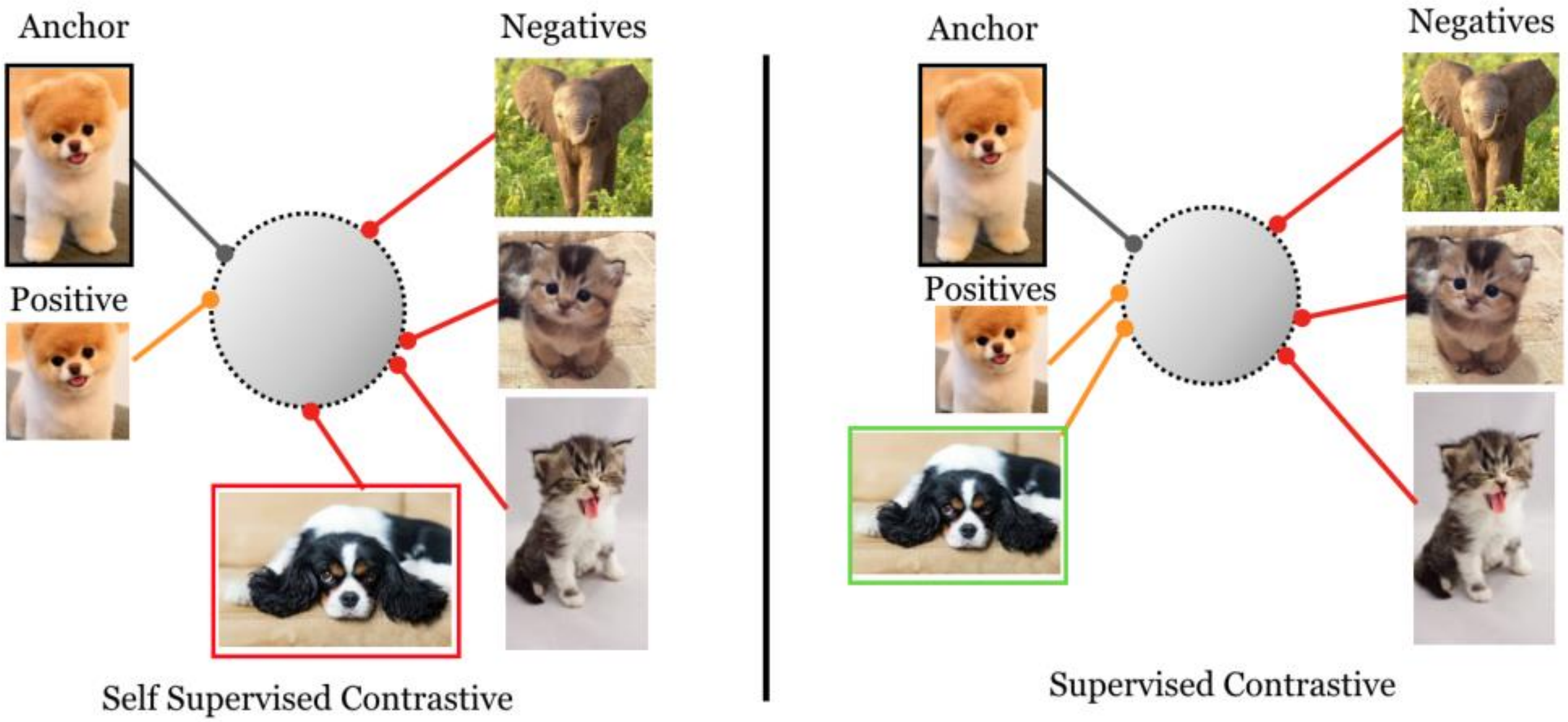


Figure 2. A simple framework for contrastive learning of visual representations. Two separate data augmentation operators are sampled from the same family of augmentations ($t \sim \mathcal{T}$ and $t' \sim \mathcal{T}$) and applied to each data example to obtain two correlated views. A base encoder 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.

