NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

School of Computer Science and Engineering
College of Engineering

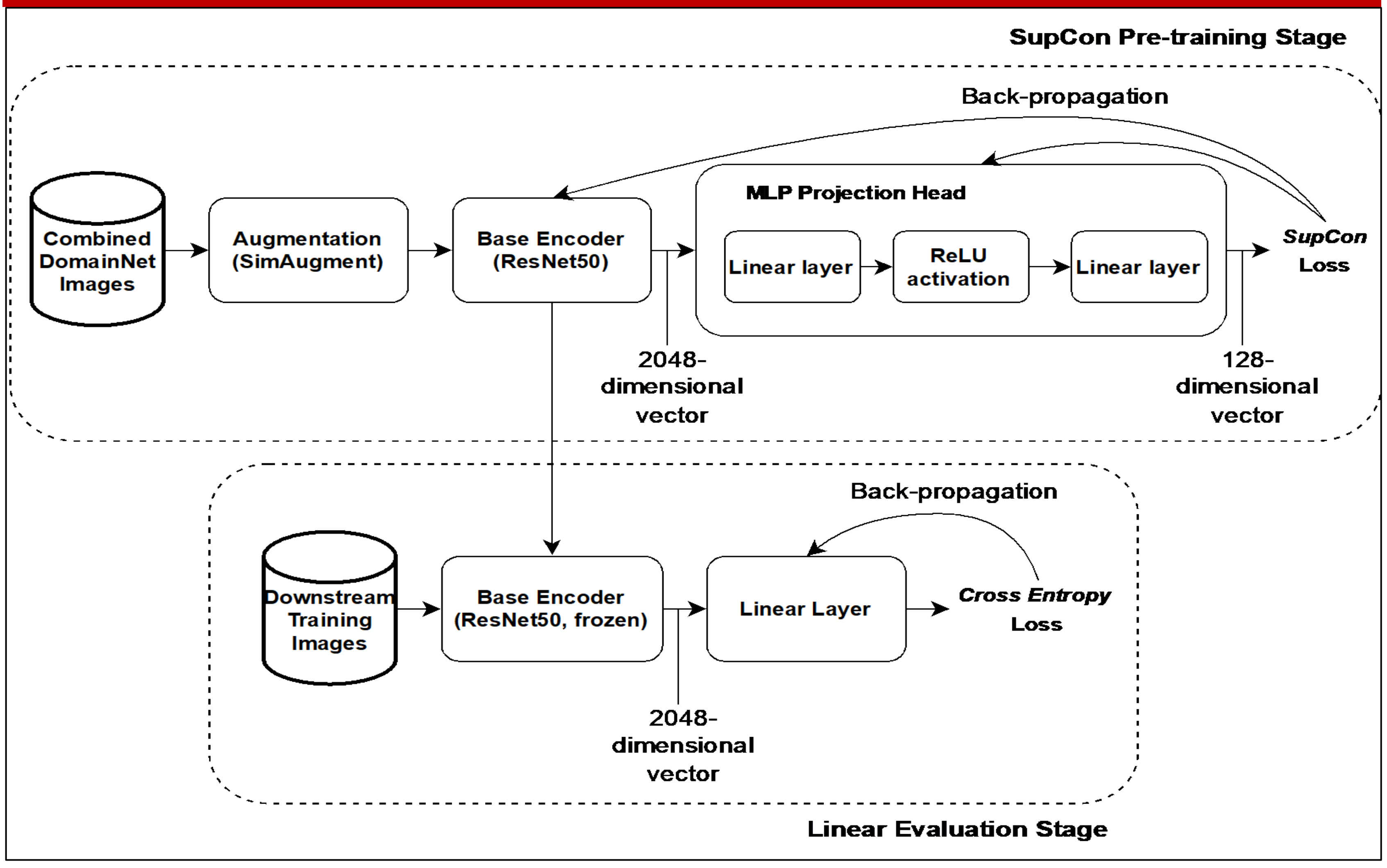
Self-Supervised and Supervised Contrastive Learning

Student: Alvin Tan De Jun   Supervisor: A/P Yeo Chai Kiat

Introduction

- Contrastive learning is a method to guide a model to learn an embedding space, where samples from the same class will be pulled closer together and samples from a different class will be pulled apart from each other.
- This project explored contrastive learning in self-supervised (using SimCLR) and supervised (using Supervised Contrastive Learning) manner. I will present my findings for the Supervised Contrastive Learning part in this poster.
- Using Supervised Contrastive Learning (SupCon), I attempted to learn representations from the multi-domain DomainNet dataset and then evaluate the transferability of the representations learned on other downstream datasets. The results obtained will be compared to a baseline model that was trained using the widely used cross entropy (CE) loss.

