A Simple Framework for Contrastive Learning of Visual Representations (SimCLR)

School of Industrial and Management Engineering, Korea University

Insung Baek

Contents

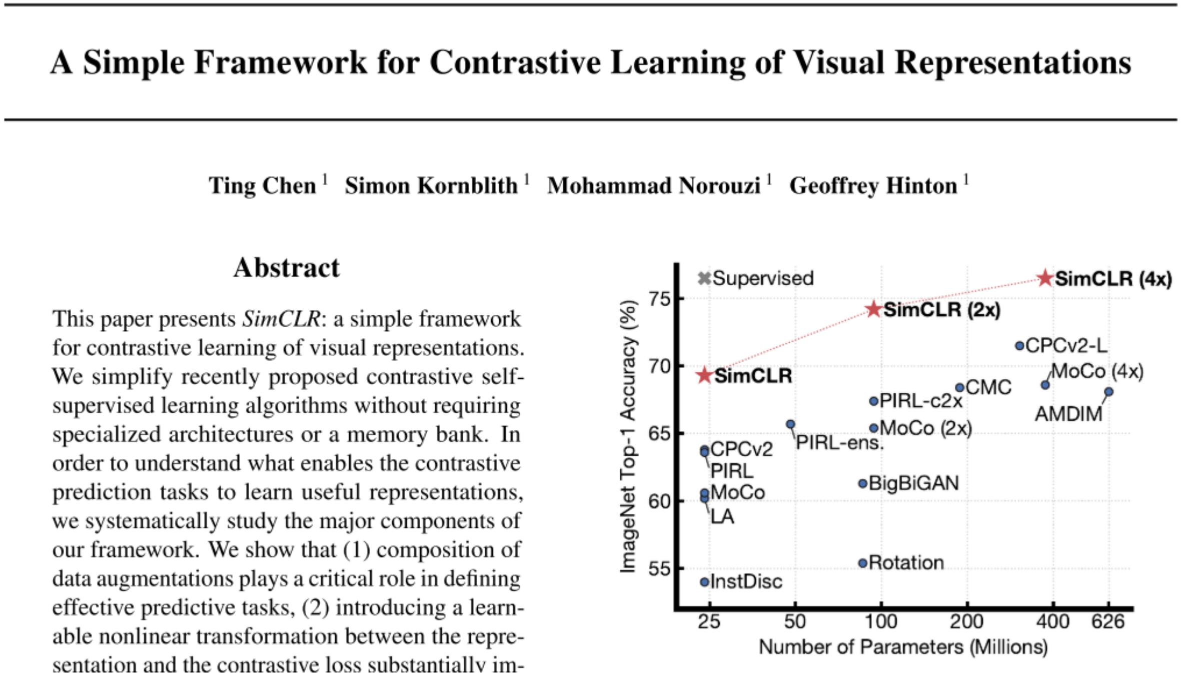
- Research Purpose
- Proposed Method
- Experiments
- Conclusion

Research Purpose

SimCLR

❖ A Simple Framework for Contrastive Learning of Visual Representations (ICML, 2020)

- Google Brain에서 진행한 연구, 2022년 5월 27일 기준 4557회 인용됨
- 특정 이미지에서 augmentation을 적용한 데이터는 positive pair, 그 외의 이미지에 augmentation을 적용한 데이터는 negative pair로 보고 contrastive learning을 진행한 연구임



A Simple Framework for Contrastive Learning of Visual Representations

Ting Chen[1] Simon Kornblith[1] Mohammad Norouzi[1] Geoffrey Hinton[1]

