The Families and Origins of Self-Supervised Learning

Shengjie Niu

Department of Statistics and Data Science, SUSTech

July 13, 2023

Self-Supervised Learning

Main Reference

R Balestriero et al, A Cookbook of Self-Supervised Learning

Background

- Dubbed "The dark matter of intelligence".
- Self-Supervised approaches can learn from vast unlabeled data.
- NLP: advances from automated machine translation to large language models trained on web-scale corpora of unlabeled text.
- CV: push new bounds on data size & match or in some cases surpass models trained on labeled data.
- Vast applicability in video, audio, and time series.

Mechanism

- Define a **pretext task** based on unlabeled inputs to produce descriptive and intelligent representation (typically no need for manual labels)
- NLP: mask a word in the text and predict the surrounding words
  - to capture relationships among words in the text
  - without any labels & vast applicability in downstream tasks
- CV: analogous objective
  - predict masked patches of an image or representation
  - encourage two view of the same image to be mapped to similar representations

SSL V.S. SSL

Self-Supervised Learning

- Create and solve a pretext task to learn a robust representation of the data
- Then be used for a variety of downstream tasks, often with additional fine-tuning.

Semi-Supervised Learning

- Combination of unsupervised learning and supervised learning
- Utilize the unlabeled data to improve the model performance. (understand the underlying structure of unlabeled data)

The Origins of SSL

Time Point

- SSL methods have enjoyed a renaissance since 2020
- Availability of extremely large datasets and high-memory GPUs

Why Origins

- Contemporary methods build upon the knowledge we gained from early experiments.
- Ideas from previous papers form the foundation for many of the modern methods.

The Origins of SSL

Early progress in SSL focused on the development of methods that fell into the following (sometimes overlapping) categories:

- Information restoration
- Learning spatial context
- Grouping similar images together
- Multi-view invariance
- etc.

The Origins of SSL

Information restoration

- Mask or remove something from an image
- Convert image to grayscale - objective segmentation

Learning spatial context

- Trains a model to understand the relative positions and orientations of objects within a scene
- RotNet: random rotation and then asks the model to predict the rotation
- jigsaw method: break an image into an array of disjoint patches and predict the relative location of each
- Train to output the number of objects in an image

The Origins of SSL

Grouping similar images together

- Learn rich features by grouping semantically similar images together.

Multi-view invariance

- Utilize contrastive learning to create feature representations that are invariant to simple transforms.
- Create feature representations that are invariant to simple transforms.
- Later works focused on maintaining invariance to data augmentation transforms.

Families of SSL

With these origins, SSL can be categoried into four broad families:

- Self-Distillation Family
- Deep Metric Learning Family
- Canonical Correlation Analysis Family
- Masked Image Model Family