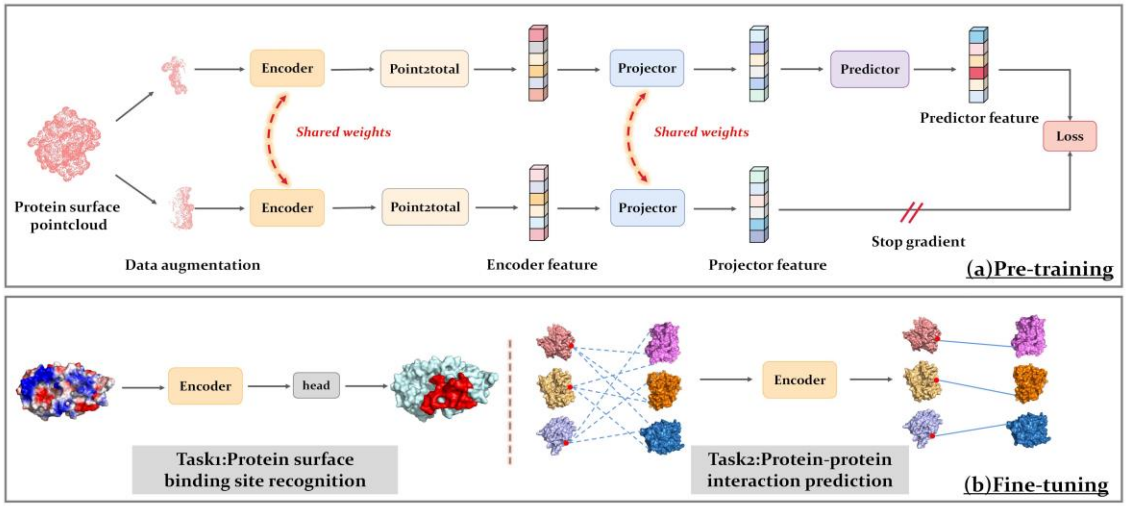
SS-Pro: A Simplified Siamese Contrastive Learning Approach for Protein Surface Representation

Ao SHEN, Mingzhi YUAN, Yingfan MA, Manning WANG

Frontiers of Computer Science, DOI: 10.1007/s11704-024-3806-9

Problems & Ideas

- Problems of conventional stereo matching approaches:
  - Data-driven deep learning methods in protein surface representation face challenges of label scarcity.
  - Current deep learning protein representation methods are not suitable for self-supervised learning.
- Ideas: We propose a contrastive learning framework can be adapted to various point cloud feature extraction backbone networks, demonstrating strong generalization capabilities.



Pre-training and fine-tuning of the SS-Pro framework. (a) Contrastive pre-training using protein surface point cloud. (b) Fine-tuning on downstream tasks using the pre-trained encoder.

Main Contributions

- Contributions:
  - We design an efficient and straightforward contrastive learning framework called SS-Pro for protein surface representation learning, which is versatile and can be adapted to various point cloud feature extraction backbone networks, demonstrating strong generalization capabilities.
  - We apply our method to four point cloud feature extraction backbone networks and conduct experiments on a fine-grained task, protein surface binding site recognition, and a coarse-grained task, protein-protein interaction prediction. The proposed method shows improved performance across multiple backbone networks.

	Transformer	DGCNN	PointNet++	dMaSIF
From scratch	0.859	0.777	0.787	0.862
Pre-train	0.867	**0.800**	**0.798**	**0.866**
ProteinMAE	**0.871**	-	-	-

	Transformer	DGCNN	PointNet++	dMaSIF
From scratch	0.940	0.976	0.962	0.866
Pre-train	**0.950**	**0.985**	**0.974**	**0.871**
ProteinMAE	0.948	-	-	-

Protein surface binding site recognition results (upper) and protein-protein interaction prediction results (lower).