Abstract

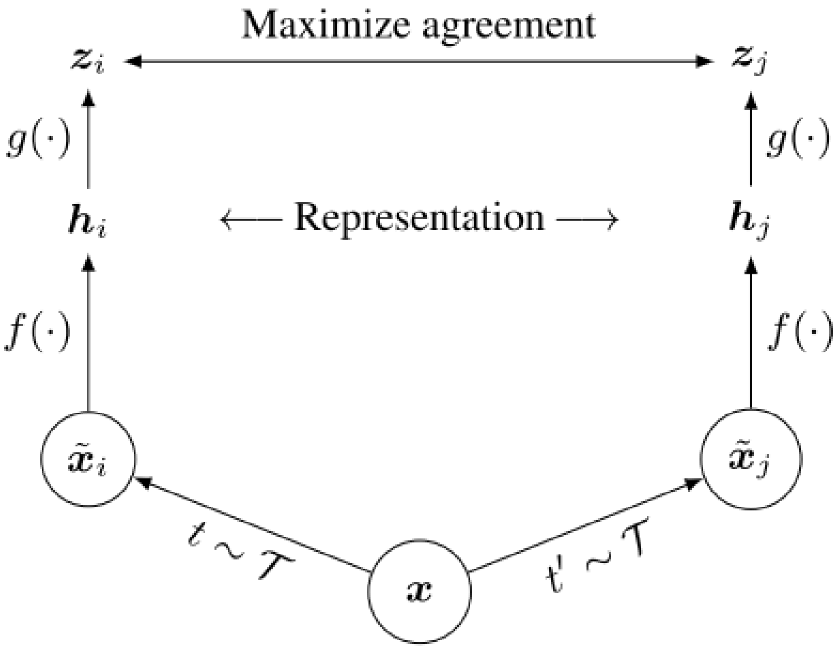
This paper presents *SimCLR*: a simple framework for contrastive learning of visual representations. We simplify recently proposed contrastive self-supervised learning algorithms without requiring specialized architectures or a memory bank. In order to understand what enables the contrastive prediction tasks to learn useful representations, we systematically study the major components of our framework. We show that (1) composition of data augmentations plays a critical role in defining effective predictive tasks, (2) introducing a learnable nonlinear transformation between the representation and the contrastive loss substantially im-

Reference: Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." International conference on machine learning. PMLR, 2020.

Proposed Method

Contrastive learning framework

- ❖ A simple framework for contrastive learning of visual representations
  1) 원본 이미지 x에 대해서 augmentation을 통해 $\tilde{x}_i$와 $\tilde{x}_j$를 생성함
  2) Encoder network $f(\cdot)$와 projection head $g(\cdot)$를 통해 z_i와 z_j를 산출함
  3) z_i와 z_j의 일치성이 최대화 될 수 있도록 $f(\cdot)$와 $g(\cdot)$를 학습함

