**Contrastive Learning with Boosted Memorization**

In the fields of computer vision and natural language processing, self-supervised learning has had tremendous success by learning the robust representation for downstream tasks. Most of the existing research on self-supervised long-tailed learning is from a loss or model perspective. The loss perspective approach uses loss reweighting to emphasize tail samples during training, such as the focal loss in hard example mining or SAM using the sharpness of the loss surface. The precision of the tail sample finding, however, is sensitive to and constrained by the efficacy of these procedures. In order to help the model better capture the semantics of the tail samples, the latter generally uses specialized model designs like the divide-and-contrast ensemble or self-damaged-contrast via pruning. These designs, which are typically black boxes to understand the possible working dynamics for further improvement, call for the use of empirical heuristics. A study of self-supervised long-tailed learning from a data perspective is suggested in this research. The framework was developed in response to the memorizing impact that deep neural networks have on data, where the simple patterns are typically retained before the complex ones. As an alternative to loss reweighting or model redesign, data augmentation is a very effective way to improve self-supervised long-tailed learning by creating information discrepancy between two viewpoints. A novel Boosted Contrastive Learning method from the data perspective is proposed. A momentum loss to take the cues from the DNNs' memorizing effect and anchor most tail sample possibilities is proposed. The momentum loss is then used to create separate information discrepancies for the head and tail samples in order to drive an instance-wise augmentation.

BCL can be conceptualized as a contrastive learning approach that adaptively assigns the necessary augmentation strength for each individual sample based on feedback from the memorization hints. The memorizing effect of DNNs is controlled via boosted contractive learning, which makes use of a momentum loss proxy. Tail samples will receive higher levels of augmentation because they typically learn more slowly. The model is then instructed to draw out more data from the enhanced views of tail samples for better generalization. β is a hyper-parameter used to control the degree that is smoothed by the historical losses. A set of the momentum losses for each sample is obtained and normalized. The intensity of memorization effect is used as an indicator to control the occurrence and strength of the augmentation. K types of augmentations are randomly selected from RandAugment and each augmentation is applied with the probability Mi (intensity of memorization effect) and the strength respectively. The memorization-boosted augmentation is formulated. Up until its momentum loss declines, BCL continuously relies on memorization-boosted augmentation to emphasize the training samples to which DNNs exhibit the poor memorization effect. The model can be adaptively motivated to learn "residual" information contained in tail samples by iteratively optimizing the model and creating the memorization enhanced information discrepancy.

Due to BCL's lack of explicit model structure requirements, it has been compatible with a variety of self-supervised learning techniques in recent years. BCL captures the memorization cues to drive the creation of the information discrepancy for the implicit re-weighting but does not alter the loss directly. This literature is the first to investigate self-supervised long-tailed learning from the data perspective, which uses the DNN memorization impact on data and the augmentation efficiency in self-supervised learning, in contrast to earlier works in the loss and model views. The suggested approach of boosted contrastive learning creates a momentum loss to capture memorizing effect cues and drive instance-wise augmentation to dynamically preserve learning of head samples and improve learning of tail samples. To the present self-supervised approaches on long-tailed data, the suggested BCL is orthogonal. Extensive tests on a variety of benchmark datasets show that BCL performs better than other algorithms.

**Efficiently Identifying Task Groupings for Multi-Task Learning**

Through multi-task learning, one task's knowledge can be used to help other tasks learn. Despite this capability, training all tasks in one model naively frequently leads to performance degradation, and it might be prohibitively expensive to thoroughly search through all possible task groups. While novel multi-task learning optimization algorithms have been created recently, the issue of selecting which tasks should be trained jointly is a complex and understudied one. Also, multi-task learning depends on numerous non-trivial decisions, such as dataset features, model architecture, hyperparameters, capacity, and convergence. A subset of options evaluated by approximation task grouping algorithms may become excessively expensive and time-consuming to analyze. The goal of this research is to develop an effective framework for choosing task groupings without losing performance. The study suggests measuring the degree to which one task's gradient update would affect another task's loss by training all tasks concurrently in a single multi-task network. To optimize the affinity onto each task, tasks are then grouped together once this per-step quantity is averaged across training. The suggested method is applicable to any paradigm in which shared parameters are updated with respect to numerous losses and makes no assumptions about the model architecture.

In this work, a method to group tasks by looking at how the gradient of one task would affect the loss of another activity is proposed. The paper discusses how to choose which tasks in multi-task neural networks that should be trained together using a technique called Task Affinity Groupings (TAG). TAG is a wonderful method for selecting the tasks that should be trained concurrently. The approach examines how tasks interact during training, specifically how changing the model's parameters during training on one task will affect the network's other tasks' loss values. In order to maximize performance across all tasks, the technique tries to break a set of tasks into smaller subgroups. To achieve this, it trains all tasks simultaneously in a single multi-task model and assesses how much the loss of one task will be impacted by a gradient update to the model's parameters. The term "inter-task affinity" refers to this amount. According to the experimental results, choosing task groups that optimize inter-task affinity highly corresponds with overall model performance. To better understand the training dynamics of multi-task neural networks, TAG uses a similar method. In specifically, it modifies the parameters of the model with respect to just one task, considers how this modification will impact the other tasks in the multi-task neural network, and then reverses this modification. To learn more about how each activity in the network might interact with every other task, this procedure is then repeated for every other task. The model's common parameters are then updated with respect to each task in the network as part of regular training. The inter-task affinity onto a specific task is approximately determined for task groupings of three or more tasks by averaging the pairwise affinities onto that task.

The main contribution of the proposed work is that an inter-task affinity measure could be applied to systematically and effectively group tasks for multi-task learning. Theoretical research demonstrates that, in the convex setting, under mild conditions, task grouping by maximizing inter-task affinity will outperform all other task groupings. The empirical study further reveals that this strategy outperforms training all tasks separately, training all tasks jointly and is competitive with a state-of-the-art task grouping method while reducing runtime by more than an order of magnitude. The results are backed by comprehensive analysis, which indicates that inter-task affinity scores can identify nearly ideal auxiliary tasks and measure generalization ability among tasks in an implicit manner.