Supervised Contrastive Learning

School of Industrial and Management Engineering, Korea University

Go Eun Chae

Contents

Research Purpose

❖ Supervised Contrastive Learning (2020)

- Google Research 연구, 841회 인용 (2022.05.24 기준)
- **Supervised 와 Self-supervised Contrastive Learning 의 각 장점을 통합하여 활용**

Supervised Contrastive Learning

Pranney 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

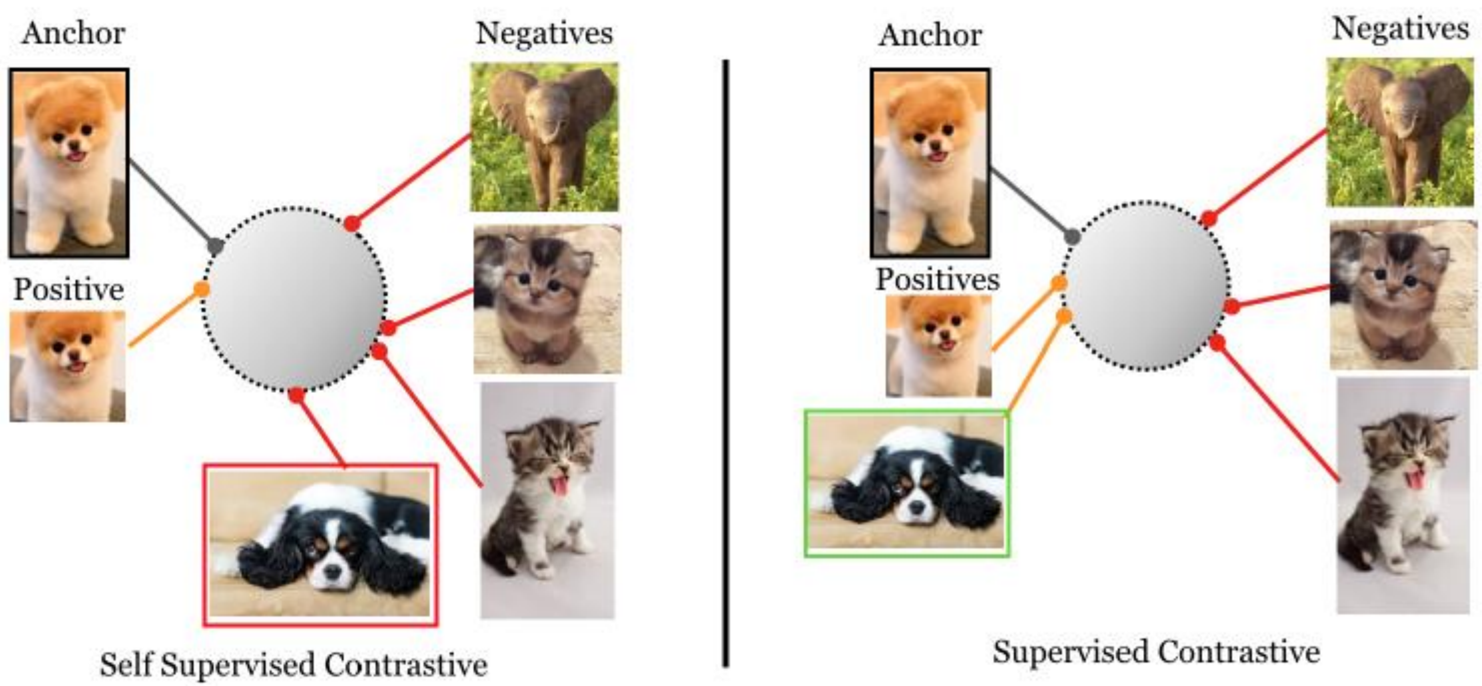
Abstract

Contrastive learning applied to self-supervised representation learning has seen a resurgence in recent years, leading to state of the art performance in the unsupervised training of deep image models. Modern batch contrastive approaches subsume or significantly outperform traditional contrastive losses such as triplet, max-margin and the N-pairs loss. In this work, we extend the self-supervised batch contrastive approach to the *fully-supervised* setting, allowing us to effectively leverage label information. Clusters of points belonging to the same class are pulled together in embedding space, while simultaneously pushing apart clus-

