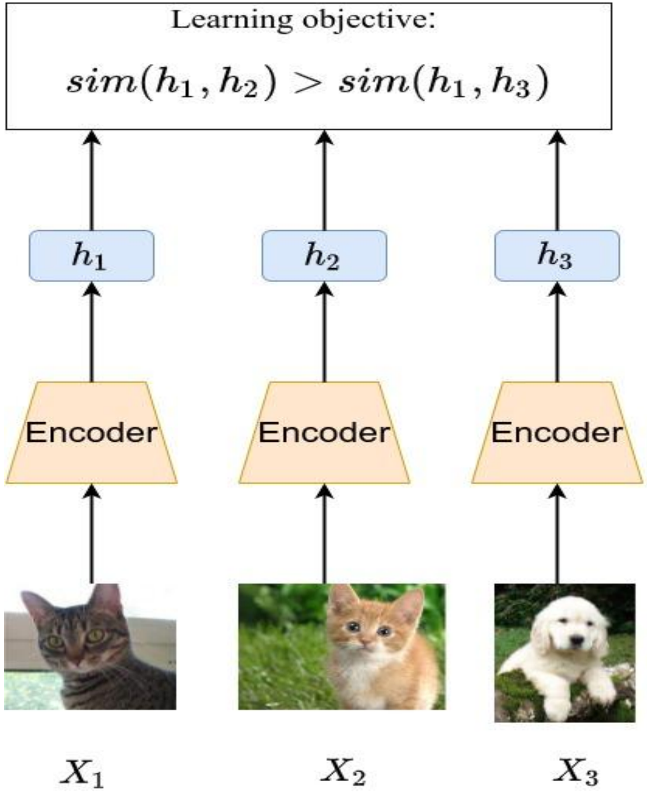
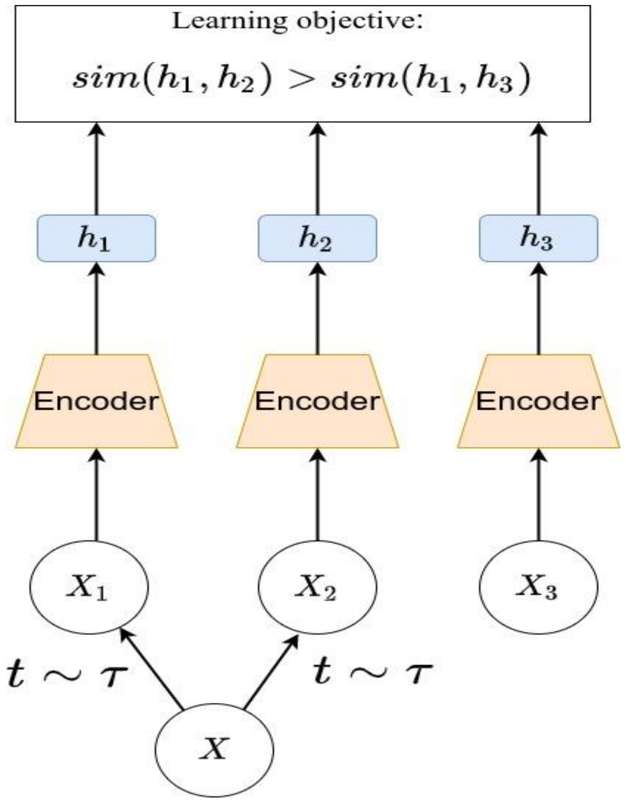
Towards Domain-Agnostic Contrastive Learning

Vikas Verma, Minh-Thang Luong, Kenji Kawaguchi, Hieu Pham, Quoc V. Le

Contrastive Learning



Contrastive Learning



Limitations

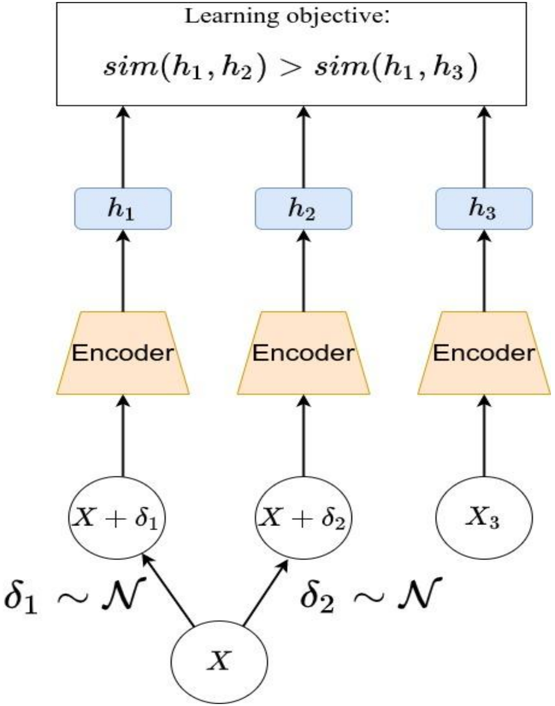
Requires “semantic preserving” augmentations.

