Contrastive learning: Big Data Foundations and Applications

Sandhya Tripathi
sandhyat@wustl.edu
Department of Anesthesiology, Washington University
School of Medicine in St Louis
St Louis, Missouri, USA

Christopher R. King
christopherking@wustl.edu
Department of Anesthesiology, Washington University
School of Medicine in St Louis
St Louis, Missouri, USA

ABSTRACT

Contrastive learning (CL) has exploded in popularity due to its ability to learn effective representations using vast quantities of unlabelled data across multiple domains. CL underlies some of the most impressive applications of generative AI for the general public. We will review the fundamentals and applied work on contrastive learning representations focusing on three main topics: 1) CL in supervised, unsupervised and self-supervised setup and its revival in AI research as an instance discriminator. In this part, we will focus on learning about the nuts and bolts, such as different augmentation techniques, loss functions, performance evaluation metrics, and some theoretical understanding of contrastive loss. We will also present the methods supporting DALL·E 2, a popular generative AI. 2) Learning contrastive representations across vision, text, time series, tabular data and knowledge graph modalities. Specifically, we will present the literature representative of solution approaches regarding new augmentation techniques, modification in the loss function, and additional information. The first two parts will also have small hands-on session on the application shown and some of the methods learned. 3) Discussing the various theoretical and empirical claims for CL's success, including the role of negative examples. We will also present some work that challenges the shared information assumption of CL and propose alternative explanations. Finally, we will conclude with some future directions and applications for CL.

CCS CONCEPTS

• **Computing methodologies → Unsupervised learning**; **Knowledge representation and reasoning**.

KEYWORDS

contrastive learning, multi-view, multi-modal, augmentations, time-series, graphs, tabular datasets, distillation, clustering, noise estimation loss

ACM Reference Format:
Sandhya Tripathi and Christopher R. King. 2024. Contrastive learning: Big Data Foundations and Applications. In *7th Joint International Conference on Data Science & Management of Data (11th ACM IKDD CODS and 29th COMAD) (CODS-COMAD 2024), January 04–07, 2024, Bangalore, India.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3632410.3633291

This work is licensed under a Creative Commons Attribution International 4.0 License.

CODS-COMAD 2024, January 04–07, 2024, Bangalore, India
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1634-8/24/01.
https://doi.org/10.1145/3632410.3633291

1 BACKGROUND

Traditionally learning paradigms have been understood in the context of supervised, semi-supervised and unsupervised learning where the presence of supervision refers to the availability of annotations. However, annotation can be time-consuming, cost-ineffective or erroneous. To circumvent this issue, self-supervised learning provides a framework (of pretraining) where the goal is to learn good feature representations without annotations. These representations can then be used for various downstream tasks in both discriminative and generative forms.

Recently, multi-view self-supervised learning (MVSSL) where the goal is to learn similar representations for the different views of the same data has gained a lot of attention due to its success demonstrated in computer vision [1], speech recognition [2] and other domains. Three main categories of MVSSL are: 1) contrastive loss-based where a suitable similarity metric between the representations of two views from the same example should be maximized and from different examples should be minimized [7], 2) clustering-based where the clusters assignments for one view are to be predicted from the representations of another view [6], 3) distillation-based where a teacher-student model approach is taken for while training [17]. In this tutorial, our focus is on contrastive learning methods across a variety of modalities. In particular, we focus on how different components of contrastive learning vary across modalities in terms of their unique and shared properties.

2 GOALS AND RELEVANCE

Our goal is to review foundations and applied practices in multi-modal contrastive learning for representation learning. We present the major breakthroughs in the following modalities: 1) vision, 2) text, 3) time series, 4) tabular data, and 5) knowledge graphs. To complement the theory, we also provide a code base for the presented work[1].

Self-supervised and semi-supervised learning have long been important for learning representations from big data. While auto-encoder models are successful in some domains, contrastive approaches learning to discriminate encoded representations have exploded in their applicability and power in many domains, especially in those where generative models and loss functions on objects are difficult. However, most available applications are in the image and text domains; other modalities that stand to benefit from CL-based encoders include tabular databases, clinical times series in Electronic Health Records (EHRs), and structured (graph-encoded) data. This strategy is fundamental to several state-of-the-art generative methods. Our tutorial will provide a summary of state-of-the-art solutions and open problems.