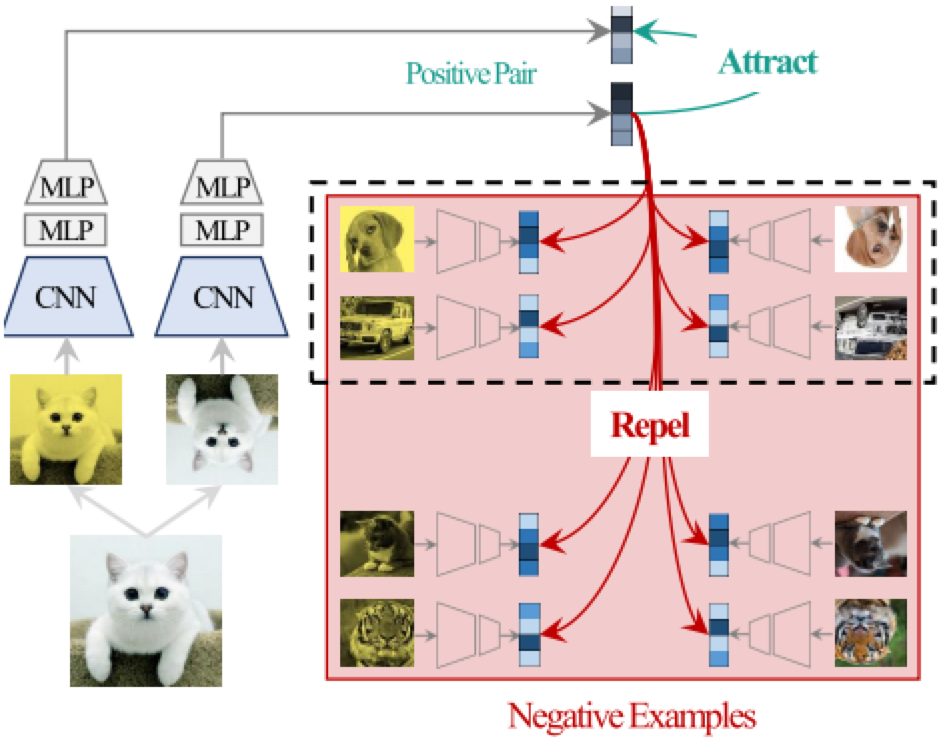


Reference: Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." International conference on machine learning. PMLR, 2020.

Proposed Method

Contrastive learning framework of SimCLR

- ❖ 특정 이미지 (고양이)에 대해 augmentation을 적용한 두 이미지를 생성해 positive pair로 봄
- ❖ 학습시 배치 내 나머지 이미지에 augmentation을 적용한 모든 이미지를 negative pair로 봄
- ❖ Feature space에서 positive pair는 가깝게 negative pair는 멀어지도록 학습을 유도함



Positive pair

$$Loss_{i,j} = -log\frac{\exp(\frac{sim(\boldsymbol{z_i},\boldsymbol{z_j})}{\tau})}{\sum_{k=1(k\neq i)}^{2N}\exp(\frac{sim(\boldsymbol{z_i},\boldsymbol{z_k})}{\tau})}$$

Negative pair

<Loss Function>

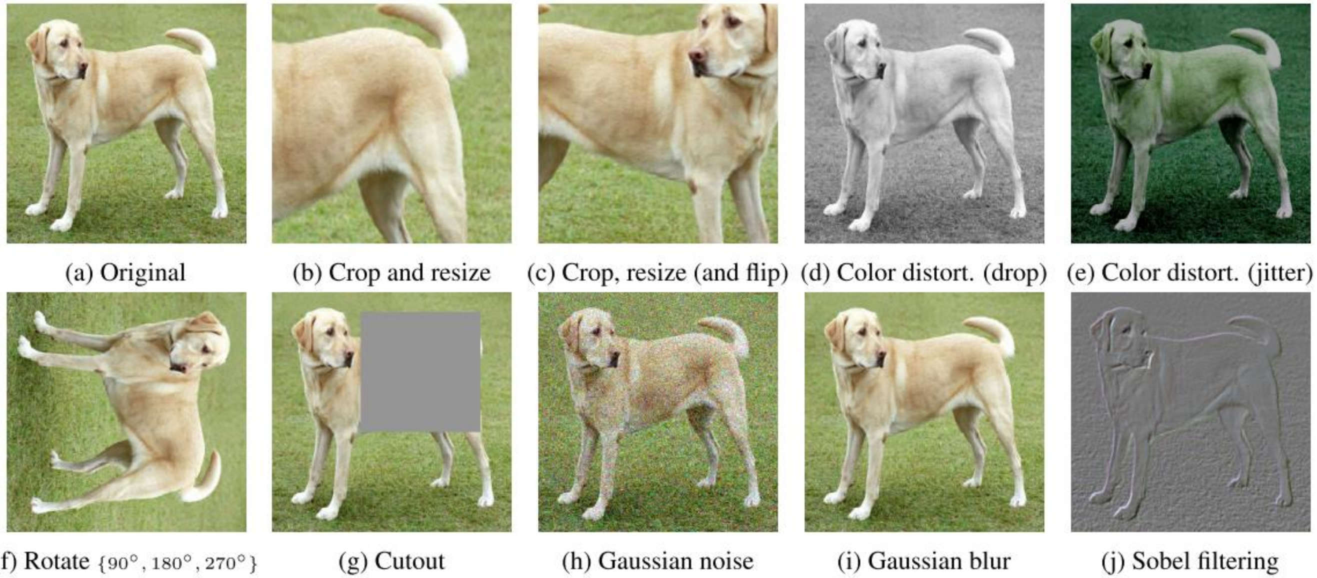
Reference: "Applications of Self-Supervised Learning", 이영재 연구원 (DMQA 세미나)

