KG Priors for Self Supervised Learning

Navodita Mathur

Intro

Self-Supervised Learning (SSL) underpins modern foundation models.

Common paradigms:

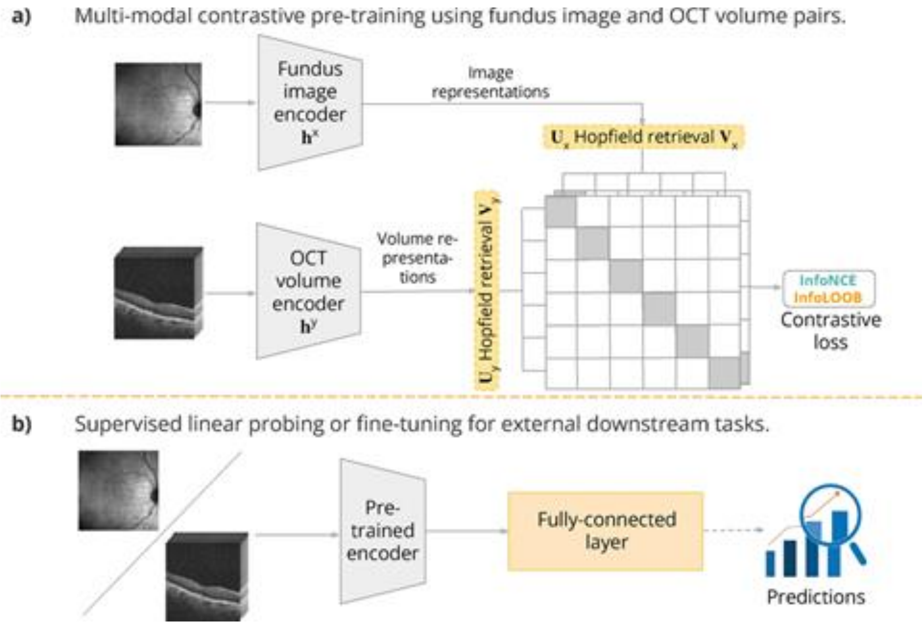
- Contrastive learning
- Masked modeling
- Generative objectives
- Distillation

Why contrastive learning?

- Brings modalities together via semantic alignment
- Label-efficient and scalable
- Strong zero-shot transfer

Examples in science

- MedCLIP
- BioCLIP

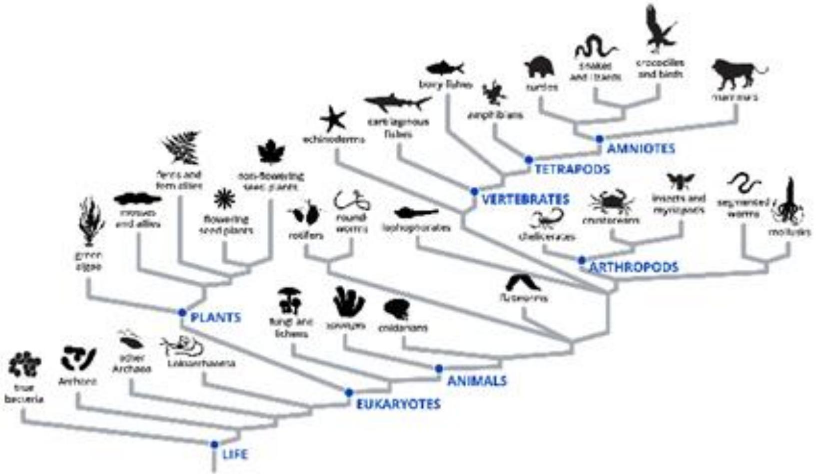


Contrastive Learning

Contrastive learning aligns based on co-occurrence.

This leads to problems in scientific domains:

- Can learn spurious correlations
- Treats all negatives as equally unrelated
- Operates in a flat similarity space