[1] https://github.com/sandhyat/ContrastiveLearning_Tutorial

3 GENERAL DESCRIPTION OF TUTORIAL CONTENT

Contrastive learning (CL) is a self-supervised strategy to learn a dense numeric representation of a given feature space, usually by discriminating augmentations of a sample (positive pairs) versus other samples (negative pairs). CL methods differ based on the strategy for generating augmentations, strategy for selecting negative pairs, and loss function. By manipulating the contrastive task, differentiating corrupted versions of the same object from other objects will train embedding functions to represent the major features of the feature space. Because paired representations of the same object in different domains (such as text description of an image) can be used as the positive pair, multimodal representation learning in the same space is possible without significant modifications. Although the concept of contrastive losses was introduced by [18], it was not broadly used by ML community until the InfoNCE loss was defined [34] and found to work well as an instance discriminator [50]. In the computer vision and natural language processing domain, the CLIP model [35] shows the potential of contrastive learning by pre-training representations of images and text that are zero-shot transferred to generative methods. Due to CL's real world popularity in vision and language, the existing survey papers and tutorials (Neurips'21, ECCV'22, NAACL'22) focus on these domains. We will present a comprehensive overview of CL foundations and applications outside of images and text.

The goal of data augmentation is to increase the variability of the data and expose the model to different perspectives for learning similar or dissimilar representations. Data augmentation techniques in CL are modality-specific, and hence each data type has its own particular strategies which may not apply elsewhere. For example, there is obviously workable analogy of "color distortion" from images into text. Common data augmentation techniques include cropping, flipping, rotation, adding noise, and random trimming in sequences. In addition to pre-determined augmentations, one can also create augmentations adaptively [31]. Specific augmentation techniques are also available in time series [11, 23, 57] and tabular data [3, 44, 54]. Learning contrastive representation over graphs has focused on downstream task performance [19] and the need for better augmentations [55, 56].

In new data modalities, the optimal augmentation is difficult to identify; in those cases, the loss function selected becomes more important [33]. Some loss functions are optimized for a particular modality, such as in images [58] and time series [33, 49, 52]. Generally, there are a variety of loss functions that are available for CL, the main ones being 1) contrastive loss that maximizes the similarity between positive pairs, 2) InfoNCE loss that treats CL as a binary classification problem between positive and negative pairs, 3) triplet loss which relies on the relative distance between the examples.

Unlike most supervised learning methods, forming batches in an appropriate way is a key step in the contrastive learning framework. Because the examples selected into a particular step of the loss function can make the underlying discriminative task easier or harder, "hard negative mining" of example pairs which force the embedding to represent more subtle features is a long-standing approach [21, 22, 37, 45]. However, some variations are successful with no negative examples at all [8, 58, 59]. Alternatively, large batch sizes can make likely that at least some negative pairs are "hard" and improve performance [50]. To circumvent the computational complexity and optimization instability of these algorithms, approaches which cluster examples based on an intermediate representation or an auxillary network can accomplish much the same goal [6, 17]. Batch selection is therefore intimately tied to the optimization strategy, which can have a large impact on the quality of representations. For example, large memory banks [50], momentum encoders as in MoCo [20], and simple agreement maximization on large batch sizes for a longer time as in SimCLR [7] were found to work well in images, but may not be as important in tabular data. Most CL methods assume that positive pairs (from augmentations or views) share underlying information and maximize the agreement between the latent representation as training progresses. FactorCL [29] challenges this assumption and demonstrates that there exist cases where task-relevant information can be factorized into shared and unique representation.

The two views in contrastive learning can be augmented views (noisy images) of the same data point [7, 20], two different views (images from different angles) of a source of information[38, 42], or two different sources of information about the same object (image and its text description) [14, 62]. In addition to contrasting in the native input space, similar to VAEs, one can implement contrastive loss on the perturbations of embeddings [26, 56].

Understanding why CL is so successful in some cases is still an active area of research. Some work has elucidated the relationship between mutual information and CL via InfoNCE [43]. Another line of work explores the geometry of embeddings being on a hypersphere [45] and of the modality gap between views [30]. The network architecture and training algorithm are also important variables in the success of CL; this is especially important in less investigated modalities [39]. A popular use case of contrastive learning, DALL·E 2, which generates images from a text prompt, is a prominent example which illustrates the working of the above components together to form a highly performant representation used in a complex downstream task [35, 36].