Proposed Method

Types of image augmentation

❖ 이미지에 적용하는 augmentation 기법은 크게 3가지로 나눌 수 있음

1) Spatial/geometric transformation: crop and resize, flip, rotation, cutout
2) Color appearance transformation: color distortion (color drop, jitter, brightness)
3) Appearance transformation except for color: gaussian noise/blur, Sobel filtering



(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

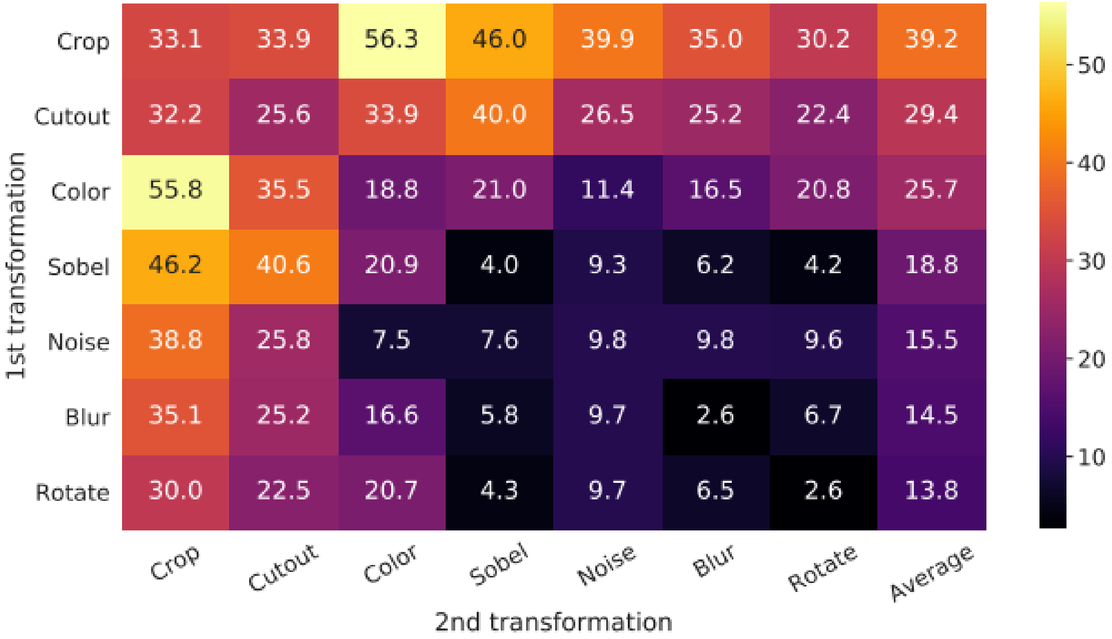
Figure 4. Illustrations of the studied data augmentation operators. Each augmentation can transform data stochastically with some internal parameters (e.g. rotation degree, noise level). Note that we *only* test these operators in ablation, the *augmentation policy used to train our models* only includes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

Reference: Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." International conference on machine learning. PMLR, 2020.

Experiment

Accuracy by augmentation combination

- ❖ 총 7개의 data augmentation 기법을 1개 또는 2개 조합을 사용해 성능을 도출한 결과
- ❖ 1개의 augmentation 기법을 사용했을 때보다 2개 조합해서 사용했을 때 더 좋은 성능 보임
- ❖ 2개 augmentation 조합했을 때 prediction task 난이도가 높아지면서 representation quality 상승