Supervised Contrastive Learning



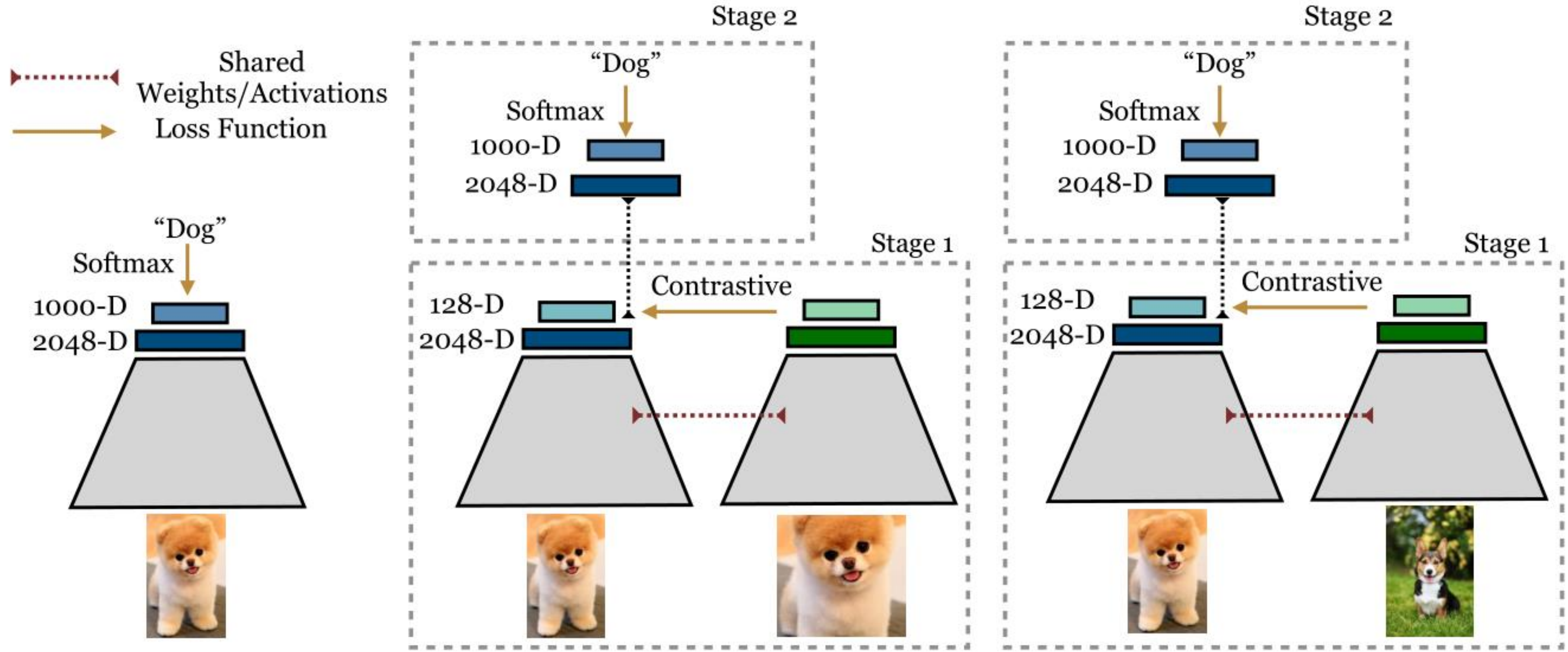
Method

Representation Learning Framework:

- *Data Augmentation* module, $Aug(\cdot)$.
- *Encoder Network*, $Enc(\cdot)$
- *Projection Network*, $Proj(\cdot)$

Method

Representation Learning Framework:



(a) Supervised Cross Entropy
(b) Self Supervised Contrastive
(c) Supervised Contrastive

Method

Contrastive Loss Functions

$$\mathcal{L}^{self} = \sum_{i \in I} \mathcal{L}_i^{self} = -\sum_{i \in I} \log \frac{\exp\left(\boldsymbol{z}_i \cdot \boldsymbol{z}_{j(i)}/\tau\right)}{\sum_{a \in A(i)} \exp\left(\boldsymbol{z}_i \cdot \boldsymbol{z}_a/\tau\right)} \tag{1}$$

Here, $\boldsymbol{z}_\ell = Proj(Enc(\tilde{\boldsymbol{x}}_\ell)) \in \mathcal{R}^{D_P}$, the $\bullet$ symbol denotes the inner (dot) product, $\tau \in \mathcal{R}^+$ is a scalar temperature parameter, and $A(i) \equiv I \setminus \{i\}$. The index i is called the *anchor*, index $j(i)$ is called the *positive*, and the other $2(N-1)$ indices ($\{k \in A(i) \setminus \{j(i)\}\}$) are called the *negatives*.