4 OUTLINE

(1) Part 1: Overview of contrastive learning foundations and applications
  - Contrastive learning in different paradigms, history and popularity [18, 34, 50]
  - Different data augmentation techniques: cutout, adding noise, trimming, etc.
  - Different loss functions: InfoNCE, Triplet loss, NT-Xent loss, etc.
  - Performance evaluation methods: linear methods, transfer learning abilities
  - Theoretical understanding of contrastive learning in terms of inductive biases [39], geometry of embedding [46], mutual information and entropy [16]
  - Understanding CLIP [35] and UnCLIP [36] that led to DALL·E 2 (AI artist)

(2) Part 2: How contrastive learning differs in different data modalities?

- Need of modality-specific augmentations and loss functions.
- Contrastive learning methods for visual representations that rely on
  - optimization parameters (large batch size, more training steps etc[7] , adding a momentum encoder [20]) or using a memory bank [50]
  - using embeddings from nearest neighbour set [12]
- Contrastive learning methods for text representations
  - earliest work modified noise contrastive estimation [32]
  - based on classification task [15, 51], text generation task [27] or pretraining task [41]
- Contrastive learning methods for time series representations
  - using another source of information in addition to the temporal one: temporal and context [13], temporal and spatial [25], temporal and frequency [61]
  - modifying the loss function using the neighbourhood information [52] or expert feature information [33]
  - utilizing data augmentation techniques [23, 49, 57]
- Contrastive learning methods for tabular data representation
  - augmentations created based on addition of noise [3, 54], some use complex models such as transformers [40, 47]
  - modelling as a multiview problem by subsetting the features [44]
- Contrastive learning methods for graph based representations: whether the augmentations are needed or not [55, 56], modelling the real world structures [24]
- Hands-on experience on one method for each data modality

(3) Part 3: What works and why? Further possibilities for CL!
- Decoupling and debiased contrastive loss functions [10, 53]
- (Un)Importance of negative samples in contrastive learning [21, 22, 37, 45, 59, 60]
- Learning in the presence of noisy views by proposing a noise-robust loss function [9], in the presence of noisy labels by selecting confident pairs [28]
- Multi-modal contrastive learning [14, 35, 62]
- Importance of shared and unique information in multimodal CL [29], Connecting pairs of modalities [48]
- Possible attacks [5] and mitigation strategies on multimodal CL [4]
- Non-contrastive (without negative examples) methods: BYOL [17], using stop gradient [8], using correlation metric [58]; clustering to improve the representations[6]
- CL for learning joint representations from two databases for feature matching task
- CL application in healthcare AI and implementation challenges

5 AUDIENCE, REQUISITES AND DURATION

The tutorial duration will be 1.5 hours. The tutorial could be useful for researchers, students, and practitioners interested in adding CL-based methods to experiments and applications. Audience background can be intermediate with knowledge of neural network training (loss functions, gradient descent) and widely used architectures (CNN, RNN, transformers). Some topics will be more valuable to those with prior exposure (graph neural networks).

6 TUTORIAL PRESENTERS

- Dr. Christopher Ryan King, MD, PhD is an assistant professor in the Department of Anesthesiology at Washington University School of Medicine. His PhD in biostatistics from the University of Chicago focused on the analysis of rare variants in whole-genome sequencing. His current research centers on the applied use of artificial intelligence for clinical decision support. Email id: christopherking@wustl.edu
- Sandhya Tripathi, PhD is a postdoctoral research associate in the Department of Anesthesiology at Washington University School of Medicine. Her PhD at IIT Bombay, India focused on developing solutions for label noise (incorrect annotations) problems in classification and interpretable subset selection method using Shapley value. Currently, she has been interested in developing and applying machine learning models for real-world clinical problems, such as combining EHRs across hospitals and evaluating fairness. Email id: sandhyat@wustl.edu

REFERENCES

[1] Philip Bachman, R Devon Hjelm, and William Buchwalter. 2019. Learning representations by maximizing mutual information across views. *Advances in neural information processing systems* 32 (2019).

