Unsupervised Contrastive Representation Learning

Kelsey Ball

The University of Texas at Austin

September 24, 2022

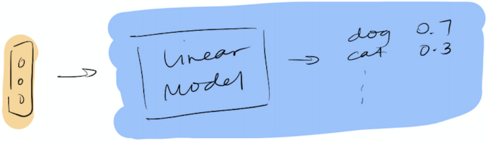
Overview

- Motivation
- Framework for Vision
- Adaptations to NLP

Motivation

We want an image encoder f:

whose image representations can be used to train a simple, linear classifier:



We can evaluate f using the **ImageNet Linear Benchmark**: (1) train a linear classifier on representations from f (2) evaluate on ImageNet.

Motivation

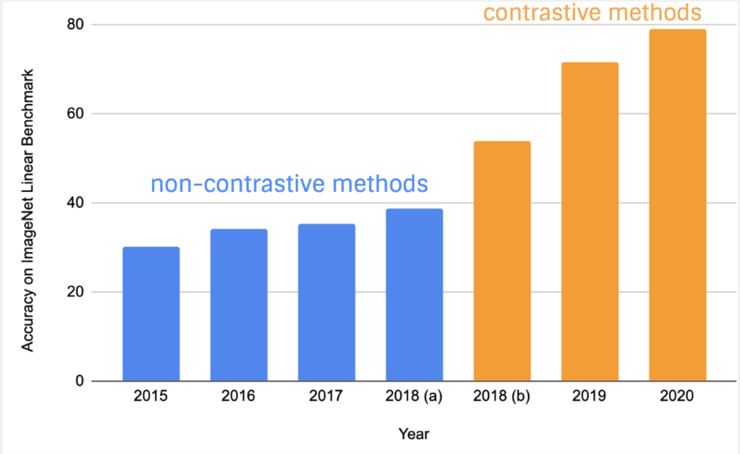
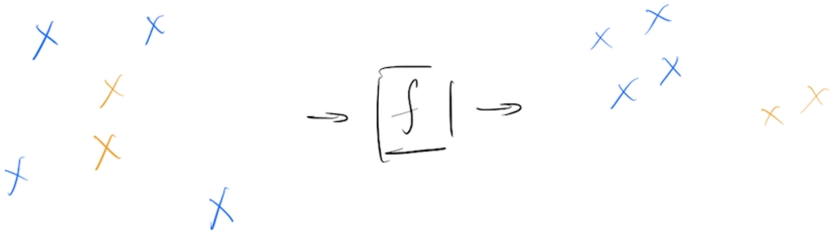
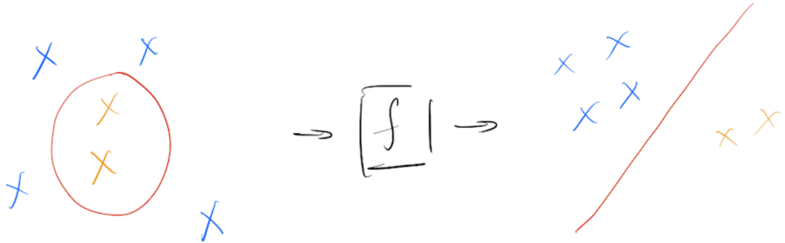


Figure: Accuracy of contrastive vs. non-contrastive methods on ImageNet Linear Benchmark.

Contrastive Learning: Goal



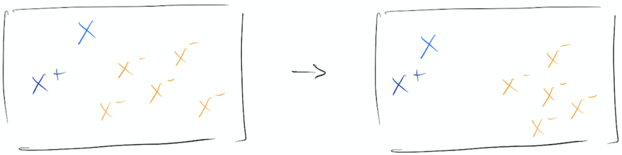
Contrastive Learning: Goal



Contrastive Learning: Approach

Contrastive loss with three inputs:

1. anchor example
2. positive example (similar)
3. negative examples (dissimilar)



$$\text{loss} = -\log \frac{\exp(\mathrm{sim}(x, x^+))}{\exp(\mathrm{sim}(x, x^+) + \sum_{i=1}^{k} \exp(\mathrm{sim}(x, x_i^-))}$$

Data Augmentation

How to identify similar (positive) examples without knowledge of underlying class?

Apply meaning-preserving transformation to anchor example.
For images, e.g.:

- Random cropping
- Discoloration
- Blur

The result is an "augmentation" of the original image.

Image Contrastive Learning

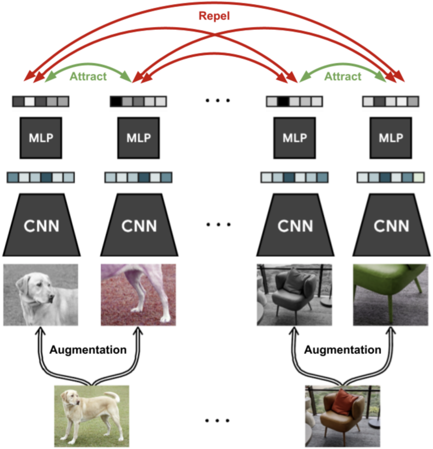


Figure: A Simple Framework for Contrastive Learning of Visual Representations [CKNH20]

Does This Work For Language?

