Self-Supervised / Representation Learning

Zhonglin Wang, Hao Wu, Zhanwang Liu

Duke PRATT SCHOOL of ENGINEERING

Introduction

Self-supervised learning is a rapidly growing area in the field of artificial intelligence. The core idea of it is to use the data itself as its own supervision to learn effective visual representations without human supervision.

In this work, we implemented SimCLR and RotNet frameworks on the CIFAR-10 dataset with different architectures like NIN, ResNet.

Then we compare the different performances of the two frameworks on the different portion of the labeled data.

Contributions:

- **Zhonglin Wang**: Implement contrastive learning using SimCLR on ResNet-50 and ResNet-20; perform linear evaluation to reproduce results of the simCLR paper and compare accuracy with parameters
- **Hao Wu**: Implement Rotnet on the NIN structure, comparing the performance of each feature maps generated and conduct limited label test
- **Zhanwang Liu:** implement the SimCLR and RotNet on Resnet-20, finetune two methods on linear evaluation and semi-supervised learning

Two random augmentations of the same image on CIFAR-10 for SimCLR training

RotNet rotation images for 0,90,180,270 degree

Methodology

SimCLR: A Simple Framework for Contrastive Learning of Visual Representations

- Use data augmentation module that transforms any given data example randomly generates two correlated views.
- Extract representations with ResNet models.
- Map the representations with projection head and compute the contrastive loss
- Through away the projection head and finetune the pretrained model on CIFAR-10 dataset.

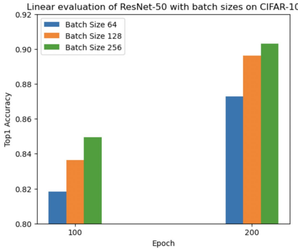
RotNet : unsupervised representation learning by predicting image rotations

- Transform the data with 4 different degrees rotations.
- Extract representations with ResNet models.
- Train the model to predict the image rotations to get the representations of the data.
- Finetune the pretrained model on CIFAR-10 dataset.

Rotnet - NIN:

- Construct self-supervised NIN model by feeding image with 4 degrees rotations
- Find the top feature map generated by specific conv block
- Keep top feature map freezed, and add additional conv/fc blocks to do the semi-supervised learning

Experimental Evaluation



Linear evaluation of SimCLR based on ResNet-50 with different batch size and number of epochs

Method	Architecture	Label fraction	
		1%	10%
Supervised base line	ResNet50	48.4	80.4
	NIN (4 blocks)	52.3	76.7
RotNet	ResNet-20	58.6	74.8
	NIN + Conv	**64.8**	**80.7**
SimCLR	ResNet-20	68.8	75.3
	ResNet-50	**74.5**	**79.4**

Performance of RotNet and SimCLR with limited labels under different architecture

Conclusion

- SimCLR (ResNet-50) with batch size 256 achieved the best accuracy in our experiments, which is consistent with the trend in the paper.
- The errors of accuracy for batch size 256 between ours and the paper is about 1% and 0.6% on the 100 and 200 epoch, respectively.
- Overall, both ResNet and NIN architecture suited better than supervised based line under limited labels if applying RotNet or SimCLR
- SimCLR performs better than RotNet if there is only extremely limited labels applied on training
- Within the same method, NIN+Conv performs better than ResNet-20 in Rotnet, and ResNet-50 performs better than ResNet20 in SimCLR

References

References in 14pt font

1. Chen et al., “A Simple Framework for Contrastive Learning of Visual Representations,” (https://arxiv.org/pdf/2002.05709.pdf)

2. Gidaris et al., “Unsupervised Representation Learning by Predicting Image Rotations”, (https://arxiv.org/abs/1803.07728)