

Case ID:M22-261P^

Published: 6/2/2023

## Inventors

Jae-Sun Seo

Jian Meng

Li Yang

Deliang Fan

## Contact

Physical Sciences Team

# Learning Sparse Features for Self-Supervised Learning with Contrastive Dual Gating

The success of conventional supervised learning relies on large-scale labeled datasets to achieve high accuracy. However, annotating millions of data samples is labor-intensive and time-consuming. This promotes self-supervised learning as an attractive solution with artificial labels being used instead of human-annotated ones for training. Contrastive learning (or its variants) has recently become a promising direction in the self-supervised learning domain, achieving similar performance as supervised learning with minimum fine-tuning.

Despite the labeling efficiency, wide and large networks are required to achieve high accuracy which incurs a high amount of computation and hinders the pragmatic merit of self-supervised learning. To effectively reduce the computation of insignificant features or channels, recent dynamic pruning algorithms for supervised learning employed auxiliary saliency predictors. However, the saliency predictors cannot be easily trained when that are naively applied to contrastive learning from scratch. There is a need for a dynamic pruning algorithm that skips the uninformative features during contrastive learning without hurting the trainability of the networks.

Researchers at Arizona State University (ASU) have developed a dynamic pruning algorithm designed for contrastive self-supervised learning. This algorithm exploits spatial redundancy by using a spatial gating function with full awareness of the saliency difference between contrastive branches. The ASU algorithm learns the sparse features in both contrastive branches during the unsupervised learning process. Furthermore, it can exploit the sparse features in both structured and unstructured manner. Aided by the efficient and optimized sparsification, the ASU algorithm achieves high floating-point operations reduction and high inference accuracy, without any auxiliary predictors.

This algorithm has been verified on multiple benchmark datasets and various SSL frameworks. For example, the algorithm was evaluated for RESNET models across multiple datasets and achieved up to 2.25x and 1.65x computation reduction for CIFAR-10/-100 and ImageNet-100 datasets, respectively. Compared to other dynamic pruning algorithms for self-supervised learning, the ASU algorithm achieves up to 15% accuracy improvement for CIFAR-10 dataset with higher computation reduction.

Related publication: [Contrastive Dual Gating: Learning Sparse Features with Contrastive Learning](#)

Potential Applications:

- Training algorithm for neural networks (NNs)

#### Benefits and Advantages:

- Algorithm skips the uninformative features during contrastive learning without hurting trainability of the networks
- Algorithm achieves high floating-point operations reduction and high inference accuracy without any auxiliary predictors
- Algorithm successfully demonstrated with ResNet models for CIFAR-10, CIFAR-100, and ImageNet-100 datasets