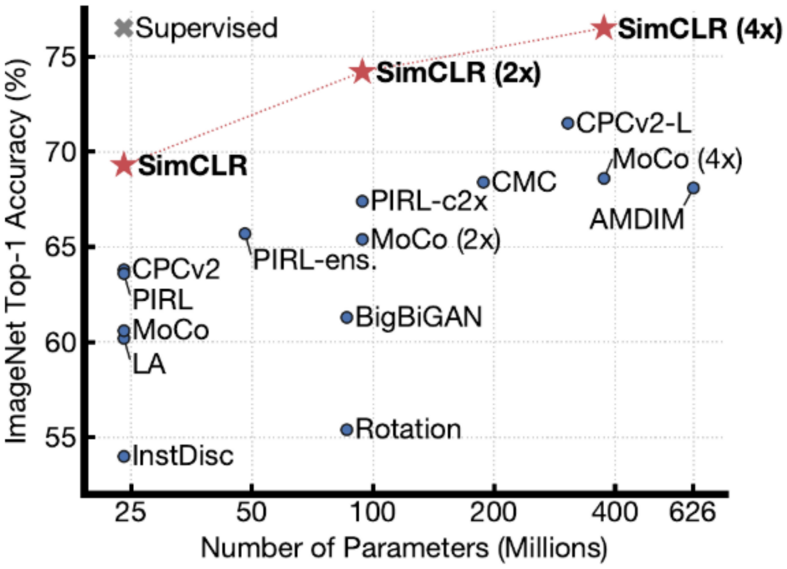
	Crop	Cutout	Color	Sobel	Noise	Blur	Rotate	Average
Crop	33.1	33.9	56.3	46.0	39.9	35.0	30.2	39.2
Cutout	32.2	25.6	33.9	40.0	26.5	25.2	22.4	29.4
Color	55.8	35.5	18.8	21.0	11.4	16.5	20.8	25.7
Sobel	46.2	40.6	20.9	4.0	9.3	6.2	4.2	18.8
Noise	38.8	25.8	7.5	7.6	9.8	9.8	9.6	15.5
Blur	35.1	25.2	16.6	5.8	9.7	2.6	6.7	14.5
Rotate	30.0	22.5	20.7	4.3	9.7	6.5	2.6	13.8

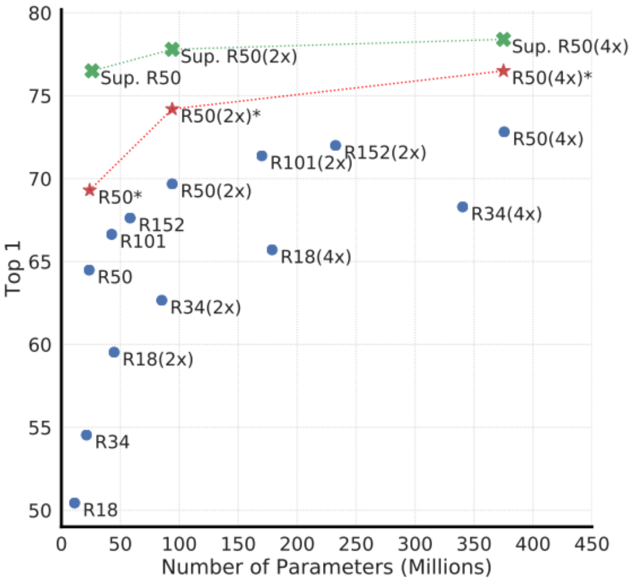
Reference: Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." International conference on machine learning. PMLR, 2020.

Experiment

Model performance

- ❖ 다양한 self-supervised 방법론 중 본 연구에서 제안한 SimCLR가 가장 우수한 성능을 보임
- ❖ ResNet 모델을 기준으로 depth와 width가 증가할수록 더 좋은 성능을 보임
- ❖ ResNet-50(4x) 모델의 경우 supervised learning 결과에 가깝게 좋은 성능을 보임





