IEEE Access, vol.14, pp.40311-40327, December, 2026.

Raman Spectroscopy Pre-Trained Encoder: A Self-Supervised Learning Approach For Data-Efficient Domain-Independent Spectroscopy Analysis

A. Eranti; Y. Tewari; R. Palacios Hielscher; A. Gupta

Abstract-

Deep-learning methods have boosted the analytical power of Raman spectroscopy, yet they still require large, task-specific, labeled datasets and often fail to transfer across application domains. The study explores pre-trained encoders as a solution. Pre-trained encoders have significantly impacted Natural Language Processing and Computer Vision with their ability to learn transferable representations that can be applied to a variety of datasets, significantly reducing the amount of time and data required to create capable models. The following work puts forward a new approach that applies these benefits to Raman Spectroscopy. The proposed approach, RSPTE (Raman Spectroscopy Pre-Trained Encoder), is designed to learn generalizable spectral representations without labels. RSPTE employs a novel domain adaptation strategy using unsupervised Barlow Twins decorrelation objectives to learn fundamental spectral patterns from multi-domain Raman Spectroscopy datasets containing samples from medicine, biology, and mineralogy. Transferability is demonstrated through evaluation on several models created by fine-tuning RSPTE for different application domains: Medicine (detection of Melanoma and COVID), Biology (Pathogen Identification), and Agriculture. As an example, using only 20% of the dataset, models trained with RSPTE achieve accuracies ranging 50%–86% (depending on the dataset used) while without RSPTE the range is 9%–57%. Using the full dataset, accuracies with RSPTE range 81%–97%, and without pretraining 51%–97%. Current methods and state-of-the-art models in Raman Spectroscopy are compared to RSPTE for context, and RSPTE exhibits competitive results, especially with less data as well. These results provide evidence that the proposed RSPTE model can effectively learn and transfer generalizable spectral features across different domains, achieving accurate results with less data in less time (both data collection time and training time).

Index Terms- Raman Spectroscopy, self-supervised learning, pre-trained encoder, multi-domain data, clinical diagnostics

Due to copyright restriction we cannot distribute this content on the web. However, clicking on the next link, authors will be able to distribute to you the full version of the paper:
Request full paper to the authors

If you institution has a electronic subscription to IEEE Access, you can download the paper from the journal website:
Access to the Journal website

Citation:
Eranti, A.; Tewari, Y.; Palacios, R.; Gupta, A. "Raman Spectroscopy Pre-Trained Encoder: A Self-Supervised Learning Approach For Data-Efficient Domain-Independent Spectroscopy Analysis", IEEE Access, vol.14, pp.40311-40327, December, 2026.