
REGCLR: A Self-Supervised Framework for Tabular Representation Learning in the Wild

Weiyao Wang*
Johns Hopkins University
wwang121@jhu.edu

Byung-Hak Kim*,†
AKASA, Inc.
hak.kim@AKASA.com

Varun Ganapathi
AKASA, Inc.
varun@AKASA.com

Abstract

Recent advances in self-supervised learning (SSL) using large models to learn visual representations from natural images are rapidly closing the gap between the results produced by fully supervised learning and those produced by SSL on downstream vision tasks. Inspired by this advancement and primarily motivated by the emergence of tabular and structured document image applications, we investigate which self-supervised pretraining objectives, architectures, and fine-tuning strategies are most effective. To address these questions, we introduce REGCLR, a new self-supervised framework that combines contrastive and regularized methods and is compatible with the standard Vision Transformer [1] architecture. Then, REGCLR is instantiated by integrating masked autoencoders [2] as a representative example of a contrastive method and enhanced Barlow Twins as a representative example of a regularized method with configurable input image augmentations in both branches. Several real-world table recognition scenarios (e.g., extracting tables from document images), ranging from standard Word and Latex documents to even more challenging electronic health records (EHR) computer screen images, have been shown to benefit greatly from the representations learned from this new framework, with detection average-precision (AP) improving relatively by 4.8% for Table, 11.8% for Column, and 11.1% for GUI objects over a previous fully supervised baseline on real-world EHR screen images. More information on the framework and the experimental results can be found in [3].

References

- [1] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *9th International Conference on Learning Representations (ICLR 2021)*, 2021.
- [2] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of Computer Vision and Pattern Recognition (CVPR 2022)*. IEEE, 2022.
- [3] W. Wang*, B.-H. Kim*, and V. Ganapathi. RegCLR: A self-supervised framework for tabular representation learning in the wild. *arXiv preprint arXiv:2211.01165*, 2022.

*equal technical contribution

†project lead