Research Purpose

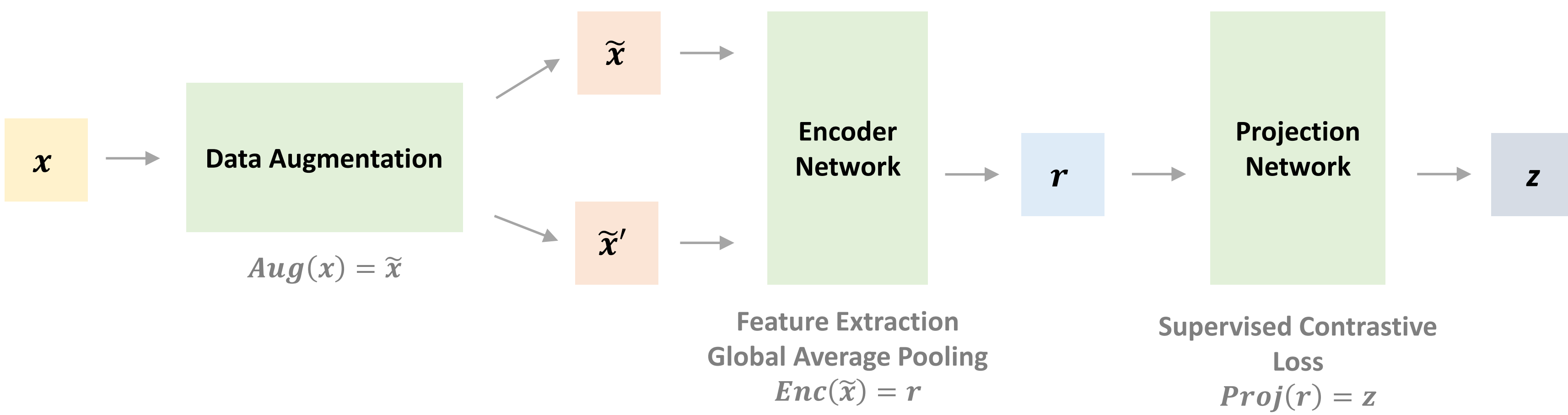
❖ Main Idea

- Anchor 의 **Multiple Positives** 를 허용하는 Contrastive Loss Function 제안
- 같은 Class 의 Samples 는 **모두 Positives** 로 생각
- **Fully Supervised** 설정에 Contrastive Learning 적용



Proposed Method

- ❖ **Step 1: Supervised Contrastive Loss 활용한 Pre-training**
- ❖ **Step 2: Categorical Cross Entropy Loss 활용한 Fine-tuning**
- ❖ **Main Components of Framework (Step 1)**
  - Data Augmentation: 데이터 증강 기법 적용하여 다양한 패턴 반영, Data의 다른 관점 표현
  - Encoder Network: ResNet 구조 적용하여 **Representation Vector** r 추출
  - Projection Network: 정규화 한 **Representation Vector** z 추출 후 Supervised Contrastive Loss 적용



Proposed Method

❖ Self-Supervised Contrastive Loss

- $L = \sum_{i \in I} L_i^{self} = -\sum_{i \in I} log \frac{\exp(z_i \cdot z_{j(i)}/\tau)}{\sum_{a \in A(i)} \exp(z_i \cdot z_a/\tau)}$
- j(i): **Positive 1개** (Anchor i 로 증강한 데이터)

