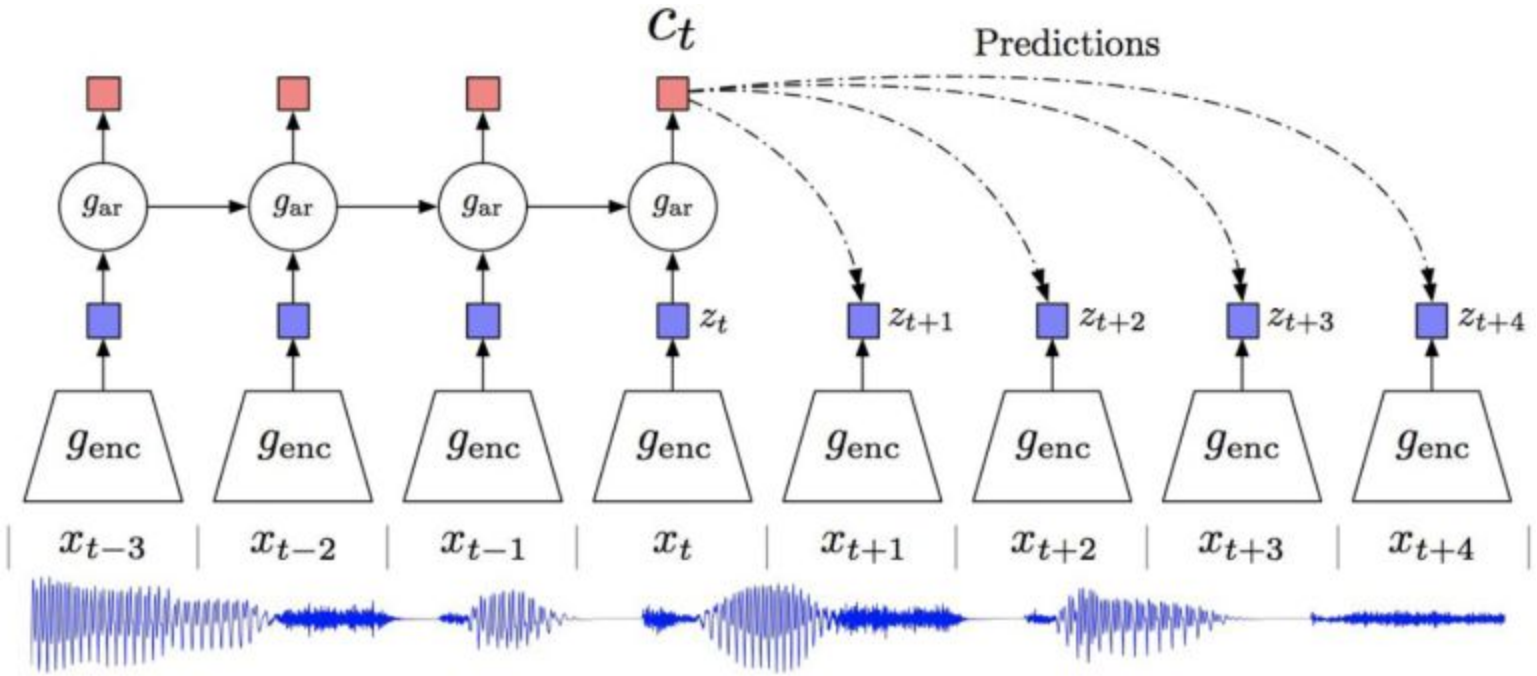
Deep Metric Learning

University of Victoria - PHYS-555

Contrastive Predictive Coding



v.d. Oord et al (2018)

What is Contrastive Learning?

Contrastive learning is a learning paradigm where we want to learn *distinctiveness.*

- *What makes two objects similar or different?*
- *When I train a network for some task, say classification, I am already forcing my network to learn discriminative features, right?*

Sometimes high-level features alone aren't enough to learn good representations, especially when *semantics* come into play.

Features like shape and color of the tail of a whale aren't enough to uniquely identify its species because the semantics for the tails of all whales are very similar.

From: Blog on Medium

Learning similarity between samples with a distance

Goal: build a function $d_\theta(\mathbf{x}_1, \mathbf{x}_2)$ to quantify how “similar” two sample of data are

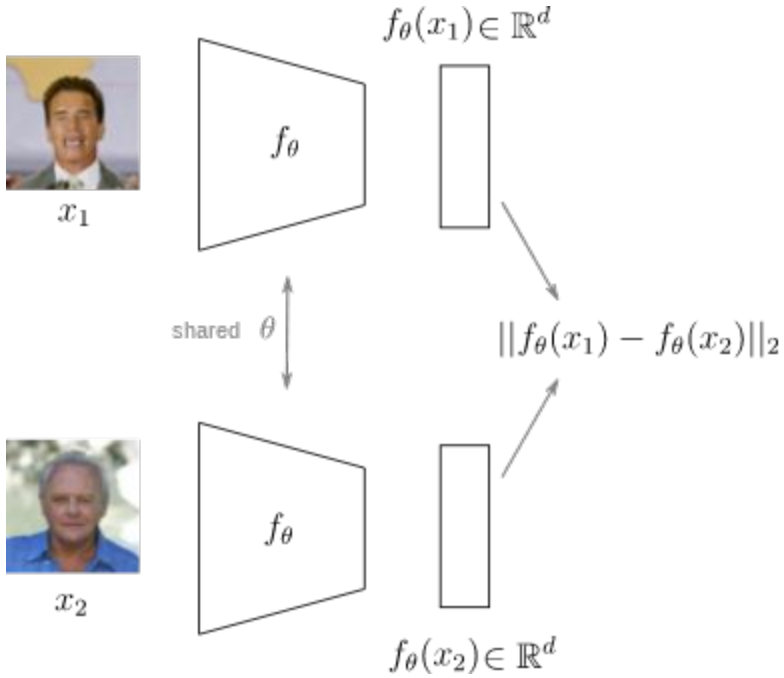
Example: a Euclidean distance between two representations of a NN f_θ

$$d_\theta(\mathbf{x}_1, \mathbf{x}_2) = \|f_\theta(\mathbf{x}_1) - f_\theta(\mathbf{x}_2)\|_2$$