[2] Alexei Baevski, Yuhao Zhou, Abdelrahman Mohamed, and Michael Auli. 2020. wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in neural information processing systems* 33 (2020), 12449–12460.

[3] Dara Bahri, Heinrich Jiang, Yi Tay, and Donald Metzler. 2021. Scarf: Self-supervised contrastive learning using random feature corruption. *arXiv preprint arXiv:2106.15147* (2021).

[4] Hritik Bansal, Nishad Singhi, Yu Yang, Fan Yin, Aditya Grover, and Kai-Wei Chang. 2023. CleanCLIP: Mitigating Data Poisoning Attacks in Multimodal Contrastive Learning. *arXiv preprint arXiv:2303.03323* (2023).

[5] Nicholas Carlini and Andreas Terzis. 2021. Poisoning and backdooring contrastive learning. *arXiv preprint arXiv:2106.09667* (2021).

[6] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. 2020. Unsupervised learning of visual features by contrasting cluster assignments. *Advances in neural information processing systems* 33 (2020), 9912–9924.

[7] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*. PMLR, 1597–1607.

[8] Xinlei Chen and Kaiming He. 2021. Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 15750–15758.

[9] Ching-Yao Chuang, R Devon Hjelm, Xin Wang, Vibhav Vineet, Neel Joshi, Antonio Torralba, Stefanie Jegelka, and Yale Song. 2022. Robust contrastive learning against noisy views. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16670–16681.

[10] Ching-Yao Chuang, Joshua Robinson, Yen-Chen Lin, Antonio Torralba, and Stefanie Jegelka. 2020. Debiased contrastive learning. *Advances in neural information processing systems* 33 (2020), 8765–8775.

[11] Berken Utku Demirel and Christian Holz. 2023. Finding Order in Chaos: A Novel Data Augmentation Method for Time Series in Contrastive Learning. *arXiv preprint arXiv:2309.13439* (2023).

[12] Debidatta Dwibedi, Yusuf Aytar, Jonathan Tompson, Pierre Sermanet, and Andrew Zisserman. 2021. With a little help from my friends: Nearest-neighbor contrastive learning of visual representations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9588–9597.

[13] Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and Cuntai Guan. 2021. Time-Series Representation Learning via