<Accuracy with different self-supervised methods>

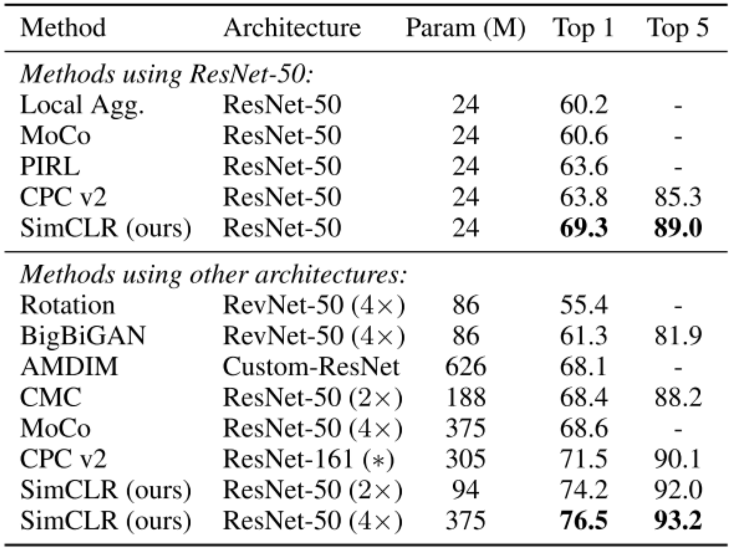
<Accuracy with varied depth and width>

Reference: Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." International conference on machine learning. PMLR, 2020.

Experiment

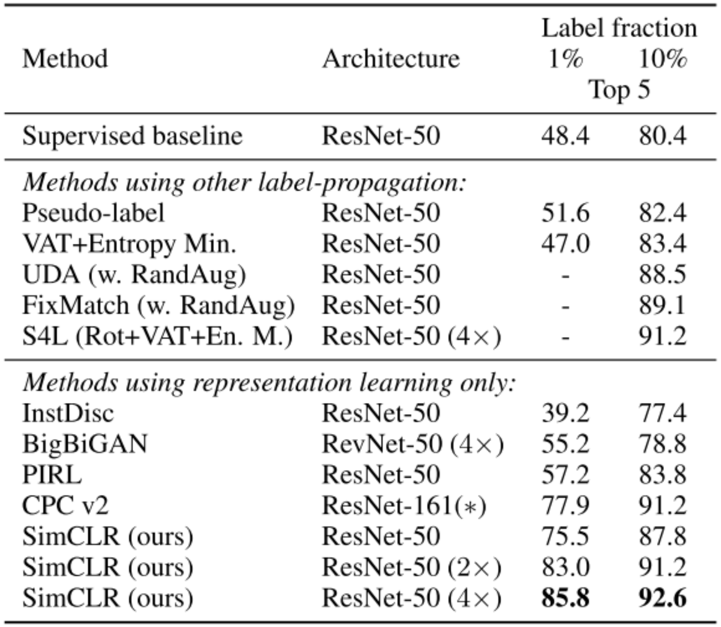
Model performance

- ❖ 다양한 self-supervised learning 기법 중 Top1, Top5 accuracy 모두 SimCLR가 가장 우수함
- ❖ Label 비율이 1%, 10%에서도 SimCLR가 가장 우수한 성능을 보임



Method	Architecture	Param (M)	Top 1	Top 5
Methods using ResNet-50:				
Local Agg.	ResNet-50	24	60.2	-
MoCo	ResNet-50	24	60.6	-
PIRL	ResNet-50	24	63.6	-
CPC v2	ResNet-50	24	63.8	85.3
SimCLR (ours)	ResNet-50	24	**69.3**	**89.0**
Methods using other architectures:				
Rotation	RevNet-50 (4×)	86	55.4	-
BigBiGAN	RevNet-50 (4×)	86	61.3	81.9
AMDIM	Custom-ResNet	626	68.1	-
CMC	ResNet-50 (2×)	188	68.4	88.2
MoCo	ResNet-50 (4×)	375	68.6	-
CPC v2	ResNet-161 (∗)	305	71.5	90.1
SimCLR (ours)	ResNet-50 (2×)	94	74.2	92.0
SimCLR (ours)	ResNet-50 (4×)	375	**76.5**	**93.2**

Table 6. ImageNet accuracies of linear classifiers trained on representations learned with different self-supervised methods.