Siamese Networks

- Can be used for classification: define a threshold *T* to decide when two samples belong to the same class:

$$d_\theta(\mathbf{x}_1, \mathbf{x}_2) < T$$



Example: SimCLR

SimCLR (Chen et al, 2020) proposed a simple framework for contrastive learning of visual representations. It learns representations for visual inputs by maximizing agreement between differently augmented views of the same sample via a contrastive loss in the latent space.

SimCLR works in the following three steps:

1. Randomly sample a mini-batch of N samples and each sample is applied with two different data augmentation operations, resulting in 2N augmented samples in total.

2. Given one positive pair, other 2(N-1) data points are treated as negative samples. The representation is produced by a base encoder NN

3. The contrastive loss is defined using cosine similarity. The loss operates on top of an extra projection of the representation rather than on the representation from the latent space directly. But only the representation h is used for downstream tasks.

How to generate pairs of similar data points?

A supervised approach would be to manually label them as similar or not


Similar


Dissimilar

Generation of similar images

Random Data Augmentation

Dataset

Framework

Encoder

Encoder

$$\begin{bmatrix} h_{0_0} \\ h_{0_1} \\ \vdots \\ h_{0_n} \end{bmatrix}$$

$$\begin{bmatrix} h_{1_0} \\ h_{1_1} \\ \vdots \\ h_{1_n} \end{bmatrix}$$

Projection Head

Projection Head

$$\begin{bmatrix} z_{0_0} \\ z_{0_1} \\ \vdots \\ z_{0_n} \end{bmatrix}$$

$$\begin{bmatrix} z_{1_0} \\ z_{1_1} \\ \vdots \\ z_{1_n} \end{bmatrix}$$

Compare Similarity of the two embeddings!

Resnet50 used as Encoder

Two layer MLP used to get embedding

Loss Calculation

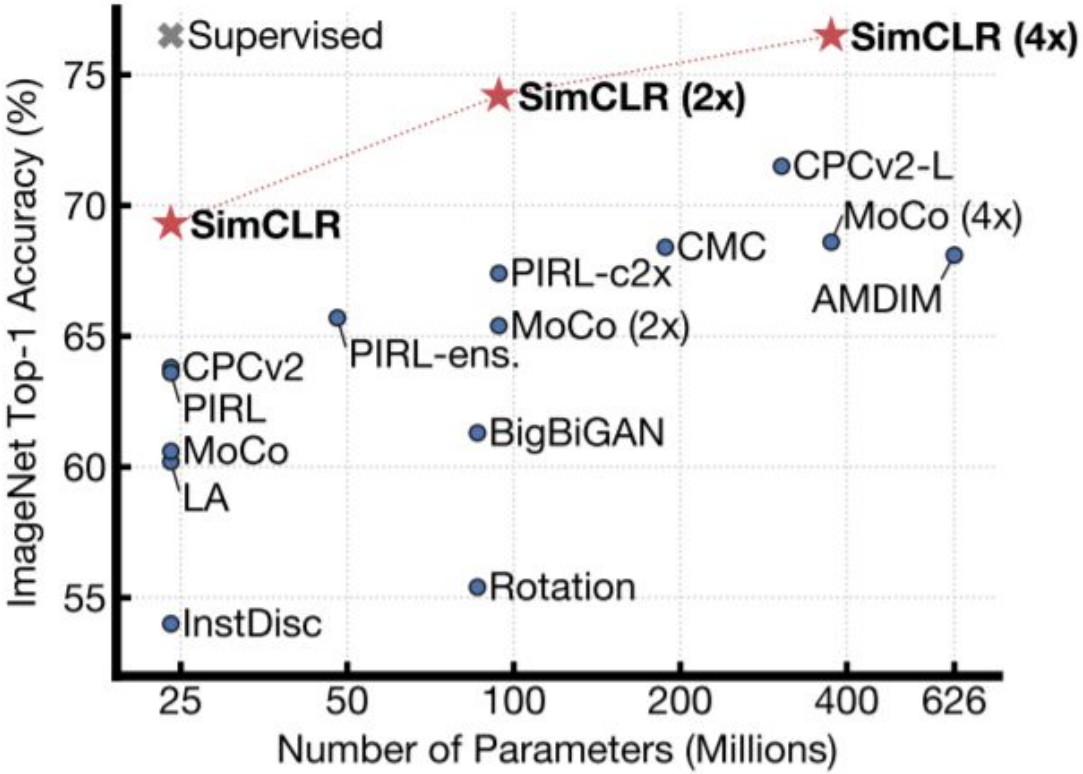
For each data pair (embeddings z):

Compute Pairwise Similarity $$s_{i,j} = \frac{z_i^T z_j}{\tau \ || \ z_i \ || \ || \ z_j \ ||}$$

Compute loss $$l_{i,j} = -log \frac{exp(s_{i,j})}{\sum_{k=1,k!=i}^{2N} exp(s_{i,k})}$$

Batch wise loss for positive pairs $$L = \frac{1}{2N} \sum_{k=1}^{2N} [l(2k-1, 2k) + l(2k, 2k-1)]$$

SimCLR Classification Results on ImageNet



Example on Galaxies 3-band images

UMAP of the SimCLR representation

[Hayat et al. 2021]