$$\mathcal{L}_{out}^{sup} = \sum_{i \in I} \mathcal{L}_{out,i}^{sup} = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp\left(\boldsymbol{z}_i \cdot \boldsymbol{z}_p/\tau\right)}{\sum_{a \in A(i)} \exp\left(\boldsymbol{z}_i \cdot \boldsymbol{z}_a/\tau\right)} \tag{2}$$

$$\mathcal{L}_{in}^{sup} = \sum_{i \in I} \mathcal{L}_{in,i}^{sup} = \sum_{i \in I} -\log \left\{ \frac{1}{|P(i)|} \sum_{p \in P(i)} \frac{\exp\left(\boldsymbol{z}_i \cdot \boldsymbol{z}_p/\tau\right)}{\sum_{a \in A(i)} \exp\left(\boldsymbol{z}_i \cdot \boldsymbol{z}_a/\tau\right)} \right\} \tag{3}$$

Method

Loss	Top-1
$\mathcal{L}_{out}^{sup}$	78.7%
$\mathcal{L}_{in}^{sup}$	67.4%

Table 1: ImageNet Top-1 classification accuracy for supervised contrastive losses on ResNet-50 for a batch size of 6144.

Experiments

Classification Accuracy

Dataset	SimCLR[3]	Cross-Entropy	Max-Margin [32]	SupCon
CIFAR10	93.6	95.0	92.4	**96.0**
CIFAR100	70.7	75.3	70.5	**76.5**
ImageNet	70.2	78.2	78.0	**78.7**

Loss	Architecture	Augmentation	Top-1	Top-5
Cross-Entropy (baseline)	ResNet-50	MixUp [61]	77.4	93.6
Cross-Entropy (baseline)	ResNet-50	CutMix [60]	78.6	94.1
Cross-Entropy (baseline)	ResNet-50	AutoAugment [5]	78.2	92.9
Cross-Entropy (our impl.)	ResNet-50	AutoAugment [30]	77.6	95.3
SupCon	ResNet-50	AutoAugment [5]	**78.7**	**94.3**
Cross-Entropy (baseline)	ResNet-200	AutoAugment [5]	80.6	95.3
Cross-Entropy (our impl.)	ResNet-200	Stacked RandAugment [49]	80.9	95.2
SupCon	ResNet-200	Stacked RandAugment [49]	**81.4**	**95.9**
SupCon	ResNet-101	Stacked RandAugment [49]	80.2	94.7

Experiments

Robustness

Loss	Architecture	rel. mCE	mCE
Cross-Entropy (baselines)	AlexNet [28]	100.0	100.0
	VGG-19+BN [44]	122.9	81.6
	ResNet-18 [17]	103.9	84.7
Cross-Entropy (our implementation)	ResNet-50	96.2	68.6
	ResNet-200	69.1	52.4
Supervised Contrastive	ResNet-50	**94.6**	**67.2**
	ResNet-200	**66.5**	**50.6**