Method	Architecture	Label fraction 1%	 10%
		Top 5	
Supervised baseline	ResNet-50	48.4	80.4
Methods using other label-propagation:			
Pseudo-label	ResNet-50	51.6	82.4
VAT+Entropy Min.	ResNet-50	47.0	83.4
UDA (w. RandAug)	ResNet-50	-	88.5
FixMatch (w. RandAug)	ResNet-50	-	89.1
S4L (Rot+VAT+En. M.)	ResNet-50 (4×)	-	91.2
Methods using representation learning only:			
InstDisc	ResNet-50	39.2	77.4
BigBiGAN	RevNet-50 (4×)	55.2	78.8
PIRL	ResNet-50	57.2	83.8
CPC v2	ResNet-161(∗)	77.9	91.2
SimCLR (ours)	ResNet-50	75.5	87.8
SimCLR (ours)	ResNet-50 (2×)	83.0	91.2
SimCLR (ours)	ResNet-50 (4×)	**85.8**	**92.6**

Table 7. ImageNet accuracy of models trained with few labels.

<Accuracy with different self-supervised methods>

<Accuracy with few labels>

Reference: Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." International conference on machine learning. PMLR, 2020.

Experiment

Model performance with different batch size and epochs

- ❖ Contrastive learning은 안정적인 학습을 위해 충분한 negative sample이 필수적임
- ❖ 배치 크기가 커질수록 충분한 negative sample을 확보해 모델 성능이 향상됨
- ❖ 학습 시간이 길어질수록 충분한 negative sample을 볼 수 있어서 모델 성능이 향상됨

