

SUMMARY ON SSL IN AI AND MY LEARNING EXPERIENCE

Self-Supervised Learning (SSL): An Introduction and My Learning Experience

What is Self-Supervised Learning?

Self-Supervised Learning (SSL) is a machine learning approach where models learn from unlabeled data by creating their own supervision signals. Unlike traditional supervised learning (which requires expensive labeled datasets), SSL relies on *pretext tasks* that automatically generate labels from the data itself. This approach allows models to extract rich, general-purpose representations.

SSL has seen remarkable success in AI, particularly in NLP (BERT, GPT) and computer vision (SimCLR, MoCo, DINO). It is considered a key step toward truly scalable, human-like learning.

Why is SSL Important?

- Reduces the cost of annotation.
 - Leverages large-scale unlabeled data.
 - Learns general representations that transfer to many downstream tasks.
 - Drives cutting-edge results in many AI benchmarks.
-

Example Pretext Tasks:

- Predicting missing pixels (inpainting)
 - Image colorization (grayscale → color)
 - Rotation prediction
 - Contrastive learning to group similar representations
-

My Mini-Implementation: Image Colorization

As a practical exercise, I implemented a simple SSL task: Image Colorization using PyTorch. The idea is to take grayscale images as input and predict their original color version. This task is self-supervised because the grayscale input can be created automatically from any color image, without human labeling.

I used the CIFAR-10 dataset, converted images to grayscale, and trained a small CNN to predict color versions. The model minimized MSE loss between predicted and true color

images. Even with a simple architecture, the network learned basic colorization, showing how self-supervision can extract meaningful visual features without labeled data.

My Learning Experience

Through this project, I learned that:

- SSL is a powerful way to utilize unlabeled data.
- Defining a good pretext task is crucial.
- Even simple models can learn meaningful representations with self-supervised signals.
- SSL bridges the gap between unsupervised and supervised learning, making it a key enabler for scalable AI.

I believe SSL is a promising direction for future AI systems, offering both practical efficiency and strong performance.