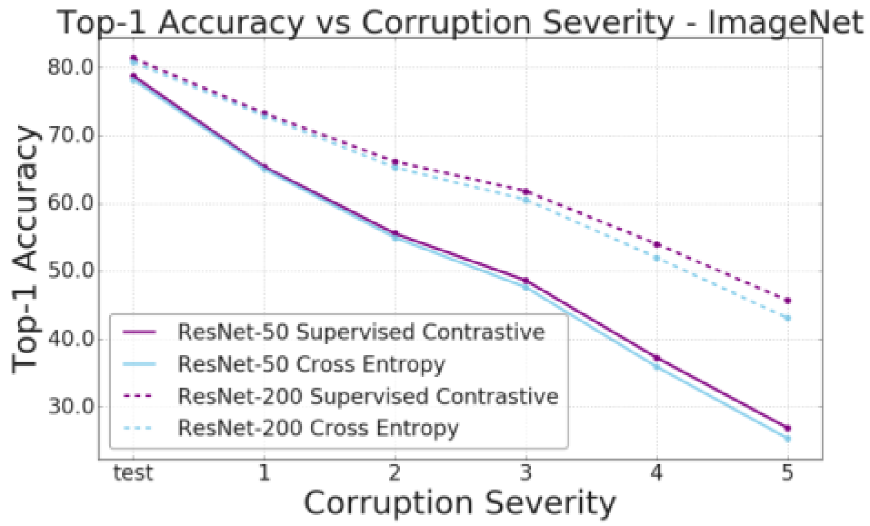
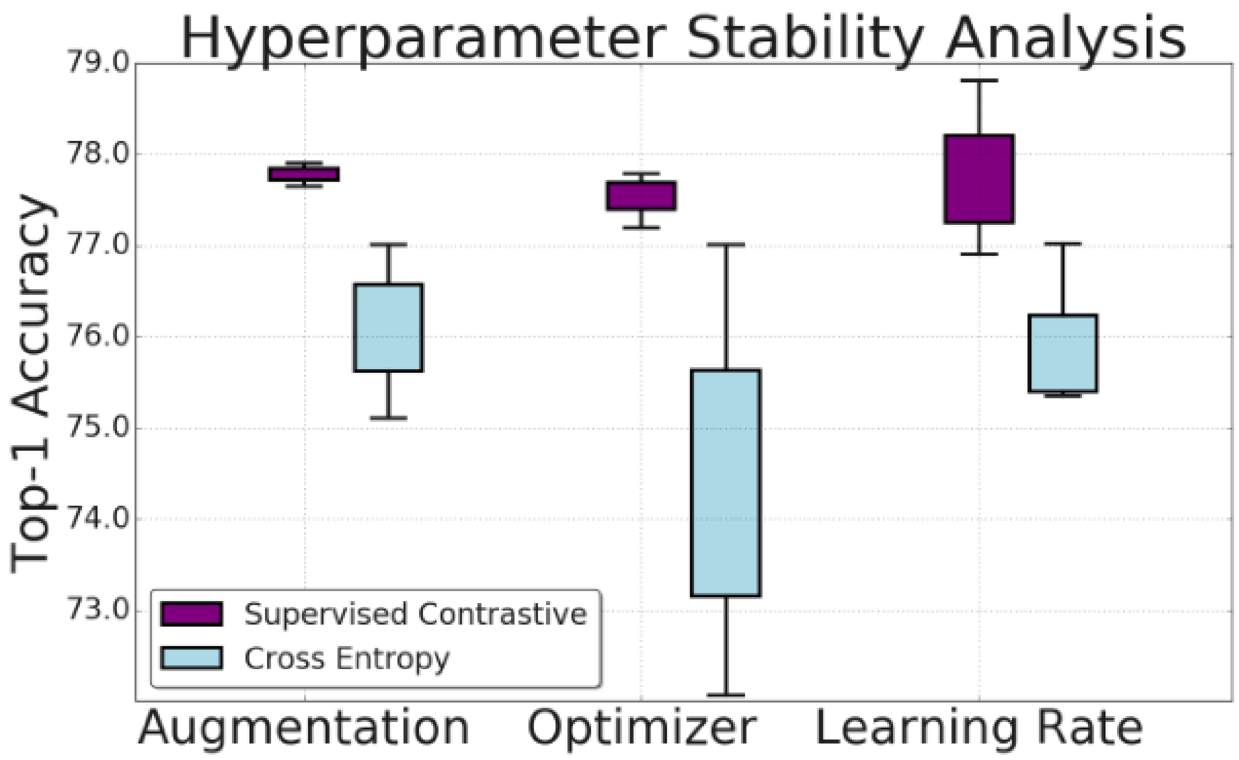