Weighted Contrastive Learning

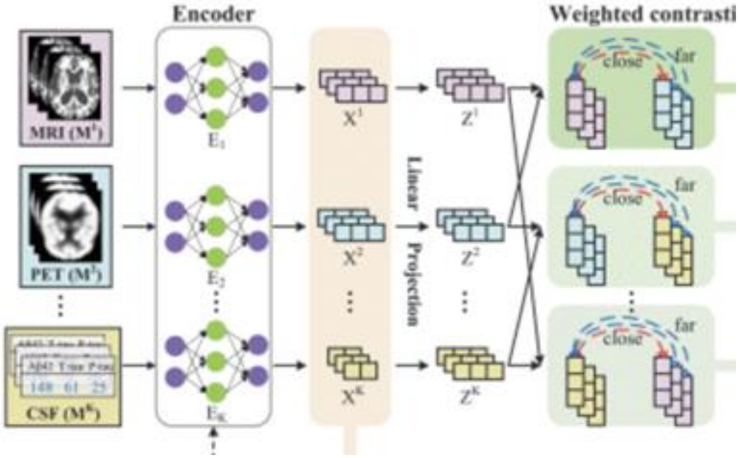
Idea: Not all pairs should contribute equally to the contrastive loss. Weights encode prior knowledge about relationships.

Example: Two cats from different species should be closer than a cat and a dog

Weighted contrastive loss

- Softens negatives
- Preserves graded similarity
- Improves representation geometry

But there's a problem…



Problem of relations

Strength cannot model the kind of relationship. In Science, there are many kinds: causal, hierarchical, spatial, temporal etc. A single weight can encode: how strong a relationship is. But cannot encode:

- causal vs hierarchical
- spatial vs temporal
- functional vs ecological

In science, relationship type matters. Scalar similarity is too weak to represent scientific structure. Enter Knowledge Graphs

Knowledge graph Embeddings

KG embeddings turn typed relations into geometry.

TransE

- translational, directional
- good for hierarchy & causal chains

RotE

- rotational structure
- good for temporal & compositional relations

ComplEx

- asymmetric, many-to-many
- good for interaction networks

GNNs

- powerful but expensive
- often unnecessary if geometry suffices

KG embeddings encode how entities are related and not just how much.

Key Insight

Contrastive learning defines similarity. KG embeddings define relational structure

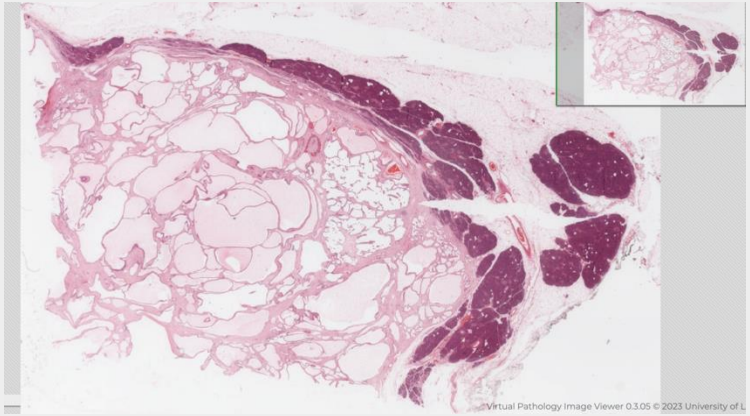
By deriving weights from KG geometry, we can:

- preserve hierarchy
- respect causal direction
- avoid collapsing incompatible entities
- reduce spurious correlations

All without modifying model architecture.

Case Study

Hispathology



Histopathology - Aligning pathology images and spatial transcriptomics

Problem

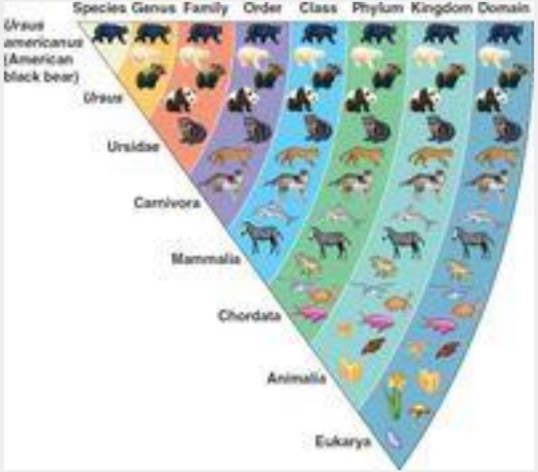
- Weak supervision
- Mixed tissue types
- Spurious correlations (stain, scanner, background)

How KG priors help

- Preserve phenotype–pathway–tissue relations
- Prevent semantic collapse in embeddings
- Improve downstream tasks

Case Study

Biodiversity



Biodiversity - Classification

Problem

- Long-tailed species
- Strong geographic and ecological confounders
- Taxonomic hierarchy ignored by contrastive SSL

How KG priors help

- Encode phylogenetic distance
- Preserve taxonomic structure
- Improve few-shot and zero-shot generalization

Thanks!