Temporal and Contextual Contrasting. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, Zhi-Hua Zhou (Ed.). International Joint Conferences on Artificial Intelligence Organization, 2352–2359. https://doi.org/10.24963/ijcai.2021/324 Main Track.
[14] Benjamin Elizalde, Soham Deshmukh, Mahmoud Al Ismail, and Huaming Wang. 2023. Clap learning audio concepts from natural language supervision. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1–5.
[15] Hongchao Fang, Sicheng Wang, Meng Zhou, Jiayuan Ding, and Pengtao Xie. 2020. Cert: Contrastive self-supervised learning for language understanding. *arXiv preprint arXiv:2005.12766* (2020).
[16] Borja Rodrıguez Gálvez, Arno Blaas, Pau Rodríguez, Adam Golinski, Xavier Suau, Jason Ramapuram, Dan Busbridge, and Luca Zappella. 2023. The role of entropy and reconstruction in multi-view self-supervised learning. In *International Conference on Machine Learning*. PMLR, 29143–29160.
[17] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. 2020. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems* 33 (2020), 21271–21284.
[18] Michael Gutmann and Aapo Hyvärinen. 2010. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 297–304.
[19] Kaveh Hassani and Amir Hosein Khasahmadi. 2020. Contrastive Multi-View Representation Learning on Graphs. In *Proceedings of the 37th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 119)*, Hal Daumé III and Aarti Singh (Eds.). PMLR, 4116–4126. https://proceedings.mlr.press/v119/hassani20a.html
[20] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 9729–9738.
[21] Chih-Hui Ho and Nuno Nvasconcelos. 2020. Contrastive learning with adversarial examples. *Advances in Neural Information Processing Systems* 33 (2020), 17081–17093.
[22] Yannis Kalantidis, Mert Bulent Sariyildiz, Noe Pion, Philippe Weinzaepfel, and Diane Larlus. 2020. Hard negative mixing for contrastive learning. *Advances in Neural Information Processing Systems* 33 (2020), 21798–21809.
[23] Konstantinos Kallidromitis, Denis Gudovskiy, Ohama Iku, and Luca Rigazio. 2021. Contrastive neural processes for self-supervised learning. In *Asian Conference on Machine Learning*. PMLR, 594–609.
[24] Thomas Kipf, Elise Van der Pol, and Max Welling. 2019. Contrastive learning of structured world models. *arXiv preprint arXiv:1911.12247* (2019).
[25] Dani Kiyasseh, Tingting Zhu, and David A Clifton. 2021. Clocs: Contrastive learning of cardiac signals across space, time, and patients. In *International Conference on Machine Learning*. PMLR, 5606–5615.
[26] Seanie Lee, Dong Bok Lee, and Sung Ju Hwang. 2020. Contrastive learning with adversarial perturbations for conditional text generation. *arXiv preprint arXiv:2012.07280* (2020).
[27] Seanie Lee, Dong Bok Lee, and Sung Ju Hwang. 2020. Contrastive learning with adversarial perturbations for conditional text generation. *arXiv preprint arXiv:2012.07280* (2020).
[28] Shikun Li, Xiaobo Xia, Shiming Ge, and Tongliang Liu. 2022. Selective-supervised contrastive learning with noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 316–325.
[29] Paul Pu Liang, Zihao Deng, Martin Ma, James Zou, Louis-Philippe Morency, and Ruslan Salakhutdinov. 2023. Factorized Contrastive Learning: Going Beyond Multi-view Redundancy. *arXiv preprint arXiv:2306.05268* (2023).
[30] Victor Weixin Liang, Yuhui Zhang, Yongchan Kwon, Serena Yeung, and James Y Zou. 2022. Mind the gap: Understanding the modality gap in multi-modal contrastive representation learning. *Advances in Neural Information Processing Systems* 35 (2022), 17612–17625.
[31] Dongsheng Luo, Wei Cheng, Yingheng Wang, Dongkuan Xu, Jingchao Ni, Wenchao Yu, Xuchao Zhang, Yanchi Liu, Yuncong Chen, Haifeng Chen, et al. 2023. Time series contrastive learning with information-aware augmentations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 4534–4542.
[32] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems* 26 (2013).
[33] Manuel T Nonnenmacher, Lukas Oldenburg, Ingo Steinwart, and David Reeb. 2022. Utilizing expert features for contrastive learning of time-series representations. In *International Conference on Machine Learning*. PMLR, 16969–16989.
[34] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748* (2018).
[35] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PMLR, 8748–8763.
[36] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125* (2022).
[37] Joshua Robinson, Ching-Yao Chuang, Suvrit Sra, and Stefanie Jegelka. 2020. Contrastive learning with hard negative samples. *arXiv preprint arXiv:2010.04592* (2020).
[38] Shuvendu Roy and Ali Etemad. 2021. Self-supervised contrastive learning of multi-view facial expressions. In *Proceedings of the 2021 International Conference on Multimodal Interaction*. 253–257.
[39] Nikunj Saunshi, Jordan Ash, Surbhi Goel, Dipendra Misra, Cyril Zhang, Sanjeev Arora, Sham Kakade, and Akshay Krishnamurthy. 2022. Understanding contrastive learning requires incorporating inductive biases. In *International Conference on Machine Learning*. PMLR, 19250–19286.