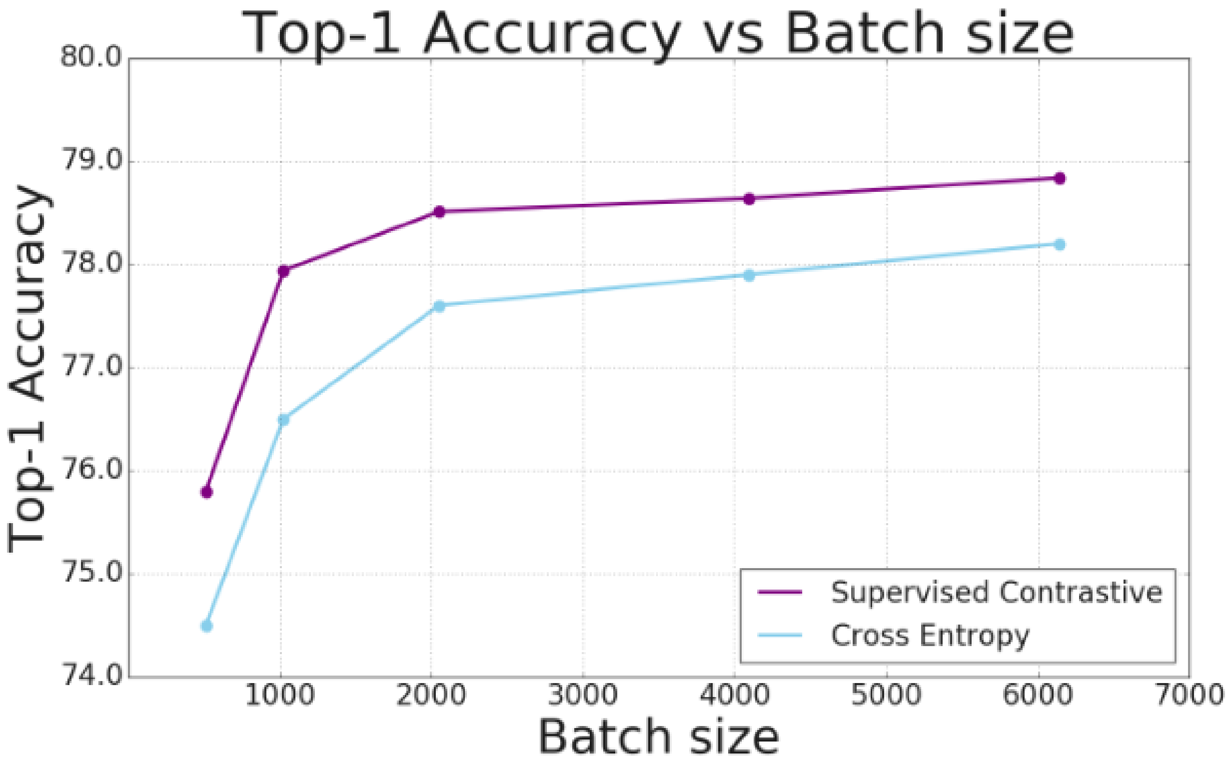