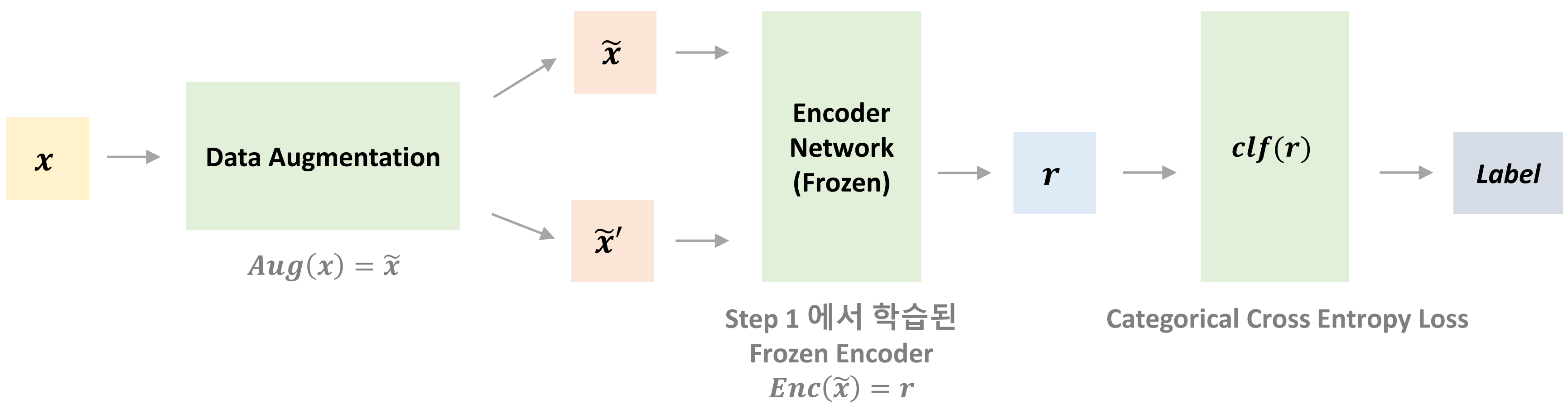
❖ Supervised Contrastive Loss

- $L = \sum_{i \in I} L_i^{sup} = -\sum_{i \in I} log\{\frac{\mathbf{1}}{|\boldsymbol{P(i)}|} \sum_{i \in I} \frac{\exp(z_i \cdot z_p/\tau)}{\sum_{\boldsymbol{a \in A(i)}} \exp(z_i \cdot z_a/\tau)}\}$
- i : Anchor, $i \in I \equiv \{1, \ldots, 2N\}$ (Input Data N 개 + Augmented Data N개, Multiviewed Batch)
- $P(i) \equiv \{p \in A(i): \tilde{y}_p = \tilde{y}_i\}$: Anchor i 의 **모든 Positives 의 Index 집합**
- $P(i)$: Anchor와 **같은 Label**을 가지는 Anchor의 증강 데이터
- Postivie (같은 Label) 와 유사하도록, Negative (다른 Label) 와 유사하지 않도록 학습
- $z = (feaetrues, features^T) \rightarrow Supervised\ Contrastive\ Loss \rightarrow Label$ 부여

Proposed Method

❖ Step 2: Fine-tuning

- Pre-training 에서 학습한 **Encoder 고정**
- Projection Network 대신 **Categorical Cross Entropy Loss** $clf(r)$ 사용
  - ✓ Supervised Contrastive Loss 의 성능을 살펴보기 위해 1-layer 신경망, 선형모델로 구성
- **Supervised Learning**으로 Fine-tuning
- 실질적인 Image Classification 수행



Experiments

❖ Image Classification Accuracy

- CIFAR-10, CIFAR-100, ImageNet Datasets 활용
- Classification Accuracy 통해 **SupCon Loss** 평가 진행
- SupCon 이 Cross-Entropy, Margin-Classifier 에 비해 잘 적용, 높은 성능
- ResNet50 + AutoAugment 통해 **78.7% SOTA** 달성

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 [60]	77.4	93.6
Cross-Entropy (baseline)	ResNet-50	CutMix [59]	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 to Image Corruptions and Reduced Training Data

- **Corruption에 대한 모델 성능 측정**을 위해 ImageNet-C Dataset 활용
- Mean Corruption Error (mCE) & Relative Mean Corruption (rel.mCE) 기준으로 성능 비교
  - ✓ ImageNet Test Set 과 비교하여 성능의 Average Degradation 측정
  - ✓ 다른 Acc 가지는 모델 비교에는 rel.mCE 가 더 적합
  - ✓ Corruption에 대한 절대적인 Robustness 측정에는 mCE 가 더 나은 지표
- SupCon은 다른 모델에 비해 낮은 mCE 가짐, **Robustness 증명**

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**

Conclusion

❖ Conclusion

- Anchor에 대해 **Multiple Positives 허용하는** 새로운 Contrastive Loss Function 제안
- 결과적으로 Fully Supervised 설정에 Contrastive Learning 적용
- SupCon Loss 활용했을 때 다양한 Datasets 에 대해 **분류 성능 증가** 확인
- Supervised Contrastive Learning 을 통해 분류기의 **Accuracy & Robustness** 발전
- SupCon의 Gradient **가 Hard Positives, Hard Negatives 로 부터의 학습** 장려함을 증명
- 실험적으로 SupCon 이 Cross-entropy 보다 **Hyperparameter 변화에 덜 민감**한 결과

Thank You