Demixing latent representations in generative self-supervised learning

Context: Self-supervised learning (SSL) has been one of the main pillars of recent progress in AI and in the so-called foundation models, allowing algorithms to leverage vast amounts of unlabeled data (text, image, video, remote sensing, etc.) to learn representations that can later be adapted to different predictive (downstream) tasks [LZH+21, HCX+22]. However, despite the extent of the practical impact of such approaches, theoretically understanding how such approaches work, especially in challenging settings such as with small or heterogeneous datasets, is still difficult.

A special problem comes with interpreting or explaining representations learned from unsupervised or SSL, which is related to understanding under which conditions one can obtain *disentangled* representations [WCW+24]. Beyond the theoretical understanding of such approaches, this is also a key question in many applications, as it can be used in explaining the results of downstream (e.g., classification) tasks. Moreover, many problems directly involve directly interpreting the learned representations as they can be linked to physical variables of interest, such as in spectral unmixing [BDPD+12].

Despite its importance, unsupervised disentanglement is a difficult subject and impossible without additional assumptions on the model or data [LBL+19]. Recent advances in identifiable variational autoencoders (VAEs) (e.g., using side information) [KKMH20, LF22] provide a rigorous theoretical and empirical framework to tackle this problem by exploiting *side information* (i.e., exogenous variables), offering both strong theoretical guarantees and practical algorithms.

Goals: This goal of this internship is to explore identifiable VAEs in the context of source separation problems, targeting the application to hyperspectral unmixing, which consists in separating the spectrum of an image pixel measured over multiple wavelengths into the spectra of its constituent materials and their contributing proportions [BDPD+12].

A first goal will be to investigate the link between exogenous variables used in the formulation [KKMH20] and related frameworks based on contractive self-supervised learning as a pretext task [LF22]. A second goal will be to develop a SSL approach for hyperspectral unmixing within this framework, exploring how the hypothesis of the unmixing problem can be formally related to the identifiable SSL framework. Specific questions will be theoretically principled approaches to generate the exogenous variables or pretext tasks for training the model, and benefiting from the mathematical background of such approaches to provide identifiability guarantees to the unmixing results.

Work conditions: The internship will take place in the Simul Research Group (`https://cran-simul.github.io`) of Centre de Recherche en Automatique de Nancy. International scientific collaborations with the United States can also be planned.

Applications (cv +motivation letter+references) should be sent to

R. Borsoi : ricardo.borsoi@univ-lorraine.fr
M. Clausel : marianne.clause@univ-lorraine.fr
K. Usevich : konstantin.usevich@univ-lorraine.fr

References

[BDPD+12] José M Bioucas-Dias, Antonio Plaza, Nicolas Dobigeon, Mario Parente, Qian Du, Paul Gader, and Jocelyn Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE journal of selected topics in applied earth observations and remote sensing*, 5(2):354–379, 2012.

[HCX+22] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022.

[KKMH20] Ilyes Khemakhem, Diederik Kingma, Ricardo Monti, and Aapo Hyvarinen. Variational autoencoders and nonlinear ICA: A unifying framework. In *International conference on artificial intelligence and statistics*, pages 2207–2217. PMLR, 2020.

[LBL+19] Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Raetsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. Challenging common assumptions in the unsupervised learning of disentangled representations. In *international conference on machine learning*, pages 4114–4124. PMLR, 2019.

[LF22] Qi Lyu and Xiao Fu. On finite-sample identifiability of contrastive learning-based nonlinear independent component analysis. In *International Conference on Machine Learning*, pages 14582–14600. PMLR, 2022.

[LZH+21] Xiao Liu, Fanjin Zhang, Zhenyu Hou, Li Mian, Zhaoyu Wang, Jing Zhang, and Jie Tang. Self-supervised learning: Generative or contrastive. *IEEE transactions on knowledge and data engineering*, 35(1):857–876, 2021.

[WCW+24] Xin Wang, Hong Chen, Zihao Wu, Wenwu Zhu, et al. Disentangled representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.