Figure 3: Training with supervised contrastive loss makes models more robust to corruptions in images. **Left**: Robustness as measured by Mean Corruption Error (mCE) and relative mCE over the ImageNet-C dataset [19] (lower is better). **Right**: Mean Accuracy as a function of corruption severity averaged over all various corruptions. (higher is better).

Experiments

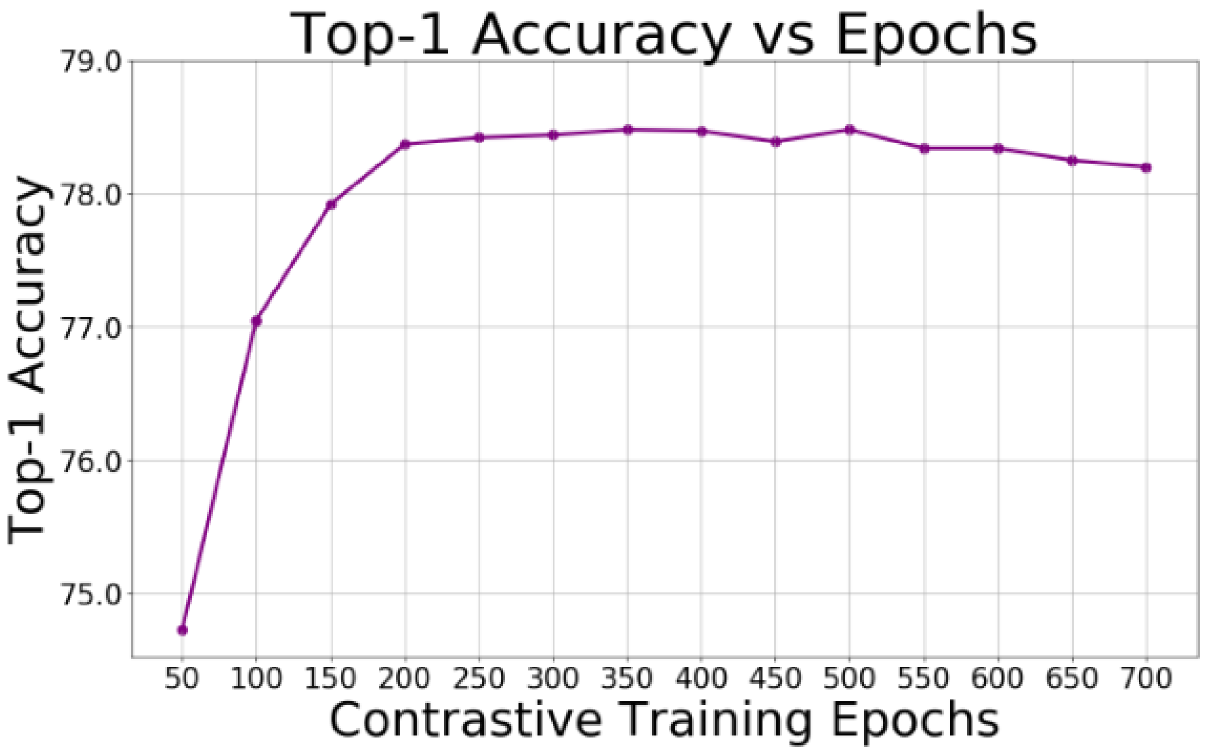
Hyperparameter Stability

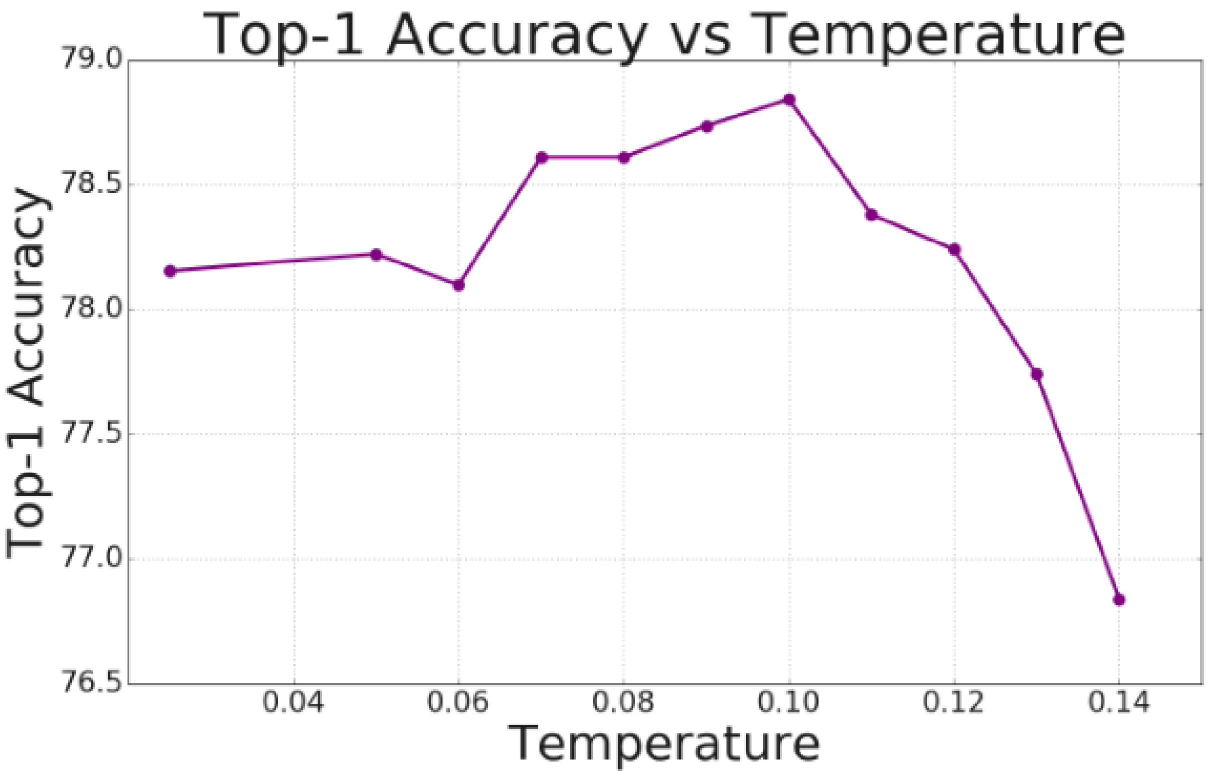




Experiments

Training Details





THANKS