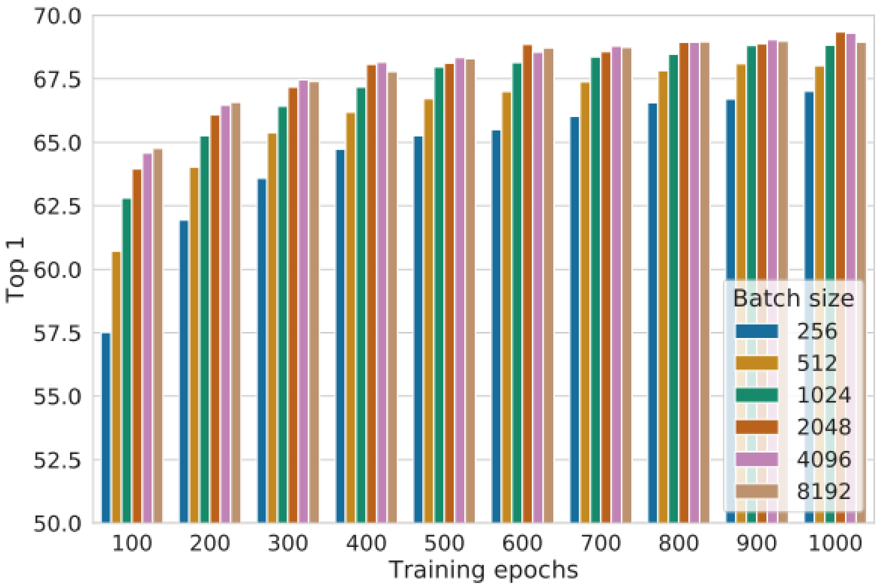


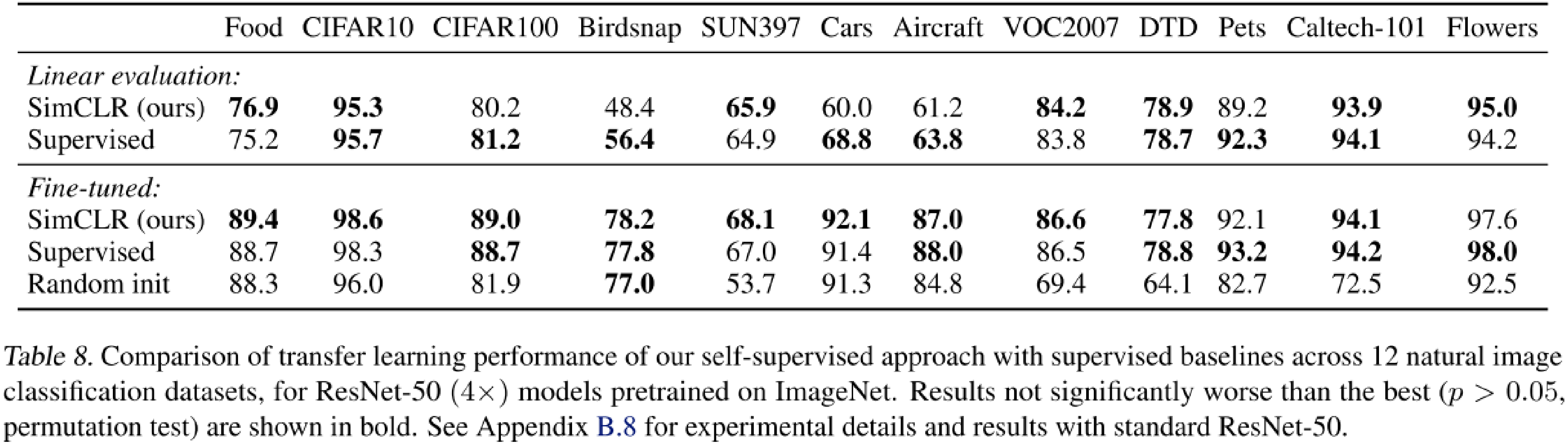
Figure 9. Linear evaluation models (ResNet-50) trained with different batch size and epochs. Each bar is a single run from scratch.[10]

Reference: Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." International conference on machine learning. PMLR, 2020.

Experiment

Comparison of SimCLR and supervised learning

- ❖ 다양한 데이터셋에 대해서 transfer learning으로 학습했을 때 모델 성능을 도출해 비교함
- ❖ SimCLR의 성능이 supervised learning의 성능에 준하거나 그 이상인 경우가 있었음



	Food	CIFAR10	CIFAR100	Birdsnap	SUN397	Cars	Aircraft	VOC2007	DTD	Pets	Caltech-101	Flowers
Linear evaluation:												
SimCLR (ours)	**76.9**	**95.3**	80.2	48.4	**65.9**	60.0	61.2	**84.2**	**78.9**	89.2	**93.9**	**95.0**
Supervised	75.2	**95.7**	**81.2**	**56.4**	64.9	**68.8**	**63.8**	83.8	**78.7**	**92.3**	**94.1**	94.2
Fine-tuned:												
SimCLR (ours)	**89.4**	**98.6**	**89.0**	**78.2**	**68.1**	**92.1**	**87.0**	**86.6**	**77.8**	92.1	**94.1**	97.6
Supervised	88.7	98.3	**88.7**	**77.8**	67.0	91.4	**88.0**	86.5	**78.8**	**93.2**	**94.2**	**98.0**
Random init	88.3	96.0	81.9	**77.0**	53.7	91.3	84.8	69.4	64.1	82.7	72.5	92.5

Table 8. Comparison of transfer learning performance of our self-supervised approach with supervised baselines across 12 natural image classification datasets, for ResNet-50 (4×) models pretrained on ImageNet. Results not significantly worse than the best ($p > 0.05$, permutation test) are shown in bold. See Appendix B.8 for experimental details and results with standard ResNet-50.

Reference: Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." International conference on machine learning. PMLR, 2020.

Conclusion

❖ Conclusion

- SimCLR은 contrastive self-supervised learning 방법론의 하나로 학습시 배치내에서 원본 데이터에 augmentation을 적용한 데이터는 positive pair로 그 외에 데이터에 augmentation을 적용한 데이터는 negative pair로 보고 학습을 진행함
- Feature space에서 positive pair는 가까워지게 negative pair는 멀어지도록 학습을 진행함
- Augmentation은 1개보다 2개를 사용했을 때 prediction task 난이도가 높아지면서 representation quality가 높아져 모델 성능이 더 향상됨
- 충분한 negative sample을 확보하는 것이 중요하기 때문에 배치 크기가 커질수록 학습시간이 길어질수록 모델 성능이 더 향상됨
- SimCLR을 사용했을 때 supervised learning의 성능에 준하거나 그 이상인 경우를 도출할 수 있음

Thank You

Appendix

Appendix

Reference

- Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." International conference on machine learning. PMLR, 2020.