- Can we adapt the data augmentation approach to language?
- Close in surface form, far in semantics:

I like cake

I don't like cake

Alice likes Bob

Bob likes Alice

- Unlike images, there are no obvious, semantic-preserving transformations for text...
- ...or are there??

Does This Work For Language?

- Can we adapt the data augmentation approach to language?
- Close in surface form, far in semantics:

I like cake
I don't like cake

Alice likes Bob
Bob likes Alice

- Unlike images, there are no obvious, semantic-preserving transformations for text...
- ...or are there??

Does This Work For Language?

- Can we adapt the data augmentation approach to language?
- Close in surface form, far in semantics:

I like cake
I don't like cake

Alice likes Bob
Bob likes Alice

- Unlike images, there are no obvious, semantic-preserving transformations for text...
- ...or are there??

Does This Work For Language?

- Can we adapt the data augmentation approach to language?
- Close in surface form, far in semantics:

I like cake
I don't like cake

Alice likes Bob
Bob likes Alice

- Unlike images, there are no obvious, semantic-preserving transformations for text...
- ...or are there??

QuickThought [Bra18]

Main Idea: Consecutive sentences are positive pairs, non-consecutive sentences are negative pairs

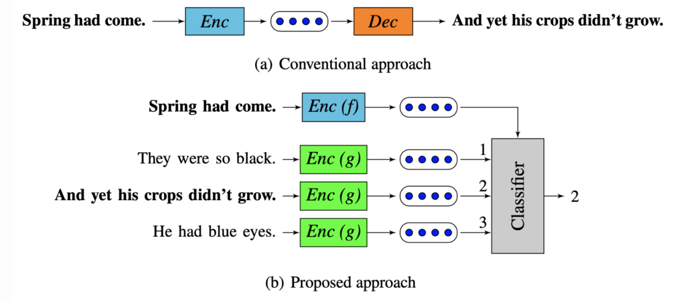


Figure: Maximize the probability of identifying context (preceding) sentence for each sentence in the training data. At inference, representations from both encoders are concatenated.

Declutr [GNBW20]

Main Idea:

- Generalize QuickThought beyond sentence boundary
- Nearby spans are positive pairs, distant spans are negative pairs

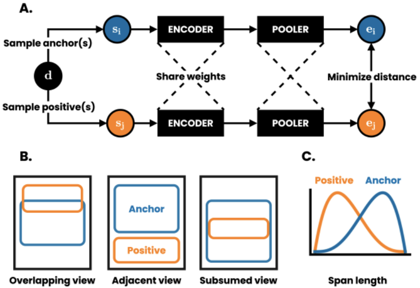


Figure: (1) Sample an anchor span (2) Sample a shorter, nearby positive span (3) Compute contrastive loss, using other examples as negatives

CLEAR [WWG+20]

Main Idea:

- Apply random edits to sentence to get augmented sentence
- Edits include: deletion, substitution, and re-ordering

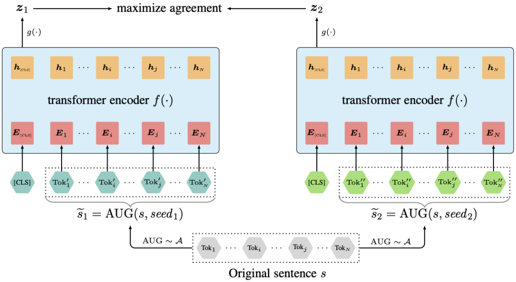


Figure: (1) Sample two augmentations from deletion, synonym-substitution, re-ordering (2) Apply to input sentence to derive two augmentations (3) Compute contrastive loss

CERT [FWZ+20]

Main Idea:

- Paraphrases are positive examples, derived using back-translation
- Back-translation: translate sentence to another language then back to get a paraphrase of the original sentence

Figure: (1) Derive positives using back-translation with different target languages (2) Encode positives using language encoder (3) Compute contrastive loss

Summary

- Unsupervised contrastive representation learning has a simple and mature framework for image learning [CKNH20]
- Equivalent techniques are still being explored for sentence representation learning

Reference I

[Bra18] Siddhartha Brahma. Unsupervised learning of sentence representations using sequence consistency. arXiv preprint arXiv:1808.04217, 2018.

[CKNH20] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In International conference on machine learning, pages 1597–1607. PMLR, 2020.

[FWZ+20] Hongchao Fang, Sicheng Wang, Meng Zhou, Jiayuan Ding, and Pengtao Xie. Cert: Contrastive self-supervised learning for language understanding. arXiv preprint arXiv:2005.12766, 2020.

[GNBW20] John M Giorgi, Osvald Nitski, Gary D Bader, and Bo Wang. Declutr: Deep contrastive learning for unsupervised textual representations. arXiv preprint arXiv:2006.03659, 2020.

[WWG+20] Zhuofeng Wu, Sinong Wang, Jiatao Gu, Madian Khabsa, Fei Sun, and Hao Ma. Clear: Contrastive learning for sentence representation. arXiv preprint arXiv:2012.15466, 2020.