[40] Gowthami Somepalli, Micah Goldblum, Avi Schwarzschild, C Bayan Bruss, and Tom Goldstein. 2021. Saint: Improved neural networks for tabular data via row attention and contrastive pre-training. *arXiv preprint arXiv:2106.01342* (2021).
[41] Yixuan Su, Fangyu Liu, Zaiqiao Meng, Tian Lan, Lei Shu, Ehsan Shareghi, and Nigel Collier. 2021. Tacl: Improving BERT pre-training with token-aware contrastive learning. *arXiv preprint arXiv:2111.04198* (2021).
[42] Yonglong Tian, Dilip Krishnan, and Phillip Isola. 2020. Contrastive multiview coding. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*. Springer, 776–794.
[43] Yao-Hung Hubert Tsai, Yue Wu, Ruslan Salakhutdinov, and Louis-Philippe Morency. 2020. Self-supervised learning from a multi-view perspective. *arXiv preprint arXiv:2006.05576* (2020).
[44] Talip Ucar, Ehsan Hajiramezanali, and Lindsay Edwards. 2021. SubTab: Subsetting Features of Tabular Data for Self-Supervised Representation Learning. In *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (Eds.), Vol. 34. Curran Associates, Inc., 18853–18865. https://proceedings.neurips.cc/paper_files/paper/2021/file/9c8661befae6dbcd08304dbf4dcaf0db-Paper.pdf
[45] Feng Wang and Huaping Liu. 2021. Understanding the behaviour of contrastive loss. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2495–2504.
[46] Tongzhou Wang and Phillip Isola. 2020. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In *International Conference on Machine Learning*. PMLR, 9929–9939.
[47] Zifeng Wang and Jimeng Sun. 2022. Transtab: Learning transferable tabular transformers across tables. *Advances in Neural Information Processing Systems* 35 (2022), 2902–2915.
[48] Zehan Wang, Yang Zhao, Xize Cheng, Haifeng Huang, Jiageng Liu, Li Tang, Linjun Li, Yongqi Wang, Aoxiong Yin, Ziang Zhang, et al. 2023. Connecting Multi-modal Contrastive Representations. *arXiv preprint arXiv:2305.14381* (2023).
[49] Kristoffer Wickstrøm, Michael Kampffmeyer, Karl Øyvind Mikalsen, and Robert Jenssen. 2022. Mixing up contrastive learning: Self-supervised representation learning for time series. *Pattern Recognition Letters* 155 (2022), 54–61.
[50] Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. 2018. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3733–3742.
[51] Seonghyeon Ye, Jiseon Kim, and Alice Oh. 2021. Efficient contrastive learning via novel data augmentation and curriculum learning. *arXiv preprint arXiv:2109.05941* (2021).
[52] Hugo Yèche, Gideon Dresdner, Francesco Locatello, Matthias Hüser, and Gunnar Rätsch. 2021. Neighborhood contrastive learning applied to online patient monitoring. In *International Conference on Machine Learning*. PMLR, 11964–11974.
[53] Chun-Hsiao Yeh, Cheng-Yao Hong, Yen-Chi Hsu, Tyng-Luh Liu, Yubei Chen, and Yann LeCun. 2022. Decoupled contrastive learning. In *European Conference on Computer Vision*. Springer, 668–684.
[54] Jinsung Yoon, Yao Zhang, James Jordon, and Mihaela van der Schaar. 2020. Vime: Extending the success of self-and semi-supervised learning to tabular domain. *Advances in Neural Information Processing Systems* 33 (2020), 11033–11043.
[55] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. 2020. Graph contrastive learning with augmentations. *Advances in neural information processing systems* 33 (2020), 5812–5823.
[56] Junliang Yu, Hongzhi Yin, Xin Xia, Tong Chen, Lizhen Cui, and Quoc Viet Hung Nguyen. 2022. Are graph augmentations necessary? simple graph contrastive learning for recommendation. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*. 1294–1303.
[57] Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and Bixiong Xu. 2022. Ts2vec: Towards universal representation of time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 8980–8987.
[58] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. 2021. Barlow twins: Self-supervised learning via redundancy reduction. In *International Conference on Machine Learning*. PMLR, 12310–12320.
[59] Chaoning Zhang, Kang Zhang, Trung X Pham, Axi Niu, Zhinan Qiao, Chang D Yoo, and In So Kweon. 2022. Dual temperature helps contrastive learning without

many negative samples: Towards understanding and simplifying moco. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14441–14450.

[60] Chaoning Zhang, Kang Zhang, Chenshuang Zhang, Trung X Pham, Chang D Yoo, and In So Kweon. 2022. How does simsiam avoid collapse without negative samples? a unified understanding with self-supervised contrastive learning. *arXiv preprint arXiv:2203.16262* (2022).

[61] Xiang Zhang, Ziyuan Zhao, Theodoros Tsiligkaridis, and Marinka Zitnik. 2022. Self-supervised contrastive pre-training for time series via time-frequency consistency. *Advances in Neural Information Processing Systems* 35 (2022), 3988–4003.

[62] Yuhao Zhang, Hang Jiang, Yasuhide Miura, Christopher D Manning, and Curtis P Langlotz. 2022. Contrastive learning of medical visual representations from paired images and text. In *Machine Learning for Healthcare Conference*. PMLR, 2–25.