Limitations

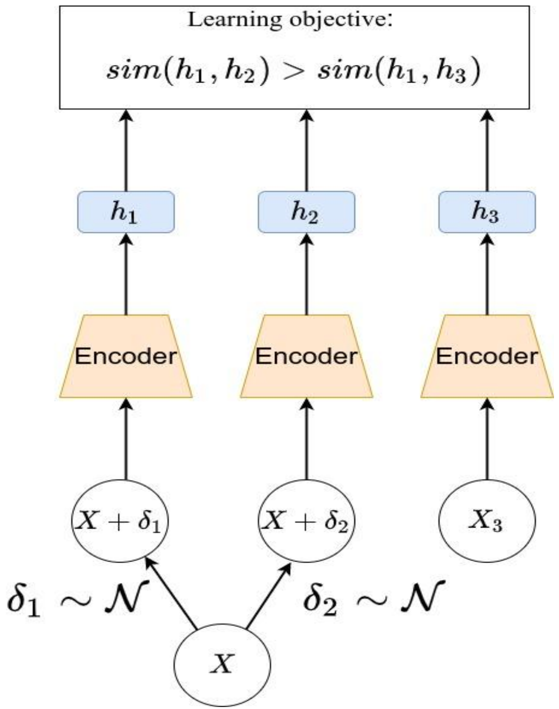
Requires “semantic preserving” augmentations.

Might not be available for some application domains such as tabular data, graph-structured data.

Gaussian Noise



Gaussian Noise



Let us consider X to be an 2D image.

In this case, to maximize the similarity between positive pair, the network can learn just to take an average over the neighboring pixels to remove the noise.

Our Method

A network is forced to learn better features if the noise captures the structure of the data manifold rather than being independent of it.

Our Method

A network is forced to learn better features if the noise captures the structure of the data manifold rather than being independent of it.

The simplest form to construct positive pair such that it captures the structure of data manifold is to add a small noise to a sample in the direction of other sample.

Our Method

A network is forced to learn better features if the noise captures the structure of the data manifold rather than being independent of it.

The simplest form to construct positive pair such that it captures the structure of data manifold is to add a small noise to a sample in the direction of other sample.

Positive sample w.r.t. X : $X^{+} = \lambda X + (1 - \lambda)\tilde{X}$

where $\lambda \sim U(\alpha, 1.0)$

Our Method

A network is forced to learn better features if the noise captures the structure of the data manifold rather than being independent of it.

The simplest form to construct positive pair such that it captures the structure of data manifold is to add a small noise to a sample in the direction of other sample.

Positive sample w.r.t. X : $X^{+} = \lambda X + (1-\lambda)\tilde{X}$

where $\lambda \sim U(\alpha, 1.0)$

For, samples with irregular topology: $h^{+} = \lambda h + (1-\lambda)\tilde{h}$

Our Method

A network is forced to learn better features if the noise captures the structure of the data manifold rather than being independent of it.

The simplest form to construct positive pair such that it captures the structure of data manifold is to add a small noise to a sample in the direction of other sample.

Positive sample w.r.t. X : $X^+ = \lambda X + (1 - \lambda)\tilde{X}$

where $\lambda \sim U(\alpha, 1.0)$

For, samples with irregular topology: $h^+ = \lambda h + (1 - \lambda)\tilde{h}$

Domain-Agnostic Contrastive Learning (DACL)

Our Method

Other form of Mixup-based Noise:

Our Method

Other form of Mixup-based Noise:

Geometric-Mixup : $X^{+} = X^{\lambda} \odot \tilde{X}^{(1-\lambda)}$

Our Method

Other form of Mixup-based Noise:

Geometric-Mixup : $X^+ = X^\lambda \odot \tilde{X}^{(1-\lambda)}$

Binary-Mixup : $X^+ = X \odot M + \tilde{X} \odot (1 - M)$

Results

- Tabular Data, Image Data, Graph-Structured Data

Results

- Tabular Data, Image Data, Graph-Structured Data

- For tabular and graph structured datasets, DACL and DACL+ improves upon baselines.

Results

- Tabular Data, Image Data, Graph-Structured Data

- For tabular and graph structured datasets, DACL and DACL+ improves upon baselines.

- For images DACL falls short off the sota methods such as SimCLR, however SimCLR+DACL can improve the performance of SimCLR alone.

Theoretical Analysis

Properties of **Gaussian-Noise** and **Mixup-Noise** based Contrastive Learning in a **binary classification task**.

Contrastive loss $\approx$ Modified classification loss+ error term

We then prove that Mixup-Noise induces a better regularization effect when compared to Gaussian Noise

Thank You for your attention!