Supervised Contrastive Learning Framework



Datasets Used

DomainNet:

- Consists of common objects in six different domains: sketch, real, quickdraw, painting, infograph, clipart
- Each domain contains 345 classes
- Combined all domains into one dataset for pre-training of the SupCon model

Downstream Datasets:

- CIFAR10
- CIFAR100
- Flowers102
- Aircraft
- SVHN
- Kaokore
- DTD

Results

Table 1: Accuracy (%) for SupCon and Cross Entropy Model on the Downstream Datasets for Linear Evaluation. (Note: Mean-per-class accuracy is provided for Aircraft and Flowers102 while the rest are top-1 accuracy. Mean and standard deviation over 5-runs are provided.)

	CIFAR10	CIFAR100	Aircraft	Flowers102	SVHN	Kaokore	DTD	Mean
SupCon	92.31 ± 0.04	75.74 ± 0.05	36.53 ± 0.27	75.09 ± 0.09	70.03 ± 0.05	76.63 ± 0.2	51.26 ± 0.13	**68.23**
CE	90.16 ± 0.06	70.99 ± 0.07	26.95 ± 0.35	65.92 ± 0.11	64.6 ± 0.07	71.22 ± 0.25	45.43 ± 0.3	62.18

- Table 1 reports the average test accuracy along with its standard deviation. The SupCon model performed, on average, 6.05% better than the cross entropy model on the 7 downstream datasets when trained with a multi-domain dataset.

Ablation Studies

Table 2: Effect of Augmentations

Augmentation	CIFAR10	CIFAR100	Aircraft	Flowers102	SVHN	Kaokore	DTD	Mean
AutoAugment	82.64 ± 0.02	59.5 ± 0.08	33.72 ± 0.15	52.77 ± 0.24	86.4 ± 0.04	70.58 ± 0.23	39.07 ± 0.19	60.67
RandAugment	87.35 ± 0.03	67.52 ± 0.03	33.47 ± 0.19	66.67 ± 0.04	83.8 ± 0.1	74.42 ± 0.15	43.38 ± 0.16	65.23
SimAugment	92.31 ± 0.04	75.74 ± 0.05	36.53 ± 0.27	75.09 ± 0.09	70.03 ± 0.05	76.63 ± 0.2	51.26 ± 0.13	68.23
Stacked RandAugment	91.99 ± 0.02	76.04 ± 0.03	35.77 ± 0.21	75.35 ± 0.09	73.73 ± 0.09	75.94 ± 0.14	50.26 ± 0.05	**68.44**

- SimAugment and Stacked RandAugment, which are stronger augmentations, performed better than AutoAugment (ImageNet policy) and RandAugment in terms of the mean accuracy, except for SVHN
- I conjecture that in SVHN, the house numbers are often sheared or skewed, and the transformations in AutoAugment (ImageNet policy) and RandAugment include shearing, translation and rotation, which could potentially boost the transfer performance for SVHN

Ablation Studies

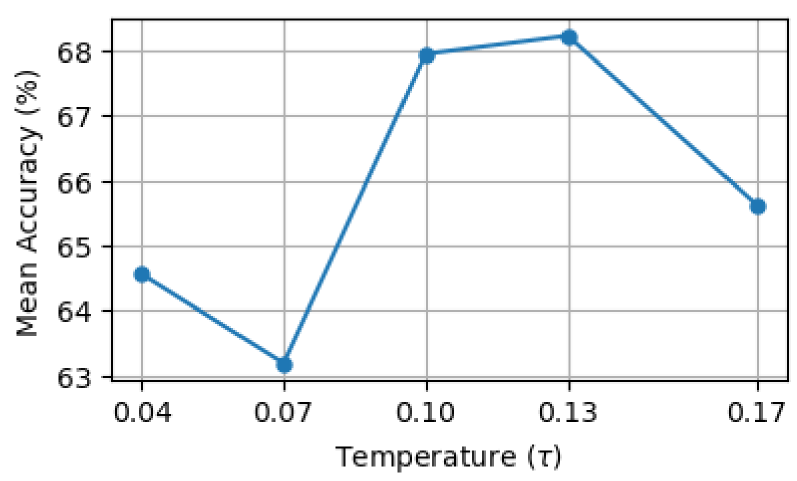


Figure 1: Plot of Mean Accuracy over all the Downstream Datasets against Temperature

- The temperature (τ) parameter used in the SupCon loss is adjustable and smaller τ values can benefit training more, but a very small value of τ can lead to numerical instability
- At lower temperature values of 0.04 and 0.07, the mean accuracy was lower than that of higher temperature values like 0.10 and 0.13. However, it is also observed that when temperature is increased further to 0.17, there is a drop in the mean accuracy
- Important to select an optimal temperature value that can benefit the training process so that the representations learned can give better transfer performance

Conclusion

- I empirically showed that supervised contrastive learning can give better transfer performance than cross entropy loss when trained on the multi-domain DomainNet dataset
- Representations learned from supervised contrastive learning could perhaps be more robust and capture more domain invariant features that are more transferable to downstream datasets across different domains